The role of Phase II detoxification genes towards modulating risk for lung cancer and its association with clinical outcomes

A

Thesis submitted

in the partial fulfilment of the requirements for the award of degree of

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY



THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Submitted by
HARLEEN KAUR WALIA
(Reg. No- 901700001)

Under the supervision of
Dr. SIDDHARTH SHARMA
Associate Professor

Department of Biotechnology

Thapar Institute of Engineering and Technology, Patiala, 147004, Punjab

July, 2022

Dedicated to my family...

CERTIFICATE

Certified that the thesis entitled "**The role of Phase II detoxification genes towards modulating risk for lung cancer and its association with clinical outcomes**" submitted by Ms. Harleen Kaur Walia, Reg. no. 901700001 in the partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy in the Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab is a record of candidate's own independent and original research work carried out by herself under my supervision and guidance. The material embodied in this thesis has not been submitted in part or full to any other University or institute for the award of any degree.



Dr. Siddharth Sharma

(Associate Professor)

Supervisor

Department of Biotechnology

Thapar Institute of Engineering & Technology

Patiala, Punjab

CANDIDATE'S DECLARATION

I hereby declare that the work being presented in the thesis entitled **"The role of Phase II detoxification genes towards modulating risk for lung cancer and its association with clinical outcomes"** in partial fulfilment of the requirements for the award of degree of Doctor of Philosphy, Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab is an authentic record of my work during the period from July 2017 to October 2021, under the conception and supervision of Dr. Siddharth Sharma, Assoc. Professor, Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab. This report has not been submitted for the award of any other degree or certificate in this or any other university.

Harleen.

Place : Patiala, Punjab
Date : 01/08/2022

Harleen Kaur Walia
(Reg no.: 901700001)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge

Siddh

Date : 01/08/2022

Dr. Siddharth Sharma
Assoc. Professor / Supervisor
Department of Biotechnology
Thapar Institute of Engineering & Technology
Patiala, Punjab

ACKNOWLEDGMENT

In pursuit of this academic endeavour, I feel that I have been singularly fortunate because inspiration, guidance, direction, cooperation, love and care- all came in my way in abundance and it seems almost an impossible task for me to acknowledge the same in adequate term.

With profound sense of indebtedness and heartfelt gratitude, I thank **Dr. Siddharth Sharma**, Associate Professor, Department of Biotechnology (DBT), Thapar Institute of Engineering and technology (TIET), Patiala, Punjab (Pb) who guided and inspired me in pursuance of this work. His association with this endeavour of mine will remain a beacon light to me throughout my life. I sincerely thank him for making me a better person, guiding me and improving silliest of my mistakes. His deep insights helped me at various stages of my research. I also remain indebted for him understanding and support me during the times when I was really down. His determined and enthusiastic nature has always inspired me to be consistent in my work. Words would be less to express my thanks to him.

My sincere gratitude to visionary director of TU, **Prof. Dr. Prakash Gopalan**, for providing advanced infrastructure and excellent research environment. I thank **Dr. Inderveer Chana**, Dean, Student affairs and **Prof. Ajay Batish**, Dean, Partnership and accreditations, TIET, Patiala for their precious guidance and constant motivation.

Constructive suggestions and critical remarks of **Dr. M.S. Reddy**, Head, DBT, **Dr. M Vasundhara**, Assistant Professor, DBT, **Dr. Manoj Baranwal**, Assistant Professor, DBT, and **Dr. Kamaldeep Paul**, Associate Professor, School of Chemistry and Biochemistry towards amelioration of this work has been of immense help in successfully achieving the objectives. I am highly obliged to them for their suggestions and constant encouragement.

A special thanks to Mr. Ram Naval, Mr. Surinder Pal, Mr. Prabhat and other lab staff members of DBT, TIET without whom's help the experiments would not be have been possible.

I am especially indebted to my seniors **Dr. Charu Bahl** and **Dr. Shweta Lawania** for their constant love, cooperation and help. I would also like to thank my lab mates Mr. Sidhartha Singh, and **Ms. Parul Sharma** who have helped in my work. A special thanks to Parul with whom I share a good repertoire and a deep mutual understanding.

I am also indebted to my friends Davinder Maan, Yadveer Kaur, Dr. Vagish Dwibedi, Sheetal, Ramandeep Kaur, Gurleen for being there to listen when i needed an ear, who became my strength during tough times and keeping faith in me.

I like to express my deep joy to work with all the M.Sc and M.tech students who gave a chance to learn more about my field. Their presence kept my enthusiasm high and my passion for this field grew during all these years. Prabhjot Kaur, Heena, Prerna, Isha, who were always eager to learn more and shared a good bond with them.

Nonetheless, friends and well-wishers know me my thick and thin and there are no ways to express any form of thanking to them. Words cannot express the feelings I have for my parents and my in-laws for their constant unconditional support - both emotionally and financially. I would not be here if it not for them.

There are times when you are at verge to fall into pieces but the presence of one single inspiration can always keep you going such was my husband **Lt Cdr Sehaj Shah Singh**. He has been a constant source of strength and inspiration. I can honestly say that it was only his determination and constant encouragement during thesis writing that ultimately made it possible for me to see this thesis through to the end.

Siblings **Dr. Brahmjot Singh Walia, Rehatjit Kaur, Jaskirat Singh, Japneet & Harjap** are always a source of motivation for whom I want to set the example. My love for them has been my one of the driving force to perform well and give them happiness. I would like to thank my grandfather, **Late Satwant Singh Walia,** who has blessed all of us and always taught us the right virtues of living.

Thanks to almighty for bestowing blessings on me.

Date: 01/08/2022

Place: Patiala

Harleen Kaur Walia

LIST OF PUBLICATIONS

1) Walia, H. K., Verma, U., Singh, N., & Sharma, S. (2019). Polymorphisms in GSTM1 and GSTT1 influence the response and treatment outcome in lung cancer patients treated with platinum-based chemotherapy. *British journal of biomedical science*, *76*(4), 198-200. https://doi.org/10.1080/09674845.2019.1634784
2) Walia, H. K., Singh, N., & Sharma, S. (2021). Association of NQO1Pro187Ser polymorphism with clinical outcomes and survival of lung cancer patients treated with platinum chemotherapy. *Personalized Medicine*, (0). https://doi.org/10.2217/pme-2020-0119
3) Walia, H. K., Singh, N., & Sharma, S. (2021). Genetic polymorphism of Arg213His variant in the SULT1A1 gene is associated with reduced susceptibility to lung cancer in North Indian population. *Xenobiotica*, (just-accepted), 1-28. https://doi.org/10.1080/00498254.2021.1963008
4) Walia, H. K., Singh, N., & Sharma, S. (2021). Polymorphism in Epoxide Hydrolase gene in relation to lung cancer susceptibility and clinical outcomes in North Indians. *Future Oncology*. 10.2217/fon-2021-0412
5) Walia, H. K., Singh, N., & Sharma, S. (2021). MTHFR polymorphism as a predictive biomarker for Gastrointestinal and Hematological toxicity in North Indian adenocarcinoma patients after treatment with pemetrexed and platinum-based doublet chemotherapy. *Journal of Chemotherapy*. 10.1080/1120009X.2021.1997008
6) Walia, H. K., Singh, N., & Sharma, S. (2021). GSTP1 Ile 105 Val polymorphism among North Indian lung cancer patients treated using monotherapy and poly-pharmacy. *Human & Experimental Toxicology*, *40*(12_suppl), S739-S752. https://doi.org/10.1177%2F09603271211059496
7) Walia, H. K., Singh, N., & Sharma, S. (2021). Association of *NAT-2* gene polymorphisms toward lung cancer susceptibility and prognosis in North Indian patients treated with platinum-based chemotherapy. *Pharmacogenomics*, *23*(2), 97-118. 10.2217/pgs-2021-0080
8) Walia, H. K., Sharma, P., Singh, N., & Sharma, S. (2022). Immunotherapy in Small Cell Lung Cancer Treatment: a Promising Headway for Future Perspective. *Current Treatment Options in Oncology*, 1-27. 10.1007/s11864-022-00949-1
9) Walia, H., Sharma, P., Singh, N., & Sharma, S. (2022). Predictive role of polymorphic variants of phase II drug metabolising enzyme in modulating toxicity in North Indian lung cancer patients undergoing chemotherapy. *Xenobiotica*, (just-accepted), 1-22. 10.1080/00498254.2022.2069527

WORKSHOPS AND CONFERENCES

- Attended the 3rd LUNG MCMC (Multidisciplinary Cancer Management Course) at PGIMER, Chandigarh (March 24-25, 2018).
- 3 days' national workshop on "COMPUTATIONAL DRUG DESIGN USING MOLECULAR DOCKING AND VIRTUAL SCREENING" sponsored by Schrodinger (26-29 April 2018) at TIET Patiala, Punjab

Abstract

Background:

Lung cancer is the leading cause of cancer mortality globally, and the key risk factors are smoking and occupational exposure. The presence of polymorphic genes affecting the levels of metabolic activation and carcinogen detoxification influences lung cancer susceptibility.

Objectives

To evaluate the role of single nucleotide polymorphic variants of Phase II detoxification genes *NQO1* (rs1800566), *SULT1A1* (rs9282861), *EPHX1* (rs1051740, rs2234922), *NAT2* (rs1799929, rs1799930, rs1799931, rs1208), *MTHFR* (rs1801133, rs1801131) and *GSTP1* (rs1695).

Methodology

Genotyping of genomic DNA was carried out using PCR-RFLP (polymerase chain reaction-restriction fragment length polymorphism) for each polymorphic site under study. Total number of subjects genotyped in case of *NQO1, SULT1A1, EPHX1,* and *NAT2* were 1100 (550 cases, 550 controls); in *MTHFR* were 253 cases and in case of *GSTP1* 682 cases, respectively. Following this association study, logistic regression was used to get the adjusted odds ratio and significance. To find the possible connection between interacting SNP-SNP and gene-gene interactions, data mining analysis was performed that included both Multi-dimensionality reduction (MDR) and Classification and Regression tree (CART) analysis. Overall survival analysis was carried out using Kaplan-Meier survival analysis and Cox-regression analysis, which yielded the adjusted hazards ratio.

Results

The *TT* genotype for *NQO1 609C>T* polymorphism exhibited a good association with LC risk in regards of clinical-pathological parameters. A 3.5 fold higher odds in subjects with stageIII (p=**0.0006**),5 fold higher probability of lymph-node invasion(p=**0.007**) and an odds of less than one in case of metastasis(p=**0.0028**). Furthermore, 473 subjects were analyzed in regards of overall survival, wherein the *TT* genotype exhibited a better OS than *CC* genotype (9.43vs.8.13months). Whereas, none of the chemotherapeutic regimen exhibited a significant association with OS, except for the subjects possessing *TT* genotype and administered with paclitaxel+cis/carboplatin, these exhibited a very poor survival(3.57vs.12.20months).In such patients a significant higher death rate was obtained using Cox-proportional analysis(HR'=7.95;p=**0.0098**). For *SULT1A1 Arg*^{213}*His* polymorphism the variant genotype exhibited no association with LC risk, even after stratification on basis of histological subtypes. The male

LC patients carrying the variant *His*^{213} allele (**p=0.02**) did not exhibited an increased risk towards LC. The smokers harboring either the *Arg/His* (**p=0.019**) or *Arg/His+His/His* (**p=0.01**) genotype, exhibit a reduced risk towards LC. Furthermore, the LC subjects who were heavy smokers and harboring the *Arg/His* (**p=0.01**) genotype did not show a genetic predisposition towards LC susceptibility. The subjects who smoked pack years of above 40 and carrying the *His/His* (**p=0.036**) and *Arg/His+His/His* (**p=0.015**) genotype were found to pose no effect on the risk of LC. Furthermore, 473 subjects were analyzed in regards of overall survival, wherein the *His/His* genotype exhibited better OS than *Arg/Arg* genotype (11.30 vs.8.07months). Variant genotype (*His/His*) of *Tyr*^{113}*His* polymorphism exhibited a two-fold increased odds of lymph node invasion (p=**0.04**). In *EPHX1 Tyr*^{113}*His* polymorphism, the *Tyr/His* (p=**0.01**) and *Tyr/His+His/His* (p=**0.01**) genotype was a risk factor for smokers. Further, subjects harboring *His/His* (p=**0.04**) and *Tyr/His* (p=**0.04**) genotype in case of light and heavy smokers, respectively were found to be associated with an increased risk of LC. Further, the *Tyr/His* (p=**0.03**) and *Tyr/His+His/His* (p=**0.03**) genotype were found to be associated with risk of LC in male subjects. The subjects carrying the combined genotype for *His*^{139}*Arg* (*His/Arg+Arg/Arg*) exhibited a better OS using Cox-proportional hazards (9.80 *vs*.7.60months, p=**0.04**) and heterozygous genotype exhibited a better OS in case of SCLC(11.30 vs.6.73months,log rank test p=**0.02**). Further, it was found that heterozygous genotype of *EPHX1 Tyr*^{113}*His* gene had a longer survival time in those patients who were receiving cisplatin/carboplatin along with irinotecan(11.30 *vs*.7.23months,log rank test p=**0.007**). The *NAT2 857G>A* polymorphism exhibited a risk towards lung cancer (p=**0.005**). Whereas, variant alleles for the *481C>T* polymorphism had a decreased risk for LC (**p**=**0.0003**). Further, *NAT2*4/*7* polymorphism conferred a positive association between genotype and ADCC (p=**0.001)** and *481C>T* polymorphism had a decreased risk for SQCC (AOR=0.39**, p=0.0006**) and SCLC (**p=0.001**) subjects. The smokers carrying mutant genotype for the *481C>T* polymorphism had a decreased risk towards LC (**p<0.0001**) even in light (**p=0.002**) as well as heavy smokers **(p=0.001**). In case of female, 2.59 folds and 3.66 folds increased risk of LC development was observed in subjects with intermediate and slow acetylator for the *857G>A* polymorphism. Whereas, in case of males this polymorphism depicts a reduced risk for lung cancer. On the other hand, *803A>G* depicted a 2.82 folds risk of lung cancer in case of female subjects who were slow acetylators. Our study exhibits a significant difference in the overall haplotype distribution between cases and controls. In our study overall, (*857G>A, 481C>T, 803A>G*) was found to be best model, but was not significant using MDR. Considering the CART results *481C>T* polymorphism came out to be the most significant factor in determining the LC risk. For the *803A>G* polymorphism a 3-fold odds of lymph node invasion were observed for mutant genotype, the recessive model exhibited an odd of 2.8. *590G>A* appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy as the survival time for such patients was better. The patients who were carrying both the mutant

alleles (*C/C*) for the *MTHFR* (*1298A>C*) polymorphism showed a higher trend of median survival time (MST) as compared to the patients bearing the wild type genotype (*A/A*) (MST=13.93 vs. 7.97, p=0.12). On basis of toxicity profiling, we observed that lung cancer patients with the *AT* (heterozygous) genotype for *MTHFR A1298C* polymorphism had a 12-fold risk of diarrhoea (AOR=12.54, 95% CI=1.54-101.86, **p=0.018**). The patients with the heterozygous genotype (*CT*) of the *MTHFR C677T* polymorphism had a 5.34-fold increased risk of developing neutropenia toxicity (AOR=5.34, 95% CI= 1.49-19.06, **p=0.009**). On basis of toxicity profiling, we observed that lung cancer patients with mutant genotype of *GSTP1 $Ile^{105}Val$* had an increased risk of leukopenia (OR=2.41; 95%CI=1.39-4.18, **p=0.001**). Similarly, data also suggested that that patients with the heterozygous genotype (*Ile/Val*) had a 2.14-fold increased risk of developing severe anemia (OR=2.14, 95% C.I.=0.97-4.62, **p=0.03**). Further our data also showed that in SCLC patients polymorphism of *GSTP1* was associated with thrombocytopenia ($\chi 2$ test=7.32, p=0.02). The five-factor model *(NQO1 $P^{187}S$, EPHX1 $Y^{113}H$, NAT2*6 $R^{197}Q$, NAT2*5B $L^{161}L$, NAT2*5C $K^{268}R$)* was the overall best model for the interaction of selected Phase II polymorphisms. The CART results demonstrates that *NAT2 481C>T* caused the initial split and genotypic combination of *NAT803A>G(M)/NAT857G>A(M)/EPHX1337T>C(W)/NQO1609C>T(M)/EPHX1415A>G(W)/ NAT590 G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)* conferred the highest risk towards lung cancer (OR=24.35, 95% C.I.=2.94-201.26, p= **0.003**). In terms of analyzing toxicity profile, the mutant genotype for *NQO1* exhibits reduced risk for gastrointestinal toxicity, including constipation and anorexia. Similarly, the heterozygous and combined genotype for *SULT1A1* polymorphism had decreased risk for developing neutropenia. Further, the heterozygous and combined genotype for *NAT2*6* polymorphism was found to be associated with an increased risk for leukopenia and anemia, *i.e.*, hematological toxicity as well as neutropenia. At the same time, *NAT2*5C* was associated with a reduced risk of developing hematological toxicity, especially anemia. The heterozygous genotype for *NAT2*7* was found to be associated with 2 folds increased risk of thrombocytopenia but decreased risk of anemia (hematological toxicities) as well as decreased risk of alopecia and anorexia.

Conclusion

These results suggest that *NQO1* variant genotype may modulate LC risk, especially those with stage III LC and as poor prognostic factor. The *SULT1A1 $Arg^{213}His$* polymorphism provides no epidemiologic evidence of genetic predisposition towards LC risk in relation to tobacco smoking. The *EPHX1 $Tyr^{113}His$* polymorphism is associated with LC risk in smokers and appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy and might increase the likelihood of lymph node metastasis. Whereas, the *EPHX1 $His^{139}Arg$* exhibit better survival especially, in SCLC patients. The *NAT2* variant genotype for *590G>A* and *803A>G* was not found to modulate risk towards LC, but *857G>A*

polymorphism exhibited a risk towards LC and *481C>T* polymorphism had a decreased risk for LC. *NAT2 590G>A* appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy and *480C>T* came out to be significant factor using CART. Our results suggest that *MTHFR* polymorphism is associated with severe hematological toxicity like leukopenia and also with gastrointestinal toxicity. Thus, evidence suggests that *MTHFR* polymorphism might have an impact on the development of pemetrexed and platinum related toxicities but not as a clinical predictor of efficiency or as a prognostic marker. Our results suggest that *GSTP1 $Ile^{105}Val$* polymorphism could act as a predictive biomarker for certain hematological toxicity, like leukopenia and anaemia, but not thrombocytopenia or neutropenia.

Table of Contents

Title **Page No.**

Abstract

List of publications

Table of contents

List of figures

List of Tables

List of Notations

List of abbreviations

Chapter 1 Introduction……1

1.1 Introduction……1

Chapter 2 Review of Literature……..9
2.1 Lung cancer……..9
2.2 TNM staging…….....11
2.3 Histopathological classification of lung cancer…….....13
2.3.1 Squamous cell carcinoma (SQCC)……...13
2.3.2 Adenocarcinoma (ADCC)…….....14
2.3.3 Large cell carcinoma (LCC)……15
2.3.4 Small cell lung cancer (SCLC)……16
2.4 Risk factors for lung cancer…….....17
2.4.1 Smoking…….....17
2.4.2 Passive smoking…….....20
2.4.3 Diet and alcohol…….....21
2.4.4 Occupational hazards…….....21
2.4.4.1 Asbestos…….....21
2.4.4.2 Radon gas exposure…….....22
2.4.4.3 Metals and other occupation exposures…….....23
2.4.4.4 Indoor and outdoor air pollutants……..23
2.5 Global epidemiology of Lung cancer…….23
2.6 Epidemiology of Lung cancer in India…….....25
2.7 Lung cancer pathogenesis at the molecular level…….....27

2.7.1 Proto-oncogene activation……27
2.7.2 Tumor suppressor genes inactivation……29
2.7.3 Escaping the programmed cell death……30
2.8 Diagnosis of lung cancer…….....32
2.8.1 Screening and early detection of lung cancer……....33
2.8.2 Lung cancer oncopanels……34
2.9 Treatment of lung cancer…….....36
2.9.1 Platinum-based chemotherapy…….....37
2.9.1.1 Docetaxel + cisplatin/ carboplatin……..…….....37
2.9.1.2 Irinotecan + cisplatin/ carboplatin……38
2.9.1.3 Pemetrexed + cisplatin/ carboplatin……...38
2.9.2 Targeted therapy and immunotherapy……...39
2.10 Genome wide Association Study (GWAS)……...42
2.11 Detoxification enzymes and Lung cancer……..42
2.11.1 Phase II metabolizing enzymes……..45
2.11.1.1 NAD(P)H:Quinone oxidoreductase 1 (NQO1)……..47
2.11.1.1.1 *NQO1* gene regulation……48
2.11.1.1.2 Structure of NQO1……..49
2.11.1.1.3 Catalytic cycle and Substrate specificity of *NQO1*……50
2.11.1.2 Sulfotransferases (SULTs)……51
2.11.1.2.1 *SULT1A1* gene regulation……52
2.11.1.2.2 Structure of SULT1A1 protein……53
2.11.1.2.3 Reaction catalysis and substrate specificity of SULT1A1……...54
2.11.1.3 Epoxide Hydrolase (EPHX)……...55
2.11.1.3.1 *EPHX1* gene regulation……....56
2.11.1.3.2 Structure of EPHX protein……...57
2.11.1.3.3 Reaction catalysis and substrate specificity of EPHX1……...58
2.11.1.4 *N-acetyltransferase (NAT)*……60
2.11.1.4.1 *NAT* gene regulation……61
2.11.1.4.2 Structure of NAT2 protein……...62
2.11.1.4.3 Reaction catalysis and substrate specificity of NAT2……63
2.11.1.5 *Methylenetetrahydrofolate reductase (MTHFR)*……...65
2.11.1.5.1 *MTHFR* gene regulation……..65
2.11.1.5.2 Structure of MTHFR protein……...66
2.11.1.5.3 Reaction catalysis and substrate specificity of MTHFR……68
2.11.1.6 Glutathione S-transferase……..70
2.11.1.6.1 *GSTP1* gene regulation……71
2.11.1.6.2 Structure of GST protein……73
2.11.1.6.3 Reaction catalysis and substrate specificity of GST……...75

2.12 Genetic variability in Phase II detoxification genes and its correlation with lung cancer susceptibility……76
2.12.1 Genetic polymorphisms in *NQO1* gene in association with lung cancer risk and overall survival of LC patients…….....76
2.12.2 Genetic polymorphisms in *SULT1A1* gene in association with lung cancer risk and overall survival of LC patients…….80
2.12.3 Genetic polymorphisms in *EPHX1* gene in association with lung cancer risk and overall survival of LC patients…….82
2.12.4 Genetic polymorphisms in *NAT2* gene in association with lung cancer risk and overall survival of LC patients…….87
2.12.5 Genetic polymorphisms in *GSTP1* gene in association for clinical outcomes in lung cancer and toxicity after chemotherapy……...92
2.12.6 Genetic polymorphisms in *MTHFR* gene in association for clinical outcomes in lung cancer and toxicity after chemotherapy……95

Chapter 3 Gaps in study and objectives……99

3.1 Gaps in study…….....99
3.2 Objectives of the study……...100

Chapter 4 Materials and methods……101

4.1 Design of the study……101
4.1.1. Chemotherapeutic regimen…….....103
4.1.2. Follow-up and response assessment……....103
4.2 DNA extraction……104
4.2.1. Genomic DNA extraction……104
4.2.2. Qualitative and quantitative estimation of genomic DNA……..105
4.3 Genotyping of polymorphic sites *NQO1*, *SULT1A1*, *EPHX1*, *NAT2*, *MTHFR*, and *GSTP1* gene……..106
4.3.1 Genotyping of *NQO1* polymorphism by PCR-RFLP……..106
4.3.2 Genotyping of *SULT1A1* polymorphism by PCR-RFLP……107
4.3.3 Genotyping of *EPHX1* polymorphisms by PCR-RFLP……...108
4.3.4 Genotyping of *NAT2* polymorphisms by PCR-RFLP……..109
4.3.5 Genotyping of *MTHFR* polymorphism by PCR-RFLP……...110
4.3.6 Genotyping of *GSTP1* polymorphism by PCR-RFLP……111
4.4 Statistical analysis……..112

Chapter 5 Results
5.1 Genomic DNA isolation……115
5.2 PCR amplification……115

5.3 Genotyping of polymorphic variants by RFLP……116
5.3.1 Genotyping of *NQO1* genetic variants……116
5.3.2 Genotyping of *SULT1A1* genetic variants……...117
5.3.3 Genotyping of *EPHX1* genetic variant……117
5.3.4 Genotyping of *NAT2* genetic variants……118
5.3.5 Genotyping of *GSTP1* genetic variant……120
5.3.6 Genotyping of *MTHFR* genetic variants……..120
5.4 Epidemiological statistics of North Indian population under study……..121
5.4.1 Regional Distribution of Lung cancer patients……....121
5.4.2 Distribution of study subjects based on smoking……123
5.4.3 Distribution of Study subjects based on age……....124
5.4.4 Distribution of lung cancer patients based on stage of tumor and TNM…….125
5.4.5 Distribution of lung cancer patients based on histology of the tumor……126
5.4.6 Distribution of lung cancer patients based o performance status……127
5.4.7 Distribution of lung cancer patients based on doublet platinum-based chemotherapeutic regimen……128
5.5 Demographic characteristics of the study subjects for *NQO1*, *SULT1A1*, *EPHX1*, and *NAT2* gene ……..130
5.6 Association of PhaseII detoxification genes towards lung cancer susceptibility…….132
5.6.1 Role of *NQO1* polymorphism towards lung cancer susceptibility……..132
5.6.1.1.Genotypic and allelic frequencies of NQO1 polymorphic variant among cases and controls and their role in modifying the susceptibility towards lung cancer…….....132
5.6.1.2.Genetic and allelic distribution of NQO1 based on histopathological subtypes……..133
5.6.1.3.The risk associated with NQO1 polymorphism stratified by smoking…...……135
5.6.1.4.Relation between NQO1 genotypic distribution and clinic-pathologic characteristics…….137
5.6.2 Statistical Analysis of *SULT1A1 Arg^{213}His* polymorphism towards lung cancer susceptibility……139
5.6.2.1 Overall association of *SULT1A1* polymorphic variant with lung cancer susceptibility……...139
5.6.2.2 Genetic and allelic distribution of *SULT1A1 Arg^{213}His* polymorphism based on histological subtypes……141
5.6.2.2.1 Role of SULT1A1 polymorphism in different histological subtypes of lung cancer based on gender…...……142
5.6.2.3 Stratification analysis by gender……..143
5.6.2.4 Interactive effect of smoking on the SULT1A1 polymorphism and their association with lung cancer risk……147
5.6.2.5 Association of SULT1A1 genotype with clinical-pathological feature...151

5.6.3 Statistical analysis for polymorphic sites of *EPHX1* gene towards lung cancer susceptibility……....154
5.6.3.1 Overall association of *EPHX1* polymorphic variants with lung cancer susceptibility……154
5.6.3.2 Association of *EPHX1* polymorphisms (*Tyr*^{113}*His* and *His*^{139}*Arg*) with lung cancer risk based on histological subtypes……156
5.6.3.3 Effect of smoking status on association of EPHX1 polymorphism with lung cancer risk……...158
5.6.3.4 Stratification analysis of *EPHX1* polymorphic variants with lung cancer risk based on gender……..……..162
5.6.3.5 Combinatorial analysis of the risk associated with different polymorphic sites of *EPHX1* gene……..164
5.6.3.6 Variant forms of *EPHX1* as surrogate of enzyme activity……...166
5.6.3.7 Haplotype analysis of *EPHX1* variants……170
5.6.3.8 Relationship between *EPHX1* genotypic variants and clinic-pathological characteristics……...171
5.6.3.9 Multiple dimensionality reduction analysis of genetic variants within *EPHX1* gene……...…….176
5.6.4 Statistical analysis for polymorphic sites of *NAT2* gene……...177
5.6.4.1 Genetic model analysis between *NAT2* variants and lung cancer risk……...……177
5.6.4.2 Association of *NAT2* polymorphisms with lung cancer risk based on histological subtypes……181
5.6.4.3 Association of *NAT2* gene with lung cancer risk based on smoking…..188
5.6.4.4 Association of *NAT2* polymorphisms with lung cancer risk based on stratification by gender……194
5.6.4.5 Combinatorial analysis of the risk associated with different polymorphic sites of *NAT2* gene……197
5.6.4.6 Association of the *NAT2* polymorphism & clinic-pathological characteristics…….…..203
5.6.4.7 Haplotype and LD analysis of *NAT2* variants and its association with lung cancer susceptibility…….……...208
5.6.4.8 MDR analysis within *NAT2* gene……209
5.6.4.9 CART analysis of *NAT2* variants……211
5.6.4.10 CART analysis and its association with smoking status……..212
5.7 Clinical characteristics of lung cancer patients evaluated for detecting the prognostic relevance of *NQO1, SULT1A1, EPHX1,* and *NAT2* polymorphisms.....216
5.8 The association of PhaseII detoxification genes towards outcome of lung cancer……..218
5.8.1 Association of polymorphic *NQO1* towards lung cancer outcomes…….218

5.8.1.1 Association of polymorphic *NQO1* genetic variant with overall survival of lung cancer patients……...218
5.8.1.2 Association of NQO1 polymorphism and overall survival according to a histological subtype of lung cancer patients……....219
5.8.1.3 Prognosis of LC patients treated with a different chemotherapeutic regimen……220
5.8.1.4 Association of *NQO1 Pro^{187}Ser* polymorphism with treatment response towards chemotherapeutic treatment and their correlation with objective response……...223
5.8.1.5 Association between *NQO1* polymorphism and toxicity……224
5.8.2 Association of polymorphic *SULT1A1* towards lung cancer outcomes……230
5.8.2.1 Association of SULT1A1 polymorphism with overall survival of LC patients……..230
5.8.2.2 Association of *SULT1A1 Arg^{213}His* with OS in LC histological subtypes.....230
5.8.2.3 Prognosis of LC patients treated with the different chemotherapeutic regimen and its relation with *SULT1A1 Arg^{213}His* polymorphism…….....233
5.8.2.4 *SULT1A1* polymorphism and its response towards chemotherapy……..…...235
5.8.2.5 Association between *SULT1A1* polymorphism and toxicity……...235
5.8.3 Association of polymorphic *EPHX1* towards lung cancer outcomes……...241
5.8.3.1 Overall association of *EPHX1* SNPs with overall survival of lung cancer patients……241
5.8.3.2 Association of *EPHX1* SNPs with overall survival in lung cancer histological subtypes……...244
5.8.3.3 Association of genetic determinants of *EPHX1* and overall survival of lung cancer patients treated with different chemotherapeutic regimens…….……248
5.8.3.4 Association of *EPHX1* polymorphism and chemotherapy response……...252
5.8.4 Association of *NAT2* polymorphism with lung cancer outcomes…….....261
5.8.4.1 Association between the *NAT2* polymorphisms and prognosis in the lung cancer patients……..261
5.8.4.2 Association of *NAT2* polymorphism and overall survival according to the histological subtype of lung cancer patients……...263
5.8.4.3 Association of *NAT2* SNPs with chemotherapy regimens and OS…….269
5.8.4.4 Association of *NAT2* polymorphism and chemotherapy response……..278
5.8.4.5 Association between *NAT2* polymorphism and toxicity…….280
5.9 Statistical analysis of polymorphic sites of *MTHFR* gene……..…...297
5.9.1 Patients characteristics and clinical predictors…….297
5.9.2 Genotype and genetic equilibrium test for *MTHFR* variants……297
5.9.3 Overall survival analysis of *MTHFR 677C>T* and *1298A>C* genotype……..299
5.9.4 Association of *MTHFR* polymorphism with chemotherapy regimens and OS....301
5.9.5 Association of *MTHFR* polymorphisms with gender and OS……..302

5.9.6 Association of the *MTHFR* polymorphism & clinic-pathological characteristics with OS……302
5.9.7 Association of *MTHFR* polymorphism and chemotherapy response……..…309
5.9.8 Cox proportional hazards regression model analysis of overall survival in adenocarcinoma……..309
5.9.9 Association between *MTHFR* genotypes and toxicity…….310
5.10 Statistical analysis of *GSTP1* gene……...……..……319
5.10.1 Patient characteristics and clinical predictors……...319
5.10.2 Genotypic distribution of *GSTP1*…….319
5.10.3 Association of the *GSTP1 Ile^{105}Val* polymorphism & clinic-pathological characteristics……..….321
5.10.4 Association of *GSTP1 Ile105Val* polymorphism and chemotherapy response....322
5.10.5 Survival analysis of *GSTP1* genotype……..323
5.10.6 Survival analysis of *GSTP1* polymorphism based on histological subtype…….324
5.10.7 Association of *GSTP1 Ile105Val* polymorphism with chemotherapy regimens and survival……...325
5.10.8 Association of *GSTP1 Ile^{105}Val* polymorphism with gender and OS……..327
5.10.9 Association of *GSTP1 Ile^{105}Val* on overall survival according to clinic-pathological features……..……...328
5.10.10 Association between *GSTP1 Ile^{105}Val* genotypes and toxicity……330
5.11 High order gene-gene interaction……...……335
5.11.1 Demographic characteristics…….335
5.11.2 Genotypic combination of Phase II detoxification genes and their association with lung cancer…….336
5.11.3 Distribution of polymorphic phase II detoxification genes and their association with histological subtypes……...…...342
5.11.4 Genotypic distribution of triple combinations and combination of four SNPs between Phase II detoxification genes (*NQO1*, *SULT1A1*, *EPHXI*, and *NAT2*) and their association with lung cancer predisposition……...…...348
5.11.5 GMDR analysis and risk of lung cancer……...352
5.11.6 GMDR and association with histological subtypes of lung cancer……..355
5.11.7 GMDR analysis and role of gene-gene environment interaction in lung cancer risk……...358
5.11.8 CART analysis of Phase II detoxification genes……..360
5.11.9 Association of unfavorable genotypic combination of Phase II detoxification genes with overall survival of lung cancer patients……..……378

Chapter 6 Discussion

6.1 *NQO1* polymorphism and its relationship with lung cancer susceptibility……...384

6.1.1 Relation of *NQO1* polymorphism with clinic-pathological features and prognosis…….386
6.2 *SULT1A1* polymorphism and its relationship with lung cancer susceptibility……..388
6.2.1 Relation of *SULT1A1* polymorphism in association with clinic-pathological parameters and prognosis…….390
6.3 *EPHX1* polymorphism and its relationship with lung cancer susceptibility…….391
6.3.1 Relation of *EPHX1* polymorphism in association with clinic-pathological features and prognosis……..397
6.4 *NAT2* polymorphism and its relationship with lung cancer susceptibility……..398
6.4.1 Relation of *NAT2* polymorphism with clinic-pathological features and prognosis……..404
6.5 *MTHFR* polymorphism and its relationship with overall survival of lung cancer patients and toxicity……..405
6.6 *GSTP1* polymorphism and its relationship with overall survival of lung cancer patients and toxicity……..410
6.7 High order gene-gene interaction and toxicity profile of *NQO1*, *SULT1A1*, *NAT2*, and *EPHX1*…….414
6.8 Strengths of the study……...416
6.9 Limitations of the study…….…..417

Chapter 7 Conclusion……418

Chapter 8 Bibliography…….423

Appendix-I……...474

Appendix-II…….477

Appendix-III……....480

Appendix-IV……....483

List of figures

Figure No.	Title	Page No.
Review of Literature		
2.1	Schematic representation of Lung tissue; a) Anatomy of lung showing trachea, bronchus, bronchioles, alveoli, lobes; b) Lung tissue harboring a solid cancerous mass	10
2.2	Microscopic images of different histological subtypes of lung cancer; a): squamous cell carcinoma; b): adenocarcinoma; c): large cell carcinoma; d): small cell lung carcinoma	17
2.3	Scheme linking nicotine addiction and lung cancer via tobacco smoke carcinogens and inducing multiple mutations in critical genes. PAH 4-polycyclic aromatic hydrocarbons; NNK 4 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone.	18
2.4	Region-Specific Incidence Age-Standardized Rates by Sex for Lung Cancer Among Men and Women in 2020. Rates areshown in descending order of the world (W) age-standardized rate in men, and the highest national rates among men and women are superimposed.	24
2.5	Relative proportion (%) of patients according to clinical extent of disease, 2012-2016 (proportion [%] may not total 100% because of rounding)	26
2.6	Annual percentage change (APC) in age-adjusted incidence rates (AAR) over the time period. Calendar years of incidence data for each population-based cancer registry (PBCR) used in trend analysis.	26
2.7	Schematic representation of significant molecular abnormalities taking place during lung cancer pathogenesis	32
2.8	Detoxification pathway followed in the human body	44
2.9	The multiple mechanisms by which carcinogens can cause lung cancer	45
2.10	The phase II drug-metabolizing enzymes participating in the metabolism.	47
2.11	Quinone activation and deactivation as a result of NQO1-mediated reduction.	48
2.12	Chromosome 16 showing the position of the *NQO1* gene on it.	48
2.13	Human NQO1 structure in complex with FAD (the apoenzyme) (PDB 1D4A). (A) NQO1 apoenzyme overall structure. One monomer is shown in green, andthe other in orange. FAD's carbon atoms are shown in purple. (B) An NQO1 apoenzyme monomer. The catalytic domain is depicted in purple, while the C-terminal domain is depicted in blue. The carbon atoms of FAD are depicted in green.	49
2.14	The general reaction catalyzed by quinone oxidoreductase in which NADH or NADPH reduces quinone to hydroquinone	50
2.15	Chromosome 16 showing the position of the *SULT1A1* gene on it	53
2.16	Human SULT1A1 crystal structure complexed with 3 -phosphoadenosine-5 -phosphate (PAP; dark pink) and two p-nitrophenol molecules (dark grey and white). The color of the C trace ranges from blue (N terminus) to red (C terminus). Coils for helices and arrows for strands are used to reflect secondary structural components	53

2.17	The general reaction catalyzed by sulfotransferases	55
2.18	The *EPHX1* gene sequence and regulatory elements are located on chromosome 1 at 1q42.12. (Note: E= exons; HS= DNaseI hypersens5-c97ivity sites, kb= kilobases and rs= reference number SNP ID number). Coding exons are highlighted in grey.	57
2.19	α/β hydrolase fold EHs 3-D structure and enzymatic mechanism. The structural model of *Aspergillus niger* EH, a close relative of mammalian mEH, was created using PyMOL and coordinates from PDB. The complete subunit with the α/β hydrolase fold is highlighted in red, the lid domain highlighted in green, and the N-terminus meander highlighted in gold	58
2.20	The two-step catalytic mechanism of epoxide hydrolysis	59
2.21	Chromosome 8 showing the position of the *NAT* gene on it	62
2.22	The structure of the human NAT2 is depicted as a ribbon diagram. All helices and strands are labeled. The carboxyl and amino termini are denoted by the letters CT and NT, respectively. CoA coupled to NAT2 are depicted as sticks, with carbon atoms depicted in gray. Human NATs have insertions and carboxyl-terminal tails absent in other species' NAT proteins and are magenta	63
2.23	Reaction catalyzed by NAT involving acetylation of arylamine	64
2.24	Chromosome 1 showing the position of the *MTHFR* gene on it	66
2.25	MTHFR ortholog domain arrangement across evolution	67
2.26	Overview of the human MTHFR structural gene depicting the regulatory domain (yellow), the catalytic domain (cyan), and the linker (red). FAD (green) and SAM (pink) bound in the sticks are depicted in the structure. The dotted lines represent chaotic regions that are not represented in the structure. The α-helices and β-sheets are also labeled	68
2.27	MTHFR in one-carbon metabolism at the folate and methionine cycle interconnection	69
2.28	Reaction catalysis carried out by MTHFR	70
2.29	The harmful molecules may diffuse across the plasma membrane and be targeted by the enzyme of Phase I metabolism inside cells. The majority of them are members of the cytochrome P450 family, which includes many enzymes that catalyze various processes such as hydroxylation (the primary one), reduction and, oxidation. GSTs, which catalyze the conjugation of Phase I modified xenobiotics to endogenous GSH, thus play a vital role in ensuing Phase II metabolism. Different transmembrane efflux pumps actively move the conjugate out of the cell (Phase III). Some compounds may directly enter Phase II metabolism	72
2.30	GSTP1 regulatory elements	73
2.31	The structure of human GSTP1 in association with GSH and its inhibitor NBDHEX is depicted. The pink-colored helical is the N-terminal thioredoxin-like domain, while the one-colored cyan is all helical C-terminal domain. A GSH molecule fills the G-site, and the H-site is occupied by an NBDHEX molecule depicted as sticks	74

2.32	Reaction catalysis by GST leading to the formation of glutathione conjugate	75
Results		
5.1	0.8% agarose gel representing genomic DNA isolated from blood; Lane 1: 1kb ladder, Lane 2-5: genomic DNA from the blood of cases, Lane 6-8: genomic DNA isolated from healthy controls.	115
5.2	Representative example of agarose gel electrophoresis showing PCR amplification for *NQO1 Pro^{187}Ser*	116
5.3	Restriction digestion pattern of *NQO1 Pro^{187}Ser* (Lane 1,5,8: Heterozygous genotype (Pro/Ser); Lane 2: mutant (*Ser/Ser*); Lane 3,4,6,7: wild genotype (*Pro/Pro*); Lane 9: uncut; Lane 10: marker (100 bp), and 35 bp band is visible at the lower most of the gel)	116
5.4	Restriction digestion pattern of *SULT1A1Arg^{213}His* (Lane 1: Ladder (50 bp); Lane 2: uncut PCR product; Lane 3: Heterozygous genotype (*Arg/His*); Lane 3: Homozygous variant genotype (*His/His*); Lane 5-8: Homozygous wild genotype (*Arg/Arg*))	117
5.5	Representative example of 8.0% Native-PAGE showing PCR-RFLP pattern of a) *EPHX1 Tyr^{113}His* (rs2854455), and b) *EPHX1 His^{139}Arg* (rs2234922)	118
5.6	Representative example of an agarose gel showing PCR-RFLP pattern of *481C>T*, and 8.0% Native-PAGE showing PCR-RFLP pattern of b) *590G>A*, c) *803A>G*, and d) *857G>A*	119
5.7	Representative example of 6% Native-PAGE exhibiting restriction digestion pattern of *GSTP1 Ile^{105}Val*	120
5.8	Representative example of 6% Native-PAGE exhibiting restriction digestion pattern of a) *MTHFR C677T*, and b) *MTHFR A1298C*	121
5.9	The regional distribution of North Indian lung cancer patients	122
5.10)	Pie chart illustrating the regional distribution of North Indian Lung cancer patients of males.	122
5.10 b)	Pie chart illustrating the regional distribution of North Indian Lung cancer patients of females.	123
5.11	Graph showing the distribution of study subjects based on smoking status. a): Among cases, b): Among controls	124
5.12	Graph showing the distribution of male study subjects based on smoking status	124
5.13	Graph showing the age-wise distribution of lung cancer based on gender	125
5.14	Stage based distribution of the study subject	126
5.15	Pie chart illustrating the distribution of North Indian Lung cancer patients based on histological subtypes	127
5.16	Pie chart illustrating the distribution of North Indian Lung cancer patients based on karnofsky performance status (KPS)	128
5.17	Pie chart illustrating North Indian Lung cancer patients' distribution based on Eastern Cooperative Oncology Group -performance status (ECOG-PS).	128
5.18	Pie chart illustrating a distribution of North Indian Lung cancer patients based on Doublet platinum-based chemotherapeutic regimen	129
5.19	Showing pairwise linkage disequilibrium between the two variants of *EPHX1* polymorphism	170
5.20	MDR interactions with *EPHX1* variants; X1: *EPHX1 Tyr^{113}His*; X2: *EPHX1 His^{139}Arg*	176

5.21	Showing pairwise linkage disequilibrium (*D'*) between the four variants of *NAT2* gene, wherein pink shows the average LD and white depicts the minimum LD between the polymorphic sites	209
5.22	The multifactor dimensionality reduction (MDR) interaction dendrogram a) overall, b) smokers and c) non-smokers	211
5.23	CART analysis in polymorphic *NAT2*. W= homozygous wild type genotype, M = heterozygous + homozygous variant genotype	214
5.24	CART analysis in polymorphic *NAT2* based on histology, a) ADCC, b) SQCC, and c) SCLC. W= homozygous wild type genotype, M = heterozygous + homozygous variant genotype	215
5.25	CART analysis in polymorphic *NAT2* based on smoking status, a) smokers, b) non-smokers. W= homozygous wild type genotype, M = heterozygous + homozygous variant genotype	216
5.26	Cox proportional hazards-regression curve illustrating the association between lung cancer subjects with *variant (TT)* genotype treated with Paclitaxel cisplatin/carboplatin regimen showing *NQO1* 609C>T polymorphism	223
5.27	Cox hazards overall survival curve of heterozygous (*His/Arg*) genotype depicting *EPHX1 His^{139}Arg* polymorphism	242
5.28	Kaplan Meier curves illustrating the association between overall survival in patients with a) heterozygous (*His/Arg*) genotype, b) combined genotype (*His/Arg+Arg/Arg)*, for SCLC showing *EPHX1 His^{139}Arg* polymorphism	245
5.29	Kaplan Meier curves illustrating the association between overall survival in patients with heterozygous (*Tyr/His*) genotype for patients treated with cisplatin/carboplatin alongwith irinotecan depicting *EPHX1Tyr^{113}His* polymorphism	252
5.30	**a)** Cox hazards curve of mutant (*NAT2*7/*7*) genotype of *857G>A* for subjects treated with cisplatin/carboplatin along with paclitaxel; **b)** Kaplan Meier curves illustrating the association between overall survival in patients with heterozygous (*NAT2*4/*5C*) genotype for patients treated with cisplatin/carboplatin along with paclitaxel	272
5.31	**a)** Cox proportional hazards curve illustrating mutant (*NAT2*6/*6*) genotype for patients treated with cisplatin/carboplatin alongwith irinotecan depicting *590G>A* polymorphism; **b)** Cox proportional hazards curve illustrating mutant (*NAT2*7/*7*) genotype of *857G>A* polymorphism for patients treated with cisplatin/carboplatin alongwith irinotecan.	275
5.32	**a)**Kaplan Meier curves illustrating the association between overall survival in different genotypes of *MTHFR A1298C* polymorphism in ADCC subjects. (Note: *AA* represents wild-type genotype, *AC* represents heterozygous genotype, and *CC* represents homozygous for mutant genotype). **b)** Kaplan Meier curves illustrating the association between overall survival in different genotypes of *MTHFR C677T* polymorphism in ADCC subjects. (Note: *CC* represents wild-type genotype, *CT* represents heterozygous genotype, and mutant genotype).	300
5.33	Kaplan Meier curves illustrating the association between overall survival in different genotypes of *GSTP1 Ile^{105}Val* polymorphism in ADCC subjects.	324
5.34	GMDR interaction dendrogram between Phase II detoxification genes (X1: *NQO1*, X2: *SULT1A1*, X3: *EPHX1Tyr^{113}His*, X4: *EPHX1 His^{139}Arg*, X5: *NAT2*6*, X6: *NAT2*7*, X7: *NAT2*5B*, X8: *NAT2*5C*)	353

5.35	Entropy interaction map of lung cancer risk between Phase II detoxification genes. (*X1: NQO1, X2: SULT1A1, X3: EPHX1Tyr113His, X4: EPHX1 His139Arg, X5: NAT2*6, X6: NAT2*7, X7: NAT2*5B, X8: NAT2*5C*)	354
5.36	Distribution of high-risk and low-risk genotypes in the best three-locus model. (*X1: NQO1, X2: SULT1A1, X3: EPHX1Tyr113His, X4: EPHX1 His139Arg, X5: NAT2*6, X6: NAT2*7, X7: NAT2*5B, X8: NAT2*5C)*	355
5.37	GMDR interaction dendrogram between Phase II detoxification genes (a) in ADCC subjects, (b) in SCLC subjects *(X1: NQO1, X2: SULT1A1, X3: EPHX1Tyr113His, X4: EPHX1 His139Arg, X5: NAT2*6, X6: NAT2*7, X7: NAT2*5B, X8: NAT2*5C)*	358
5.38	GMDR interaction dendrogram between Phase II detoxification genes in smokers (*X1: NQO1, X2: SULT1A1, X3: EPHX1Tyr113His, X4: EPHX1 His139Arg, X5: NAT2*6, X6: NAT2*7, X7: NAT2*5B, X8: NAT2*5C)*	359
5.39	Overall risk estimates based on CART analysis of Phase II detoxification genes for ADCC subgroup (W: Homozygous wild type genotype, M: Heterozygous + Homozygous variant genotype)	371
5.40	Overall risk estimates based on CART analysis of Phase II detoxification genes for non-smokers subgroup (W: Homozygous wild type genotype, M: Heterozygous + Homozygous variant genotype)	377
5.41	**a)** Kaplan Meier curves illustrating the association between overall survival in heterozygous and wild genotypes for the genotypic combination of *NQO1 $Pro^{187}Ser$* with *EPHX1 $His^{139}Arg$*. (Note: *AA* and *CC* represent wild-type genotype, *CT* and *AG* represent heterozygous genotype).	383
	b) Kaplan Meier curves illustrating the association between overall survival in heterozygous and wild genotype for the combination of *SULT1A1 $Arg^{213}His$* with *EPHX1 $His^{139}Arg$*. (Note: *GG* and *AA* represent wild-type genotype, *GA* and *AG* represent heterozygous genotype).	

List of tables

Table No.	Tables	Page No.
Review of Literature		
2.1	TNM stage grouping (Eighth edition)	11-12
2.2	Tumor, node, and metastasis (TNM) definition	12-13
2.3	List of occupational hazards	22
2.4	The phase II detoxification activities in human	46
2.5	The different studies conducted in *NQO1* polymorphism towards lung cancer risk and included in the meta-analysis	77-78
2.6	Principal characteristics of the studies included in the meta-analysis	84
2.7	Distribution of *EPHX1 $His^{139}Arg$* among lung cancer cases and controls	85
2.8	The different studies conducted in *NAT2* polymorphism and included in the meta-analysis	90
Materials and Methods		
4.1	List of primers and restriction enzymes used for *NQO1* genetic variant	107
4.2	List of primers and restriction enzymes used for *SULT1A1* genetic variant	108
4.3	List of primers and restriction enzymes used for *EPHX1* genetic variants	109
4.4	List of primers and restriction enzymes used for *NAT2* genetic variants	110
4.5	List of primers and restriction enzymes used for *MTHFR* genetic variants	111
4.6	List of primers and restriction enzymes used for *GSTP1* genetic variant	111
Results		
5.1	Demographic characteristics among cases and controls	131
5.2	Genotypic and allelic distribution of the *NQO1* genetic variant and its association with risk of lung cancer	132
5.3	*Ad*-hoc analysis of the *NQO1* genetic variant and its association with risk of lung cancer	133
5.4	Genotypic and allelic distribution of the *NQO1* genetic variant and its association with risk of lung cancer on histology	133
5.5	Relationship of different *NQO1* genotypes with the smoking status of cases and controls	136
5.6	Relationship of different genotypes with the clinical-pathological parameters	138-139
5.7	Genotypic and allelic distribution of the *SULT1A1* genetic variant and its association with risk of lung cancer	140
5.8	*Ad*-hoc analysis of the *SULT1A1* genetic variant and its association with risk of lung	141
5.9	Genotypic and allelic distribution of the *SULT1A1* genetic variant and its association with risk of lung cancer on histology	142
5.10	Genotypic frequencies of *SULT1A1* among cases and controls and their relationship with risk of SQCC histological subtype in males	144
5.11	Relationship of different genotypes based on gender	145
5.12	Relationship of different *SULT1A1* genotypes of males and females based on age	146
5.13	Relationship of different *SULT1A1* genotypes with the smoking status of cases and controls	148
5.14	Relationship of different *SULT1A1* genotypes with the pack years of heavy smokers amongst cases and controls	150
5.15	Relationship of different genotypes with the clinical-pathological parameters	153
5.16	Genotypic and allelic distribution of the *EPHX1* genetic variants and its association with risk of lung cancer	155
5.17	*Ad*-hoc analysis of the *EPHX1* genetic variant and its association with risk of lung cancer	155-156

5.18	Genotypic and allelic distribution of the *EPHX1exon3* and *exon4* genetic variant and its association with risk of lung cancer according to tumor histology	157
5.19	Relationship of different *EPHX1* genotypes with the smoking status of cases and controls	160-161
5.20	Relationship of different *EPHX1* polymorphisms based on gender	163
5.21	Combinatorial analysis of the risk associated with different polymorphic sites of *EPHX1* gene	165
5.22	Combined effects of $Tyr^{113}His$ (rs2854455) and $His^{139}Arg$ (rs2234922) on the risk of lung cancer	167
5.23	Genotypic and allelic distribution of the *EPHX1* genetic variant and its association with risk of lung cancer as Surrogate of Enzyme Activity	168-169
5.24	Haplotype frequencies and lung cancer risk associated with *EPHX1* haplotypes	170
5.25	Relationship of different genotypes with the clinical-pathological parameters	174-175
5.26	Multifactor dimensionality reduction analysis showing interactions of *EPHX1* variants with lung cancer risk	176
5.27	Genotypic and allelic distribution of the *NAT* genetic variant and its association with risk of lung cancer	179
5.28	*Ad*-hoc analysis of the *NAT2* genetic variant and its association with risk of lung cancer	180-181
5.29 a)	Genotypic and allelic distribution of the *NAT2*6 590G>A* genetic variant and its association with risk of lung cancer on basis of tumor histology	183
5.29 b)	Genotypic and allelic distribution of the *NAT2*7 857G>A* genetic variant and its association with risk of lung cancer on basis of tumor histology	184
5.29 c)	Genotypic and allelic distribution of the *NAT2*5B 481C>T* genetic variant and its association with risk of lung cancer on basis of tumor histology	186
5.29 d)	Genotypic and allelic distribution of the *NAT*5C 803A>G* genetic variant and its association with risk of lung cancer on basis of tumor histology	187
5.30 a)	Relationship of different *NAT2* genotypes with the smoking status of cases and controls (smokers)	190
5.30 b)	Relationship of different *NAT2* genotypes with the smoking status of cases and controls (light-smokers)	191
5.30 c)	Relationship of different *NAT2* genotypes with the smoking status of cases and controls (heavy smokers)	192
5.30 d)	Relationship of different *NAT2* genotypes with the smoking status of cases and controls (non-smokers)	193
5.31 a)	Relationship of different *NAT2* polymorphisms based on gender (males)	195
5.31 b)	Relationship of different *NAT2* polymorphisms based on gender (females)	196
5.32	Combinatorial analysis of the risk associated with different polymorphic sites of *NAT2* gene	198
5.33	Combined effects of *NAT2* variants on the risk of lung cancer	200-202
5.34 a)	Relationship of different *NAT2* genotypes with the clinical-pathological parameters (clinical stage)	204
5.34 b)	Relationship of different *NAT2* genotypes with the clinical-pathological parameters (primary tumor extension)	205
5.34 c)	Relationship of different *NAT2* genotypes with the clinical-pathological parameters (lymph node invasion)	206
5.34 d)	Relationship of different *NAT2* genotypes with the clinical-pathological parameters (metastasis)	207
5.35	Haplotype frequencies and lung cancer risk associated with *NAT2* haplotypes	208
5.36	Showing pairwise *D'* (r^2) values for the four polymorphic sites among cases and controls	209

5.37	Multifactor dimensionality reduction (MDR) analysis showing association of high-order interactions with lung cancer	210
5.38	Risk estimates based on classification and regression tree analysis (*NAT2* variants) overall, on the basis of histology, and smoking status	213
5.39	Patients characteristics and clinical features	217-218
5.40	Association of *NQO1* SNP (rs1800566) on overall survival in lung cancer cases	218
5.41	Association of *NQO1* SNP (rs1800566) on survival in lung cancer cases according to tumor histology	219-220
5.42	Association of *NQO1* SNP (rs1800566) on overall survival according to regimen	222
5.43	Relationship of genotypic distribution with clinic-pathological parameter based on response to chemotherapy	224
5.44 a)	Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NQO1* polymorphisms in multivariate analysis	226-227
5.44 b)	Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NQO1* polymorphisms in multivariate analysis	228
5.44 c)	Gastrointestinal Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NQO1* polymorphisms in multivariate analysis	228-229
5.45	Univariate and Multivariate analysis for *SULT1A1*	230
5.46	Association of *SULT1A1* SNP (rs9282861) on overall survival according to tumor histology	232
5.47	Association of *SULT1A1* SNP (rs9282861) on overall survival according to chemotherapeutic regimen	234
5.48	Relationship of genotypic distribution with lung cancer based on response to chemotherapy	235
5.49 a)	Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *SULT1A1* polymorphisms in multivariate analysis	237-238
5.49 b)	Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *SULT1A1* polymorphisms in multivariate analysis	239
5.49 c)	Gastrointestinal Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *SULT1A1* polymorphisms in multivariate analysis	239-240
5.50	Association of *EPHX1* on overall survival in lung cancer cases and based on histology	243
5.51	Association of *EPHX1* SNPs on overall survival according to tumor histology	246-247
5.52	Association of *EPHX1*genotype on survival in lung cancer cases based on the chemotherapeutic regimen	250-251
5.53	Relationship of genotypic distribution with clinic-pathological parameter based on response to chemotherapy	253
5.54 a)	Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *EPHX1* polymorphisms in multivariate analysis	254-256
5.54 b)	Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *EPHX1* polymorphisms in multivariate analysis	258
5.54 c)	Gastrointestinal Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *EPHX1* polymorphisms in multivariate analysis	259-260
5.55	Association of *NAT2* on overall survival in lung cancer cases and on basis of histology	262
5.56 a)	Association of *NAT2*6 590G>A* polymorphism on overall survival according to tumor histology	264
5.56 b)	Association of *NAT2*7 857G>A* polymorphism on overall survival according to tumor histology	265
5.56 c)	Association of *NAT2*5B 481C>T* polymorphism on overall survival according to tumor histology	267

5.56 d)	Association of *NAT2*5C 803A>G* polymorphism on overall survival according to tumor histology	268
5.57 a)	Association of *NAT2* genotype on survival in lung cancer cases based on the chemotherapeutic regimen (Paclitaxel + cis/carboplatin)	270-271
5.57 b)	Association of *NAT2* genotype on survival in lung cancer cases based on the chemotherapeutic regimen (Irinotecan + cis/carboplatin)	274
5.57 c)	Association of *NAT2* genotype on survival in lung cancer cases based on the chemotherapeutic regimen (Docetaxel + cis/carboplatin)	276
5.57 d)	Association of *NAT2* genotype on survival in lung cancer cases based on the chemotherapeutic regimen (Pemetrexed + cis/carboplatin)	277
5.58	Relationship of genotypic distribution with clinic-pathological parameter based on response to chemotherapy	279
5.59 a)	Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NAT2* polymorphisms in multivariate analysis	281-288
5.59 b)	Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NAT2* polymorphisms in multivariate analysis	290-291
5.59 c)	Gastrointestinal **t**oxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NAT2* polymorphisms in multivariate analysis	291-296
5.60	Patient characteristics	298
5.61	Results of genetic equilibrium test	299
5.62	Univariate and Multivariate analysis for *MTHFR* variants	300
5.63	Association of *MTHFR* SNPs on overall survival according to Chemotherapeutic regimen	301
5.64	Association of *MTHFR* SNPs on overall survival based on gender	303
5.65 a)	Association of *MTHFR* SNPs on overall survival according to clinic-pathological features	304
5.65 b)	Association of *MTHFR* SNPs on overall survival according to clinic-pathological features	306
5.65 c)	Association of *MTHFR* SNPs on overall survival according to clinic-pathological features	307-308
5.66	Association of *MTHFR* SNPs and chemotherapeutic response	309
5.67	Cox proportional hazards regression model analysis of overall survival in ADCC subjects	310
5.68 a)	Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4: grades 0-2 vs. 3-4) associated with *MTHFR* polymorphisms in multivariate analysis	312-314
5.68 b)	Gastrointestinal Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4: grades 0-2 vs. 3-4) associated with *MTHFR* polymorphisms in multivariate analysis	316-317
5.68 c)	Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4: grades 0-2 vs. 3-4) associated with *MTHFR* polymorphisms in multivariate analysis	318
5.69	Clinical predictor among cases	320
5.70	Genotypic distribution of *GSTP1* gene	321
5.71	Relationship of different genotypes with the clinic-pathological parameters	322
5.72	Relationship of *GSTP1 Ile^{105}Val* polymorphism based on response to chemotherapy	323
5.73	Survival among patients of lung cancer categorized with *GSTP1 Ile^{105}Val* genotype	324
5.74	Association of *GSTP1 IIe^{105}Val*polymorphism on overall survival according to tumor histology	325
5.75	Association of *GSTP1 IIe^{105}Val* on overall survival according to chemotherapeutic regimen	326-327
5.76	Association of *GSTP1 IIe^{105}Val* on overall survival according to gender	327
5.77	Association of *GSTP1* SNP (rs1695) on overall survival according to clinic-pathological features	329-330

5.78 a)	Hematological Toxicities (grades 0 vs. 1-4: grades 0-1 vs. 2-4 : grades 0-2 vs 3-4) associated with *GSTP1 IIe^{105}Val* polymorphisms in multivariate analysis	332-333
5.78 b)	Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4 : grades 0-2 vs 3-4) associated with *GSTP1 IIe^{105}Val* polymorphisms in multivariate analysis	333
5.78 c)	Gastrointestinal toxicity (grades 0 *vs*. 1-4: grades 0-1 *vs*. 2-4 : grades 0-2 *vs* 3-4) associated with *GSTP1 IIe^{105}Val* polymorphisms in multivariate analysis	334
5.78 d)	Hematological Toxicities (grades 0 vs. 1-4: grades 0-1 vs. 2-4 : grades 0-2 vs 3-4) associated with *GSTP1 IIe^{105}Val* polymorphisms in multivariate analysis for SCLC patients	335
5.79	Demographic characteristics among cases and controls	336
5.80	Minor allele frequencies and risk associated with single-locus sites among cases and controls	338
5.81	Distribution of genotypes depicting dual combination and their association with lung cancer risk	340-342
5.82	Distribution of genotypes depicting dual combination and their association with histological subtypes of lung cancer	344-348
5.83	Distribution of genotypes depicting triple combination and their association with lung cancer risk	350
5.84	Genotypic distribution based on different genotypic combinations and their association with lung cancer risk	351
5.85	Multifactor Dimensionality Reduction (MDR) analysis showing interactions of Phase II detoxification variants with lung cancer risk	353
5.86	Multifactor Dimensionality Reduction (MDR) analysis showing interactions of Phase II detoxification variants with SCLC risk	356
5.87	Multifactor Dimensionality Reduction (MDR) analysis showing interactions of Phase II detoxification variants with ADCC risk	357
5.88	Multifactor Dimensionality Reduction (MDR) analysis showing interactions of Phase II detoxification variants with SQCC risk	357
5.89	Multifactor Dimensionality Reduction (MDR) analysis showing gene-environment interactions (smokers) of Phase II detoxification variants with lung cancer risk	359
5.90	Multifactor Dimensionality Reduction (MDR) analysis showing gene-environment interactions (non-smokers) of Phase II detoxification variants with lung cancer risk	360
5.91	Overall risk estimates based on CART analysis of Phase II detoxification genes	361-364
5.92	Overall risk estimates based on CART analysis of Phase II detoxification genes for SCLC subgroup	367-370
5.93	Overall risk estimates based on CART analysis of Phase II detoxification genes for ADCC subgroup	371
5.94	Overall risk estimates based on CART analysis of Phase II detoxification genes for smokers	373-376
5.95	Overall risk estimates based on CART analysis of Phase II detoxification genes for non-smokers subgroup	376
5.96	Genotypic combinations and overall survival	380-382

List of Abbreviations

Abbreviated form	Full form
NQO1	NAD(P)H Quinone Dehydrogenase 1
SULT1A1	Sulfotransferase family 1A member 1
EPHX1	Epoxide Hydrolase 1
NAT2	N-Acetyltransferase 2
GSTP1	Glutathione-S-Transferase Pi 1
MTHFR	Methylenetetrahydrofolate Reductase
NSCLC	Non-Small Cell Lung Cancer
SNP	Single Nucleotide Polymorphism
MDR	Multifactor Dimensionality Reduction
CART	Classification And Regression Tree Analysis
DNA	Deoxyribonucleic Acid
IASLC	International Association for the Study of Lung Cancer
FNAC	Fine Needle Aspiration Cytology
PET	Positron Emission Tomography
CT	Computed Tomography
TNM	Tumour Node and Metastasis
IARC	International Agency For Research On Cancer
ICMR	Indian Council Of Medical Research
PBCR	Population Based Cancer Registry
NNK	4-(Methylnitrosamino)-1-(3-Pyridyl)-1-Butanone
PAH	Polycyclic Aromatic Hydrocarbon
SCLC	Small Cell Lung Cancer
LCC	Large Cell Carcinoma
ADCC	Adenocarcinoma
SQCC	Squamous Cell Carcinoma
CAM	Calcium/calmodulin-dependent protein
EMA	Epithelial Membrane Antigen
Rb	Retinoblastoma
EGFR	Epidermal Growth Factor Receptor
PIK3CA	Phosphatidylinositol 3-Kinase, Catalytic, Alpha
mTOR	Mammalian target of rapamycin

RAF	Rapidly Accelerated Fibrosarcoma
MAPK	Mitogen-Activated Protein Kinase
ATP	Adenosine Triphosphate
TKI	Tyrosine Kinase Inhibitor
PTEN	Phosphatase Tensin
TGF-β	Transforming Growth Factor
E2F	Elongation factor
BCL-2	B-Cell Lymphoma-2
PIP3	Phosphatidylinositol (3,4,5)-Trisphosphate
CR	Complete Response
PR	Partial Response
SD	Stable Disease
PD	Progressive Disease
EDTA	Ethylene Diaminetetra Acetic Acid
TBE	Tris/Borate/EDTA
TE	Tris-EDTA buffer
RBC	Red Blood Cells
HR’	Adhusted Hazards Ratio
HR	Hazards Ratio
OR	Odds Ratio
AOR	Adjusted Odds Ratio
C.I	Confidence Interval
PGIMER	Post Graduate Institute of Medical Education and Research
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
OD	Optical Density
Dntp	Deoxyribose Nucleoside triphosphates
BSA	Bovine Serum Albumin
NEB	New England biolab
PAGE	Polyacrylamide Gel Electrophoresis
APS	Ammonium Persulfate
TEMED	N,N,N’.N’ Tetramethylethylene-1,2-Diamine
HWE	Hardy Weinberg equilivrium
LD	Linkage disequilibrium

FDR	Flase Discovery rate
CVC	Cross validationconsistency
MST	Median survival time
OS	Overall Survival
WHO	World Health Organization
W	Wild
H	Heterozgote
M	Mutant
Pro	Proline
Ser	Serine
Arg	Arginine
His	Histidine
Tyr	Tyrosine
Leu	Leucine
Gln	Glutamine
Lys	Lysine
Gly	Glycine
Ala	Alanine
Val	Valine
Glu	Glutamate
Ile	Isoleucine
MAF	Minor Allele Frequency
RECIST	Response Evaluation Criteria In Solid Tumors
KPS	Karnofsky Performance Status
ECOG	Eastern Cooperative Oncology Group
WBC	White blood cells

List of Symbols

Symbol	
%	Percentage
<	Less than
≥	Equals to or greater than
°	Degree
°C	Degree Celsius
µg	Microgram
µm	Micrometer
cm	Centimeter
g	Gram
h	Hour
rpm	Revolutions per minute
mg/m^2	Milligram per square meter
L	Litre
M	Molar
χ^2	Chi-square
mg	Milligram
min	Minute
ml	Milliliter
mm	Millimeter
mM	Millimolar
Mg	Magnesium
nm	Nanometer
df	Degree of freedom
s	Second
U	Unit (activity)
v	Volume
Cl	Chloride
D	Coefficient of linkage disequilibrium
r^2	Extent of linkage disequilibrium
µg/ml	Microgram per milliliter
µM	Micromolar

EtBr	Ethidium Bromide
HCl	Hydrochloride
ng/µl	Nanograms per microliter
bp	Base pair
α	Alpha
β	Beta
γ	Gamma
±	Plus minus
NaCl	Sodium chloride
A	Adenine
T	Thymine
G	Guanine
C	Cytosine

CHAPTER 1

INTRODUCTION

Lung cancer is a disease in which some cells in the lungs become aberrant and develop uncontrollably, resulting in the formation of a tumor. Cancer is the leading cause of death worldwide, accounting for an estimated 10 million deaths in 2020 (WHO, 2021). The most common causes of cancer death are cancers of the lung (18.0% of all the cancer deaths), followed by colorectal (9.4% of deaths), liver (8.3%), stomach (7.7% deaths), and breast cancer (6.9% deaths). Unregulated cell growth is the hallmark of cancer (Sung *et al.*, 2021). Lung cancer is one of the significant causes of mortality and morbidity in developed and developing countries (Torre *et al*., 2015; Li and Wang, 2014). The most common cancers among males in the Indian subcontinent were lung cancer and oral/mouth cancer. The data states that 1 in 68 males and 1 in every 201 females might develop lung cancer during their lifetime (Mathur *et al*., 2020).

Both smoking and exposure to different carcinogenic compounds are risk factors for a predisposition towards lung cancer. Exposure to environmental carcinogens or tobacco smoke may even cause lung cancer in people who are non-smokers (Kiyohara *et al*., 2002). It has been estimated that one billion men and 250 million women are characterized as smokers worldwide (Hecht *et al.*, 2016). As the number of years spent smoking and the number of cigarette packs smoked per day increases, the risk of lung cancer too increases (Ettinger *et al*., 2012). Lung cancer develops approximately one in ten lifelong smokers, which implies that the host factor may influence the individual susceptibility to tobacco smoke (Kiyohara *et al*., 2002).

Lung cancer is classified as non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for about 85% of lung cancer and includes adenocarcinoma (ADCC), squamous cell carcinoma (SQCC), and large cell carcinoma (LCC). At the same time, SCLC accounts for the remaining 15% of lung cancer cases. SCLC metastasizes much faster to other organs than NSCLC types, and if untreated within a few weeks, SCLC can be fatal compared to metastases of NSCLC in most cases (Travis, 2011). Among the several chemotherapeutic agents available for treatment, platinum-based doublet chemotherapy remains

the standard initial treatment for NSCLC patients; however, the efficacy of platinum-based chemotherapy varies among individuals, with a response rate of 26–60%. Thus the overall survival is still dismal (Travis, 2011). Although the significant prognostic determinant for lung cancer is the disease's stage, there are differences in survival among patients with a similar disease.

The human body, throughout its lifespan, is exposed to a variety of xenobiotics, including food components and pharmaceuticals. There are various environmental risk factors for lung cancer susceptibility, with smoking being one of the most prominent. In recent years, the altered functions of Phase I and phase II detoxifying enzymes have been found to modify susceptibility to environmental carcinogens. The primary hypothesis is that there might be an accumulation of activated genotoxic metabolites, increasing the risk of lung cancer due to an increase in the activity of Phase I enzymes that produce activated metabolites from environmental procarcinogens and a decrease in the activity of phase II enzymes that detoxify these activated metabolites by conjugating with them (Nyberg *et al.*, 1998).

Detoxification or biotransformation of xenobiotics *viz.* pharmaceuticals and food components and endogenous toxins are essential. The human body accomplishes various detoxifying enzymes to remove toxins from the body (Liska, 1998). These xenobiotics are detoxified using two systems, *viz.* Phase I and Phase II detoxification (Jancova *et al.*, 2010). The role of these enzymes in cellular defense is still being investigated in various experiments. The modern advancement in food (*viz.* dietary supplements, preservatives, and artificial substitutes) and chemical (*viz.* environmental pollutants and synthetic compounds) industries challenges the ability of these enzymes to provide a cellular defense. Both Phase I and Phase II detoxification enzymes are involved in the activation of xenobiotics and endogenous toxins. The imbalance between both may lead to genomic instability, impairing cellular integrity and potentially promoting several diseases, including cancer (Coleman, 2010; Jancova *et al.*, 2010).

Phase I enzymes convert non-polar compounds to more polar compounds, while Phase II enzymes convert them to polar and water-soluble forms that can be eliminated from the body *via* urine or feces. Cytochrome P450 is the most common enzyme family among Phase I detoxifying enzymes. The reactive molecules which are produced by the Phase I enzymes need to be metabolized by the phase II conjugation; otherwise, these may cause damage to DNA, RNA, and

proteins (Rushmore and Tony, 2002; Croom, 2012). Individual predisposition to developing chemically induced cancer is influenced by a cellular imbalance between Phase I carcinogen activating enzymes and Phase II detoxifying enzymes (Coleman, 2010). The various conjugation reactions are present in the body, *viz.* sulfation, glucuronidation, acetylation, amino acid, and glutathione conjugation. The various enzymes involved in Phase II detoxification are mostly transferases. They include UDP-glucuronosyl transferases (UGTs), sulfotransferases (SULTs), N-acetyl transferases (NATs), glutathione S-transferases (GSTs), and various methyltransferases (mainly thiopurine S-methyl transferase (TPMT) and catechol O-methyl transferase (COMT) (Jancova *et al.,* 2010). The genes studied are discussed in brief below.

The Phase II detoxification enzymes function as inactivating enzymes, catalyzing the binding of intermediate metabolites to co-factors, converting them into more hydrophilic products, and facilitating their removal (Chen *et al*., 2006). N-acetyltransferase (NAT) is a Phase II enzyme that catalyzes the acetylation of aromatic amines and hydrazines, including therapeutic drugs and carcinogenic compounds, and acts as a modulator of drug and carcinogen detoxification agents (Mota *et al*., 2015). N-acetyltransferase 2 (*NAT2*) gene has no intron and is located on chromosome 8 (8q22). The *NAT2* gene contains an 870bp open reading frame that encodes for a protein of 290 amino acids. It is involved in the metabolism of heterocyclic, aromatic amines, and hydrazines *via* O-acetylation and N-acetylation, as these carcinogens are linked to initiating cancer development (Tian *et al*., 2014; Liu *et al*., 2015). It catalyzes the conversion of acetyl from acetyl CoA to nitrogen of the substrate. Polymorphisms in the *NAT2* gene locus result in impaired enzyme activity (Borlak and Reamon-Buettner, 2006). The human *NAT2* gene is polymorphic and results in a wide range of N-acetylation activity in humans, from rapid to slow acetylators. Several studies have been conducted to evaluate the role of NAT2 genetic polymorphisms in lung cancer risk, but the results have been inconsistent. Some studies show no overall association between NAT2 rapid or slow acetylators and lung cancer risk (Martínez *et al*., 1995; Bouchardy *et al*., 1998; Saarikoski *et al*., 2000). Some studies state that the homozygous rapid acetylator (Cascorbi *et al*., 1996) or slow acetylator genotype (Cascorbi *et al*., 1996; Oyama et al. *al*., 1997; Hou *et al*., 2000; Sobti *et al*., 2009) to be a risk factor for lung cancer. In a meta-analysis conducted by Cui *et al*. (2011) and Liu *et al*. (2015), no association was observed in the overall analysis and subgroup analysis based on ethnicity, gender, smoking status, and histological subtypes.

Human sulfotransferases (SULTs) are vital phase II detoxifying enzyme that catalyzes the sulfonate conjugation of a wide range of endogenous compounds such as neurotransmitters, many drugs, steroid and thyroid hormones, xenobiotics and pro-carcinogenic agents (Dash *et al.*, 2019; Xiao *et al.*, 2014). The sulfotransferases have been identified to be differentially expressed in different types of human tissues, *viz.* lung, brain, colon, liver, platelet, and other sites (Jones et al., 1992; Li et al., 2008). Sulfonation is a detoxifying mechanism that produces more water-soluble and hence fewer toxic metabolites; however, this conjugation reaction of some compounds produces electrophiles that may bind to DNA to form carcinogen-DNA adducts (Coughtrie *et al.*, 1998; Banoglu, 2000). The most expressed member of the phenol SULT1 family is SULT1A1. The SULT1A1 enzyme is expressed by the *SULT1A1* gene, which maps to chromosome 16p11.2-p12.1. Apart from toxic compounds inactivation, the SULT1A1 isoform is also involved in the bioactivation of procarcinogens such as polycyclic aromatic hydrocarbons (PAHs) and heterocyclic amines (HCAs), both of which are found in tobacco smoke (Bardakci *et al.*, 2008; Koike *et al.*, 2008; Kumar *et al.*, 2013). Among the different types of polymorphic variants found in the *SULT1A1* gene, $Arg^{213}His$ is the most studied polymorphism, which results in Arg to His amino acid replacement at exon 7. *SULT1A1*2* (^{213}His variant allele) derived protein has reduced thermal stability and twice a lower catalytic activity, comparatively to the wild type (*SULT1A1*1*) derived protein (Koike *et al.*, 2008; Xiao *et al.*, 2014). Several epidemiological studies revealed that *SULT1A1*2* isoform is associated with increased risk of lung cancer (Pachouri *et al.*, 2006; Arslan *et al.*, 2009; Liao *et al.*, 2012), urothelial cancer (Rouprêt *et al.*, 2007), stomach cancer (Boccia *et al.*, 2005), breast cancer (Lee *et al.*, 2012; Kumar *et al.*, 2013). In a study conducted by Ihsan *et al.* (2011), statistically significant lung cancer risk in *SULT1A1*2* isoforms has been observed. Therefore, the results from previous studies are conflicting and need to be corroborated further.

Microsomal epoxide hydrolase (mEH) is a phase II metabolic enzyme that catalyzes the hydrolysis of epoxide intermediates to trans-dihydrodiols so that they can be excreted from the cells. Under certain conditions, trans-dihydrodiols derived from chemical carcinogens, such as polycyclic aromatic hydrocarbons (PAHs), are highly mutagenic and toxic (Fretland and Omiecinski, 2000; Tan *et al.*, 2014). The mEH plays a dual role in tumorigenesis-detoxification and stimulating various procarcinogens, including chemical compounds and toxicological drugs (Yu *et al.*, 2015). The *mEH* gene is also known as *EPHX1* and is mapped on chromosome 1q42.

The *EPHX1* gene has 8 introns and 9 exons in its locus and encodes the proteins that vary widely in expression pattern, function, and specificity for the substrate (Skjelbred *et al*., 2011). In the *EPHX1* coding region, two specific polymorphisms are characterized in exon 3 and 4 and have been studied extensively. In exon 3, $Tyr^{113}His$ (rs2854455) results in a T to C substitution, and *His* allele is found to be associated with at least 50% (slow allele) of reduced enzyme activity whereas, in exon4, $His^{139}Arg$ (rs2234922) results in A to G substitution and $Arg^{139}Arg$ allele has been found to increase the enzyme activity by 25% (fast allele) (Hassett *et al*., 1994; Yu *et al*., 2015). Various studies have been carried out to examine the role of *EPHX1* genotypes in susceptibility to lung cancer with inconsistent results (Lee *et al*., 2002; Gsur *et al*., 2003; Kiyohara *et al*., 2006; Erkisi *et al*., 2010; Yu *et al*., 2015). The two SNP's of *EPHX1*, *viz*. rs2234922 ($His^{139}Arg$), which causes increased enzyme activity and non-functional rs2854455, were shown to be associated with increased risk of lung cancer in heavy smokers (>20 pack-years) (Timofeeva *et al*., 2010). In a meta-analysis conducted by Wang *et al.* (2013), the $T^{113}C$ polymorphism is associated with the risk of lung cancer, and there is a race-specific effect in the association.

NAD(P)H: quinone oxidoreductase 1 (NQO1) is a critical flavoenzyme which is a phase II cytosolic enzyme and is involved in catalyzing the metabolism of quinones (Mandal *et al*., 2012; Yadav *et al*., 2018). It is involved in reducing two-electrons of the carcinogenic quinones to theirreduced form, hydroquinones (Pey *et al*., 2019). The occurrence of lung cancer is implicated in the accumulation of the toxicants, thus leading to an increased reactive oxygen species (ROS) production in human tissues, resulting in up-regulation of antioxidant selective genes (Jayanthi *et al*., 2015). Thus, the role of phase II antioxidant enzymes such as NQO1 and glutathione-S-transferase (GST), which eliminate the reactive, oxidized metabolites, has played a profound role in the elimination of the carcinogens (Malec *et al*., 2010; Jayanthi *et al*., 2015). The catalyzing property of NQO1 in reducing quinones to hydroquinones can be either characterized as a detoxification or bioactivation process (Zhang *et al*., 2018; Pey *et al*., 2019). NQO1 also plays a vital role in alleviating the adverse drug reactions (ADRs) by reducing the quinine-imine metabolites produced by cytochromes (CYP) and arbitrating the protective detoxification mechanism (den Braver-Sewradj *et al*., 2018). *NQO1* is mapped to chromosome 16q22.1 and contains 6 exons, 5 introns, and is 17.2 kb in length. There are several polymorphic sites of *NQO1* that have been reported, but among the different types of polymorphism's studied,

the *NQO1609C>T* (*Pro*^{187}*Ser*, rs1800566) is the most commonly studied single nucleotide polymorphism (SNP). This polymorphism is present in exon 6 at position 609 of the *NQO1* gene (Tian *et al*., 2014; Masroor *et al*., 2015). The substitution leads to the variable enzyme activity, and thus, this variant form is volatile and can be easily denatured by the proteasome and is rapidly ubiquitinated (Siegel *et al*., 2001).Several studies have been carried out in different populations to study the correlation between *NQO1* polymorphism and lung cancer. In a meta-analysis conducted by Lou *et al*. (2013) and Huang *et al.* (2017), they observed no association between *NQO1 609C>T* polymorphism and overall risk of lung cancer in African-Americans, Hispanics, or Asian populations. Many other studies have observed that NQO1 polymorphism or its over-expression is associated with cancer progression. Thus, it might play as a possible biomarker for poor prognostic assessment in various cancers, *viz.* lung cancer (Kolesar *et al*., 2011; Cui *et al*., 2014; Li *et al*., 2015), uterine cervix cancer (Hou and Lin, 2016), breast cancer (Yang *et al*., 2014), pancreatic adenocarcinoma (ADCC) (Ji *et al*., 2017), *etc*. *NQO1* polymorphism was discovered to significantly impact the clinical outcome of NSCLC subjects who received radiotherapy following surgery and platinum-based chemotherapy (Tian *et al*., 2014; Kolesar *et al*., 2011; Song *et al*., 2010). Thus, various studies have reported a correlation between *NQO1 609C>T* polymorphism and lung cancer risk, but the results were found to be inconsistent (Sunaga *et al*., 2002; Lin *et al*., 2003; Sameer *et al*., 2010; Lin *et al*., 2014; Tong *et al*., 2017).

Methylenetetrahydrofolate reductase (MTHFR) is an enzyme that converts 5,10-methylenetetrahydrofolate (5,10-MTHF) to 5-methylenetetrahydrofolate (5-MTHF). MTHFR's role in cancer susceptibility stems from its involvement in two folate metabolism pathways: decreased MTHFR activity can reduce homocysteine methylation to methionine, lowering the level of S-adenosyl-methionine (SAM), resulting in DNA hypomethylation; and a decrease in MTHFR substrate can result in uracil misincorporation into DNA, decreased DNA repair, and an increase in the frequency of chromosomal breaks and damage (Krajinovic *et al*., 2004). The *MTHFR* gene is 2.2 kb long, situated on chromosome 1 (1p36.6), and has 11 exons. Several single nucleotide polymorphisms (SNPs) in the *MTHFR* gene have been found, including *C677T* and *A1298T*. At nucleotide 677, the C-to-T transition results in an amino acid change of valine for alanine, reducing MTHFR's activity and thermal stability (Krajinovic *et al*., 2004; Yang *et al*., 2016). The *MTHFR A1298C* polymorphism (rs1801131) in exon7, which results in the

substitution of glutamate with alanine at amino acid 429 ($E^{429}A$), may increase serum folate levels, potentially influencing the risk of cancer (Yang *et al*., 2016; Parle-McDermott *et al*., 2006). Patients with the mutant genotype had a greater chance for progression-free survival when treated with pemetrexed (Smit *et al*., 2009). It was also found that colorectal cancer patients with the mutant genotype for *MTHFR C677T* respond better to either 5-FU-based chemotherapy or FOLFOX therapy (Boige *et al*., 2010). The role of double platinum-containing regimens in treating advanced NSCLC patients is one of the gold treatment standards. The combination of platinum and pemetrexed is one of the most commonly utilized regimens in treating adenocarcinoma patients. Pemetrexed is a novel cytostatic drug effective in patients with ADCC and malignant pleural mesothelioma for first and second lines of treatment and maintenance therapy (Scagliotti *et al*., 2008; Scagliotti *et al*., 2009). Pemetrexed extends and improves the quality of life of NSCLC patients; however, this drug is also responsible for developing adverse drug reactions (ADRs) and reduced survival in certain groups of NSCLC subjects (Corrigan *et al*., 2014). In NSCLC patients receiving pemetrexed therapy, ADRs like hematological toxicity, gastrointestinal toxicity, nephrotoxicity, and hepatotoxicity have been observed (Pérez-Ramírez *et al*., 2016). Extracellular folate levels have been observed to reduce pemetrexed activity in human lung and colon cancer cell lines (Chattopadhyay *et al*., 2007). The therapeutic effect and ADRs of pemetrexed-based therapy are shown to be individually specific, explaining the genetic association between different patients and their related ADRs.

Glutathione-S-Transferase (GST) enzymes are a group of cytosolic enzymes that have been revealed to play an essential role in the metabolism and detoxification of chemotherapeutic drugs. One of the GST isoforms involved in the metabolism of anticancer drugs like platinum-based drugs is Glutathione S transferase P1 (GSTP1). GSTP1 catalyzes the binding of cisplatin to glutathione so it can be converted to a water-soluble form and hence can be easily excreted (De Luca *et al*., 2013). The gene coding for the *GSTP1* enzyme is found on chromosome 11q13 and is polymorphic. A non-synonymous A/G single nucleotide polymorphism at nucleotide position 313 in exon 5 results in the replacement of adenine to guanine (SNP's ID: rs1695), which results in the substitution of the amino acid isoleucine to valine at position 105. This substitution in the GSTP1 protein encoded results in decreased enzyme activity and reduced thermal stability and substrate specificity, thus affecting the biotransformation reactions of chemotherapeutic drugs mediated by GSTP1 (Goričar *et al*., 2015). The effect of the

*GSTP1*polymorphism can be seen during the downstream detoxification process of cisplatin and the formation of platinum-GSH conjugates. These conjugates may accumulate in the body, which then directly alter the response rates of the patients and may increase the risk of hematologic and gastrointestinal toxicity (Yuan *et al*., 2015). The previous research has suggested that the genetic variability of the *GSTP1 IIe*^{105}*Val* genes may play a role in the treatment outcomes of LC patients; however, the results are inconclusive (Booten *et al*., 2006; Ada *et al*., 2010; Yang *et al*., 2014).

The primary reason for evaluating the sequence variants of PhaseII detoxification genes (*NQO1, SULT1A1, NAT2, EPHX1, MTHFR, GSTP1*) and their cumulative effect is that there are no or very few reports available so far with regards to lung cancer. We thought it would be interesting to screen some of these genetic variabilities within the PhaseII detoxification genes for any evidence to show if the polymorphic variants within the *NQO1, SULT1A1, NAT2, EPHX1, MTHFR,* and *GSTP1* genes play a role in changing the predisposition of North Indians towards lung cancer. As far our knowledge is concerned, this is the first attempt to investigate the role of polymorphic sites of *NQO1* (rs1800566), *SULT1A1* (rs9282861), *EPHX1* (rs1051740, rs2234922), *NAT2* (rs1799929, rs1799930, rs1799931, rs1208), *MTHFR* (rs1801133, rs1801131) and *GSTP1* (rs1695) genes together. This study aimed to identify vulnerable subgroups within the sample population by stratifying the population-based on histology, clinic-pathological characteristics, gender, and smoking habits. Furthermore, we investigated whether these genes' haplotypes contribute to host vulnerability to lung cancer. The current study was also intended to investigate high-order gene-gene and gene-environment interactions that may account for the differences in lung cancer susceptibility in North Indians. Multifactor dimensionality reduction (MDR) and Classification and regression tree analysis (CART) were employed extensively to investigate the cumulative effects of all these polymorphic sites and their interactions with environmental variables such as smoking. These single nucleotide polymorphisms (SNP's) have been examined for their association with the overall survival of North Indian lung cancer patients. Also, these SNPs were screened to find their ability to act as predictive and prognostic markers. Furthermore, these genes' hematological, gastrointestinal, and nephrotoxicity were also evaluated in North Indian lung cancer patients.

CHAPTER 2

REVIEW OF LITERATURE

2.1. Lung cancer

Lung cancer is a disease that occurs when specific cells in the lungs become abnormal and grow uncontrollably, resulting in the formation of a tumor. In the early stages of lung cancer, there may be no signs or symptoms. These dysfunctional neoplastic cells have the unique property of infiltrating surrounding tissues and, as a result, experiencing metastasis and thus spreading to other organs of the body (Lung cancer-Genetics Home Reference, 2017). The human body has a respiratory system that allows for the exchange of gases. The lungs, found in the thoracic cavity, are a complex pair of organs linked to the nasal passage by a tracheal system. The tracheal system is divided further into bronchi and bronchioles, which contribute to the alveolar sacs, where the actual exchange of O_2 and CO_2 occurs (Nagaishi and Nagasawa, 1972; Figure 2.1). The mediastinum is the space between the two lungs. The presence of lymph nodes in these organs is a particular feature that increases the immunological capacity of the lungs. The lungs are positioned in the pleural cavity, surrounded by pleural fluid and bounded by the pleural membrane. The abnormal growth in the various regions of the lung tissue contributes to the development of a solid tumor, which derives all of the blood supply and inherently impairs the proper functioning of the lung tissue by invading adjacent cells and, as a result, another organ *via* lymph infiltration.

Any error or mutation in a cell's DNA might lead to the development of cancer (Loeb and Loeb, 2000). These mutations affect cells so that there is a disturbance in control over cell death and cell development, resulting in tumor growth as a mass. A tumor mass is classified either as "malignant"or "benign"(Type of tumors, 2016). Benign tumors do not spread to other areas of the body; on the contrary malignant tumors can invade other tissues and spread to other body areas after entering the blood or lymphatic system. Metastasis is the mechanism by which a tumor spreads to other parts of the body. Bones, adrenal glands, brain, and liver are the most

common areas for lung cancer metastasis. The following are the main clinical effects that are apparent in the condition of lung carcinoma:

- coughing,
- bronchitis,
- change in voice,
- shortness of breath,
- harsh sounds while breathing,
- loss of appetite,
- weakness,
- hemoptysis,
- dysphonia and
- asthenia

Lung cancer can develop in any part of the lung, but epithelial cells are responsible for 90-95% of lung cancers. Cell lines cover larger and smaller airways (bronchi and bronchioles); lung cancers are often called bronchogenic carcinomas or bronchogenic cancers. Mesotheliomas are cancers that develop from the pleura or, in extreme cases, from supporting tissues inside the lungs, such as blood vessels (Tamura *et al*., 2015).

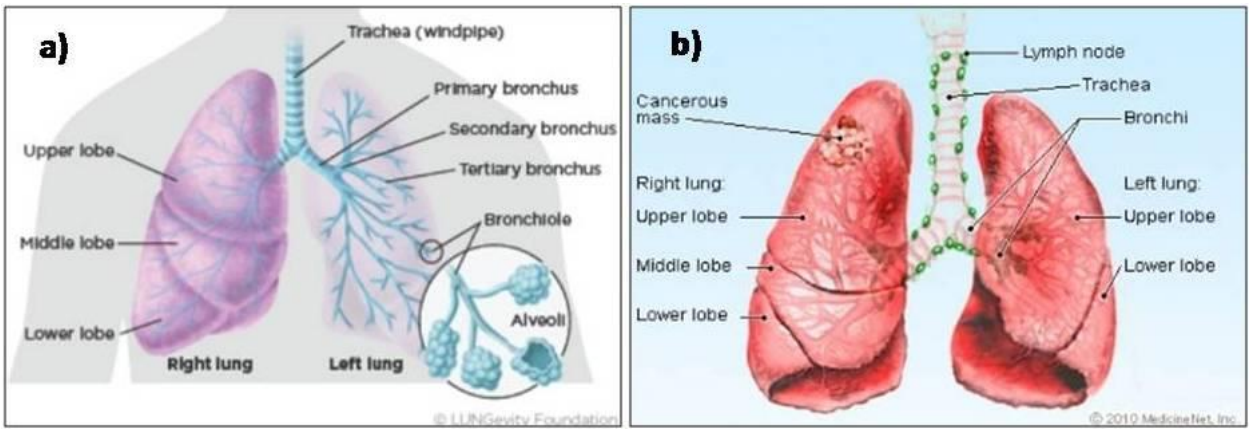


Figure 2.1: Schematic representation of Lung tissue; a) Anatomy of lung showing trachea, bronchus, bronchioles, alveoli, lobes; b) Lung tissue harboring a solid cancerous mass (Source: https://www.medicinenet.com/lung_cancer/article.htm)

2.2. TNM staging

The staging of lung tumors follows the 8th edition of the Tumor-Node-Metastasis (TNM) classification of lung cancer tumors given by the International Association for the Study of Lung Cancer (IASLC) (Lim *et al.*, 2018). The staging involves the size and frequency of the primary tumor's loco-regional invasions (T), the extent of lymph node involvement (N), and the presence or absence of distant metastasis (M). Because of the asymptomatic nature of early-stage lung tumors, most lung tumors tend to bediagnosed at advanced stages. Both pathological and clinical factors arehelpful to assess the stage of a lung tumor. The result of biopsies, imaging tests, and physical exams are appliedto determine the clinical stage and the pathological stage determined after surgery (Santos *et al.,* 2007). Staging is usually performed after a confirmed diagnosis and is made using microscopic detection of tissue by immunohistochemistry, fiber optic bronchoscopy, Positron emission tomography (PET), scan and computed tomography (CT), and guided fine-needle aspiration cytology (FNAC). This technique distinguishes applicants eligible for surgery, those with a loco-regional disease, and those with metastatic disease. The lower the staging of the tumor leads to better patient prognosis (Rami-Porta *et al.*, 2009). Table 2.1 summarizes the parameters commonly used by physicians to identify lung tumor stages. TNM classification and parameters aid in determining tumor level, and the specifics of the same are tabulated in table 2.2.

Table 2.1: TNM stage grouping (Eighth edition) (Source: Detterbeck*et al.,*2017)

Stage	**Tx**	**N0**	**M0**
Stage 0	Tis	N0	M0
Stage IA	T1a	N0	M0
	T1b	N0	M0
	T1c	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0	M0
Stage IIB	T3	N0	M0
	T1a	N1	M0
	T1b	N1	M0
	T1c	N1	M0
	T2a	N1	M0
	T2b	N1	M0
Stage IIIA	T4	N0	M0
	T4	N1	M0
	T3	N1	M0
	T1a	N2	M0
	T1b	N2	M0
	T1c	N2	M0
	T2a	N2	M0

	T2b	N2	M0
Stage IIIB	T3	N2	M0
	T4	N2	M0
	T1a	N3	M0
	T1b	N3	M0
	T1c	N3	M0
	T2a	N3	M0
	T2b	N3	M0
Stage IIIC	T3	N3	M0
	T4	N3	M0
Stage IVA	T4	N0	M1a
	T4	N0	M1b
	T4	N1	M1a
	T4	N1	M1b
	T4	N2	M1a
	T4	N2	M1b
	T4	N3	M1a
	T4	N3	M1b
Stage IVB	T4	N0	M1c
	T4	N1	M1c
	T4	N2	M1c
	T4	N3	M1c

Table 2.2: Tumor, node, and metastasis (TNM) definition (Source: Detterbeck *et al*., 2017)

Primary tumor (T)	
Tx	Tumor in sputum/ bronchial washings but not be assessed in imaging or bronchoscopy
T0	There is no primary tumor on the imaging
T1	A tumor is 3 cm or less in the most significant dimension, surrounded by lung/visceral pleura, not involving the main bronchus
T1a(mi)	Minimally invasive carcinoma
T1a	Tumor 1 cm or less in most significant dimension
T1b	Tumor more remarkablethan 1 cm but 2 cm or less in most significant dimension
T1c	Tumor more significantthan 2 cm but 3 cm or less in most significant dimension
T2	Tumor more remarkablethan 3 cm but 5 cm or less in most significant dimension, *or* involvement of the main bronchus without carina, regardless of distance from carina or invasion visceral pleural *or* atelectasis or post obstructive pneumonitis extending to the hilum
T2a	Tumor more significantthan 3 cm but 4cm or less in greatest dimension
T2b	Tumor more significant than 4 cm but five or more in most significant dimension
T3	Tumor more significant than 5 cm but 7 cm or less in most significant dimension *or* tumor of any size that involves chest wall, pericardium, phrenic nerve,*or* satellite nodules in the same lobe
T4	Tumor more remarkablethan 7 cm in most significant dimension or any tumor with invasion of the mediastinum, diaphragm, heart, great vessels, recurrent laryngeal nerve, carina, trachea, esophagus,

	spine,*or* separate tumor in the different lobe of ipsilateral lung
Regional lymph nodes (N)	
N1	Ipsilateral peribronchial and hilar nodes and intrapulmonary nodes
N2	Ipsilateral mediastinal or subcarinal nodes
N3	Contralateral mediastinal or hilar; ipsilateral/contralateral scalene/supraclavicular
Distant metastasis (M)	
M1	Distant metastasis
M1a	Tumor in contralateral lung or pleural/pericardial nodule/malignant effusion
M1b	Single extrathoracic metastasis, including single non-regional lymphnode
M1c	Multiple extrathoracic metastases in one or more organs

2.3. Histopathological classification of lung cancer

Lung cancer isa highly heterogeneous disease that is classified based uponits histopathological features.There are two types of lung cancers: small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC). The classifications of NSLC and SCLC can be evaluated by the microscopic appearance of the tumor cells, precisely the cell's size. These two forms of cancer develop and spread in different ways, and their treatment options may also differ (Travis and Harris, 2004). SCLC accounts for about 20% of all lung cancers and is the most active and rapidly developing lung cancer. SCLC is closely associated with cigarette smoking, with just 1% of these tumors occurring in non-smokers (Cruz *et al*., 2011). SCLC metastasizes quickly to several locations in the body and is usually detected after it has spread widely. NSCLC isthe most common form of lung cancer, accounting for roughly 85% of all lung cancers (Travis, 2011). Furthermore, NSCLC is further sub-divided into many groups, the most common of which are adenocarcinoma (ADCC), squamous cell carcinoma (SQCC), and large cell carcinoma (LCC). Many of the subtypes vary in terms of clinical and biological characteristics (Lung Cancer, 2019).

2.3.1. Squamous cell carcinoma (SQCC)

SQCC is a type of NSCLC that occurs in epithelial cells and is characterized by keratinization and the formation of intercellular connections between bronchial epithelial cells. Cigarette smoking is responsible for 90% of the SQCC cases (Chu and Weiss, 2002). Arsenic is another significant occupational agent regarded as the second most important causative agent for SQCC

(Spiro and Porter, 2002). SQCC develops in the central part of the lung (lobar), the mainstem, or the segmental bronchi. It all starts with a precancerous lesion known as dysplasia or squamous metaplasia in the bronchial epithelium. This tumor often has an irregular friable, grey-white cut surface with a wide area of central necrosis with or without cavity formation, as illustrated in figure 2.2 (a). This tumor is aggressive locally, but metastasis to distant organs is rare (Funai *et al*., 2003). SQCC ranges histopathologically from well-differentiated squamous cell neoplasm with keratin formation to poorly differentiated intracellular bridges with limited residual squamous cell features. SQCCs scatter locally throughout the chest using the mediastinal structure. On the other hand, poorly differentiated carcinomas metastasize to distant sites early (Funai *et al*., 2003). Histochemically, they seem to be thyroid transcription factor (TTF1) negative and CK 5, 6, P63, P40 positive. Immunohistochemical properties reveal the presence of keratin (34βE12), cytokeratins 5/6, and carcinoembryonic antigen (CEA) in SQCC cells (Chu and Weiss, 2002). The SQCC subtype exhibits a lesserfrequency of metastasisthan other histological subtypes (Quint *et al.*, 1996).

2.3.2. Adenocarcinoma (ADCC)

The term adenocarcinoma has come from the term "adeno" (referring to a gland) and "carcinoma," which refers to cancer that develops in epithelial cells. They produce mucin, which results in various growth patterns, like acinar, bronchio-vascular, and papillary. It is the most common subtype among people who never smoke or have little exposure to smoke (Khuder, 2001). Adenocarcinomas are located near the periphery of nodules rather than the central hilar or perihilar nodes. The pleural membrane and chest wall presence occurs in 15% of cases, the highest of all histological subtypes (Kondo *et al*., 2002). Cytological recognition of adenocarcinoma cells is dependent on both cell morphology and cell cluster architecture. The cells are arranged in three-dimensional acini, morulae, papillae, fibrovascular cores, or sheets. The cytoplasm inthe cells varies, but it is usually abundant and more transparent than SQCC (Walter and Pryce, 1955; Kondo *et al*., 2002). Another distinguishing feature is that the cellular boundaries of these cells are well defined. These cells have single, eccentric, and oval-shaped nuclei. There are prominent, single macronuclei circular or irregular shapes, as shown in figure 2.2 (b). Necrosis, fibrosis, hemorrhagic area, and cavitationare familiar with such tumors (Walter and Pryce, 1955). The presence of epithelial markers such as AE1/AE3, CAM 5.2, epithelial

membrane antigen (EMA), and carcinoembryonic antigen (CEA) reveals typical immunohistochemical characteristics of these cells (Colby, 1995). The immune-phenotype for cytokeratin (CK)7 and CK20 can be used to distinguish primary ADCC from metastatic ADCC. The immuno-phenotype of pulmonary adenocarcinoma tumor is distinguished because it stains positive for CK7 and negative for CK20 (Rubin *et al.*, 2001). ADCC is categorized into grades based upon how the architectural frame of tumor cells differs from that of normal lung tissue. Grade 1 is well-differentiated, grade 2 is moderately differentiated, and grade 3 is poorly differentiated (Travis and Harris, 2004).

2.3.3. Large cell carcinoma (LCC)

Large cell carcinoma is an undifferentiated type of NSCLC that lacks small cell carcinoma's cytological and architectural characteristics, glandular or squamous carcinoma. LCC accounts for almost 9% of all cases (Travis and Harris, 2004). This form of tumor has no distinguishing cytological features. Since the cell boundaries are not well established, hence no particular pattern in syncytium formation is found. The shape of the nucleus is also undefined, ranging from circular to irregular. The distribution of chromosomes is also haphazard, but a prominent nucleolus is visible. The cytoplasm is not surplus, basophilic, and has an improved nuclear to cytoplasm (N/C) ratio, as shown in figure 2.2 (c). Except for established nucleoli, some of these characteristics are similar to neuroendocrine tumors (Takei *et al.*, 2002). These tumors are large, peripheral masses that can invade a segment or large bronchi on chest radiographs. This type of tumor is invasive in the viscera, pleural wall, and surrounding areas (Silverman and Geisinger, 1996). Hemorrhage, necrosis, and cavitation are all symptoms of this tumor. Immunohistochemical markers such as synaptophysin, chromogranin, and N-CAM (CD56) are useful to distinguish this type of tumor from others (Lantz *et al.*, 1989). Since it is a poorly differentiated form of NSCLC, the molecular and genetic changes are almost identical to those redundant in other NSCLC forms. These changes include p53 mutations, K-ras mutations, cyclin D1 overexpression, and Rb signaling (Travis and Harris, 2004). The prognostic and predictive response for one form of LCC is large neuroendocrine carcinoma, almost similar to SCLC (Takei *et al.*, 2002).

2.3.4. Small cell lung cancer (SCLC)

Small cell lung carcinoma is an aggressive form of lung cancer closely linked to cigarette smoking and early spread. SCLC typically manifests in the central airways and infiltrates the submucosa, narrowing bronchial airways (Carter *et al*., 2014). Because of the flat shape of the cell and scanty cytoplasm, andreferred to as "oat cell carcinoma."Oat cell carcinomas seem to have originated in the bronchus from neuroendocrine cells (Amine Precursor Uptake and Decarboxylation (APUD) cells) called Feyrter cells (Carter *et al*., 2014). They express several neuroendocrine markers and may result in ectopic production of a hormone, such as adrenocorticotropic hormone (ACTH) and atypical ductal hyperplasia (ADH), which can cause paraneoplastic syndromes. SCLC also spreads more quickly and broadly than NSCLC (staged differently) (Underwood and Glass, 2021). SCLC cells are soft friable peri-hilar masses that are white-tan and have extensive nodal involvement. They have a scant cytoplasm with ill-defined cell boundaries, giving them a flat appearance, and they are also often referred to as "oat cell carcinoma."They have finely granular chromatin and or few nucleoli. SCLC cells may be round, oval, or spindle-shaped (Carter *et al*., 2014).

There is much necrosis in them, as shown in figure 2.2 (d). SCLC is positive for CD56, all neuroendocrine markers, and transcriptional intermediary factor (TIF) (Nicholson *et al*., 2002). The cells in SCLC are aneuploid and chromosomal imbalances are found in most SCLC populations, resulting in significant 3p depletions associated with 3q isochromosome formation. As the disease progresses, SCLC cells typically cause brain metastasis. SCLC tumors have a high propensity towards mutations in the *p53* gene (Underwood and Glass, 2021). Methylation of the caspase8 and myc amplification, an anti-apoptotic gene, are the chief characteristics of SCLC. Inactivation of the retinoblastoma (*RB*) gene and overexpression of E2F1 are found in SCLC (Nicholson *et al*., 2002).

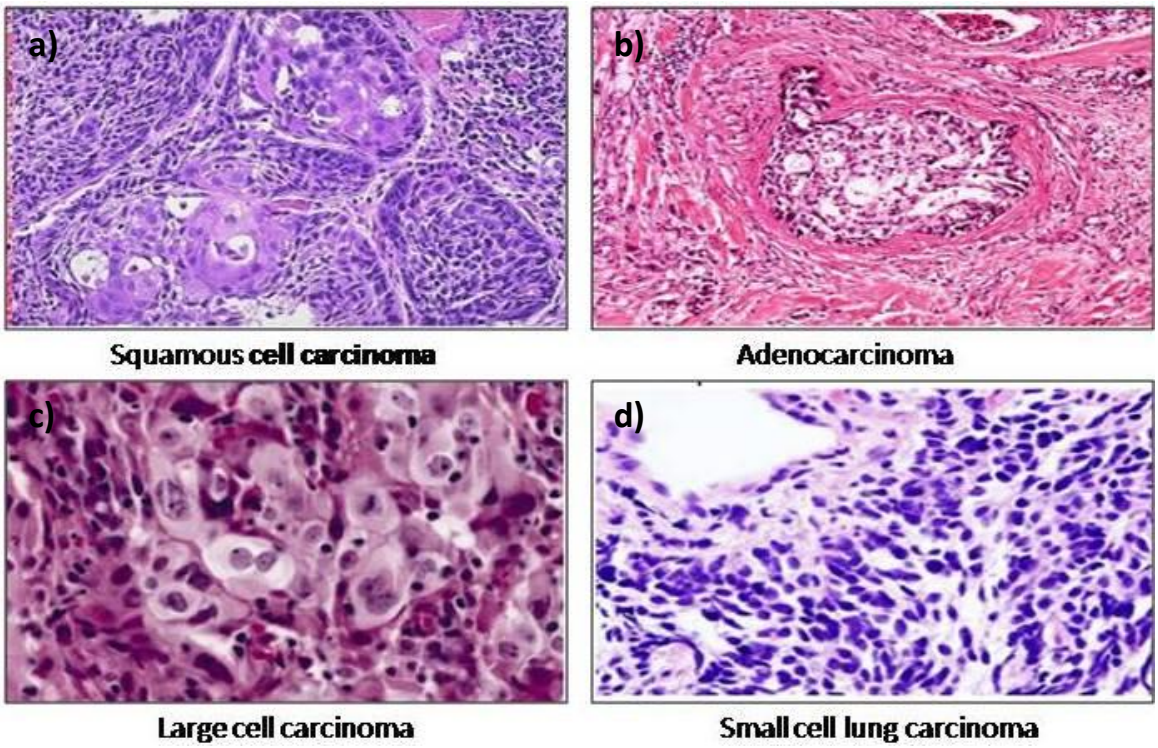


Figure 2.2: Microscopic images of different histological subtypes of lung cancer; **a)**: squamous cell carcinoma; **b)**: adenocarcinoma; **c)**: large cell carcinoma; **d)**: small cell lung carcinoma (Source:Weiss, 2010).

2.4. Risk factors for lung cancer

2.4.1. Smoking

Lung cancer is one of the most common solid tumors among tobacco-related malignancies. One of the leading causes of lung tissue neoplasm is tobacco smoking (Gandini *et al*., 2008). The risk increases with the period of smoking, which is quantified in terms of pack years. It has been reported that 75-80% of female and 90% of male lung cancer deaths worldwide are due to smoking (Ankathil, 2010; Alberg *et al*., 2013). Many epidemiological studies have found that tobacco smoke is carcinogenic (Fleiss and Gross, 1991; Sasco *et al*., 2004; Mishra *et al*., 2012). The smoke emitted by various tobacco products contains numerous potentially carcinogenic compounds such as isoprene, benzene, nitrogen oxides, butadiene, formaldehyde, styrene, acrolein, acetaldehyde, and furan in the vapor phase. Specific components present in cigarette smoke has been classified as being carcinogenic for the pulmonary tract, which includes aza-arenes, tobacco-specific nitrosamines, *e.g.*, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone

(NNK), ethyl carbonate, 1,3-butadiene, ethylene oxide, chromium, nickel, cadmium, polonium-210, hydrazine, and arsenic (Hoffman *et al*., 2001). Many compounds are found in particulate matter of smoke, including polycyclic aromatic hydrocarbons (PAH), nicotine, N-nitrosamines, metals, benzo (a) pyrene, and aromatic amines, are carcinogenic (Pfeifer *et al*., 2002). Figure 2.3 represents the scheme of how tobacco smoke carcinogens and their activation of multiple mutations in essential genes are found to correlate between nicotine addiction and lung cancer. Cigarettes, hookah, and bidis are the most commonly used tobacco products in India. Cigarettes usually contain slightly less nicotine (13.3-16.5 mg/g) than bidi (21.2 mg/g) (Malson *et al*., 2001). Individuals who consume bidi appear to inhale greater depth, increasing their susceptibility to carbon monoxide and other harmful constituents. Bidi smoking is more responsible for initiating and developing bronchial and oesophagealcancers (Kumar *et al*., 2010). It has been suggested that bidi smokers have an increased propensity towards lung cancer. Bidis, tobacco flakes wrapped in a tendu plant leaf, are common in the Indian subcontinent (Kumar *et al*., 2010; Mishra *et al*., 2012). In India, the rural population (31%) has a higher proportion of smokers than the urban population (21.5%) (Mishra *et al*., 2016). The "pack years" or the Brinkman index isused to calculate an individual's smoking status.

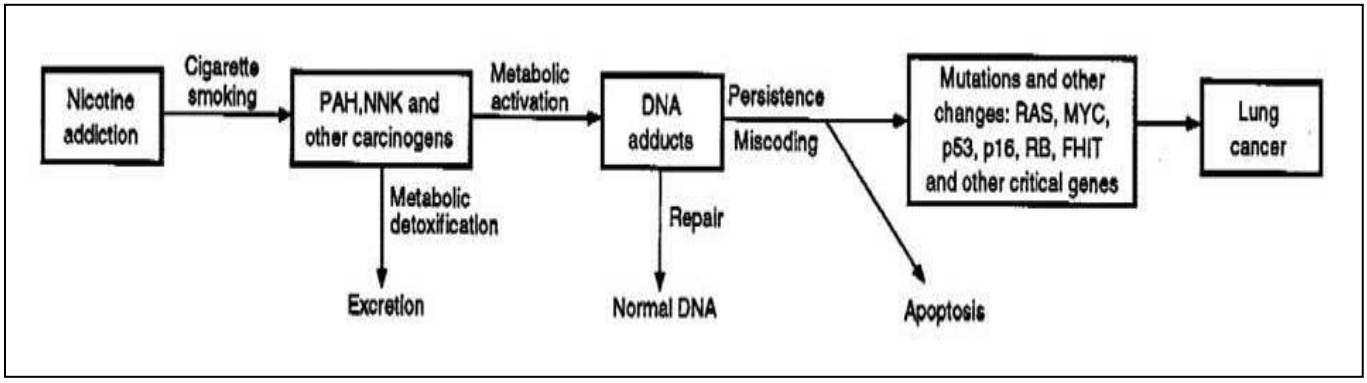


Figure 2.3: Scheme linking nicotine addiction and lung cancer via tobacco smoke carcinogens and inducing multiple mutations in critical genes. PAH 4-polycyclic aromatic hydrocarbons; NNK 4 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. (Source: Hecht, 1999)

Pack years can be calculated by multiplying the number of cigarettes/bidi smoked packs by the number of years smoked. It provides a total calculation of the smoking dose (Asami *et al*., 1997). In India, tobacco is consumed in chutta, chillum, and hukkah, in addition to cigarettes and bidi (Jindal and Behera, 1990).

Bidi is an ancient form of indigenous smoking made of raw, dried, and crushed tobacco flakes rolled in either white paper or tendu leaf that is widely used in rural India and low socioeconomic urban counterparts (Rahman and Fukui, 2000; Pandey *et al*., 2020). Bidi smoking has been linked to several cancers, including oral cancers that are similar to those caused by cigarettes (Rahman *et al*., 2003). Low socioeconomic status, illiteracy, and increasing age are all important factors in areas where bidi smoking prevails over cigarette smoking (Gupta, 2006; Jindal *et al*., 2006; Mishra *et al*., 2016). Bidi smoke is thought to have 2–3 times the nicotine and tar inhalation of cigarette smoke, increasing the risk of several cancers and fatal cardiovascular events (Rahman and Fukui, 2000; Rahman *et al*., 2003). A greater amount of bidi smoked per day and a longer duration of smoking directly increase the risk of lung cancer (Jayalekshmi *et al*., 2008). Although bidi contains one-fourth the amount of tobacco as a cigarette, the mainstream smoke of bidi contains significantly more toxic agents such as carbon monoxide, hydrogen cyanide, ammonia, other volatile phenols, and carcinogenic hydrocarbons such as benz(a)anthracene and benzopyrene (Jayant and Pakhale, 1989). Bidi is said to contain 1.5 times more carcinogenic hydrocarbons than cigarettes. Because of the tendu leaves' low combustibility and non-porous nature, smokers must take more frequent and deeper puffs to keep bidi lit, making it harder on the smoker's lungs than cigarettes rolled in paper (Prasad *et al*., 2009).

Notani and Sanghvi (1974) conducted a retrospective study of 520 lung cancer patients at Tata Memorial Hospital in Bombay discovered that the relative risk of all types of smokers to nonsmokers was 2.45, among cigarette smokers 2.23, and bidi smokers 2.65. Notani *et al*. (1977) examined the relative risks of bidi and cigarette smoking in 683 male lung cancer cases and 1,279 male non-cancer patients from Mumbai. When compared to non-smokers, the relative risk for bidis was 3.38, which was higher than the relative risk for cigarette smoking (2.36). In another study when compared to nonsmokers, all smokers appear to be at high risk (16.8). However, the relative risk in bidi smokers of 19.3 was even higher than in cigarette smokers (Jussawalla *et al*., 1979). Further, in a study conducted in Lucknow by Prasad *et al*. (1998) the bidi smokers had a 5.05 odds ratio (2.21-11.7). The authors were also able to demonstrate a significant smoking duration and dose-relationship between bidi smoking and lung cancer. The risk of lung cancer in bidi smokers of 11-20 bidis per day was seven times that of nonsmokers (7.06; 5.41-11.24), while the risk was more than ten times that of smokers of more than 20 bidis per day (10.60; 6.82-15.36). The effect of increasing bidi smoking duration is reflected in the

increasing odds ratios, which range from 3.43 (1.62-6.84) for bidi smokers of 1-39 years to 14.24 (8.34-24.31) for bidi smokers of more than 40 years. Gupta *et al.* (2001) conducted study in Chandigarh and found that bidi smoking had an odds ratio of 5.76 (3.42-9.70), while cigarette smoking had an odds ratio of 3.86 (3.42-9.70). (2.11-7.06).

A study from Bhopal (Dikshit and Kanhere, 2000) found that bidi smokers (Odds ratio (OR)=11.6 (6.4-21.3)) had a higher risk of developing lung cancer than cigarette smokers (OR= 7.7. (3.2-18.4)). They also found that the risk of lung cancer increases with the duration of cigarette smoking. In another case-control study conducted by Gajalakshmi *et al*. (2003) from Chennai and Thirvananthapuram found that more than 30 years of exclusive cigarette smoking and exclusive bidi smoking, the odds ratios were 4.54 (2.96-6.95) and 6.45 (4.38-9.50), respectively. The study also found that former cigarette smokers have a lower risk of lung cancer after quitting smoking than former bidi smokers. In a cross-sectional study of tobacco use among healthy subjects in rural Bihar, smoking prevalence was found to be 28%, with bidi smoking being the most common form (80%) (Sinha *et al*., 2003). In a study conducted by Pandey *et al* (2020) 48% of lung cancer patients and 39% of head and neck cancer patients smoked, primarily bidi, which is comparable to the smoking prevalence seen in another major epidemiological Indian study of lung cancer (48%) and higher than that seen in rural Indian oral cavity cancer (27%), respectively (Noronha *et al*., 2012; Pandey *et al*., 2016). Thus, it can be very well demonstrated that amongst all other risk factors for lung cancer bidi smoking is considered an important risk factor for the development of lung cancer.

2.4.2. Passive smoking

Second-hand smoke, also known as passive smoking, has been identified as a risk factor for lung cancer in never-smokers. Passive smoking, or inhalation of tobacco smoke by non-smokers sharing working quarters or living with smokers, is also a known risk factor for lung cancer (McCarthy *et al*., 2012). In never-smokers, 10-15% of cases of lung cancer are registered. In conjunction with other genetic and environmental causes, passive smoking is a significant contributor to the development of lung cancer in never-smokers (Noronha *et al*., 2016). The research has shown that non-smokers who live with a smoker have a 24% higher risk of developing lung cancer than non-smokers who do not live with a smoker (Samet*et al*., 2009; Noronha *et al*., 2016). Individuals inhaling secondhand smoke tend to have a 4-amino-biphenyl

adduct of hemoglobin produced by nitrosamines such as NNK found in tobacco smoke. Passive smoke has four times higher B[α]P than in the primary smoke inhaled by the smoker (de Groot and Munden, 2012).

2.4.3. Diet and Alcohol

Our health has suffered as a result of changes in our eating habits over time. Dietary preferences have shifted over time, from home-cooked meals to processed foods laden with spices and oils. Iso-thiocyanates, which are present in fruits, vegetables, and cruciferous vegetables, are responsible for lowering the risk of lung cancer (Lam *et al*., 2009). Smokers who drink more alcohol are known to be at a greater risk of developing lung cancer (Malhotra *et al*., 2016).

2.4.4. Occupational hazards

In addition to smoke, different occupational hazards are also responsible for lung cancer risks, such as asbestos, radon, crystalline silica, mixtures of PAHs, and heavy metals. Since the 1950s, several studies have established the importance of exposure to these agents in the etiology of lung cancer. The International Agency for Research on Cancer (IARC) has compiled a list of such occupational hazards, as shown in table 2.3.

2.4.4.1. Asbestos

Asbestos fibers are silicate fibers that can remain in lung tissue for a lifetime. Asbestos is found in different workplaces such as manufacturing, carpentry, insulation, and demolition. Asbestos fibers become lodged in the lung lining and might be responsible for causing mesothelioma, a form of lung cancer (Berman and Crump, 2008). However, the risk associated with different workers varies greatly depending on the degree of exposure and type of fibers. Cigarette smoking significantly raises the risk of developing asbestos-related lung cancer in asbestos workers. Asbestos workers who smoke are five times more likely to develop lung cancer than non-smokers (Reddy *et al*., 2004).

Table 2.3: List of occupational hazards (Source: IARC, 2004)

Agents, mixture, circumstances	**Main Industry, use**
Arsenic and arsenic compounds	Glass, metals, pesticides
Asbestos	Insulation, filters, textiles
Beryllium and beryllium compounds	Aerospace
Bis(chloromethyl)ether	Chemical intermediate
Chloromethyl methyl ether	Dye/pigment
Cadmium and cadmium compounds	Metal plating, dye/pigment
Chromium (VI) compounds	Chemical industry
Dioxin (TCDD)	Metallurgy, alloy, catalyst
Nickel compounds	Nuclear
Plutonium-239	Minning
Radon-222 and its decay products	Stone cutting, mining, glass, paper
Silica, crystalline	Paper, paints
Talc containing asbestiform fibers	Medical, nuclear
X- and gamma-radiation	Construction, electrodes
Coal-tar pitches	Fuel
Coal-tars	Pigments
Soots	
Exposure circumstances	
Aluminum production	
Coal gasification	
Coke production	
Haematite mining (underground) with exposure to radon	
Iron and steel founding	
Painter (occupational exposure)	

2.4.4.2. Radon gas exposure

Radon, a radioactive contaminant, was first identified as an occupational carcinogen for the respiratory tract. Radon gas is a radioactive natural gas that is a uranium decay product and releases ionizing radiation. Radon gas is a recognized cause of lung cancer, accounting for 12% of all lung cancer deaths worldwide. The alpha-particles, high energy, and high mass particles are released by inhaled radon progeny, causing unregulated division of lung tissue (National Research Council, 1999). The mechanisms underlying the impact of these particles on lung tissue include the introduction of DNA base mutations through the formation of DNA adducts and the breaking of the chromosome to affect genomic integrity (Samet, 1989). The different epidemiological studies have linked radon exposure in mine workers to the development of lung cancer. A meta-analysis review of 2,798 never smokers from 11 separate cohorts found a 3-fold increased risk in never-smokers relative to smokers. Since radon is a natural indoor air containment, it is primarily responsible for lung cancer in non-smokers (Samet *et al*., 2009).

2.4.4.3. Metals and other occupation exposures

Arsenic, a heavy metal, is another candidate whose exposure modifies the risk of lung cancer in copper smelting workers. Non-smoker smelter workers exposed to ≥15g/m^3 arsenic per year had an increased lung cancer risk (Järup and Pershagen, 1991). People exposed to arsenic, which may be present in drinking water, have a higher risk of developing lung cancer. People who deal with Chromium (IV) compounds, nickel miners, have an increased lung cancer risk (Hayes, 1997). People working in silica-related occupations such as brick-making, poultry, earth mining, and stone cutting have also reported increased lung cancer risk, with evidence of an exposure-response relationship in the high-exposure range (Steenland *et al.*, 2001).

2.4.4.4. Indoor and outdoor air pollutants

Indoor and outdoor air pollutants, such as biomass combustion and solid fuel, coal, arsenic in drinking water, and cooking gases, play a cumulative role in increasing susceptibility to this disease, alongside tobacco and environmental agents (Samet *et al.*, 2009). Exposure to particulate matters emitted by solid fuel combustion, such as benzo[a]-pyrene, benzene, and formaldehyde, increases the ability of this indoor smoke to interfere with the bronchial tract since they have an inherited capacity to form DNA adducts (Loprieno, 1975). All of the above causes might be attributed to the increasing incidence of lung cancer in females who are never smokers or non-smokers (Loprieno, 1975). Polycyclic aromatic hydrocarbons (PAHs) are formed during the combustion of organic materials. PAH are generated by coal gasification, aluminum, coke manufacturing, tar distillation, and roofing, and people exposed to high PAH levels have an increased lung cancer risk (Rota *et al.*, 2014). Coal-burning in non-ventilated homes, working on chullahs, and fumes from high-temperature cooking with unrefined vegetable oil contribute to the rising incidence of lung cancer among women (Melville and Eastwood, 1998).

2.5. Global Epidemiology of Lung cancer

Lung cancer is one of the most common forms of cancer diagnosed in the world. Cancer is the world's leading cause of death, responsible for almost 10 million deaths in 2020. There were 2.21 million new lung cancer cases reported and 1.80 million deaths due to lung cancer (WHO, 2021). Lung cancer is the leading cause of cancer death (18.0% of all cancer deaths), followed by colorectal (9.4%), liver (8.3%), stomach (7.7%), and female breast (6.9%) cancers. Lung cancer

is the most common cancer in men and the leading cause of cancer death, followed by prostate and colorectal cancer in terms of incidence and liver and colorectal cancer in terms of mortality. Breast cancer is the most frequently diagnosed cancer in women and the leading cause of cancer death, followed by colorectal and lung cancer in terms of prevalence and mortality, respectively (Sung *et al*., 2021).

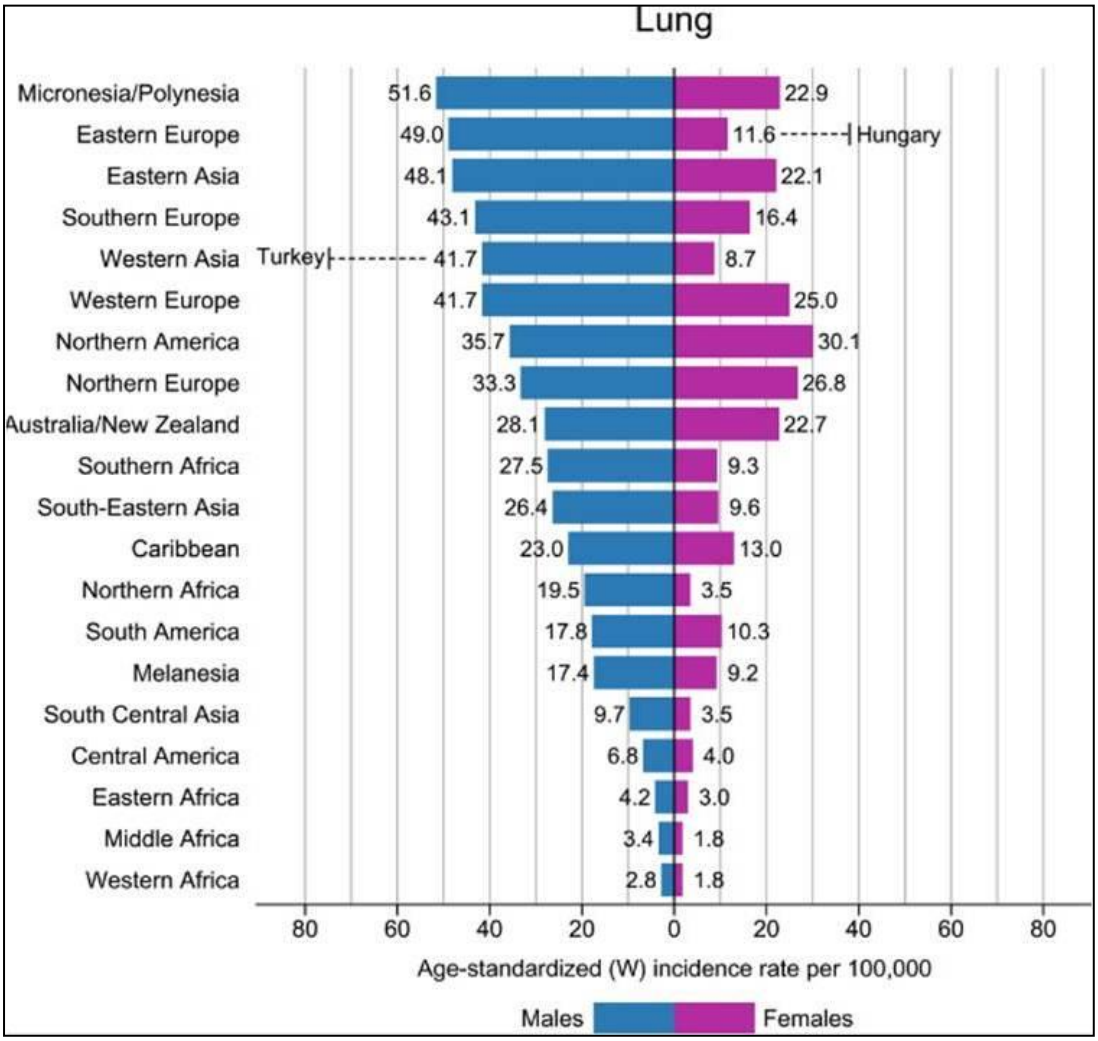


Figure 2.4: Region-Specific Incidence Age-Standardized Rates by Sex for Lung Cancer Among Men and Women in 2020. Rates areshown in descending order of the world (W) age-standardized rate in men, and the highest national rates among men and women are superimposed. (Source: Sung *et al*., 2021)

As per the Globocan survey for incidence in men, lung cancer ranks first (39 per 100,000) in higher HDI (Human Development Index) countries. Lung cancer is the most common cancer among men in 36 countries, and it is the leading cause of cancer death in 93

countries. Eastern and Southern Europe, Micronesia/Polynesia, Eastern Asia, and Western Asia have the highest incidence rates, with Turkey's highest prevalence among men globally (Sung *et al*., 2021). Lung cancer is the leading cause of cancer death among women in 25 countries across Oceania, North America, and Europe. Northern America, Micronesia/Polynesia, Australia/New Zealand, and Northern and Western Europe have the highest incidence rates, with Hungary having the highest country-specific rates (Sung *et al*., 2021). The occurrence of lung cancer differs geographically between males and females, as shown in figure 2.4. This trend showcases the historical scheme of tobacco smoking worldwide since the prevalence of lung cancer is easily associated with the evolving pattern of smoking in developed and developing countries.

2.6. Epidemiology of lung cancer in India

Lung cancer was found to be exceedingly rare in India during the early 19th century. In the early 1950s, 27.4 per million cases were registered; by 1959, the figure had risen to 78.6 per million (Behera and Balamugesh, 2004). The Indian Council of Medical Research (ICMR) launched the National Cancer Registry Programme (NCRP) with the primary goal of collecting data from all urban and sub-urban areas of the country. NCRP has developed a nationwide network of cancer registries. There are three population-based cancer registries (PBCR) in Bangalore, Mumbai, and Chennai and three hospital-based cancer registries (HBCR) in Chandigarh Thiruvananthapuram, and Dibrugarh, all of which began on January 1, 1982.

In 2012, there were 1.8 million confirmed cases of lung cancer worldwide. Lung cancer accounts for 6.9% of all new cancer cases in India and 9.3% of cancer-related deaths in both sexes (Behera and Balamugesh, 2012). It is the most common cancer in men and the leading cause of cancer-related mortality. Mizoram had the highest recorded incidences in both males and females (in males and females, the age-adjusted rate (AAR) was 28.3 and 28.7 per 100,000 population, respectively).

According to the Cancer statistics, 2020: Report from National Cancer Registry Programme, India, lung cancer has reported 44.0% of males and 47.6% of females in Manipur and Mizoram were diagnosed with distant metastasis in the majority of lung cancer patients (Figure 2.5). Lung cancer was associated with a substantial rise in 11 PBCRs in females, and annual percentage change was more in Thiruvananthapuram tulak (7.1%; 2005-2016), followed

by Nagpur (6.8%; 2005-2016) and minimum in Mizoram state (0.5%; 2004-2016) (Figure 2.6). The data states that 1 in 68 males and 1 in every 201 females might develop lung cancer during their lifetime (Mathur *et al*., 2020). The most common cancers among males in the Indian subcontinent were lung cancer and oral/mouth cancer. Lung cancer was the most common site in the southern region and metropolitan cities. Air pollution and tobacco use are the two leading risk factors for lung cancer in India. A hospital-based study in northern India found that 90% of patients with lung cancer were diagnosed at an advanced stage of the disease and that diagnosis and care were delayed (Mathur *et al*., 2020).

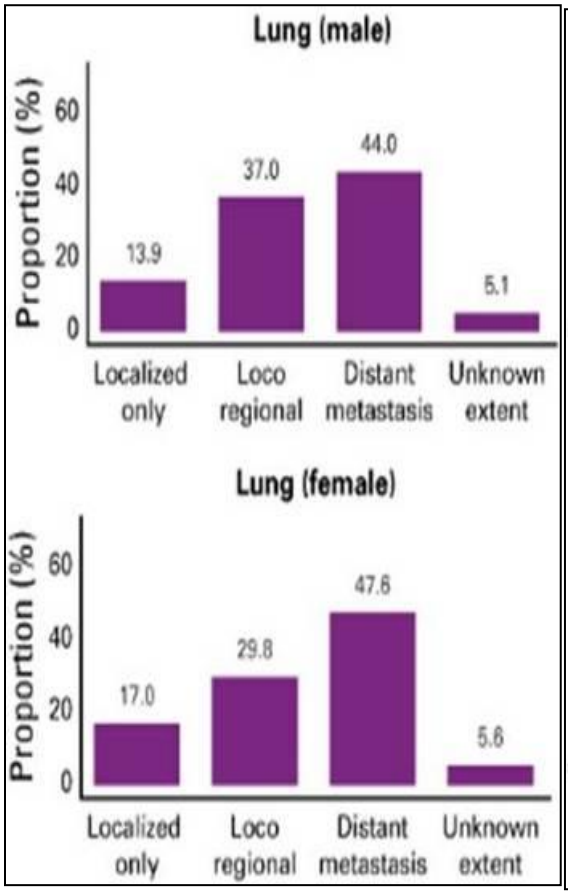


Figure 2.5: Relative proportion (%) of patients according to clinical extent of disease, 2012-2016 (proportion [%] may not total 100% because of rounding) (Source: Mathur *et al*., 2020)

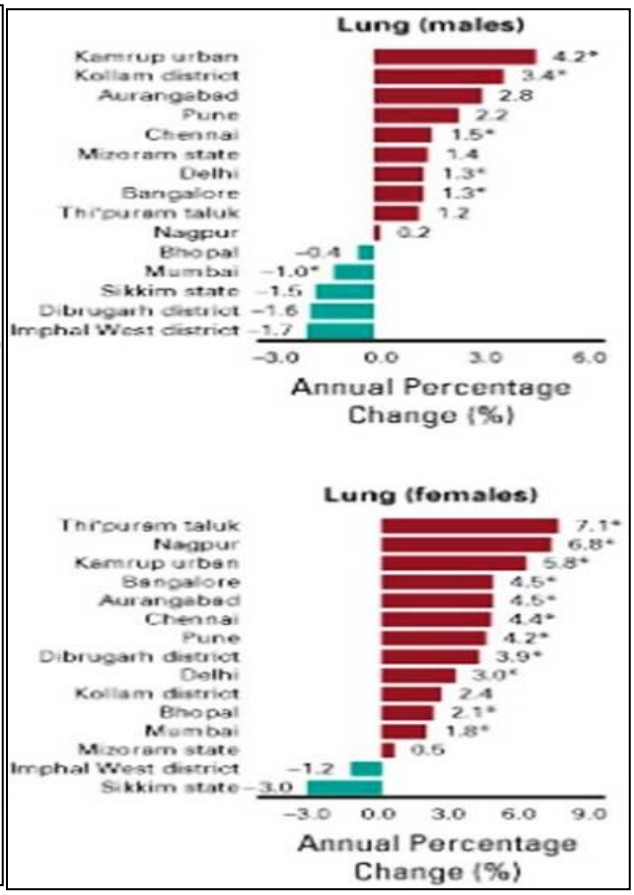


Figure 2.6: Annual percentage change (APC) in age-adjusted incidence rates (AAR) over the time period. Calendar years of incidence data for each population-based cancer registry (PBCR) used in trend analysis. (Source: Mathur *et al*., 2020)

2.7. Lung cancer pathogenesis at the molecular level

The accumulation of various alterations at both the genetic and molecular levels islinked to the pathogenicity of lung carcinoma (Travis and Harris, 2004). The most significant property gained by cells undergoing the path of hyperplasia, dysplasia, and malignant neoplasia is the capacity to proliferate without reacting to any regulatory factors, evade the cellular apoptotic pathway, overexpress the factors responsible for vascular angiogenesis, undergo unregulated replication due to telomerase hyperactivity, and eventually invade adjacent tissues and far away sites (Hanahan and Weinberg, 2000). Many of these functional differences are due to differences in the expression of genes and proteins that regulate multiple homeostasis processes.

Another distinguishing feature is that these heterogeneous lung tissue cells originate from cancer stem cells of various lineages that dedifferentiate to form these vigorously dividing cells (Nuciforo and Fraggetta, 2004). Different forms of molecular changes in cells during carcinogenesis include tumor suppressor gene dysregulation, proto-oncogene activation in response to any environmental carcinogen, loss of heterozygosity, and genomic instability caused by gain or loss of heterozygosity loss of whole chromosome or fragments. These changes can be confirmed in a tumor's karyotyping and molecular profiles (Massion and Carbone, 2003). Allelic loss of many chromosomal loci containing proto-oncogenes, such as 3p14-23, 8q21-23, 9p21, 13q, 17q, 18q, and 22p, isthe common genetic determinant seen in almost all histological subtypes of lung cancer (Panov, 2005).

2.7.1. Proto-oncogene activation

Tumor cells produce growth factors (GF) that promote autocrine and paracrine cell growth. Autocrine growth occurs in cells that contain GF and express their receptors. When GF induces growth in neighboring cells, this is calledparacrine growth. The ERBB family is a group of receptor tyrosine kinases responsible for NSCLC growth (Kirschbaum and Yarden, 2000).

The unregulated proliferative ability of lung tumor cells can be assigned to this class of growth-stimulating genes. One such example is protein tyrosine kinases, the main regulatory force for cellular pathways that regulate cell division and differentiation. They also play a role in regulating cell-to-cell communication. As a result, they provide a favorable microenvironment for cancer cell division by forming paracrine and autocrine cycles of growth-stimulating factors

(Fong *et al*., 2003). Several NSCLC tissues have shown increased expression of cyclin B (Fukuse *et al*., 2000), cyclin D1 (Keum *et al*., 1999), and cyclin E (Soria *et al*., 2000). Another type of molecular modulator known as EGFR (Epidermal growth factor receptor) is a type of tyrosine kinase that augments a ligand-activated cascade of auto-phosphorylation of mediators within cells to eventually trigger the effector molecules responsible for important events such as cell division, apoptosis, motility, and neovascularization (Huang and Chang, 2011). EGFR is found on chromosome 7p11.2 and spanning approximately 200kb, with 28 exons is stimulated by the EGFR signal transduction that occurs through the PIK3/AKT/mTOR RAS/RAF/MAPK, and JAK/STAT pathways (Cohen and Arnold, 2011). Tyrosine kinase is constitutively expressed as a result of EGFR gene mutations. Mutated EGFR is responsible for oncogenic differentiation, proliferation, neovascularization, invasion, andmetastasis. EGFR gene mutations arepresent in 40-50% of advanced NSCLC cases, especially ADCC (Kitamura *et al*., 2008). Furthermore, non-smokers are more likely to have these mutations (Kitamura *et al*., 2008). Four major mutations are present in ADCC cases and are located in exons 18-21 of the EGFR gene. The most common mutation is a frame deletion mutation in exon 19. The second most common mutation is a missense substitution in exon 21 (Kosaka *et al*., 2004). Most of these genetic modifications modify the EGFR protein's ATP binding cleft, resulting in unregulated proliferation. EGFR inhibitors, also known as tyrosine kinase inhibitors (TKIs), have gained enormous clinical significance because they compete with ATP to bind to the active site, thus blocking the active cleft and the cascade (Pao *et al*., 2005). These genetic changes are strongly correlated to prognosis and resistance to platinum-based chemotherapy. The most commonly used TKIs are gefitinib, erlotinib, and afatinib, which have a 70% objective response rate (Ansari *et al*., 2009; Wu *et al*., 2012). In a study conducted by Tamiya *et al.* (2019), the results depict that Afatinib followed by osimertinib may be a therapeutic option for NSCLC with the EGFR mutation.

Her-2/neu and MYC are two other significant proto-oncogenes that have led role to the development of lung cancer. Her-2/neu is a 185kDa transmembrane glycoprotein that shares homology with EGFR, is a member of the ERBB protein family, and is found in NSCLCs, primarily in adenocarcinomas (Fong *et al*., 2003). MYC proto-oncogenes are members of the transcription factor family, which regulates genes that control cellular proliferation and apoptosis. The nucleus contains these phosphor-proteins, activating the cascade by forming heterodimers with other proteins that regulate cellular differentiation. It also interacts with

histone deacetylase, which directly regulates the chromatin conformation that affects gene transcription (Forgacs *et al.*, 2001). Overexpression and gene amplification are two common c-myc gene abnormalities found in approximately 80-90% of SCLC cases (Rom *et al*., 2000).

The *RAS* gene is another significant proto-oncogene found to be dysregulated during lung carcinogenesis. It is involved in signaling pathways that lead to differentiation, survival, and cell proliferation. Ras is inactive and GDP (guanosine diphosphate) bound in its natural state (Forgacs *et al*., 2001). It is activated when GDP is converted to GTP and upstream growth factor receptors are activated. Ras GTP that is found to be activated binds to and activates MAPK (Mitogen-activated protein kinase), RAS/RAF/MEK/MAPK (Goodsell, 1999). With a mutation in K-RAS, the above signal cascade plays a role in lung cancer (Downward, 2003). Epithelial cells can transform after a K-ras mutation by activating ERK-MAP pathways (Massion and Carbone, 2003). In 15-20% of adenocarcinomas are found to have mutated form of K-ras gene. The most common K-ras mutations occur in codons 12, 13, and 61 (Riely *et al*., 2008). K-ras mutations with a G-T transversion are found predominantly in smokers (Graziano *et al*., 1999).

2.7.2. Tumor suppressor genes inactivation

Tumor suppressor genes (TSGs) control cellular proliferation, the response to DNA damage, and the subsequent process of DNA repair mechanism. The inactivation of anti-proliferative genes is a common feature of lung tumorigenesis. Dysfunctional TSGs result fromthe allelic loss at specific chromosomal loci, genetic alterations, or epigenetic changes, such as hypermethylation of promoter regions of these genes (Wistuba *et al*., 2000). The changes in the *p53* gene, which plays the most critical role in preserving the integrity of the human genome, are the most common redundant mutation. It performs multidimensional functions for which it has several specific features such as DNA binding proficiency and certain specific sequences that control the expression of genes that are active growth arrest phaseof the cells under genotoxic conditions that may arise due to DNA damage hypoxia switching oncogenes on. It interacts with the MDM2 protein, which is substantially responsible for cell division regulation (Oren, 2003). *Tp53* is found on the 17p13 chromosome and encodes a 53kDa protein that functions as a transcription factor (Rivlin *et al*., 2011). Carcinogenic load or damaged DNA causes p53 to dysregulate the cell cycle, causing apoptosis and, eventually, tumor formation (Zilfou and Lowe, 2009). More

than two-thirds of lung cancer cases have the *p53* gene mutated, which is likely to be used as a biomarker.

Tp53 mutations are correlated to a patient's smoking history in NSCLCs (Sholl, 2016). Missense mutations (mostly G to T transversions) clustered in the center of the gene at codons 157, 245, 248, and 273 are the most common mutations that trigger p53 protein malfunction. Cigarette smoking is related to genetic changes that result in tumor suppressor activity loss (Sholl, 2016). *TSGp53* gene is frequently found to be mutated in both SQCC and SCLC subtypes of lung cancer. *Tp53* mutations in NSCLCs have been linked to a patient's smoking history (Wistuba *et al*., 2002).

Retinoblastoma (Rb) protein (13q14) is the next essential TSG involved in lung carcinogenesis. *Rb* collaborates with *p53* at the transcriptional level in cell cycle progression and maintaining cell differentiation and proliferation balance. Non-sense mutations and chromosomal deletions are examples of natural gene behavior seen in lung cancer (Forgacs *et al.*, 2001). *Rb* gene is overexpressed in 90% of SCLC cases and 20% of NSCLC cases. *Rb* controls the cell's transition from the G1 to S phase of the cell cycle (Sholl, 2016).

PTEN (Phosphatase Tensin Homolog Deleted on Chromosome Ten) is a tumor suppressor gene located on chromosome 10q23. It codes for a lipid phosphatase that mediates the removal of the phosphate moiety from (Phosphatidylinositol (3,4,5)-trisphosphate) PIP_3 and has an anti-proliferative effect (Forgacs *et al.*, 2001) and Transforming growth factor (TGF-β), which has multiple functions as a protein wherein, it inhibits the proliferation of several epithelial cells by binding to a series of cell receptors. Immunohistochemical staining revealed that this protein was lower in NSCLC tissues (Panov, 2005).

2.7.3. Escaping the programmed cell death

Apoptosis or programmed cell death has been the essential metabolic tool for controlling the repercussions of exposure to carcinogens, DNA damage, and necrotic cells. The inherent mechanism of the cells to undergo the apoptotic pathway maintains the homeostasis of the human body. Apoptotic pathways are found to be dysregulated whenever a tumor formation initiates (Panov, 2005). The *Bcl-2* (*B-cell lymphoma-2*) gene was the first oncogene discovered to act by producing an apoptosis inhibitor. The *Bcl-2* gene family contains over 15 members that

facilitate or inhibit apoptosis (Reed, 1997). Bcl-2 overexpression, as observed by immunohistochemistry, was found in 75-90% of SCLC, 25-30% of SQCC, and 10% of ADCC tumors (Kaiser *et al.*, 1996). Figure 2.7, based on previous studies, shows the pathogenesis of lung tissue.

Tobacco and other toxins cause the removal of the 3p21.3 allele in several epithelial cells of the respiratory system at different locations. Furthermore, chromosomal failure of the tumor suppressor gene 3p21.3 occurs, rendering it haplo insufficient. The succeeding sequence of events is defined by inactivating growth-regulating genes such as *p53*, *Rb*, or *p16* due to mutation or epigenetic silencing, thus allowing the clonal proliferation of initially transformed cells. A hypothesis shows that during the pathogenesis of SCLC, neoplastic cells emerge directly from either standardor hyperplastic epithelial cells without going through typical preneoplastic intermediate pathological stages (parallel theory of lung cancer pathogenesis). The pathogenesis of NSCLC, on the other hand, is followed by sequential morphological changes (sequential theory) (Panov, 2005). Some oncogenic driver mutations have gained clinical significance and are now translationally used in the therapeutic regimen of ADCC patients alongside the ongoing chemotherapy, which primarily consists of platinum-based drugs. These potential targets have played a significant role in improving the average survival time of lung cancer patients, especially ADCC patients (Sholl, 2016).

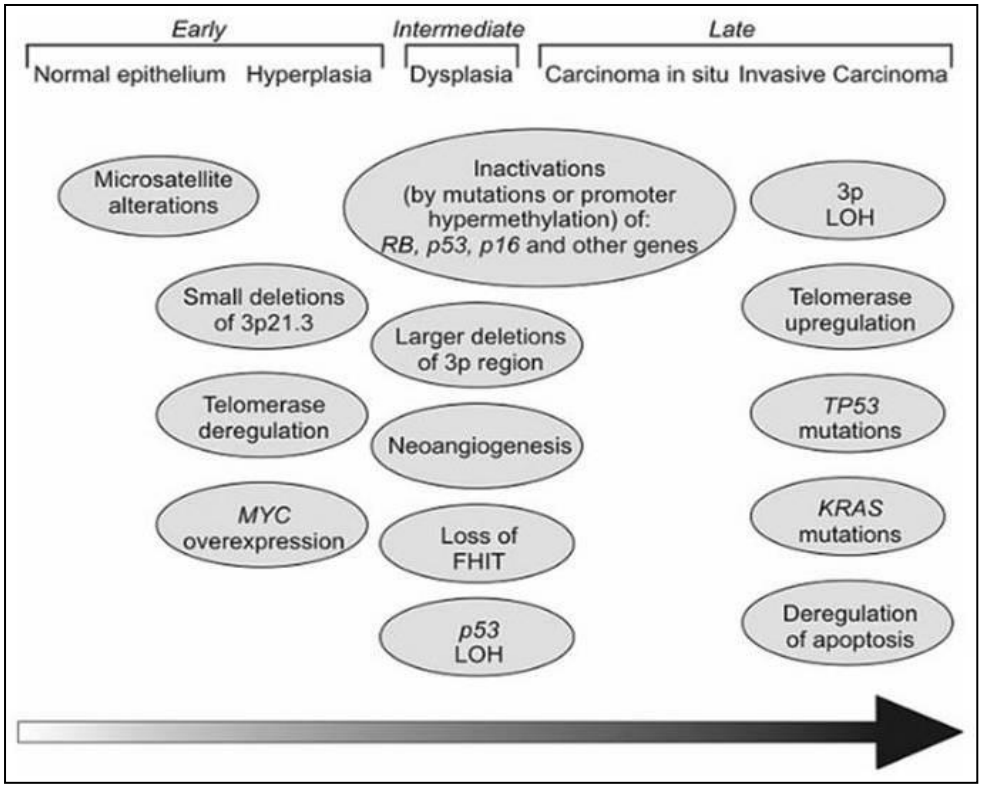


Figure 2.7: Schematic representation of significant molecular abnormalities taking place during lung cancer pathogenesis. (Source: Panov, 2005)

2.8. Diagnosis of lung cancer

At the time of initial diagnosis, lung cancer is typically symptomatic. The primary diagnostic protocol involves a review of the patient's history and a preliminary physical examination. After a physical examination, the series of tests include complete blood count, spirometry, measurement of alkaline phosphatase and hepatic transaminase, calcium levels, and determination of physiological parameters such as electrolytes, creatinine, and urea concentration in blood and urine, and chest radiography. Following the initial screening, contrast-enhanced computed tomography (CT) accompanied by positron emission tomography (PET) to determine the stage of the tumor according to the 8th updated edition of IASLC and the exact status of the tumor's distribution across the lung tissue is performed to confirm the existence of the tumor, which may be lying beneath the mediastinum or at some other location within the lungs. The subsequentstep is to confirm the lung tumor's histology using molecular identification techniques such as immuno-histochemical characterization. A portion of lung tumor tissue is required to

perform this molecular testing. The method used for tissue extraction is influenced by the position and size of the tumor and its metastatic accessibility. Fiberoptic bronchoscopy is used to diagnose centrally located lesions, whereas CT-guided transthoracic needle aspiration is standard for lesions near the periphery. Modern diagnostics use endobronchial ultrasound and electromagnetic navigation (Latimer and Mott, 2015). In developing countries, there are a large number of pulmonary tuberculosis cases, with India reporting an incidence of 204 per 100,000 cases in 2017 (Tuberculosis country profile). It's always difficult to develop an effective lung cancer screening method programme in India due to a high number of cases of tuberculosis of the lungs and other chest infections (Shankar *et al*., 2019).

2.8.1. Screening and early detection of lung cancer

Early detection of lung cancer is defined as strategies that detect lung cancer at a stage where surgery or stereotactic body radiation therapy (SBRT) with the goal of curing can be offered. Aside from when the lung cancer is detected, a number of factors influence the outlook for any lung cancer patient, including the type of lung cancer, the patient's overall health, and the responsiveness of the patient's lung cancer to treatment. LC, on the other hand, is most easily and effectively treated when discovered at an early stage. Statistically, those with lung cancer who are diagnosed early have a much better chance of surviving at least five years after diagnosis than those who are diagnosed later in the disease's progression (SEER Stat Fact Sheets). Screening can detect lung cancer at an early stage. Screening for lung cancer entails looking for cancer before any symptoms appear. Mammograms for breast cancer, colonoscopies for colon cancer, and the PAP smear and human papillomavirus (HPV) tests for cervical cancer are examples of commonly used cancer screening tests (Guidelines for the Early Detection of Cancer). Early-detection screening for lung cancer with a low-dose computed tomography (LDCT) scan (also known as a low-dose spiral CT scan or a helical CT scan) has been shown to be effective in individuals considered to be at high risk. It is also only advised for these people. The National Lung Screening Trial (NLST) and the Nederlands-Leuvent Longkanker Screenings Onderzoek (NELSON) trials both demonstrated the utility of LDCT in high-risk individuals. LDCT is far more sensitive than chest X-rays in detecting smaller abnormalities that may be lung cancer. Chest X-rays are not recommended for lung cancer screening because they frequently miss early-stage lung cancers and have not reduced mortality. This is also true for sputum

cytology, a test that looks for abnormal cells in sputum (Lung Cancer Screening). There are many trials on LDCT for lung cancer screening being done or ongoing in many developed and developing countries (Shankar *et al*., 2019).

A study conducted in India, where 350 smokers were scanned with LDCT in a retrospective study (Parang and Bhavin, 2021). The average age was 61 years (32–90 years), and the average smoking history was 22 pack-years (1–160 pack-years). Lung nodules were found in 335 (93%) smokers, with Lung-RADS category 1 nodules seen in 117 (36%) of positive scans, category 2 in 133 (41%), category 3 in 29 (9%), and category 4 in 46 (14%). Seven of the seven category 4 patients who underwent biopsy had carcinoma, with a smoking history of 30 pack years on average. The scans also revealed superimposed interstitial lung disease, airway diseases, and infections. In a tuberculosis-endemic country like India, smokers with a history of more than 20 pack years have a higher incidence of lung nodules, which may represent carcinoma, as has been demonstrated in other Western countries. Larger studies and trials may be conducted to expand on the findings of this proof-of-concept study.

2.8.2. Lung cancer oncopanels

Multiplex hotspot sequencing, including mass spectrometry–based sequencing methods6, and multiplex single-base pair extension sequencing were previously used to detect mutations across multiple genes (Su *et al*., 2011). Next-generation sequencing (NGS) technology has increased the number of genes and types of mutations detectable by a high-throughput assay, and multiple laboratories have successfully validated NGS tests to detect somatic alterations in cancer (Frampton *et al*., 2013; Cottrell *et al*., 2014; Cheng *et al*., 2015). In a study conducted by Garcia *et al*. (2017) the OncoPanel was created to detect single-nucleotide variants, insertions and deletions, copy number changes, and structural variants in 282 genes implicated in cancer biology. A validation strategy that compared results from clinically validated orthogonal technologies to formalin-fixed, paraffin-embedded, fresh or frozen samples was used. When compared to single gene assays and mass spectrometry–based genotyping, OncoPanel achieved 98% sensitivity and 100% specificity for single-nucleotide variant detection and 84% sensitivity and 100% specificity for insertions and deletions. OncoPanel's analytical validation demonstrates the ability of targeted next-generation sequencing to detect multiple types of genetic alterations across a panel of cancer-related genes.

The PMDA approved three NGS assays in June 2019: the Oncomine Dx TT, the FoundationOne CDx (F1CDx), and the OncoGuide NCC OncoPanel (NCC Oncopanel). The first two are well-known throughout the world, but the NCC-OP is unique to Japan. The National Cancer Center and Sysmex Corporation collaborated to develop the NCC Oncopanel, a comprehensive genomic profiling test (Center NC). The test covers 114 genes and 12 fusions and is distinguished by simultaneous blood sampling, which is used to address accurate somatic and germline mutations as well as tumour mutation burden. In terms of the ratio to lead to CGP-guided treatment, the report of the first 230 cases analysed with this panel (Sunami *et al*., 2019) was similar to those of the other panels (Sholl *et al*., 2016; Zehir *et al*., 2017). Based on the number of genes and the need for genetic consultation for germline mutations, the PMDA divided panel tests into two categories: conventional IVD and comprehensive genomic panel (CGP) tests. The F1CDx and NCC Oncopanel are CGP tests because they examine over 100 genes and have the potential to suggest or detect germline mutations. The Japanese Ministry of Health, Labor, and Welfare (MHLW) organises a nation-wide network for cancer genome medicine, with 11 core institutes, 34 facilitating, and 122 affiliated hospitals designated for CGP testing.

Oncomine DxTT is classified as a hot spot panel test that uses the amplicon method to analyse mutations in 46 genes and fusions in 21 genes, with *EGFR*, *ALK*, *ROS1*, and *BRAF* approved as companion diagnostics. Although this test can detect *KRAS, NRAS, MET* exon 14 skipping, *FGFR*, *KIT*, *RET, PIK3CA*, and *NTRK*, it is only reimbursed for patients with lung cancer. The FoundationOne CDx Cancer Genome Profile (F1CDx) is a CGP test. Only DNA extracted from tumour tissues was examined. This test covers 324 genes in total, with access to base substitution, insertion, deletion mutation, and copy number alteration in 309 genes, gene fusion in 36 genes, microsatellite instability (MSI), and tumour mutation burden (TMB). OncoGuide NCC OncoPanel (NCC OncoPanel) is another CGP test. This test looks for base substitutions, insertions/deletions, mutations, and gene amplification in 114 genes, as well as fusion in 12 genes and TMB. This test is distinguished by the simultaneous analysis of blood DNA, which allows for the complete exclusion of uncommon polymorphisms and the definitive determination of germline mutations in 13 genes (Yatabe *et al*., 2020).

Oncomine (testing EGFR, BRAF, and ROS1), MSK IMPACT (Integrated Mutation Profiling of Actionable Cancer Targets), and Foundation One CDx are FDA-approved molecular testing platforms based on next-generation sequencing (NGS). Broader NGS-based assays can evaluate emerging biomarkers in a single test, allowing for lower costs, more patients to be tested, and fewer recurrent biopsies (Nooreldeen and Bach, 2021). Several biomarkers are emerging as tool for early detection. Serological biomarkers are suitable for mass screening because they can be analysed easily and cheaply. Carcinoembryonic antigen (CEA) and serum cytokeratin 19 fragments are currently available serologic biomarkers for NSCLC (CYFRA 21-1) (Cho, 2007). Several antigens found in blood have been studied as potential biomarkers of lung cancer over the years. CYFRA 21-1, carcinoembryiogenic antigen (CEA), neuron specific enolase (NSE), and squamous cell carcinoma antigen are among the most studied biomarkers (SCC-Ag) (Nooreldeen and Bach, 2021).

2.9. Treatment of lung cancer

Cancers that have spread only locally are better treated with a local procedure that removes or kills the entire tumor. Local care includes surgery and radiation. Surgery is usually more effective than radiation in removing tumors, but not all patients can tolerate surgery safely. Chemotherapy can be administered before or after local procedures, depending on the spread of cancer cells to other organs. Adjuvant therapy refers to chemotherapy administered following surgery andis provided to help destroy any cancer cells that have not been removed by surgery (Ellis *et al*., 2008). When chemotherapy is administered before surgery, it is referred to as neoadjuvant therapy. Chemotherapy for NSCLC usually consists of a platinum drug (either cisplatin or carboplatin) combined with a second drug (paclitaxel, docetaxel, gemcitabine, or vinorelbine). Following first-line chemotherapy, a few medications, such as pemetrexed, are approved for second-line therapy of NSCLC. Radiation therapy is used routinely in patients with limited-stage cancer, i.e., a tumor restricted to the chest. In limited-stage patients, radiation therapy to the brain is widely used to avoid brain metastasis (Behera, 2007).

2.9.1. Platinum-based chemotherapy

2.9.1.1. Docetaxel + cisplatin/ carboplatin

Docetaxel is an anti-mitotic chemotherapy drug commonly used to treat breast, ovarian, and NSCLC cancers. It binds to microtubules reversibly and with high affinity. Docetaxel is a semi-synthetic taxoid derived from 10-diacetyl vacation III, which results in microtubule assembly and inhibition of tubulin depolymerization, resulting in microtubule stabilization (Aihara *et al*., 2002). This medication has a 23-33% treatment success rate and a 21% objective response rate in NSCLC patients and new cancer cases. Cisplatin, which has a different mechanism than docetaxel, is often commonly used in the care of NSCLC patients. According to studies, cisplatin-based chemotherapy in NSCLC has a higher response rate and survival rate than regimens that do not contain cisplatin (Donnadieu *et al*., 1991). As a result, the use of cisplatin is known as an independent indicator of enhanced survival (Aihara *et al*., 2002).

Mechanism of action

Cisplatin inhibits DNA replication, halting the cell proliferation of carcinogenic cells. Aquation occurs when one of the two chloride ligands is displaced by water, resulting in the aquo-complex cis-$[PtCl(NH_3)_2(H_2O)]^+$. The N-heterocyclic bases on DNA displace the aqua ligand in cis-$[PtCl(NH_3)_2(H_2O)]^+$. Guanine preferentially binds subsequently, resulting in the formation of $[PtCl(\text{guanine-DNA})(NH_3)_2]^+$, which interferes with mitosis through DNA crosslinking. When the damage cannot be repaired, the damaged DNA triggers the DNA repair mechanism, which causes apoptosis. While cisplatin is commonly referred to as an alkylating agent, it lacks an alkyl group and cannot perform alkylating reactions (Yvon *et al*., 1999). When combined with another chemotherapy drug, cisplatin has become a mainstay in the treatment of many cancers. While platinum responsiveness is high at first, most cancer patients ultimately relapse due to cisplatin resistance. This resistance is manifested through various mechanisms, including increased drug uptake and efflux, increased drug detoxification, inhibition of apoptosis, and increased DNA repair.

Docetaxel promotes and stabilizes microtubule assembly by preventing microtubules depolymerization and stabilizing microtubule assembly in the absence of GTP (Williamson *et al*., 2005). As a result, the number of free tubulins necessary for microtubule formation

decreases, inhibiting mitotic cell division between metaphase and anaphase and preventing cancer cell progeny (Yvon *et al*., 1999). Since microtubules do not disassemble in the presence of docetaxel, they accumulate within the cell, causing apoptosis to begin. Apoptosis has been aided by the inhibition of the apoptosis-inhibiting Bcl-2 oncoprotein. *In vitro* and *in vivo* studies show that docetaxel has anti-neoplastic activity against many identified cancer cells, cooperates with other anti-neoplastic agents, and has higher cytotoxicity.

2.9.1.2. Irinotecan + cisplatin/ carboplatin

Irinotecan (Camposter) is a camptothecin-derived compound in which the active compound SN-38 interacts with topoisomerase-I to form a DNA stabilizing complex. In addition to its novel cytotoxic mechanism, Irinotecan differs from other chemotherapeutic agents in the multidrug-resistant cell lines that maintain irinotecan sensitivity. It is an authorized drug in the United States for the treatment of metastatic colon cancer. It is either injected alone or in combination with fluorouracil for colon cancer treatment. It is also used in the treatment of SCLC in conjunction with cisplatin. When used as a single agent, the response rate in NSCLC can range between 15-34% (Pillot *et al.*, 2006).

Mechanism of action

Irinotecan is a prodrug that, once within the cell, is metabolized into its active form, SN-38. Multidrug transporters such as ABCB1 may transport irinotecan and SN-38 out of the cell. Irinotecan is converted to SN-38 by human carboxylesterases. However, it can also be converted to inactive metabolites by CYP3A4 and CYP3A5, and UGT1A1 can inactivate SN-38 through glucuronidation to SN-38G (Artac *et al*., 2010). Both DNA replication and transcription are inhibited as a result of this inactivation. The UGT1A1*28 allele provides the most evidence for the role of pharmacogenetics in predicting irinotecan toxicity (Fukuda *et al*., 2004).

2.9.1.3. Pemetrexed + cisplatin/carboplatin

Pemetrexed is an anti-metabolite inhibitor that inhibits thymidylate synthase (TS) and other folate-dependent enzymes. It is primarily used to treat NSCLC and pleural mesothelioma (Metro *et al*., 2011). Corticosteroids and vitamin B12, and folic acid supplementation minimize its toxicity. Fatigue, nausea, vomiting, mucositis, and rash are the most common side effects.

Pemetrexed is approved as second-line therapy for NSCLC (Tomasini *et al*., 2016). for Treatment of NSCLC patients, pemetrexed is typically given in conjunction with the platinum-based drug carboplatin.

Mechanism of action

Pemetrexed is a chemotherapy drug that is chemically identical to folic acid. It belongs to a class of chemotherapy drugs known as folate antimetabolites. Pemetrexed inhibits three enzymes that are found to be involved in purine and pyrimidine synthesis,*viz*. dihydrofolate reductase (DHFR), thymidylate synthase (TS), and glycinamide ribonucleotide formyltransferase (GRAFT). Pemetrexed inhibits DNA and RNA formation by inhibiting the formation of purine and pyrimidine nucleotides, which are typically necessary for the survival of both standard and cancer cells (Metro *et al*., 2011).

2.9.2. Targeted therapy and immunotherapy

Unlike chemotherapy, targeted therapy targets the molecules expressed on the surface of cancerous cells. The therapy targets these targets by either blocking or killing them. Anti-angiogenesis therapy, also known as anti-VEGF therapy, restricts blood supply to tumors to prevent their development and eventually starve them to death (Al-Abd *et al*., 2017). Angiogenesis in tumors is caused by vascular endothelial growth factor (VEGF). VEGF facilitates cell survival while also inhibiting apoptosis. It is a growth factor that controls vascular endothelial cell permeability. VEGF is produced by cancer cells. Anti-angiogenic compounds are thought to either restore natural blood flow within the tumor or enhance treatment efficacy. Bevacizumab is a VEGF inhibitor that the FDA has approved for the treatment of colon cancer. Erlotinib, Gefitinib, and Afatinib are EGFR inhibitors that stop the growth of cancerous cells that are EGFR positive. EGFR promotes the growth and development of cancerous cells (Cabanero *et al*., 2017). It was approved recently for the treatment of advanced NSCLC combined with carboplatin/paclitaxel, demonstrating a two-month improvement in overall survival over chemotherapy alone (Wu *et al*., 2012).

Immune checkpoint inhibitors

Immunotherapy, which uses immune checkpoint inhibitors such as Nivolumab and Pembrolizumab, is another forthcoming treatment. Immunotherapy is now recommended as a first-line treatment. Checkpoint immunotherapy employs highly selective humanized monoclonal antibodies directed toward checkpoint signals such as programmed cell death receptors and ligands (Darvin *et al*., 2018; Desnoyer *et al*., 2020). The immune system can be reactivated to fight the tumor by blocking these receptors and signals. Immunotherapy for advanced NSCLC has resulted in a new model of treatment choices, with increased survival and response rates and a less severe yet distinct toxicity profile compared to chemotherapy. Regulatory authorities have currently approved the programmed cell death ligand inhibitors nivolumab and pembrolizumab and the programmed death ligand-receptor inhibitor atezolizumab for the treatment of advanced NSCLC as SCLC (Leventakos and Mansfield, 2016; Xu *et al*., 2019). Many biomarkers can be used to assess the probability of response to anti-PD-L1 inhibitors in NSCLC and SCLC patients. Checkpoint inhibition (*i.e.*, antibodies targeting adaptive immune suppression pathways such as PD-L1, PD-1, and CTLA-4) is a novel anti-tumor strategy that appears to be clinically relevant across many tumor types (Pardoll, 2012; Buchbinder and Desai, 2016). Furthermore, the prolonged clinical response observed due to memory T-cell responses can provide long-term immunity. Furthermore, based on improved understanding of the biological pathways underlying cancer-immune interactions, a wave of novel immunotherapeutic procedures targeting activating and inhibitory T cell receptors, chimeric antigen receptors (CARs), adoptive cell therapies (ACT) involving tissue-infiltrating lymphocytes (TILs), and TCR-modified T cells, and specific antibodies are now being tested in clinical trials (Phan and Rosenberg, 2013; Garfall et al., 2014; Gastman et al., 2020). Two chimeric CAR-T cell drugs have recently been approved in the United States to treat patients with refractory hematologic malignancies (specifically acute lymphoblastic leukemia and diffuse large B cell lymphoma). These biomarkers include PD-L1 expression, microsatellite instability, and the number of tumor mutations. For Pembrolizumab indication, PD-L1 expression (Tumor Proportion Score, i.e., TPS >1%) is a criterionin ADCC NSCLC patients. Another significant milestone is the inclusion of tumor mutation burden as an emerging market in guidelines due to the Checkmate 227 trial, which suggests that ten mutations per megabase will better respond to nivolumab or ipilimumab and nivolumab (Prelaj *et al*., 2019).

Ipilimumab is a monoclonal human antibody that activates the immune system by CTLA-4 targeting. CTLA-4 is a protein receptor that is involved in the downregulation of the immune system. Generally, cytotoxic T lymphocyte (CTL) is able to recognize as well as destroy tumor cells, but inhibitory mechanism hinders this destruction process; therefore, ipilimumab switches off this inhibitory mechanism, thus boosting the body's immune response against tumor cells (Reck *et al*., 2013). Nivolumab is a monoclonal human IgG4 antibody that is targeted against PD-1. Pembrolizumab is a humanized IgG4 antibody, which targets PD1 and has also been studied in SCLC. Atezolizumab is an IgG1 antibody, which is a humanized monoclonal antibody that binds to PD-L1 and blocks the interactions with receptors, PD-1 and B7-1. Durvalumab is a human IgG1 monoclonal antibody (mAb) that targets PD-L1, whereas tremelimumab targets CTLA-4 (CD152) and is a fully human IgG2 mAb (Walia *et al*., 2022). Lefitolimod (MGN1703) is a Toll-like receptor 9 (TLR9) agonist, which is DNA-based and expressed in the dendritic cells, inducing immune surveillance that induces IFN-α secretion and thus stimulates NK cells, monocytes, NKT cells, and T cells (König-MeredizSven, 2015). Great interest has been developed by the inhibitory Notch ligand Delta-like protein 3 (DLL3) (George *et al*., 2015; Sabari *et al*., 2017). Many clinical trials test antibodies that target DLL3 (an atypical family ligand of Notch receptors), which are expressed on the surface of the cell of lung cancer and other neuroendocrine cancer. Tesirine (Rova-T) is an antibody-drug conjugate that consists of the humanized DLL3-specific IgG1 monoclonal antibody which is associated to an internalized toxic DNA cross-linking agent that stimulates cell death (Saunders *et al*., 2015).

The other upcoming therapy is anti-ganglioside therapy which involves fucosyl GM1, GD2 ganglioside, and anti CD47. BMS-986012, a first-class human monoclonal IgG1 antibody with improved antibody-dependent cell-mediated cytotoxicity (ADCC), has higher specificity and affinity for fucosyl GM1 and has shown preliminary activity in SCLC (Chu *et al*., 2016). Dinutuximab binds to disialogangloside (GD2) and is a monoclonal GD2-binding human/mouse chimeric Ab. It is expressed on the cell's surface, is tumor-associated Ag, and also induces cell lysis via ADCC and complementary cytotoxicity (Mueller *et al*., 1990; Navid *et al*., 2010). CD47 axis targeted therapies have shown efficacy in preclinical models and are currently being investigated for both hematological and solid malignancies in clinical studies (Weiskopf, 2017).

2.10. Genome-Wide association study (GWAS)

A genome-wide association study is a method of rapidly scanning markers across the entire sets of DNA, or genomes, of many people in order to find genetic variations associated with a specific disease. Once new genetic associations are discovered, researchers can use the information to develop better strategies for disease detection, treatment, and prevention. These studies are especially useful in identifying genetic variations that contribute to common, complex diseases like asthma, cancer, diabetes, heart disease, and mental illnesses. GWAS that used single-locus models to test each single nucleotide polymorphism (SNP) for association with a phenotype proved useful in identifying thousands of genetic variants linked to human traits and disorders (Manolio *et al*., 2009; Hirschhorn and Daly, 2005; McCarthy *et al*., 2008; Bush and Moore, 2012). However, the majority of the findings only explained a small portion of the genetic effects on diseases and traits (Maher, 2008; Manolio *et al*., 2009). Researchers were motivated to study not only the main additive effects of single genetic variations, but also interactions between multiple variants with non-additive effects to explain more of the heritability of complex diseases (Moore and Williams, 2009; Park and Hastie, 2008; Andrew *et al*., 2009; Luyapan *et al*., 2020). As smoking is the main contributing risk factor towards lung cancer risk, so we want to access the role of detoxification genes towards lung cancer susceptibility and chemotherapy based toxicity.

2.11.Detoxification enzymes and Lung cancer

Tobacco smoking, including second-hand smoke, is the principal etiological agent towards risk for lung cancer. Tobacco smoke contains many carcinogens, including polycyclic aromatic hydrocarbons (PAH), nicotine-derived nitrosamine ketone (NNK), volatiles such as 1,3-butadiene, ethylene oxide, and others (Hecht, 2011). These carcinogens may be detoxified by the human body using a detoxification system. The detoxification mechanism is divided into phase I and phase II. The vast majority of the carcinogens found in cigarette smoke are procarcinogens that act as substrates for phase I and phase II enzymes. Phase I activates the procarcinogens or the by-products of Phase I, and when polar, they are directly excreted out of the body. However, otherwise Phase I is followed by conjugating a polar compound to the Phase I by-product

(Goldstein and Faletto, 1993). The first step is the activation of carcinogenic compounds by phase I enzymes such as cytochrome P450 (CYP), monoamine oxidase (MAO), flavin-containing mono-oxygenases (FMO), xanthine oxidase/ aldehyde oxidase (XO/AO) (Hecht, 2011). N- and O-dealkylation, aliphatic and aromatic hydroxylation, N- and S-oxidation, oxidative deamination, dehydrogenation, reductive dehalogenation, hydrolysis (ester, epoxide, amide, carbamate), and reduction (azo, nitro, keto, olefin, disulfide, keto) are all processes carried out by phase I enzymes (Taxak and Bharatam, 2014). These reactions add or remove a functional group (e.g., -CO_2H, -OH, -SH, or –NH_2). The activated carcinogens generated during the process can actively bind to DNA and form DNA adducts (Castell *et al*., 2005; Philippe and Dominique, 2010). Further, the conjugation of a polar compound to a phase I by-product is carried out by Phase II "transferase"enzymes, as shown in figure 2.8 (Xue *et al*., 2014). Glutathione S-transferases (GSTs), UDP-glucuronosyl transferases (UGTs), sulfatases (STs), arylamine N-acetyl transferases (NATs), and methyltransferases transform activated carcinogens into harmless, more hydrophilic substances that are quickly excreted from the body (Nishikawa *et al*., 2004). GSTs are responsible for cytosolic dimerization, UGTs for glucuronidation, and STs for aliphatic alcohol and phenol sulfation. Phase II detoxification is followed by Phase III, during which membrane transporters become active. *ABC* transporters are critical genes in Phase III (Omiecinski *et al*., 2010). There must be a balance between carcinogen activation and detoxification. There is an inter-individual disparity in the balance of carcinogen activation and detoxification, which affects cancer susceptibility. Certain carcinogens form adducts without being activated (Cohen and Arnold, 2011).

Environmental factors such as tobacco smoke, diet, and occupational exposures can trigger upto 90% of all cancers. Most chemical carcinogens require metabolic activation before interacting with cellular macromolecules and cause lung cancer initiation through DNA damage (Zhang *et al*., 2006). Exogenous genotoxic carcinogens are a diverse group of chemically distinct compounds that share the ability to form chemical bonds with DNA, resulting in the formation of "DNA-adducts." Furthermore, the initial stage of genetic damage is caused by endogenous

processes that result in mutations or even gene deletions.

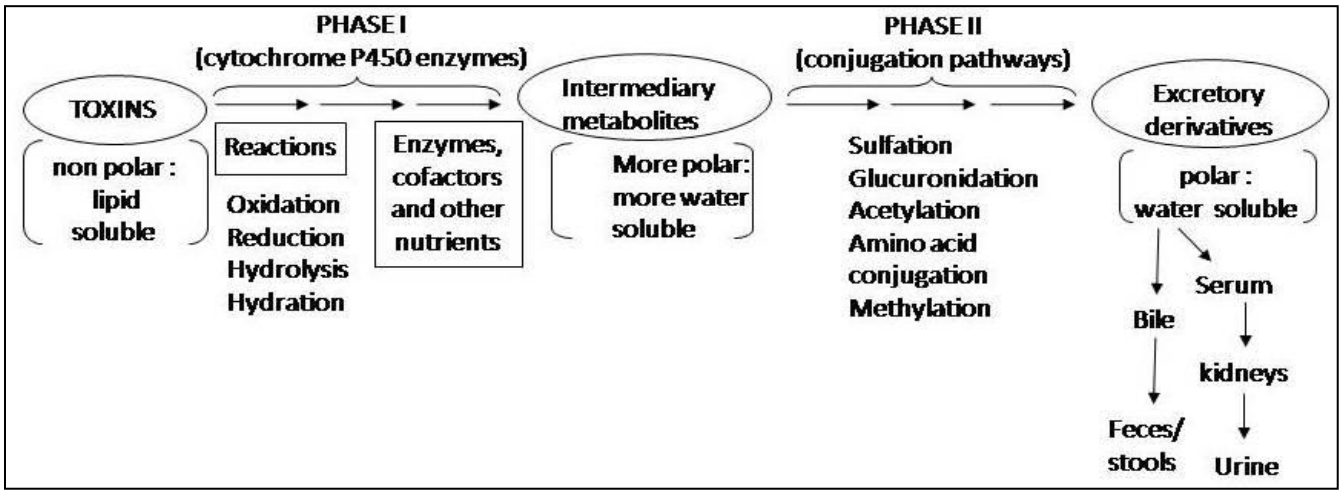


Figure 2.8: Detoxification pathway followed in the human body

Most genotoxic carcinogens such as PAHs, alkyl, and arylnitrosamines, aromatic amines, are activated by enzymes whose regular role is to metabolize, detoxify and eliminate potentially toxic compounds. The pro-carcinogens are activated phase I enzymes, most of which are members of the CYP450-dependent mono-oxygenases superfamily. Phase I reactions normal function is to inject oxygen into relatively inert and typically non-polar substrates, and phase II enzymes then conjugate the intermediates produced by oxidative metabolism with endogenous polar substrates such as glutathione, glucuronides, or sulfate to allow excretion of this polar product formed (Goldstein and Faletto, 1993; Jancova *et al*., 2010). As a result, phase-II enzymes can minimize cellular exposure to carcinogens, whereas phase-I enzymes can increase it. Some evidence indicates that polymorphisms and altered functions of these enzymes can play a role in cancer susceptibility (Bennett *et al*., 1999; Ji *et al*., 2012; Walia *et al*., 2019; Walia *et al*., 2021).

Cellular DNA repair systems can remove DNA adducts and restore standard DNA structure, thus counteracting mutational consequences. These intricate systems include direct base repair by alkyltransferase, DNA damage repair by nucleotide and base excision repair, mismatch repair, and double-strand repair (Christmann *et al*., 2003). Polymorphisms exist in genes that code for certain DNA repair enzymes. Suppose these variants result in poor DNA repair, the likelihood of cancer developing increases (Liu *et al*., 2005). Another protective mechanism is apoptosis, or programmed cell death, which can remove cells with DNA damage,

acting as a counterbalance to mutational events. The balance of apoptotic and those repressing apoptotic mechanisms will impact tumor development (Bode and Dong, 2005; Hecht, 2011). Figure 2.9 demonstrates the mechanism by which the carcinogens cause lung cancer.

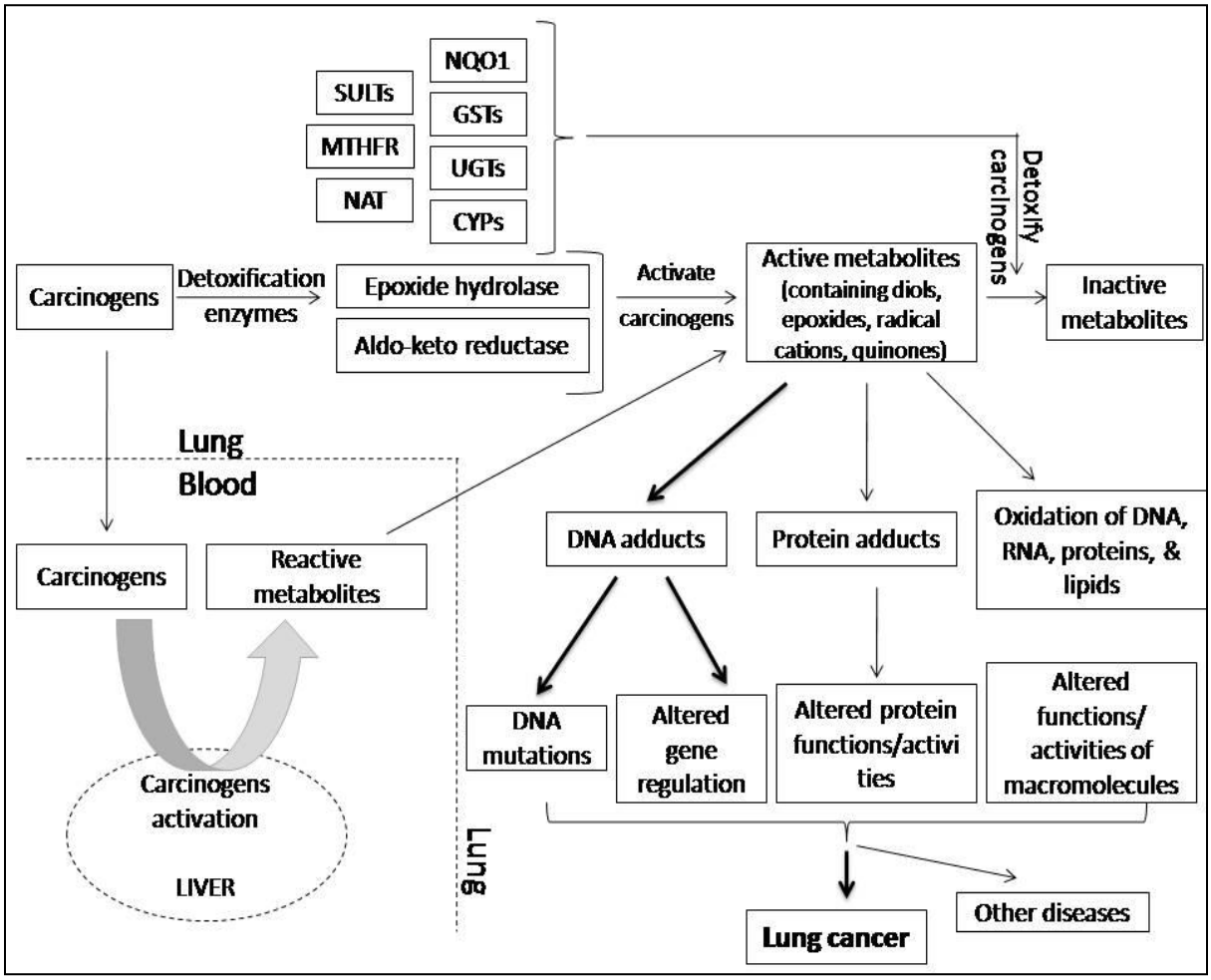


Figure 2.9: The multiple mechanisms by which carcinogens can cause lung cancer

2.11.1 Phase II metabolizing enzymes

The phase II enzymes play a principal role in the biotransformation of endogenous compounds (that originate from the system within such as cell, tissue, or an organism) and xenobiotics (drugs, carcinogens) to more easily excretable forms, as well as in the metabolic inactivation of pharmacologically active substances (Sheweita and Tilmisany, 2003; Jancova *et al*., 2010). Phase II biotransformation aims to carry out conjugation reactions, like sulfation, acetylation,

glucuronidation, methylation, amino acid conjugation, and glutathione are examples (Table 2.4). The respective conjugates are generally more hydrophilic than the parent compounds (Jancova *et al*., 2010).

Since drug interactions involving phase II enzymes are uncommon, they have received much less attention in clinical pharmacology than cytochromes P450. Nonetheless, the reduced metabolizing ability of phase II enzymes will manifest clinical drug toxicity. While phase II reactions are typically detoxifying, the conjugates produced can also have adverse consequences (for example, conjugates acting as carriers for potentially carcinogenic compounds in activating polycyclic aromatic hydrocarbons, benzylic alcohols, hydroxamic acid, aromatic hydroxylamines, and nitroalkanes by sulphotransferases). Figure 2.10 depicts the involvement of phase II drug-metabolizing enzymes in the metabolism.

Table 2.4: The phase II detoxification activities in human

Enzyme	**Reaction**	**Localization**[a]	**Substrates**
Glutathione transferases	Glutathione	Microsomes	Electrophiles
Glucuronyl transferase	Glucuronic acid (UDPGA)[b]	Microsomes	Phenols, thiols, amines, carboxylic acid
Epoxide hydrolase	H_2O	Microsomes Cytosol	Epoxides
N-acetyl transferases	Acetic acid (Acetyl-CoA)[b]	Cytosol	Amines
N- and O-methyltransferases	Methyl group (SAM)[b]	Cytosol Microsomes	Phenols, amines
Sulfotransferase	Sulfuric acid (PAPS)[b]	Cytosol	Phenols, amines, thiols
Amino acid transferases	Amino acid (Acetyl-CoA, taurine, glycine)	Microsomes	Carboxylic acids

[a]Cytosol refers to soluble activities present in the cytosolic portion of the cell; Microsome refers to membrane associated activities but localized to the cellular membrane or internal membrane.
[b] UDPGA = uridine-3',5'- diphosphoglucuronic acid; CoA = coenzyme A; SAM = S- adenosylmethionine; PAPS = phosphoadenosine 5'- phosphosulfate. (Source: Liska, 1998)

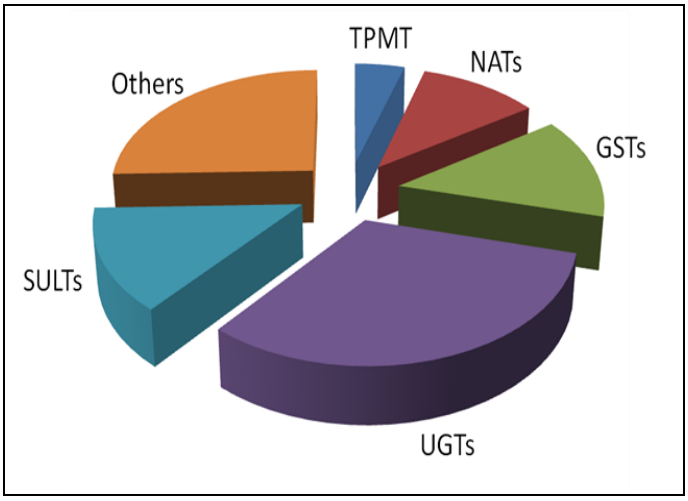


Figure 2.10: The phase II drug-metabolizing enzymes participating in the metabolism. (Source: Jancova*et al*., 2010)

2.11.1.1. NAD(P)H:Quinone oxidoreductase 1 (NQO1)

Ernster and Navazio reported the catalytic properties of NAD(P)H:quinone oxidoreductase 1 (NQO1) for the first time in 1958 (Eranster, 1958). NQO1 is an obligate two-electron reductase distinguished by its ability to use either NADH or NADPH as a reducing cofactor and its inhibition by dicoumarol (Ross *et al*., 2000). The observation that menadione reductase (later discovered to be NQO1) levels were induced in response to low doses of certain carcinogens provided some defense against subsequent treatments was one of the first examples of its protective nature (Huggins, 1979). NQO1, also known as DT-diaphorase, is a flavoenzyme that protects against endogenous and exogenous quinones by catalyzing two-or-four-electron reductions of these substrates. Quinone compounds can be found in our bodies (for example, vitamin K) as well as in the environment (e.g., urushiol, the active chemical in poison ivy) (Jayanthi *et al*., 2015). In a single two-electron step, NQO1 converts quinones to hydroquinones. In addition to producing substrates for Phase II conjugation reactions and promoting excretion, this two-electron method avoids potentially toxic semiquinone radical intermediates. However, not all hydroquinones are redox-stable, and in some cases, NQO1 metabolism produces a more active substance. The redox-unstable hydroquinone formed can further react with molecular oxygen, form semiquinone, and produce reactive oxygen species (ROS), or be produced by comproportionation reactions (Cadenas, 1995; Ross *et al*., 2000). In addition to potentially

inducing oxidative stress, the reduction of the quinone moiety may result in a compound capable of alkylating nucleophilic sites as well as DNA (Figure 2.11).

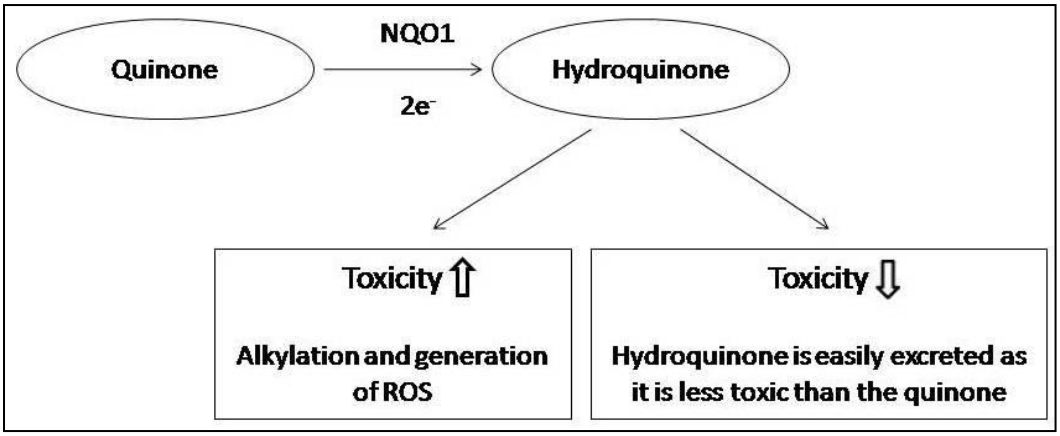


Figure 2.11: Quinone activation and deactivation as a result of NQO1-mediated reduction.

2.11.1.1.1. *NQO1* gene regulation

The isolation of NQO1 cDNAs from humans (Jaiswal *et al*., 1988), mice (Vasiliou *et al*., 1994), and rats (Robertson *et al*., 1986) indicated that *NQO1* is a single copy gene located on human chromosome 16q22.1 (Figure 2.12) (Jaiswal, 1991; Chen *et al*., 1991) and chromosome 8 in case of the mouse (Petersen *et al*., 1989). The *NQO1* gene has six exons and five introns and is approximately 20 kb in length based upon restriction mapping and sequencing. The 5'-UT, the first two amino acids, and the first nucleotide of the third amino acid are encoded by exon 1, while the remaining 272 amino acids and the 3'-UT are encoded by exons 2-6. The rat and human *NQO1* coding regions share a high degree of homology (85%) (Vasiliou *et al*., 1994). Transcriptional start sites vary between the human, mouse, and rat genes, implying that NQO1 can be regulated differently depending on the species (Kepa *et al*., 1997). The *NQO2* gene, which is distantly affiliated, has seven exons and is found on chromosome 6p25; instead of NAD(P)H, its gene product utilizes dihydro-nicotinamide riboside (NRH) as its electron donor. Both the *NQO1* and *NQO2* are induced by dioxin, oxidative stress, and PAHs contained in combustion processes (e.g., urban smog or smoke of cigarette) (Nebert *et al*., 2002).

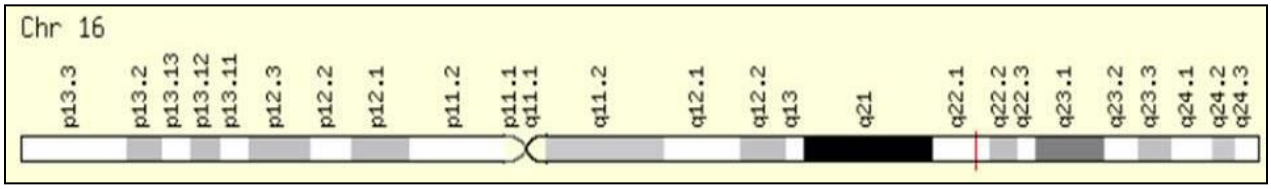


Figure 2.12: Chromosome 16 showing the position of the *NQO1* gene on it. (Source: Chen *et al*., 1991)

2.11.1.1.2. Structure of NQO1

NQO1 is a 60 kDa protein composed of two highly similar subunits entangled antiparallel to one another, each with 274 amino acids and one noncovalently bound FAD. Each subunit has two distinct domains, with residues 1-220 constituting a sizeable catalytic domain and 221-273 constituting a small C-terminal domain. The catalytic domain is made up of a central parallel β-sheet surrounded by helices, while the C terminal domain is made up of an antiparallel hairpin motif, α-helix, and multiple loops, as shown in figure 2.13. Each catalytic domain has three binding sites: (1) FAD binding site; (2) NAD(P)H binding site's hydrophilic adenine-ribose portion; and (3) the hydride donor/ acceptor binding site, that is a catalytic pocket that binds either the nicotinamide portion of NAD(P)H, the hydride donor/ acceptor substrate (Faig *et al*., 2001; Bianchet *et al*., 2008) as NQO1 is a homodimer with two active sites, one at each interface between the subunits. As a result, both active sites contain residues from both subunits. The FAD cofactor is a component of these active sites, and the NAD(P)H substrate binds so that the nicotinamide ring is parallel to the FAD, allowing for efficient electron transfer (Zhang *et al*., 2018).

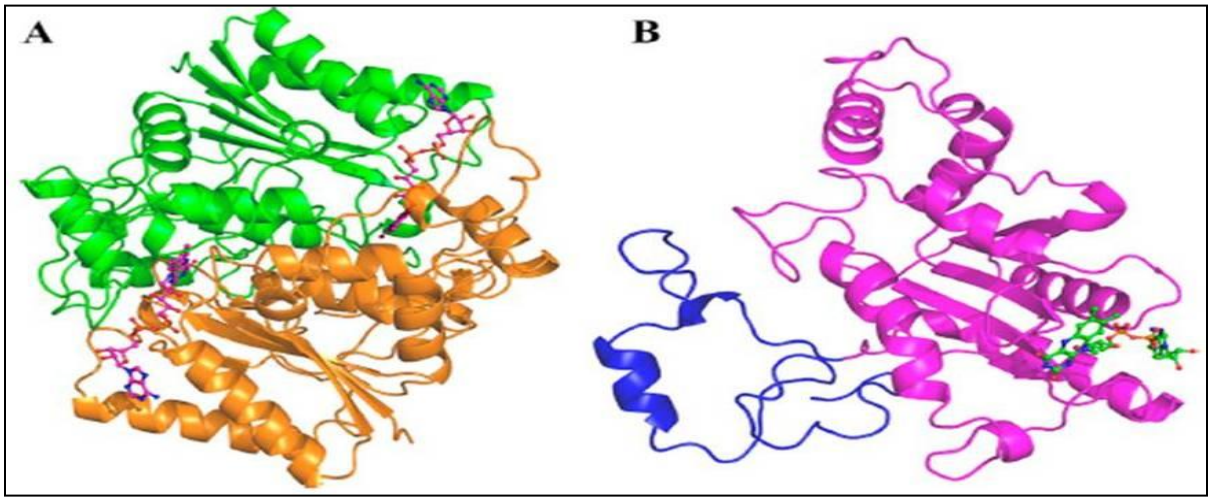


Figure 2.13: Human NQO1 structure in complex with FAD (the apoenzyme) (PDB 1D4A). (A) NQO1 apoenzyme overall structure. One monomer is shown in green, andthe other in orange. FAD's carbon atoms are shown in purple. (B) An NQO1 apoenzyme monomer. The catalytic domain is depicted in purple, while the C-terminal domain is depicted in blue. The carbon atoms of FAD are depicted in green. (Source: Zhang *et al*., 2018)

The NQO1-catalyzed reaction proceeds through an enzyme-substituted ping-pong mechanism in which NAD(P)H and the substrate alternately occupy the same binding site. It entails a double

hydride transfer which involves the transfer of one hydride from NAD(P)H to N5 of the FAD, and this is followed by the release of $NAD(P)^+$ from the catalytic site; whereas the other hydride is transferred over a 4A˚ distance from FADH2's N5 to the quinone substrate, followed by the release of hydroquinone produced from catalytic binding pocket (Li *et al*., 1995; Faig *et al*., 2000; Hassani *et al*., 2005).

2.11.1.1.3. Catalytic cycle and Substrate specificity of *NQO1*

The cytosolic flavoenzyme NAD(P)H: quinone oxidoreductase 1 (NQO1) catalyzes the two-electron reduction of quinones into hydroquinones. The *NQO1* gene encodes an obligate two-electron reductase that catalyzes a wide variety of substrates by using either NADH or NADPH as reducing cofactors and is effectively inhibited by the anticoagulant dicumarol. It is mainly a cytosolic enzyme (about 90% of the time) and exists as a homodimer with one FAD molecule per monomer (Zhang *et al*., 2018; Pey *et al*., 2019). Increased expression of NQO1 in response to oxidative stress is shared across different organisms and provides the cell with defense strategies, most likely through its ability to reduce reactive quinones and quinone-imines to less reactive and less toxic hydroquinones types (Figure 2.14). A two-electron reduction also bypasses toxic and reactive semiquinone production, thus preventing ROS formation resulting from the semiquinone's interaction with molecular oxygen (Dinkova-Kostova and Talalay, 2010).

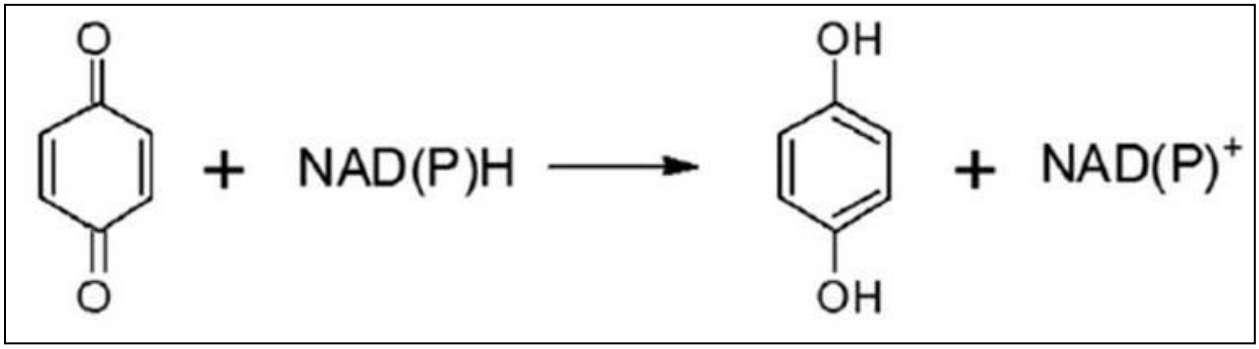


Figure 2.14: The general reaction catalyzed by quinone oxidoreductase in which NADH or NADPH reduces quinone to hydroquinone. (Source: Pey *et al*., 2019)

The ESR experiment supports this mechanism as it failed to detect the semiquinone radicals during the metabolism of naphthoquinone and benzoquinone substrates and verified by stop-flow and steady-state kinetic methods. The oxidoreductase functions *via* the "ping-pong"mechanism

where the reduced pyridine nucleotide binds the active site, reduces the flavin cofactor to $FADH_2$, and before binding the substrate, the oxidized pyridine nucleotide is released and completelyreducedby hydride transfer (Bianchet *et al.,* 2004).

NQO1 exhibits activity against a wide range of reactive species, including quinones, quinone-imines, dichlorophenolindolphenol, methylene blue, glutathionyl-substituted, naphthoquinones, azo, and nitro compounds, demonstrating its importance as a chemoprotective enzyme (Talalay and Dinkova-Kostova, 2004; Cenas *et al*.,2004). Both the para and ortho quinones are substrates for NQO1 (Segura-Aguilar and Lind, 1989). The enzyme's metabolism is not limited to quinones, and it also works well as a nitro-reductase, using substrates such as nitrobenzamides, nitrophenylaziridines, and dinitropyrenes (Knox *et al*., 1988; Hajos and Winston, 1991; Lambert *et al*., 1992). NQO1 can catalyze four-electron reductions of azo dyes and nitro substances in addition to a single two-electron reduction (Boland *et al*., 1991; Huang *et al*., 1979). By applying NQO1 heterodimers, it was discovered that the NQO1 subunits function independently in metabolizing two-electron substrates and dependently in metabolizing four-electron substrates (Cui *et al*., 1995).

2.11.1.2. Sulfotransferases (SULTs)

Baumann described sulfonate conjugation for the first time in 1876. SULTs are most likely the primary detoxification enzyme mechanism in the developing human fetus: no UGT transcripts were found in the fetal liver at 20 weeks of gestation (Strassburg *et al*., 2002). Sulfotransferases (SULTs) are classified into two types: membrane-bound SULTs and cytosolic SULTs. Membrane-bound SULTs in the Golgi apparatus catalyze peptide, protein, lipid, and glycosaminoglycan's sulfonation (Negishi *et al*., 2001; Nimmagadda *et al*., 2006). The cytosolic SULT superfamily all sulfonate xenobiotics, poisonous agents, therapeutic medicines, chemical carcinogens, peptides, hormones, bile acids, lipids, and neurotransmitters (Negishi *et al*., 2001; Gamage *et al*., 2006). Hepatic sulfonation is a typical phase II metabolic mechanism for increasing molecular hydrophilicity before biliary excretion and clearance (Banoglu, 2000; Blanchard *et al*., 2004). As a result, sulfonation is widely regarded as a detoxification pathway. SULTs, on the other hand, can convert such pro-mutagenic compounds found in food and the

atmosphere into mutagens and carcinogens. Sulfation is a metabolic activation mechanism for xenobiotics that results in highly reactive electrophiles that interact with DNA and proteins (Banoglu, 2000; Gamage *et al*., 2006). Many cytosolic SULTs have been identified at the mRNA level and classified into several gene families based on their amino acid sequences and catalytic properties in mammals (Glatt *et al*., 2001). 13 human cytosolic SULT genes have been identified, divided into four families: *SULT1, SULT2, SULT4,* and *SULT6* (Negishi *et al*., 2001; Blanchard *et al*., 2004; Freimuth *et al*., 2004; Nimmagadda *et al*., 2006). Even though the family members have a high degree of sequence and structural similarity, they have distinct biological functions. The SULT1 family has nine members divided into four subfamilies (1A1, 1A2, 1A3, and 1A4; 1C1, 1C2, and 1C3; 1B1, and 1E1). The sulfonation process, usually initiated by sulfotransferases, metabolizes various endogenous and exogenous compounds in the body. SULT1 isoform SULT1A1 (Sulfotransferase under the Family of 1A Member 1) is the most abundant and widely distributed member of the sulfotransferase family, and it is distributed throughout different tissues (Glatt, 2000; Gamage *et al*., 2006; Huang *et al*., 2015).

2.11.1.2.1. *SULT1A1* gene regulation

The *SULT1A1* gene found on chromosome 16p11.2-12.1, near the related isoform *SULT1A2* (Figure 2.15). The SULT1A1 cDNA was isolated from a liver cDNA library, and its high affinity distinguishes its 295 amino acid, 32 kDa protein for the model substrate p-nitrophenol (Km~1M), susceptibility to inhibition by 2,6-dichloro-4-nitrophenol (DCNP), and thermostability at 45˚C. As a result, it was initially known as aryl/phenol or thermostable sulfotransferase (P-PST or TS-PST) (Wilbom *et al*., 1993; Veronese *et al*., 1994). The *SULT1A1* gene encodes four allozymes: SULT1A1*1 (wild-type), SULT1A1*2, SULT1A1*3, and SULT1A1*4. The SULT1A1 enzyme is a 295 amino acid protein that is found in almost all human tissues. In humans, three closely related SULT1A members (SULT1A1, SULT1A2, and SULT1A3) have been isolated, with >90% sequence identity (Hempel *et al*., 2007). The SULT1A subfamily has the broadest tissue range, with SULT1A1 being the main SULT present in the human liver and expressed in the gastrointestinal tract (GIT), brain, placenta, and platelets (Jancova *et al*., 2010).

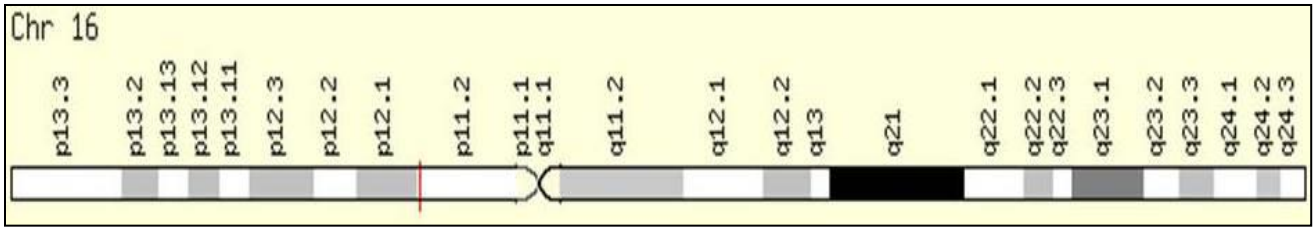


Figure 2.15: Chromosome 16 showing the position of the *SULT1A1* gene on it. (Source: https://www.genecards.org/cgi-bin/carddisp.pl?gene=SULT1A1)

2.11.1.2.2. Structure of SULT1A1 protein

Human SULT1A1 has a single α/β domain that forms a central five-stranded parallel-sheet surrounded on either side by α-helices, forming a fold that is common to all recognized cytosolic SULT structures (Figure 2.16). The strands contribute to the PAPS binding site residues as well as the essential catalytic residues. Surprisingly, the structural fold of SULT is close to that of nucleotide kinases such as uridylate kinase, implying an evolutionary association between sulfonation and phosphorylation (Kakuta *et al*., 1998).

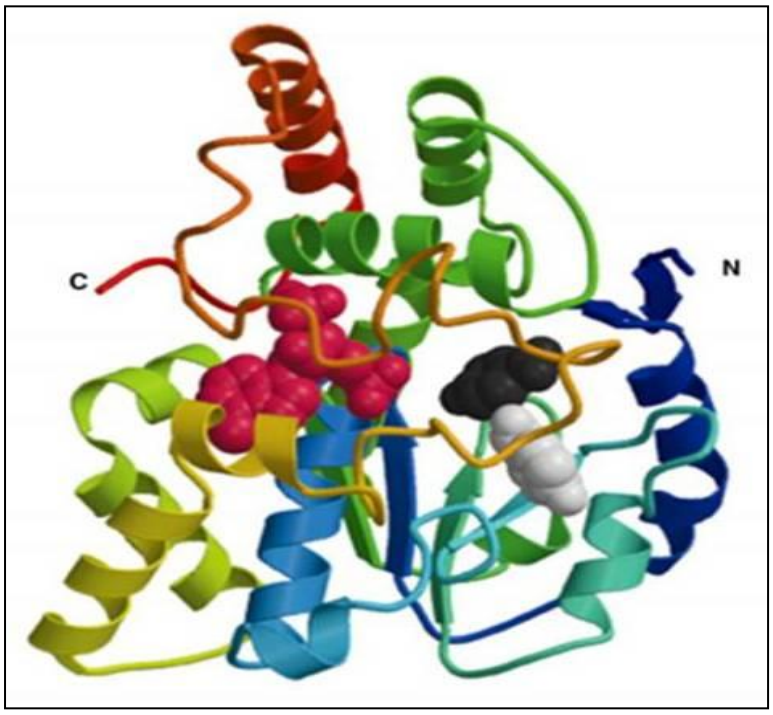


Figure 2.16: Human SULT1A1 crystal structure complexed with 3 -phosphoadenosine-5 -phosphate (PAP; dark pink) and two p-nitrophenol molecules (dark grey and white). The color of the C trace ranges from blue (N terminus) to red (C terminus). Coils for helices and arrows for strands are used to reflect secondary structural components. (Source: Hempel *et al.*, 2007)

The SULT1A1 residues form the phosphosulfate binding (PSB) loop 45-TYPKSGT-51, which provides the main binding interactions for the 5'-phosphate group of the sulfate donor PAPS (Kakuta *et al*., 1998; Hempel *et al*., 2007). The PSB loop is critical for orienting the cofactor during in-line sulfuryl transfer to the acceptor substrate (Kakuta *et al*., 1998; Pedersen *et al*., 2002). The side chains of aromatic residues Trp53, Thr227, and Phe229 specify the location of PAPS's adenine ring. PAPS-binding site's structural features are well conserved among sulfotransferases (Hempel *et al*., 2007).

SULT1A1's crystal structure, like that of most cytosolic sulfotransferases, involves a covered hydrophobic substrate-binding site, wherein PAP and twomolecules of model xenobiotic substrate p-nitrophenol are bound in a hydrophobic L-shaped binding pocket. SULT1A1 has a "plastic" binding site that can take on different structures to interact with different compounds (Gamage *et al*., 2003; Hempel *et al*., 2007).

2.11.1.2.3. Reaction catalysis and substrate specificity of SULT1A1

Sulfonation of small endogenous substrates and xenobiotics like hormones, bile acids, and neurotransmitters occurs in a wide range of species, from microbes to humans (Nagata and Yamazoe, 2000). The transition of a sulfonate group (SO_3^-) from PAPS to the substrate's acceptor group is catalyzed by SULTs (Pang *et al*., 1994). ATP sulfurylase conjugates inorganic sulfate with adenosine monophosphate in the cytosol to form 5'-phosphosulfate (APS), forming PAPS after phosphorylation APS kinase, which is the donor of sulfonate groups in all the sulfate conjugation reactions (Strott, 2002). Sulfotransferases catalyze the conjugation of 3'-phosphoadenosine 5'-phosphosulphate (PAPS) with an O-, N-, or S- acceptor group of an acceptable molecule (Figure 2.17). In general, O-sulfation is the most crucial cellular sulfonation reaction. Nonetheless, N-sulfation is a significant reaction in modifying carbohydrate chains in macromolecules such as heparin and heparan sulfate, both of which are standard proteoglycan components (Habuchi, 2000). N-Sulfoconjugation also plays a role in the metabolism of xenobiotics, including quinolones and amino drugs (Senggunprai *et al*., 2009). PAPS is a sulfate (or, more accurately, sulfonate) donor molecule necessary for all sulfonation reactions (Strott, 2002).

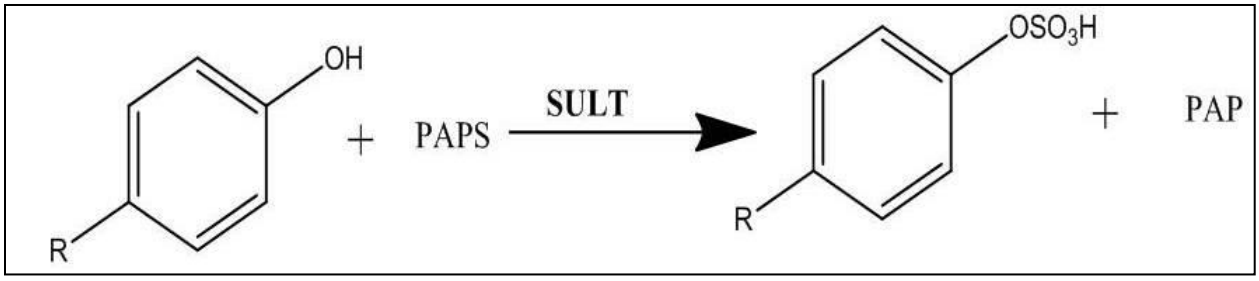


Figure 2.17: The general reaction catalyzed by sulfotransferases. (Source: Jancova *et al*., 2010)

The sulfated products produced by the reaction carried out by sulfotransferase usually have lower biological activity than the parent compounds since they are water-soluble and serve as substrates for one or more of the many transporters involved in uptake into the cells, removal from organs, and blood-brain barrier traversal (Coughtrie, 2016). This is valid for the vast majority of sulfated xenobiotics and endogenous compounds, although several molecules widely, when sulfated, either maintain or develop any pharmacological, harmful, or other biological activity. Examples are neurosteroids, steroid sulfates, various dietary and environmental chemicals, and their metabolites such as hydroxylated alkylbenzenes and N-aryl hydroxamic acids N-hydroxylamines benzylic alcohols of PAHs (Miller, 1994; Coughtrie, 2016).

SULT1A1 is a xenobiotic-conjugating enzyme with a comprehensive substrate set. It is also known as thermostable phenol sulfotransferase (TS PST1) and phenol sulfotransferase (P-PST) (Glatt and Meinl, 2004). It is also responsible for the sulfoconjugation of phenolic compounds, including benzylic alcohols, monocyclic phenols, aromatic amines, naphthols, hydroxylamines, iodothyronines, and dopamine (Jancova*et al*., 2010).

2.11.1.3. Epoxide Hydrolase (EPHX)

Watabe and Kanehira (1970) were the first to purify microsomal epoxide hydrolase 1 (EPHX1, EC 3.3.2.9) from rabbit liver, and is also known as *MEH*, *EPOX*,*EPHX*, or *HYL1*. Human *EPHX1* was later identified in the human liver by Oesch*et al*. (1974). EPHX1 is an evolutionarily conserved biotransformation enzyme found in nearly all tissues and mainly in the microsomal part of eukaryotic cells, endoplasmic reticulum. Human *EPHX1* gene orthologues have been discovered in 127 other species. The conversion of epoxides to trans-dihydrodiols is

the most common EPHX1 reaction (Václavíková *et al*., 2015). Two mammalian enzymes, microsomal (mEH) and soluble epoxide hydrolase (sEH),are traditionally essential in xenobiotics metabolism. With a few exceptions, hydrolysis of epoxide leads to increased metabolite water solubility and the end of its genotoxic potential. As a result, mammalian epoxide hydrolases, particularly the mEH, with its unusually broad substrate selectivity against a comprehensive set of epoxides, perform admirably as xenobiotic-metabolizing enzymes (Decker *et al*., 2009). The HUGO (Human Genome Organisation) gene nomenclature committee has designated the two well-studied mammalian EHs as EPHX1 (mEH) and EPHX2 (sEH). It is recognized that the organism uses many endogenous epoxides as signaling molecules, particularly epoxides generated from unsaturated fatty acids, such as epoxyeicosatrienoic acid (EETs) and leukotriene A4 (LTA4). Human sEH, which is expressed in the body, is the central enzyme in the metabolism of such endogenous epoxides. The phosphatase activity of human sEH emphasizes the enzyme's significance in regulatory processes rather than xenobiotic metabolism. Human sEH, which is expressed widely in the body, is currently thought to be the central enzyme in the metabolism of such endogenous epoxides. The phosphatase activity of human sEH emphasizes the enzyme's significance in regulatory processes rather than xenobiotic metabolism. However, because it accepts some trans-substituted epoxides, it functions to mEH in xenobiotics metabolism (Decker *et al*., 2009).

2.11.1.3.1. *EPHX1* gene regulation

Human *EPHX1* (Gene ID: 2052), located on chromosome1 (1q42.12), has nine exons and spans approximately 35 kb (Figure 2.18) (Jackson *et al*., 1987; Hartsfield *et al*., 1998; Václavíková *et al*., 2015). Exons 2-9 of EPHX1 encode three transcription variants that differ in the 5' untranslated region, with each translated protein product containing 455 amino acids (Václavíková *et al*., 2015). The presence of different promoters has been initially attributed to the complex control of *EPHX1* gene expression (Gaedigk *et al*., 1997), but later the contribution of numerous post-transcriptional processes, such as upstream open reading frames, was established (Liang *et al*., 2005; Nguyen *et al*., 2013). The proximal E1 promoter causes human EPHX1 expression in the liver, but an alternate promoter region (E1-b promoter) induces expression in other organs from adult and fetal sources (Liang *et al*., 2005).

A wide variety of transcriptional factors binding to regulatory regions adds to the intricacy of *EPHX1* gene regulation. GATA4 is the most important activator of EPHX1 expression, although HNF3 functions as a co-repressor in HepG2 cells (Liang *et al.*, 2005). Furthermore, CEBPA acts as an additional regulator of EPHX1 expression by interacting with DNA-bound NFY (A subunit, B subunit, C subunit) (Zhu *et al.*, 2004). Other nuclear receptors (CAR (NR113), HNF4A, and RXR (RXRA, and RXRB)) have been demonstrated to bind to the proximal EPHX1 promoter region and regulate its expression in human hepatocytes (Peng *et al.*, 2013).

The human liver and skin have the highest amounts of EPHX1 transcript and protein expression. EPHX1 expression is tissue, gender, and age-specific, but it also varies significantly between individuals in humans (Oesch *et al.*, 1977; Hammock *et al.*, 1997; Václavíková *et al.*, 2015). Transcripts of EPHX1 are expressed in human primary bronchial epithelial cells but not in alveolar macrophages (Willey *et al.*, 1996; Václavíková *et al.*, 2015).

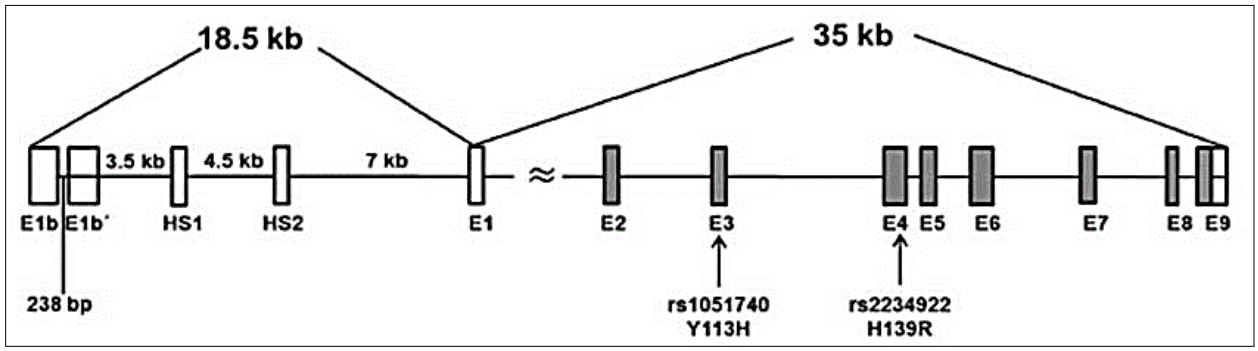


Figure 2.18: The *EPHX1* gene sequence and regulatory elements are located on chromosome 1 at 1q42.12. (Note: E= exons; HS= DNaseI hypersens5-c97ivity sites, kb= kilobases and rs= reference number SNP ID number). Coding exons are highlighted in grey. (Source: Václavíková*et al*., 2015).

2.11.1.3.2. Structure of EPHX1 protein

EPHX1 is a member of the α/β hydrolase family (Ollis *et al.*, 1992). Comparisons of homologies between microsomal epoxide hydrolases from phylogenetically distinct taxa imply a common ancestor (Arand *et al.*, 1994; van Loo *et al.*, 2006). The N-terminus of the EPHX1 protein attaches it to the membrane (Craft *et al.*, 1990), whereas the C-terminus contains catalytic

residues (Zou *et al*., 2000). Although the 3D structure of human EPHX1 has yet to be determined, the crystal structure of its orthologue from *Aspergillus niger* is known (Zou *et al*., 2000). A subsequent quantitative structure-activity relationship (QSAR) model reveals that this structure is somewhat helpful for predicting the binding of small organic compounds such as styrene epoxide to human EPHX1. On the other hand, congeners containing bulky side groups are likely to disturb the charge-relay phase of the catalytic mechanism (Lewis *et al*., 2005). Figure 2.19 depicts a schematic illustration of the critical structural motifs of epoxide hydrolases.

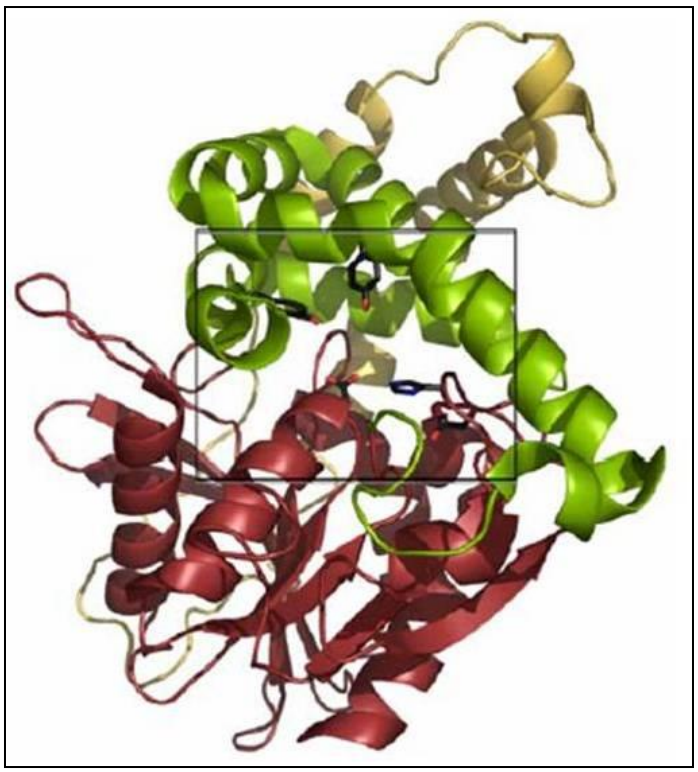


Figure 2.19: α/β hydrolase fold EHs 3-D structure and enzymatic mechanism. The structural model of *Aspergillus niger* EH, a close relative of mammalian mEH, was created using PyMOL and coordinates from PDB. The complete subunit with the α/β hydrolase fold is highlighted in red, the lid domain highlighted in green, and the N-terminus meander highlighted in gold (Source: Decker *et al*., 2009).

2.11.1.3.3. Reaction catalysis and substrate specificity of EPHX1

The overall reaction catalyzes the addition of water to the oxirane ring, resulting in the creation of vicinal diol as a reaction product and the production of an ester intermediate. The catalytic triad is formed by the EPHX1 catalytic cycle (Armstrong, 1987; Decker *et al*., 2009). This triad consists of the EPHX1-Asp^{226} residue rapidly nucleophilically attacking the substrate, producing an enzyme-substrate ester intermediate, and then hydrolyzing this complex with activated water

(Armstrong *et al*., 1981; Decker *et al*., 2009). Proton abstraction from the EPHX1-His^{431}-Glu^{404} charge relay pathway powers water activation (Oesch *et al*., 2000). Furthermore, Yamada *et al*. (2000) demonstrated that Tyr^{374} of human *EPHX1* plays a crucial mechanistic role in substrate activation. The active site is composed up of a catalytic triad comprised of the active site residues aspartic acid (catalytic nucleophile), histidine (broadbase), and aspartic or glutamic acid (charge relay acid). The catalytic nucleophile attacks the oxirane ring (with certain enantio- and stereoselectivity depending enzyme) at one of the two carbon atoms to generate a covalent intermediate in the first chemical reaction step (Lacourciere *et al*., 1993). A water molecule hydrolyzes the enzyme-substrate ester intermediate in a second step, activated by proton abstraction from the catalytic triad's histidine, as shown in figure 2.20. The ensuing positive charge is protected by an acidic amino acid residue, either glutamic acid (in mammalian EHs) or aspartic acid (in most other EHs), and the oxyanion hole stabilizes the intermediate state. Because the catalytic nucleophile is on a nucleophilic elbow, it can adapt to the position of the substrate epoxide with some flexibility (Decker *et al*., 2009).

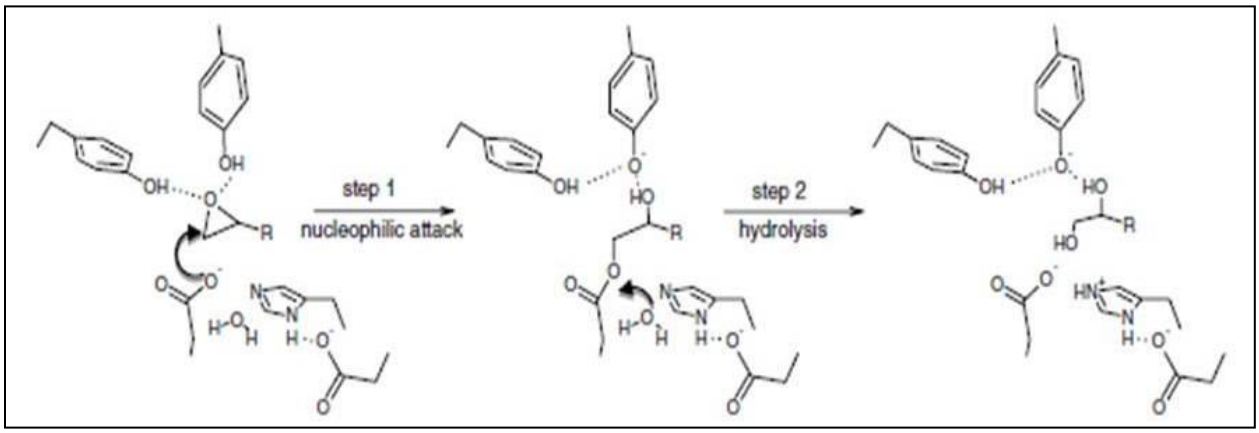


Figure 2.20: The two-step catalytic mechanism of epoxide hydrolysis (Source: Decker *et al*., 2009).

EPHX1 transforms xenobiotics more rapidly than endogenous substrates and primarily serves a detoxifying function. There is some evidence, however, that EPHX1 plays a crucial role in organ-specific human physiology. EPHX1 was shown to convert epoxides such as styrene oxide, 1-methyl-1-phenyloxirane, indene 1,2-oxide, and cyclohexane oxide into diols (Oesch 1974). Later, it was discovered that EPHX1 has more considerable substrate specificity (Fretland and Omiecinski, 2000), and there is a significant substrate-dependent variation in EPHX1 enzymatic activity among various species (Kitteringham *et al*., 1996).

Overall, microsomal EPHX1 has a dual role in xenobiotic biotransformation. While it detoxifies carcinogenic chemicals such as benzene, butadiene, styrene, and others (Decker *et al*., 2009), it can also activate procarcinogens such as PAHs (El-Sherbeni and El-Kadi, 2014; Václavíková *et al*., 2015). Toxic and carcinogenic substances, as well as routinely used anticonvulsant drugs, are examples of typical substrates. Epoxide derivatives of PAHs are the most prevalent environmental compounds digested by mEH. Other xenobiotic compounds processed by mEH include 1,3-butadiene epoxide derivatives, chrysene, benzene, nitropyrene, anthracene, aflatoxin B1, and naphthalene (Fretland and Omiecinski, 2000).

2.11.1.4. *N-acetyltransferase (NAT)*

N-acetyltransferases (NATs, EC 2.3.1.5) (Weber and Hein, 1985) are the drug-metabolizing enzymes that catalyze the conjugation of an acetyl group from acetyl Co-enzyme A (AcCoA) to N-hydroxy arylamine, hydrazines, and arylamines. They are found in a wide range of eukaryotic organisms such as humans, and their presence has also been demonstrated in the prokaryote *Salmonella typhimurium* and has a role in xenobiotic detoxification and metabolic activation (Sim *et al*., 2014). (HUMAN)NAT1 and (HUMAN)NAT2 are two polymorphic *NAT* genes found in the human genome. The biotransformation of heterocyclic and aromatic amine procarcinogens to reactive forms is a complicated procedure that requires at least two enzyme-mediated steps: N-hydroxylation, which is mediated by cytochrome P450 (CYP) 1A2, and subsequent O-acetylation of the N-hydroxyarylamine product by NAT to form N-acetoxyarylamines (Hein, 1988; Windmill *et al*., 1999).

NAT1 expression is widespread throughout the body, whereas NAT2 expression is thought to be more restricted. Activity studies have revealed that both forms are present in the liver and other tissues such as uroepithelial cells, gut, and lymphocytes (Windmill *et al*., 2000). In 2000, 25 human *NAT1* alleles and 27 human *NAT2* alleles were discovered (Hein *et al*., 2000). *NATs* have different tissue distribution and substrate selectivity despite their high level of similarity. NAT1 has a widespread tissue distribution, and its expression has been linked to malignancies. NAT2 activity is found in the colon, liver, and intestinal epithelium (Jancova *et al*., 2010).

The N-acetylation of isoniazid and other drugs like sulfamethazine divided the human population into rapid, intermediate, and slow acetylator phenotypes, whereas the N-acetylation of other drugs like p-aminosalicylic acid produced a normal distribution as per an early paradox (Jenne, 1965). Based on this scrutiny, drugs like sulfamethazine and isoniazid were labeled as "polymorphic,"whereas p-aminosalicylic acid was labeled as "monomorphic" substrates for N-acetylation (Hein *et al*., 2000). The isoniazid slow acetylator phenotype(s) is/are caused, at least in part, by NAT2 protein reduction(s), which occurs at a frequency of around 50–60% in Caucasian populations but exhibits substantial ethnic heterogeneity. For example, the slow acetylator phenotype is substantially more common in Egyptians but considerably less common in Asians (Weber, 1987). Because of its fundamental function in activating and deactivating a vast and diversified number of hydrazine and aromatic amine drugs used in clinical medicine, the NAT2 acetylation polymorphism is significant in clinical pharmacology and toxicology (Weber, 1987; Hein *et al*., 2000).

2.11.1.4.1. *NAT* gene regulation

NAT1 and NAT2 result from single, intronless protein-coding exons of 870 bp open reading frames that encode 290 amino acids. *NAT1*, *NAT2*, and a pseudogene, *NATP*, are all found on human chromosome 8p22, a location where heterozygosity is found to be frequently lost in human malignancies (Hickman *et al*., 1994; Matas *et al*., 1997) (Figure 2.21). Between *NAT1* and *NAT2*, the coding exon is a pseudogene (*NATP*). A third *NAT* gene (*NAT3*) has been discovered (Fretland *et al*., 1997). *NAT1* and *NAT2* have 85% nucleotide homology in the coding region, resulting in 55 amino acid variations. NAT2 mRNA forms from the protein-coding exon and a second non-coding exon of 100 bp located about 8 kb upstream of the translation start site, whereas NAT1 mRNA is obtained from a single exon (Blum *et al*., 1990; Hein *et al*., 2000). In a study conducted by Sugamori *et al*. (2007), NAT3 is unlikely to have a substantial role in arylamine N-acetylation in wild-type mice or mice lacking NAT1 and NAT2 activity.

The open reading frames of the *NAT1* and *NAT2* genes were intronless when the cDNA and genomic DNA clones were compared. Human *NAT1* contains a non-coding exon, and we now know that the human *NAT1* gene contains several upstream non-coding exons, and splicing of these into proximal region directly in front of coding exon may be tissue-specific (Boukouvala

and Sim, 2005; Barker *et al*., 2006; Minchin *et al*., 2007). In addition to its single coding-exon, human *NAT2* has a non-coding exon spliced in the frame. The extended *NAT2* gene is substantially more straight forward than the human *NAT1* gene, with several alternative transcripts. Human NAT1's widespread tissue expression, as opposed to human NAT2's more restricted expression, may explain the more complexgene organization (Minchin *et al*., 2007; Sim *et al*., 2010; Sim *et al*., 2014).

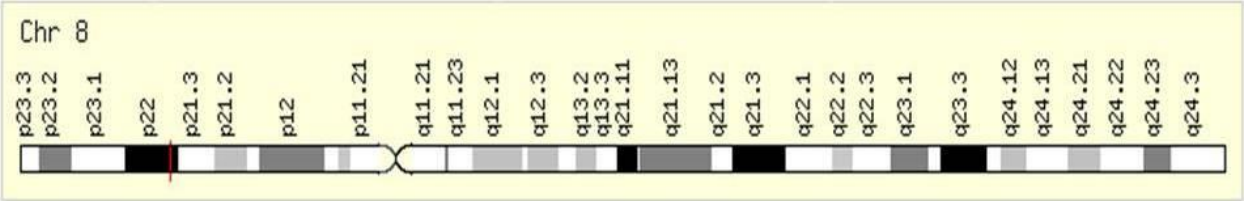


Figure 2.21: Chromosome 8 showing the position of the *NAT* gene on it. (Source: de Figueiredo Teixeira *et al*., 2013)

2.11.1.4.2. Structure of NAT2 protein

NAT1 and NAT2 have similar three-dimensional structures, and the Cα atoms of residues 1-290 are superimposed with an r.m.s.d of 0.7Å. The fold of human NAT1 and NAT2 closely stimulates prokaryotic orthologues' general structure (Figure 2.22), traditionally characterized in three domains. The NAT enzyme comprises of three about equal-sized domains, α-domain, β-domain, α/β-C- terminal domain. The first two domains contain all active site Cys, His107, and Asp122 residues, while a third α-β lid folds over the active site. The length and function of the third C-terminal domain vary the greatest between NATs (Sim *et al*., 2014).

The amino-terminal domain (residues 1 to 83) consists of five helices (α1-α5) and one short β-strand (which is between helices α2 and α3). The second domain is composed of nine β-strands (β2 to β11, and two short helices α-6 and α-7) which comprises residues from 85 to 192. These two domains via the α-helical inter-domain (helices α8 to α10) are linked to the third domain, which has four anti-parallel β-strands (β12 to β15) and α11 helix (Wu *et al*., 2007). The α11 helix is followed by a stretch of well-resolved residues that lead over the surface of the protein molecule to a buried carboxyl terminus. Residues 167 and 183 in human NATs surround a 17-residue insertion that is lacking in prokaryote NAT structures. A study conducted by Walraven *et al*. (2007) found that insertion helps maintain the stability of human NATs. The

initial few residues of this insertion create a short β-strand (β10) and extend the previously mentioned 4-stranded anti-parallel β-sheet (domain III) by parallel alignment to its fourth strand (β15), in comparison to prokaryotic NATs, the structures of NAT1 and NAT2 exhibit a considerable change in the arrangement of their carboxyl-terminal residues. The carboxyl-terminal helix in prokaryotic NATs is dissolved ("stretched out"), beginning at residue 274. On the other hand, the carboxyl terminus reaches deep within the folded human NATs, close to the buried catalytic triad (Figure 2.22), and can interact with CoA (Wu *et al*., 2007).



Figure 2.22: The structure of the human NAT2 is depicted as a ribbon diagram. All helices and strands are labeled. The carboxyl and amino termini are denoted by the letters CT and NT, respectively. CoA coupled to NAT2 are depicted as sticks, with carbon atoms depicted in gray. Human NATs have insertions and carboxyl-terminal tails absent in other species' NAT proteins and are magenta. (Source: Wu et al., 2007)

The GCN5-related N-acetyltransferase and amino(α)-terminal acetyltransferase superfamilies exhibit little resemblance to the human NAT1 and NAT2 folds (Dyda *et al*., 2000).

2.10.1.4.3. Reaction catalysis and substrate specificity of NAT2

The liver arylamine N-acetyltransferases (acetyl CoA dependent N-acetyltransferases, NATs) participate in the biotransformation of aromatic amines and hydrazines through the transfer of

the acetyl group from acetyl-coenzyme A to the free amino group of the parent molecule (Figure 2.23). O-acetylation of aromatic and heterocyclic amines (4-aminobiphenyls) is catalyzed by NATs, whereas N-acetylation of the parent amines is considered the detoxification (Hein *et al*., 2000; Jancova *et al*., 2010).

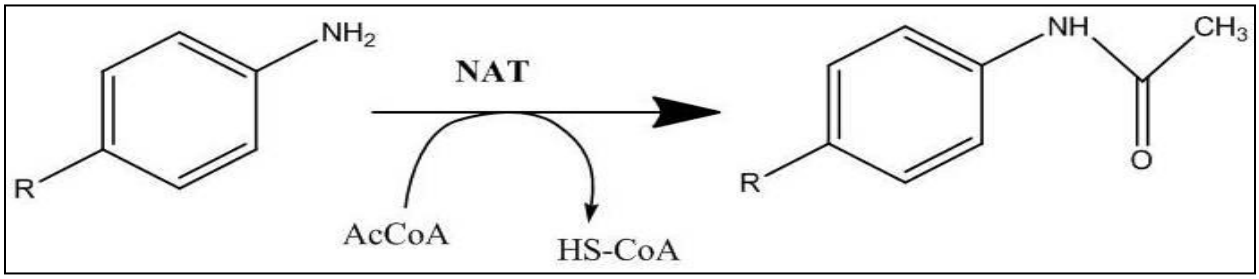


Figure 2.23: Reaction catalyzed by NAT involving acetylation of arylamine (Source: Jancova *et al*.; 2010)

The hepatic cytochrome P4501A2 catalyzes the N-oxidation of arylamines to generate the N-hydroxyarylamine proximal carcinogen metabolite (Butler *et al*., 1989). The arylamine carcinogens N-acetylation can compete with N-oxidation, implying that N-acetylation can operate as a deactivation mechanism for carcinogenic arylamines (Hein, 1988). After N-oxidation, acetyl coenzyme A-dependent Cytosolic acetyltransferase(s) may metabolically activate N-hydroxyarylamine proximal carcinogens, resulting in O-acetylation of the N-hydroxy moiety directly (Jancova *et al*., 2010). The formed N-acetoxyarylamine metabolites are volatile and spontaneously hydrolyze to arylnitrenium ions, which bind to nucleophiles such as DNA, forming covalent adducts with the potential to cause mutations in genes that control cellular proliferation, resulting in a more malignant phenotype (Flammang and Kadlubar, 1986; Hein, 1988; Hein *et al*., 1993).

The substrate profiles of the relevant enzymes are distinct: HUMAN(NAT1) acetylates the arylamines 4-aminobenzoyl glutamate (4ABglu), 4-aminobenzoic acid (4ABA), and 4-aminosalicylate (4AS), whereas (HUMAN)NAT2 preferentially N-acetylates the hydralazine (HDZ), and hydrazines isoniazid (INH), as well as the arylamine sulf (SMZ) (Jancova*et al*., 2010; Laurieri*et al*., 2014). Heterocyclic and N-hydroxy aromatic amines (O-acetylation) and hydrazines and aromatic amines (N-acetylation) are two types of acceptor substrates that are either activated (O-acetylation) or deactivated (N-acetylation) by NAT1 and NAT2. N, O-

acetyltransfer of N-hydroxy-N-acetyl-aromatic amines is also catalyzed by NAT1 and NAT2 (Hein, 1988; Hein *et al*., 2000).

2.11.1.5. *Methylenetetrahydrofolate reductase (MTHFR)*

The *MTHFR* gene encodes for an enzyme known as Methylenetetrahydrofolate reductase. This enzyme is involved in the processing of amino acids, which are the building blocks of proteins. MTHFR is used for a chemical reaction involving vitamin folate (also known as vitamin B9) (Medline Plus, *MTHFR* gene, 2021). Both folate and methionine cycles in humans produce compounds that are required for cellular viability. Folate, the primary cellular carrier of single carbon units, is essential for purine and thymidine monophosphate production. The homocysteine's methylation to methionine carried out by methionine synthase within the methionine cycle produces an essential amino acid that can be used for the synthesis of a protein or canconvert to S-adenosylmethionine (SAM), a critical donor for methylation of proteins, RNA, and DNA, as well as the formation of various methylated compounds. The enzyme 5,10-methylenetetrahydrofolate reductase connects these two cycles (MTHFR; E.C. 1.5.1.20) (Froese *et al*., 2018).

Epidemiologic studies have shown that lack of dietary folate is a risk factor for the development of cancers; such as tumors of the lung, colorectum, breast, pancreas, and bladder (Shen *et al*., 2003; Larsson *et al*., 2006; Schabath *et al*., 2005; Shrubsole *et al*., 2004; Choi *et al*., 2002; Arslan *et al*., 2011). A variety of fruits and vegetables are rich in dietary constituents and help to negate the carcinogenesis of these harmful chemicals, thus providing a protective effect against tumorigenesis. Among the different chemical constituents present in these dietary products is folate, usually present in folic acid (Shi *et al*., 2005). Folate deficiency can lead to misincorporation of uracil and subsequently instability of DNA, as well as a reduced ability to repair alkylating or oxidative damage, and global and proto-oncogenic DNA hypomethylation (Duthie, 1999; Kim, 1999). MTHFR exists as homodimers in humans and other eukaryotic organisms, whereas *MTHFR* in *E. coli* exists as a homotetramer (Shan *et al*., 1999).

2.11.1.5.1. *MTHFR* gene regulation

MTHFR is an enzyme involved in DNA synthesis and methylation and plays a crucial rolein the folic acid metabolic process. *MTHFR* is present on the short arm of chromosome 1 (1p36.3) (Figure 2.24). Eukaryotic MTHFR is a homodimer with a regulatory domain and catalytic domain in each subunit (70~77 kDa) (Igari *et al*., 2011). The human gene encodes a protein with a calculated molecular mass of 74.5 kDa and 656 amino acids (Goyette *et al*., 1994). A catalytically active 70 kDa protein isproduced by expressing a 2.2 kb human MTHFR cDNA in a bacterial expression system (Frosst *et al*., 1995). The mouse *MTHFR* cDNA is isolated, and it specifies a protein that shares 90% amino acid identity with its human counterpart (Goyette *et al*., 1998).

Mammalian MTHFR polypeptides have two domains. The NH2 terminal catalytic domain (approximately 40 kDa) shares 30% sequence identity with E. coli MetF and, like MetF, contains FAD as a non-covalently attached prosthetic group (Matthews *et al*., 1998). The AdoMet binding site islocated in the COOH- terminal domain; [methyl3H] AdoMet photoaffinity labeling identified roughly 50 residues from the domain junction (Goyette *et al*., 1994; Matthews *et al*., 1998).

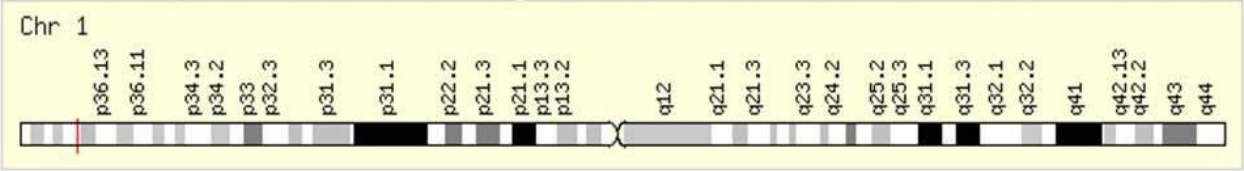


Figure 2.24: Chromosome 1 showing the position of the *MTHFR* gene on it. (Source: https://www.genecards.org/cgi-bin/carddisp.pl?gene=MTHFR&keywords=MTHFR)

2.11.1.5.2. Structure of MTHFR protein

MTHFR is a 656 amino acid multi-domain protein found in humans. The catalytic domain has remained to be conserved during evolution (Figure 2.25). The catalytic domain forms a β8α8 (TIM) barrel and hasrevealed residues essential for binding of FAD (the cofactor) (Guenther *et al*., 1999), NADPH (the electron donor) (Pejchal *et al*., 2005), and CH_3-THF (the product) (Pejchal *et al*., 2005; Pejchal *et al*., 2006; Lee *et al*., 2009). The bacterial structures demonstrate that the catalytic domain is adequate for the entire catalytic cycle (Matthews *et al*., 1984).

MTHFR orthologs in eukaryotes also have a C-terminal regulatory domain linked to the catalytic domain *via* a linker sequence. This C-terminal domain can bind S-adenosylmethionine (SAM), resulting in allosteric inhibition of enzyme activity (Sumner *et al*., 1986), which is very slow (Kutzbach *et al*., 1971) and reversed by attaching to S-adenosylhomocysteine (SAH) (Daubner and Matthews, 1982; Yamada *et al*., 2001), the demethylated form of SAM.

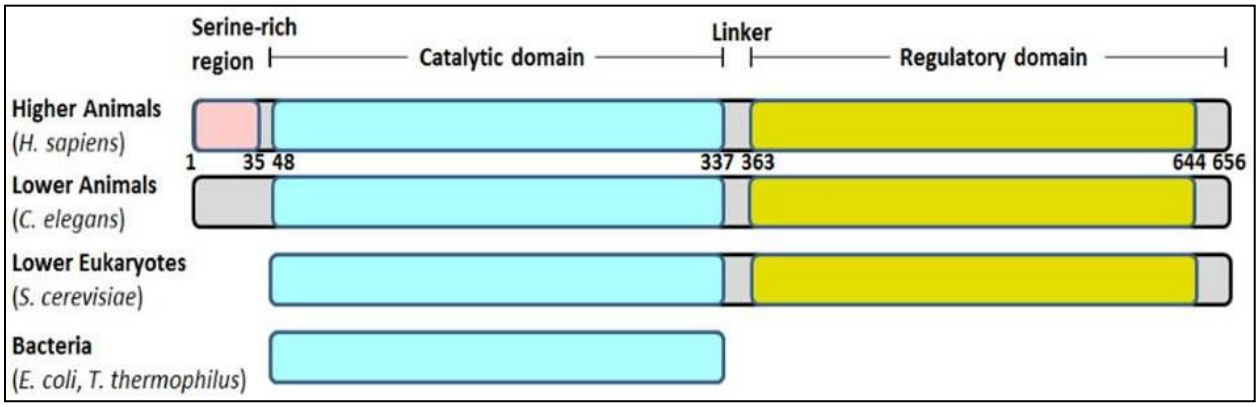


Figure 2.25: MTHFR ortholog domain arrangement across evolution. (Source: Froese *et al*., 2018)

The promoter of human $MTHFR_{38-644}$ folds into two globular domains to generate an overall elongated protein. The N-terminal catalytic domain (amino acids (aa) 40-337), as predicted from bacterial structures, is made up of 8α/8β TIM barrel, enhanced with three additional α helices (α8, α9, and α11). The C-terminal regulatory domain (aa 363 to 644) is a unique fold consisting of 2, and five stranded β-sheets positioned side by side in core (Froese *et al*., 2018) (Figure 2.26).

The interaction between two domains is not direct but linked by a long linker region spanning from amino acid 338 to 362, with amino acid sequence rich in Pro (n=4), Arg (4), Leu (3), and Glu (3) (Figure 2.26). This linker, mediated by three turns, establishes several complicated interactions with both domains and changes direction twice while traveling between regulatory and catalytic domains (Froese *et al*., 2018).

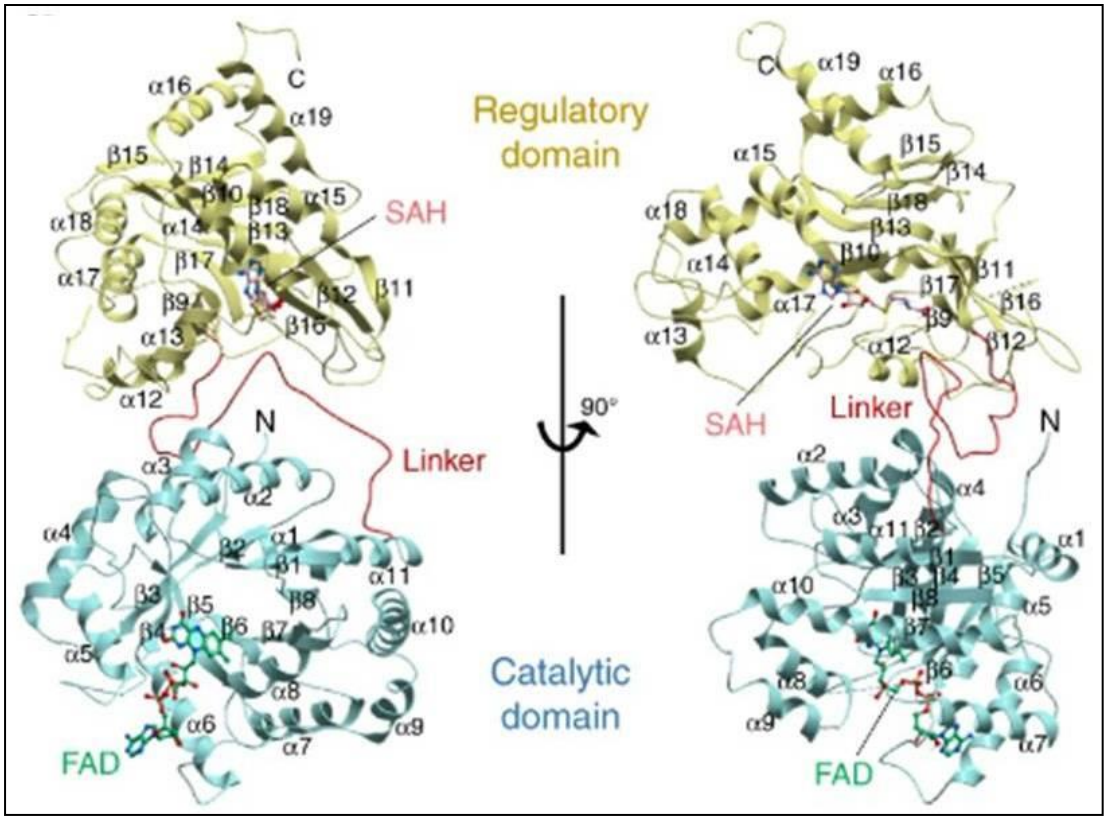


Figure 2.26: Overview of the human MTHFR structural gene depicting the regulatory domain (yellow), the catalytic domain (cyan), and the linker (red). FAD (green) and SAM (pink) bound in the sticks are depicted in the structure. The dotted lines represent chaotic regions that are not represented in the structure. The α-helices and β-sheets are also labeled. (Source: Froese *et al*., 2018)

2.11.1.5.3. Reaction catalysis and substrate specificity of MTHFR

The two cycles involved in DNA synthesis and production of ubiquitous methyl donor S-adenosylmethionine (SAM) are the critical metabolic pathways for life, the folate and methionine cycles. The enzyme 5,10-methylenetetrahydrofolate reductase (MTHFR) is a crucial regulatory link between these cycles and significantly impacts different diseases. Therefore, the enzyme MTHFR is present at the interlinkage between the folate cycle, which is essential for synthesizing thymidylate and purine, and the methionine cycle is responsible for synthesizing ubiquitous methyl donor SAM (Figure 2.27). The flow of folate-mediated one-carbon unit, which derives critical processes for cell viability, is regulated by these two metabolic cycles together.

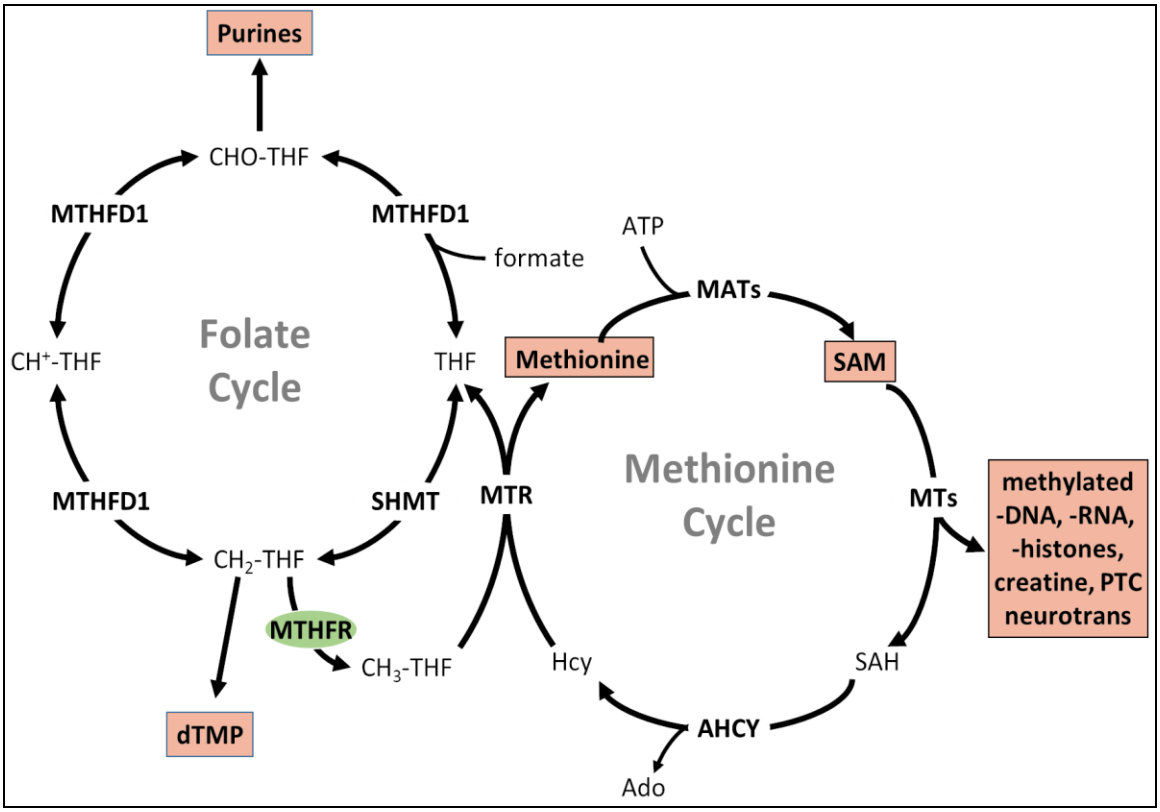


Figure 2.27: MTHFR in one-carbon metabolism at the folate and methionine cycle interconnection (Source: http://doi.org/10.5281/zenodo.3584528).

MTHFR is responsible for the physiologically irreversible conversion of Methylenetetrahydrofolate (CH_2-H_4folate) to methyltetrahydrofolate (CH_3-H_4folate), which requires FAD as a cofactor and NADPH as an electron donor (Guenther *et al*., 1999) (Figure 2.28). Because the product CH_3-H_4folate is employed only in the methionine cycle, MTHFR devotes folate-mediated one-carbon units to methionine and SAM synthesis by carrying out this physiologically irreversible process. MTHFR requires FAD as a non-covalently bound cofactor for the catalytic function (Matthews, 1986). To reduce FAD, MTHFR uses NAD(P)H, which then reduces CH_2-H_4folate. As the electron transfer between the two, *i.e.*, CH_3-H_4folate and FAD's oxidized form, is reversible, the enzyme is assayed by monitoring CH_3-H_4folate oxidation in the presence of an electron acceptor (Igari *et al*., 2011).

Previous research with porcine MTHFR has looked at the mechanism and stereochemistry of the NAD(P)H-CH_2- H_4 folate oxidoreductase reaction. NAD(P)H attaches to the *si* face of the FAD in the reductive half-reaction, and the 4S-hydrogen of NAD(P)Has a hydride is transferred to N5 of FAD cofactor (Vanoni and Matthews, 1984). The enzyme is

dissociated from NAD(P)+. CH_2-H_4folate binds at the *si* face of reduced FAD in oxidative half-reaction (Sumner and Matthews, 1992), which is consistent with the ping-pong bi-bi mechanism. The five-membered imidazolidine ring of CH_2-H_4 folate is proposed to be open by protonation of N10, forming a 5-iminium cation. The product CH_3-H_4folate is obtained by transferring a hydride from the reduced FAD's N5 to the exocyclic methylene group (C11) (Sumner and Matthews, 1992; Trimmer *et al*., 2001).

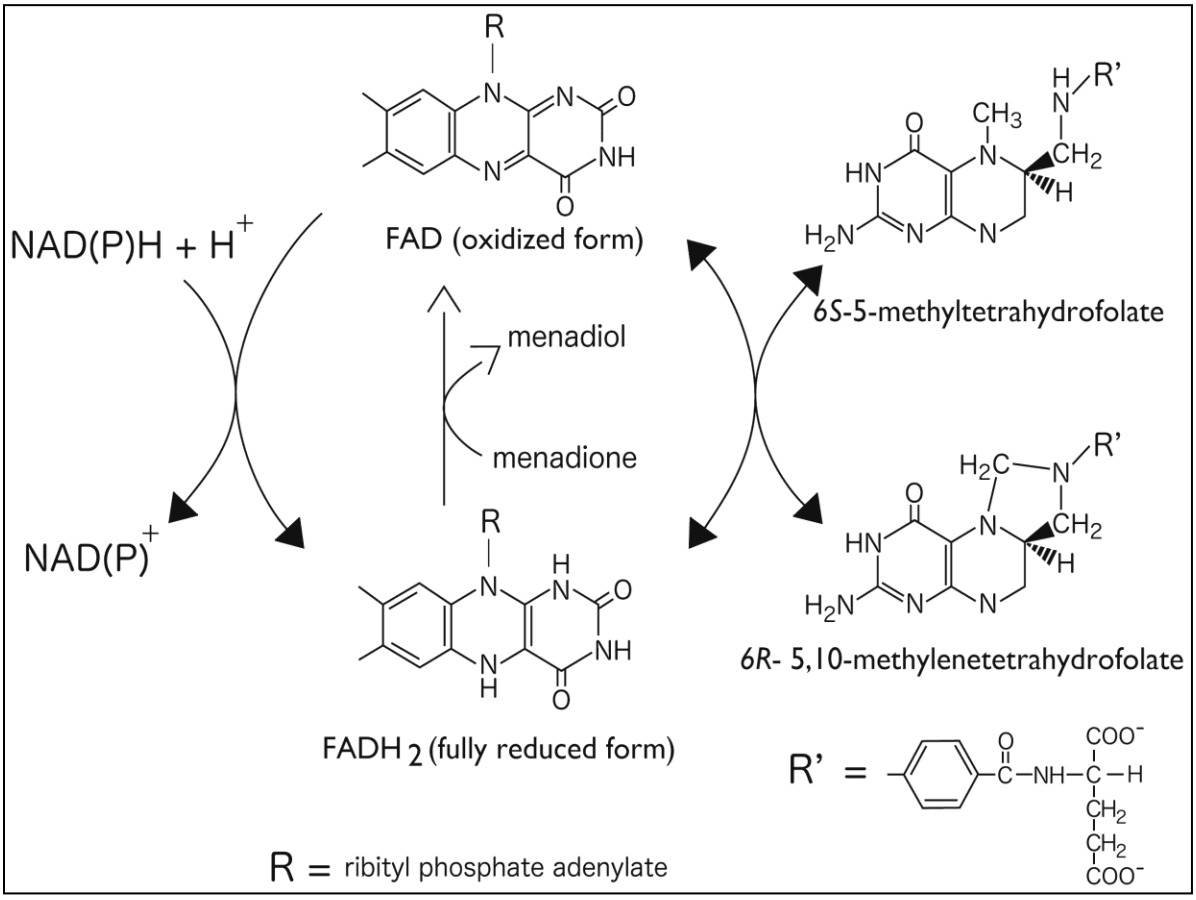


Figure 2.28: Reaction catalysis carried out by MTHFR (Source: Igari *et al*., 2011).

2.11.1.6. Glutathione S-transferases (GSTs)

Glutathione transferases (GSTs) are a multifunctional protein superfamily found in practically all the prokaryotic and eukaryotic cells that can detoxify various exogenous and endogenous toxic compounds (Sheehan *et al*., 2001; Hayes *et al*., 2005). GSTs are broadly classified into three primary categories in mammalian organisms: Cytosolic GSTs, mitochondrial GSTs, and

microsomal GSTs (Hayes *et al*., 2005). The cytosolic GSTs comprise several gene-independent classes; each group of GST isoenzymes has identical sequences and structural features. For example, seven classes are present in humans and mammals: Alpha (α), Mu (μ), Pi (π), Theta (θ), Omega (ω), Sigma (σ), and Zeta (ζ). While the Alpha class contains A1-1, A2-2, A3-3, and A4-4 isoenzymes, the Pi class contains only one enzyme, GSTP1 (Sheehan *et al*., 2001; Bocedi *et al*., 2019). Around sixty years ago, these enzymes were discovered (Booth *et al*., 1961).

The GSTs (EC 2.5.1.18) catalyze glutathione (GSH, -Glu-Gys-Gly) conjugation with a wide range of electrophilic substrates. In mammalian cells, they function as phase II detoxification enzymes, and the conjugated products mediated through the GST enzyme reactions are made water-soluble to eliminatethrough the urinary tract (Prade *et al*., 1997). GSTs are dimeric enzymes. Typically, dimmers are formed from identical chains, but heterodimers are formed from two different chains of the same class are to be found. Each GST monomer contains two unique domains: the C-terminal alpha-helical domain and N-terminal thioredoxin-like domain (Allocati *et al*., 2018). The same class members share more than 40-50% of sequence identity, but GSTs from other classes share less than 25-30% sequence identity (Board and Menon, 2013; Singh, 2015).

GSTs have various biological activities, including cell defense against oxidative stress and various harmful chemicals andcreating and modifying prostaglandins and leukotrienes (Hayes *et al*., 2005). GSTs, for example, protect cellular DNA against oxidative damage, which can cause an increase in DNA mutation or produce DNA damage that promotes carcinogenesis (Li *et al*., 2009). GSTs can conjugate glutathione (γ-L-glutamyl-L-cysteinyl-glycine, GSH) to a wide range of electrophilic and hydrophobic molecules, including many therapeutic drugs, carcinogens, and oxidative metabolism products, making them less toxic and more susceptible to further modification for excretion from the cell (Hayes *et al*., 2005) (Figure 2. 29).

2.10.1.6.1. *GSTP1* gene regulation

One of the well-studied variations of the GST metabolism is cytosolic GSTP1 isoenzyme. The *GSTP1* coding region is located on chromosome 11 and is controlled by a vast CpG island upstream of the transcription start point in the promoter region. Both sections are separated by a lengthy ATAAA repeating stretch, which most likely serves as an insulator between various

epigenetic states (Millar *et al*., 2000). Furthermore, activator protein 1 (AP-1), specificity protein 1 (SP1), GATA1, and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) are found to beimplicated in the control of GSTP1 expression (Figure 2.30) (Moffat *et al*., 1996; Schnekenburger *et al*., 2003; Duvoix *et al*., 2004; Morceau *et al*., 2004).

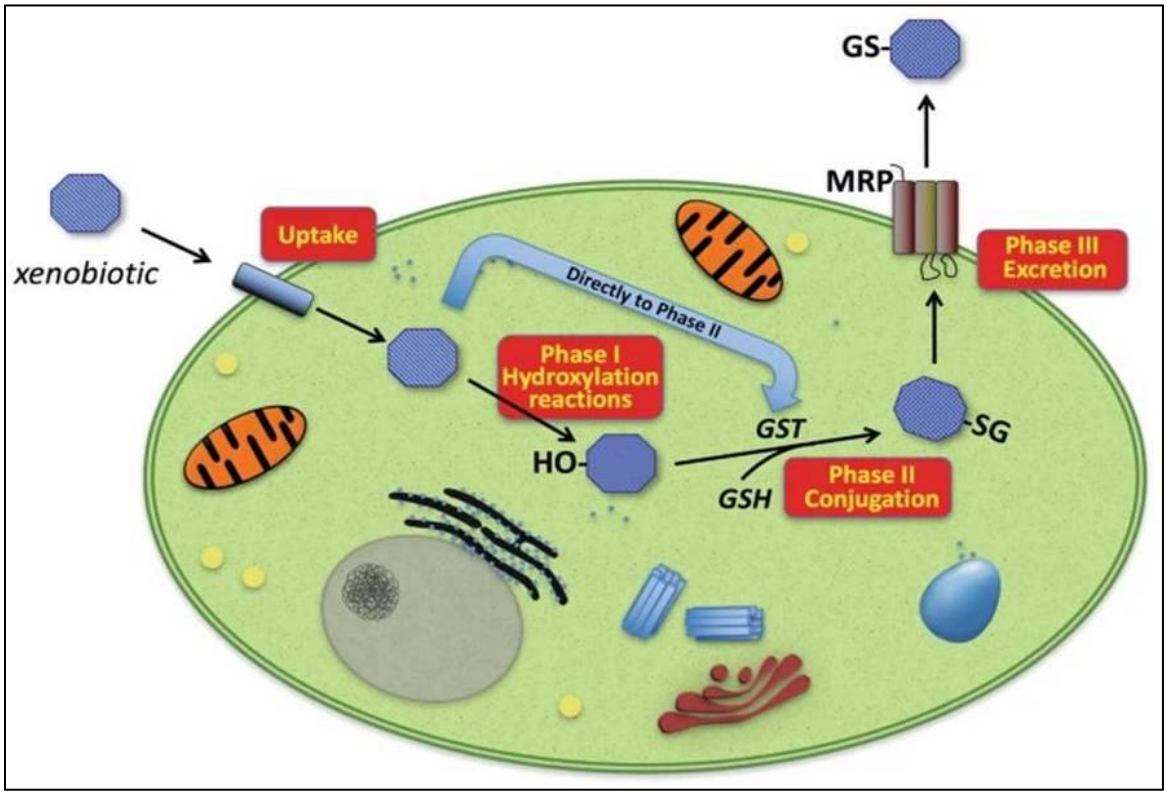


Figure 2.29: The harmful molecules may diffuse across the plasma membrane and be targeted by the enzyme of Phase I metabolism inside cells. The majority of them are members of the cytochrome P450 family, which includes many enzymes that catalyze various processes such as hydroxylation (the primary one), reduction and, oxidation. GSTs, which catalyze the conjugation of Phase I modified xenobiotics to endogenous GSH, thus play a vital role in ensuing Phase II metabolism. Different transmembrane efflux pumps actively move the conjugate out of the cell (Phase III). Some compounds may directly enter Phase II metabolism. (Source: Allocati *et al*., 2018).

GSTP1 is expressed in most cells, particularly those in contacts with the external environment, such as those in the digestive, urinary, and respiratory tracts, which correlates with its biotransformation and detoxifying ability (Yuan *et al*. 2008). Increased GSTP1 expression may also indicate increased detoxifying activity due to xenobiotic or endobiotic exposure implicating oxidative stress (Kanwal *et al*., 2014). As a result, increased GSTP1 expression is detected in many cancers (e.g., bladder, breast, colon, glioma, head and neck, lung, meningioma,

ovary and cervix, pancreas, stomach, soft tissue sarcoma, and embryonic testicular carcinoma), which is associated with enhanced detoxification activity, protecting cancer cells from cytostatic and cytotoxic drugs (Ruzza *et al*., 2009).

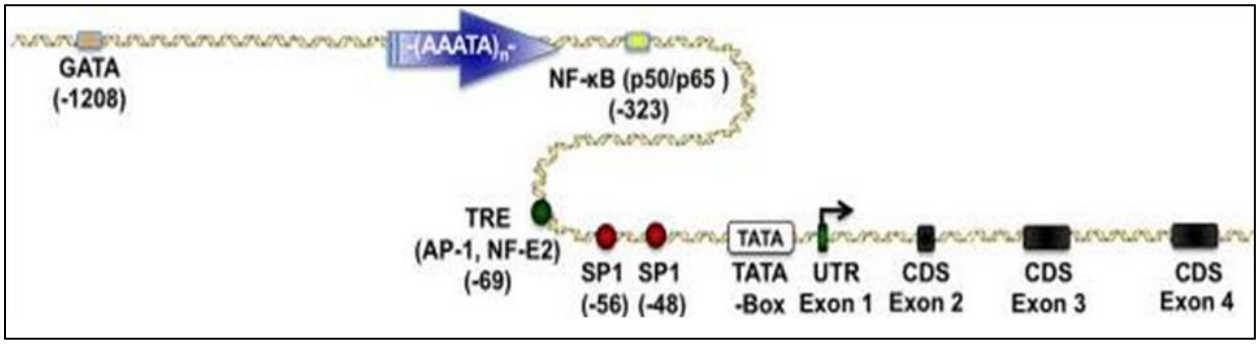


Figure 2.30: GSTP1 regulatory elements. (Source: Schnekenburger *et al*., 2014)

2.11.1.6.2. Structure of GST protein

The GST enzymes are dimeric. Usually, dimmers are formed from identical chains, although heterodimers formed from two distinct chains of the same classes are too found. Each GST monomer contains two unique domains: an N-terminal thioredoxin-like domain and a C-terminal alpha-helical domain (Figure 2.31). The first domain is in charge of GSH binding, with a specialized binding duty known as G-site. A particular residue within this location activates the GSH cysteinyl side chain *via* hydrogen bonding. This residue is a tyrosine in some classes and serine or cysteine in others. In humans, the G-site of alpha, mu, pi, and sigma isoenzymes contains tyrosine, whereas the other classes have serine or cysteine. Mitochondrial kappa class enzymes are similar to theta class enzymes in that they have a serine in the G-site. However, they differ in the overall fold due to a DsbA-like domain inserted within the N-terminal thioredoxin-like domain (Oakley, 2011; Allocati *et al*., 2018). Because of the hydrophobic nature of co-substrates, the C-terminal domain works with the N-terminal domain to shape the co-substrate binding site, which is known as the H-site. The diverse nature of GST co-substrates isreflected in the various H-site geometries and chemical properties identified among classes (Allocati *et al*., 2018).

The enzymes in complex with GSH are represented as GSH analogs, transition state analogs, and products, as well as the unliganded enzyme. All of the structures have the same basic protein fold, which consists of two domains. The N-terminal domains overall fold belongs

to the thioredoxin superfamily fold, which contains disulfide bond formation facilitator, glutaredoxin, and glutathione peroxidase.

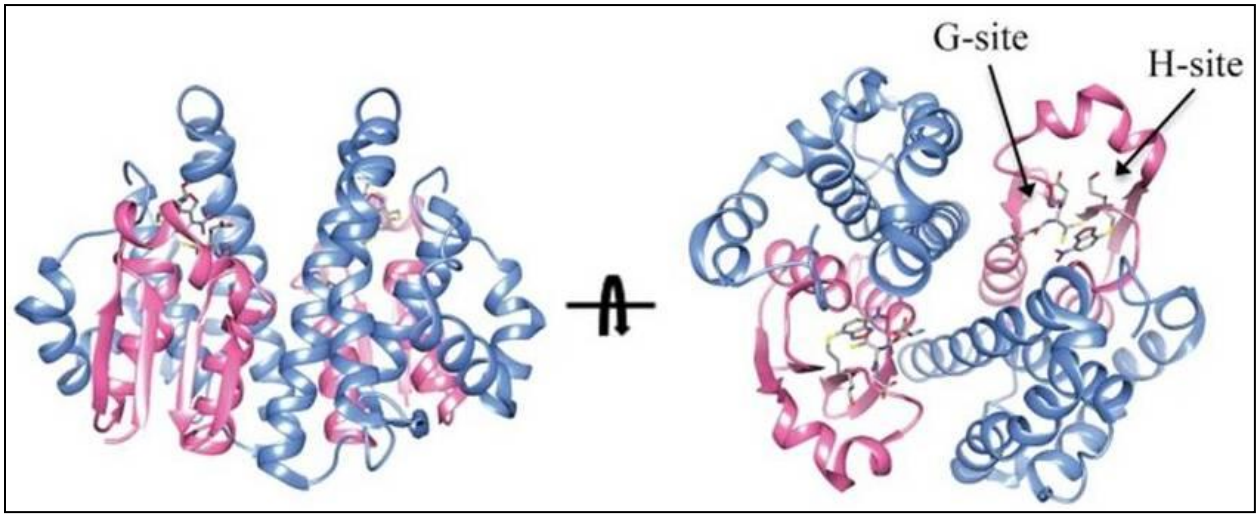


Figure 2.31: The structure of human GSTP1 in association with GSH and its inhibitor NBDHEX is depicted. The pink-colored helical is the N-terminal thioredoxin-like domain, while the one-colored cyan is all helical C-terminal domain. A GSH molecule fills the G-site, and the H-site is occupied by an NBDHEX molecule depicted as sticks. (Source: Allocati *et al*., 2018)

This domain accounts for around one-third of protein and a structural motif (α-β-α-β-α-β-β-α), which in order 4312 forms mixed four-strand sheet, with strand threeantiparallels to others. The domain's core is made up of three layers, with the β-sheet sandwiched between α-helices (α/β/α). The protein's C-terminal two-thirds are all α-helices with a distinct protein fold, and the center consists of a bundle of 4 helices (Armstrong, 1997).

The principal role of the N-terminal domain is to provide a binding site for GSH. The peptidyl component is located at the end of the β-sheet and interacts with protein through electrostatic and hydrogen-bonding interactions. The product's xenobiotics moiety is present in the gap between the two domains and forms several interactions with residues in domain II, especially along the face of the 4-helix and C-terminal tail. Therefore, domain II seems to provide structural features for xenobiotic substrate recognition and aids in defining the substrate selectivities of various isoenzymes. Domains I and II are also known as the GSH and xenobiotics substrate-binding domains (Armstrong, 1997; Prade *et al*., 1997).

2.11.1.6.3. Reaction catalysis and substrate specificity of GST

GSTs are enzymes that catalyze the production of thioether conjugates between endogenous tripeptide glutathione and xenobiotic substances (Figure 2.32). GSTs accelerate GSH's nucleophilic attack on electrophilic substrates by bringing the substrate in close vicinity with GSH in the active site of a protein and activating GSH's sulfhydryl group, which acts as a base catalyst (Dourado *et al.*, 2008). The GSTA/M/P classes highly conserved tyrosine-7 and GSTT/Z classes serine-17 play a crucial role in the catalytic activation of GSH. The hydroxyl group of tyrosine/serine works as a hydrogen bond donor to the thiol group of GSH, enabling the production and stability of the highly reactive thiolate anion, which is the target of nucleophilic attack on an electrophilic substrate (Oakley, 2011). GSTs undergo nucleophilic aromatic substitutions, nucleophilic addition to epoxides, Michael additions to α,β unsaturated ketones, hydroperoxide reduction, and isomerization (Henderson *et al*. 2005; Di Pietro *et al*., 2010).

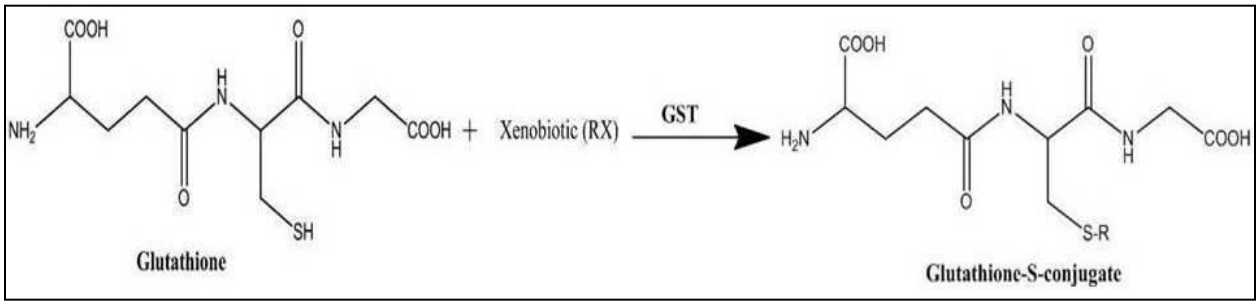


Figure 2.32: Reaction catalysis by GST leading to the formation of glutathione conjugate

GSTs are essential for the detoxification of epoxides produced from PAHs and α-β unsaturated ketones. Furthermore, glutathione conjugation is used to metabolize various endogenous substances like steroids and prostaglandins (van Bladeren, 2000). The primary biological role of glutathione transferases appears to be a defense against reactive and poisonous electrophiles such as ROS (hydrogen peroxide and superoxide radicals) produced by normal metabolic processes. Many of these are produced due to cellular oxidative processes catalyzed by cytochrome P450 and other oxidases (Sheehan *et al*., 2001).

GST substrates are any substances that can react with glutathione's thiol moiety. α,β unsaturated ketones, epoxides, quinones, esters, sulfoxides, ozonides, and peroxides are examples of electrophilic compounds (Jancova *et al*., 2010). The GSTs Alpha (A), Mu (M), and Pi (P) detoxify commonly occurring harmful α,β unsaturated carbonyls (e.g., 4-hydroxynonenal, acrolein, thymine propenals, adenine) (Hayes and Strange, 2000). A variety of GST-specific

substrates have been evaluated. Ethacrynic acid is a highly selective substrate for GST-P1 (Jancova *et al*., 2010).

2.12. Genetic variability in Phase II detoxification genes and its correlation with lung cancer susceptibility

2.12.1. Genetic polymorphisms in *NQO1* gene in association with lung cancer risk and overall survival of LC patients

NQO1, a Phase II detoxification enzyme, is a cancer preventive and chemoprotective due to the metabolism of xenobiotics and is highly expressed in human tissue, particularly cancer tissue. The NQO1 enzyme has several functions, such as detoxifying xenobiotics, scavenging superoxides, repairing endogenous antioxidants, p53 modulation, and the degradation of the proteasome. *NQO1* C609T (Pro187Ser, rs1800566) is the most commonly studied SNP. This polymorphism is present on exon six at position 609 of the *NQO1* gene, and it involves C- to -T allele base-pair mutation (Tian *et al.*, 2014; Masroor *et al*., 2015). The mutation leads to the variable activity of the enzyme, and this variant form is volatile and is readily denatured by the proteasome, and is ubiquitinated rapidly (Siegel *et al*., 2001). *Ser/Ser* that is the homozygous variant exhibits significantly less enzymatic activity, *i.e*., 2-4% as that of wild type (*Pro/Pro*) (Eom *et al*., 2009; Tian *et al*., 2014; Kim and Hong, 2015). It has been demonstrated that NQO1 regulates the activity of tumor suppressor gene *p53* and thus, in turn, influences the cancer cell's life (Pey*et al*., 2019). Various studies carried out earlier in different populations have produced inconsistent results on the association between *NQO1 609C>T* polymorphism and lung cancer risk (Sunaga *et al*., 2002; Lin *et al*., 2003; Tian *et al*., 2014; Kim and Hong, 2015; Tong *et al*., 2017). A meta-analysis conducted by Huang *et al*. (2017) provides evidence that *NQO1* polymorphism is not correlated with lung cancer risk, and the studies included in this analysis are tabulated in table 2.5.

In several earlier studies, the mutant (*TT*) genotype did not affect lung cancer risk (Saldivar *et al*., 2005; Timofeeva*et al*., 2010; Guo *et al*., 2012). On the contrary, other studies have reported an association of the *NQO1* polymorphism towards an increased risk for lung cancer (Eom*et al*., 2009; Tian *et al*., 2014; Masroor*et al*., 2015).

Table 2.5: The different studies conducted in *NQO1* polymorphism towards lung cancer risk and included in the meta-analysis (Huang *et al*., 2017)

Name	Year	Country	Ethnicity	Case				Control				HWE	Score
				CC	CT	TT	All	CC	CT	TT	All		
Wiencke	1997	USA	Caucasian	29	32		61	52	109		161	—	10
Wiencke	1997	USA	African-American	77	39		116	83	53		136	—	10
Chen	1999	USA(Hawaii)	Asian	54	48	7	109	64	78	25	167	0.877	10
Chen	1999	USA(Hawaii)	Caucasian	81	49	5	135	105	62	4	171	0.137	10
Chen	1999	USA(Hawaii)	Hawaiian	61	18	4	83	60	39	3	102	0.258	10
Lin	2000	China	Asian	12	63	20	95	41	73	22	136	0.268	3
Yin	2001	China	Asian	28	39	17	84	26	41	17	84	0.909	6
Xu	2001	USA	Caucasian	513	246	21	780	715	341	40	1096	0.933	6
Xu	2001	USA	Mixed	18	14	2	34	20	6	1	27	0.534	7
Lewis	2001	UK	Caucasian	56	24	2	82	111	32	2	145	0.858	4
Benhamou	2001	France	Caucasian	85	55	10	150	105	62	5	172	0.243	10
Sunaga	2002	Japan	Asian	83	93	22	198	52	77	23	152	0.526	6
Hamajima	2002	Japan	Asian	87	71	34	192	240	286	114	640	0.076	6
Lin	2003	China	Asian	57	141		198	95	237		332	—	9
Yin	2003	China	Asian	26	41	17	84	28	39	17	84	0.613	4
Lan	2004	China	Asian	37	57	25	119	32	54	23	109	0.980	8
Liang	2004	China	Asian	37	79	36	152	53	71	28	152	0.624	7
Alexandrie	2004	Sweden	Caucasian	345	168	11	524	368	153	9	530	0.124	7
Sorensen	2005	Denmark	Caucasian	162	83	9	254	176	80	11	267	0.618	10
Saldivar	2005	USA	Caucasian	454	205	24	683	480	186	17	683	0.839	11
Saldivar	2005	USA	Hispanics	15	17	4	36	15	14	7	36	0.275	8
Saldivar	2005	USA	African-American	67	33	7	107	69	35	3	107	0.563	9
Skuladottir	2005	Denmark	Caucasians	108	45		153	227	119		346	—	6
Chan	2005	China	Asian	25	37	13	75	45	83	34	162	0.708	5
Bock	2005	USA	Caucasian	93	37		130	87	57		144	—	7
Bock	2005	USA	African-American	21	10		31	21	8		29	—	7

Lawson	2005	Finland	Caucasian	244	109		353	243	117		360	—	10
Demirkan	2005	Turkey	Caucasian	47	34	7	88	80	34	3	117	0.785	4
Yang	2007	Korea	Asian	110	158	46	314	120	166	61	347	0.784	10
Eom	2009	Korea	Asian	122	265		387	148	239		387	—	9
Cote	2009	USA	Caucasian	271	97	19	387	271	119	15	405	0.668	11
Cote	2009	USA	African-American	77	32	4	113	79	36	6	121	0.478	10
Su	2009	China	Asian	102	199	95	396	139	244	82	465	0.158	8
Timofeeva	2010	Germany	Caucasian	429	188		617	856	411		1267	-	10
Guo	2012	China	Asian	187	327	168	682	171	282	144	597	0.192	8
Tian	2014	China	Asian	88	171	132	391	215	307	141	663	0.109	10
Masroor	2015	India	Asian	45	48	7	100	71	26	3	100	0.743	6

Note: HWE, Hardy-Weinberg equilibrium.

In a study conducted by Saldivar *et al*. (2005), the *NQO1* genotype was associated with a marginally increased risk of lung cancer. The contradictory results were observed from some of the previous studies conducted in which the *NQO1* variant genotype exhibited a potential protective effect (Bock *et al*., 2005; Chao *et al*., 2006; Yang *et al*., 2007). In a previous study by Lewis *et al*. (2001) almost 4-fold increase in risk towards SCLC was observed in patients with at least one variant allele. Similarly, in NSCLC subjects, *TT* genotype has been reported with an increased risk of NSCLC (Tian *et al*., 2014; Masroor *et al*., 2015). Eom *et al*. (2009) reported no association for any different lung cancer histological cell types. This deviation may be due to the various factors *viz.* study population, age, gender, exposure levels of carcinogens may affect the *NQO1* enzyme activity. It has also been reported that the association of the *NQO1* C609T polymorphism in other types of malignancies had conflicting results wherein some studies observed an increased risk and other studies showed a lack of association for several other malignancies, e.g., colorectal cancer, breast cancer, bladder cancer, prostate cancer, cervical cancer (Chao *et al*., 2006; Yang *et al*., 2007; Wang *et al*., 2008; Sameer *et al*., 2010; Yuan *et al*., 2011; Lajin and Alachkar, 2013).

The earlier studies have reported an increased risk of lung cancer in heavy smokers and carrying heterozygous (Eom *et al*., 2009) and mutant genotype (Masroor *et al*., 2015). In a study conducted by Masoor *et al*. (2015), the *NQO1* polymorphism exhibited a protective effect in lung cancer progression to other organs. Various studies have reported the role of NQO1 protein expression in various cancers, *viz.* NSCLC (Li *et al*., 2015), gastric ADCC (Nunobiki *et al*., 2015), ovarian carcinoma (Lin *et al*., 2014), and pancreatic ADCC (Ji *et al*., 2017), indicating that NQO1 expression was associated with the occurrence of tumors and that it may be a significant predictive or a prognostic marker.

Regarding the *NQO1* polymorphism with treatment response for different chemotherapeutic regimens and its correlation with objective response, Tian *et al*. (2014) reported that the *TT* genotype carriers had less chance of receiving chemotherapy than *CC* genotype carriers.

Few studies have been studied to examine the overall survival of patients with LC or other cancers. A previous study by Kolesar *et al*. (2011) reported that the *NQO1* mutant genotype in NSCLC patients was associated with poor survival. Masroor *et al*. (2015) observed

that NSCLC subjects carrying the *CT* and *TT* genotype were also associated with poor survival. Various studies have been conducted in several other cancers *viz.* lung cancer (Kolesar *et al*., 2011; Cui *et al*., 2014; Li *et al*., 2015), breast cancer (Yang *et al*., 2014), uterine cervix (Hou and Lin, 2016), pancreatic adenocarcinoma (Ji *et al*., 2017) have indicated that *NQO1* polymorphism or over-expression is associated with cancer progression and maybe a possible biomarker for inadequateprognostic assessment. Further, in the study carried on NSCLC patients, the *NQO1* polymorphism played an essential role in clinical outcomes in patients who underwent radiotherapy after surgery and platinum-based chemotherapy (Song *et al*., 2010; Tian *et al*., 2014). The loss in NQO1 activity may increase the cancer risk due to a reduction in antioxidant activity in the cell and the increased probability of one electron reduction of quinones to semiquinones, which may attack proteins, DNA, and lipids (Cui *et al*., 2015).

2.12.2. Genetic polymorphisms in *SULT1A1* gene in association with lung cancer risk and overall survival of LC patients

Tobacco smoking is considered a significant risk factor for the development of lung cancer, and 90% of lung cancer cases are found to be attributed to smoking (de Groot *et al*., 2018). SULT1A1, a phase II detoxification enzyme, plays a vital role in detoxifying carcinogens in cigarettes and hazardous chemicals (Wang *et al*., 2008). It has been postulated that polymorphism in the genes coding for carcinogen metabolizing enzymes affects the individual's susceptibility to cancer initiation. Among the different types of single nucleotide polymorphism found in the gene, $Arg^{213}His$ is the most studied polymorphism that replaces Arginine with Histidine amino acid in exon 7. SULT1A1*2 (^{213}His variant allele) derived protein has reduced thermal stability and twice a lower catalytic activity, comparatively to the wild type (SULT1A1*1) derived protein (Koike *et al*., 2008; Xiao *et al*., 2014). Based on crystal structure modeling, it has been postulated that the $Arg^{213}His$ amino acid substitution induces structural changes in the SULT1A1 protein, which affects substrate binding ability (Koike *et al*., 2008; Arslan *et al*., 2009; Kumar *et al*., 2013). In a study conducted by Ihsan *et al*. (2011), a reduced susceptibility towards lung cancer was seen in subjects in a population from India. The study by Tamaki *et al*. (2011) found no association with lung cancer risk. Various case-control studies were conducted to find out the association between *SULT1A1* $Arg^{213}His$ polymorphism and different cancer types were contradictory. The *SULT1A1* His^{213} allele has been associated with

an increased risk of stomach cancer (Boccia *et al*., 2005), breast cancer (Lee *et al*., 2012), and urothelial cancer (Liang *et al*., 2004). On the other hand, reduced risks for prostate cancer (Nowell *et al*., 2004), colorectal cancer (Nowell *et al*., 2002), bladder cancer (Hung *et al*., 2004; Zheng *et al*., 2003), and lymphocytic leukemia (Sak *et al*., 2019) have been reported with the *His*^{213} genotype. Chung *et al*. (2009) studied the association between *SUTT1A1* polymorphisms and oral squamous cell carcinoma (OSCC) susceptibility and found that the presence of *Arg*^{213}*His SULT1A1* polymorphisms was not associated with the risk of developing oral cancer. These inconsistencies in different studies may be due to differences in selection criteria of patients and controls, differences in ethnic composition among populations, and sample size. Smoking and exposure to other environmental pollutants areother major confounding factors.

In a study conducted by Rouprêt *et al*. (2007), the histological types exhibited an increased risk of lung cancer. In earlier studies conducted by Wang *et al*. (2002) and Liang *et al*. (2004), stratification based on gender exhibited an increased risk towards lung cancer. As SULT1A1 also carries out the bioactivation of tobacco smoke procarcinogens and smoking is a prime risk factor for lung cancer, the *SULT1A1 Arg*^{213}*His* polymorphism can be related to lung cancer carcinogenesis in smokers. In a study conducted by Ihsan *et al*. (2011), the authors reported a reduced risk for lung cancer for smokers. The variant genotype also did not exhibit an increased risk of lung cancer in the Japanese population (Tamaki *et al*., 2011). In a study conducted by Zheng *et al*. (2003), the result exhibited a reduced bladder cancer risk among light smokers (30 pack-years) with the *His*^{213} allele genotypes than among heavy smokers, but this was not statistically significant. These findings indicate that the low enzyme activity variant *His/His* may contribute to a decrease in bioactivation/or detoxification of carcinogens (such as PAHs); however, it might also be possible that other Phase II detoxification enzymes may play an essential role in reducing the risk of smoking-related cancers. The current smokers carrying *His/His* and *Arg/His* genotype exhibited an increased risk of developing lung cancer compared to never smokers (Liang *et al*., 2004; Chung *et al*., 2009; Tasnim *et al*., 2017). In a study evaluated by Silva Santos *et al*. (2012) in the Brazilian population, the authors reported that the association between cigarette smoking and oral cancer was higher in individuals with a *SULT1A1*1* isoform (wildtype genotype *Arg/Arg*) than in those individuals with at least one *SULT1A1*2* allele (*Arg/His* + *His/His*), therefore suggesting the latter would modulate the influence of cigarette smoking on oral cancer. An increased risk was found in smoking and *SULT1A1* polymorphism in

other cancers,*viz.* urothelial cancer (Wang *et al*., 2008), breast cancer (Lilla *et al*., 2005), and bladder cancer (Wang *et al*., 2017).

The prognosis of patients in association with *SULT1A1* polymorphism, the study conducted in breast cancer patients found that carriers of *SULT1A1*2 /*2* genotype had a higher risk of death in tamoxifen-treated women's (Nowell *et al*., 2002)

2.12.3. Genetic polymorphisms in *EPHX1* gene in association with lung cancer risk and overall survival of LC patients

Microsomal epoxide hydrolase (EPHX1) is a PhaseII enzyme that is not only involved in the general oxidative protection against a myriad of substances in the environment but is also primarily responsible for the xenobiotic activation of tobacco smoke carcinogens (Gresner*et al*., 2007). In the *EPHX1* coding region, two specific polymorphisms are characterized in exon 3 and 4 and studied extensively. In exon 3, $Tyr^{113}His$ (rs2854455) results in T to C substitution whereas, exon4, $His^{139}Arg$ (rs2234922) results in A to G substitution, and these polymorphic variants have been found to modulate the enzyme activity. The $His^{113}His$ allele has been found to be associated with at least 50% (slow allele) of reduced enzyme activity whereas the $Arg^{139}Arg$ allele increased the enzyme activity by 25% (short allele) (Hassett *et al*., 1994; Yu *et al*., 2015). Variations in the EPHX1 enzymatic activity due to such polymorphisms can lead to inter-individual variations in xenobiotics detoxification, thus conferring differences in susceptibility to cancers induced by chemical carcinogens (Lin *et al*., 2007; Mittal and Srivastava, 2007).

Several earlier studies conducted in the two polymorphic variants of this gene have been studied, and the results have been quite inconclusive (To-Figueras *et al*., 2001; Wu *et al*., 2001; Gsur *et al*., 2003; Park *et al*., 2005; Rosenberger *et al*., 2008; Graziano *et al*., 2009; Erkisi *et al*., 2010; Timofeeva *et al*., 2010; Ihsan *et al*., 2011; Tilak *et al*., 2011). These inconsistencies may be attributed to differences in selection criteria of patients and controls, differences in ethnic composition among populations, and sample size. Smoking or exposure to other environmental pollutants is another major confounding factor (Graziano *et al*., 2009; Fathy *et al*., 2014). Several earlier studies found no association between exon3 polymorphism and LC risk (Zhao *et al*., 2002; Cajas-Salazar *et al*., 2003; Graziano *et al*., 2009). Similarly, the study conducted by Timofeeva *et al*. (2010) in the German population also reported that the *EPHX1* $Tyr^{113}His$

polymorphism did not exhibit any significant association towards LC risk. In a study conducted in African-Americans by Wu *et al*. (2001), no association was observed between *EPHX1* $Tyr^{113}His$ (exon3) polymorphism and LC risk. However, an increased risk for lung cancer was observed in studies conducted in other populations for the variant genotype of *EPHX1* $Tyr^{113}His$ polymorphism, especially in Asian subjects (Tan *et al*., 2014). In a study conducted in Chinese, the exon3 $Tyr^{113}His$ polymorphism was associated with susceptibility for lung cancer (Persson *et al*., 1999). Similarly, in a study conducted on 175 lung cancer patients from north India, the analysis exhibited an association of $Tyr^{113}His$ polymorphism towards susceptibility for lung cancer and the subjects with haplotype His^{113}-His^{139} were at slightly higher risk for lung cancer. However, they were not significant (Tilak *et al*., 2011). On the contrary, several other studies have evaluated the exon3 $Tyr^{113}His$ polymorphism in other populations *viz.* South Indian (Ihsan *et al*., 2011), German (Rosenberger *et al*., 2008), Finnish (Voho *et al*., 2006), Austrian (Gsur *et al*., 2003), and Spanish (To-Figueras *et al*., 2001), have reported a decreased risk for lung cancer with the variant genotype.

In a meta-analysis study that was evaluated in 24 studies, the pooled findings showed no association between $Tyr^{113}His$ polymorphism and lung cancer risk (Wang *et al*., 2013). In contrast, subgroup analysis of Caucasians and Asians suggested that $Tyr^{113}His$ polymorphism was associated with decreased and increased risk for lung cancer. A meta-analysis conducted by Tan *et al*. (2014) reported no association between the *EPHX1* $Tyr^{113}His$ polymorphism and risk for lung cancer susceptibility. However, on the contrary, the analysis conducted by Yu *et al*. (2015) has reported that the $Tyr^{113}His$ polymorphic variant was a risk factor for lung cancer susceptibility. Further, in a meta-analysis conducted by Zhang *et al.* (2015), the patients carrying *CC* variant genotype (*i.e., His/His*) for $Tyr^{113}His$ polymorphism may be a risk factor in Asians, and the studies included in the analysis are tabulated in table 2.6. Thus, we could say that the results from different lung cancer subjects amongst different populations have been relatively inconclusive.

Table 2.6: Principal characteristics of the studies included in the meta-analysis (Zhang et al., 2015)

Firstauthor/Year	Country	Ethnicity	Case/Control			HWE	Score
			TT	TC	CC		
Benhamou/1998	France	Caucasian	82/64	46/77	22/31	0.36	6
Cajas-Salazar/2003	America	African	67/62	37/52	6/5	0.14	6
Erkisi/2010	Turkey	Caucasian	6/22	26/12	26/7	0.04	7
Fathy/2014	Egypt	African	28/78	14/18	8/4	0.04	6
Gemignani/2007	Mixed	Caucasian	136/151	83/82	31/27	0.00	8
Graziano/2009	Italy	Caucasian	26/40	14/32	2/0	0.02	7
Gsur/2003	Australia	Caucasian	147/224	114/218	16/54	0.93	5
Ihsan/2011	India	Asian	82/94	51/133	55/63	0.22	7
Liang/2004	China	Asian	36/48	87/76	29/28	0.83	5
London/2000	USA	Caucasian	85/237	82/184	15/37	0.88	6
London/2000	USA	African	106/153	48/77	1/12	0.57	6
Park/2005	USA	Mixed	81/138	72/147	25/80	0.01	7
Perez-Morales/2014	Mexico	Mixed	61/128	64/134	65/120	0.94	5
Persson/1999	China	Asian	21/41	33/59	20/22	0.92	5
Rosenberger/2008	Germany	Caucasian	55/49	38/43	7/8	0.74	6
Smith/1997	UK	Caucasian	25/91	20/99	5/13	0.04	5
Sun/2007	China	Asian	38/61	40/38	144/123	0.97	8
Tilak/2011	India	Asian	62/134	85/157	28/31	0.12	6
Timofeeva/2010	Germany	Caucasian	316/627	238/520	57/119	0.46	8
To-Figueras/2001	Spain	Caucasian	97/87	70/85	8/15	0.36	7
Voho/2006	Finland	Caucasian	133/1029	81/865	13/189	0.71	6
Wang/2012	China	Asian	54/97	70/82	85/77	0.96	5
Wu/2001	USA	Mixed	20/28	26/29	5/7	0.90	5
Wu/2001	USA	African	40/38	22/20	3/4	0.54	5
Yin/2001	China	Asian	15/24	54/46	15/14	0.31	7
Yoshikawa/2000	Japan	Asian	24/35	35/51	12/21	0.76	5
Zhao/2002	USA	Caucasian	86/77	56/54	20/22	0.02	7
Zhou/2001	USA	Caucasian	465/581	332/355	177/206	0.97	6

Earlier studies conducted in different populations that are Chinese (Yin *et al*., 2001), Caucasian's (Park *et al*., 2005; Erkisi *et al*., 2010), Austrian (Gsur *et al*., 2003), American (Cajas-Salazar *et al*., 2003), and North Indians (Tilak *et al*., 2011), no positive association was found between exon4 and LC risk. In astudy conducted in African-Americans by Wu *et al*. (2001), no association was observed between *EPHX1 His^{139}Arg* (exon4) polymorphism and LC risk.

On the contrary, the *EPHX1 His^{139}Arg* variant genotype depicted an increased risk for lung cancer in a study from North East Indian lung cancer subjects, and further, no haplotype was significantly associated with lung cancer risk (Ihsan *et al*., 2011). Prior studies conducted in Italian and Mexican-American lung cancer patients (Graziano *et al*., 2009; Cajas-Salazar *et al*., 2003) have reported a 3-fold increased risk for lung cancer susceptibility for subjects harboring the variant *EPHX1 His^{139}Arg* genotype. However, in a study conducted in Spanish lung cancer patients, exon4 was associated with decreased LC risk (To-Figueras *et al*., 2001). In a meta-analysis conducted by Liu *et al*. (2013), both the heterozygous and combined genotypes of *EPHX1 His^{139}Arg* polymorphism were associated with an increased risk for lung cancer. In another study in the Asian population, the *EPHX1 His^{139}Arg* polymorphism was associated with risk towards lung cancer susceptibility (Yu *et al*., 2015). In a meta-analysis conducted by Xu *et al*. (2015), the *EPHX1 His^{139}Arg* polymorphism was associated with the risk of lung cancer among the Asian population, and the studies included are demonstrated in table 2.7.

In a study conducted on SQCC subjects study carried out by Yin *et al*. (2001), the authors reported an association between variant genotype of *Tyr^{113}His* and SQCC, and further, no association was observed between gender and lung cancer risk. In a study by Zienolddiny *et al*. (2008) carried out in NSCLC subjects, the *His^{139}Arg* polymorphism was associated with NSCLC. A study by Park *et al*. (2005) found a significant increase in risk for ADCC for the high activity genotype of *EPHX1*. Similarly, Cajas-Salazar *et al*. (2003) also pointed out an increase in risk for ADCC among those subjects harboring the high enzyme activity genotypes. A study by Park *et al*. (2005) revealed that when the subjects were stratified based upon predicted activity phenotype, subjects with high activity genotypes exhibited an increase in LC risk compared to the subjects having low genotype activity.

Table2.7: Distribution of *EPHX1 His^{139}Arg* among lung cancer cases and controls (Xu et al., 2015)

Firstauthor,year	Ethnicity (country oforigin)	Totalsample (case/control)	Lungcancer			Controls			OR(95%CI)
			AA	AG	GG	AA	AG	GG	
WangN,2012	Asian(China)	209/256	161	45	3	207	46	3	1.26(0.80-1.97)
IhsanR,2011	Asian(India)	188/290	121	59	8	212	70	8	1.50(1.01-2.24)
TilakAR,2011	Asian(India)	175/322	103	62	10	187	117	18	1.05(0.72-1.52)
SunXW,2007	Asian(China)	222/222	168	48	6	182	36	4	1.46(0.92-2.32)
LiangGY,2004	Asian(China)	152/152	115	36	1	128	23	1	1.72(0.97-3.04)
YinL,2001	Asian(China)	84/84	65	18	1	69	14	1	1.34(0.63-2.87)
YoshikawaM,2000	Asian(Japan)	71/107	45	24	2	75	28	4	1.35(0.72-2.56)
PerssonI,1999	Asian(China)	74/117	54	20	0	97	17	3	1.80(0.89-3.63)

AA,indicateswild-type;AG,indicatesheterozygote;GG,indicatesvarianthomozygote.EPHX1,epoxidehydrolase1.

A significant increase in LC risk was observed based on enzymatic activity phenotype for the high enzyme activity genotype (Cajas-Salazar *et al*., 2003). The study conducted by Zhao *et al*. (2002) also found that high enzymatic activity phenotype had an elevated LC risk. In a study conducted in Caucasians,the predicted activity exhibited no association (London *et al*., 2000). On similar lines, Zhou *et al*. (2001) reported no association between *EPHX1* activity phenotypes and lung cancer risk. The subjects who were smokers and carrying the single variant allele for *EPHX1 His*^{113}*His* and *Tyr*^{113}*His* genotypes in the North Indian population showed more than 4-fold and a 2.3-fold risk of LC, respectively (Tilak *et al*., 2011). The *EPHX1 Tyr*^{113}*His* and *His*^{139}*Arg* polymorphisms were also associated with an increased risk of lung cancer in the group of heavy smokers in German lung cancer patients, and no association was observed between gender and LC risk (Timofeeva *et al*., 2010). A clear relationship was observed between the genotypes of *EPHX1 Tyr*^{113}*His* and the incidence of lung cancer among smokers in a study conducted by Yin *et al*. (2001). On the contrary, earlier case-control study, the subjects who were smokers with both the *His*^{113} variant alleles tended to have a decreased risk for lung cancer compared to subjects who were smokers but carrying the homozygous wild-type *Tyr*^{113}*/Tyr*^{113} genotype (Voho *et al*., 2006). Lack of association for the *EPHX1 Tyr*^{113}*His* polymorphism and smoking towards modulating risk for LC has to been reported by earlier studies (Lee *et al*., 2002; Zhao *et al*., 2002).

In Turkish lung cancer subjects, the study carrieda 2-fold increased odds of lymph node invasion was reported (Erkisi *et al*., 2010). In a study on breast cancer subjects treated with tamoxifen, it was found that mEH might have a role in suppressing tamoxifen response and causing poor prognosis (Lin *et al*., 2007). A study conducted by Lin *et al*. (2007) found a significant difference in patient survival when mEH expression and chemotherapy containing Adriamycin were administered. Therefore, more studies on a large scale need to be conducted to evaluate the role of polymorphisms in the *EPHX1* gene and its role in chemotherapeutic response and toxicity.

2.12.4. Genetic polymorphisms in *NAT2* gene in association with lung cancer risk and overall survival of LC patients

N-acetyltransferase (NAT) are necessary Phase II enzymes involved in catalyzing the acetylation of aromatic amines and hydrazines, including therapeutic drugs and carcinogenic compounds,

and function as modulator activity of drugs and carcinogen detoxifying agents (Mota *et al*., 2015). The polymorphisms in the *NAT2* gene locus result in impaired enzyme activity (Borlak and Reamon-Buettner, 2006). There are three NAT2 phenotypes known: a rapid acetylation phenotype consisting of two copies of wild type allele, intermediate phenotype with the single allele of wild type and a mutant allele, and a slow acetylation phenotype consisting of two mutant alleles. Thus, the rapid acetylators are the subjects homozygous or heterozygous for the *NAT2* wild type allele, while those homozygous for the mutant allele are classified as slow acetylators (Seow *et al*., 1999; Mahasneh *et al*., 2012). The functional status of the phenotype is due to impaired translation or stabilization of the protein. No changes in mRNA levels have been observed (da Silva *et al*., 2011).

Genetic and environmental factors are the criticalfactors in the etiology of lung cancer. The lung cancer risk associated with exposure to endogenous or exogenous xenobiotics compounds may be modified by genetic variation in metabolic detoxification activities (Liu *et al*., 2015). Henceforth, the relevance of *NAT2* polymorphisms to the risk of lung cancer is therefore of particular importance. Several earlier studies were conducted to evaluate the *NAT2* polymorphic variants of this gene, and the results have been quite inconclusive (Cascorbi *et al*., 1996; Oyama *et al*., 1997; Chen *et al*., 2006; Lee *et al*., 2009; Sobti *et al*., 2009). These inconsistencies may be attributed to differences in selection criteria of patients or controls, differences in ethnic composition among populations, and sample size. Smoking or exposure to other environmental pollutants is another major confounding factor.

In several earlier studies done, no association was found between *590G>A (NAT2*6)* and *803A>G (NAT2*5C)* polymorphism and LC risk (Martinez *et al*., 1995; Bouchardy *et al*., 1998; Hansen *et al*., 2001; Zhou *et al*., 2002; Borlak *et al*., 2006). Perera *et al.* (2006) conducted the study in Caucasians and have reported no association between the *590G>A (NAT2*6)* and *803A>G (NAT2*5C)* polymorphisms and LC risk. In another study in Caucasian lung cancer subjects, these two polymorphic variants of the *NAT* gene did not appear to modify the risk for LC (Wikman *et al*., 2001). However, on the contrary, some studies suggest the role of the *NAT2* polymorphisms in modulating lung cancer risk. Rusinova *et al*. (2014) reported that the variant genotype of *NAT2*6 590G>A* impacted lung cancer risk. Studies done in Mongolian and Chinese Han subjects have shown that the slow acetylators exhibit a higher risk of lung cancer.

Sobti *et al*. (2009) have also suggested that slow acetylators for *NAT2*5* and *NAT2*6* exhibited a higher risk of lung cancer in the North Indian population. Meta-analysis data reported a lack of association between the *NAT2* polymorphisms slow acetylator phenotypes and lung cancer risk by Cui *et al*. (2011) and Liu *et al*. (2015).

In a study conducted by Gemignani *et al*. (2007), lung cancer risk was observed in subjects harboring heterozygous genotype for *857G>A* polymorphism of *NAT2*. Similarly, in a meta-analysis conducted, an association was observed between the *NAT2* polymorphism and lung cancer susceptibility for *857G>A* polymorphism of *NAT2* (Cui *et al*., 2011). The acetylation status for the synonymous variation c.*481C>T* is not clear because it is also associated with *NAT2*12C* (c.*481C>T*+c.*803A>G*), which is a rapid allele. In a study conducted in Central Europe, the authors reported that *NAT2*4/*5* conferred a reduced risk towards lung cancer for the *481C>TNAT2* polymorphism (McKay *et al*., 2008). Cascorbi *et al*. (1996) reported that fast acetylators were over-represented of the fast acetylators (*NAT2*4*/4*) in German lung cancer patients *vs*. healthy subjects, thus supporting the role of ultra-rapid acetylation capacity in lung cancer risk. Further, in a meta-analysis conducted by Zhu *et al.* (2021), the study includes 18 studies (Table 2.8), and the result depicts that the slow + intermediate phenotype of *NAT* has been found to significantly increase the risk of lung cancer. Table 2.8 demonstrates the earlier studies included in this meta-analysis and includes the OR of different studies carried out earlier.

Based on histological subtypes varying results were observed by different studies. In an earlier study conducted, a 2-fold increase in the risk of ADCC was observed for the *NAT2*7* polymorphism in Chinese lung cancer subjects of Singapore (Seow *et al.* 1999). A study conducted on SQCC subjects in Chinese lung cancer patients reported a decreased risk of SQCC for the slow acetylators (Tian *et al*., 2014). In North Indian lung cancer patients, an increased risk for SQCC, SCLC, and ADCC was observed (Sobti *et al*., 2009). In a study conducted in the Slovakian population, *NAT2*5B/*6* was found to confer a risk for SQCC (Habalova *et al*. 2005). Whereas, in several earlier studies conducted, no association was observed between the *NAT2* polymorphism and different histological subtypes (Saarikoski *et al*., 2000; Zhou *et al*., 2002; Skuladottir *et al*., 2005; Perera *et al*., 2006; Zienolddiny *et al*., 2008).

Table 2.8: The different studies conducted in *NAT2* polymorphism and included in the meta-analysis (Zhu et al., 2021)

First author	Year	Coutry	Ethnicity	Slow (case)	Rapid (case)	Slow (control)	Slow (control)	OR (95%CI)
Mota	2015	Portugal	European	153	36	168	61	1.54(0.97-2.46)
Zhang	2014	China	Asian	135	795	113	887	1.33(1.02-1.74)
Mahasneh	2012	Jordan	Asian	22	27	61	38	0.51(0.25-1.02)
Zupa	2009	Italy	European	43	32	50	71	1.91(1.06-3.42)
Sobti	2009	India	Asian	50	252	33	269	1.62(1.01-2.59)
Lee	2009	China	Asian	22	95	36	83	0.53(0.29-0.98)
Osawa	2007	Japan	Asian	14	99	9	112	1.76(0.73-4.24)
Chen	2006	China	Asian	18	79	43	154	0.82(0.44-1.51)
Chiou	2005	China	Asian	27	135	64	144	0.45(0.27-0.75)
Sorensen	2005	Denmark	European	156	99	143	121	1.33(0.94-1.89)
Skuladottir	2005	Denmark	European	154	87	321	219	1.21(0.88-1.65)
Belogubova	2005	Norway Russia	European	99	79	426	289	0.85(0.61-1.18)
Wikman	2001	German	European	237	151	196	149	1.19(0.80-1.60)
Hou	2000	Norway	European	169	112	237	138	0.88(0.64-1.21)
Saarikoski	2000	Finland	European	102	93	152	140	1.01(0.70-1.45)
Seow	1999	China	Asian	60	93	36	105	1.88(1.14-3.10)
Nyberg	1998	Sweden	European	113	70	96	62	1.04(0.67-1.61)
Martinez	1995	Spain	European	61	47	136	107	1.02(0.65-1.61)

Note: Phenotype distribution: slow represents the slow acetylation status and rapid represents th erapid+intermediate acetylation status.

In another study conducted in the Chinese population, a marginal increase in risk for adenocarcinoma was observed (Seow *et al.*, 2001). Similarly, Chiou *et al*. (2005) also pointed out an increase in risk for ADCC and SQCC in subjects harboring fast acetylator phenotype. In a meta-analysis conducted, no significant association between *NAT2* genotypes and histological subtypes by Cui *et al*. (2011). Wikman *et al*. (2001) discovered that combining slow acetylators of *NAT2* with rapid acetylators of *NAT1* increased the risk of ADCC. It has been demonstrated that combining the slow genotype of *NAT2* with the null genotype of *GSTM1* may act as a catalyst, conferring higher vulnerability to DNA adduct production, gene mutations, and LC susceptibility (Hou *et al*., 2000). Similarly, the combination of both *NAT2* slow acetylators and *CYP1A1* rapid acetylator were at the highest risk for lung adenocarcinoma (Perera *et al*., 2006).

The gene-environment interaction between *NAT2* and smoking, particularly cumulative exposure to cigarette smoke over time, may significantly increase the elevated risk. An earlier investigation in the Indian population found that slow acetylators for both *NAT2*5* and *NAT2*6* polymorphisms had a 2.8-fold and 3-fold incidence of lung cancer in both smokers and non-smokers (Sobti *et al*., 2009). Seow *et al*. (1999) found no link between the *NAT2* polymorphism and smokers among Chinese lung cancer patients. According to Sørensen *et al*. (2005), the *NAT2* rapid acetylator phenotype of *NAT2* appeared to be protective against lung cancer in light smokers. A Japanese study found that the risk of lung cancer was considerably higher among light smokers who were intermediate-slow *NAT2* acetylators (Osawa *et al*., 2007). Slow acetylators, according to Nyberg *et al*. (1998), increase the risk of non-smokers, whereas rapid acetylators increase the risk of smokers. Smokers with *NAT2* slow acetylator had a 1.84-fold risk of lung cancer in another study conducted in the Mongolian and Han populations (Zhang *et al*., 2014).

Furthermore, the *NAT2* polymorphism did not appear to change lung cancer risk based on smoking status in Caucasian and Taiwanese populations (Wikman *et al*., 2001; Lee *et al*., 2009). A previous study has shown that cigarette smoking was not a risk factor for breast cancer among rapid acetylators (Hunter *et al*., 1997). Thus, high CYP1A2 activity has been observed to raise the risk of lung cancer in nonsmokers, but this impact was only seen when high CYP1A2 activity was paired with slow NAT2 activity. Data from a prior study corroborated the concept that high CYP1A2 activity and NAT2 intermediate-slow acetylator activity synergize lung cancer risk

(Dik *et al*., 2013). In one study conducted by Zhou *et al*. (2002) in the Caucasian population, no significant association was found when disease stages were compared.

Thus more studies need to be conducted in the North Indian population to evaluate their association with lung carcinogenesis and the interactive effect of smoking on *NAT2* variants and its association with lung cancer. Further, the role of these SNPs towards the overall survival (OS) of patients treated with platinum-based Chemotherapy and clinical outcomes too need to be evaluated.

2.12.5. Genetic polymorphisms in *GSTP1* gene in association for clinical outcomes in lung cancer and toxicity after chemotherapy

Glutathione-S-Transferase (GST) is a family of cytosolic enzymes that have been linked to play chemotherapeutic drug metabolism and detoxification. Glutathione S transferase P1 (GSTP1) is one of the GST isoforms implicated in the metabolism of anticancer drugs like platinum-based drugs. GSTP1 catalyzes the binding of cisplatin to glutathione, converting it to a water-soluble form that can be easily excreted (De Luca *et al*., 2013). A non-synonymous *A/G* single nucleotide polymorphism at nucleotide position 313 in exon 5 results in the substitution of adenine to guanine (SNP ID: rs1695), which changes the amino acid isoleucine to valine at position 105. This change in the encoded GSTP1 protein reduces enzyme activity, as well as substrate selectivity and thermal stability, reducing the biotransformation processes of chemotherapeutic drugs mediated by GSTP1 (Goričar *et al*., 2015). The *GSTP1* polymorphism influences the downstream detoxification process of cisplatin and the production of platinum-GSH conjugates. These conjugates may accumulate in the body, altering patient response rates and increasing the risk of hematologic and gastrointestinal toxicity (Yuan *et al*., 2015). Previous research has suggested that genetic variability in the *GSTP1 IIe105Val* genes may affect lung cancer treatment outcomes. However, the results are inconclusive (Booten *et al*., 2006; Ada *et al*., 2010; Yang and Xian, 2014).

Published data have indicated that the *GSTP1 IIe*^{105}*Val* variant might be associated with the efficacy of platinum-based chemotherapy in a lung cancer patient. The *GSTP1 IIe*^{105}*Val* polymorphism was associated with NSCLC patient survival in research conducted by Booton *et al.* (2006). A Chinese study found that *GSTP1 IIe*^{105}*Val* influences the clinical prognosis of

NSCLC patients (Liu *et al*., 2015). Similarly, Deng *et al*. (2015) observed that NSCLC patients with the mutant genotype of *GSTP1 IIe*^{105}*Val* polymorphism had recurrence-free survival. Jia *et al*. (2016) discovered that advanced NSCLC patients with mutant genotype (*Val/Val*) had a better OS and clinical response. However, studies that have looked into the relationship between *GSTP1 IIe*^{105}*Val* polymorphism and the prognosis of advanced NSCLC found ambiguous results (Ada *et al*., 2010; Zhou *et al*., 2011; Ke *et al*., 2012; Lv *et al*., 2014; Yang *et al*., 2014; Han *et al*., 2015; Liu *et al*., 2015; Deng *et al*., 2018). It has been observed that patients with Asian lung cancer who contain the *Val* allele of *GSTP1 IIe*^{105}*Val* polymorphism had a better survival time and clinical response than those with the wild type genotype (Ada *et al*., 2010; Zhou *et al*., 2011; Ke *et al*., 2012; Lv *et al*., 2014; Ha *et al*., 2015). Others have observed that the *Val* allele of *GSTP1 IIe*^{105}*Val* has been linked to poor survival and clinical response (Deng *et al*., 2015; Liu *et al*., 2015). On the other hand, other studies have revealed contradictory results. For example, a study conducted in the Chinese population by Lu *et al*. (2006) and in the Turkish population by Ada *et al*. (2010) found that *GSTP1 IIe*^{105}*Val* polymorphism was not related with a better OS in patients with advanced NSCLC. Furthermore, individuals with the mutant form of the *GSTP1* gene have a lower ability to detoxify the drug metabolites, promoting the OS of NSCLC subjects (Lu *et al*., 2006; Ke *et al*., 2012; Lv *et al*., 2014; Deng *et al*., 2015; Han *et al*., 2015).

The studies have also been conducted to investigate the impact of *GSTP1* polymorphism and response concerning the treatment outcome of lung cancer patients who underwent platinum-based chemotherapy. So far, data have suggested that the variant allele of the *GSTP1* may be linked to the efficacy and clinical response to platinum-based chemotherapy in lung cancer patients (Lv *et al*., 2014; Deng *et al*., 2015; Han *et al*., 2015; Liu *et al*., 2015; Zhao *et al*., 2015; Bu *et al*., 2016; Chen *et al*., 2016; Jia *et al*., 2016). The *GSTP1 IIe*^{105}*Val* polymorphism was significantly associated with clinical response in NSCLC patients (Booten *et al*., 2006). In a Chinese study, Liu *et al*. (2015) discovered that the *GSTP1 IIe*^{105}*Val* polymorphism affects the clinical outcome of lung cancer patients who have received platinum-based chemotherapy. Similarly, other studies in Chinese lung cancer patients have found that *GSTP1* polymorphism correlates with chemotherapy response (Lv *et al*., 2014; Han *et al*., 2015). Meta-analysis research found that a *GSTP1* polymorphism predicts the response to platinum-based chemotherapy in East-Asian lung cancer patients (Yang *et al*., 2014). Vlachogeorgos *et al*. (2008) found that the *GSTP1* polymorphism may predict therapy response in individuals with advanced NSCLC in

Caucasian participants, particularly Greek lung cancer patients. On the contrary, some studies have found no link between the *GSTP1 IIe^{105}Val* polymorphism and response. *GSTP1* polymorphism was not associated with treatment outcome in advanced NSCLC by Kalikaki *et al*. (2009) or Ada *et al*. (2010). The inconsistent results in different populations might be caused by the differences in ethnicities, source of patients, and sample size. Lung cancer is a complicated disease, and different ethnicities have different genetic backgrounds, which may affect the anti-cancer therapeutic outcome of platinum-based chemotherapy. Some studies have reported that platinum resistance is not associated with *GSTP1* (Galluzzi *et al*., 2012).

GSTP1 appears to develop resistance to various treatments, including etoposide, adriamycin, cisplatin, and carboplatin (Nakagawa *et al*., 1988). As a result, the beneficial effect of the *GSTP1 IIe^{105}Val* variant allele cannot be connected to a specific medicine that is concerned with treatment response. A possibility is that a tumor's response to cisplatin can also be related to different drug combinations. However, the benefit of specific chemotherapeutic treatments might remain unknown until it is proved that GSTP1 detoxifies the drugs in question.

The existence of changes in the GSTP1 proteins, which results in a reduction of the transport capacity for cisplatin in patients with mutant genotype (*Val/Val*), may explain inter-individual variability in platinum-based toxicity (Stoehlmacher *et al*., 2004). GSTP1 is a phase II detoxification enzyme that primarily functions in the detoxification of platinum-based substances. *GSTP1* polymorphism results in *GSTP1* modification, which results in GSTP1's altered ability to metabolize chemotherapeutic drugs and alters drug responses (Allan *et al*., 2001). It has been documented that individuals who carry both mutant alleles for the *GSTP1 IIe^{105}Val* polymorphism are more sensitive to inflammatory symptoms (Pincinato *et al*., 2019). The decrease in GSTP1 activity in lung cancer patients with the heterozygous or mutant genotype may increase exposure to platinum-based compounds. Thus, a proportional increase in drug toxicity due to reduced GSTP1 mediated catalyzed reactions, resulting in decreased metabolism of platinum-based drugs.

The association of *GSTP1 IIe^{105}Val* genotypes with the clinical outcomes of platinum-based chemotherapy is still controversial. Individuals with mutant or heterozygous genotypes had a lower risk of anemia, according to Deng *et al*. (2015). According to Zhang *et al*. (2011), patients with breast cancer who have the *GSTP1 Val/Val* genotype are more likely to have severe

toxicity. According to Yoshihama *et al*. (2018), patients with the *GSTP1* wild-type genotype had higher grade 3 or 4 hematological toxicity rates than patients with the mutant *GSTP1* genotype. Therefore more such studies in different populations (especially the Indian population) need to be carried out to get an idea of chemotherapeutic related toxicities in the lung cancer patient.

2.12.6. Genetic polymorphisms in *MTHFR* gene in association for clinical outcomes in lung cancer and toxicity after chemotherapy

Methylenetetrahydrofolate reductase (MTHFR) is an enzyme that catalyzes the conversion of 5,10-methylenetetrahydrofolate (5,10-MTHF) to 5-methylenetetrahydrofolate (5-MTHF). MTHFR's role in cancer susceptibility stems from its involvement in two folate metabolism pathways: decreased MTHFR activity can reduce homocysteine methylation to methionine, lowering the level of S-adenosyl-methionine (SAM), resulting in DNA hypomethylation; decreased MTHFR substrate can result in uracil misincorporation into DNA, decreased DNA repair, and an increase in the frequency of chromosomal breaks and damage (Krajinovic *et al*., 2004).

Several single nucleotide polymorphisms (SNPs) in the *MTHFR* gene have been found, including *C677T* and *A1298C*. At nucleotide 677, the C-to-T transition results in an amino acid change of valine for alanine, reducing MTHFR's activity and thermal stability (Krajinovic *et al*., 2004; Yang *et al*., 2016). The *MTHFR A1298C* polymorphism (rs1801131) in exon7, which results in the substitution of glutamate with alanine at amino acid 429 (*E429A*), may increase serum folate levels, potentially influencing the risk of cancer (Yang *et al*., 2016; Parle-McDermott *et al*., 2006). Lung cancer patients with the mutant genotype had a higher chance for progression-free survival when treated with pemetrexed (Smit *et al*., 2009). It was found that colorectal cancer patients with the mutant genotype for *MTHFR 677T* show a better response to either 5-FU-based chemotherapy or FOLFOX therapy (Boige *et al*., 2010). It has previously been reported that the therapeutic effect and ADRs of pemetrexed-based therapy are found to be individually specific, thus explaining the genetic association between different patients and their related ADRs.

Several studies have been conducted to examine the prognosis of cancer patients. The *MTHFR A1298C* polymorphism was an independent prognostic factor associated with a poor

prognosis in a study of German gastric cancer patients (Blank *et al*., 2014). The *MTHFR C677T* polymorphism did not affect survival in patients with the mutant genotype. At present, the relationship between *MTHFR* SNPs and 5-FU efficacy is inconclusive, and few studies have looked into the relationship between *MTHFR* SNPs and pemetrexed efficacy. Jakobsen *et al*. (2005) discovered that patients with the *MTHFR TT* genotype for the *C677T* polymorphism had a considerably greater response rate in metastatic colorectal cancer patients treated with 5-FU when compared to two other genotype groups. However, several large-scale clinical trials failed to find a direct association (Ruzzo *et al*., 2006; Goekkurt *et al*., 2009; Boige *et al*., 2010).

Furthermore, a large meta-analysis discovered that *MTHFR C677T* is not a strong predictor of FU-based therapy efficacy (Zintzaras *et al*., 2009). Patients with an *MTHFR TT* homozygous mutation exhibited a more prolongedprogression-free survival (PFS) compared to patients with wild-type or heterozygous NSCLC who were administeredwith pemetrexed, according to Smit *et al*. (2009). Argiris *et al*. (2011), on the other hand, presented contradictory results. Kucharczyk *et al*. (2016) discovered a modest (but not statistically significant) difference in OS among bearers of the *MTHFR C677T* polymorphism. In a breast cancer investigation, the results depict that the variant allele (*CC*) of *A1298C* had lower survival relative to the wild genotype (*AA*), whereas the variant allele of the *C677T* polymorphism had higher survival (not significant) (Martin *et al*., 2006). The correlation of *C677T* polymorphism with OS in glioblastoma multiforme (GBM) did not show any therapy-specific effect (Linnebank *et al*., 2008). According to Faria *et al*. (2020), patients with recurrent glioblastoma who had the mutant (*TT*) genotype for the *C677T* polymorphism and were receiving perillyl alcohol (POH)-based therapy had a longer survival time than those who had the wild genotype (*CC*).

Pemetrexed is becoming more widely utilized as first and second-line therapy and maintenance therapy for advanced NSCLC patients with “non-squamous” (i.e., ADCC) histology. Tiseo *et al*. (2013) discovered a substantial correlation between the *TT* genotype (*C677T*) and increased OS in NSCLC patients treated second-line with pemetrexed or pemetrexed/carboplatin. Patients receiving cisplatin plus pemetrexed had a longer OS duration than those receiving cisplatin and gemcitabine in phase III clinical trial done by Scagliotti *et al*. (2008). Li *et al*. (2013) discovered no relationships between the *MTHFR C677T* polymorphism and pemetrexed effectiveness in advanced ADCC LC subjects. The *TT* genotype of the *MTHFR*

C677T polymorphism is related with prolonged PFS and OS when compared to other genotypes (*CC* and *CT*) in two clinical studies for polymorphisms associated with pemetrexed effectiveness, GOIRC 02–2006 (Ardizzoni *et al*., 2012) and NVALT-7 (Smit *et al*., 2009). According to Krawczyk *et al*. (2014), the *MTHFR C677T* polymorphism did not affect OS or PFS. In a study of patients with advanced CKD and ESRD, Jamison *et al*. (2009) discovered that the *TT* genotype of the *C677T* polymorphism was associated with worse survival, and the *CC* genotype of the *A1298C* polymorphism was associated with improved survival in patients with ESRD but not patients with advanced CKD. A study in the Netherlands indicated that the *MTHFR C677T* polymorphism was related to outcomes in which individuals bearing the *TT* genotype had poorer progression-free survival in HER2-negative gastric cancer patients treated with bevacizumab combined with docetaxel, oxaliplatin, and capecitabine (Meulendijks *et al*., 2016). The previous investigation by Takehara*et al*. (2005) found no statistically significant link between clinic-pathological features of *MTHFR C677T* polymorphisms and their role in the OS of LC participants. However, they revealed a novel relationship between *TS* available type, *MTHFR* genotype, and prognosis in stage I NSCLC. These findings suggested that the MTHFR genotype could influence the effect of the TS functional form on NSCLC tumor biology. As *TS* and *MTHFR* compete for CH_2FH_4 in folate metabolism, this gene-gene interaction can be comparedto the folate status of NSCLC tissue. Tang and colleagues (2018) investigated the impact of the *C677T* polymorphism in patients with gastric cancer receiving 5-fluorouracil (5-FU)-based chemotherapy. There was no link between the *C677T* polymorphism and response rate or overall survival (Tang *et al*., 2018).

Studies on lung cancer and other cancers and diseases have investigated *MTHFR* polymorphism's role in toxicity among chemotherapy-treated patients. In a study by Lan *et al*. (2017), patients with the *AC/CC* genotype had a higher incidence of leucopenia, neutropenia, fatigue, and nausea than patients with *CC* genotype in non-squamous NSCLC patients treated with pemetrexed. Six patients experienced extreme toxic effects (grade 4 according to NCI-CTC criteria) during the first cycle of chemotherapy in the mono-institutional community of pre and postmenopausal patients with early breast cancer treated with classical CMF (cyclophosphamide, methotrexate, and fluorouracil) (Lan *et al*., 2017). In another study conducted by Toffoli *et al*. (2003) in ovarian cancer, 12 patients of 43 experienced G3/4 whomtoxicity (6 were G3/4 thrombocytopenia, 1 with G3 neutropenia, 1 with G3 anemia, 9 were G3 mucositis, and 1 G4

mucositis cases) and a strong correlation were seen between toxicity and *TT* genotype of *MTHFR C677T* polymorphism. The patients with *MTHFR C677T* genotype *CC* had significantly higher liver toxicity, calculated using CTCAE G3/G4. Nearly half of the patients experienced hematologic toxicity (leukopenia, neutropenia) to grade G3/G4, and even an increased tendency for mutant genotype patients to have G3/G4 hematologic toxicity was also observed.

A study conducted by Tang *et al*. (2018) reported that gastric cancer patients treated with 5-fluorouracil and carrying the mutant genotype (*TT*) for *C677T* polymorphism tended to present more hematologic toxicity than those with the *CC* or *CT* genotype. It has been found that the *MTHFR* is associated with the metabolism of 5-FU (Zhao *et al*., 2015). Therefore, subjects carrying the mutant genotype for *C677T* polymorphism reduce MTHFR activity, thus leading to inhibition of thymidylate synthase (TS) and an increase in efficiency 5-fluorouracil. This might lead to more toxicity to patients since 5-fluorouracil is cytotoxic (Tang *et al*., 2018). In patients treated with methotrexate for treatment of rheumatoid arthritis, overall toxicity was more common in patients with *TT* genotype, however for the *A1298C* polymorphism, the proportion of patients who experienced toxicity did not vary between genotypes (Urano *et al*., 2002). In context to the studies mentioned above, the polymorphism in the *MTHFR* genotype is linked with toxicity among the patients treated using chemotherapy. It has been reported that patients who were negative for Helicobacter pylori infection and carrying the mutant genotype for *C677T* had a higher probability of developing a severe form of atrophic gastritis, a precursor lesion for gastric cancer (Kong *et al*., 2020). These studies somewhat give information about the potential role of MTHFR as a predictive marker for hematological and gastrointestinal toxicity. However, before arriving at any definitive conclusion, more studies need to be carried out that will ultimately help to confirm the findings.

CHAPTER 3

GAPS IN STUDY AND OBJECTIVES

3.1 Gaps in study

The critical reason for cancer susceptibility lies in the polymorphism of the detoxification genes. In India, the cases of lung cancer and genetic variations are considerably high. Besides this, Indians consume local cigarettes more frequently containing exceptionally more xenobiotic compounds; therefore, there might be a high probability that the North Indian population may be more susceptible to risk for developing lung cancer. On the other hand, only a few polymorphism studies have been conducted in the Indian population. Hence it is necessary to study the association of polymorphism in Phase II detoxification genes with lung cancer. In addition, there have been no studies conducted in the Indian population regarding the role of the SNPs in Phase II detoxification genes and its association with clinical outcomes of lung cancer patients undergoing platinum-based chemotherapy. Therefore, identifying polymorphisms of Phase II genes and understanding their correlation with survival in lung cancer patients and susceptibility to lung cancer would yield insights into how these gene polymorphisms adversely or favorably influence the chemotherapy outcomes among North Indians. The selection of a chemotherapeutic regimen could improve response and hence the overall survival of cancer patients. Personalized chemotherapy as per the molecular marker could help improve lung cancer patients' response rate and survival. Further, these SNPs may be used as biomarkers, help design personalized chemotherapy regimens based on the genotype, and determine the association between these polymorphisms concerning clinical outcomes — objective response rates (ORRs) using RECIST criteria and overall survival (OS) among patients as personalized therapeutics.

3.2 Objectives of the study

1) To study the single nucleotide polymorphic variants of Phase II genes.
2) To evaluate the role of polymorphic variants of Phase II gene towards lung cancer susceptibility.
3) To test the association between Phase II gene polymorphisms and the clinical outcomes of lung cancer patients.

CHAPTER 4

MATERIALS AND METHODS

4.1 Design of the study

A case-control study design was chosen to evaluate the north Indian population for lung cancer predisposition and its association with the genetic variants of Phase II detoxification genes. Five hundred fifty lung cancer patients were recruited to study four genes, including *NQO1, SULT1A1, EPHX1,* and *NAT2* genes, from the Department of Pulmonary Medicine, Post Graduate Institute of Medical Education and Research (PGIMER) Chandigarh, India. The study also includes 550 healthy controls. Whereas for *MTHFR* gene 253 adenocarcinoma patients (here only cases that too diagnosed with ADCC are studied because *MTHFR* has been found to be correlated with pemetrexed and we want to access the toxicity profile and survival outcomes only) and for *GSTP1* 682 patients were studied (here the patients with any histology were taken and earlier the case-control study has been conducted in North Indian population, so there is no point repeating same kind of study), only case studies were conducted for these two genes. The subjects who visited PGIMER from February 2019 to March 2020 are included in this study. This study has been reviewed and approved by the institutional ethics committee of PGIMER. The ethical clearance letter is attached in Appendix-I. Informed written consent was obtained from all participants or their representatives, as attached in Appendix-II. The information of patients was also obtained from participants or their representatives, as attached in Appendix-III.

The enrolled patients had NSCLC or SCLC subtypes identified histopathologically. The study included newly diagnosed lung cancer patients with no limits on age, gender, stage, histology, or TNM. Patients who were under observation or had a history of cancer were excluded from the study.

Patients meeting all the following requirements shall be eligible for enrollment:

- Diagnosis of lung cancer (NSCLC or SCLC) is confirmed either by histology or cytology.
- Stage I, II, III or IV disease (although patients diagnosed with stage I/II were few).
- No age, gender, smoking, histology, and staging restrictions were applied.

- Untreated and intent to treat with definitive chemotherapy (Treated with platinum agents cisplatin/ carboplatin, either as the first or second line).
- An Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 0–2.
- At least one bi-dimensionally measurable lesion, according to the RECIST criteria
- Adequate organ function, defined as absolute neutrophil count>1500/μL, platelet count >100,000/μl, and levels of creatinine, liver enzymes, and alanine aminotransferase (ALT) less than two times the upper limits of normal (ULN).
- Written informed consent was obtained.

The exclusion criteria included the following:

- Patients had a prior history of any other carcinoma.
- Active infection or immunesuppression (HIV).
- Patients were receiving any systemic steroids.
- Patients are suffering from a chronic form of diarrhea due to any cause.
- Patients who are not undergoing chemotherapy.

However, numerous physiological tests were performed before recruitment, including a complete blood count, which determines neutrophil count, and biochemical tests such as liver and renal function tests. A chest X-ray and a contrast-enhanced computed tomography (CT) scan of the thorax and upper abdomen region were also performed on the patient.

Alongside, 550 unrelated individuals with no history of lung or any other cancer who visited PGIMER for regular health check-ups were recruited as a control group. An attempt was made to match controls based on age (±ten years), gender, and smoking parameters concerning cases (*i.e.*, to include patients from all age group, both genders and with less as well as high pack-years). A detailed questionnaire was filled by a trained interviewer for each case and control. The questionnaire included crucial demographic information such as age, gender, the patient's region, smoking, and drinking habits. Smokers reported tobacco habits such as cigarette and/or beedi smoking (a native cigarette-like stick of coarse tobacco hand-rolled in a dry tembuhurni leaf). Pack-years were calculated as an indicator of cumulative smoking exposure using the following formula:

$$\frac{\text{number of cigarettes/bidis smoked per day}}{20} \times number\ of\ years\ smoked$$

The essential medical information of cases, such as histology, TNM classification, clinical staging, performance status (KPS and ECOG), chemotherapeutic regimen, and objective response following the Response Evaluation Criteria in Solid Tumor Group (RECIST) criteria, was obtained from the hospital's medical records.

4.1.1 Chemotherapeutic regimen

All the lung cancer patients recruited in the current study received a doublet platinum-based chemotherapy drug combination. This combination included different drugs such as docetaxel, paclitaxel, irinotecan, and pemetrexed along with a platinum-based drug, such as cisplatin or carboplatin. These drugs were administered intravenously at a fixed dosage based on the body's surface area. The combination chemotherapy dosage consisted of administration of docetaxel (75 mg/m^2), pemetrexed (500 mg/m^2), or 100mg/m^2 irinotecan as a 1 hr infusion, followed by *iv.* infusion of 70 mg/m^2 cisplatin for over 3 hr. Standard folate and Vitamin B12 formulations were given to each patient as supplements. Certain drugs such as dexamethasone, granisetron, and ranitidine were also administered as a part of the premedication routine. Before carrying out tumor response assessment, four cycles of chemotherapy were administered as is standard procedure at PGIMER. If there was unacceptable toxicity or clinic-radiological evidence of disease progression before the completion of four cycles, the tumor response was assessed, and, if necessary, chemotherapy was discontinued. Patients who demonstrated an objective response to treatment were given two additional cycles (i.e., maximum of six cycles).

4.1.2 Follow-up and response assessment

After every two months, all subjects were followed up on telephonic conversation until the end of the study period. The survival time was considered from the date of diagnosis of lung cancer till the last date of follow-up or death due to uncertain reasons. The study's primary endpoint was overall survival, and the second was the relationship between genetic polymorphism and treatment responses. Response Evaluation Criteria in Solid Tumors (RECIST) criteria were used to evaluate tumor response. Based on RECIST criteria, responses are broadly classified as "responders" who exhibit complete response (CR) or partial response (PR) and "non-responders" who exhibit stable disease (SD) and progression disease (PD).

- CR: Complete disappearance of all target lesions.

- PR: At least a 30% decrease in the sum of diameters of target lesions, taking the baseline sum diameter as reference.
- PD: At least a 20% increase in the sum of diameters of target lesions, taking the smallest sum on study (this includes the baseline sum if that is the smallest on the study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. (Note: the appearance of one or more new lesions is also considered progression).
- SD: Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the minor sum diameters while on a study (Eisenhauer *et al.*, 2009).

Adverse events (AEs) were recorded and classified according to the standard toxicity criteria (CTC) version 3.0. Toxicity was assessed for patients receiving three or more cycles of chemotherapy as well as for all patients who receive at least one chemotherapy cycle. PGIMER's standard protocols were used to handle AEs. Patients with febrile neutropenia or grade 3 or higher gastrointestinal AEs were hospitalized because outpatient management was either not possible or was ineffective. Furthermore, any other AEs caused by chemotherapy were recorded.

4.2 DNA Extraction

4.2.1 Genomic DNA Extraction

Genomic DNA was isolated from peripheral blood samples of cases and controls according to the protocol of Bartlett and White (2003) with slight modifications.

- Firstly, 3ml of whole blood was treated with an equal volume of washing buffer containing 320 mM sucrose, 1.5% Triton X-100, 20 mM Tris-HCl (pH 8.0), and five mM $MgCl_2$ separating red blood cells from the blood lymphocytes. This step was repeated thrice. The sample was centrifuged at 3,500 rpm for 10 min at 4˚C each time.
- Further, the RBC-free pellet was lysed at 45˚C overnight using an equal volume of WBC lysis buffer having 400 mM Tris-HCl (pH 8.0), 150 mM NaCl, 60 mM EDTA, 1% SDS, and 100 µg/ml proteinase-K.
- Removal of proteins was done by adding an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) solution. The mixture, after gentle mixing, was centrifuged at 8000 rpm

for 7 min at 4˚C, two transparent layers were formed. The upper aqueous layer was separated into the new vial, and this step was repeated.

- An equal volume of chloroform/isoamyl alcohol (24:1) solution was added to the aqueous layer. The mixture was centrifuged at 6500 rpm for 5 min at 4˚C to separate the aqueous layer into another vial.
- DNA was precipitated finally using an equal volume of chilled isopropanol or 2.5 times volume of ethanol which was obtained after centrifugation at 10,000 rpm for 5 min after freezing it at -20˚C for 1-2 h.
- The pellet obtained was washed with 1 ml of ice-cold 70% ethanol by centrifugation at 10,000 rpm for 5 min. The pellet obtained was then air-dried.
- DNA pellet was then suspended in 200 µl of sterile Tris-EDTA buffer.

4.2.2 Qualitative and quantitative estimation of genomic DNA

The quality of DNA isolated was assessed using agarose gel electrophoresis. 0.8% agarose gel with 0.5 µg/ml ethidium bromide immersed in 0.5X TBE was used to see the isolated DNA samples. Before loading the wells, the sample was combined with 6X loading dye and diluted with sterile water. The gel was examined under a UV-transilluminator to determine the presence or absence of DNA (Johansson, 1972).

The quantity of DNA was measured by spectrophotometric analysis using Nanodrop (NanoDrop ND-1000 spectrophotometer). At 260 nm, a single unit of absorbance (1OD) of double-stranded DNA is comparable to 50 ng/l of DNA. The amount of DNA in each sample will now be calculated by estimating its absorbance at 260 nm. The sample purity was also determined by calculating the absorbance ratio at 260 nm to absorbance at 280 nm. In the case of a pure DNA sample, the ratio should be 1.8. A ratio less than 1.8 suggests the presence of protein contaminants in the DNA sample, whereas a ratio more than 1.8 indicates RNA or phenol impurities (Desjardins and Conklin, 2010). After a complete examination of concentration and purity, DNA was diluted to 100 ng/l and stored at -20˚C for further use as a template.

4.3 Genotyping of polymorphic sites *NQO1, SULT1A1, EPHX1, NAT2, MTHFR,* and *GSTP1* gene

4.3.1 Genotyping of *NQO1* polymorphism by PCR-RFLP

Genotyping analysis was performed using the PCR-RFLP technique using the primers, annealing temperature, and restriction enzymes as previously reported (Malec *et al*., 2010; Mandal *et al*., 2012) which are tabulated in table 4.1. A standard PCR was used for the amplification of the polymorphic site. The amplification was carried out in a 25 µl mixture which consisted of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.5 µM of forward and reverse primer, 200 µM dNTPs, 100 g/ml bovine serum albumin (BSA), and 1U Taq polymerase (DNAzyme, Thermo Scientific) and 200 ng DNA. PCR profile consisted of melting step of 94°C for 5 min and followed by 30 cycles of 94°C for 45s,annealing step at an annealing temperature of the primer for the 30s, primer extension at 72°C for 30 s and final extension step at 72°C for 5 mins.

Restriction fragment length polymorphism

After the successful amplification, the PCR product was treated with suitable restriction enzymes obtained from New England Biolabs, Ipswich, MA, USA. 10 µl of each PCR product was treated with 5U of enzyme in the presence of 1X respective NEB buffer, making a reaction volume of 20 µl. This reaction was incubated at the desired temperature, which was 37˚C, and incubated for 12-16 hours. The amplified products were resolved using polyacrylamide gel electrophoresis based on size. Two different people analyzed the results to remove biasedness. A further 20% of the samples were chosen randomly, and the results' reproducibility was 100%.

Polyacrylamide gel electrophoresis

- Polyacrylamide gels are used to separate nucleic acids and proteins in solid support under an electric field in a buffer. PAGE gel is a cross-linked polymer of acrylamide and bis-acrylamide. Pore size is determined by the total concentration of monomeric units (acrylamide and bis-acrylamide) and the ratio of acrylamide: bis acrylamide. Polymerization is initiated by ammonium persulphate (APS) and catalyzed by TEMED (N, N, N, N-Tetramethylethylenediamine). Polyacrylamide gels have a relatively small range of separation but very high resolving power. In the case of DNA, polyacrylamide is used for separating fragments of less than about 500 bp. It is used to resolve fragments of

DNA differing in length by a single base pair also. The gels were visualized using ethidium bromide staining. This protocol used was given by Sambrook *et al.* (2001) with slight modifications.

- A free radical-generating system initiates vinyl addition polymerization.
- TEMED accelerates the rate of formation of free radicals from persulfate, and these, in turn, catalyze polymerization.
- The persulfate free radicals convert acrylamide monomers to free radicals which react with inactivated monomers to begin the polymerization chain reaction.
- The elongating polymer chains are randomly cross-linked by bis-acrylamide, resulting in a gel.
- The chemicals required include 30% Acrylamide solution [(29:1) acrylamide: bis-acrylamide], 5X TBE (Tris borate EDTA), 10% APS, TEMED and 6X loading buffer.

Ethidium bromide staining

The Native-PAGE gel was stained in ethidium bromide (EtBr) solution, which was prepared by adding 10 µl of EtBr (concentration 0.5 µg/ml) in 50 ml of sterile water. The gel was kept in EtBr staining solution for 20-25 min in the dark. After staining, the gel was visualized under a UV-transilluminator to determine the bands after digestion.

Table 4.1: List of primers and restriction enzymes used for *NQO1* genetic variant

Codon for protein	SNP	SNP ID #	Primers	PCR product size (bp)	Restriction enzyme	Fragments identifying genotypes (bp)
Pro^{187}Ser	*609C>T*	rs1800566	5'-TCC TCA GAG TGG CAT TCT GC-3' – forward primer 5'-TCT CCT CAT CCT GTA CCT CT-3' – reverse primer	230	*Hinf I*	CC - 195, 35 TT - 151, 44, 35 CT - 195, 151, 44, 35

4.3.2 Genotyping of *SULT1A1* polymorphism by PCR-RFLP

The PCR-RFLP-based assay was used to determine the genotype of *SULT1A1* as previously described (Peng *et al*., 2003; Arslan, 2010) with slight modifications. The forward and reverse primers used are tabulated in table 4.2. The PCR mixture of 25 µl consists of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.5 µM of forward and reverse primer, 200 µM dNTPs, 100 mg/ml bovine serum albumin (BSA) and 1U Taq polymerase (DNAzyme, Thermo Scientific), and 200 ng DNA. PCR

conditions used for the mixture were: 95 ˚C for 5 min and 94 ˚C for the 30s (denaturation), 62 ˚C for 45s (annealing) followed by 72 ˚C for 30s (extension) for 29 cycles and the final extension for 5 min at 72 ˚C. After amplification, the PCR product of 281 bp was checked on 1.5% agarose gel and then digested with 5U of *Hha I* restriction enzyme (Thermo Fisher Scientific Inc.) at 37 ˚C.

The digested samples were run on a 2.0% agarose gel that was dissolved in 0.5X TBE with 0.5µg/ml ethidium bromide. The electrophoresis is conducted in 0.5X TBE at 50 volts/cm. The digested samples are mixed with 2 µl of 6X loading buffer containing sucrose, bromophenol blue, andxylene cyanol. The resolved gel is visualized under the UV trans-illuminator, and a restriction pattern is observed. Further, the results were documented using Gel Doc™XR+ Gel Documentation System by Bio-rad. To avoid biases in results scoring of gel was done by two independent observers. The gel scoring was performed according to a defined digestion pattern for each SNP by the respective restriction enzyme based on the generation or removal of the restriction site in the presence or absence of the variant allele. The genotyping analysis was carried out for both cases and control. Further, 20% of the samples were chosen randomly, and the reproducibility was 100%.

Table 4.2: List of primers and restriction enzymes used for *SULT1A1* genetic variant

Codon for protein	SNP	SNP ID #	Primers	PCR product size (bp)	Restriction enzyme	Fragments identifying genotypes (bp)
Arg^{213}His	*638G>A*	rs9282861	5'-GGT TGA GGA GTT GGC TCT GC-3' – forward primer 5'-ATG AAC TCC TGG GGG ACG GT-3' – reverse primer	281	*Hha I*	GG – 281 AA - 175, 106 GA - 281, 175, 106

4.3.3 Genotyping of *EPHX1* polymorphisms by PCR-RFLP

The polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) based assay was used to determine the genotype of *EPHX1* as previously described (Zhou *et al.*, 2001; Ghattas and Amer, 2012) with slight modifications. For amplification of exon three variant and exon 4 using the primer pairs used were are tabulated in table 4.3. The PCR mixture of 25 µl consists of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.5 µM of forward and reverse primer, 200 µM dNTPs, 100 µg/ml bovine serum albumin (BSA) and 1U Taq polymerase (DNAzyme, Thermo Scientific) and 200 ng DNA. PCR conditions used for the mixture were: 95 ˚C for 5 min and 94 ˚C for the 30s (denaturation), 52 ˚C (exon 3) and 59 ˚C (exon 4) for 45s (annealing) followed by

72 ˚C for 30s (extension) for 29 cycles and the final extension for 5 min at 72 ˚C. The PCR product of 162 bp for exon three variant and 357 bp for exon four after amplification was checked on 2.0% agarose gel and then digested with 5U of *EcoRV* and *RsaI* restriction enzyme (New England Biolabs, Ipswich MA), respectively at 37 ˚C. The digested samples were run on 8% Native-PAGE and were stained in ethidium bromide and visualized under UV trans-illuminator (procedure as mentioned in section 4.3.1). Two different people analyzed the results to remove biasedness. A further 20% of the samples were chosen randomly, and the results' reproducibility was 100%.

Table 4.3: List of primers and restriction enzymes used for *EPHX1* genetic variants

Codon for protein	SNP	SNP ID #	Primers	PCR product size (bp)	Restriction enzyme	Fragments identifying genotypes (bp)
Tyr^{113}His (exon 3)	*337T>C*	rs1051740	5'GATCGATAAGTTCCGTTTCACC-3' – forward primer 5'-ATCCTTAGTCTTGAAGTGAGGA T-3' – reverse primer	162	*EcoRV*	TT - 140, 22 CC - 162 TC - 162, 140, 22
His^{139}Arg (exon 4)	*415A>G*	rs2234922	5'-GGGGTACCAGAGCCTGACCGT-3' – forward primer 5'-AACACCGGGCCCACCCTTGGC-3' – reverse primer	357	*RsaI*	AA - 295, 62 GG - 174, 121, 62 AG - 295, 174, 121, 62

4.3.4 Genotyping of *NAT2* polymorphisms by PCR-RFLP

The polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) based assay was used to determine the genotype of *NAT2* as previously described (Smith *et al*., 1997; Lotfi *et al*., 2018) with slight modifications. For amplification of *NAT2* variants, the primer pairs used are tabulated in table 4.4. The PCR mixture of 25 µl consists of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.5 µM of forward and reverse primer, 200 µM dNTPs, 100 µg/ml bovine serum albumin (BSA) and 1U Taq polymerase (DNAzyme, Thermo Scientific) and 200 ng DNA. PCR conditions used for the mixture were: 95 ˚C for 5 min and 94 ˚C for the 30s (denaturation), 63 ˚C for 45s (annealing) followed by 72 ˚C for 30s (extension) for 29 cycles and the final extension for 5 min at 72 ˚C. AFTER AMPLIFICATION, the PCR product of 547 bp for the NAT2 variant was checked on 1.5% agarose gel. The mutations at positions 481, 590, 803, and 857 were distinguished by the electrophoretic or Native-PAGE banding pattern produced after digestion of the PCR product with 5U of *KpnI*, *TaqI*, *DdeI*, and *BamHI* restriction enzyme (New England

Biolabs, Ipswich MA), respectively at 37 ˚C and for *TaqI* at 65 ˚C (Table 4.4). Two different people analyzed the results to remove biasedness, and further 20% of the samples were chosen randomly, and the results' reproducibility was 100%.

Table 4.4: List of primers and restriction enzymes used for *NAT2* genetic variants

Codon for protein	SNP	SNP ID #	Primers	PCR product size (bp)	Restriction enzyme	Fragments identifying genotypes (bp)
Leu^{161}*Leu*	*481C>T*	rs1799929	5'-GCTGGGTCTGGAAGCTCC TC-3' – forward primer 5'-TTGGGTGATACATACACAAGGG-3' – reverse primer	547	*KpnI*	CC – 547 TT - 433, 114 CT - 547, 433, 114
Arg^{197}*Gln*	*590G>A*	rs1799930		547	*TaqI*	GG - 392, 155 AA - 222, 170, 155 GA -392, 222, 170, 155
Lys^{268}*Arg*	*803A>G*	rs1208		547	*DdeI*	AA - 345, 114, 65 GG - 345, 137, 65 AG - 345, 137, 114, 65
Gly^{286}*Gln*	*857G>A*	rs1799931		547	*BamHI*	GG – 547 AA - 490, 57 GA - 547, 490, 57

4.3.5 Genotyping of *MTHFR* polymorphisms by PCR-RFLP

The PCR-RFLP assay was used to determine the MTHFR genotype as previously reported (Le Marchand *et al*., 2002; Siemianowicz *et al*., 2003) with slight modifications. For amplification of the *C677T* variant and *A1298C,*the primer pairs used are mentioned in table 4.5.The PCR mixture of 25 µl consists of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.5 µM of forward and reverse primer, 200 µM dNTPs, 100 µg/ml bovine serum albumin (BSA) and 1U Taq polymerase (DNAzyme, Thermo Scientific) and 200 ng DNA. PCR conditions used for the mixture were: 95 ˚C for 5 min and 94 ˚C for 30s (denaturation), 61 ˚C (*C677T*) and 55 ˚C (*A1298C*) for 45s (annealing) followed by 72 ˚C for 30s (extension) for 29 cycles and the final extension for 5 min at 72 ˚C. The PCR product of 198 bp for *C677T* variant and 163 bp for *A1298C* variant after amplification was checked on 2.0% agarose gel and then digested with 5U of *HinfI* and *MboII* restriction enzyme (Takara Bio Inc.), respectively at 37 ˚C. The digested samples were run on 6% Native-PAGE and were stained in ethidium bromide and visualized under UV trans-illuminator. Two different people analyzed the results in order to remove biasedness, and further 20% of the samples were chosen randomly, and the results' reproducibility was 100%.

Table 4.5: List of primers and restriction enzymes used for *MTHFR* genetic variants

Codon for protein	SNP	SNP ID #	Primers	PCR product size (bp)	Restriction enzyme	Fragments identifying genotypes (bp)
Ala^{222}Val	*677C>T*	rs1801133	5'-TGAAGGAGAAGGTGTCT GCGGGA-3'– forward primer 5'-AGGACGGTGCGGTGAGAGTG-3' – reverse primer	198	*Hinf I*	CC - 198 TT - 175, 23 CT - 198, 175, 23
Glu^{429}Ala	*1298A>C*	rs1801131	5'-CTTTGGGGAGCTGAAGG ACTACTAC-3'- forward primer 5'-CACTTTGTGACCATTCCGGTTT G-3' – reverse primer	163	*Mbo II*	AA - 56,31,30,28,18 CC - 84,31,30,18 AC - 84,56,31,30,28,18

4.3.6 Genotyping of *GSTP1* polymorphism by PCR-RFLP

The PCR-RFLP assay was used to determine the genotype of *GSTP1 IIe^{105}Val*, as previously reported by Ibarrola-Villava *et al.* (2012), with slight modifications. For amplification of *GSTP1* variants, the primer pairs used are tabulated in table 4.6. The 25 µl PCR mixture consists of 1X PCR buffer with 1.5 mM $MgCl_2$, 0.5 µM each forward and reverse primer, 200 µM dNTPs, 100 µg/ml of bovine serum albumin (BSA) and 1U Taq polymerase (DNAzyme, Thermo Scientific) and 200 ng DNA. The following PCR conditions were used: 95˚C for 5 min and 94˚C for 30s (denaturation), 59˚C for 45s (annealing) followed by 72˚C for 30s (extension) for 29 cycles and the final extension for 5 min at 72˚C. The PCR product of 568bp was checked on 1.5% agarose gel. And then PCR product was digested *BsmA1* restriction enzyme (New England Biolabs, Ipswich MA) at 55˚C.The digested samples were run on 6% Native-PAGE stained in ethidium bromide and visualized under UV trans-illuminator. Without prior knowledge genotyping of a 15% random sample of cases was carried out twice by different people to ensure quality control and reproducibility was100%.

Table 4.6: List of primers and restriction enzymes used for *GSTP1* genetic variant

Codon for protein	SNP	SNP ID #	Primers	PCR product size (bp)	Restriction enzyme	Fragments identifying genotypes (bp)
Ile^{105}Val	*313A>G*	rs1695	5'-AATACCATCCTGCGTCACCT-3'- forward primer 5'-TGAGGGCACAAGAAGCCCCTT-3' – reverse primer	568	*Bsm*AI	AA - 305, 135, 128 bp GG - 222, 135, 128, 83 bp AG - 305, 222, 135, 128, 83 bp

4.4 Statistical analysis

Demographic analysis

Demographic variables such as age, gender, smoking status, and pack-years were classified into categorical and continuous variables. Chi-square test was used to compare the cases and controls for the selected demographic characteristic such as gender and smoking status, which are the categorical variables. Whereas, t-test was used to compare groups about continuous variables such as age and smoking.

Genotypic and allelic Frequency analysis

The Hardy-Weinberg equilibrium theory (p2+2pq+q2=1; where p is the frequency of the wild-type allele and q is the frequency of the variant allele) was used to compute the genotype frequencies of all SNPs using the χ^2 test in both cases and controls. The Pearson's χ^2 test was used to see a statistically significant difference in allelic and genotypic frequencies between cases and controls. For each polymorphic location, the minor allele frequency was also estimated.

Association analysis

To examine the risk of lung cancer and the polymorphisms, odds ratios (ORs) and 95% confidence intervals (CIs) were determined. Using logistic regression, adjusted odds ratios (AORs) were produced for subsequent investigation of potential risk factors such as age, smoking, and gender. Except for the continuous variables such as age, all of the variables utilized in the study were dichotomous. The occurrence of lung cancer was chosen as the dependent variable in the logistic regression analysis, and genotype was chosen as the independent variable that can predict susceptibility to lung cancer. Aside from the comprehensive association study, a stratified analysis was carried out to estimate risk for subgroups based on smoking status and tumor histology. The above statistical analyses were performed with Medcalc version14.8.1 (Medcalc Software, Ostend, Belgium).

Haplotype analysis

SHEsis software (http://202.120.31.177/myAnalysis.php) was used to estimate haplotype frequency and linkage disequilibrium between groups (Yong and Lin, 2005). To achieve D and r^2

values, polymorphic sites that share a common chromosomal locus and belong to the same gene were assessed for LD. Only haplotype blocks with a case and control frequency greater than 0.03 were assessed; the remaining haplotypes were not considered for their association with lung cancer risk. The global chi-square test was used to calculate the frequency difference. The odds ratio for each haplotype was used to calculate the connection.

The false discovery rate analysis

Multiple comparison corrections for genotypic combinations (SNP- SNP) were performed using FDR (False discovery rate) method by online software based on Benjamini and Hochberg methodology (http://sdmproject.com/ utilities/?show=FDR). The adjusted p values are mentioned, and FDR adjusted $p<0.01$ was deemed significant (Benjamini and Hochberg, 1995).

MDR

To discover interaction models, we used the multifactor dimensionality reduction (MDR) approach. MDR is a non-parametric genetic model method for detecting and characterizing SNP interactions that overcome some of the limitations of logistic regression. MDR uses cross-validation consistency (CVC) and permutation testing to determine disease status. CVC necessitates the separation of data into training and testing sets. The study is repeated ten times in order to get the interaction models. Using MDR, multi-locus genotypes are grouped into two groups, high and low risk, lowering the number of genotype predictors from n dimensions to one. Only the genotypic combinations with the highest cross-validation consistency (CVC) and the lowest prediction error were chosen as the optimal interaction model from the several genotype models created. MDR tests were carried out using version 0.5.1 of the open-source MDR software tool Epistasis (http://www.epistasis.org) (Ritchie and Motsinger, 2005).

CART analysis

Finally, we used CART software to perform a classification and regression tree (CART) analysis to detect and characterize the high-order interactions (6.0, Salford Systems, San Diego, California). CART is a tree-based model that uses the binary recursive partitioning approach to generate a decision tree to discover sub-groups at higher risk that are less obvious when using logistic regression methods. The analysis is carried out so that the most significant predictor is

used to divide the sample into subgroups, and the process is repeated until the differences are no longer significant. Finally, the findings are obtained in the form of classification or decision trees with nodes and sub-nodes. The risk of all genotypes sets was estimated by considering the node with a low case rate was as the reference to calculate the ORs and 95% CIs (Gupta *et al.*, 2016).

Overall survival and Hazard rate analysis

The Kaplan Meier and Log-rank tests were used to assess the relationship between overall survival and genetic polymorphism. The log-rank test was used to compare survival curves. After adjusting for other covariates, the Cox proportional hazard model was used to compute the hazard rate and examine the influence of genetic polymorphism on overall survival. Cox proportional hazards models were used to calculate the hazard ratios (HRs) and 95% confidence intervals (CIs). The hazard is the response (dependent) variable in this model. The hazard is the probability of death that patients have survived upto that point in time, or the risk of death then. The relationship between response and genetic polymorphisms was evaluated using a logistic regression model adjusted for age, gender, smoking, stage, histology, performance status, and chemotherapy regimen. All tests were two-sided, and the threshold for statistical significance was set at $p<0.05$. MedCalc version 14.8.1 was used for all statistical analyses (Medcalc Software, Ostend, Belgium).

CHAPTER 5

RESULTS

5.1 Genomic DNA isolation

The isolated genomic DNA was estimated both qualitatively and quantitatively. A 0.8% agarose gel was used to perform a qualitative estimation of genomic DNA. The bands obtained were checked for integrity. Figure 5.1 represents agarose gels for genomic DNA isolated from blood.

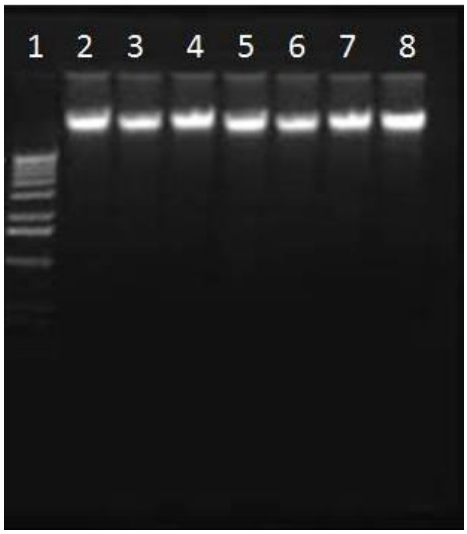


Figure 5.1: 0.8% agarose gel representing genomic DNA isolated from blood; Lane 1: 1kb ladder, Lane 2-5: genomic DNA from the blood of cases, Lane 6-8: genomic DNA isolated from healthy controls.

5.2 PCR amplification

PCR carried out the amplification of the gene. Amplification of each SNP after PCR was confirmed using 1.7% agarose gel electrophoresis to assure successful amplification of the appropriate length, which was validated using marker DNA of either 100 bp or 50 bp size. Figure 5.2 shows a representative example of agarose gel electrophoresis illustrating the amplification of genomic DNA for *NQO1 Pro*^{187}*Ser*. An amplified product of 230 bp was obtained as indicated by the 100 bp ladder run along.

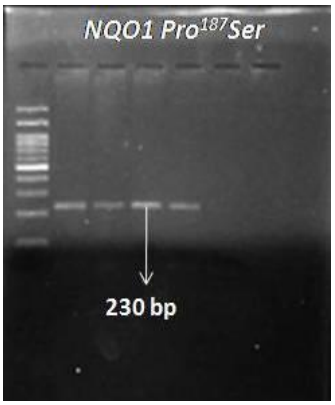


Figure 5.2: Representative example of agarose gel electrophoresis showing PCR amplification for *NQO1 Pro*^{187}*Ser*

5.3 Genotyping of polymorphic variants by RFLP

5.3.1 Genotyping of *NQO1* genetic variant

The polymorphic variant of *NQO1* studied is *Pro*^{187}*Ser* (C609T, rs1800566). The polymorphism is located in exon 6 of the gene. The restriction pattern of a polymorphic variant of *NQO1 Pro*^{187}*Ser* was confirmed using 8.0% Native-PAGE and stained in ethidium bromide and visualized under UV trans-illuminator. PCR product of 230 bp after amplification was digested with 5U of *Hinf I* restriction enzyme (New England Biolabs, Ipswich MA) at 37 ˚C. The wild-type homozygous genotype (*CC* or *Pro/Pro*) has one site for *HinfI* and gave two fragments (195, 35 bp). In contrast, the variant genotype (*TT* or *Ser/Ser*) has an additional site for *HinfI* and gave three fragments (151, 44, 35 bp). The heterozygous variant (*CT* or *Pro/Ser*) gave four fragments (195, 151, 44, 35 bp), respectively, as shown in Figure 5.3.

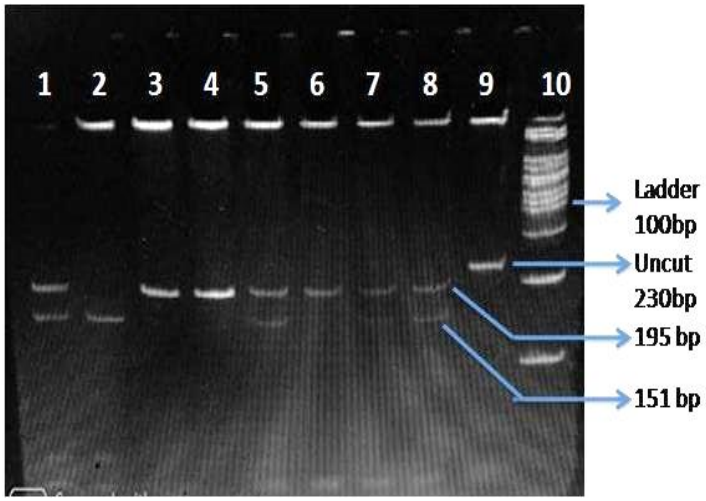


Figure 5.3: Restriction digestion pattern of *NQO1 Pro*^{187}*Ser* (Lane 1,5,8: Heterozygous genotype (Pro/Ser); Lane 2: mutant (*Ser/Ser*); Lane 3,4,6,7: wild genotype (*Pro/Pro*); Lane 9: uncut; Lane 10: marker (100 bp), and 35 bp band is visible at the lower most of the gel)

5.3.2 Genotyping of *SULT1A1* genetic variant

The polymorphic variant of *SULT1A1* studied is *Arg*^{213}*His* (rs9282861). The polymorphism is located at exon 7 of the gene. After amplification, the PCR product of 281 bp was digested with 5U of *Hha I* restriction enzyme (Thermo Fisher Scientific Inc.) at 37 ˚C. The digested samples were run on 2.0% agarose gel and visualized under a UV trans-illuminator. The wild-type homozygous genotype (*GG* or *Arg/Arg*) has one site for *HhaI* and gave two fragments (175, 106 bp). In contrast, the variant genotype (*AA* or *His/His*) has no site for *HhaI* and gave one fragment (281 bp), and the heterozygous variant (*GA* or *Arg/His*) gave three fragments (281, 175, 106 bp), respectively as shown in Figure 5.4.

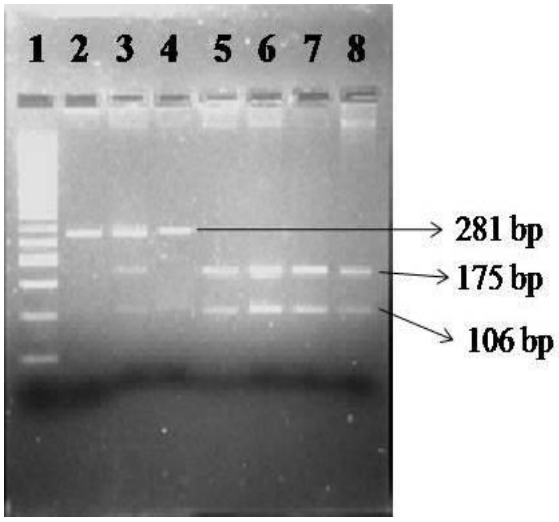


Figure 5.4: Restriction digestion pattern of *SULT1A1Arg*^{213}*His* (Lane 1: Ladder (50 bp); Lane 2: uncut PCR product; Lane 3: Heterozygous genotype (*Arg/His*); Lane 3: Homozygous variant genotype (*His/His*); Lane 5-8: Homozygous wild genotype (*Arg/Arg*))

5.3.3 Genotyping of *EPHX1* genetic variant

In the *EPHX1* coding region, two specific polymorphisms are characterized in both exon 3 and 4. In exon 3, *Tyr*^{113}*His* (rs2854455) results in a T to C substitution and, in exon4, *His*^{139}*Arg* (rs2234922) results in A to G substitution. The PCR product of 162 bp for exon three variant and 357 bp for exon four after amplification digested with 5U of *EcoRV* and *RsaI* restriction enzyme (New England Biolabs, Ipswich MA), respectively at 37 ˚C. The digested samples were run on 8% Native-PAGE stained in ethidium bromide and visualized under UV trans-illuminator. For exon 3 (*Tyr*^{113}*His*), wild-type produced two bands (140 and 22 bp), variant alleles were identified by a single band (undigested 162 bp), and heterozygous displayed three bands (162, 140, and 22

bp), as shown in Figure 5.5 a). For exon 4 (*His*^{139}*Arg*), wild-type alleles were identified by two bands (295 and 62 bp), variant alleles displayed three bands (174, 121, and 62 bp), and heterozygous allele produced all of the four bands (295, 174, 121, and 62 bp) as shown in Figure 5.5 b).

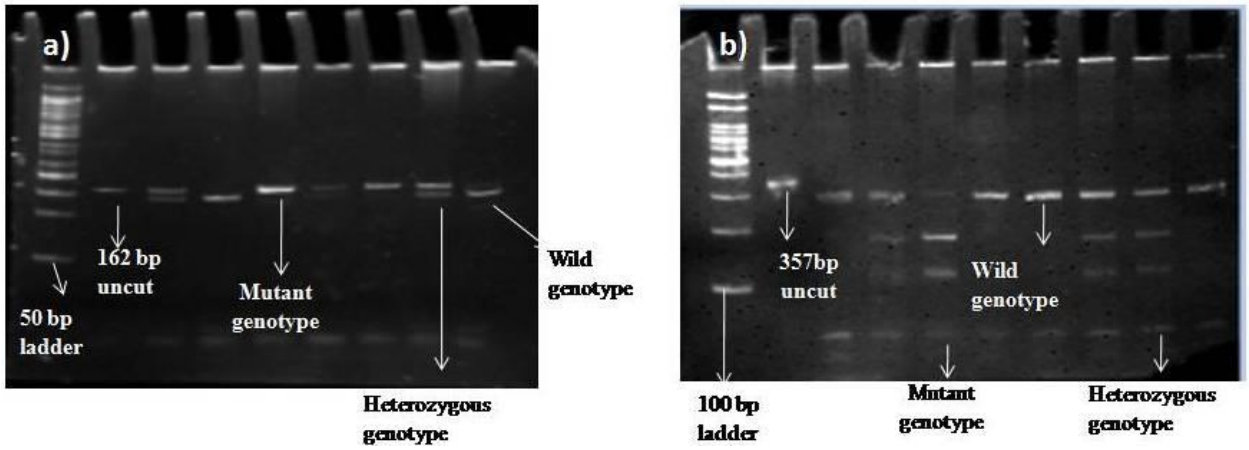


Figure 5.5: Representative example of 8.0% Native-PAGE showing PCR-RFLP pattern of a) *EPHX1 Tyr*^{113}*His* (rs2854455), and b) *EPHX1 His*^{139}*Arg* (rs2234922)

5.3.4 Genotyping of *NAT2* genetic variants

There are three NAT2 phenotypes known: a rapid acetylation phenotype consisting of two copies of wild type allele, intermediate phenotype with a single allele of wild type and a mutant allele, and a slow acetylation phenotype consisting of two mutant alleles. In this study we have examined four variant alleles of *NAT2* (*viz. NAT2*5B (481C>T,* rs1799929), *NAT2*5C (803A>G,* rs1208*), NAT2*6 (590G>A,* rs1799930*),* and *NAT2*7* (*857G>*A, rs1799931)) polymorphisms. The mutations at positions 481, 590, 803, and 857 were distinguished by the electrophoretic or Native-PAGE banding pattern produced after digestion of the PCR product of 547 bp with 5U of *KpnI, TaqI*, *DdeI*, and *BamHI* restriction enzyme (New England Biolabs, Ipswich MA), respectively at 37 ˚C and for *TaqI* at 65 ˚C.

For *481C>T,* the wild type produced two bands (433, 114 bp), variant allele produced a single band of 547 bp, and heterozygous displayed three bands (547, 433, 114 bp) as shown in figure 5.6 a). The digested samples were run on 2.0% agarose gel and were stained in ethidium bromide and visualized under UV trans-illuminator. For *590G>A*, wild type displayed three

bands (222, 170, 155 bp), variant genotype produced two bands (392,155bp), and heterozygous genotype displayed four bands (392, 222, 170, 155 bp) as depicted in figure 5.6 b). For *803A>G*, the wild type produced three bands (345, 137, 65 bp), variant allele also produced three fragments (345, 114, 65 bp), and heterozygous displayed four bands (345, 137, 114, 65 bp) as depicted in figure 5.6 c). For *857G>A*, the wild type produced two bands (490, 57 bp), variant allele produced a single band of 547bp, and heterozygous genotype produced three bands (547, 490, 57 bp) as shown in figure 5.6 d).The digested samples were run on 8% Native-PAGE stained in ethidium bromide and visualized under UV trans-illuminator.

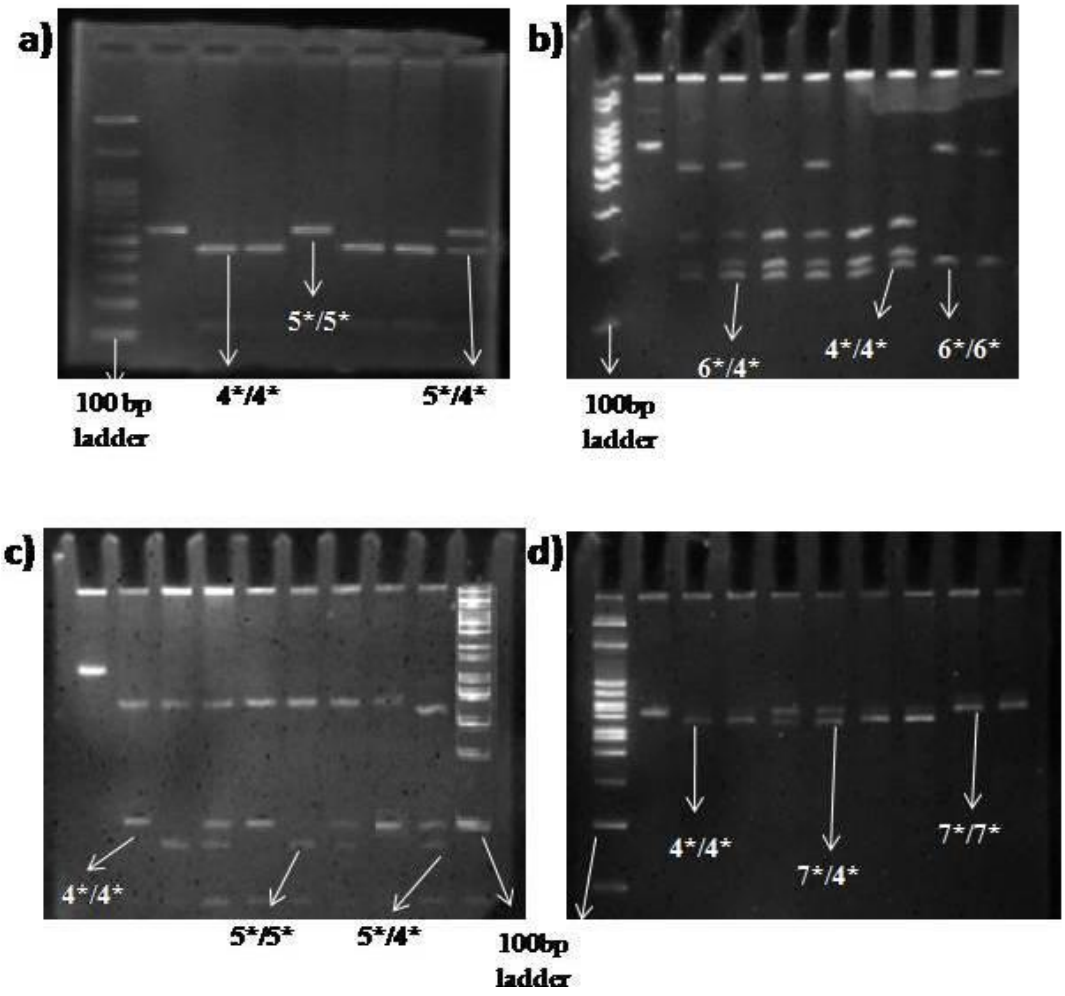


Figure 5.6: a) Representative example of an agarose gel showing PCR-RFLP pattern of *481C>T*, and 8.0% Native-PAGE showing PCR-RFLP pattern of b) *590G>A*, c) *803A>G*, and d) *857G>A*

5.3.5 Genotyping of *GSTP1* genetic variant

A non-synonymous *A/G* single nucleotide polymorphism at nucleotide position 313 in exon 5 results in the replacement of adenine to guanine (SNP's ID: rs1695), which results in the substitution of the amino acid isoleucine to valine at position 105. The PCR product of 568-bp after amplification was digested with *BsmA1* restriction enzyme (New England Biolabs, Ipswich MA) at 55˚C. The PCR product was digested into 222 bp, 135 bp, 128 bp, and 83 bp, representing mutant genotype (*Val/Val*). In contrast, the presence of five bands of 305 bp, 222 bp, 135 bp, 128 bp, and 83 bp was indicative of the heterozygous (*Ile/Val*) genotype, and the presence of three bands of 305 bp, 135 bp, and 128 bp represented the wild genotype (*Ile/Ile*). The difference between the bands of size 135 bp and 128 bp is too tiny to differentiate. Thus they were represented as one band, as shown in figure 5.7. The digested samples were run on 6% Native-PAGE stained in ethidium bromide and visualized under UV trans-illuminator.

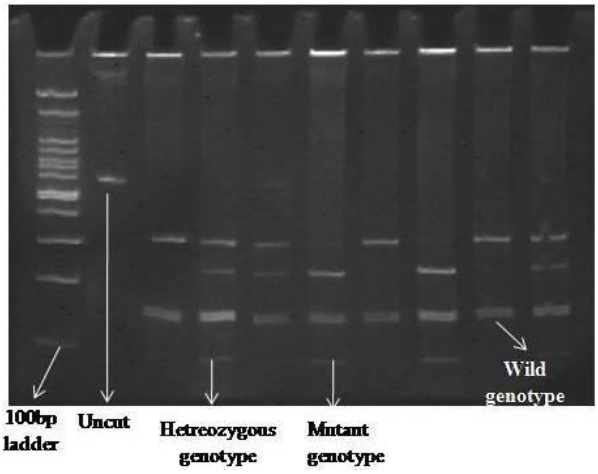


Figure 5.7: Representative example of 6% Native-PAGE exhibiting restriction digestion pattern of *GSTP1 Ile^{105}Val*

5.3.6 Genotyping of *MTHFR* genetic variants

The *MTHFR* gene has 11 exons. Several single nucleotide polymorphisms (SNPs) in the *MTHFR* gene, including *C677T* and *A1298T*. At nucleotide 677, the C-to-T transition results in an amino acid change of valine for alanine. The *MTHFR A1298C* polymorphism (rs1801131) in exon7 is identified due to substituting glutamate with alanine at amino acid 429 (*Glu^{429}Ala*). The PCR product of 198 bp for *C677T* variant and 163 bp for *A1298C* variant after amplification were digested with 5U of *HinfI* and *MboII* restriction enzyme (Takara Bio Inc.), respectively at 37 ˚C. The digested samples were run on 6% Native-PAGE stained in ethidium bromide and visualized

under UV trans-illuminator. For *C677T*, variant-type produced two bands (175 and 23 bp), wild alleles were identified by a single band (undigested 198 bp), and heterozygous displayed three bands (198, 175, and 23 bp) (Figure 5.8 a) b).For *A1298C*, wild-type alleles were identified by five bands (56, 31, 30, 28, and 18 bp), variant alleles displayed four bands (84, 31, 30, and 18 bp), and heterozygous allele produced all of the six bands (84, 56, 31, 30, 28 and 18 bp) (Figure 5.8 b)). The difference between the size 31, 30, 28, and 18 bp bands is significantly less, so they were all represented as one band.

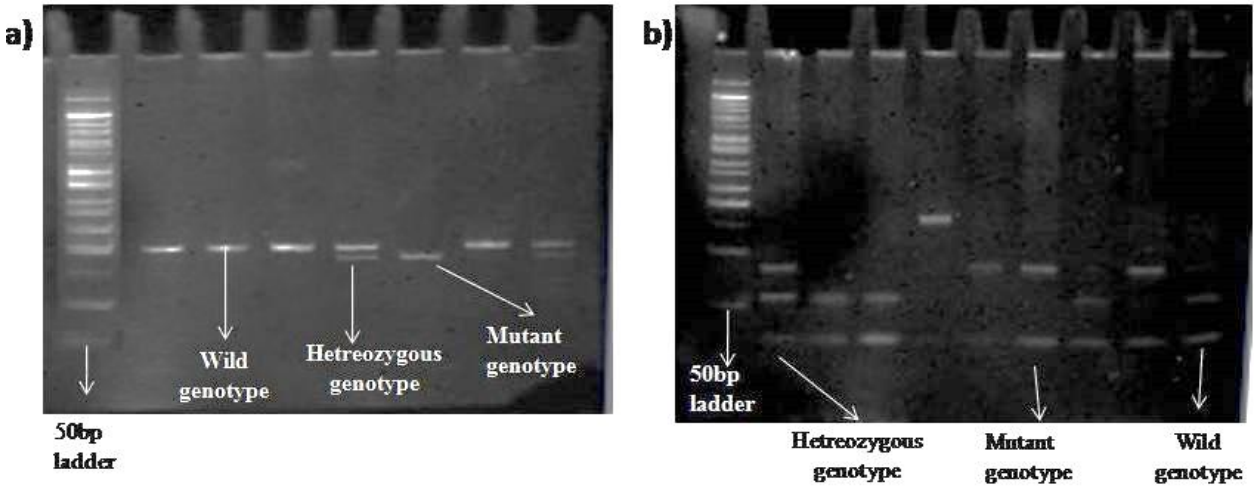


Figure 5.8: Representative example of 6% Native-PAGE exhibiting restriction digestion pattern of a) *MTHFR C677T*, and b) *MTHFR A1298C*

5.4 Epidemiological statistics of North Indian population under study

5.4.1 Regional Distribution of Lung cancer patients

Blood samples were collected from lung cancer patients visiting the PGIMER Department of Pulmonary Medicine. PGIMER, Chandigarh is a multi-specialist government hospital to which patients from various regions of North India are referred. Punjab, Haryana, Himachal Pradesh Uttar Pradesh, Uttarakhand, Jammu and Kashmir, and Chandigarh were prominent regions. This study has been reviewed and approved by the institutional ethics committee of PGIMER. A total of 550 patients were enrolled in the current study, with Himachal Pradesh (142, 25.82%), Punjab (133, 24.18%), and Haryana (114, 20.73%) having the highest proportion of patients, as indicated in figure 5.9. The remaining lung cancer patients recruited in the study were from states and Union territories like Uttar Pradesh (56, 10.18%), Chandigarh (38, 6.91), Uttarakhand

(23, 4.18%), and Jammu and Kashmir (18, 3.27%). The most minor lung cancer patients were from Bihar (3, 0.55%), Delhi (1, 0.18%), and one patient was from Nepal.

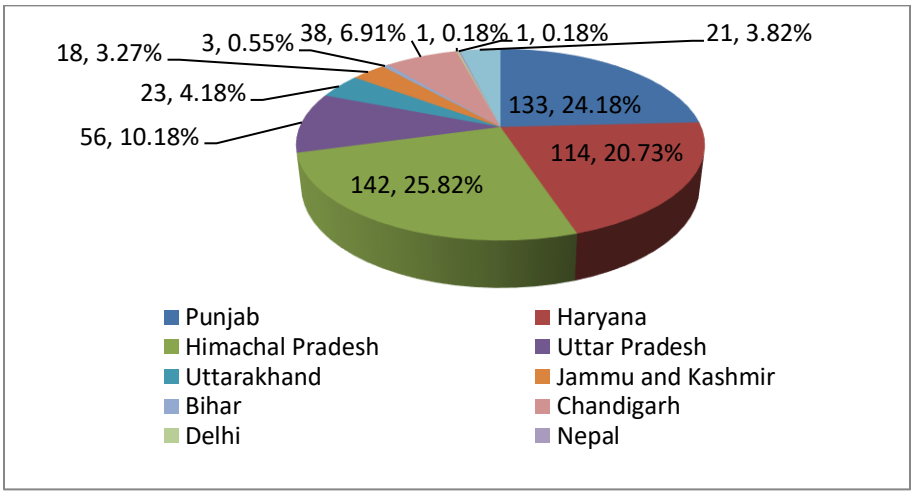


Figure 5.9: The regional distribution of North Indian lung cancer patients

The subjects were also analyzed for their regional distribution based on gender. In cases, the study population comprises 458 (83.27%) males and 92 (16.73%) females. Out of 550 control subjects, 459 (83.45%) were recorded as males and 91 (16.55%) as females. The diagrammatic representation of regional distributional of males and female study population is shown in figure 5.10. The maximum proportion of males and females belonged to Himachal Pradesh, Punjab, Himachal Pradesh, Haryana, respectively.

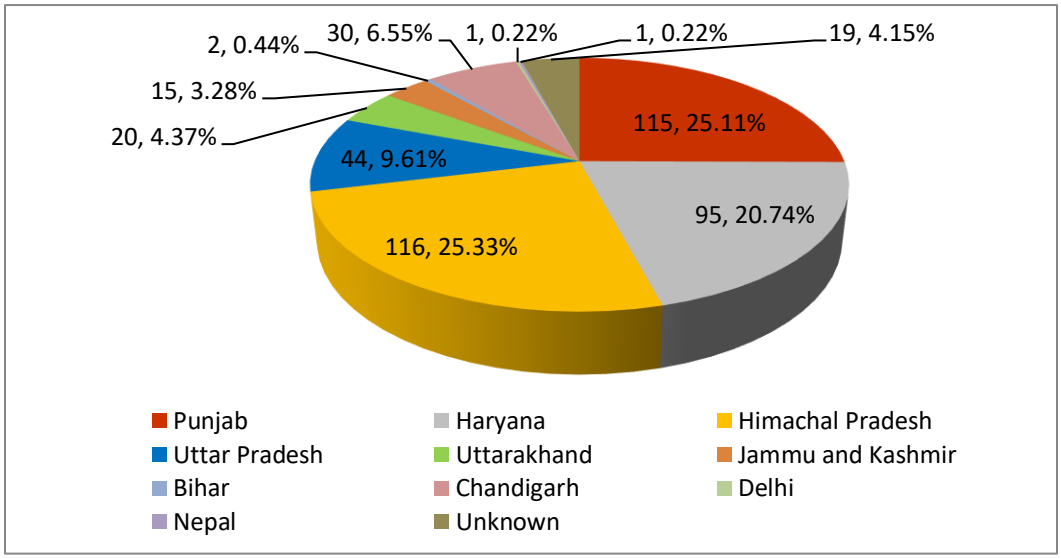


Figure 5.10 a): Pie chart illustrating the regional distribution of North Indian Lung cancer patients of males.

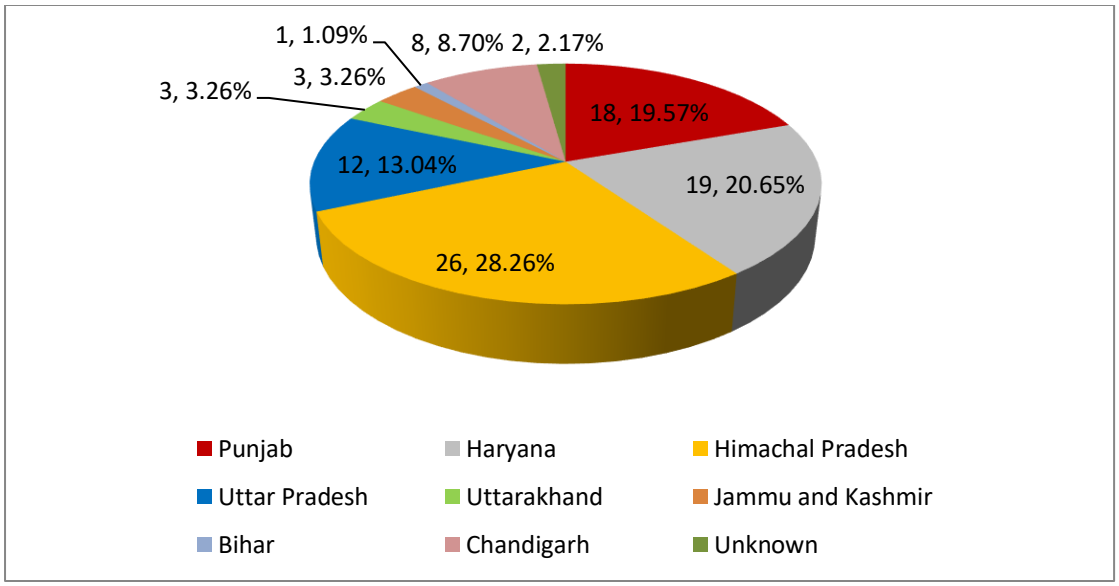
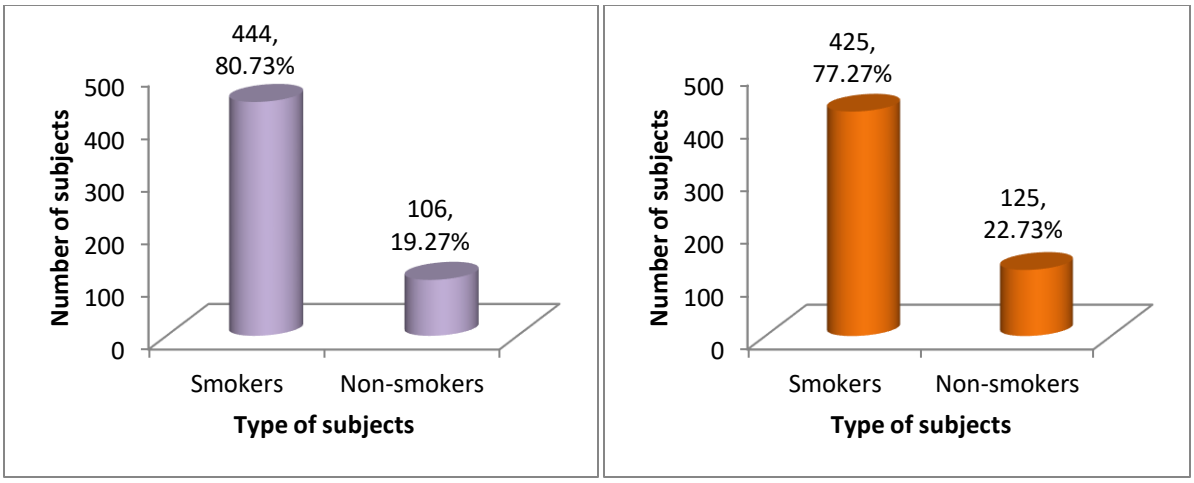


Figure 5.10 b): Pie chart illustrating the regional distribution of North Indian Lung cancer patients of females.

5.4.2 Distribution of study subjects based on smoking

The smoking status of all cases and controls was one of the essential epidemiological data obtained. The study population was bifurcated into smokers and nonsmokers based on their smoking habits. Because smoking is a known cause of lung cancer, the proportion of smokers among cases was higher than that of non-smokers. Similarly, an attempt was made to match the controls in terms of smoking status to eliminate sample bias. The percentage of smokers among cases and controls was 80.73% and 77.27%, respectively, as shown in figure 5.11. The quantitative estimation of smoking was done by calculating the pack-years for each study subject. The mean pack-year was higher in the case group when compared to the control group (31.16 ± 28.28 years *vs*. 20.76 ± 15.95 years). Thus, this indicates that the prevalence of smokers is much higher in cases as compared to controls. As the number of smokers and pack years between two groups was different it was considered as a covariate for analysis.



a) **b)**

Figure 5.11: Graph showing the distribution of study subjects based on smoking status. a): Among cases, b): Among controls

In addition, specific information regarding the type of smoking was recorded. This data reveals that the various types of smoking practiced by study individuals included beedi, cigarette, and a combination of beedi and cigarette. However, in both cases and controls, most research subjects were reported to have been exposed to beedi (63.09% and 73.45%, respectively).The smoking pattern of the study subjects was also analyzed based on their gender. The percentage of smokers among males in cases (93.24%) is lower than those in the control group (99.53%), as shown in figure 5.12.

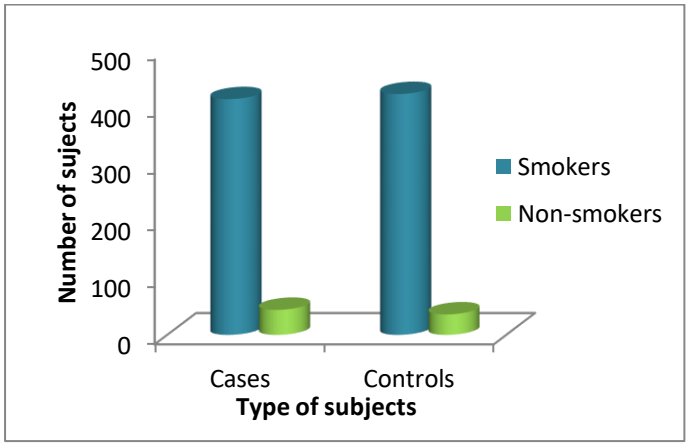


Figure 5.12: Graph showing the distribution of male study subjects based on smoking status

5.4.3 Distribution of Study subjects based on age

Age is another crucial demographic factor that has been linked to the onset of many diseases. The age of all subjects included in the study, both cases and controls, were recorded. The mean

age of the cases was estimated to be 59.23 ± 10.28 years, with a range of 28-86 years. A sincere attempt was made to match the age of cases and controls with (±5 years) difference. The mean age calculated for control subjects was 58.49 ± 11.67 years with a range of 28-84 years. The median age of both cases and controls was found to be 59 years. But all the cases and controls were not matching one on one for the aspect of age and was considered as covariate for further analysis. The diagrammatic representation of the distribution of study subjects based on their gender is shown in figure 5.13.

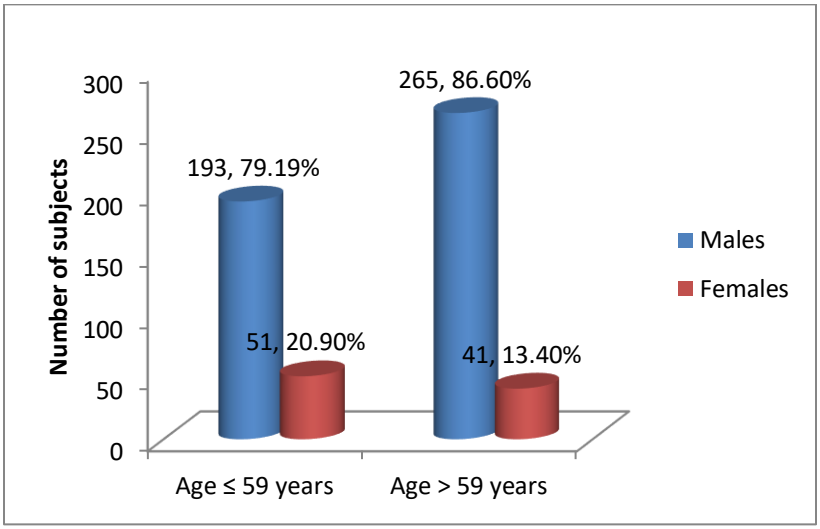


Figure 5.13: Graph showing the age-wise distribution of lung cancer based on gender

A slightly higher percentage of males were reported to have age >59 years than those less than 59 years (86.60% *vs.* 79.19%) among lung cancer patients. Whereas, 20.90% of females were found to have age≤59 years, and 13.40% were reported with age more than 59 years.

5.4.4 Distribution of lung cancer patients based on the stage of tumor and TNM

The clinico- pathological parameters of all the lung cancer subjects included in the study were recorded from the hospital's medical records. The patients were classified based on the clinical stage of the tumor into stage I, stage II, stage III, and stage IV. The proportion of stage I and stage II patients was deficient compared to stage III and stage IV, as shown in figure 5.14. TNM stage data were available in 500 cases (90.91%) out of 550 patients [stage I: 5 (0.91%), II: 19 (3.45%), III: 229 (41.64%), IV: 247 (44.91%)]. Further, the patients were also stratified based on the various parameters of TNM, such as primary tumor extension (T), lymph node involvement (N), and metastasis (M). The detailed distribution of patients based on tumor size is as follows: Tumor size T1 and T2 had very low frequencies of 4.91 and 7.64%, whereas T3 and T4 had

frequencies of 20.91 and 54.36%, respectively. For lymph-node involvement, N0 had 11.82%, whereas N1, N2, N3, and N4 had frequencies of 8.00, 42.36, 27.82, and 0.73%, respectively. Among all the cases for which TNM data were available, 46.18% had no metastasis involvement (M0), whereas 44.73% had distant metastasis (M1).

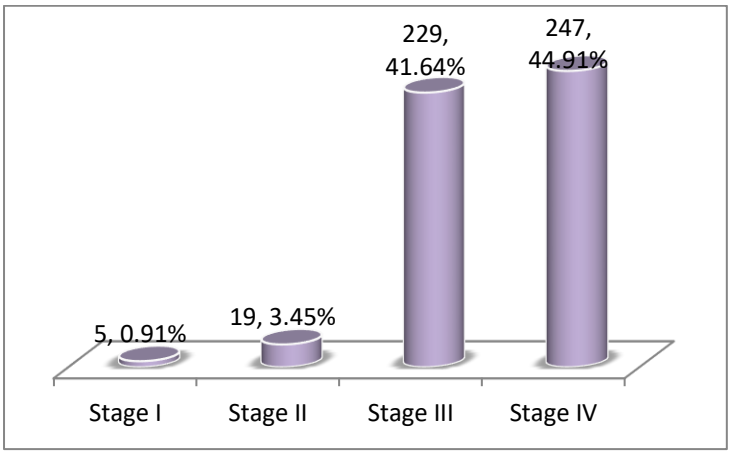


Figure 5.14: Stage based distribution of the study subject

5.4.5 Distribution of lung cancer patients based on histology of the tumor

SQCC, ADCC, and SCLC are the three significant classifications of lung cancer based on histology. The patients in this study were all stratified in the same way. In the present study, the prevalence of SQCC patients (39.27%) was highest among all other histological subtypes, followed by ADCC (36.36%) and SCLC (23.09%), as illustrated in figure 5.15. The remaining patients (1.09%) were diagnosed with miscellaneous histological subtypes such as giant cell carcinoma, mesothelial, and NSCLC undifferentiated, and 0.18% (*i.e.,* one subject) of the subjects was those whose histology was not clear. Therefore, it is evident that SQCC is the most frequently occurring type of lung cancer in North Indians. An evaluation of several smokers in each histological category was evaluated, which revealed that the SQCC subgroup has 94.91% smokers, followed by SCLC having 93.70% smokers (almost similar to SQCC), and the ADCC subgroup has a minuscule percentage of smokers (57.50%). Hence, it is evident that smoking is the most prevalent risk factor in SQCC and SCLC patients.

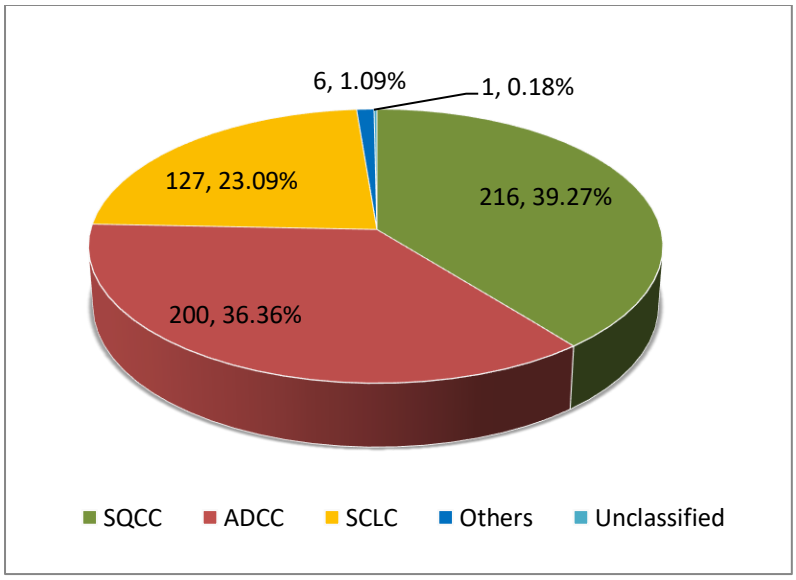


Figure 5.15: Pie chart illustrating the distribution of North Indian Lung cancer patients based on histological subtypes.

5.4.6 Distribution of lung cancer patients based on performance status

Patients are classified using two standard ways to evaluate their performance throughout therapy and estimate their disease's prognosis: Karnofsky performance status (KPS) and Eastern Cooperative Oncology Group-performance status (ECOG-PS). These methods aid in determining the functional status of patients. The distribution of patients in accordance to their KPS showed that 51.82% had a KPS score of (70-80), 28.18% had KPS of (90-100), and the minor percentage(13.82%) of patients had a KPS score of (30-60) indicating poor performance status as illustrated in figure 5.16. Similarly, patients were classified according to ECOG performance status, and it was seen that 10.55% of patients had ECOG status (0), 33.64% had ECOG status (1), 37.54% had ECOG status (2), and 12.36% had ECOG status (3-4) as shown in figure 5.17.

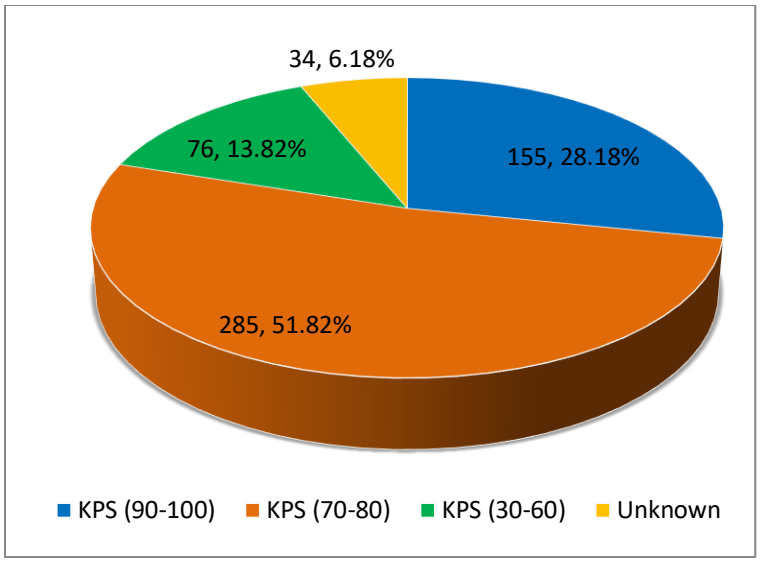


Figure 5.16: Pie chart illustrating the distribution of North Indian Lung cancer patients based on karnofsky performance status (KPS)

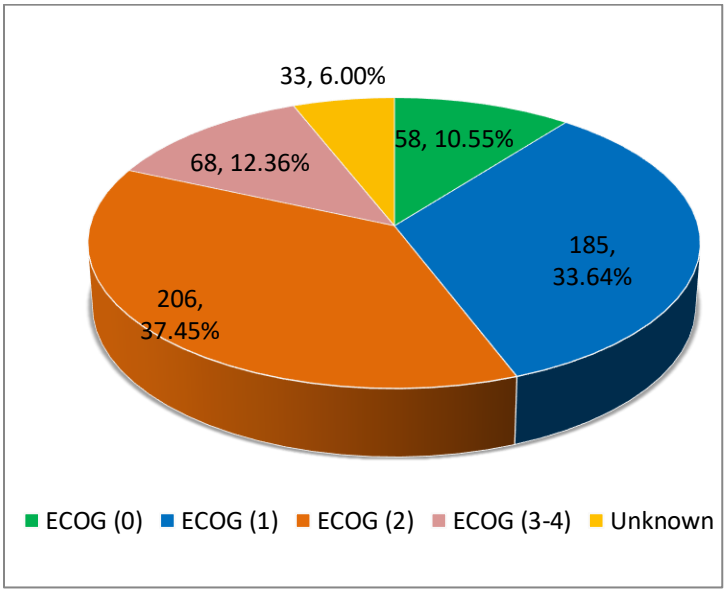


Figure 5.17: Pie chart illustrating North Indian Lung cancer patients' distribution based on Eastern Cooperative Oncology Group -performance status (ECOG-PS).

5.4.7 Distribution of lung cancer patients based on doublet platinum-based chemotherapeutic regimen

The current study treated lung cancer patients with platinum-based doublet chemotherapy and a second-line chemotherapeutic agent based on the patient's histological subtype of lung cancer and other fundamental clinical parameters. Platinum-based drugs, such as cisplatin or

carboplatin, were combined with additional chemotherapeutic agents such as docetaxel, paclitaxel, pemetrexed, gemcitabine, and irinotecan. The chemotherapeutic regimen data was available for 382 subjects. The highest proportions (21.64%) of patients were treated with pemetrexed and cis/carboplatin regimen followed by irinotecan with cis/carboplatin (13.82%) and docetaxel with cis/carboplatin (13.64%). The regimen paclitaxel and cis/carboplatin were administered to 10.73% of the subjects; very few were treated with gemcitabine and cis/carboplatin (3.64%) shown in figure 5.18. Further, 2.18% of the patients were treated with gefitinib along with cis/carboplatin followed by etoposide with cis/carboplatin (2.00%) and 0.18% with ceritinib along with cis/carboplatin.1.64% of the subjects were treated with other chemotherapeutic regimens. Due to the unavailability of records about the chemotherapeutic regimen of some patients, 30.55% had fallen in the unknown category.

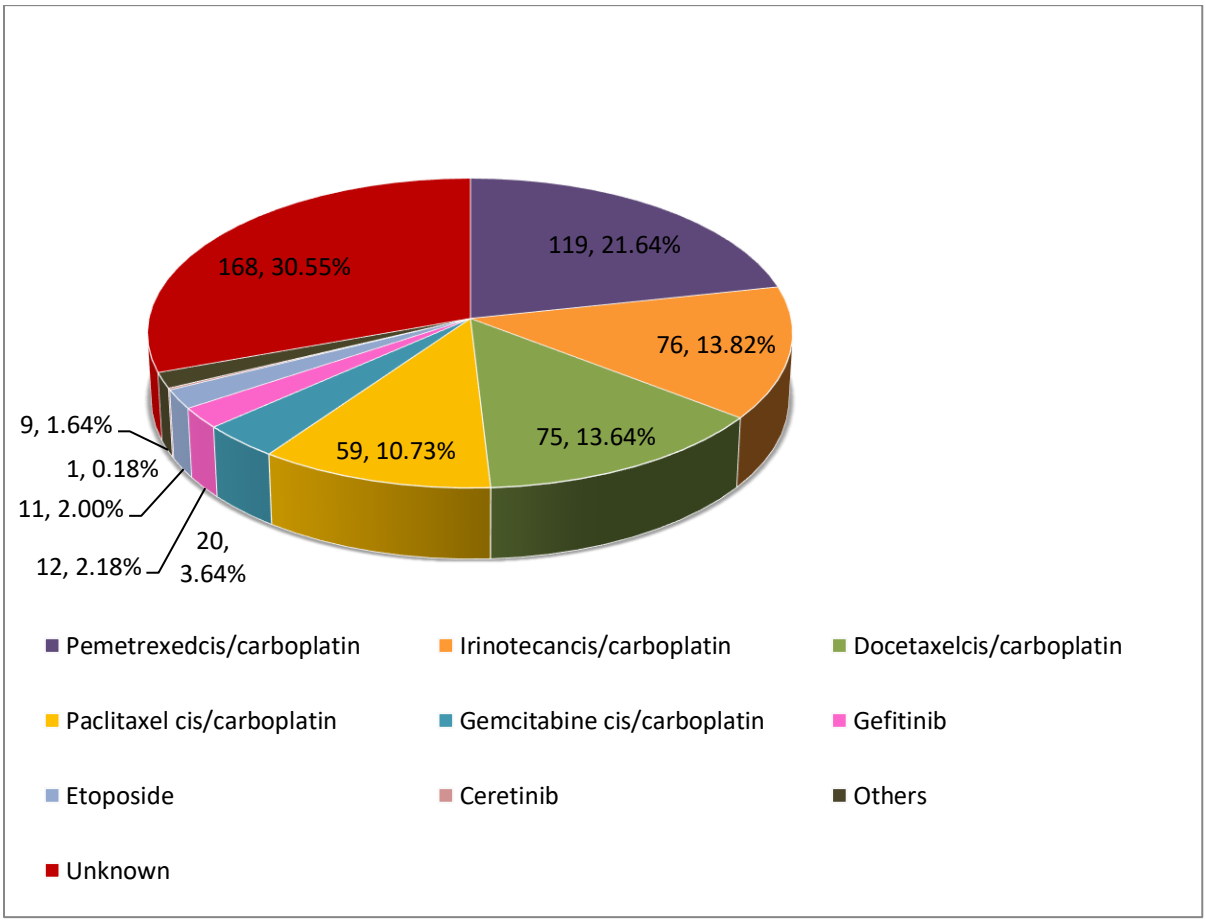


Figure 5.18: Pie chart illustrating a distribution of North Indian Lung cancer patients based on Doublet platinum-based chemotherapeutic regimen

5.5. Demographic characteristics of the study subjects for *NQO1, SULT1A1, EPHX1,* and *NAT2* gene

The sample population in the current study has been characterized by various epidemiological and demographic parameters, including age, gender, smoking status, and the clinical specifications of the cases, as tabulated in Table 5.1. The mean age came out to be 59.23± 10.28 years and 58.49± 11.67 years in cases and controls, respectively, and hence there was no significant difference between the mean ages of both cases and controls (p=0.27). The study was performed on 550 cases and 550 controls, including 458 (83.27%) males in cases and 459(83.45%) males in controls whereas 92 (16.73%) and 91 (16.55%) were females both in case and control groups, respectively. No significant difference was found between the distribution of males and females in cases and controls (p=0.18). Hence there is no sampling bias based on gender. The number of smokers and non-smokers was compared between the two study groups. The case group comprised 80.73% smokers, whereas controls had only 77.27% smokers; no significant difference was observed in the proportion of smokers and non-smokers (p=0.18). An attempt was made to adequately match the number of smokers and non-smokers in our study. Smoking was quantified in terms of pack years [(cigarettes or beedis per day/20) X years smoked], the means of pack years were compared among cases and controls, and a significant difference was observed (p<**0.0001**).

Further, the different clinical specifications such as histology, clinical stage, tumor size extensions, lymph node invasion, and distant metastasis were recorded from medical reports. Out of the 550 cases, 216(39.27%) were histologically identified as SQCC, 200 (36.36%) as ADCC, and 127 (23.1%) as SCLC. The TNM data was available for 500 cases, and when classified based on clinical stage, only a few cases were reported with stages I, II. In contrast, most of them were categorized into stages III and IV, respectively. (I: 5, II:19, III:229, IV:247) as shown in Table 5.1.

Table 5.1: Demographic characteristics among cases and controls

Variable	Total (N)	Cases, n (%)	Controls, n (%)	[a]*p*-value
Age (years)	550			
Mean ± SD		59.23 ± 10.28	58.49 ± 11.67	0.27
Range		28-86	27-89	
Gender	550			
Male		458 (83.27)	459 (83.45)	
Female		92 (16.73)	91 (16.55)	0.94
Smoking status	550			
Smokers		444 (80.73)	425 (77.27)	
Non-smokers		106 (19.27)	125 (22.73)	0.18
Pack years	550			
Mean ± SD		31.16 ± 28.28	20.76 ± 15.95	***p*<0.0001**
Histological types	550			
SQCC		216 (39.27)		
ADCC		200 (36.36)		
SCLC		127 (23.1)		
Others		6 (1.09)		
Unknown		1 (0.18)		
Overall survival	473			
Dead		336(71.04)		
Alive		137(28.96)		
TNM staging	550			
I		5 (0.91)		
II		19 (3.45)		
III		229 (41.64)		
IV		247 (44.91)		
Unclassified		50 (9.09)		
Tumor Size	550			
Tx		17 (3.09)		
T1		27 (4.91)		
T2		42 (7.64)		
T3		115 (20.91)		
T4		299 (54.36)		
Unknown		50 (9.09)		
Lymph Node	550			
Nx		1 (0.18)		
N0		65 (11.82)		
N1		44 (8)		
N2		233 (42.36)		
N3		153 (27.82)		
N4		4 (0.73)		
Unknown		50 (9.09)		
Metastasis	550			
Mx		2 (0.36)		
M0		254 (46.18)		
M1		246 (44.73)		
Unknown		48 (8.73)		

Abbreviations: SD = Standard Deviation, n = total number of lung cancer cases or control subjects, a) *p*-values were derived from Pearson Chi-square test except for age and pack-years; Student t-test was used for age and pack-years. All *p*-values are two-sided. $p < 0.05$ was considered statistically significant

5.6 Association of PhaseII detoxification genes towards lung cancer susceptibility

5.6.1 Role of *NQO1* polymorphism towards lung cancer susceptibility

5.6.1.1. Genotypic and allelic frequencies of *NQO1* polymorphic variant among cases and controls and their role in modifying the susceptibility towards lung cancer

Table 5.2 demonstrates the allelic and genotypic frequency distribution of *NQO1*polymorphic variant *609C>T* (rs1800566, *Pro*^{187}*Ser*). The genotypic frequencies of *NQO1* 609C>T did not exhibit any significant difference between cases and controls (χ2=1.84; df=2; p=0.39), as shown in Table 5.2. No significant departure was observed from Hardy-Weinberg equilibrium (HWE) for *NQO1* genotype among controls (χ^2=0.46; df=1; p=0.49) and cases (χ^2=0.89; df=1; p=0.34). It was observed that the mutant (*TT*) genotype frequency was over-represented in cases than controls (9.64% vs. 7.45%). The minor allele frequency (MAF) was 0.31 and 0.28 for cases and controls, respectively. The homozygous *CC* (wild) genotype was used as a reference. The association analysis revealed that subjects with mutant genotype *TT* were at slightly higher risk (AOR= 1.42; 95% C.I.= 0.91-2.22) of acquiring lung cancer but were not statistically significant. In the dominant model, no significant association was observed towards lung cancer (AOR= 1.11, 95% C.I.= 0.88-1.41).

Table 5.2: Genotypic and allelic distribution of the *NQO1* genetic variant and its association with risk of lung cancer

Genotype rs1800566 C/T	**Controls (550) N(%)**	**Cases (550) N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model				
CC	279 (50.73)	266 (48.36)	1.00 (Reference)	
CT	230 (41.82)	231 (42.00)	1.06 (0.83-1.36)	0.65
TT	41 (7.45)	53 (9.64)	1.42 (0.91-2.22)	0.12
Dominant model				
CC	279 (50.73)	266 (48.36)	1.00 (Reference)	
CT+TT	271 (49.27)	284 (51.64)	1.11 (0.88-1.41)	0.38
Recessive model				
CC+CT	509 (92.55)	497 (90.36)	1.00 (Reference)	
TT	41 (7.45)	53 (9.64)	1.38 (0.89-2.12)	0.14
C (Allele)	788	763		
T (Allele)	312	337		
MAF	0.28	0.31		
	χ2=1.84, df=2			0.39

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender, and smoking status. (b)-Two-sided χ2 test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table

Further, the analysis was carried out using the recessive model wherein heterozygous genotype was used as the reference. It was observed that subjects carrying the two copies of the mutant allele (*TT*) exhibited no significant association towards risk for lung cancer (AOR= 1.38, 95% C.I.= 0.89-2.12) (Table 5.2). Further, the *ad*-hoc analysis was carried out wherein crude OR and OR after adjustment with smoking status only (smoking being an important risk factor for lung cancer) was calculated as tabulated in Table 5.3. The results depict that there were not much difference in OR with adjustment and without adjustment with smoking status.

Table 5.3: *Ad*-hoc analysis of the *NQO1* genetic variant and its association with risk of lung cancer

Genotype	Without adjustment				With adjustment			
rs1800566 *C/T*	OR[a] (95% CI)	Coeffecient	SE	P[c]	OR[b] (95% CI)	Coeffecient	SE	P[c]
Codominant model								
CC	1.00 (Reference)				1.00 (Reference)			
CT	1.05(0.82-1.35)	0.052	0.126	0.68	1.05(0.82-1.35)	0.051	0.126	0.685
TT	1.356(0.87-2.11)	0.304	0.225	0.176	1.372(0.88-2.13)	0.316	0.225	0.161
Dominant model								
CC	1.00 (Reference)				1.00 (Reference)			
CT+TT	1.099(0.87-1.39)	0.094	0.121	0.433	1.101(0.87-1.39)	0.097	0.121	0.423
Recessive model								
CC+CT	1.00 (Reference)				1.00 (Reference)			
TT	1.324(0.86-2.03)	0.280	0.217	0.197	1.346(0.88-2.06)	0.297	0.218	0.173

a) Odds ratios, 95% confidence intervals, b) adjusted Odds ratio, 95% confidence interval calculated, and c) corresponding *p*-values were calculated by logistic regression analysis b) after adjusting for smoking status.

5.6.1.2.Genetic and allelic distribution of *NQO1* based on histological subtypes

Patients were categorized into categories according to their histological subtypes to segregate the population into high-risk subgroups. The genotypic frequency of *NQO1*variant was compared to determine the prevalence of high-risk subgroups across all categories, as shown in Table 5.4. In the case of *NQO1*(*C/T*), an increased frequency of subjects with mutant (TT) genotype was observed in ADCC (11.50%) and SCLC (9.45%) patients as compared to controls(7.45%) as shown in Table 5.4.In comparison, the genotypic frequency for mutant (*TT*) genotype in the case of SQCC was found to be similar to the control group (7.87%*vs.* 7.45%, respectively).When compared with the wild (*CC*) genotype, the risk was detected to be associated in ADCC patients having mutant (*TT*) genotype, but the risk was not statistically significant(AOR=1.66; 95% C.I.= 0.92-2.98; p= 0.09), nonetheless a visible trend was observed. Among all the models evaluated,

the recessive model was the one that did show some association among ADCC and SCLC but was not significant (Table 5.4).

Table 5.4: Genotypic and allelic distribution of the *NQO1* genetic variant and its association with risk of lung cancer on histology

SQCC				
Genotype rs1800566 C/T	**Controls (550) N(%)**	**Cases (216) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model				
CC	279 (50.73)	110(50.93)	1.0(Reference)	
CT	230 (41.82)	89(41.2)	1.02(0.72-1.42)	0.93
TT	41 (7.45)	17(7.87)	1.15(0.61-2.17)	0.67
Dominant model				
CC	279 (50.73)	110(50.93)	1.0(Reference)	
CT+TT	271 (49.27)	106(49.07)	1.03(0.75-1.42)	0.83
Recessive model				
CC+CT	509(92.55)	199(92.13)	1.0(Reference)	
TT	41(7.45)	17(7.87)	1.16(0.63-2.13)	0.63
ADCC				
Genotype rs1800566 C/T	**Controls (550) N(%)**	**Cases (200) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model				
CC	279 (50.73)	89(44.5)	1.0(Reference)	
CT	230 (41.82)	88(44.0)	1.22(0.87-1.74)	0.25
TT	41 (7.45)	23(11.5)	1.66(0.92-2.98)	0.09
Dominant model				
CC	279 (50.73)	89(44.5)	1.0(Reference)	
CT+TT	271 (49.27)	111(55.5)	1.30(0.93-1.81)	0.12
Recessive model				
CC+CT	509(92.55)	177(88.5)	1.0(Reference)	
TT	41(7.45)	23 (11.5)	1.52(0.88-2.65)	0.13
SCLC				
Genotype rs1800566 C/T	**Controls (550) N(%)**	**Cases (127) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model				
CC	279 (50.73)	64(50.39)	1.0(Reference)	
CT	230 (41.82)	51(40.16)	1.01(0.66-1.54)	0.97
TT	41 (7.45)	12(9.45)	1.61(0.78-3.33)	0.20
Dominant model				
CC	279 (50.73)	64(50.39)	1.0(Reference)	
CT+TT	271 (49.27)	63(49.61)	1.08(0.72-1.62)	0.70
Recessive model				
CC+CT	509(92.55)	115(90.55)	1.0(Reference)	
TT	41(7.45)	12(9.45)	1.62(0.81-3.24)	0.17

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender, and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table

5.6.1.3.The risk associated with *NQO1* polymorphism stratified by smoking

The relationship between *NQO1* polymorphism and smoking, was also evaluated in the current study; patients were segregated into smokers and non-smokers based on their smoking status. Table 5.5 summarizes the genotypic frequency of smokers and non-smokers, along with an odds ratio, 95% C.I and p-values. The study population who were smokers was further sub-divided into light smokers with consumption of up to 25 pack-years and heavy smokers with pack-years of more than 25 years. The percentage of mutants for the *NQO1* polymorphism among smokers was 8.78%, which was slightly higher than controls (6.82%). As a result, a slight trend of increased risk was seen in smokers who were carrying mutant (*TT*) genotype (AOR=1.51, 95% C.I= 0.89-2.53). Similarly, the frequency of mutant genotype was found to be over-represented in non-smokers in cases compared to controls (13.21% vs. 9.6%). A slightly increased risk was observed in non-smokers carrying mutant (*TT*) genotype (AOR=1.47, 95% C.I.= 0.61-3.57), but lacks statistical significance. Thus, no significant association was observed between smoking and *NQO1* polymorphism after stratification of lung cancer subjects based on smoking status and neither for light smokers nor for heavy smokers in case of any model evaluated, as shown in Table 5.5.

Table 5.5: Relationship of different *NQO1*genotypes with the smoking status of cases and controls

Genotype rs1800566 C/T	SMOKER				NON-SMOKER			
	Controls (425) N(%)	**Cases (444) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (125) N(%)**	**Cases (106) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model								
CC	218(51.30)	215(48.42)	1.0(Reference)		61(48.8)	51(48.11)	1.0(Reference)	
CT	178(41.88)	190(42.80)	1.11(0.83-1.47)	0.48	52(41.6)	41(38.68)	0.99(0.56-1.76)	0.99
TT	29(6.82)	39(8.78)	1.51(0.89-2.53)	0.12	12(9.6)	14(13.21)	1.47(0.61-3.57)	0.39
Dominant model								
CC	218(51.30)	215(48.42)	1.0(Reference)					
CT+TT	207(48.7)	229(51.58)	1.16(0.88-1.51)	0.29	64(51.2)	55(51.89)	1.09(0.64-1.87)	0.73
Recessive model								
CC+CT	396()	405()	1.0(Reference)		113(90.40)	92(86.79)	1.0(Reference)	
TT	29()	39()	1.42(0.86-2.34)	0.17	12(9.60)	14(13.21)	1.55(0.66-3.61)	0.31
	LIGHT SMOKER				**HEAVY SMOKER**			
rs1800566 C/T	**Controls (300) N(%)**	**Cases (248) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (125) N(%)**	**Cases (196) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model								
CC	149(49.67)	120(48.39)	1.0(Reference)		69(55.20)	95(48.47)	1.0(Reference)	
CT	131(43.66)	107(43.14)	1.03(0.72-1.47)	0.88	47(37.60)	83(42.35)	1.32(0.82-2.13)	0.25
TT	20(6.67)	21(8.47)	1.45(0.75-2.81)	0.27	9(7.20)	18(9.18)	1.71(0.71-4.12)	0.23
Dominant model								
CC	149(49.67)	120(48.39)	1.0(Reference)		69(55.20)	95(48.47)	1.0(Reference)	
CT+TT	151(50.33)	128(51.61)	1.08(0.77-1.53)	0.65	56(44.80)	101(51.53)	1.36(0.86-2.15)	0.18
Recessive model								
CC+CT	280(93.33)	227(91.53)	1.0(Reference)		116(92.80)	178(90.82)	1.0(Reference)	
TT	20(6.67)	21(8.47)	1.44(0.76-2.74)	0.26	9(7.20)	18(9.18)	1.41(0.61-3.28)	0.42

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.1.4. Relation between *NQO1* Genotypic Distribution and Clinic-pathologic Characteristics

The influence of *NQO1* variant SNP on several clinic-pathological features such as tumor stage, primary tumor extension, lymph node invasion, and distant metastasis is shown in Table 5.6. The patients were divided into two groups based on clinical stage: stage III and stage IV and tumor extension: TX+T1+T2 and T3+T4, and the data are reported in Table 5.6.To study the association between *NQO1* gene polymorphism and various clinic-pathological features, the genotypic frequencies for *NQO1 609C>T* polymorphism were calculated for stage III and stage IV patients. The genotypic frequencies were compared to see any significant difference in the genotypic frequencies between the two categories based on the stage of the tumor. The frequency of lung cancer subjects carrying the mutant allele for stage III was double (13.97%) than the subjects with stage IV lung cancer (6.88%). The subjects diagnosed with stage III lung cancer and carrying the T allele were associated with 3.57 fold increased lung cancer risk (AOR=3.57; 95% C.I.=1.72-7.41; p=**0.0006**). Using the dominant model the combined genotype (*CT+TT*) exhibit 1.58 fold increased risk for lung cancer (95% C.I.= 1.08-2.31; p=**0.018**). Whereas, when *CC+CT* was taken as reference in recessive model the patients carrying *TT* genotype was not found to be associated with an increased risk of lung cancer (AOR=0.58; 95% C.I.=0.41-0.81; p=**0.0015).** The patients were classified into two groups based on lymph node invasion: N0 and N1+N2+N3. The subjects with *TT* genotype were overrepresented in the N2+N3+N4 category compared to the Nx+N0+N1 group (12.05% *vs.* 2.73%). Our data exhibited a five-fold increased probability of lymph node invasion in subjects with variant genotype than subjects with no lymph node involvement (AOR=5.28; 95% C.I.=1.58-17.68; p=**0.007**). Further, the recessive model exhibited a four-fold increased probability of lymph node invasion in subjects with variant genotype than subjects with no lymph node involvement when *CC+CT* genotype was taken as reference (AOR=4.88; 95% C.I.=1.48-16.04; p=**0.009**). The further analysis consisted of finding an association of *NQO1* variants with metastasis (progressive disease). The subjects were divided into two groups: the one involving metastasis and the other not involving metastasis. Our results demonstrated that the *NQO1 TT* genotype was present at a higher frequency in patients without metastasis (M0, 12.5%) than in patients with metastasis (M1, 7.32%). When we compared the two using the unconditional logistic regression model, we discovered that the

NQO1TT genotype was associated with a substantial reduction in the risk of metastasis (AOR=0.35, 95% C.I. 0.17–0.69, p=**0.028**) (Table 5.6).

Table 5.6: Relationship of different genotypes with the clinical-pathological parameters

CLINICAL STAGE				
Genotype rs1800566 C/T	**III 229, N(%)**	**IV 247, N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model				
CC	95(41.49)	128(51.82)	1.0(Reference)	
CT	102(44.54)	102(41.30)	1.35(0.90-2.01)	0.14
TT	32(13.97)	17(6.88)	3.57(1.72-7.41)	**0.0006**
Dominant model				
CC	95(41.49)	128(51.82)	1.0(Reference)	
CT+TT	134(58.51)	119(48.18)	1.58(1.08-2.31)	**0.018**
Recessive model				
CC+CT	197(86.03)	230(93.12)	1.0(Reference)	
TT	32(13.97)	17(6.88)	0.58(0.41-0.81)	**0.0015**
PRIMARY TUMOR EXTENSION				
Genotype rs1800566 C/T	**T_3 115, N(%)**	**T_4 299, N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model				
CC	57(49.57)	141(47.16)	1.0(Reference)	
CT	48(41.74)	127(42.47)	1.08(0.68-1.71)	0.74
TT	10(8.69)	31(10.37)	1.27(0.58-2.79)	0.55
Dominant model				
CC	57(49.57)	141(47.16)	1.0(Reference)	
CT+TT	58(50.43)	158(52.84)	1.09(0.70-1.68)	0.71
Recessive model				
CC+CT	105(91.30)	268(89.63)	1.0(Reference)	
TT	10(8.70)	31(10.37)	1.20(0.56-2.54)	0.64
LYMPH NODE INVASION				
Genotype rs1800566 C/T	**N_x+N_0+N_1 110, N(%)**	**N_2+N_3+N_4 390, N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model				
CC	61(55.45)	182(46.67)	1.0(Reference)	
CT	46(41.82)	161(41.28)	1.17(0.75-1.82)	0.48
TT	3(2.73)	47(12.05)	5.28(1.58-17.68)	**0.007**
Dominant model				
CC			1.0(Reference)	
CT+TT	49(44.55)	208(53.33)	1.41(0.92-2.17)	0.11
Recessive model				
CC+CT	107(97.27)	343(87.95)	1.0(Reference)	
TT	3(2.73)	47(12.05)	4.88(1.48-16.04)	**0.009**

	METASTASIS			
Genotype rs1800566 C/T	**No 256, N(%)**	**Yes 246, N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model				
CC	109(42.58)	134(54.47)	1.0(Reference)	
CT	115(44.92)	94(38.21)	0.64(0.44-0.95)	**0.026**
TT	32(12.50)	18(7.32)	0.35(0.17-0.69)	**0.0028**
Dominant model				
CC			1.0(Reference)	
CT+TT	147(57.42)	112(45.53)	0.58(0.39-0.84)	**0.0037**
Recessive model				
CC+CT	224(87.50)	228(92.68)	1.0(Reference)	
TT	32(12.50)	18(7.32)	0.44(0.23-0.84)	**0.0135**

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding p-values were calculated by logistic regression analysis after adjusting for age, smoking and gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Furthermore, combining the two risk genotypes, *i.e*., heterozygous and mutant genotype (*CT + TT*) for the dominant model, we observed decreased risk for metastasis (AOR=0.58, 95% C.I.=0.39–0.84, p= **0.0037**). Similarly, for recessive model the variant genotype exhibited the decreased risk for metastasis (AOR=0.44, 95% C.I.=0.23–0.84, p= **0.0135**). Thus, our findings suggest that these genotypes may have a protective effect on lung cancer progression to other organs.

5.6.2. Statistical Analysis of *SULT1A1 Arg^{213}His* polymorphism towards lung cancer susceptibility

5.6.2.1. Overall Association of *SULT1A1* polymorphic variant with lung cancer susceptibility

The allelic and genotypic distribution of *SULT1A1* polymorphism, *i.e., Arg^{213}His,* is shown in Table 5.7. No significant departure was observed from Hardy-Weinberg equilibrium (HWE) for *SULT1A1* genotype among cases (χ^2=1.72; df=1; p=0.19) and controls (χ^2=3.40; df=1; p=0.07), signifying no sample bias. Statistical analysis revealed that lung cancer patients had a higher representation of wild-type genotype (*Arg/Arg*) than controls (66.18% vs. 61.09%). In contrast, heterozygous (*Arg/His*) and homozygous variant genotype (*His/His*) were over-represented in controls as compared to cases (*Arg/His*: 32.36% vs. 29.27%, and *His/His*: 6.55% vs. 4.55%) as shown in Table 5.7. Therefore, overall no significant difference was observed for genotypic distribution amongst cases and controls ($\chi 2$=3.96; df=2; p=0.14). The minor allele frequency

(MAF) for controls (0.23) was higher as compared to cases (0.19). Table 5.7 further showed the association of lung cancer risk with the *SULT1A1* polymorphic variant. Adjusted ORs and 95% C.I. were calculated using logistic regression analysis. To further explore the correlations between *SULT1A1* variants and lung cancer risk, three different genetic models (co-dominant, dominant, and recessive) were studied. The most common genotype, *i.e.*, *Arg/Arg* (wild), was considered the reference. It was observed that in the co-dominant model, the *His/His* genotype did not exhibit an increased risk of lung cancer (AOR=0.66; 95% C.I.= 0.38-1.12). In the dominant model, a reduced lung cancer risk was observed but marginally significant (AOR=0.80, 95% C.I.= 0.62-1.02). Further, the analysis was also evaluated using the recessive model wherein wild and heterozygous genotype (*Arg/His+His/His*) combined was used as the reference; it was observed that the subjects carrying the two copies of the variant allele (*His/His*) exhibited no significant association towards risk for lung cancer (AOR= 0.70; 95% C.I.= 0.41-1.19) (Table 5.7).

Table 5.7: Genotypic and allelic distribution of the *SULT1A1* genetic variant and its association with risk of lung cancer

Genotype rs9282861,Arg^{213}His	**Controls (550) N(%)**	**Cases (550) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Codominant model				
Arg/Arg	336 (61.09)	364 (66.18)	1.00 (Reference)	
Arg/His	178 (32.36)	161 (29.27)	0.82 (0.63-1.07)	0.15
His/His	36 (6.55)	25 (4.55)	0.66 (0.38-1.12)	0.12
Arg (Allele)	850	889		
His (Allele)	250	211		
MAF	0.23	0.19		
	χ^2= 3.96, df= 2, p= 0.14			
Dominant model				
Arg/Arg	336 (61.09)	364 (66.18)	1.00 (Reference)	
Arg/His+His/His	214 (38.91)	186 (33.82)	0.80 (0.62-1.02)	0.07
Recessive model				
Arg/His+Arg/Arg	510 (92.73)	521(94.73)	1.00 (Reference)	
His/His	40 (7.27)	29(5.27)	0.70 (0.41-1.19)	0.19

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender, and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the TABLE.

Further, the *ad*-hoc analysis was carried out wherein crude OR and OR after adjustment with smoking status only (smoking being an important risk factor for lung cancer) was calculated as tabulated in Table 5.8. The results depict that there were not much difference in OR with adjustment and without adjustment with smoking status.

Table 5.8: *Ad*-hoc analysis of the *SULT1A1* genetic variant and its association with risk of lung cancer

Genotype rs9282861 *Arg*^{213}*His*	Without adjustment				With adjustment			
	OR[a] (95% CI)	Coeffecient	SE	p[c]	OR[b] (95% CI)	Coeffecient	SE	p[c]
Codominant model								
Arg/Arg	1.00 (Reference)				1.00 (Reference)			
Arg/His	0.835(0.64-1.08)	-0.180	0.132	0.173	0.829(0.64-1.07)	-0.187	0.133	0.158
His/His	0.641(0.38-1.09)	-0.445	0.271	0.101	0.646(0.38-1.10)	-0.436	0.272	0.109
Dominant model								
Arg/Arg	1.00 (Reference)				1.00 (Reference)			
Arg/His+ His/His	0.802(0.63-1.03)	-0.220	0.125	0.079	0.798(0.62-1.08)	-0.226	0.125	0.073
Recessive model								
Arg/His+ His/His	1.00 (Reference)				1.00 (Reference)			
His/His	0.679(0.40-1.15)	-0.386	0.268	0.149	0.685(0.40-1.16)	-0.379	0.268	0.157

a) Odds ratios, 95% confidence intervals, b) adjusted Odds ratio, 95% confidence interval calculated, and c) corresponding *p*-values were calculated by logistic regression analysis b) after adjusting for smoking status.

5.6.2.2. Genetic and allelic distribution of *SULT1A1 Arg*^{213}*His* polymorphism based on histological subtypes

The study group was further comprehended into histological subgroups to evaluate the severity of risk towards the disease based on lung cancer histology. As shown in Table 5.9, the frequency of mutant (*His/His*) genotype was overrepresented in SQCC (5.09%) compared to both ADCC (4.00%) and SCLC (4.73%), respectively. Three alternative genetic models (co-dominant, dominant, and recessive) were assessed to investigate further the associations between *SULT1A1* polymorphisms and lung cancer risk among different histological subtypes. After applying statistics, none of the three genetic models had any association with lung cancer susceptibility. In the case of the co-dominant and dominant model, a trend was visible for patients carrying the mutant genotype exhibiting a protective effect towards SCLC with AOR of 0.66 (p= 0.08, and 0.06, respectively). Similarly, the SQCC patients carrying heterozygous genotype and combined genotype (*Arg/His+ His/His*) had a trend, exhibiting protective effect towards SQCC in case of the co-dominant and dominant model, respectively, with an AOR of 0.72 (95% C.I.=0.51-1.04,) and 0.73 (95% C.I.=0.52-1.02).

Table 5.9: Genotypic and allelic distribution of the *SULT1A1* genetic variant and its association with risk of lung cancer on histology

Genotype	**SCLC**			
Arg^{213}His	**Controls (550) N(%)**	**Cases (127) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Codominant model				
Arg/Arg	336 (61.09)	87 (68.50)	1.0(Reference)	
Arg/His	178 (32.36)	34 (26.77)	0.66(0.42-1.05)	0.08
His/His	36 (6.55)	6 (4.73)	0.65(0.25-1.68)	0.38
Dominant model				
Arg/Arg	336 (61.09)	87 (68.50)	1.0(Reference)	
Arg/His+ His/His	214 (38.91)	40 (31.50)	0.66(0.43-1.02)	0.06
Recessive model				
Arg/His+ His/His	514 (93.45)	121 (95.27)	1.0(Reference)	
His/His	36 (6.55)	6 (4.73)	0.74(0.29-1.90)	0.54
Genotype	**ADCC**			
Arg^{213}His	**Controls (550) N(%)**	**Cases (200) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Codominant model				
Arg/Arg	336 (61.09)	129(64.50)	1.0(Reference)	
Arg/His	178 (32.36)	63 (31.50)	0.98(0.68-1.40)	0.90
His/His	36 (6.55)	8 (4.00)	0.57(0.25-1.27)	0.17
Dominant model				
Arg/Arg	336 (61.09)	129(64.50)	1.0(Reference)	
Arg/His+ His/His	214 (38.91)	71 (35.50)	0.91(0.64-1.28)	0.57
Recessive model				
Arg/His+ His/His	514 (93.45)	192 (96.00)	1.0(Reference)	
His/His	36 (6.55)	8 (4.00)	0.57(0.26-1.27)	0.17
Genotype	**SQCC**			
Arg^{213}His	**Controls (550) N(%)**	**Cases (216) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Codominant model				
Arg/Arg	336 (61.09)	144 (66.67)	1.0(Reference)	
Arg/His	178 (32.36)	61 (28.24)	0.72(0.51-1.04)	0.08
His/His	36 (6.55)	11 (5.09)	0.74(0.36-1.53)	0.42
Dominant model				
Arg/Arg	336 (61.09)	144 (66.67)	1.0(Reference)	
Arg/His+ His/His	214 (38.91)	72 (33.33)	0.73(0.52-1.02)	0.07
Recessive model				
Arg/His+ His/His	514 (93.45)	205 (94.91)	1.0(Reference)	
His/His	36 (6.55)	11 (5.09)	0.82(0.40-1.68)	0.60

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender, and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table

5.6.2.2.1. Role of *SULT1A1* polymorphism in different histological subtypes of lung cancer based on gender

We also evaluated the role of *SULT1A1* in the different histological subtypes of lung cancer after stratification based on gender (Table 5.10). Among the several histological subtypes of lung cancer, we discovered that in the case of SQCC, male subjects with the Arg/*His* genotype (AOR=0.70; 95% C.I.= 0.48-1.01; p=**0.05**) in the co-dominant model and *His^{213}* allele

(*Arg/His+His/His*) genotype (AOR=0.69; 95% C.I.=0.49-0.98; p=**0.04**) in the dominant model showed a reduced risk towards lung cancer susceptibility (Table 5.10). The SCLC and ADCC subjects did not exhibit any risk towards lung cancer predisposition (Table 5.10).

5.6.2.3. Stratification analysis by gender

The lung cancer cases and healthy controls were further stratified based on gender and tabulated in Table 5.11. The frequency distribution of wild (*Arg/Arg*) genotype and mutant (*His/His*) genotype was higher in controls as compared to cases in females (*Arg/Arg*: 69.23% *vs*. 63.04%; *His/His*: 5.50% *vs*. 2.18%). Whereas, in the case of males, the wild genotype (*Arg/Arg*) was over-represented in cases than controls (66.81% *vs*. 59.48%) (Table 5.11). Further, when the co-dominant model was used, it was discovered that male lung cancer patients bearing the *Arg/His* genotype had no higher risk of lung cancer when compared to the reference (*Arg/Arg*) genotype(AOR= 0.74; 95% C.I.= 0.56-0.99; p=**0.04**). In addition, when we applied the dominant model, wherein the *His*^{213} allele genotypes (*Arg/His+His/His*) were combined and compared to the *Arg/Arg* genotype, it was pronounced that male lung cancer subjects carrying the variant allele showed a significantly decreased genetic predisposition towards lung cancer (AOR= 0.73; 95% C.I.=0.55-0.95; p=**0.02**) (Table 5.11). On the other hand, the recessive model showed no association towards lung cancer risk (AOR=0.72; 95% C.I.=0.41-1.26).

Further, we also observed the correlation between lung cancer susceptibility and *SULT1A1* variants after stratification analysis based on 59 years, the mean age of the study group (Table 5.12). None of the three models found no association between *SULT1A1* genotype and lung cancer risk in male participants aged more than or less than 59 years. ORs for *SULT1A1 His*^{213} allele could not be determined due to the absence of *His*^{213} allele in female lung cancer individuals under the age of 59 and control subjects above 59.Using the co-dominant model, female subjects carrying the *Arg/His* genotype, both above (AOR=1.84; 95% C.I.=0.59-5.72;) and below (AOR=1.45; 95% C.I.=0.58-3.65) the age of 59, had an increased risk of lung cancer. When applying the dominant model, the combined genotype (*Arg/His+His/His*) enhanced lung cancer susceptibility by 2.23 times in female patients over 59 (AOR=2.23; 95% C.I.=0.74-6.73), although this was not statistically significant. In contrast, in male subjects over the age of 59, the dominant model did not show increased susceptibility to lung cancer (AOR=0.71; 95% C.I.=0.49-1.04; p=0.08) (Table 5.12).

Table 5.10: Genotypic frequencies of *SULT1A1* among cases and controls and their relationship with risk of SQCC histological subtype in males

Genotype rs9282861 *Arg^{213}His*	***SQCC***				***SCLC***			
	Controls (459) N(%)	**Cases (203) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (459), N(%)**	**Cases (106) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model								
Arg/Arg	273 (59.48)	137 (67.49)	1.00 (Reference)		273 (59.48)	72 (67.92)	1.00 (Reference)	
Arg/His	155 (33.77)	56 (27.59)	0.70 (0.48-1.01)	**0.05**	155 (33.77)	28 (26.42)	0.66 (0.41-1.07)	0.09
His/His	31 (6.75)	10 (4.92)	0.70 (0.33-1.48)	0.35	31 (6.75)	6 (5.66)	0.82 (0.33-2.08)	0.68
Dominant model								
Arg/Arg	273 (59.48)	137 (67.49)	1.00 (Reference)		273 (59.48)	72 (67.92)	1.00 (Reference)	
Arg/His+His/His	186 (40.52)	66 (32.51)	0.69 (0.49-0.98)	**0.04**	186 (40.52)	34 (32.08)	0.68 (0.44-1.08)	0.10
Recessive model								
Arg/His+Arg/Arg	428 (93.25)	193 (95.07)	1.00 (Reference)		428 (93.25)	100 (94.34)	1.00 (Reference)	
His/His	31 (6.75)	10 (4.92)	0.77 (0.37-1.63)	0.50	31 (6.75)	6 (5.66)	0.94 (0.38-2.35)	0.89

Genotype rs9282861 *Arg^{213}His*	***ADCC***			
	Controls (459) N(%)	**Cases (142) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model				
Arg/Arg	273 (59.48)	93(65.49)	1.00 (Reference)	
Arg/His	155 (33.77)	42(29.58)	0.86(0.56-1.31)	0.47
His/His	31 (6.75)	7(4.93)	0.57(0.23-1.38)	0.21
Dominant model				
Arg/Arg	273 (59.48)	93(65.49)	1.00 (Reference)	
Arg/His+His/His	186 (40.52)	49(34.51)	0.80(0.53-1.19)	0.27
Recessive model				
Arg/His+Arg/Arg	428 (93.25)	135(95.07)	1.00 (Reference)	
His/His	31 (6.75)	7(4.93)	0.58(0.24-1.40)	0.22

[a]Adjusted Odds ratios, [b] 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender and smoking. Abbreviation: AOR, Adjusted Odds ratio; CI, confidence interval.

Table 5.11: Relationship of different genotypes based on gender

Genotype rs9282861 *Arg*^{213}*His*	FEMALE				MALE			
	Controls (91) N(%)	**Cases (92) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (459) N(%)**	**Cases (458) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Co-dominant model								
Arg/Arg	63(69.23)	58(63.04)	1.0(Reference)		273(59.48)	306(66.81)	1.0(Reference)	
Arg/His	23(25.27)	32(34.78)	1.70(0.84-3.44)	0.14	155(33.77)	129(28.17)	0.74(0.56-0.99)	**0.04**
His/His	5(5.50)	2(2.18)	0.35(0.05-2.67)	0.31	31(6.75)	23(5.02)	0.66(0.38-1.17)	0.15
Dominant model								
Arg/Arg	63(69.23)	58(63.04)	1.0(Reference)		273(59.48)	306(66.81)	1.0(Reference)	
Arg/His+His/His	28(30.77)	34(36.96)	1.47(0.75-2.87)	0.26	186(40.52)	152(33.19)	0.73(0.55-0.95)	**0.02**
Recessive model								
Arg/His+Arg/Arg	86(94.50)	90(97.83)	1.0(Reference)		428(93.25)	435(94.98)	1.0(Reference)	
His/His	5(5.50)	2(2.17)	0.33(0.05-2.28)	0.26	31(6.75)	23(5.02)	0.72 (0.41-1.26)	0.25

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding p-values were calculated by logistic regression analysis after adjusting for age and smoking. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.12: Relationship of different *SULT1A1* genotypes of males and females based on age

Genotype rs9282861 *Arg^{213}His*	Age ≤ 59 years				Age > 59 years			
Males								
	Controls (234) N(%)	**Cases (193) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (225) N(%)**	**Cases (265) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model								
Arg/Arg	143 (61.11)	130 (67.36)	1.00 (Reference)		130 (57.78)	176 (66.41)	1.00 (Reference)	
Arg/His	75 (32.05)	52 (26.94)	0.75 (0.49-1.15)	0.19	80 (35.55)	77 (29.06)	0.74 (0.49-1.10)	0.13
His/His	16 (6.84)	11 (5.70)	0.78 (0.35-1.76)	0.55	15 (6.67)	12 (4.53)	0.58 (0.26-1.29)	0.18
Dominant model								
Arg/Arg	143 (61.11)	130 (67.36)	1.00 (Reference)		130 (57.78)	176 (66.41)	1.00 (Reference)	
Arg/His+His/His	91 (38.89)	63 (32.64)	0.75 (0.50-1.12)	0.16	95 (42.22)	89 (33.59)	0.71 (0.49-1.04)	0.08
Recessive model								
Arg/His+Arg/Arg	218 (93.16)	182 (94.30)	1.00 (Reference)		210 (93.33)	253 (95.47)	1.00 (Reference)	
His/His	16 (6.84)	11 (5.70)	0.83 (0.37-1.84)	0.64	15 (6.67)	12 (4.53)	0.65 (0.30-1.43)	0.28
Females								
Genotype rs9282861 *Arg^{213}His*	**Age ≤ 59 years**				**Age > 59 years**			
	Controls (58) N(%)	**Cases (51) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (33) N(%)**	**Cases (41) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model								
Arg/Arg	38 (65.52)	32 (62.75)	1.00 (Reference)		25 (75.76)	26 (63.41)	1.00 (Reference)	
Arg/His	15 (25.86)	19 (37.25)	1.45(0.58-3.65)	0.43	8 (24.24)	13 (31.71)	1.84 (0.59-5.72)	0.29
His/His	5 (8.62)	0 (0.00)	-	0.99	0 (0.00)	2 (4.88)	-	0.99
Dominant model								
Arg/Arg	38 (65.52)	32 (62.75)	1.00 (Reference)		25 (75.56)	26 (63.41)	1.00 (Reference)	
Arg/His+His/His	20 (34.48)	19 (37.25)	0.99 (0.41-2.39)	0.98	8 (24.24)	15 (36.59)	2.23 (0.74-6.73)	0.16
Recessive model								
Arg/His+Arg/Arg	53 (91.38)	51 (100.00)	1.00 (Reference)		33 (100.00)	39 (95.12)	1.00 (Reference)	
His/His	5 (8.62)	0 (0.00)	-	0.99	0 (0.00)	2 (4.88)	-	0.99

[a]Adjusted Odds ratios, [b] 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting smoking. Abbreviation: AOR, Adjusted Odds ratio; CI, confidence interval.

5.6.2.4. Interactive effect of smoking on the *SULT1A1* polymorphism and their association with lung cancer risk

To better understand the effect of gene-environment interaction, the study group was segregated into two groups based on smoking status as smokers and non-smokers (Table 5.13). The smokers were classified as light, and heavy smokers, based on the number of pack-years smoked. Individuals who had smoked less than 25 pack-years were classified as light smokers, while those who had smoked more than 25 pack-years were classified as heavy smokers. In our study, participants who were smokers and carrying mutant genotype (*His/His*) were shown to be less represented in cases when compared to controls (4.50% vs. 6.12%). Three alternative genetic models (co-dominant, dominant, and recessive) were analyzed based on smoking exposure to investigate the associations between *SULT1A1* polymorphisms and lung cancer risk. When both co-dominant and dominant models were applied, it was discovered that smokers with lung cancer who possessed either heterozygous genotype (*Arg/His*) (AOR= 0.70; 95% C.I.=0.52-0.95; p=**0.019**) or His^{213} allele genotype (*Arg/His+His/His*) (AOR=0.70; 95% C.I.=0.53-0.92; p=**0.01**), had a significantly reduced risk of developing lung cancer. Further, the lung cancer subjects who were heavy smokers and had heterozygous genotype (*Arg/His*) (AOR= 0.28; 95% C.I.= 0.10-0.81; p= **0.019**) or the combined (*Arg/His+His/His*) genotype (AOR= 0.56; 95% C.I.= 0.35-0.90; p= **0.017**) showed a reduced probability of developing lung cancer (Table 5.13).

Table 5.13: Relationship of different *SULT1A1* genotypes with the smoking status of cases and controls

Genotype rs9282861 *Arg^{213}His*	SMOKER				NON-SMOKER			
	Controls (425) N(%)	Cases (444) N(%)	AOR (95% C.I.) [a]	p [b]	Controls (125) N(%)	Cases (106) N(%)	AOR (95% C.I.) [a]	p [b]
Co-dominant model								
Arg/Arg	249 (58.59)	298 (67.12)	1.0(Reference)		87 (69.60)	66 (62.26)	1.0(Reference)	
Arg/His	150 (35.29)	126 (28.38)	0.70(0.52-0.95)	**0.019**	28 (22.40)	35 (33.02)	1.82(0.99-3.37)	**0.05**
His/His	26 (6.12)	20 (4.50)	0.66(0.36-1.22)	0.18	10 (8.00)	5 (4.72)	0.65(0.20-2.09)	0.47
Dominant model								
Arg/Arg	249 (58.59)	298 (67.12)	1.0(Reference)		87 (69.60)	66 (62.26)	1.0(Reference)	
Arg/His+His/His	176 (41.41)	146 (32.88)	0.70(0.53-0.92)	**0.01**	38 (30.40)	40 (37.74)	1.51(0.85-2.65)	0.15
Recessive model								
Arg/His+Arg/Arg	*399 (93.88)*	424 (95.50)	1.0(Reference)		115(92.00)	101(95.28)	1.0(Reference)	
His/His	*26 (6.12)*	20 (4.50)	0.74(0.40-1.36)	0.33	10 (8.00)	5 (4.72)	0.58(0.18-1.81)	0.34
Genotype rs9282861 *Arg^{213}His*	**LIGHT SMOKER**				**HEAVY SMOKER**			
	Controls (300) N(%)	**Cases (248) N(%)**	**AOR (95% C.I.) [a]**	**p [b]**	**Controls (125) N(%)**	**Cases (196) N(%)**	**AOR (95% C.I.) [a]**	**p [b]**
Co-dominant model								
Arg/Arg	176 (58.67)	159 (64.11)	1.0(Reference)		73 (58.40)	139 (70.92)	1.0(Reference)	
Arg/His	108 (36.00)	75 (30.24)	0.79(0.54-1.15)	0.22	42 (33.60)	51 (26.02)	0.63(0.38-1.04)	0.07
His/His	16 (5.33)	14 (5.65)	1.10(0.52-2.33)	0.81	10 (8.00)	6 (3.06)	0.28(0.10-0.81)	**0.019**
Dominant model								
Arg/Arg	176 (58.67)	159 (64.11)	1.0(Reference)		73 (58.40)	139 (70.92)	1.0(Reference)	
Arg/His+His/His	124 (41.33)	89 (35.89)	0.83(0.58-1.18)	0.29	52 (41.60)	57 (29.08)	0.56(0.35-0.90)	**0.017**
Recessive model								
Arg/His+Arg/Arg	284(94.67)	234(94.35)	1.0(Reference)		115(92.00)	190 (96.94)	1.0(Reference)	
His/His	16 (5.33)	14 (5.65)	1.18(0.56-2.46)	0.67	10 (8.00)	6 (3.06)	0.32(0.11-0.94)	**0.04**

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

We further stratified this group of heavy smokers based on pack years. According to Table 5.14, subjects who smoked 25-30 pack-years and carried the heterozygous (*Arg/His*) genotype did not show any relation to lung cancer risk (AOR=1.10; 95% C.I.=0.43-2.80). Using the co-dominant model, patients who smoked 30-40 pack-years and had the mutant (*His/His*) genotype did not have an increased risk of lung cancer susceptibility (AOR=0.48; 95% C.I.=0.07-3.27). Furthermore, the results show that lung cancer subjects who smoked pack years of more than 40 years and carried the mutant (*His/His)* genotype have a decreased effect on the risk of lung cancer (AOR=0.21; 95% C.I.=0.05-0.90; p=**0.036**). Similarly, in the co-dominant model, the *Arg/His* genotype showed a borderline significant reduction in lung cancer risk (AOR=0.42; 95% C.I.=0.17-1.01; p=**0.05**). On the other hand, when the dominant model was also applied, and it was observed that the lung cancer cases harboring the combined (*Arg/His+His/His*) genotype for the $Arg^{213}His$ polymorphism showed a protective effect towards lung cancer susceptibility (AOR=0.37; 95% C.I.=0.17-0.83; p=**0.015**) (Table 5.14).

Table 5.14: Relationship of different *SULT1A1* genotypes with the pack-years of heavy smokers amongst cases and controls

Genotype rs9282861 *Arg[213]His*	Pack years >25 - ≤30				Pack years >30 - ≤40				Pack years above 40			
	Controls (36) N(%)	Cases (55) N(%)	AOR (95% C.I.) [a]	*p*[b]	Controls (50) N(%)	Cases (48) N(%)	AOR (95% C.I.) [a]	*p*[b]	Controls (39) N(%)	Cases (93) N(%)	AOR (95% C.I.) [a]	*p*[b]
Co-dominant model												
Arg/Arg	24 (66.67)	38 (69.09)	1.0 (Reference)		28 (56.00)	31 (64.58)	1.00 (Reference)		21 (53.85)	70 (75.27)	1.00 (Reference)	
Arg/His	10 (27.78)	17 (30.91)	1.10 (0.43-2.80)	0.85	19 (38.00)	15 (31.25)	0.66 (0.27-1.59)	0.35	13 (33.33)	19 (20.43)	0.42 (0.17-1.01)	**0.05**
His/His	2 (5.55)	0 (0.00)	0.00	0.99	3 (6.00)	2 (4.17)	0.48 (0.07-3.27)	0.45	5 (12.82)	4 (4.30)	0.21 (0.05-0.90)	**0.036**
Dominant model												
Arg/Arg	24 (66.67)	38 (69.09)	1.0 (Reference)		28 (56.00)	31 (64.58)	1.00 (Reference)		21 (53.85)	70 (75.27)	1.00 (Reference)	
Arg/His+His/ His	12 (33.33)	17 (30.91)	0.91 (0.37-2.25)	0.84	22 (44.00)	17 (35.42)	0.64 (0.27-1.50)	0.30	18 (46.15)	23 (24.73)	0.37 (0.17-0.83)	**0.015**
Recessive model												
Arg/His+Arg/ Arg	34 (94.44)	55 (100.0)	1.0 (Reference)		47 (94.00)	46 (95.83)	1.00 (Reference)		34 (87.18)	89 (95.70)	1.00 (Reference)	
His/His	2 (5.56)	0 (0.00)	0.00	0.99	3 (6.00)	2 (4.17)	0.64 (0.10-4.09)	0.63	5 (12.82)	4 (4.30)	0.27 (0.07-1.13)	0.07

[a]Adjusted Odds ratios, [b] 95% confidence intervals, and corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender and smoking. Abbreviation: AOR, Adjusted Odds ratio; CI, confidence interval.

5.6.1.5.Association of *SULT1A1* genotype with clinical-pathological features

In the current study, we have evaluated the association of *SULT1A1* polymorphism on various clinic-pathological features; the subjects were segregated based on the stage of cancer (III vs. IV), direct tumor extension (T_3 vs. T_4), lymph node involvement (N_x+N_0+N_1 vs. N_2+N_3+N_4), and metastatic status (M_0 vs. M_1) (Table 5.15).

Based on the clinical stage, the patients were stratified into stage III and stage IV. The genotypic distribution for each polymorphic site has been estimated. The frequency distribution for subjects carrying the wild genotype was almost the same for both stage III and stage IV cases (65.94%, and 66.00%, respectively). In the case of *Arg*^{213}*His* polymorphism, the heterozygous patients (*Arg/His*) were found at a slightly higher frequency among stage IV cases (29.55%) as compared to stage III patients (28.82%), as shown in Table 5.15. For homozygous variant type, the prevalence of subjects having *His/His* genotype was higher among stage III (5.24%) subgroup than stage IV (4.45%) subgroup. Logistic regression analysis showed that patients in stage IV and carrying mutant genotype (*His/His*) were at lower risk when compared with patients in stage III of disease (AOR= 0.80, 95% C.I.= 0.33-1.94).

The following parameter evaluated for the *SULT1A1* polymorphism was primary tumor extension; the patients were stratified into two groups, namely, T4 and T3. The frequency of heterozygous genotype (*Arg/His*) was over-represented in patients with tumor size for T4 extension (29.43%) as compared to the patients having T3 (27.83%) tumor stage. A 3.4-fold high risk was observed for lung cancer patients carrying mutant genotype (*His/His*) and having T4 tumor stage with T3 stage as a reference (AOR=3.44; 95% C.I.= 0.77-15.37), but lacks statistical significance. Further, the analysis was carried out after stratifying patients based on lymph node invasion (N_x+N_0+N_1 vs. N_2+N_3+N_4). The frequency distribution of subjects carrying heterozygous genotype (*Arg/His*) was over-represented in the N_x+N_0+N_1 group compared to patients in the N_2+N_3+N_4 group (33.63% vs. 29.23%). At the same time, the homozygous variant genotype was highly represented in the N_2+N_3+N_4 group compared to the subjects in theN_x+N_0+N_1 group (4.87% vs. 2.73%) (Table 5.15). There was no significant association observed between *SULT1A1* polymorphism and lymph node invasion in all the three models evaluated.

The relationship of different genotypes of *Arg*^{213}*His* with metastatic status is demonstrated in Table 5.15. The genotypic frequencies were evaluated, and the comparison showed that 66.41% of lung cancer subjects did not show any metastasis (M0), and 64.63% had positive metastasis for wild genotype patients (*Arg/Arg*). The homozygous variant genotype (*His/His*) was also in almost similar frequency for both metastatic and non-metastatic groups (4.07% vs.4.69%). The frequency of subjects exhibiting heterozygous genotype (*Arg/His*) was 31.30% for positive metastasis (M1) and 28.90% for subjects lacking metastasis (M0). Since there was no significant difference in the frequency distribution among the two above discussed genotypic groups, it was evident that this polymorphic site did not show any relation with metastasis or spread of lung cancer to other organs, as shown in Table 5.15.

Table 5.15: Relationship of different genotypes with the clinical-pathological parameters

Genotype	CLINICAL STAGE				PRIMARY TUMOR EXTENSION			
rs9282861	**III**	**IV**	**AOR (95% C.I.)** [a]	***p*** [b]	**T3**	**T4**	**AOR (95% C.I.)** [a]	***p*** [b]
Arg*^{213}*His	**229**	**247**			**115**	**299**		
Co-dominant model								
Arg/Arg	151(65.94)	163(66.00)	1.0(Reference)		81(70.43)	195(65.22)	1.0(Reference)	
Arg/His	66(28.82)	73(29.55)	0.95(0.62-1.44)	0.80	32(27.83)	88(29.43)	1.14(0.70-1.86)	0.59
His/His	12(5.24)	11(4.45)	0.80(0.33-1.94)	0.61	2(1.74)	16(5.35)	3.44(0.77-15.37)	0.11
Dominant model								
Arg/Arg	151(65.94)	163(66.00)	1.0(Reference)		81(70.43)	195(65.22)	1.0(Reference)	
Arg/His+His/His	78(34.06)	84(34.00)	0.97(0.65-1.44)	0.88	34(29.57)	104(34.78)	1.28(0.80-2.05)	0.30
Recessive model								
Arg/His+His/His	217(94.76)	236(95.55)	1.0(Reference)		113(98.26)	283(94.65)	1.0(Reference)	
His/His	12(5.24)	11(4.45)	0.84(0.35-2.03)	0.71	2(1.74)	16(5.35)	3.38(0.76-14.97)	0.11
	LYMPH NODE INVASION				**METASTASIS**			
rs9282861	**Nx+N0+N1**	**N2+N3+N4**	**AOR (95% C.I.)** [a]	***p*** [b]	**No**	**Yes**	**AOR (95% C.I.)** [a]	***p*** [b]
Arg*^{213}*His	**110**	**390**			**256**	**246**		
Co-dominant model								
Arg/Arg	70(63.64)	257(65.90)	1.0(Reference)		170(66.41)	159(64.63)	1.0(Reference)	
Arg/His	37(33.63)	114(29.23)	0.84(0.53-1.33)	0.47	74(28.90)	77(31.30)	1.06(0.70-1.58)	0.79
His/His	3(2.73)	19(4.87)	1.74(0.50-6.09)	0.39	12(4.69)	10(4.07)	0.85(0.35-2.09)	0.72
Dominant model								
Arg/Arg	70(63.64)	257(65.90)	1.0(Reference)		170(66.41)	159(64.63)	1.0(Reference)	
Arg/His+His/His	40(36.36)	133(34.10)	0.91(0.58-1.42)	0.68	86(33.59)	87(35.37)	1.04(0.71-1.53)	0.84
Recessive model								
Arg/His+Arg/Arg	107(97.27)	371(95.13)	1.0(Reference)		244(95.31)	236(95.93)	1.0(Reference)	
His/His	3(2.73)	19(4.87)	1.83(0.53-6.33)	0.34	12(4.69)	10(4.07)	0.86(0.35-2.11)	0.74

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.2. Statistical analysis for polymorphic sites of *EPHX1* gene towards lung cancer susceptibility

5.6.2.5. Overall Association of *EPHX1* polymorphic variants with lung cancer susceptibility

Table 5.16 demonstrates the allelic and genotypic frequency distribution of the *EPHX1* polymorphic variants, *Tyr*^{113}*His* and *His*^{139}*Arg*. The genotypic frequencies of *EPHX1 Tyr*^{113}*His* among controls (n=550, χ2=1.72; df=1; p=0.19) and cases (n=550, χ2=0.01; df=1; p=0.91) were in Hardy-Weinberg equilibrium, suggesting no sample bias for both control and cases. Similarly, the genotypic frequencies for *EPHX1 His*^{139}*Arg* polymorphism for both control (n=550; χ2= 0.18; df=1; p=0.67) and cases (n=550; χ2=2.87; df=1; p=0.09) were in Hardy-Weinberg equilibrium. The wild genotype (*Tyr/Tyr*) for *EPHX1 Tyr*^{113}*His* gene was less prevalent in lung cancer subjects than the controls (34.18% *vs*. 39.09%). However, as shown in Table 5.16, the mutant genotype (*His/His*) frequency for *EPHX1 Tyr*^{113}*His* was 17.46% for cases compared to controls (16.18%). There was no significant difference in the genotypic distribution between cases and controls (χ^2=2.85, df=2, p=0.24). Minor allelic frequency (MAF) calculated for *EPHX1 Tyr*^{113}*His* in controls and cases were 0.38 and 0.31, respectively, suggesting that both variant alleles were more pronounced in the controls than in the controls cases. In the case of *EPHX1 His*^{139}*Arg*, the frequency of the mutant genotype (*Arg/Arg*) was more represented in the control group in comparison with cases (2.54% vs. 1.82%), whereas heterozygous genotype (*His/Arg*) was more presented in cases (30.00%) than controls (28.36%). Overall there was no significant difference in the distribution of genotypes between the controls and cases (χ^2=0.95, df=2, *p*=0.62). The MAF calculated it was the same for both controls and cases, *i.e*., 0.17. The three genetic models used showed no overall relationship between the *EPHX1* polymorphism (*Tyr*^{113}*His* and *His*^{139}*Arg*) and the risk of developing lung cancer (Table 5.16).

Further, the *ad*-hoc analysis was carried out wherein crude OR and OR after adjustment with smoking status only (smoking being an important risk factor for lung cancer) was calculated as tabulated in Table 5.17. The results depict that there were not much difference in OR with adjustment and without adjustment with smoking status.

Table 5.16: Genotypic and allelic distribution of the *EPHX1* genetic variants and its association with risk of lung cancer

Genotype ***Tyr*^{113}*His*** **(exon3)**	**Controls (550) N(%)**	**Cases (550) N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model				
Tyr/Tyr	215 (39.09)	188 (34.18)	1.00 (Reference)	
Tyr/His	246 (44.73)	266 (48.36)	1.23 (0.95-1.60)	0.12
His/His	89 (16.18)	96 (17.46)	1.23 (0.87-1.75)	0.24
Tyr (Allele)	676	763		
His (Allele)	424	337		
MAF	0.38	0.31		
		χ2=2.85, df=2 0.24		
Dominant model				
Tyr/Tyr	215 (39.09)	188 (34.18)	1.00 (Reference)	
Tyr/His+His/His	335 (60.91)	362 (65.82)	1.23 (0.96-1.57)	0.10
Recessive model				
Tyr/Tyr+Tyr/His	461 (83.82)	454 (82.54)	1.00 (Reference)	
His/His	89 (16.18)	96 (17.46)	1.10 (0.80-1.51)	0.55
Genotype ***His*^{139}*Arg*** **(exon4)**	**Controls (550) N(%)**	**Cases (550) N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model				
His/His	380 (69.10)	375 (68.18)	1.0(Reference)	
His/Arg	156 (28.36)	165 (30.00)	1.07 (0.82-1.39)	0.61
Arg/Arg	14 (2.54)	10 (1.82)	0.77 (0.34-1.77)	0.54
His (allele)	916	915		
Arg (allele)	184	185		
MAF	0.17	0.17		
χ2=0.95, df=2				0.62
Dominant model				
His/His	380 (69.10)	375 (68.18)	1.0(Reference)	
His/Arg + Arg/Arg	170 (30.90)	175 (31.82)	1.05 (0.81-1.35)	0.73
Recessive model				
His/His + His/Arg	536 (97.46)	540 (98.18)	1.0(Reference)	
Arg/Arg	14 (2.54)	10 (1.82)	0.75 (0.33-1.72)	0.50

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender, and smoking status. (b)-Two-sided χ2 test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.17: *Ad*-hoc analysis of the *EPHX1* genetic variant and its association with risk of lung cancer

Genotype ***Tyr*^{113}*His*** **(exon3)**	**Without adjustment**				**With adjustment**			
	OR[a] (95% CI)	**Coeffecient**	**SE**	**p[c]**	**OR[b] (95% CI)**	**Coeffecient**	**SE**	**p[c]**
Codominant model								
Tyr/Tyr	1.00 (Reference)				1.00 (Reference)			
Tyr/His	1.237(0.95-1.61)	0.212	0.133	0.111	1.238(0.95-1.61)	0.214	0.133	0.109
His/His	1.234(0.87-1.75)	0.210	0.178	0.238	1.234(0.87-1.75)	0.210	0.178	0.238
Dominant model								
Tyr/Tyr	1.00 (Reference)				1.00 (Reference)			
Tyr/His+ His/His	1.236(0.97-1.58)	0.212	0.125	0.091	1.238(0.97-1.58)	0.213	0.125	0.089

Recessive model								
Tyr/Tyr+ Tyr/His	1.00 (Reference)				1.00 (Reference)			
His/His	1.095(0.79-1.50)	0.091	0.161	0.573	1.094(0.79-1.50)	0.090	0.161	0.577
Genotype *His[139]Arg* (exon4)	**Without adjustment**				**With adjustment**			
	OR (95% CI)	**Coeffecient**	**SE**	**p**	**OR (95% CI)**	**Coeffecient**	**SE**	**P**
Codominant model								
His/His	1.00 (Reference)				1.00 (Reference)			
His/Arg	1.072(0.82-1.39)	0.069	0.133	0.603	1.078(0.83-1.40)	0.075	0.133	0.572
Arg/Arg	0.724(0.32-1.65)	-0.323	0.420	0.442	0.745(0.33-1.70)	-0.295	0.421	0.485
Dominant model								
His/His	1.00 (Reference)				1.00 (Reference)			
His/Arg+ Arg/Arg	1.043(0.81-1.34)	0.042	0.130	0.745	1.051(0.81-1.36)	0.049	0.130	0.701
Recessive model								
His/His+ His/Arg	1.00 (Reference)				1.00 (Reference)			
Arg/Arg	0.709(0.31-1.61)	-0.344	0.418	0.411	0.727(0.32-1.65)	-0.319	0.419	0.447

a) Odds ratios, 95% confidence intervals, b) adjusted Odds ratio, 95% confidence interval calculated, and c) corresponding *p*-values were calculated by logistic regression analysis b) after adjusting for smoking status.

5.6.2.6. Association of *EPHX1* polymorphisms (*Tyr[113]His* and *His[139]Arg*) with lung cancer risk based on histological subtypes

The lung cancer cases were segregated into histological subgroups to evaluate the risk of susceptibility amongst these subgroups, as shown in Table 5.18. SQCC patients with heterozygous genotype (*Tyr/His*) for *EPHX1 Tyr[113]His* were more represented as compared to wild genotype (*Tyr/Tyr*) (49.08% vs. 31.94%). 18.98% of cases with SQCC had mutant genotype (*His/His*). Subjects diagnosed with ADCC, 46.00% had heterozygous genotype, 37.50% had wild-type genotype, and 16.50% were mutants for *EPHX1 Tyr[113]His* polymorphism. In the SCLC subgroup, 32.28%, 51.18%, and 16.54% of patients had wild, heterozygous, mutant genotypes, respectively. For *Tyr[113]His* polymorphism, using the dominant model, we found a weak but significant increase in the risk of SQCC (AOR=1.44; 95% C.I.=1.02-2.03; p=**0.04**) for subjects carrying the *Tyr/His+His/His* genotype. Nonetheless, as indicated in Table 5.18, we found no significant association between the ADCC and SCLC subtypes of lung cancer and *Tyr[113]His* polymorphism. There was no association between any genotype and any histological subtypes, including SQCC, ADCC, and SCLC for the *His[139]Arg* polymorphism of the *EPHX1* genotype.

Table 5.18: Genotypic and allelic distribution of the *EPHX1exon3* and *exon4* genetic variant and its association with risk of lung cancer according to tumor histology

	SQCC				ADCC			SCLC		
Genotype *Tyr[113]His* (exon3)	**Controls (550) N(%)**	**Cases (216) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (200)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (127)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Co-dominant model										
Tyr/Tyr	215 (39.09)	69 (31.94)	1.00 (Reference)		75 (37.50)	1.0(Reference)		41 (32.28)	1.00(Reference)	
Tyr/His	246 (44.73)	106 (49.08)	1.41 (0.98-2.03)	0.06	92 (46.00)	1.03 (0.72-1.49)	0.86	65 (51.18)	1.40 (0.89-2.20)	0.15
His/His	89 (16.18)	41 (18.98)	1.51 (0.94-2.42)	0.08	33 (16.50)	1.00(0.61-1.64)	0.99	21 (16.54)	1.39 (0.76-2.54)	0.28
Dominant model										
Tyr/Tyr	215 (39.09)	69 (31.94)	1.00 (Reference)		75 (37.50)	1.0(Reference)		41 (32.28)	1.00 (Reference)	
Tyr/His+His/His	335 (60.91)	147 (68.06)	1.44 (1.02-2.03)	**0.04**	125 (62.50)	1.02(0.73-1.44)	0.89	86 (67.72)	1.40 (0.91-2.15)	0.12
Recessive model										
Tyr/Tyr+Tyr/His	461 (83.82)	175 (81.02)	1.00 (Reference)		167 (83.50)	1.0(Reference)		106 (83.46)	1.00 (Reference)	
His/His	89 (16.18)	41 (18.98)	1.25 (0.82-1.91)	0.30	33 (16.50)	1.01(0.65-1.58)	0.96	21 (16.54)	1.14 (0.67-1.96)	0.63
	SQCC				**ADCC**			**SCLC**		
Genotype *His[139]Arg*	**Controls (550)N(%)**	**Cases (216)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (200)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (127)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Co-dominant model										
His/His	380 (69.10)	148 (68.52)	1.0(Reference)		137 (68.50)	1.0(Reference)		85 (66.93)	1.00 (Reference)	
His/Arg	156 (28.36)	65 (30.09)	1.08 (0.76-1.54)	0.66	60 (30.00)	1.03 (0.71-1.48)	0.88	38 (29.92)	1.20 (0.77-1.88)	0.41
Arg/Arg	14 (2.54)	3 (1.39)	0.65 (0.18-2.36)	0.51	3 (1.50)	0.50 (0.14-1.82)	0.29	4 (3.15)	1.83 (0.55-6.02)	0.32
Dominant model										
His/His	380 (69.10)	148 (68.52)	1.0(Reference)		137 (68.50)	1.0(Reference)		85 (66.93)	1.00 (Reference)	
His/Arg + Arg/Arg	170 (30.90)	68 (31.48)	1.05 (0.74-1.49)	0.78	63 (31.50)	0.98 (0.69-1.41)	0.93	42 (33.07)	1.24 (0.81-1.91)	0.32
Recessive model										
His/His + His/Arg	536 (97.46)	213 (98.61)	1.0(Reference)		197 (98.50)	1.0(Reference)		123 (96.85)	1.00 (Reference)	
Arg/Arg	14 (2.54)	3 (1.39)	0.63 (0.17-2.30)	0.49	3 (1.50)	0.51 (0.14-1.86)	0.31	4 (3.15)	1.69 (0.52-5.48)	0.38

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.2.7. Effect of smoking status on association of *EPHX1* polymorphism with lung cancer risk

Table 5.19 shows the correlation between *EPHX1* polymorphic variants *Tyr*^{113}*His* and *His*^{139}*Arg* with smoking and its association with lung cancer. The sample population is divided into two subgroups: smokers and non-smokers. The smokers were further divided into heavy smokers and light smokers. The patients who had smoked less than or equal to 25 pack-years were grouped under light smokers, whereas more than 25 pack-years were grouped as heavy smokers. Our data showed controls who were smokers and carrying heterozygous genotype (*Tyr/His*) or mutant genotype (*His/His*) for *Tyr*^{113}*His* polymorphism were less represented in controls as compared to the cases (43.06% *vs.* 49.33% and 15.53% *vs.* 18.24%, respectively). Smokers harboring wild genotype were more prevalent in controls than in cases (41.41% *vs.* 32.4%). Lung cancer subjects who were smokers and carrying the heterozygous (*Tyr/His*) genotype for *EPHX1 Tyr*^{113}*His* gene were found to be associated towards risk for lung cancer (AOR=1.44; 95% C.I.=1.07-1.94; *p*=**0.017**) as compared to smokers having wild type genotype (*Tyr/Tyr*) in the co-dominant model. Smokers having combined genotype (*Tyr/His+His/His*) also showed a risk towards lung cancer compared with the reference genotype (AOR= 1.44; 95% C.I.= 1.09-1.90; p=**0.01**) using the dominant model. Further, stratified analysis by pack-years also suggested that light smokers harboring the homozygous mutant genotype (*His/His*) exhibited a high risk for lung cancer development (AOR=1.67, 95% C.I.=1.02-2.74, p=**0.04**). A similar trend was also visible in heavy smokers carrying the single variant allele for the *EPHX1 Tyr*^{113}*His* (*Tyr/His*) polymorphism. They were also associated with the risk of lung cancer (AOR=1.66, 95% C.I.=1.00-2.74, p=**0.04**) compared with the wild reference genotype (*Tyr/Tyr*). In the case of non-smokers, a non-significant protective effect was observed for lung cancer risk, as shown in Table 5.19.

In the case of *His*^{139}*Arg* polymorphism, no such association was observed between smoking status and lung cancer susceptibility with any genotypic combination, as shown in Table 5.19. Smokers with heterozygous genotype (*His/Arg*) were more pronounced in cases than controls (29.28% *vs*.27.76%). The *Arg/Arg* genotype was over expressed in controls than cases (2.12 % *vs.* 1.58%). A protective effect was observed for the mutant genotype (*Arg/Arg*; AOR= 0.80; 95% C.I.= 0.29-2.19) in the co-dominant model and as well as in recessive model also

Arg/Arg genotype showed protective effect (AOR= 0.79; 95% C.I.= 0.29-2.15) as compared to combined genotype (*His/Arg+Arg/Arg*) taken as reference. Among heavy smokers, the mutant genotype was more prevalent in controls when compared with cases (1.60% *vs.* 0.51%). Light smokers exhibited an over-representation of wild genotype (*His/His*) in controls as compared to cases (70.34% *vs.*68.15%), whereas heterozygous genotype was more expressed among cases than controls (29.43% *vs.* 27.33%). A non significant protective effect was observed in mutant genotype in cases who were heavy smokers *i.e. Arg/Arg* (AOR= 0.33; 95% C.I.=0.03-3.67), combined genotype (*His/Arg+Arg/Arg*) (AOR=0.98; 95% C.I.= 0.60-1.61). Using the recessive model, the mutant genotype for *His*^{139}*Arg* polymorphism who were heavy smokers also showed a protective effect (AOR= 0.33; 95% C.I.= 0.03-3.67) but was not significant. Light smokers did not exhibit any significant association with the risk for lung cancer. Non-smokers did not show any major significant risk for the disease. On the other hand, for the *EPHX1 His*^{139}*Arg* polymorphism, the stratification analysis by smoking status did not exhibit any association towards lung cancer susceptibility when any of the three genetic models were applied, as shown in Table 5.19. The results obtained suggest that the *EPHX1 Tyr*^{113}*His* polymorphism is a smoking-related risk factor, whereas; *His*^{139}*Arg* does not exhibit any association between smoking status and lung cancer risk.

Table 5.19: Relationship of different *EPHX1* genotypes with the smoking status of cases and controls

	SMOKER				NON-SMOKER			
Genotype ***Tyr^{113}His***	**Controls (425) N(%)**	**Cases (444)N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (125)N(%)**	**Cases (106)N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model								
Tyr/Tyr	176 (41.41)	144 (32.43)	1.0(Reference)		39 (31.20)	44 (41.51)	1.0(Reference)	
Tyr/His	183 (43.06)	219 (49.33)	1.44 (1.07-1.94)	**0.017**	63 (50.40)	47 (44.34)	0.63 (0.35-1.15)	0.13
His/His	66 (15.53)	81 (18.24)	1.44 (0.97-2.15)	0.07	23 (18.40)	15 (14.15)	0.58 (0.26-1.27)	0.17
Dominant model								
Tyr/Tyr	176 (41.41)	144 (32.43)	1.0(Reference)		39 (31.20)	44 (41.51)	1.0(Reference)	
Tyr/His+His/His	249 (58.59)	300 (67.57)	1.44 (1.09-1.90)	**0.01**	86 (68.80)	62 (58.49)	0.63 (0.36-1.09)	0.10
Recessive model								
Tyr/Tyr+Tyr/His	359 (84.47)	363 (81.76)	1.0(Reference)		102 (81.60)	91 (85.85)	1.0(Reference)	
His/His	66 (15.53)	81 (18.24)	1.18 (0.82-1.69)	0.38	23 (18.40)	15 (14.15)	0.74 (0.36-1.54)	0.42
Genotype ***His^{139}Arg***	**Controls (425) N(%)**	**Cases (444)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (125)N(%)**	**Cases (106)N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model								
His/His	298 (70.12)	307 (69.14)	1.0(Reference)		82 (65.60)	68 (64.15)	1.0(Reference)	
His/Arg	118 (27.76)	130 (29.28)	1.06 (0.79-1.44)	0.69	38 (30.40)	35 (33.02)	1.10 (0.62-1.97)	0.74
Arg/Arg	9 (2.12)	7 (1.58)	0.80 (0.29-2.19)	0.67	5 (4.00)	3 (2.83)	0.80 (0.18-3.56)	0.77
Dominant model								
His/His	298 (70.12)	307 (69.14)	1.0(Reference)		82 (65.60)	68 (64.15)	1.0(Reference)	
His/Arg + Arg/Arg	127 (29.88)	137 (30.86)	1.04 (0.78-1.40)	0.77	43 (34.40)	38 (35.85)	1.07 (0.61-1.86)	0.81
Recessive model								
His/His + His/Arg	416 (97.88)	437 (98.42)	1.0(Reference)		120 (96.00)	103 (97.17)	1.0(Reference)	
Arg/Arg	9 (2.12)	7 (1.58)	0.79 (0.29-2.15)	0.65	5 (4.00)	3 (2.83)	0.78 (0.18-3.48)	0.75
	LIGHT SMOKER				**HEAVY SMOKER**			
Genotype ***Tyr^{113}His***	**Controls (300) N(%)**	**Cases (248)N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (125)N(%)**	**Cases (196)N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model								
Tyr/Tyr	128 (42.67)	86 (34.68)	1.0(Reference)		48 (38.40)	58 (29.59)	1.0(Reference)	
Tyr/His	127 (42.33)	108 (43.55)	1.23 (0.84-1.80)	0.29	56 (44.80)	111 (56.63)	1.66 (1.00-2.74)	**0.049**
His/His	45 (15.00)	54 (21.77)	1.67 (1.02-2.74)	**0.04**	21 (16.80)	27 (13.78)	1.10 (0.54-2.21)	0.80
Dominant model								
Tyr/Tyr	128 (42.67)	86 (34.68)	1.0(Reference)		48 (38.40)	58 (29.59)	1.0(Reference)	
Tyr/His+His/His	172 (57.33)	162 (65.32)	1.34 (0.94-1.91)	0.10	77 (61.60)	138 (70.41)	1.51 (0.93-2.43)	0.09
Recessive model								

Tyr/Tyr+Tyr/His	255 (85.00)	194 (78.23)	1.0(Reference)		104 (83.20)	169 (86.22)	1.0(Reference)	
His/His	45 (15.00)	54 (21.77)	1.50 (0.95-2.35)	0.08	21 (16.80)	27 (13.78)	0.80 (0.43-1.49)	0.47
Genotype ***His[139]Arg***	**Controls (300) N(%)**	**Cases (248)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (125)N(%)**	**Cases (196)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Co-dominant model								
His/His	211 (70.34)	169 (68.15)	1.0(Reference)		87 (69.60)	138 (70.41)	1.0(Reference)	
His/Arg	82 (27.33)	73 (29.43)	1.08 (0.74-1.60)	0.68	36 (28.80)	57 (29.08)	1.02 (0.62-1.68)	0.94
Arg/Arg	7 (2.33)	6 (2.42)	1.15 (0.38-3.51)	0.80	2 (1.60)	1 (0.51)	0.33 (0.03-3.67)	0.36
Dominant model								
His/His	211 (70.34)	169 (68.15)	1.0(Reference)		87 (69.60)	138 (70.41)	1.0(Reference)	
His/Arg + Arg/Arg	89 (29.66)	79 (31.85)	1.09 (0.75-1.58)	0.65	38 (30.40)	58 (29.59)	0.98 (0.60-1.61)	0.95
Recessive model								
His/His + His/Arg	293 (97.67)	242 (97.58)	1.0(Reference)		123 (98.40)	195 (99.49)	1.0(Reference)	
Arg/Arg	7 (2.33)	6 (2.42)	1.13 (0.37-3.41)	0.83	2 (1.60)	1 (0.51)	0.33 (0.03-3.67)	0.36

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.2.8. Stratification analysis of *EPHX1* polymorphic variants with lung cancer risk based on gender

The stratification analysis was further carried out based on gender for both cases and controls, as shown in Table 5.20. In the case of *EPHX1 Tyr*^{113}*His* polymorphism, the females carrying heterozygous genotype (*Tyr/His*) are over presented in controls compared to cases (54.95% *vs*. 48.91%). At the same time, the frequency distribution of mutant genotype and wild genotype was higher in cases than controls (*His/His*: 17.39% *vs.* 15.38%, and *Arg/Arg*: 33.70% *vs*. 29.67%) in female subjects. In male subjects, the heterozygous genotype was presented in cases compared to controls (48.25% *vs*. 42.70%) for *EPHX1 Tyr*^{113}*His* polymorphism. In comparison, the wild genotype was over-expressed in controls than the cases (40.96% *vs*. 34.28%). It was observed that male subjects carrying the *Tyr/His* genotype exhibited a risk of lung cancer (AOR=1.36, 95% C.I.=1.02-1.82, p=**0.03**) compared to subjects carrying the wild type genotype (*Tyr/Tyr*). In contrast, homozygous variant genotype (*His/His)* did not exhibit a significant association towards lung cancer (AOR=1.29, 95% C.I.=0.88-1.89). Similarly, applying the dominant model, our results suggested that the *Tyr/His+His/His* genotype also exhibited an association towards risk for lung cancer (AOR=1.34, 95% C.I.=1.02-1.75, **p=0.03**). In contrast, the recessive model did not reveal any significant association in the case of male subjects for *EPHX1 Tyr*^{113}*His* polymorphism (Table 5.20). In females, *EPHX1 Tyr*^{113}*His* did not reveal an association towards risk for lung cancer in any of the models applied.

Further, for *EPHX1 His*^{139}*Arg* polymorphism in female subjects, the frequency of heterozygous genotype was higher amongst controls than cases (34.06% *vs.* 32.61%).In male subjects carrying the heterozygous genotype (*His/Arg*), the frequency distribution was higher in controls than in cases (34.06% *vs.* 32.61%). At the same time, the frequency of wild genotype and mutant genotype was higher in controls than in cases (*His/His*: 70.15% *vs*. 68.78%, and *Arg/Arg*: 2.61% *vs*. 1.75%). The male subjects who harbored mutant genotype exhibit a protective effect towards lung cancer risk (AOR=0.69; 95% C.I.=0.28-1.72; p=0.43), but it was not significant. Similarly, after stratification based on gender, none of the models exhibited any association for the *EPHX1 His*^{139}*Arg* polymorphism (Table 5.20).

Table 5.20: Relationship of different *EPHX1* polymorphisms based on gender

	FEMALE				MALE			
Genotype *Tyr^{113}His*	**Controls (91) N(%)**	**Cases (92) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (459) N(%)**	**Cases (458) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model								
Tyr/Tyr	27 (29.67)	31 (33.70)	1.0(Reference)		188 (40.96)	157 (34.28)	1.0(Reference)	
Tyr/His	50 (54.95)	45 (48.91)	0.69 (0.34-1.39)	0.30	196 (42.70)	221 (48.25)	1.36 (1.02-1.82)	**0.03**
His/His	14 (15.38)	16 (17.39)	0.70 (0.25-1.89)	0.48	75 (16.34)	80 (17.47)	1.29 (0.88-1.89)	0.19
Dominant model								
Tyr/Tyr	27 (29.67)	31 (33.70)	1.0(Reference)		188 (40.96)	157 (34.28)	1.0(Reference)	
Tyr/His+His/His	64 (70.33)	61 (66.30)	0.70 (0.36-1.38)	0.30	271 (59.04)	301 (65.72)	1.34 (1.02-1.75)	**0.033**
Recessive model								
Tyr/Tyr+Tyr/His	77 (84.62)	76 (82.61)	1.0(Reference)		384 (83.66)	378 (82.53)	1.0(Reference)	
His/His	14 (15.38)	16 (17.39)	0.90 (0.37-2.19)	0.81	75 (16.34)	80 (17.47)	1.07 (0.76-1.52)	0.69
Genotype *His^{139}Arg*	**Controls (91) N(%)**	**Cases (92) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (459) N(%)**	**Cases (458) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model								
His/His	58 (63.74)	60 (65.22)	1.0(Reference)		322 (70.15)	315 (68.78)	1.0(Reference)	
His/Arg	31 (34.06)	30 (32.61)	0.97 (0.49-1.90)	0.92	125 (27.24)	135 (29.47)	1.10 (0.82-1.47)	0.52
Arg/Arg	2 (2.20)	2 (2.17)	1.26 (0.17-9.52)	0.82	12 (2.61)	8 (1.75)	0.69 (0.28-1.72)	0.43
Dominant model								
His/His	58 (63.74)	60 (65.22)	1.0(Reference)		322 (70.15)	315 (68.78)	1.0(Reference)	
His/Arg + Arg/Arg	33 (36.26)	32 (34.78)	1.003 (0.52-1.94)	0.99	137 (29.85)	143 (31.22)	1.06 (0.80-1.41)	0.68
Recessive model								
His/His + His/Arg	89 (97.80)	90 (97.83)	1.0(Reference)		447 (97.39)	450 (98.25)	1.0(Reference)	
Arg/Arg	2 (2.20)	2 (2.17)	1.46 (0.20-10.78)	0.71	12 (2.61)	8 (1.75)	0.66 (0.27-1.63)	0.37

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, smoking. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.2.9. Combinatorial analysis of the risk associated with different polymorphic sites of *EPHX1* gene

To evaluate SNP to SNP interactions, the genotypic combinations of the various *EPHX1* genotypes were examined to find out if there is a significant difference in frequency between cases and controls. It comprises subjects carrying wild genotype for both the SNPs and heterozygous and mutant combined genotype for both the polymorphisms of *EPHX1* genotype. The genotypic frequency of these combinations and AOR and 95% C.I.s are tabulated in Table 5.21. For *EPHX1* polymorphic sites, we combined both the heterozygous and mutant genotypes as a single group. We compared it with the wild-type genotype as it was taken as the reference genotype. The patients possessing *(Tyr/His+His/His)+(His/Arg+Arg/Arg)* genotype did not exhibit a significant association towards lung cancer risk (AOR=1.36, 95% C.I.=0.92-1.99). Further, the stratification was done based on histology, smoking status, and gender to evaluate the risk of lung cancer. The results depict that the patients with *(Tyr/His+His/His)+(His/Arg+Arg/Arg)* genotype exhibited a risk of acquiring squamous cell lung cancer, but was marginally significant (AOR=1.66, 95% C.I.=0.99-2.79, p=**0.05**) and small cell lung cancer (AOR=1.85, 95% C.I.=0.98-3.49, p=**0.05**) but did not confer any modulation in the adenocarcinoma risk in subjects with *(Tyr/His+His/His)+(His/Arg+Arg/Arg)* genotype (AOR=0.98, 95% C.I.=0.56-1.71). In addition, the stratified analysis based on smoking status exhibited that the subjects who were smokers and carried the combined genotype *((Tyr/His+His/His)+(His/Arg+Arg/Arg))* exhibited a 1.67 fold-risk of developing lung cancer (AOR=1.67, 95% C.I.= 1.07-2.60, p=**0.02**). However, no such association was seen in case of light smokers (AOR=1.62, 95% C.I.= 0.91-2.88,), and heavy smokers (AOR=1.64, 95% C.I.=0.78-3.45, p=0.19) as shown in Table 5.21. In the case of subjects who were non-smokers conferred a decreased risk towards lung cancer (AOR=0.62, 95% C.I.=0.27-1.44, p=0.27), but it was not significant (Table 5.21).

Our study observed that the male subjects who carried the heterozygous and variant combined genotype for both *EPHX1* $Tyr^{113}His$ and $His^{139}Arg$ showed a risk towards lung cancer susceptibility (AOR=1.54, 95% C.I.=1.01-2.35, p=**0.04**). However, no such association was assessed between the female subjects possessing this genotypic combination and lung cancer susceptibility (Table 5.21).

Table 5.21: Combinatorial analysis of the risk associated with different polymorphic sites of *EPHX1* gene

	OVERALL				**ADCC**			
	Controls (215) N(%)	**Cases (219) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (215) N(%)**	**Cases (74) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Tyr/Tyr+His/His	130 (60.47)	116 (52.97)	1.00 (Reference)		130 (60.47)	43 (58.11)	1.00 (Reference)	
(Tyr/His+His/His)+(His/Arg+Arg/Arg)	85 (39.53)	103 (47.03)	1.36 (0.92-1.99)	0.12	85 (39.53)	31 (41.89)	0.98 (0.56-1.71)	0.94
	SQCC				**SCLC**			
	Controls (215) N(%)	**Cases (89) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (215) N(%)**	**Cases (53) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Tyr/Tyr+His/His	130 (60.47)	45 (50.56)	1.00 (Reference)		130 (60.47)	26 (49.06)	1.00 (Reference)	
(Tyr/His+His/His)+(His/Arg+Arg/Arg)	85 (39.53)	44 (49.44)	1.66 (0.99-2.79)	**0.05**	85 (39.53)	27 (50.94)	1.85 (0.98-3.49)	**0.05**
	SMOKERS				**NON-SMOKERS**			
	Controls (163) N(%)	**Cases (175) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (52) N(%)**	**Cases (44) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Tyr/Tyr+His/His	106 (65.03)	91 (52.00)	1.00 (Reference)		24 (46.15)	25 (56.82)	1.00 (Reference)	
(Tyr/His+His/His)+(His/Arg+Arg/Arg)	57 (34.97)	84 (48.00)	1.67 (1.07-2.60)	**0.02**	28 (53.85)	19 (43.18)	0.62 (0.27-1.44)	0.27
	LIGHT SMOKERS				**HEAVY SMOKERS**			
	Controls (117) N(%)	**Cases (91) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (46) N(%)**	**Cases (84) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Tyr/Tyr+His/His	78 (66.67)	49 (53.85)	1.00 (Reference)		28 (60.87)	42 (50.00)	1.00 (Reference)	
(Tyr/His+His/His)+(His/Arg+Arg/Arg)	39 (33.33)	42 (46.15)	1.62 (0.91-2.88)	0.10	18 (39.13)	42 (50.00)	1.64 (0.78-3.45)	0.19
	Males				**Females**			
	Controls (179) N(%)	**Cases (182) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (36) N(%)**	**Cases (37) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Tyr/Tyr+His/His	115 (64.25)	98 (53.85)	1.00 (Reference)		15 (41.67)	18 (48.65)	1.00 (Reference)	
(Tyr/His+His/His)+(His/Arg+Arg/Arg)	64 (35.75)	84 (46.15)	1.54 (1.01-2.35)	**0.04**	21 (58.33)	19 (51.35)	0.68 (0.25-1.85)	0.45

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Another method used to study the interaction based on risk allele was evaluating different combinations of risk allele made, as shown in Table 5.22. None of the alleles exhibit any risk towards LC compared to the 0 risk allele used as the reference (Table 5.22).

5.6.2.10. Variant Forms of *EPHX1* as Surrogate of Enzyme Activity

The population was divided into four different enzyme activity groups (very low, low, intermediate, and high) according to the gene polymorphisms in *Tyr*^{113}*His* and *His*^{139}*Arg*, as shown in Table 5.23. Very low (VL) consisted of individuals with *His/His* in both *Tyr*^{113}*His* and *His*^{139}*Arg*. The low (L) activity genotype group consisted of *His/His Tyr*^{113}*His* and either *His/Arg* or *Arg/Arg His*^{139}*Arg* or *Tyr/His Tyr*^{113}*His* and *His/His His*^{139}*Arg*. The intermediate (I) activity genotype group consisted of *Tyr/His* and *His/Arg* or *Tyr/Tyr* and *His/His*. The high (H) activity genotype group consisted of individuals with *Tyr/Tyr* in *Tyr*^{113}*His* and either *His/Arg* or *Arg/Arg* in *His*^{139}*Arg* or *Tyr/His* in *Tyr*^{113}*His* and *Arg/Arg* in *His*^{139}*Arg*. The results show no relationship between *EPHX1* genotypes (based on *EPHX1* activity) and the lung cancer risk. Cases were stratified according to histological types, no association was found. Similarly, the gene-environment interaction (*i.e.*, based on smoking status) was also evaluated, and the results exhibit no significant association between *EPHX1* genotypes and lung cancer risk (Table 5.23).

Table 5.22: Combined effects of *Tyr*^{113}*His* (rs2854455) and *His*^{139}*Arg* (rs2234922) on the risk of lung cancer

OVERALL					ADCC			
Number of Risk alleles$	**Control (550), N(%)**	**Cases (550), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (550), N(%)**	**Cases (200), N(%)**	**AOR (95% C.I.)**	**p-value**
0	130 (23.64)	116 (21.10)	1.0 (Reference)		130 (23.64)	43 (21.50)	1.0 (Reference)	
1	260 (47.27)	251 (45.64)	1.07 (0.79-1.45)	0.66	260 (47.27)	97 (48.50)	1.10 (0.72-1.70)	0.65
2	132 (24.00)	157 (28.54)	1.32 (0.94-1.86)	0.11	132 (24.00)	53 (26.50)	1.16 (0.71-1.87)	0.55
3	28 (5.09)	26 (4.72)	1.05 (0.58-1.90)	0.88	28 (5.09)	7 (3.50)	0.57 (0.22-1.47)	0.25
SQCC					**SCLC**			
Number of Risk alleles$	**Control (550), N(%)**	**Cases (216), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (550), N(%)**	**Cases (127), N(%)**	**AOR (95% C.I.)**	**p-value**
0	130 (23.64)	45 (20.83)	1.0 (Reference)		130 (23.64)	26 (20.47)	1.0 (Reference)	
1	260 (47.27)	94 (43.52)	1.05 (0.69-1.60)	0.82	260 (47.27)	57 (44.88)	1.09 (0.64-1.89)	0.74
2	132 (24.00)	66 (30.56)	1.50 (0.94-2.38)	0.08	132 (24.00)	36 (28.35)	1.50 (0.84-2.67)	0.17
3	28 (5.09)	11 (5.09)	1.36 (0.61-3.04)	0.46	28 (5.09)	8 (6.30)	1.88 (0.72-4.93)	0.20
SMOKERS					**NON-SMOKERS**			
Number of Risk alleles$	**Control (425), N(%)**	**Cases (444), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (125), N(%)**	**Cases (106), N(%)**	**AOR (95% C.I.)**	**p-value**
0	106 (24.94)	91 (20.50)	1.0 (Reference)		24 (19.20)	25 (23.59)	1.0 (Reference)	
1	203 (47.76)	203 (45.72)	1.15 (0.81-1.62)	0.42	57 (45.60)	48 (45.28)	0.76 (0.37-1.56)	0.45
2	100 (23.53)	128 (28.83)	1.45 (0.98-2.13)	0.06	32 (25.60)	29 (27.36)	0.92 (0.43-1.97)	0.83
3	16 (3.76)	22 (4.95)	1.55 (0.76-3.15)	0.23	12 (9.60)	4 (3.77)	0.32 (0.09-1.15)	0.08
LIGHT-SMOKERS					**HEAVY-SMOKERS**			
Number of Risk alleles$	**Control (300), N(%)**	**Cases (248), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (125), N(%)**	**Cases (196), N(%)**	**AOR (95% C.I.)**	**p-value**
0	78 (26.00)	49 (19.76)	1.0 (Reference)		28 (22.40)	42 (21.43)	1.0 (Reference)	
1	141 (47.00)	112 (45.16)	1.25 (0.80-1.95)	0.31	62 (49.60)	91 (46.43)	0.97 (0.54-1.74)	0.92
2	71 (23.67)	72 (29.03)	1.47 (0.90-2.42)	0.12	29 (23.20)	56 (28.57)	1.36 (0.70-2.63)	0.36
3	10 (3.33)	15 (6.05)	2.39 (0.99-5.81)	0.05	6 (4.80)	7 (3.57)	0.74 (0.21-2.56)	0.63
MALES					**FEMALES**			
Number of Risk alleles$	**Control (459), N(%)**	**Cases (458), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (91), N(%)**	**Cases (92), N(%)**	**AOR (95% C.I.)**	**p-value**
0	115 (25.05)	98 (21.40)	1.0 (Reference)		15 (16.48)	18 (19.57)	1.0 (Reference)	
1	215 (46.84)	209 (45.63)	1.14 (0.82-1.59)	0.42	45 (49.45)	42 (45.65)	0.72 (0.30-1.69)	0.44
2	107 (23.31)	130 (28.38)	1.44 (0.99-2.09)	0.05	25 (27.47)	27 (29.35)	0.79 (0.30-2.05)	0.63

3	22 (4.79)	21 (4.59)	1.14 (0.59-2.21)	0.69	6 (6.59)	5 (5.43)	0.63 (0.15-2.73)	0.54

[$] 0 risk allele: Tyr/Tyr + His/His; 1 risk allele: Tyr/Tyr + His/Arg, Tyr/His + His/His; 2 risk allele: Tyr/Tyr + Arg/Arg, His/His + His/His, Tyr/His + His/Arg; 3 risk allele: Tyr/His + Arg/Arg, His/His + His/Arg

Table 5.23: Genotypic and allelic distribution of the *EPHX1* genetic variant and its association with risk of lung cancer as Surrogate of Enzyme Activity

	OVERALL				**ADCC**			
	Controls (550) N(%)	**Cases (550) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (550) N(%)**	**Cases (200) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Very low (VL)	64 (11.64)	73 (13.27)	1.00 (Reference)		64 (11.64)	27 (13.50)	1.00 (Reference)	
Low (L)	211 (38.36)	209 (38.00)	0.87 (0.59-1.28)	0.48	211 (38.36)	73 (36.50)	0.80 (0.47-1.36)	0.41
Intermediate (I)	187 (34.00)	193 (35.09)	0.90 (0.61-1.34)	0.61	187 (34.00)	67 (33.50)	0.84 (0.49-1.45)	0.54
High (H)	88 (16.00)	75 (13.64)	0.73 (0.46-1.16)	0.19	88 (16.00)	33 (16.50)	0.87 (0.46-1.62)	0.66
Combined (L+I+H)	486 (88.36)	477 (86.73)	0.86 (0.60-1.23)	0.41	486 (88.36)	173 (86.50)	0.82 (0.50-1.34)	0.43
	SQCC				**SCLC**			
	Controls (550) N(%)	**Cases (216) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (550) N(%)**	**Cases (127) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Very low (VL)	64 (11.64)	30 (13.89)	1.00 (Reference)		64 (11.64)	15 (11.81)	1.00 (Reference)	
Low (L)	211 (38.36)	84 (38.89)	0.88 (0.52-1.48)	0.63	211 (38.36)	50 (39.37)	0.88 (0.45-1.72)	0.71
Intermediate (I)	187 (34.00)	78 (36.11)	0.89 (0.53-1.50)	0.66	187 (34.00)	45 (35.43)	0.97 (0.50-1.87)	0.92
High (H)	88 (16.00)	24 (11.11)	0.58 (0.30-1.10)	0.10	88 (16.00)	17 (13.39)	0.77 (0.35-1.69)	0.51
Combined (L+I+H)	486 (88.36)	186 (86.11)	0.82 (0.51-1.33)	0.43	486 (88.36)	112 (88.19)	0.89 (0.48-1.65)	0.72
	SMOKERS				**NON-SMOKERS**			
	Controls (425) N(%)	**Cases (444) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (125) N(%)**	**Cases (106) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Very low (VL)	50 (11.76)	61 (13.74)	1.00 (Reference)		14 (11.20)	12 (11.32)	1.00 (Reference)	
Low (L)	158 (37.18)	175 (39.41)	0.93 (0.60-1.44)	0.74	53 (42.40)	34 (32.08)	0.67 (0.27-1.69)	0.40
Intermediate (I)	147 (34.59)	153 (34.46)	0.88 (0.56-1.37)	0.57	40 (32.00)	40 (37.73)	1.20 (0.48-2.98)	0.69
High (H)	70 (16.47)	55 (12.39)	0.66 (0.39-1.12)	0.12	18 (14.40)	20 (18.87)	1.36 (0.49-3.77)	0.56
Combined (L+I+H)	375 (88.24)	383 (86.26)	0.86 (0.57-1.29)	0.47	111 (88.80)	94 (88.68)	0.95 (0.41-2.20)	0.90
	LIGHT SMOKERS				**HEAVY SMOKERS**			
	Controls (300) N(%)	**Cases (248) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Controls (125) N(%)**	**Cases (196) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Very low (VL)	35 (11.67)	40 (16.13)	1.00 (Reference)		15 (12.00)	21 (10.71)	1.00 (Reference)	
Low (L)	108 (36.00)	94 (37.90)	0.83 (0.48-1.45)	0.52	50 (40.00)	81 (41.33)	1.12 (0.52-2.38)	0.77
Intermediate (I)	107 (35.67)	76 (30.65)	0.65 (0.37-1.14)	0.13	40 (32.00)	77 (39.29)	1.35 (0.62-2.90)	0.45
High (H)	50 (16.66)	38 (15.32)	0.71 (0.38-1.36)	0.30	20 (16.00)	17 (8.67)	0.56 (0.22-1.45)	0.23
Combined (L+I+H)	265 (88.33)	208 (83.87)	0.74 (0.45-1.23)	0.24	110 (88.00)	175 (89.29)	1.11 (0.54-2.25)	0.78

	Males				Females			
	Controls (459) N(%)	**Cases (458) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (91) N(%)**	**Cases (92) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Very low (VL)	55 (11.98)	61 (13.32)	1.00 (Reference)		9 (9.89)	12 (13.04)	1.00 (Reference)	
Low (L)	172 (37.47)	175 (38.21)	0.91 (0.60-1.39)	0.67	39 (42.86)	34 (36.96)	0.79 (0.25-2.44)	0.68
Intermediate (I)	157 (34.20)	161 (35.15)	0.92 (0.60-1.41)	0.71	30 (32.97)	32 (34.78)	1.10 (0.36-3.41)	0.86
High (H)	75 (16.34)	61 (13.32)	0.71 (0.43-1.17)	0.17	13 (14.28)	14 (15.22)	1.44 (0.37-5.56)	0.60
Combined (L+I+H)	404 (88.02)	397 (86.68)	0.88 (0.60-1.30)	0.53	82 (90.11)	80 (86.96)	0.94 (0.33-2.66)	0.90

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.2.11. Haplotype analysis of *EPHX1* variants

For the current study, we analyzed haplotypes using the SHEsis program platform (Table 5.24). For two SNPs, linkage disequilibrium and haplotype frequencies were analyzed for two SNPs (*Tyr*^{113}*His* and *His*^{139}*Arg*). Both the SNPs were in weak linkage disequilibrium with each other in this study population. *EPHX1 Tyr*^{113}*His* and *EPHX1 His*^{139}*Arg* (D =0.342, r^2 = 0.016) (Figure 5.19). There was no statistically significant difference in the overall haplotype distribution between cases and controls (global $\chi 2$= 2.26; p= 0.52). We considered the individuals with -*Tyr*^{113}-*His*^{139} haplotype as the reference group for OR estimations. The analysis gave four haplotypes *viz*. -His^{113}-His^{139}, -His^{113}-Arg^{139}, -Tyr^{113}-His^{139}, and -Tyr^{113}-Arg^{139} with an OR of 1.13 (95% C.I.=0.95-1.34, p=0.18), 1.09 (95% C.I.=0.73-1.64, p=0.68), 0.89 (95% C.I.=0.75-1.06, p=0.18), and 0.97(95% C.I.=0.76-1.26, p=0.84), respectively. All the exploratory haplotypes did not significantly affect lung cancer susceptibility, as depicted in Table 5.24.

Table 5.24: Haplotype frequencies and lung cancer risk associated with *EPHX1* haplotypes

Haplotype	**Case frequency (%)**	**Control frequency (%)**	**OR (95% C.I.)**[$]	**p-value**[#]
EPHX1 Tyr^{113}*His* and *EPHX1 His*^{139}*Arg*				
-His^{113}-His^{139}	407.38 (0.37)	377.39 (0.34)	1.13 (0.95-1.34)	0.18
-His^{113}-Arg^{139}	50.62 (0.05)	46.61 (0.04)	1.09 (0.73-1.64)	0.68
-Tyr^{113}-His^{139}	507.62 (0.46)	538.61 (0.49)	0.89 (0.75-1.06)	0.18
-Tyr^{113}-Arg^{139}	134.38 (0.12)	137.39 (0.13)	0.97 (0.76-1.26)	0.84

[$]ORs, 95% C.I.s, and their corresponding p-values were calculated by logistic regression analysis.

[#] Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls.

OR: Odds ratio.

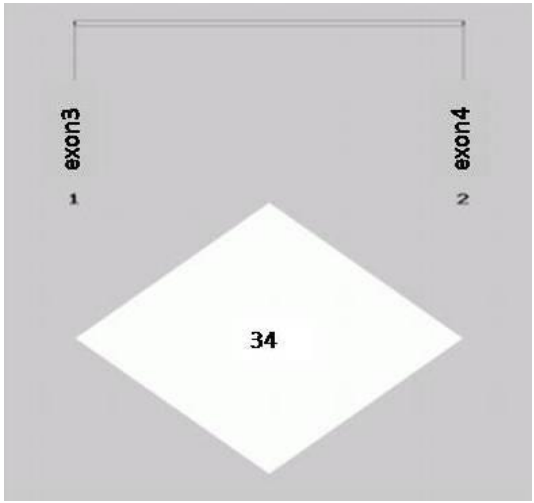


Figure 5.19: Showing pairwise linkage disequilibrium between the two variants of *EPHX1* polymorphism

5.6.2.12. Relationship between *EPHX1* Genotypic variants and Clinico-pathological Characteristics

To determine the association between the two polymorphic variants of *EPHX1* (*Tyr*^{113}*His*, *His*^{139}*Arg*) and clinic-pathological parameters, we divided patients based on the stage of cancer (III *vs*. IV), primary tumor extension (T3 *vs*. T4), lymph node invasion (Nx+N0+N1 *vs*. N2+N3+N4), and metastatic status (positive *vs*. negative) (Table 5.25).

To investigate the relationship of different genotypes of *EPHX1* polymorphic variants, the genotypic frequencies were calculated for each polymorphic site for both stage III and stage IV patients, as shown in Table 5.25. The genotypic frequencies were compared to determine whether any significant difference lies in the genotypic frequencies among the two subgroups made based on stage. Overall, none of the polymorphic sites depicted any significant difference in the genotypic frequencies among the two subgroups, thus, the lack of evident association between these genotypes and the tumor stage. This can be seen in *Tyr*^{113}*His*, where the genotypic frequencies of the subjects carrying variant genotype (*His/His*) were 16.16% in stage III and 18.62% among the stage IV category. Similarly, the genotypic frequencies of heterozygotes in cases were 51.09% among stage III and 46.96% in stage IV patients. No significant AOR was observed for both heterozygous (AOR=0.93, 95% C.I.=0.61-1.43,) and mutant (AOR= 1.22, 95% C.I.=0.70-2.13) genotypes for *Tyr*^{113}*His* polymorphism. Further, *His*^{139}*Arg* genetic variants were evaluated, and there was no association with tumor stage that can be established. The wild-type genotype was more represented in stage III (70.74%) than stage IV (66.80%), while heterozygous genotype was expressed more in stage IV (31.18%) than stage IV (27.08%) patients.

The next clinical parameter which has been accounted for consideration is the primary tumor extension. The patients were divided into two categories based on tumor size, as shown in Table 5.25. A higher proportion of lung cancer patients were reported in the T4 category (299), whereas the T3 subgroup consists of 115 patients. The genotypic frequencies were estimated for each polymorphic site and it was observed that *EPHX1 Tyr*^{113}*His* heterozygous genotype (*Tyr/His*) was represented in 50.43% patients with T3 tumor as compared to 48.49% patients with a T4 tumor. It was also observed that the frequency of mutant genotype (*His/His*) in subjects with T3 extension was almost similar to the lung cancer cases with T4 extension (16.52% *vs*. 17.06%), indicating no correlation with tumor extension. None of the genotypes of

EPHX1 Tyr^{113}*His* polymorphism showed a correlation towards the risk of lung cancer. In the *EPHX1 His*^{139}*Arg*, the percentage of a heterozygous genotype was less in the T3 subgroup than the T4 tumor subgroup (26.09% *vs*. 29.77%). The mutant genotype was over expressed in the T4 subgroup than in patients with the T3 tumor subgroup (3.01% *vs*. 0.87%). At the same time, wild genotype was highly prevalent in the T3 subgroup compared with the T4 subgroup (73.04% *vs*. 67.22%). A 3.6 fold tumor extension (T4) for subjects carrying variant genotype was observed compared to patients with T3 tumor size (AOR=3.69; 95% C.I.= 0.46-29.76) for the *EPHX1 His*^{139}*Arg* polymorphism but was not significant. Similarly, in the recessive model, a non-significant 3.3 fold tumor extension (T4) was observed for subjects carrying variant genotype compared to patients with T3 tumor size (AOR= 3.34, 95% C.I.= 0.42-26.87). It was observed that the percentage of heterozygotes (*His/Arg*) is lesser in the T3 subgroup as compared to the T4 tumor subgroup (26.09% *vs*.29.77%). A non-significant association was observed for subjects carrying the heterozygous genotype (AOR=1.27, 95% C.I.= 0.78-2.07) (Table 5.25).

Table 5.25 also demonstrates the relationship of these different genotypes with the metastatic status of the patients. These were further classified based on metastatic status into M0 and M1groups. Patients showing positive distant metastasis (M1) were compared with those lacking metastasis (M0), also regarded as the reference group. Among *Tyr*^{113}*His* genotypic variants (*His/His*), 18.70% of patients showed metastasis, while 16.01% did not show any metastasis. The frequency of subjects harboring wild genotype (*Tyr/Tyr*) was 33.33% for distant metastasis and 35.16% for subjects lacking metastasis. The combined variants (*Tyr/His+His/His*) illustrated a higher frequency for the metastatic group than the non-metastatic group (66.67% vs. 64.84%). Since there was no significant difference in the frequency distribution among the two above discussed genotypic groups, it was evident that this polymorphic site did not show any relation with metastasis or spread of lung cancer to other organs. In *EPHX1 His*^{139}*Arg* polymorphism, the percentage of subjects carrying homozygous wild genotype (*His/His*) and lacking metastasis were 69.14%, while 67.07% of lung cancer patients had distant metastasis. The representation of double variant allele (*Arg/Arg*) was almost near in both M1 (1.63%) and M0 (2.34%) groups. The frequency of combined variants (*His/Arg+Arg/Arg*) was 32.93% for patients with distant metastasis and 30.86% for those lacking metastasis. The patients with variant genotype for *His*^{139}*Arg* polymorphism showed a protective effect of lung cancer metastasis to other organs when patients who did not show metastasis were considered as

reference (co-dominant model: AOR=0.58; 95% C.I.=0.15-2.23; recessive model: AOR= 0.53, 95% C.I.= 0.13-2.09).

Further, Table 5.25 also demonstrates the relationship of these different genotypes with the lymph node invasion of the patients. These were classified based on lymph node invasion into Nx+N0+N1 and N2+N3+N4 groups. Patients with the Nx+N0+N1 tumor group were compared with the N2+N3+N4 group, also regarded as the reference group. For the *Tyr*^{113}*His* polymorphism, the subjects carrying heterozygous genotype (*Tyr/His*) were represented more amongst the N2+N3+N4 tumor group than patients with the Nx+N0+N1 tumor group (48.97% *vs*. 45.45%). Similarly, the variant genotype (*His/His*) also exhibits a higher percentage in the N2+N3+N4 group than the Nx+N0+N1 tumor group (18.72% *vs*. 12.73%). The lung cancer patients who were carrying both the variant alleles (*His/His*) for the *EPHX1 Tyr*^{113}*His* polymorphism exhibited a two-fold increased risk of lymph node metastasis towards lung cancer as compared to subjects who were homozygous for the wild type genotype (*Tyr/Tyr*) of the *EPHX1*, especially under the co-dominant model (AOR=1.94; 95% C.I.=0.90-3.80; p=**0.04**) (Table 5.25). Among *His*^{139}*Arg* polymorphism, the subjects carrying variant genotype (*Arg/Arg*), there were 1.79% patients in the N2+N3+N4 group while there were 2.73% patients among the Nx+N0+N1 tumor group. The frequency of subjects harboring wild genotype (*His/His*) was 68.72% for the N2+N3+N4 group, while there were 65.45% patients for the Nx+N0+N1 tumor group. The combined variants (*His/Arg+Arg/Arg*) illustrated a higher frequency for the Nx+N0+N1 tumor group than the N2+N3+N4 group (34.55% *vs*.31.28%). The heterozygous (*His/Arg*) as well as mutant genotype (*Arg/Arg*) showed a trend of decreased tumor progression (*His/Arg*: AOR=0.89, 95% C.I.=0.56-1.41; *Arg/Arg*: AOR=0.60, 95% C.I.=0.15-2.38), but was not significant. Thus, regarding clinical stage, tumor size, and distant metastasis, no relationship between the *Tyr*^{113}*His* and *His*^{139}*Arg* polymorphism was observed in patients in either of the three applied genetic models, as shown in Table 5.25.

Table 5.25: Relationship of different genotypes with the clinical-pathological parameters

	CLINICAL STAGE				PRIMARY TUMOR EXTENSION			
	III	**IV**			**T3**	**T4**		
Genotype *Tyr[113]His*	**N=229 (%)**	**N=247 (%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**N=115 (%)**	**N=299 (%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model								
Tyr/Tyr	75 (32.75)	85 (34.41)	1.0(Reference)		38 (33.04)	103 (34.45)	1.0(Reference)	
Tyr/His	117 (51.09)	116 (46.96)	0.93 (0.61-1.43)	0.76	58 (50.43)	145 (48.49)	0.90 (0.56-1.47)	0.69
His/His	37 (16.16)	46 (18.62)	1.22 (0.70-2.13)	0.48	19 (16.52)	51 (17.06)	0.81 (0.41-1.59)	0.54
Dominant model								
Tyr/Tyr	75 (32.75)	85 (34.41)	1.0(Reference)		38 (33.04)	103 (34.45)	1.0(Reference)	
Tyr/His+His/His	154 (67.25)	162 (65.59)	1.01 (0.67-1.50)	0.97	77 (66.96)	196 (65.55)	0.92 (0.58-1.45)	0.72
Recessive model								
Tyr/Tyr+Tyr/His	192 (83.84)	201 (81.38)	1.0(Reference)		96 (83.48)	248 (82.94)	1.0(Reference)	
His/His	37 (16.16)	46 (18.62)	1.23 (0.75-2.01)	0.41	19 (16.52)	51 (17.06)	0.98 (0.55-1.76)	0.95
Genotype *His[139]Arg*	**N=229 (%)**	**N=247 (%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**N=115 (%)**	**N=299 (%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Co-dominant model								
His/His	162 (70.74)	165 (66.80)	1.0(Reference)		84 (73.04)	201 (67.22)	1.0(Reference)	
His/Arg	62 (27.08)	77 (31.18)	1.18 (0.78-1.79)	0.43	30 (26.09)	89 (29.77)	1.27 (0.78-2.07)	0.34
Arg/Arg	5 (2.18)	5 (2.02)	0.81 (0.21-3.02)	0.75	1 (0.87)	9 (3.01)	3.69 (0.46-29.76)	0.22
Dominant model								
His/His	162 (70.74)	165 (66.80)	1.0(Reference)		84 (73.04)	201 (67.22)	1.0(Reference)	
His/Arg + Arg/Arg	67 (29.26)	82 (33.20)	1.15 (0.77-1.73)	0.50	31 (26.96)	98 (32.78)	1.35 (0.83-2.18)	0.22
Recessive model								
His/His + His/Arg	224 (97.82)	242 (97.98)	1.0(Reference)		114 (99.13)	290 (96.99)	1.0(Reference)	
Arg/Arg	5 (2.18)	5 (2.02)	0.73 (0.19-2.82)	0.65	1 (0.87)	9 (3.01)	3.34 (0.42-26.87)	0.26
	LYMPH NODE INVASION				**METASTASIS**			
	Nx+N0+N1	**N2+N3+N4**			**No (M0)**	**Yes (M1)**		
Genotype *Tyr[113]His*	**N=110 (%)**	**N=390 (%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**N=256 (%)**	**N=246 (%)**	**AOR (95% C.I.) [a]**	***p*[b]**
Co-dominant model								
Tyr/Tyr	46 (41.82)	126 (32.31)	1.0(Reference)		90 (35.16)	82 (33.33)	1.0(Reference)	
Tyr/His	50 (45.45)	191 (48.97)	1.40 (0.88-2.22)	0.15	125 (48.83)	118 (47.97)	1.14 (0.76-1.72)	0.52
His/His	14 (12.73)	73 (18.72)	1.94 (0.99-3.80)	**0.04**	41 (16.01)	46 (18.70)	1.41 (0.82-2.43)	0.21
Dominant model								
Tyr/Tyr	46 (41.82)	126 (32.31)	1.0(Reference)		90 (35.16)	82 (33.33)	1.0(Reference)	
Tyr/His+His/His	64 (58.18)	264 (67.69)	1.51 (0.98-2.34)	0.06	166 (64.84)	164 (66.67)	1.22 (0.83-1.79)	0.32
Recessive model								

Tyr/Tyr+Tyr/His	96 (87.27)	317 (81.28)	1.0(Reference)		215 (83.99)	200 (81.30)	1.0(Reference)	
His/His	14 (12.73)	73 (18.72)	1.57 (0.85-2.91)	0.15	41 (16.01)	46 (18.70)	1.27 (0.78-2.05)	0.33
Genotype ***His[139]Arg***	**N=110 (%)**	**N=390 (%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**N=256 (%)**	**N=246 (%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Co-dominant model								
His/His	72 (65.45)	268 (68.72)	1.0(Reference)		177 (69.14)	165 (67.07)	1.0(Reference)	
His/Arg	35 (31.82)	115 (29.49)	0.89 (0.56-1.41)	0.63	73 (28.52)	77 (31.30)	1.10 (0.73-1.64)	0.65
Arg/Arg	3 (2.73)	7 (1.79)	0.60 (0.15-2.38)	0.46	6 (2.34)	4 (1.63)	0.58 (0.15-2.23)	0.43
Dominant model								
His/His	72 (65.45)	268 (68.72)	1.0(Reference)		177 (69.14)	165 (67.07)	1.0(Reference)	
His/Arg + Arg/Arg	38 (34.55)	122 (31.28)	0.87 (0.55-1.36)	0.53	79 (30.86)	81 (32.93)	1.05 (0.71-1.55)	0.80
Recessive model								
His/His + His/Arg	107 (97.27)	383 (98.21)	1.0(Reference)		250 (97.66)	242 (98.37)	1.0(Reference)	
Arg/Arg	3 (2.73)	7 (1.79)	0.63 (0.16-2.50)	0.51	6 (2.34)	4 (1.63)	0.53 (0.13-2.09)	0.36

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender, smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.2.13. Multiple dimensionality reduction analysis of genetic variants within *EPHX1* gene

Non-parametric analysis was carried out using the MDR approach to find out the interaction of the *EPHX1* variants and their role in predicting the lung cancer predisposition in north Indians (Table 5.26). This approach reveals the possible interaction models comprising the genetic variants and environmental exposure, which is smoking in this case. The present analysis was carried out taking two variants of the *EPHX1* gene in predicting lung cancer susceptibility. The analysis predicted two interaction models, as summarized in Table 5.26. A minimum prediction error and higher CVC are preferred for the best results. CVC of 10/10 was observed for the model, which included *EPHX1 Tyr*^{113}*His* and *EPHX1 His*^{139}*Arg* with a prediction error of 0.42 (p=0.86), but was not significant.

Further, the interaction was also visualized using an interaction dendrogram between *EPHX1 Tyr*^{113}*His* and *EPHX1 His*^{139}*Arg* and lung cancer risk. The length of the lines connecting these variants signifies the intensity of interaction. The colors of the lines connecting these SNPs indicate the type of interaction taking place between them. A weak interaction was observed between the *EPHX1* polymorphic variants (*EPHX1 Tyr*^{113}*His* and *EPHX1 His*^{139}*Arg*) as these are connected *via* blue line (Figure 5.20).

Table 5.26: Multifactor dimensionality reduction analysis showing interactions of *EPHX1* variants with lung cancer risk

Number of risk factors	Best model	Cross-validation consistency	Prediction error	p-value
1	*EPHX1 Tyr*^{113}*His*	10/10	0.48	0.72
2	*EPHX1 Tyr*^{113}*His, EPHX1 His*^{139}*Arg*	10/10	0.42	0.86



Figure 5.20: MDR interactions with *EPHX1* variants; X1: *EPHX1 Tyr*^{113}*His*; X2: *EPHX1 His*^{139}*Arg*

5.6.3. Statistical analysis for polymorphic sites of *NAT2* gene

5.6.3.5. Genetic model analysis between *NAT2* variants and lung cancer risk

The genotypic and allelic frequencies of the four different variants of the *NAT2* gene have been tabulated in Table 5.27. The four polymorphic sites which have been evaluated in this study include *NAT2*5B* (rs1799929, *481C>T*), *NAT2*5C* (rs1208, *803A>G*), *NAT2*6* (rs1799930, *590G>A*), and *NAT2*7* (rs1799931, *857G>A*). Genotypic and allelic frequencies have been illustrated for each variant, along with minor allele frequencies in both the study groups. The observed genotype frequencies of *NAT2* polymorphism in both controls and cases in the present study were consistent with the Hardy–Weinberg equilibrium (*590G>A; NAT2*6*, χ^2*=1.38,* p=0.24 and χ^2=0.21, p=0.64, *857G>A (NAT2*7)*; χ*2*=2.26, p=0.13 and χ^2=2.88, p=0.09, *481C>T (NAT2*5B)*, χ^2=0.14, p=0.70 and χ*2*=0.002, 0.96 and for *803A>G (NAT2*5C)*, χ^2=3.35 p=0.07 and χ^2=15.58, p=**0.0008**).

The first variant of the *NAT2* gene evaluated was *590G>A*. The genotypic frequencies of cases and controls were compared, and it was found that there was no significant difference between the two study groups (χ^2=2.26, df=2) as shown in Table 5.27. The proportions of subjects carrying the variant (*NAT2*6/*6*) genotype were 12.18% in cases and 15.27% in controls. Also, the frequency of *NAT2*4/*6+*6/*6* carrying subjects was almost similar in cases (58.91%) and controls (59.82%). MAF was comparable in cases (0.35) and controls (0.37). Association analysis was carried out using *NAT2*4/*4* (wild) as a reference genotype, and it was visualized that neither of the variant genotypes, *i.e., NAT2*4/*6* (AOR=1.04, 95% C.I.=0.81-1.35) and *NAT2*6/*6* (AOR=0.78, 95% C.I.=0.54-1.14), showed any significant role in modulating the risk towards lung cancer as shown in Table 5.27. The three different genetic models used exhibit no association of *590G>A* polymorphism towards lung cancer susceptibility.

For *NAT2 857G>A* (*NAT2*7*), the proportion of subjects with wild (*NAT2*4/*4*) genotype were 48.00% in controls and 40.36% in cases, as shown in Table 5.27. The percentage of subjects carrying heterozygotes (*NAT2*4/*7*) was more in cases (48.91%) as compared to controls (40.55%). The frequency of variant genotype was almost the same in both controls and cases (11.45% and 10.73%). However, there was a significant difference between the genotypic frequencies among cases and controls (χ^2=8.06, df=2, p=**0.01**). MAF was found to be 0.32 in controls and 0.35 in cases. The individuals carrying the heterozygous (*NAT2*4/*7*) genotype

exhibit a marginal risk towards development for lung cancer (AOR= 1.43, 95% C.I.=1.11-1.85, p=**0.005**) when the co-dominant model was applied. Similarly, when the dominant model for the combined genotype (*i.e., NAT2 *4/*7 + *7/*7*) of the *857G>A (NAT2*7)* polymorphism was evaluated, here too, our data also indicated a marginal risk for lung cancer (AOR=1.36; 95% C.I.= 1.07-1.73; **p**=**0.01**).

Another variant of *NAT2*, namely *481C>T* (*NAT2*5B*) polymorphism, was also analyzed in the present study, shown in Table 5.27. The proportion of subjects with wild *(NAT2*4/*4*) genotype were considerably over-represented in cases (43.27%) than controls (33.45%). The frequency of heterozygous genotype was 45.09% in cases and 49.45% in controls. The percentage of mutant genotype *(NAT2*5B/*5B*) was higher in controls than in cases (17.10% *vs.* 11.64%). A significant difference was observed between the genotypic distribution of cases and controls (χ^2=13.71, df=2, p=**0.001**). The minor allele frequency (MAF) of 0.42 was observed in controls and 0.34 in cases. When odds ratio was calculated using wild (*NAT2*4/*4*) genotype as reference, the heterozygous (*NAT2*4/*5B*) (AOR=0.69; 95% C.I.= 0.53-0.90; **p**=**0.005**) as well as the mutant genotype (*NAT2*5B/*5B*) (AOR=0.50; 95% C.I.= 0.34-0.73; **p**=**0.0003**) demonstrate reduced risk for lung cancer development when the co-dominant model was applied. Using the dominant as well as the recessive model, both the combined (*NAT2 *4/*5B + *5B/*5B*) and mutant genotype (*NAT2*5B/*5B*) exhibited a reduced risk towards lung cancer susceptibility with an adjusted odd of 0.64 (95% C.I.=0.50-0.83; **p**=**0.0005**) and 0.63 (95% C.I.=0.44-0.88; p=**0.008**), respectively.

The last variant of the *NAT2* gene studied was *803A>G (NAT2*5C)* and tabulated in Table 5.27. The frequency of wild genotype (*NAT2*4/*4*) was 48.55% in controls and 51.09% in cases. For heterozygous (NAT2*4/*5C) (39.82% *vs.* 35.64%) and mutant genotype (NAT2*5C/*5C) (11.63% *vs.* 13.27%), not much difference was found between subjects in controls and cases. However, there was no significant difference between the genotypic frequencies among cases and controls (χ^2=2.22, df=2, p=0.33). The MAF was found to be almost similar in controls (0.32) and cases (0.31). The odds ratio was calculated using logistic regression and wild genotype as a reference, and this polymorphism of the *NAT2* gene did not exhibit any significant role in modulating the risk towards lung cancer in any of the three genetic models used, shown in Table 5.27.

Table 5.27: Genotypic and allelic distribution of the *NAT* genetic variant and its association with risk of lung cancer

Genotype ***590G>A (NAT2*6)***	**Controls (550) N(%)**	**Cases (550) N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model				
*NAT2*4/*4*	221 (40.18)	226 (41.09)	1.00 (Reference)	
*NAT2*4/*6*	245 (44.55)	257 (46.73)	1.04 (0.81-1.35)	0.74
*NAT2*6/*6*	84 (15.27)	67 (12.18)	0.78 (0.54-1.14)	0.20
NAT2*4 (Allele)	687	709		
NAT2*6 (Allele)	413	391		
MAF	0.37	0.35		
		$\chi2$=2.26, df=2		0.32
Dominant model				
*NAT2*4/*4*	221 (40.18)	226 (41.09)	1.00 (Reference)	
*NAT2*4/*6 + *6/*6*	329 (59.82)	324 (58.91)	0.98 (0.77-1.25)	0.86
Recessive model				
*NAT2*4/*4 + *4/*6*	466 (84.73)	483 (87.82)	1.00 (Reference)	
*NAT2*6/*6*	84 (15.27)	67 (12.18)	0.78 (0.55-1.10)	0.16

Genotype ***857G>A (NAT2*7)***	**Controls (550) N(%)**	**Cases (550)N(%)**	**AOR (95% C.I.)**[a]	***p*** [b]
Codominant model				
*NAT2 *4/*4*	264 (48.00)	222 (40.36)	1.00 (Reference)	
*NAT2 *4/*7*	223 (40.55)	269 (48.91)	1.43 (1.11-1.85)	**0.005**
*NAT2 *7/*7*	63 (11.45)	59 (10.73)	1.09 (0.73-1.62)	0.67
NAT2 *4 (Allele)	751	713		
NAT2 *7 (Allele)	349	387		
MAF	0.32	0.35		
	$\chi2$=8.06, df=2			**0.01**
Dominant model				
*NAT2 *4/*4*	264 (48.00)	222 (40.36)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	286 (52.00)	328 (59.64)	1.36 (1.07-1.73)	**0.01**
Recessive model				
*NAT2 *4/*4 + *4/*7*	487 (88.55)	491 (89.27)	1.00 (Reference)	
*NAT2 *7/*7*	63 (11.45)	59 (10.73)	0.93 (0.64-1.36)	0.71

Genotype ***481C>T (NAT2*5B)***	**Controls (550)N(%)**	**Cases (550)N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model				
*NAT2*4/*4*	184 (33.45)	238 (43.27)	1.00 (Reference)	
*NAT2*4/*5B*	272 (49.45)	248 (45.09)	0.69 (0.53-0.90)	**0.005**
*NAT2*5B/*5B*	94 (17.10)	64 (11.64)	0.50 (0.34-0.73)	**0.0003**
NAT2*4 (Allele)	640	724		
NAT2*5B (Allele)	460	376		
MAF	0.42	0.34		
		$\chi2$=13.71, df=2		**0.001**
Dominant model				
*NAT2*4/*4*	184 (33.45)	238 (43.27)	1.00 (Reference)	
*NAT2*4/*5B+ *5B/*5B*	366 (66.55)	312 (56.73)	0.64 (0.50-0.83)	**0.0005**
Recessive model				
*NAT2 *4/*4 + *4/*5B*	456 (82.90)	486 (88.36)	1.00 (Reference)	
*NAT2 *5B/*5B*	94 (17.10)	64 (11.64)	0.63 (0.44-0.88)	**0.008**

Genotype ***803A>G (NAT2*5C)***	**Controls (550)N(%)**	**Cases (550)N(%)**	**AOR (95% C.I.)**[a]	***p*** [b]
Codominant model				
*NAT2 *4/*4*	267 (48.55)	281 (51.09)	1.00 (Reference)	
*NAT2 *4/*5C*	219 (39.82)	196 (35.64)	0.84 (0.65-1.08)	0.18
*NAT2 *5C/*5C*	64 (11.63)	73 (13.27)	1.04 (0.86-1.25)	0.71
NAT2 *4 (Allele)	753	758		
NAT2 *5C (Allele)	347	342		
MAF	0.32	0.31		
	$\chi2$=2.22, df=2			0.33
Dominant model				
*NAT2*4/*4*	267 (48.55)	281 (51.09)	1.00 (Reference)	
*NAT2*4/*5C + *5C/*5C*	283 (51.45)	269 (48.91)	0.89 (0.70-1.13)	0.34
Recessive model				
*NAT2 *4/*4 + *4/*5C*	486 (88.36)	477 (88.73)	1.00 (Reference)	
*NAT2 *5C/*5C*	64 (11.64)	73 (13.27)	1.16 (0.81-1.66)	0.41

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided $\chi2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Further, the *ad*-hoc analysis was carried out wherein crude OR and OR after adjustment with smoking status only (smoking being an important risk factor for lung cancer) was calculated as tabulated in Table 5.28. The results depict that there were not much difference in OR with adjustment and without adjustment with smoking status.

Table 5.28: *Ad*-hoc analysis of the *NAT2* genetic variant and its association with risk of lung cancer

Genotype rs1799931 857G>A (*NAT2*7*)	**Without adjustment**				**With adjustment**			
	OR[a] (95% CI)	**Coeffecient**	**SE**	**p[c]**	**OR[b] (95% CI)**	**Coeffecient**	**SE**	**p[c]**
Codominant model								
NAT2*4/*4	1.00 (Reference)				1.00 (Reference)			
NAT2*4/*7	1.434(1.11-1.84)	0.361	0.128	**0.005**	1.427(1.11-1.84)	0.356	0.128	**0.0057**
NAT2*7/*7	1.114(0.75-1.66)	0.107	0.203	0.595	1.102(0.74-1.64)	0.097	0.203	0.633
Dominant model								
NAT2*4/*4	1.00 (Reference)				1.00 (Reference)			
NAT2*4/*7+ NAT2*7/*7	1.364(1.07-1.73)	0.310	0.122	**0.011**	1.358(1.07-1.72)	0.306	0.122	**0.012**
Recessive model								
NAT2*4/*4+ NAT2*4/*7	1.00 (Reference)				1.00 (Reference)			
NAT2*7/*7	0.929(0.64-1.35)	-0.074	0.192	0.701	0.924(0.63-1.35)	-0.079	0.192	0.681
Genotype rs1208 803A>G (*NAT2*5C*)	**Without adjustment**				**With adjustment**			
	OR[a] (95% CI)	**Coeffecient**	**SE**	**p[c]**	**OR[b] (95% CI)**	**Coeffecient**	**SE**	**p[c]**
Codominant model								
NAT2*4/*4	1.00 (Reference)				1.00 (Reference)			
NAT2*4/*5C	0.850(0.66-1.10)	-0.162	0.130	0.213	0.85(0.65-1.09)	-0.167	0.130	0.200
NAT2*5C/*5C	1.084(0.74-1.58)	0.080	0.191	0.674	1.085(0.74-1.58)	0.082	0.192	0.670
Dominant model								
NAT2*4/*4	1.00 (Reference)				1.00 (Reference)			
NAT2*4/*5C+ NAT2*5C/*5C	0.903(0.71-1.14)	-0.102	0.121	0.399	0.901(0.71-1.14)	-0.104	0.121	0.388
Recessive model								
NAT2*4/*4+NAT2*4/*5	1.00 (Reference)				1.00 (Reference)			
NAT2*5/*5	1.162(0.81-1.66)	0.150	0.183	0.411	1.165(0.81-1.67)	0.153	0.183	0.404
Genotype rs1799929 481C>T (*NAT2*5B*)	**Without adjustment**				**With adjustment**			
	OR[a] (95% CI)	**Coeffecient**	**SE**	**p[c]**	**OR[b] (95% CI)**	**Coeffecient**	**SE**	**p[c]**
Codominant model								
NAT2*4/*4	1.00 (Reference)				1.00 (Reference)			
NAT2*4/*5B	0.705(0.54-0.91)	-0.349	0.132	**0.008**	0.692(0.53-0.89)	-0.368	0.132	**0.005**

NAT2*5B/*5B	0.526(0.36-0.76)	-0642	0.189	**0.0007**	0.509(0.35-0.74)	-0.675	0.191	**0.0004**
Dominant model								
NAT2*4/*4	1.00 (Reference)				1.00 (Reference)			
NAT2*4/*5B+ NAT2*5B/*5B	0.659(0.52-0.84)	-0.417	0.125	**0.0008**	0.650(0.51-0.83)	-0.430	0.125	**0.0006**
Recessive model								
NAT2*4/*4+ NAT2*4/*5B	1.00 (Reference)				1.00 (Reference)			
NAT2*5B/*5B	0.634(0.45-0.90)	-0.448	0.175	**0.010**	0.633(0.45-0.89)	-0.457	0.175	**0.009**
Genotype rs1799930 590G>A (NAT2*6)	**Without adjustment**				**With adjustment**			
	OR[a] (95% CI)	**Coeffecient**	**SE**	**p[c]**	**OR[b] (95% CI)**	**Coeffecient**	**SE**	**p[c]**
Codominant model								
NAT2*4/*4	1.00 (Reference)				1.00 (Reference)			
NAT2*4/*6	1.026(0.79-1.32)	0.025	0.130	0.845	1.036(0.80-1.34)	0.035	0.130	0.787
NAT2*6/*6	0.780(0.54-1.13)	-0.248	0.189	0.199	0.79(0.54-1.14)	-0.241	0.189	0.204
Dominant model								
NAT2*4/*4	1.00 (Reference)				1.00 (Reference)			
NAT2*4/*6+ NAT2*6/*6	0.963(0.76-1.22)	-0.038	0.123	0.759	0.972(0.76-1.24)	-0.028	0.123	0.817
Recessive model								
NAT2*4/*4+ NAT2*4/*6	1.00 (Reference)				1.00 (Reference)			
NAT2*6/*6	0.769(0.54-1.09)	-0.262	0.176	0.137	0.772(0.55-1.09)	-0.259	0.176	0.142

a) Odds ratios, 95% confidence intervals, b) adjusted Odds ratio, 95% confidence interval calculated, and c) corresponding *p*-values were calculated by logistic regression analysis b) after adjusting for smoking status.

5.6.3.6. Association of *NAT2* polymorphisms with lung cancer risk based on histological subtypes

As lung cancer is one of the most heterogeneous solid tumors and the molecular pathology of each subtype is varied. Hence, the analysis for cancer predisposition was also carried out for different subtypes of lung cancer concerning the different genotypes of *NAT2* SNP's, and the data are summarized in Table 5.29 a), b), c) and d).

In the case of *NAT2*6* (*590G>A*), a decreased frequency of patients carrying variant (*NAT2*6/*6*) was observed in the SQCC subgroup (9.72% *vs.* 15.27% in controls), which showcased a decreased risk of lung cancer in them (AOR=0.64, 95% C.I.=0.37-1.11) but was not significant. No association was observed in ADCC and SCLC as the frequency of subjects harboring variant genotype among ADCC and SCLC groups was 13.50% and 14.17%, respectively. However, the frequency was lower in both the subgroups than in controls, but the difference was not statistically significant (Table 5.29 a)).

In the case of *NAT2*7* (*857G>A*), stratifying the subjects based on their histological subtypes, it was evident that lung cancer patients carrying the mutant genotype did not exhibit any significant association in SQCC and SCLC subgroups (Table 5.29 b)). The percentage of heterozygotes among the SCLC subgroup was 41.73%, similar to controls (40.55%) but lacking statistical significance. On the contrary, the frequency of heterozygotes increased to 52.00% and 50.46% among ADCC, and SQCC and hence, it pertained to show a significantly increased risk of developing lung cancer. The AOR for the heterozygote was found to be 1.81 (95% C.I.=1.26-2.59, p=**0.0012**) in ADCC subjects and 1.39 (95% C.I.=0.99-1.97, p=**0.05**) in SQCC subjects.

Table 5.29 a): Genotypic and allelic distribution of the *NAT2*6 590G>A* genetic variant and its association with risk of lung cancer on basis of tumor histology

	ADCC				SQCC			SCLC		
Genotype rs1799930 *590G>A (NAT2*6)*	**Controls (550) N(%)**	**Cases (200) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (216) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (127) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model										
*NAT2 *4/*4*	221 (40.18)	83 (41.50)	1.00 (Reference)		89 (41.20)	1.00 (Reference)		50 (39.37)	1.00 (Reference)	
*NAT2 *4/*6*	245 (44.55)	90 (45.00)	0.92 (0.65-1.32)	0.67	106 (49.08)	1.15 (0.81-1.62)	0.43	59 (46.46)	1.15 (0.74-1.79)	0.53
*NAT2 *6/*6*	84 (15.27)	27 (13.50)	0.82 (0.49-1.37)	0.44	21 (9.72)	0.64 (0.37-1.11)	0.11	18 (14.17)	1.12 (0.60-2.08)	0.71
Dominant model										
*NAT2 *4/*4*	221 (40.18)	83 (41.50)	1.00 (Reference)		89 (41.20)			50 (39.37)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	329 (59.82)	117 (58.50)	0.91 (0.65-1.27)	0.57	127 (58.80)	1.02 (0.73-1.41)	0.92	77 (60.63)	1.13 (0.75-1.72)	0.55
Recessive model										
*NAT2 *4/*4 + *4/*6*	466 (84.73)	173 (86.50)	1.00 (Reference)		195(90.28)	1.00 (Reference)		109(85.83)	1.00 (Reference)	
*NAT2 *6/*6*	84 (15.27)	27 (13.50)	0.87 (0.54-1.40)	0.56	21 (9.72)	0.60 (0.36-1.01)	0.05	18 (14.17)	1.01 (0.57-1.77)	0.97

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.29 b): Genotypic and allelic distribution of the *NAT2*7 857G>A* genetic variant and its association with risk of lung cancer on basis of tumor histology

					ADCC					
Genotype rs1799931 *857G>A (NAT2*7)*	**Controls (550) N(%)**	**Cases (200) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (216) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (127) N(%)**	**AOR (95% C.I.)[a]**	
Codominant model										
*NAT2 *4/*4*	264 (48.00)	73 (36.50)	1.00 (Reference)		88 (40.74)	1.00 (Reference)		59 (46.46)	1.00 (Reference)	
*NAT2 *4/*7*	223 (40.55)	104 (52.00)	1.81 (1.26-2.59)	**0.0012**	109(50.46)	1.39 (0.99-1.97)	**0.05**	53 (41.73)	0.97 (0.63-1.51)	0.91
*NAT2 *7/*7*	63 (11.45)	23 (11.50)	1.42 (0.81-2.48)	0.21	19 (8.80)	0.80 (0.45-1.44)	0.46	15 (11.81)	1.004(0.52-1.95)	0.67
Dominant model										
*NAT2 *4/*4*	264 (48.00)	73 (36.50)	1.00(Reference)		88 (40.74)	1.00 (Reference)		59 (46.46)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	286 (52.00)	127 (63.50)	1.07 (0.64-1.80)	0.79	128 (59.26)	1.27 (0.91-1.76)	0.16	68 (53.54)	0.99 (0.66-1.50)	0.99
Recessive model										
*NAT2 *4/*4 + *4/*7*	487 (88.55)	177 (88.50)	1.00(Reference)		197 (91.20)	1.00 (Reference)		112 (88.19)	1.00 (Reference)	
*NAT2 *7/*7*	63 (11.45)	23 (11.50)	1.07 (0.64-1.80)	0.79	19 (8.80)	0.70 (0.41-1.22)	0.21	15 (11.81)	1.04 (0.56-1.94)	0.89

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

The next genotypic variant was *NAT2*5B* (*481C>T*), and the subjects carrying heterozygous, mutant, or combined genotype did not exhibit any significant association in the ADCC subgroup (Table 5.29 c)). Further, the frequency of subjects carrying mutant genotype for the SQCC subgroup was more in controls than in the cases (17.10% *vs*. 10.19%). At the same time, the wild genotype was overrepresented in cases than the controls (42.59% *vs.* 33.45%). The mutant genotype (*NAT2*5B/*5B*) exhibits a highly significant protective effect towards lung cancer risk (AOR=0.39, 95% C.I.=0.22-0.67, p=**0.0006**). The percentage of subjects harboring heterozygotes was almost similar in controls and cases (49.45% and 47.22%), and the same trend of reduced risk towards lung cancer was observed (AOR=0.65, 95% C.I.=0.45-0.92, p=**0.016)**. Similarly, the combined genotype (*NAT2*4/*5B+NAT2*5B/5B*; AOR=0.58, 95% C.I.=0.41-0.81, p=**0.0013**) as well as mutant genotype (using the recessive model; AOR=0.50, 95% C.I.=0.30-0.82, p=**0.0065**) also showed a protective effect towards SQCC (Table 5.29 c)). In the case of subjects in the SCLC subgroup, the heterozygous genotype (*NAT2*4/*5B*) was over-represented in controls than cases (49.45% *vs*. 40.16%) and was found to exhibit a protective effect towards the risk of SCLC (AOR=0.52, 95% C.I.=0.33-0.80, p=**0.0034**). Whereas the frequency of mutant genotype was a bit more among controls than cases (17.10% *vs*. 11.81%), and this genotype was also found to exhibit a reduced risk of SCLC (AOR=0.43, 95% C.I.=0.23-0.82, p=**0.01**). Similarly, the combined genotype (*NAT2*4/*5B+NAT2*5B/5B*) pertained to show a significant protective effect (AOR=0.50, 95% C.I.=0.33-0.75, p=**0.001**) (Table 5.29 c)).

For *NAT2 803A>G*, when patients were classified based on their histological subtype, none of the subgroups revealed any association with lung cancer risk, as shown in Table 5.29 d). As the difference in frequency of mutant genotype was not found to be significant, the lowest frequency was observed among SQCC subjects (9.72%), followed by SCLC patients (14.96%) and ADCC (15.50%) and among the control group (5.2%). Therefore, this polymorphic site did not confer any role towards modulating lung cancer susceptibility in any of the histological subtypes.

Table 5.29 c): Genotypic and allelic distribution of the *NAT2*5B 481C>T* genetic variant and its association with risk of lung cancer on basis of tumor histology

	ADCC				SQCC			SCLC		
Genotype rs1799929 *481C>T(NAT2*5B)*	**Controls (550) N(%)**	**Cases (200) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (216) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (127) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model										
*NAT2 *4/*4*	184 (33.45)	81 (40.50)	1.00 (Reference)		92 (42.59)	1.00 (Reference)		61 (48.03)	1.00 (Reference)	
*NAT2 *4/*5B*	272 (49.45)	93 (46.50)	0.81 (0.57-1.17)	0.26	102 (47.22)	0.65 (0.45-0.92)	**0.016**	51 (40.16)	0.52 (0.33-0.80)	**0.0034**
*NAT2 *5B/*5B*	94 (17.10)	26 (13.00)	0.67 (0.40-1.12)	0.13	22 (10.19)	0.39 (0.22-0.67)	**0.0006**	15 (11.81)	0.43 (0.23-0.82)	**0.01**
Dominant model										
*NAT2 *4/*4*	184 (33.45)	81 (40.50)	1.00 (Reference)		92 (42.59)	1.00 (Reference)		61 (48.03)	1.00 (Reference)	
*NAT2 *4/*5B + *5B/*5B*	366 (66.45)	119(59.50)	0.79 (0.56-1.11)	0.17	124 (57.41)	0.58 (0.41-0.81)	**0.0013**	66 (51.97)	0.50 (0.33-0.75)	**0.001**
Recessive model										
*NAT2 *4/*4 + *4/*5B*	456 (82.91)	174(87.00)	1.00 (Reference)		194 (89.81)	1.00 (Reference)		112 (88.19)	1.00 (Reference)	
*NAT2 *5B/*5B*	94 (17.10)	26 (13.00)	0.78 (0.48-1.25)	0.30	22 (10.19)	0.50 (0.30-0.82)	**0.0065**	15 (11.81)	0.61 (0.33-1.12)	0.11

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.29 d): Genotypic and allelic distribution of the *NAT*5C 803A>G* genetic variant and its association with risk of lung cancer on basis of tumor histology

	ADCC				SQCC			SCLC		
Genotype rs1208 *803A>G(NAT2*5C)*	**Controls (550) N(%)**	**Cases (200) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (216) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Cases (127) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model										
*NAT2 *4/*4*	267 (48.55)	98 (49.00)	1.00 (Reference)		115(53.24)	1.00 (Reference)		64 (50.39)	1.00 (Reference)	
*NAT2 *4/*5C*	219 (39.82)	71 (35.50)	0.90 (0.63-1.30)	0.59	80 (37.04)	0.80 (0.57-1.13)	0.21	44 (34.65)	0.79 (0.51-1.24)	0.30
*NAT2 *5C/*5C*	64 (11.63)	31 (15.50)	1.16 (0.90-1.48)	0.25	21 (9.72)	0.84 (0.64-1.10)	0.21	19 (14.96)	1.12 (0.83-1.52)	0.46
Dominant model										
*NAT2 *4/*4*	267 (48.55)	98 (49.00)	1.00 (Reference)		115 (53.24)	1.00 (Reference)		64 (50.39)	1.00 (Reference)	
*NAT2 *4/*5C + *5C/*5C*	283 (51.45)	102 (51.00)	1.002(0.72-1.39)	0.99	101 (46.76)	0.78 (0.56-1.07)	0.13	63 (49.61)	0.89 (0.59-1.33)	0.56
Recessive model										
*NAT2 *4/*4 + *4/*5C*	486 (88.36)	169 (84.50)	1.00 (Reference)		195 (90.28)	1.00 (Reference)		108 (85.04)	1.00 (Reference)	
*NAT2 *5C/*5C*	64 (11.64)	31 (15.50)	1.40 (0.87-2.25)	0.16	21 (9.72)	0.77 (0.45-1.32)	0.35	19 (14.96)	1.38 (0.77-2.46)	0.28

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.3.7. Association of *NAT2* gene with lung cancer risk based on smoking

To evaluate the role of smoking in modulating the association between the *NAT2* gene polymorphisms and lung cancer risk; the subjects were classified based on their smoking status into smokers and non-smokers. The smokers were further classified into heavy-smokers and light-smokers. The association analysis was carried out for all four polymorphic sites of the *NAT2* gene, and the genotypic frequencies and the adjusted odds ratio with 95% C.I. within both the subgroups have been tabulated in Table 5.30 a), b), c), and d).

When we applied the co-dominant model for subjects who were smokers and carrying both the mutant alleles (*NAT2*5B/*5B*) for the *481C>T (NAT2*5B*) polymorphism, a decreased propensity towards smoking-induced lung cancer (AOR=0.38, 95% C.I.=0.25-0.59, p=**<0.0001**) was evident. Similarly, it was conferred that the presence of the heterozygous genotype (NAT2*4/*5B; AOR=0.59, 95% C.I.=0.44-0.80, p=**0.0007**) in smokers were also found to be associated with reduced risk for lung cancer. To increase the power of the association, both the heterozygous and the homozygous variant genotypes were combined (*NAT2*4/*5B + *5B/*5B*), and this single subgroup also exhibited a decreased risk towards lung cancer in case of smokers (AOR=0.53, 95% C.I.=0.40-0.71, **p=<0.0001**). Using the recessive model, wherein the mutant genotype (*NAT2*5B/*5B*) was compared to the combined genotype (NAT2*4/*5B + *5B/*5B) taken as a reference, a reduced risk towards lung cancer was observed (AOR=0.52, 95% C.I.=0.35-0.77, **p=0.001**) as shown in Table 5.30 a).

Furthermore, stratified analysis by pack-years also suggested that light smokers harboring the heterozygous (*NAT2*4/*5B*) (AOR=0.60, 95% C.I.=0.41-0.87, **p=0.008**) as well as homozygous mutant genotypes (*NAT2*5B/*5B*) (AOR=0.44, 95% C.I.=0.25-0.75, **p=0.002**) for the *481C>T (NAT2*5B*) polymorphism revealed a diminished risk for lung cancer development. Using both the dominant (AOR=0.55, 95% C.I.=0.38-0.79, **p=0.001**) as well as recessive models (AOR=0.58, 95% C.I.=0.35-0.96, **p=0.03**), our data observed that lung cancer risk for light smokers for the *481C>T (NAT2*5B*) polymorphism was found to be associated with decreased LC risk (Table 5.30 b))**.** In the case of heavy smokers and *481C>T (NAT2*5B*) polymorphism, as shown in Table 5.30 c), a weak association for lung cancer susceptibility was visible for those subjects who were carrying both the copies of mutant allele for the *NAT2*5B* polymorphism (AOR=0.28, 95% C.I.=0.13-0.61, **p=0.001**). Furthermore, a similar association was also clearly visible for the subjects who were categorized as heavy smokers and were heterozygous for the

481C>T (*NAT2*5B*) polymorphism (AOR=0.55, 95% C.I.=0.33-0.92, **p=0.02)** using the co-dominant model. The dominant model for the *481C>T (NAT2*5B*) polymorphism also expressed an attenuated risk for lung cancer for the combined genotype (AOR=0.48, 95% C.I.=0.29-0.79, **p=0.004**) for subjects who were heavy smokers, whereas for lung cancer patients who were heavy smokers and carrying the mutant genotype for the *481C>T (NAT2*5B*) polymorphism also demonstrated a diminutive risk for lung cancer susceptibility (AOR=0.43, 95% C.I.=0.22-0.83, **p=0.01**).

In case for the *NAT2 590G>A (NAT2*6*) and *803A>G (NAT2*5C*) polymorphisms after the stratification analysis by smoking status, our data did not show any association towards lung cancer susceptibility when all the three genetic models were applied as shown in Table 5.30 a), b), and c). In the case of the *857G>A (NAT2*7*) polymorphism of *NAT2* gene, here too, we failed to observe any meaningful association based on smoking; however, in the case of non-smokers, the subjects carrying the heterozygous genotype depicted a 2-fold risk towards lung cancer (AOR=2.12, 95% C.I.=1.20-3.77, **p=0.01**) using the co-dominant model (Table 5.30 d)). The dominant model for the above polymorphic variant also revealed a 2-fold risk of developing lung cancer (AOR=2.14, 95% C.I.=1.24-3.69, **p=0.006**) in the case of non-smokers (Table 5.30 d)). The results obtained here suggest that the *NAT2 481C>T* (*NAT2*5B*) and *857G>A (NAT2*7*) polymorphisms are related to smoking status, whereas; *590G>A (NAT2*6*) and *803A>G (NAT2*5C*) polymorphisms do not manifest any association between smoking status and lung cancer risk.

Table 5.30 a): Relationship of different *NAT2* genotypes with the smoking status of cases and controls (smokers)

SMOKERS									
Genotype ***590G>A (NAT2*6)***	**Controls (425) N(%)**	**Cases (444)N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]	**Genotype** ***481C>T (NAT2*5B)***	**Controls (425)N(%)**	**Cases (444)N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model					**Co-dominant model**				
*NAT2 *4/*4*	177 (41.65)	188 (42.34)	1.00 (Reference)		*NAT2 *4/*4*	126 (29.65)	194 (43.69)	1.00 (Reference)	
*NAT2 *4/*6*	184 (43.29)	203 (45.72)	1.05 (0.78-1.40)	0.75	*NAT2 *4/*5B*	219 (51.53)	200 (45.05)	0.59 (0.44-0.80)	**0.0007**
*NAT2 *6/*6*	64 (15.06)	53 (11.94)	0.81 (0.53-1.24)	0.33	*NAT2 *5B/*5B*	80 (18.82)	50 (11.26)	0.38 (0.25-0.59)	**<0.0001**
Dominant model					**Dominant model**				
*NAT2 *4/*4*	177 (41.65)	188 (42.34)	1.00 (Reference)		*NAT2 *4/*4*	126(29.65)	194 (43.69)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	248 (58.35)	256 (57.66)	0.98 (0.75-1.29)	0.91	*NAT2*4/*5B + *5B/*5B*	299(70.35)	250 (56.31)	0.53 (0.40-0.71)	**<0.0001**
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	361 (84.94)	391 (88.06)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	345 (81.18)	394 (88.74)	1.00 (Reference)	
*NAT2 *6/*6*	64 (15.06)	53 (11.94)	0.79 (0.53-1.18)	0.25	*NAT2 *5B/*5B*	80 (18.82)	50 (11.26)	0.52 (0.35-0.77)	**0.0011**
Genotype ***857G>A (NAT2*7)***	**Controls (425) N(%)**	**Cases (444)N(%)**	**AOR (95% C.I.)** [a]		**Genotype** ***803A>G (NAT2*5C)***	**Controls (425)N(%)**	**Cases (444)N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model					**Codominant model**				
*NAT2 *4/*4*	195 (45.88)	182 (40.99)	1.00 (Reference)		*NAT2 *4/*4*	200 (47.06)	230 (51.80)	1.00 (Reference)	
*NAT2 *4/*7*	177 (41.65)	216(48.65)	1.32 (0.99-1.75)	0.06	*NAT2 *4/*5C*	173 (40.70)	159 (35.81)	0.79 (0.59-1.06)	0.12
*NAT2 *7/*7*	53 (12.47)	46 (10.36)	0.91 (0.58-1.44)	0.69	*NAT2 *5C/*5C*	52 (12.24)	55 (12.39)	0.95 (0.76-1.18)	0.63
Dominant model					**Dominant model**				
*NAT2 *4/*4*	195 (45.88)	182 (40.99)	1.00 (Reference)		*NAT2 *4/*4*	200 (47.06)	230 (51.80)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	230 (54.12)	262 (59.01)	1.23 (0.94-1.62)	0.14	*NAT2 *4/*5C + *5C/*5C*	225 (52.94)	214 (48.20)	0.81 (0.62-1.07)	0.14
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	372 (87.53)	398 (89.64)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	373 (87.76)	389 (87.61)	1.00 (Reference)	
*NAT2 *7/*7*	53 (12.47)	46 (10.36)	0.82 (0.53-1.25)	0.35	*NAT2 *5C/*5C*	52 (12.24)	55 (12.39)	0.99 (0.66-1.50)	0.97

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.30 b): Relationship of different *NAT2* genotypes with the smoking status of cases and controls (light-smokers)

LIGHT-SMOKER									
Genotype ***590G>A (NAT2*6)***	**Controls (300) N(%)**	**Cases (248)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Genotype** ***481C>T (NAT2*5B)***	**Controls (300)N(%)**	**Cases (248)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	124 (41.33)	102 (41.13)	1.00 (Reference)		*NAT2 *4/*4*	91 (30.33)	110 (44.35)	1.00 (Reference)	
*NAT2 *4/*6*	127 (42.33)	118 (47.58)	1.13 (0.78-1.64)	0.51	*NAT2 *4/*5B*	153 (51.00)	107 (43.15)	0.60 (0.41-0.87)	**0.008**
*NAT2 *6/*6*	49 (16.34)	28 (11.29)	0.71 (0.41-1.23)	0.23	*NAT2 *5B/*5B*	56 (18.67)	31 (12.50)	0.44 (0.25-0.75)	**0.0027**
Dominant model					**Dominant model**				
*NAT2 *4/*4*	124 (41.33)	102 (41.13)	1.00 (Reference)		*NAT2 *4/*4*	91 (30.33)	110 (44.35)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	176 (58.67)	146 (58.87)	1.02 (0.71-1.44)	0.93	*NAT2 *4/*5B + *5B/*5B*	209 (69.67)	138 (55.65)	0.55 (0.38-0.79)	**0.0012**
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	251 (83.66)	220 (88.71)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	244 (81.33)	217 (87.50)	1.00 (Reference)	
*NAT2 *6/*6*	49 (16.34)	28 (11.29)	0.67 (0.39-1.09)	0.10	*NAT2 *5B/*5B*	56 (18.67)	31 (12.50)	0.58 (0.35-0.96)	**0.03**
Genotype ***857G>A (NAT2*7)***	**Controls (300) N(%)**	**Cases (248)N(%)**	**AOR (95% C.I.)[a]**		**Genotype** ***803A>G (NAT2*5C)***	**Controls (300)N(%)**	**Cases (248)N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	143 (47.67)	105 (42.34)	1.00 (Reference)		*NAT2 *4/*4*	144 (48.00)	129 (52.01)	1.00 (Reference)	
*NAT2 *4/*7*	123 (41.00)	116 (46.77)	1.28 (0.89-1.85)	0.18	*NAT2 *4/*5C*	119 (39.67)	89 (35.89)	0.82 (0.57-1.20)	0.31
*NAT2 *7/*7*	34 (11.33)	27 (10.89)	0.99 (0.55-1.80)	0.98	*NAT2 *5C/*5C*	37 (12.33)	30 (12.10)	0.95 (0.72-1.25)	0.70
Dominant model					**Dominant model**				
*NAT2 *4/*4*	143 (47.67)	105 (42.34)	1.00 (Reference)		*NAT2 *4/*4*	144 (48.00)	129 (52.02)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	157 (52.33)	143 (57.66)	1.23 (0.87-1.75)	0.24	*NAT2 *4/*5C + *5C/*5C*	156 (52.00)	119 (47.98)	0.84 (0.59-1.18)	0.32
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	266 (88.67)	221 (89.11)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	263 (87.67)	218 (87.90)	1.00 (Reference)	
*NAT2 *7/*7*	34 (11.33)	27 (10.89)	0.93 (0.53-1.61)	0.78	*NAT2 *5C/*5C*	37 (12.33)	30 (12.10)	0.96 (0.56-1.64)	0.88

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.30 c): Relationship of different *NAT2* genotypes with the smoking status of cases and controls (heavy smokers)

HEAVY-SMOKER									
Genotype	**Controls**	**Cases**	**AOR**	***p*[b]**	**Genotype**	**Controls**	**Cases**	**AOR**	***p*[b]**
590G>A (NAT2*6)	**(125) N(%)**	**(196)N(%)**	**(95% C.I.)[a]**		***481C>T (NAT2*5B)***	**(125)N(%)**	**(196)N(%)**	**(95% C.I.)[a]**	
Codominant model					**Codominant model**				
*NAT2 *4/*4*	53 (42.40)	86 (43.88)	1.00 (Reference)		*NAT2 *4/*4*	35 (28.00)	84 (42.86)	1.00 (Reference)	
*NAT2 *4/*6*	57 (45.60)	85 (43.37)	0.94 (0.58-1.52)	0.79	*NAT2 *4/*5B*	66 (52.80)	93 (47.45)	0.55 (0.33-0.92)	**0.02**
*NAT2 *6/*6*	15 (12.00)	25 (12.75)	1.15 (0.55-2.40)	0.71	*NAT2 *5B/*5B*	24 (19.20)	19 (9.69)	0.28 (0.13-0.61)	**0.0012**
Dominant model					**Dominant model**				
*NAT2 *4/*4*	53 (42.40)	86 (43.88)	1.00 (Reference)		*NAT2 *4/*4*	35 (28.00)	84 (42.86)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	72 (57.60)	110 (56.12)	0.97 (0.61-1.53)	0.89	*NAT2 *4/*5B + *5B/*5B*	90 (72.00)	112(57.14)	0.48 (0.29-0.79)	**0.004**
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	110 (88.00)	171 (87.25)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	101 (80.80)	177 (90.31)	1.00 (Reference)	
*NAT2 *6/*6*	15 (12.00)	25 (12.75)	1.13 (0.57-2.24)	0.73	*NAT2 *5B/*5B*	24 (19.20)	19 (9.69)	0.43 (0.22-0.83)	**0.01**
Genotype	**Controls**	**Cases**	**AOR**		**Genotype**	**Controls**	**Cases**	**AOR**	***p*[b]**
857G>A (NAT2*7)	**(125) N(%)**	**(196)N(%)**	**(95% C.I.)[a]**		***803A>G (NAT2*5C)***	**(125)N(%)**	**(196)N(%)**	**(95% C.I.)[a]**	
Codominant model					**Codominant model**				
*NAT2 *4/*4*	52 (41.60)	77 (39.29)	1.00 (Reference)		*NAT2 *4/*4*	56 (44.80)	101 (51.53)	1.00 (Reference)	
*NAT2 *4/*7*	54 (43.20)	100 (51.02)	1.24 (0.76-2.01)	0.39	*NAT2 *4/*5C*	54 (43.20)	70 (35.71)	0.68 (0.42-1.12)	0.13
*NAT2 *7/*7*	19 (15.20)	19 (9.69)	0.70 (0.34-1.45)	0.34	*NAT2 *5C/*5C*	15 (12.00)	25 (12.76)	0.83 (0.40-1.74)	0.63
Dominant model					**Dominant model**				
*NAT2 *4/*4*	52 (41.60)	77 (39.29)	1.00 (Reference)		*NAT2 *4/*4*	56 (44.80)	101 (51.53)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	73 (58.40)	119 (60.71)	1.09 (0.69-1.73)	0.70	*NAT2 *4/*5C + *5C/*5C*	69 (55.20)	95 (48.47)	0.72 (0.45-1.13)	0.15
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	106 (84.80)	177 (90.31)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	110 (88.00)	171 (87.24)	1.00 (Reference)	
*NAT2 *7/*7*	19 (15.20)	19 (9.69)	0.62 (0.31-1.22)	0.17	*NAT2 *5C/*5C*	15 (12.00)	25 (12.76)	1.06 (0.53-2.11)	0.87

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.30 d): Relationship of different *NAT2* genotypes with the smoking status of cases and controls (non-smokers)

NON-SMOKERS									
Genotype ***590G>A (NAT2*6)***	**Controls (125) N(%)**	**Cases (106)N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]	**Genotype** ***481C>T (NAT2*5B)***	**Controls (125)N(%)**	**Cases (106)N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model					**Codominant model**				
*NAT2 *4/*4*	44 (35.20)	38 (35.85)	1.00 (Reference)		*NAT2 *4/*4*	58 (46.40)	44 (41.51)	1.00 (Reference)	
*NAT2 *4/*6*	61 (48.80)	54 (50.94)	1.07 (0.60-1.92)	0.81	*NAT2 *4/*5B*	53 (42.40)	48 (45.28)	1.12 (0.64-1.98)	0.68
*NAT2 *6/*6*	20 (16.00)	14 (13.21)	0.79 (0.33-1.89)	0.60	*NAT2 *5B/*5B*	14 (11.20)	14 (13.21)	1.22 (0.51-2.89)	0.65
Dominant model					**Dominant model**				
*NAT2 *4/*4*	44 (35.20)	38 (35.85)	1.00 (Reference)		*NAT2 *4/*4*	58 (46.40)	44 (41.51)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	81(64.80)	68 (64.15)	1.03 (0.59-1.80)	0.91	*NAT2 *4/*5B + *5B/*5B*	67 (53.60)	62 (58.49)	1.14 (0.67-1.95)	0.63
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	105 (84.00)	92 (86.79)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	111 (88.80)	92 (86.79)	1.00 (Reference)	
*NAT2 *6/*6*	20 (16.00)	14 (13.21)	0.87 (0.41-1.87)	0.73	*NAT2 *5B/*5B*	14 (11.20)	14 (13.21)	1.16 (0.51-2.61)	0.72
Genotype ***857G>A (NAT2*7)***	**Controls (125) N(%)**	**Cases (106)N(%)**	**AOR (95% C.I.)** [a]		**Genotype** ***803A>G (NAT2*5C)***	**Controls (125)N(%)**	**Cases (106)N(%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model					**Codominant model**				
*NAT2 *4/*4*	69 (55.20)	40 (37.74)	1.00 (Reference)		*NAT2 *4/*4*	67 (53.60)	51 (48.11)	1.00 (Reference)	
*NAT2 *4/*7*	46 (36.80)	53 (50.00)	2.12 (1.20-3.77)	**0.01**	*NAT2 *4/*5C*	46 (36.80)	37 (34.91)	1.27 (0.70-2.31)	0.42
*NAT2 *7/*7*	10 (8.00)	13 (12.26)	2.34 (0.90-6.06)	0.08	*NAT2 *5C/*5C*	12 (9.60)	18 (16.98)	1.41 (0.92-2.15)	0.11
Dominant model					**Dominant model**				
*NAT2 *4/*4*	69 (55.20)	40 (37.74)	1.00 (Reference)		*NAT2 *4/*4*	67 (53.60)	51 (48.11)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	56 (44.80)	66 (62.26)	2.14 (1.24-3.69)	**0.006**	*NAT2 *4/*5C + *5C/*5C*	58 (46.40)	55 (51.89)	1.42 (0.83-2.43)	0.20
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	115 (92.00)	93 (87.74)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	113 (90.40)	88 (83.02)	1.00 (Reference)	
*NAT2 *7/*7*	10 (8.00)	13 (12.26)	1.59 (0.66-3.84)	0.30	*NAT2 *5C/*5C*	12 (9.60)	18 (16.98)	1.81 (0.81-4.05)	0.14

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.3.8. Association of *NAT2* polymorphisms with lung cancer risk based on stratification by gender

The stratification analysis was further carried out based on gender for both cases and controls, as shown in Table 5.31 a) for males and Table 5.31 b) for females. In the case of *590G>A (NAT2*6*) polymorphism when the recessive model was applied, it was evident that male subjects carrying the mutant genotype *(NAT2*6/*6*) depicted an attenuated risk towards lung cancer (AOR=0.60, 95% C.I.=0.42-0.86, **p=0.005**) (Table 5.31 a). At the same time, the co-dominant and dominant models did not exhibit any association with lung cancer development. For *857G>A (NAT2*7*) polymorphism, the male subjects carrying the single copy of the mutant allele (*NAT2*4/*7*) exhibited a weak association towards lung cancer susceptibility (AOR=1.32, 95% C.I.=1.00-1.74, **p=0.048**). In the co-dominant model in the case of *481C>T (NAT2*5B)* polymorphism, it was visible that those male subjects carrying the *NAT2*4/*5B* and *NAT2*5B/*5B* genotype exhibited a decreased risk towards lung cancer (AOR=0.61, 95% C.I.=0.46-0.82, **p=0.0009** and AOR=0.40, 95% C.I.=0.26-0.61, **p=<0.0001**) as compared to the male subjects carrying the wild type genotype (*NAT2*4/*4*). Similarly, applying the dominant model, our results also suggested that the *NAT2*4/*5B+*5/*5B* genotype showed a reduced and a weak association towards lung cancer susceptibility in male lung cancer patients (AOR=0.56, 95% C.I.=0.42-0.73, **p=<0.0001**), a similar trend was also seen when the recessive model was applied (AOR=0.54, 95% C.I.=0.37-0.79, **p=0.0014**) for those male lung cancer subjects carrying the *NAT2*5B/*5B* genotype (Table 5.31 a)).

However, in females, the mutant genotype (*NAT2*6/*6*) for *NAT2*6* polymorphism exhibited a diminutive risk towards lung cancer susceptibility using the co-dominant (AOR=0.30, 95% C.I.=0.09-0.99, **p=0.04**) model (Table 5.31 b)). In the case of the recessive model applied, it was evident that female subjects carrying the mutant alleles had a very less probability risk for lung cancer susceptibility (AOR=0.06, 95% C.I.=0.02-0.16, p=**<0.0001**). Our data observed a 2.6 and 3.66-folds risk for lung cancer development in female subjects harboring either the heterozygous (*NAT2*4/*7* (AOR=2.59, 95% C.I.=1.30-5.16, **p=0.007**)) or mutant genotype (*NAT2*7/*7* (AOR=3.66, 95% C.I.=1.12-11.99, **p=0.03**)) for the *857G>A (NAT2*7*) polymorphism. However, in the case of females, the *481C>T (NAT2*5B*) polymorphism did not reveal any association towards risk for lung cancer in any of the models applied.

Table 5.31 a): Relationship of different *NAT2* polymorphisms based on gender (males)

MALES									
Genotype rs1799930 *590G>A (NAT2*6)*	**Controls (459) N(%)**	**Cases (458) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Controls (459) N(%)**	**Cases (458) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	188 (40.96)	190 (41.48)	1.00 (Reference)		*NAT2 *4/*4*	140(30.50)	199 (43.45)	1.00 (Reference)	
*NAT2 *4/*6*	204 (44.44)	207 (45.20)	1.02 (0.77-1.35)	0.88	*NAT2 *4/*5B*	235(51.20)	209 (45.63)	0.61 (0.46-0.82)	**0.0009**
*NAT2 *6/*6*	67 (14.60)	61 (13.32)	0.89 (0.60-1.34)	0.59	*NAT2 *5B/*5B*	84 (18.30)	50 (10.92)	0.40 (0.26-0.61)	**<0.0001**
Dominant model					**Dominant model**				
*NAT2 *4/*4*	188 (40.96)	190 (41.48)	1.00 (Reference)		*NAT2 *4/*4*	140 (30.50)	199 (43.45)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	271 (59.04)	268 (58.52)	0.99 (0.76-1.28)	0.92	*NAT2*4/*5B + *5B/*5B*	319 (69.50)	259 (56.55)	0.56 (0.42-0.73)	**<0.0001**
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	392 (85.40)	397 (86.68)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	375 (81.70)	408 (89.08)	1.00 (Reference)	
*NAT2 *6/*6*	67 (14.60)	61 (13.32)	0.60 (0.42-0.86)	**0.005**	*NAT2 *5B/*5B*	84 (18.30)	50 (10.92)	0.54 (0.37-0.79)	**0.0014**
Genotype rs1799931 *857G>A (NAT2*7)*	**Controls (459) N(%)**	**Cases (458) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Controls (459) N(%)**	**Cases (458) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	212 (46.19)	189 (41.27)	1.00 (Reference)		*NAT2 *4/*4*	221 (48.15)	241 (52.62)	1.00 (Reference)	
*NAT2 *4/*7*	190 (41.39)	221 (48.25)	1.32 (1.00-1.74)	**0.048**	*NAT2 *4/*5C*	181 (39.43)	160 (34.93)	0.81 (0.61-1.07)	0.14
*NAT2 *7/*7*	57 (12.42)	48 (10.48)	0.93 (0.60-1.44)	0.75	*NAT2 *5C/*5C*	57 (12.42)	57 (12.45)	0.89 (0.59-1.35)	0.59
Dominant model					**Dominant model**				
*NAT2 *4/*4*	212 (46.19)	189 (41.27)	1.00 (Reference)		*NAT2 *4/*4*	221 (48.15)	241 (52.62)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	247 (53.81)	269 (58.73)	1.24 (0.95-1.61)	0.11	*NAT2 *4/*5C + *5C/*5C*	238 (51.85)	217 (47.38)	0.83 (0.64-1.07)	0.15
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	402 (87.58)	410 (89.52)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	402 (87.58)	401 (87.55)	1.00 (Reference)	
*NAT2 *7/*7*	57 (12.42)	48 (10.48)	0.83 (0.55-1.25)	0.38	*NAT2 *5C/*5C*	57 (12.42)	57 (12.45)	0.98 (0.66-1.45)	0.92

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, smoking. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.31 b): Relationship of different *NAT2* polymorphisms based on gender (females)

FEMALES									
Genotype rs1799930 *590G>A (NAT2*6)*	**Controls (91) N(%)**	**Cases (92) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Controls (91) N(%)**	**Cases (92) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	33 (36.26)	36 (39.13)	1.00 (Reference)		*NAT2 *4/*4*	44 (48.35)	39 (42.39)	1.00 (Reference)	
*NAT2 *4/*6*	41 (45.05)	50 (54.35)	1.29 (0.65-2.55)	0.47	*NAT2 *4/*5B*	37 (40.66)	39 (42.39)	1.19 (0.61-2.35)	0.61
*NAT2 *6/*6*	17 (18.68)	6 (6.52)	0.30 (0.09-0.99)	**0.04**	*NAT2 *5B/*5B*	10 (10.99)	14 (15.22)	1.37 (0.49-3.80)	0.54
Dominant model					**Dominant model**				
*NAT2 *4/*4*	33 (36.26)	36 (39.13)	1.00 (Reference)		*NAT2 *4/*4*	44 (48.35)	39 (42.39)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	58 (63.74)	56 (60.87)	1.03 (0.53-1.99)	0.93	*NAT2 *4/*5B + *5B/*5B*	47 (51.65)	53 (57.61)	1.37 (0.49-3.80)	0.54
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	74 (81.32)	86 (93.48)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	81 (89.01)	78 (84.78)	1.00 (Reference)	
*NAT2 *6/*6*	17 (18.68)	6 (6.52)	0.06 (0.02-0.16)	**<0.0001**	*NAT2 *5B/*5B*	10 (10.99)	14 (15.22)	1.27 (0.48-3.31)	0.63
Genotype rs1799931 *857G>A (NAT2*7)*	**Controls (91) N(%)**	**Cases (92) N(%)**	**AOR (95% C.I.)[a]**		**Genotype rs1208 *803A>G (NAT2*5C)***	**Controls (91) N(%)**	**Cases (92) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	52 (57.14)	33 (35.87)	1.00 (Reference)		*NAT2 *4/*4*	46 (50.55)	40 (43.48)	1.00 (Reference)	
*NAT2 *4/*7*	33 (36.26)	48 (52.17)	2.59 (1.30-5.16)	**0.007**	*NAT2 *4/*5C*	38 (41.76)	36 (39.13)	1.15 (0.57-2.29)	0.69
*NAT2 *7/*7*	6 (6.60)	11 (11.96)	3.66(1.12-11.99)	**0.03**	*NAT2 *5C/*5C*	7 (7.69)	16 (17.39)	2.82 (0.98-8.08)	**0.05**
Dominant model					**Dominant model**				
*NAT2 *4/*4*	52 (57.14)	33 (35.87)	1.00 (Reference)		*NAT2 *4/*4*	46 (50.55)	40 (43.48)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	39 (42.86)	59 (64.13)	2.72 (1.40-5.26)	**0.003**	*NAT2 *4/*5C + *5C/*5C*	45 (49.45)	52 (56.52)	1.41 (0.74-2.69)	0.29
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	85 (93.40)	81 (88.04)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	84 (92.31)	76 (82.61)	1.00 (Reference)	
*NAT2 *7/*7*	6 (6.60)	11 (11.96)	2.11 (0.70-6.34)	0.18	*NAT2 *5C/*5C*	7 (7.69)	16 (17.39)	2.62 (0.97-7.10)	0.05

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, smoking. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.3.9. Combinatorial analysis of the risk associated with different polymorphic sites of *NAT2* gene

Further, to investigate SNP to SNP interactions, the genotypic combinations of the various *NAT2* genotypes were analyzed to determine whether there is a significant difference in frequency between cases and controls. The analysis included subjects with wild-type genotype for all the four SNPs and heterozygous and mutant combined genotype for the four polymorphisms of *NAT2* genotype. The genotypic frequency of these combinations and AOR, along with 95% C.I.s, are tabulated in Table 5.32. For *NAT2* polymorphic sites, we combined both the heterozygous and mutant genotypes as a single group and compared it with the wild-type genotype. The patients possessing (*NAT2*4/*6+ NAT2*6/*6*) + (*NAT2*4/*7+ NAT2*7/*7*) + (*NAT2*4/*5B+ NAT2*5B/*5B*) and (*NAT2*4/*5C+NAT2*5C/*5C*) genotypic combinations did not exhibit any significant association towards risk for lung cancer (AOR=0.91, 95% C.I.=0.47-1.74). Further, the stratification was done based on histology, smoking status, and gender to evaluate the association towards lung cancer. The results depict that the patients with (*NAT2*4/*6+ NAT2*6/*6*) + (*NAT2*4/*7+ NAT2*7/*7*) + (*NAT2*4/*5B+NAT2*5B/*5B*) and (*NAT2*4/*5C+NAT2*5C/*5C*) genotypes did not confer any modulation towards risk for ADCC, SQCC, and SCLC respectively. In addition, the stratified analysis based on smoking status and gender also did not reveal any association of developing lung cancer with the above type of combinations of the *NAT2* polymorphism (Table 5.32).

Table 5.32: Combinatorial analysis of the risk associated with different polymorphic sites of *NAT2* gene

	OVERALL				**ADCC**			
	Controls (92) N(%)	**Cases (97) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (92) N(%)**	**Cases (38) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
*NAT2*4/*4*	24 (26.09)	27 (27.84)	1.00 (Reference)		24 (26.09)	10 (26.32)	1.00 (Reference)	
*(NAT2*4/*6+ NAT2*6/*6) +(NAT2*4/*7+ NAT2*7/*7) +(NAT2*4/*5+ NAT2*5/*5) +(NAT2*4/*5+ NAT2*5/*5)*	68 (73.91)	70 (72.16)	0.91 (0.47-1.74)	0.77	68 (73.91)	28 (73.68)	1.09 (0.44-2.68)	0.85
	SQCC				**SCLC**			
	Controls (92) N(%)	**Cases (37) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (92) N(%)**	**Cases (22) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
*NAT2*4/*4*	24 (26.09)	10 (27.03)	1.00 (Reference)		24 (26.09)	7 (31.82)	1.00 (Reference)	
*(NAT2*4/*6+ NAT2*6/*6) +(NAT2*4/*7+ NAT2*7/*7) +(NAT2*4/*5+ NAT2*5/*5) +(NAT2*4/*5+ NAT2*5/*5)*	68 (73.91)	27 (72.97)	0.72 (0.28-1.80)	0.48	68 (73.91)	15 (68.18)	0.64 (0.22-1.90)	0.42
	SMOKERS				**NON-SMOKERS**			
	Controls (73) N(%)	**Cases (80) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (19) N(%)**	**Cases (17) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
*NAT2*4/*4*	16 (21.92)	23 (28.75)	1.00 (Reference)		8 (42.11)	4 (23.53)	1.00 (Reference)	
*(NAT2*4/*6+ NAT2*6/*6) +(NAT2*4/*7+ NAT2*7/*7) +(NAT2*4/*5+ NAT2*5/*5) +(NAT2*4/*5+ NAT2*5/*5)*	57 (78.08)	57 (71.25)	0.69 (0.33-1.45)	0.33	11 (57.89)	13 (76.47)	3.64 (0.69-19.10)	0.13
	LIGHT SMOKERS				**HEAVY SMOKERS**			
	Controls (46) N(%)	**Cases (55) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (27) N(%)**	**Cases (25) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**
NAT2*4/*4	9 (19.57)	17 (30.91)	1.00 (Reference)		7 (25.93)	6 (24.00)	1.00 (Reference)	
(NAT2*4/*6+ NAT2*6/*6) +(NAT2*4/*7+ NAT2*7/*7) +(NAT2*4/*5+ NAT2*5/*5) +(NAT2*4/*5+ NAT2*5/*5)	37 (80.43)	38 (69.09)	0.50 (0.19-1.30)	0.16	20 (74.07)	19 (76.00)	1.02 (0.27-3.84)	0.97
	Males				**Females**			
	Controls (79) N(%)	**Cases (79) N(%)**	**AOR (95% C.I.) [a]**	***p*[b]**	**Controls (13) N(%)**	**Cases (18) N(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
*NAT2*4/*4*	18 (22.78)	23 (29.11)	1.00 (Reference)		6 (46.15)	4 (22.22)	1.00 (Reference)	
*(NAT2*4/*6+ NAT2*6/*6) +(NAT2*4/*7+ NAT2*7/*7) +(NAT2*4/*5+ NAT2*5/*5) +(NAT2*4/*5+ NAT2*5/*5)*	61 (77.22)	56 (70.89)	0.71 (0.35-1.46)	0.36	7 (53.85)	14 (77.78)	3.27 (0.58-18.45)	0.18

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender and smoking status. (b)-Two-sided χ2 test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Another method used for studying the interactions was based on risk alleles. None of the alleles exhibited any risk towards lung cancer compared to the 0 risk allele (the group with no variant allele), which was used as the reference (Table 5.33). Further, the analysis was carried out based on histological subtypes, gender, and smoking status. None of the risk alleles of the different *NAT2* polymorphic sites was associated with the ADCC, SQCC, and SCLC subtype of lung cancer. In addition, stratification based on smoking status depicted that the non-smokers with five risk alleles for *NAT2* genotype exhibited a 5-fold increased risk of developing lung cancer (AOR=5.10, 95% C.I.=1.13-23.04, **p**=**0.03**).

Further on basis of pack-years, light smokers with three (AOR=0.35, 95% C.I.=0.14-0.85, p=**0.02**), four (AOR=0.35, 95% C.I.=0.14-0.85, **p**=**0.02**), and five (AOR=0.35, 95% C.I.=0.12-0.99, **p**=**0.04**) risk alleles indicated a decreased lung cancer risk as indicated in Table 5.33. On the contrary, in lung cancer, subjects who were heavy smokers and with one risk allele displayed approximately 4-fold risk towards lung cancer development (AOR=3.96, 95% C.I.=1.04-15.08, **p**=**0.04**). However, stratification analysis based on gender did not exhibit any association towards modulation of lung cancer susceptibility (Table 5.33).

Table 5.33: Combined effects of *NAT2* variants on the risk of lung cancer

OVERALL					ADCC			
Number of Risk alleles[$] (Deduced phenotype)	**Control (550), N(%)**	**Cases (550), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (550), N(%)**	**Cases (200), N(%)**	**AOR (95% C.I.)**	**p-value**
0 (rapid)	24 (4.36)	27 (4.91)	1.0 (Reference)		24 (4.36)	10 (5.00)	1.0 (Reference)	
1 (intermediate)	66 (12.00)	81 (14.73)	1.02 (0.53-1.97)	0.95	66 (12.00)	22 (11.00)	0.79 (0.33-1.92)	0.61
2 (slow)	132 (24.00)	137 (24.91)	0.92 (0.50-1.69)	0.79	132 (24.00)	48 (24.00)	0.87 (0.38-1.98)	0.74
3 (slow)	151 (27.45)	138 (25.09)	0.81 (0.45-1.48)	0.49	151 (27.45)	52 (26.00)	0.86 (0.38-1.97)	0.73
4 (slow)	119 (21.64)	117 (21.27)	0.90 (0.49-1.66)	0.73	119 (21.64)	48 (24.00)	1.20 (0.50-2.89)	0.67
5 (slow)	44 (8.00)	44 (8.00)	0.87 (0.43-1.77)	0.70	44 (8.00)	17 (8.50)	1.26 (0.46-3.44)	0.66
6 (slow)	10 (1.82)	4 (0.73)	0.40 (0.10-1.49)	0.17	10 (1.82)	1 (0.50)	0.23 (0.02-2.14)	0.20
7 (slow)	2 (0.36)	1 (0.18)	0.40 (0.03-4.99)	0.48	2 (0.36)	1 (0.50)	0.95(0.06-14.09)	0.97
8 (slow)	2 (0.36)	1 (0.18)	0.45 (0.03-5.78)	0.54	2 (0.36)	1(0.50)	1.15 (0.09-14.98)	0.91
SQCC					**SCLC**			
Number of Risk alleles[$]	**Control (550), N(%)**	**Cases (216), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (550), N(%)**	**Cases (127), N(%)**	**AOR (95% C.I.)**	**p-value**
0	24 (4.36)	10 (4.63)	1.0 (Reference)		24 (4.36)	7 (5.51)	1.0 (Reference)	
1	66 (12.00)	36 (16.67)	1.24 (0.51-2.99)	0.64	66 (12.00)	21 (16.54)	0.87 (0.29-2.55)	0.79
2	132 (24.00)	59 (27.31)	1.07 (0.47-2.43)	0.88	132 (24.00)	29 (22.83)	0.71 (0.25-1.99)	0.52
3	151 (27.45)	51 (23.61)	0.75 (0.32-1.73)	0.49	151 (27.45)	35 (27.56)	0.78 (0.30-2.06)	0.62
4	119 (21.64)	45 (20.83)	0.79 (0.34-1.84)	0.58	119 (21.64)	20 (15.75)	0.50 (0.18-1.42)	0.19
5	44 (8.00)	14 (6.48)	0.48 (0.17-1.35)	0.16	44 (8.00)	13 (10.24)	0.95 (0.30-2.99)	0.92
6	10 (1.82)	1 (0.46)	0.27 (0.03-2.67)	0.26	10 (1.82)	2 (1.57)	0.92 (0.14-6.05)	0.93
7	2 (0.36)	0 (0.00)	-	-	2 (0.36)	0 (0.00)	-	-
8	2 (0.36)	0 (0.00)	-	-	2 (0.36)	0 (0.00)	-	-
SMOKERS					**NON-SMOKERS**			
Number of Risk alleles[$]	**Control (425), N(%)**	**Cases (444), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (125), N(%)**	**Cases (106), N(%)**	**AOR (95% C.I.)**	**p-value**

0	16 (3.76)	23 (5.18)	1.0 (Reference)		8 (6.40)	4 (3.77)	1.0 (Reference)	
1	48 (11.29)	72 (17.38)	1.02 (0.48-2.16)	0.96	18 (14.40)	9 (8.49)	1.05 (0.24-4.69)	0.94
2	95 (22.35)	111 (25.00)	0.85 (0.42-1.72)	0.65	37 (29.60)	26 (24.53)	1.33 (0.34-5.18)	0.68
3	114 (26.82)	108 (24.32)	0.69 (0.34-1.39)	0.29	37 (29.60)	30 (28.30)	1.75 (0.45-6.82)	0.42
4	102 (24.00)	88 (19.82)	0.60 (0.29-1.22)	0.16	17 (13.60)	29 (27.36)	5.10 (1.13-23.04)	**0.034**
5	41 (9.65)	38 (8.56)	0.62 (0.28-1.39)	0.25	3 (2.40)	6 (5.66)	3.76 (0.56-25.35)	0.17
6	6 (1.41)	3 (0.68)	0.42 (0.09-2.02)	0.28	4 (3.20)	1 (0.94)	0.55 (0.04-7.18)	0.65
7	2 (0.47)	0 (0.00)	-	-	0 (0.00)	1 (0.94)	-	-
8	1 (0.24)	1 (0.23)	0.67 (0.04-12.34)	0.79	1 (0.80)	0 (0.00)	-	-
	LIGHT-SMOKERS				**HEAVY-SMOKERS**			
Number of Risk alleles[$]	**Control (300), N(%)**	**Cases (248), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (125), N(%)**	**Cases (196), N(%)**	**AOR (95% C.I.)**	**p-value**
0	9 (3.00)	17 (6.85)	1.0 (Reference)		7 (5.60)	6 (3.06)	1.0 (Reference)	
1	36 (12.00)	38 (15.32)	0.49 (0.18-1.28)	0.15	12 (9.60)	34 (17.35)	3.96 (1.04-15.08)	**0.043**
2	71 (23.67)	62 (20.67)	0.47 (0.19-1.15)	0.10	24 (19.20)	49 (25.00)	3.13 (0.82-11.85)	0.09
3	78 (26.00)	54 (21.78)	0.35 (0.14-0.85)	**0.02**	36 (28.80)	54 (27.55)	2.11 (0.62-7.18)	0.23
4	75 (25.00)	53 (21.37)	0.35 (0.14-0.85)	**0.02**	27 (21.60)	35 (17.86)	1.50 (0.43-5.26)	0.53
5	26 (8.67)	20 (8.06)	0.35 (0.12-0.99)	**0.049**	15 (12.00)	18 (9.18)	1.45 (0.30-6.97)	0.64
6	3 (1.00)	3 (1.21)	0.95 (0.13-6.90)	0.96	3 (2.40)	0 (0.00)	-	-
7	2 (0.67)	0 (0.00)	-	-	0 (0.00)	0 (0.00)	-	-
8	0 (0.00)	1 (0.40)	-	-	1 (0.80)	0 (0.00)	-	-
	MALES				**FEMALES**			
Number of Risk alleles[$]	**Control (459), N(%)**	**Cases (458), N(%)**	**AOR (95% C.I.)**	**p-value**	**Control (91), N(%)**	**Cases (92), N(%)**	**AOR (95% C.I.)**	**p-value**
0	18 (3.92)	23 (5.02)	1.0 (Reference)		6 (6.59)	4 (4.35)	1.0 (Reference)	
1	52 (11.33)	70 (15.28)	1.02 (0.50-2.10)	0.94	14 (15.38)	11 (11.96)	0.95 (0.16-5.55)	0.96
2	107 (23.31)	117 (25.55)	0.86 (0.44-1.68)	0.65	25 (27.47)	20 (21.74)	1.74 (0.31-9.84)	0.53

3	121 (26.36)	116 (25.33)	0.75 (0.38-1.47)	0.40	30 (32.97)	22 (23.91)	1.68 (0.29-9.63)	0.56
4	108 (23.53)	91 (19.87)	0.66 (0.33-1.30)	0.23	11 (12.09)	26 (28.26)	4.68 (0.91-24.10)	0.06
5	43 (9.37)	36 (7.86)	0.65 (0.30-1.40)	0.27	1 (1.10)	8 (8.70)	18.57(0.78-442.28)	0.07
6	7 (1.53)	3 (0.65)	0.33 (0.07-1.52)	0.16	3 (3.30)	1 (1.09)	1.05 (0.06-17.36)	0.97
7	2 (0.43)	1 (0.22)	0.33 (0.02-4.34)	0.40	0 (0.00)	0 (0.00)	-	-
8	1 (0.22)	1 (0.22)	0.72 (0.04-12.66)	0.82	1 (1.10)	0 (0.00)	-	-

$ 0 risk allele: GG+GG+CC+AA; 1 risk allele: GA+GG+CC+AA, GG+GA+CC+AA, GG+GG+CT+AA, GG+GG+CC+AG; 2 risk allele: GA+GA+CC+AA, GA+GG+CT+AA, GA+GG+CC+AG, GG+GA+CT+AA, GG+GA+CC+AG, GG+GG+CT+AG, AA+GG+CC+AA, GG+AA+CC+AA, GG+GG+TT+AA, GG+GG+CC+GG; 3 risk allele: GA+GA+CT+AA, GA+GA+CC+AG, GA+GG+CT+AG, GG+GA+CT+AG, AA+GA+CC+AA, AA+GG+CT+AA, AA+GG+CC+AG, GA+AA+CC+AA, GG+AA+CT+AA, GG+AA+CC+AG, GG+GG+TT+AG, GA+GG+CT+AA, GA+GG+TT+AA, GG+GA+TT+AA, GG+GG+CT+GG, GG+GA+CC+GG, GA+GG+CC+GG; 4 risk allele: GA+GA+CT+AG, AA+AA+CC+AA, AA+GG+TT+AA, AA+GG+CC+GG, GG+AA+TT+AA, GG+AA+CC+GG, GG+GG+TT+GG, GA+GA+TT+AA, GG+GA+TT+AG, GA+GG+TT+AG, GG+GA+CT+GG, GA+GG+CT+GG, GA+GA+CC+GG, AA+GA+CT+AA, AA+GA+CC+AG, AA+GG+CT+AG, GG+AA+CT+AG, GA+AA+CC+AG, GA+AA+CT+AA; 5 risk allele: AA+AA+CT+AA, AA+GA+TT+AA, AA+GA+CC+GG, AA+AA+CC+AG, AA+GG+TT+AG, AA+GG+CT+GG, GA+AA+TT+AA, GA+AA+CC+GG, GG+AA+TT+AG, GG+AA+CT+GG, GG+GA+TT+GG, GA+GG+TT+GG, AA+GA+CT+AG, GA+AA+CT+AG, GA+GA+TT+AG, GA+GA+CT+GG; 6 risk allele: AA+AA+CT+AG, AA+GA+CC+AG, AA+GA+CT+GG, GA+AA+TT+AG, GA+AA+CT+GG, GA+GA+TT+GG, AA+AA+TT+AA, GG+AA+TT+GG, AA+GG+TT+GG, AA+AA+CC+GG; 7 risk allele: AA+AA+TT+AG, AA+AA+CT+GG, AA+GA+TT+GG, GA+AA+TT+GG; 8 risk allele: AA+AA+TT+GG.

5.6.3.10. Association of the *NAT2* polymorphism & clinic-pathological characteristics

The effect of *NAT2* variant SNPs on the various clinic-pathological features such as tumor stage, primary tumor extension, lymph node invasion, and distant metastasis was also calculated. The patients were stratified into stage III and stage IV patients (Table 5.34 a)). The subjects were also categorized based on tumor extension into T3 and T4 (Table 5.34 b)). Based on lymph node invasion, patients were divided into Nx+N0+N1 and N2+N3+N4 groups (Table 5.34 c)). Further, to classify the patients on the distant metastatic status, they were stratified into M0 and M1 groups (Table 5.34 d)). There was no relationship between the *NAT2* polymorphism in patients in any of the three applied genetic models regarding clinical stage, tumor size, and distant metastasis.

However, on the contrary, it was observed that lung cancer patients who were carrying the single variant alleles (*NAT2*4/*6*) for the *NAT2*, 590G>A (*NAT2*6*) polymorphism a decreased odds was observed for lymph node invasion, in respect to subjects who were homozygous for the wild type genotype (*NAT2*4/*4*), especially under the co-dominant model (AOR=0.53; 95% C.I.=0.33-0.85; **p**=**0.009**). Further, using the dominant model, the combined genotype (*NAT2*4/*6 + *6/*6*) also depicted a reduced odds of lymph node invasion (AOR=0.54, 95% C.I.=0.34-0.85, **p**=**0.008**) (Table 5.34 c)). In the case of the *803A>G (NAT2*5C*) polymorphism, it was observed that lung cancer patients who were carrying both the mutant alleles for the *803A>G (NAT2*5C)* polymorphism expressed a 3-fold higher odds of lymph node invasion using the co-dominant model (AOR=3.22, 95% C.I.=1.39-7.42, **p**=**0.006**) concerning lung cancer subjects who were harboring both the wild type alleles for the*803A>G (NAT2*5C*) polymorphism (Table 5.34 c)). Similarly, the recessive model also exhibited an odd of 2.8 for the lymph node invasion (AOR=2.78, 95% C.I.=1.23-6.29, **p**=**0.014**) for subjects carrying the mutant genotype in respect to combined genotype *(NAT2 *4/*5C + *5C/*5C*). Furthermore, the combinatorial results depicted that the combined genotype (*NAT2 *4/*5C + *5C/*5C*) for *803A>G* (*NAT2*5C*) polymorphism using the dominant model exhibit a 1.78-fold odds of lymph node invasion (AOR=1.78, 95% C.I.=1.15-2.75, p=**0.009**) (Table 5.34 c)).

Table 5.34 a): Relationship of different *NAT2* genotypes with the clinical-pathological parameters (clinical stage)

CLINICAL STAGE									
Genotype	**III**	**IV**	**AOR**	***p*[b]**	**Genotype**	**III**	**IV**	**AOR**	***p*[b]**
590G>A (NAT2*6)	**N= 229(%)**	**N= 247(%)**	**(95% C.I.)[a]**		***481C>T (NAT2*5B)***	**N= 229(%)**	**N= 247(%)**	**(95% C.I.)[a]**	
Codominant model					**Codominant model**				
*NAT2 *4/*4*	102 (44.54)	95 (38.46)	1.00 (Reference)		*NAT2 *4/*4*	100 (43.67)	101 (40.89)	1.00 (Reference)	
*NAT2 *4/*6*	99 (43.23)	119 (48.18)	1.24 (0.83-1.85)	0.30	*NAT2 *4/*5B*	103 (44.98)	116 (46.96)	1.10 (0.74-1.64)	0.64
*NAT2 *6/*6*	28 (12.23)	33 (13.36)	1.20 (0.65-2.21)	0.56	*NAT2 *5B/*5B*	26 (11.35)	30 (12.15)	1.10 (0.58-2.04)	0.79
Dominant model					**Dominant model**				
*NAT2 *4/*4*	102 (44.54)	95 (38.46)	1.00 (Reference)		*NAT2 *4/*4*	100 (43.67)	101 (40.89)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	127 (55.46)	152 (61.54)	1.23 (0.84-1.80)	0.29	*NAT2*4/*5B +*5B/*5B*	129 (56.33)	146 (59.11)	1.10 (0.75-1.62)	0.61
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	201 (87.77)	214 (86.64)	1.00 (Reference)		*NAT2*4/*4 +*4/*5B*	203 (88.65)	217 (87.85)	1.00 (Reference)	
*NAT2 *6/*6*	28 (12.23)	33 (13.36)	1.10 (0.62-1.93)	0.74	*NAT2 *5B/*5B*	26 (11.35)	30 (12.15)	1.06 (0.59-1.90)	0.85
Genotype	**III**	**IV**	**AOR**		**Genotype**	**III**	**IV**	**AOR**	***p*[b]**
857G>A (NAT2*7)	**N= 229(%)**	**N= 247(%)**	**(95% C.I.)[a]**		***803A>G (NAT2*5C)***	**N= 229(%)**	**N= 247(%)**	**(95% C.I.)[a]**	
Codominant model					**Codominant model**				
*NAT2 *4/*4*	94 (41.05)	96 (38.87)	1.00 (Reference)		*NAT2 *4/*4*	116 (50.65)	121 (48.99)	1.00 (Reference)	
*NAT2 *4/*7*	112 (48.91)	126 (51.01)	1.07 (0.72-1.59)	0.75	*NAT2 *4/*5C*	82 (35.81)	92 (37.24)	1.05 (0.70-1.59)	0.80
*NAT2 *7/*7*	23 (10.04)	25 (10.12)	0.98 (0.50-1.91)	0.96	*NAT2 *5C/*5C*	31 (13.54)	34 (13.77)	1.00 (0.56-1.77)	0.99
Dominant model					**Dominant model**				
*NAT2 *4/*4*	94 (41.05)	96 (38.87)	1.00 (Reference)		*NAT2 *4/*4*	116 (50.65)	121 (48.99)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	135 ()	151 ()	1.05 (0.71-1.54)	0.81	*NAT2 *4/*5C + *5C/*5C*	113 (49.34)	126 (51.01)	1.03 (0.71-1.50)	0.87
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	206 (89.96)	222 (89.88)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	198 (86.46)	213 (86.23)	1.00 (Reference)	
*NAT2 *7/*7*	23 (10.04)	25 (10.12)	0.96 (0.52-1.79)	0.90	*NAT2 *5C/*5C*	31 (13.54)	34 (13.77)	0.96 (0.55-1.66)	0.88

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.34 b): Relationship of different *NAT2* genotypes with the clinical-pathological parameters (primary tumor extension)

PRIMARY TUMOR EXTENSION									
Genotype ***590G>A (NAT2*6)***	**T3 N= 115(%)**	**T4 N= 299(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Genotype** ***481C>T (NAT2*5B)***	**T3 N= 115(%)**	**T4 N= 299(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	43 (37.39)	125 (41.81)	1.00 (Reference)		*NAT2 *4/*4*	48 (41.74)	132 (44.15)	1.00 (Reference)	
*NAT2 *4/*6*	56 (48.70)	140 (46.82)	0.83 (0.52-1.33)	0.44	*NAT2 *4/*5B*	52 (45.22)	136 (45.48)	0.98 (0.61-1.55)	0.92
*NAT2 *6/*6*	16 (13.91)	34 (11.37)	0.74 (0.36-1.50)	0.40	*NAT2 *5B/*5B*	15 (13.04)	31 (10.37)	0.70 (0.34-1.44)	0.33
Dominant model					**Dominant model**				
*NAT2 *4/*4*	43 (37.39)	125 (41.81)	1.00 (Reference)		*NAT2 *4/*4*	48 (41.74)	132 (44.15)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	72 (62.61)	174 (58.19)	0.81 (0.52-1.26)	0.35	*NAT2 *4/*5B + *5B/*5B*	67 (58.26)	167 (55.85)	0.93 (0.60-1.45)	0.76
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	99 (86.09)	265 (88.63)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	100 (86.96)	268 (89.63)	1.00 (Reference)	
*NAT2 *6/*6*	16 (13.91)	34 (11.37)	1.10 (0.63-1.93)	0.74	*NAT2 *5B/*5B*	15 (13.04)	31 (10.37)	0.79 (0.41-1.54)	0.49
Genotype ***857G>A (NAT2*7)***	**T3 N= 115(%)**	**T4 N= 299(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Genotype** ***803A>G (NAT2*5C)***	**T3 N= 115(%)**	**T4 N= 299(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	52 (45.22)	108 (36.12)	1.00 (Reference)		*NAT2 *4/*4*	58 (50.43)	154 (51.51)	1.00 (Reference)	
*NAT2 *4/*7*	57 (49.56)	160 (53.51)	1.38 (0.88-2.17)	0.16	*NAT2 *4/*5C*	39 (33.91)	106 (35.45)	1.01 (0.62-1.63)	0.97
*NAT2 *7/*7*	6 (5.22)	31 (10.37)	2.41 (0.94-6.19)	0.06	*NAT2 *5C/*5C*	18 (15.65)	39 (13.04)	0.81 (0.43-1.54)	0.52
Dominant model					**Dominant model**				
*NAT2 *4/*4*	52 (45.22)	108 (36.12)	1.00 (Reference)		*NAT2 *4/*4*	58 (50.43)	154 (51.51)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	63 (54.78)	191 (63.88)	1.49 (0.96-2.32)	0.07	*NAT2 *4/*5C + *5C/*5C*	57 (49.57)	145 (48.49)	0.96 (0.62-1.47)	0.84
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	109 (94.78)	268 (89.63)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	97 (84.35)	260 (86.96)	1.00 (Reference)	
*NAT2 *7/*7*	6 (5.22)	31 (10.37)	2.05 (0.83-5.07)	0.12	*NAT2 *5C/*5C*	18 (15.65)	39 (13.04)	0.82 (0.44-1.51)	0.52

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.34 c): Relationship of different *NAT2* genotypes with the clinical-pathological parameters (lymph node invasion)

LYMPH NODE INVASION									
Genotype	**Nx+N0+N1**	**N2+N3+N4**	**AOR**	***p*[b]**	**Genotype**	**Nx+N0+N1**	**N2+N3+N4**	**AOR**	***p*[b]**
590G>A (NAT2*6)	**N= 110(%)**	**N= 390(%)**	**(95% C.I.)[a]**		***481C>T (NAT2*5B)***	**N= 110(%)**	**N= 390(%)**	**(95% C.I.)[a]**	
Codominant model					**Codominant model**				
*NAT2 *4/*4*	33 (30.00)	171 (43.85)	1.00 (Reference)		*NAT2 *4/*4*	48 (43.64)	165 (42.31)	1.00 (Reference)	
*NAT2 *4/*6*	62 (56.36)	173 (44.36)	0.53 (0.33-0.85)	**0.009**	*NAT2 *4/*5B*	48 (43.64)	179 (45.90)	1.10 (0.69-1.73)	0.69
*NAT2 *6/*6*	15 (13.64)	46 (11.79)	0.56 (0.28-1.15)	0.11	*NAT2 *5B/*5B*	14 (12.72)	46 (11.79)	0.94 (0.47-1.88)	0.87
Dominant model					**Dominant model**				
*NAT2 *4/*4*	33 (30.00)	171 (43.85)	1.00 (Reference)		*NAT2 *4/*4*	48 (43.64)	165 (42.31)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	77 (70.00)	219 (56.15)	0.54 (0.34-0.85)	**0.008**	*NAT2 *4/*5B + *5B/*5B*	62 (56.36)	225 (57.69)	1.07 (0.69-1.64)	0.76
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	95 (86.36)	344 (88.21)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	96 (87.27)	344 (88.21)	1.00 (Reference)	
*NAT2 *6/*6*	15 (13.64)	46 (11.79)	0.83 (0.44-1.56)	0.57	*NAT2 *5B/*5B*	14 (12.72)	46 (11.79)	0.93 (0.49-1.77)	0.82
Genotype	**Nx+N0+N1**	**N2+N3+N4**	**AOR**		**Genotype**	**Nx+N0+N1**	**N2+N3+N4**	**AOR**	***p*[b]**
857G>A (NAT2*7)	**N= 110(%)**	**N= 390(%)**	**(95% C.I.)[a]**		***803A>G (NAT2*5C)***	**N= 110(%)**	**N= 390(%)**	**(95% C.I.)[a]**	
Codominant model					**Codominant model**				
*NAT2 *4/*4*	47 (42.73)	148 (37.95)	1.00 (Reference)		*NAT2 *4/*4*	67 (60.91)	183 (46.92)	1.00 (Reference)	
*NAT2 *4/*7*	55 (50.00)	199 (51.03)	1.16 (0.74-1.81)	0.52	*NAT2 *4/*5C*	36 (32.73)	146 (37.44)	1.49 (0.94-2.36)	0.09
*NAT2 *7/*7*	8 (7.27)	43 (11.02)	1.70 (0.74-3.88)	0.21	*NAT2 *5C/*5C*	7 (6.36)	61 (15.64)	3.22 (1.39-7.42)	**0.006**
Dominant model					**Dominant model**				
*NAT2 *4/*4*	47 (42.73)	148 (37.95)	1.00 (Reference)		*NAT2 *4/*4*	67 (60.91)	183 (46.92)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	63 (57.27)	242 (62.05)	1.23 (0.79-1.89)	0.35	*NAT2 *4/*5C + *5C/*5C*	43 (39.09)	207 (53.08)	1.78 (1.15-2.75)	**0.0091**
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	102 (92.73)	347 (88.98)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	103 (93.64)	329 (84.36)	1.00 (Reference)	
*NAT2 *7/*7*	8 (7.27)	43 (11.02)	1.58 (0.72-3.46)	0.26	*NAT2 *5C/*5C*	7 (6.36)	61 (15.64)	2.78 (1.23-6.29)	**0.014**

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

Table 5.34 d): Relationship of different *NAT2* genotypes with the clinical-pathological parameters (metastasis)

METASTASIS									
Genotype ***590G>A (NAT2*6)***	**No (M0) N= 256(%)**	**Yes (M1) N= 246(%)**	**AOR (95% C.I.)[a]**	***p*[b]**	**Genotype** ***481C>T (NAT2*5B)***	**No (M0) N= 256(%)**	**Yes (M1) N= 246(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	110 (42.97)	95 (38.62)	1.00 (Reference)		*NAT2 *4/*4*	115 (44.92)	100 (40.65)	1.00 (Reference)	
*NAT2 *4/*6*	117 (45.70)	119 (48.37)	0.86 (0.58-1.28)	0.47	*NAT2 *4/*5B*	110 (42.97)	117 (47.56)	1.25 (0.85-1.85)	0.25
*NAT2 *6/*6*	29 (11.33)	32 (13.01)	0.80 (0.44-1.48)	0.49	*NAT2 *5B/*5B*	31 (12.11)	29 (11.79)	1.05 (0.58-1.91)	0.87
Dominant model					**Dominant model**				
*NAT2 *4/*4*	110 (42.97)	95 (38.62)	1.00 (Reference)		*NAT2 *4/*4*	115 (44.92)	100 (40.65)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	146 (62.14)	151 (57.03)	0.86 (0.59-1.24)	0.41	*NAT2 *4/*5B + *5B/*5B*	141 (55.08)	146 (59.35)	1.22 (0.84-1.77)	0.29
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	227 (88.67)	214 (86.99)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5B*	225 (87.89)	217 (88.21)	1.00 (Reference)	
*NAT2 *6/*6*	29 (11.33)	32 (13.01)	0.88 (0.50-1.54)	0.65	*NAT2 *5B/*5B*	31 (12.11)	29 (11.79)	0.97 (0.55-1.70)	0.92
Genotype ***857G>A (NAT2*7)***	**No (M0) N= 256(%)**	**Yes (M1) N= 246(%)**	**AOR (95% C.I.)[a]**		**Genotype** ***803A>G(NAT2*5C)***	**No (M0) N= 256(%)**	**Yes (M1) N= 246(%)**	**AOR (95% C.I.)[a]**	***p*[b]**
Codominant model					**Codominant model**				
*NAT2 *4/*4*	103 (40.23)	94 (38.21)	1.00 (Reference)		*NAT2 *4/*4*	134 (52.34)	118 (47.97)	1.00 (Reference)	
*NAT2 *4/*7*	128 (50.00)	126 (51.22)	1.05 (0.72-1.55)	0.79	*NAT2 *4/*5C*	86 (33.60)	96 (39.02)	0.78 (0.53-1.16)	0.23
*NAT2 *7/*7*	25 (9.77)	26 (10.57)	1.10 (0.58-2.09)	0.77	*NAT2 *5C/*5C*	36 (14.06)	32 (13.01)	1.04 (0.60-1.83)	0.88
Dominant model					**Dominant model**				
*NAT2 *4/*4*	103 (40.23)	94 (38.21)	1.00 (Reference)		*NAT2 *4/*4*	134 (52.34)	118 (47.97)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	153 (59.77)	152 (61.79)	1.06 (0.73-1.54)	0.76	*NAT2 *4/*5C +*5C/*5C*	122 (47.66)	128 (52.03)	0.85 (0.59-1.22)	0.38
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	231 (90.23)	220 (89.43)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	220 (85.94)	214 (86.99)	1.00 (Reference)	
*NAT2 *7/*7*	25 (9.77)	26 (10.57)	1.08 (0.59-1.97)	0.80	*NAT2 *5C/*5C*	36 (14.06)	32 (13.01)	1.17 (0.68-2.01)	0.56

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

5.6.3.11. Haplotype and LD analysis of *NAT2* variants and its association with lung cancer susceptibility

Table 5.35 summarizes the frequencies for different haplotypes obtained on analyzing the four polymorphic sites of the *NAT2* gene. Block structure and linkage disequilibrium (D') were analyzed with the SHEsis online program platform. The case and control frequency of more than 0.03 were evaluated for the haplotype blocks, and the rest of the haplotypes were not considered for their correlation with lung cancer predisposition. The global chi-square was calculated to compare frequencies among cases and controls (χ^2=32.96, df=10, and **p=0.0002**). The Hap block 2 (A A T A) showed a low frequency among both the cases and controls compared with other blocks. The hap block 2 showed a protective effect for lung cancer development (OR=0.31, 95* CI=0.16-0.60, **p=0.0002**). The hap block two comprised mutant alleles for *NAT2* 590G>A, *857G>A*, *481C>T*, and wild-type alleles for *NAT2 803A>G*. Similarly, Hap block 10 (G G T A) also exhibits decreased risk towards lung cancer (OR=0.69 95% C.I.=0.48-0.98, **p=0.038**) and has wild alleles for *NAT2 590G>A or 857G>A* and allele from the remaining two polymorphic sites were of mutant type. On the contrary, Hap5 (G A C A) showed a marginal risk towards lung cancer susceptibility (OR=1.58, 95% C.I.=1.18-2.11, p=**0.002**) (Table 5.35).

Table 5.35: Haplotype frequencies and lung cancer risk associated with *NAT2* haplotypes

Haplotype	Haplotype construction (590G>A, 857G>A, 481C>T, 803A>G)	Case frequency (%)	Control frequency (%)	OR (95% C.I.)[$]	p-value[#]
1	A A C A	74.17 (0.067)	70.43 (0.064)	1.07 (0.76-1.49)	0.71
2	A A T A	11.99 (0.011)	38.02 (0.034)	0.31 (0.16-0.60)	**0.0002**
3	A G C A	204.05 (0.185)	192.95 (0.175)	1.08 (0.87-1.35)	0.48
4	A G T A	41.41 (0.038)	56.90 (0.052)	0.72 (0.48-1.09)	0.12
5	G A C A	127.86 (0.116)	85.43 (0.078)	1.58 (1.18-2.11)	**0.002**
6	G A T A	29.56 (0.027)	36.15 (0.033)	0.82 (0.50-1.34)	0.43
7	G A T G	78.35 (0.071)	71.75 (0.065)	1.11 (0.79-1.55)	0.54
8	G G C A	212.82 (0.193)	193.15 (0.175)	1.14 (0.92-1.42)	0.24
9	G G C G	38.94 (0.035)	52.31 (0.047)	0.74 (0.48-1.13)	0.16
10	G G T A	56.14 (0.051)	79.97 (0.073)	0.69 (0.48-0.98)	**0.038**
11	G G T G	139.98 (0.127)	145.78 (0.132)	0.96 (0.75-1.24)	0.76

[$]ORs, 95% C.I.s, and their corresponding p-values were calculated by logistic regression analysis.

[#] Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

OR: Odds ratio.

Table 5.36 depicts the linkage disequilibrium (*D'*) and r^2 values for all the four polymorphic sites of *NAT2*. An average LD was observed in the case of *NAT2 590G>A* and *803A>G* (*D'*=0.522, r^2=0.071), and *481C>T* and *803A>G* (*D'*=0.523, r^2=0.202). Figure 5.21 represents the pairwise

LD. The block represents the D' values. The light pink color displays an average LD between the two variants, and the white color shows the minimum LD between two polymorphic sites.

Table 5.36: Showing pairwise *D'* (r2) values for the four polymorphic sites among cases and controls

***NAT2* SNP**	***857G>A***	***481C>T***	***803A>G***
590G>A	0.039 (0.000)	0.335 (0.039)	0.522 (0.071)
857G>A	-	0.036 (0.001)	0.071 (0.004)
481C>T	-	-	0.523 (0.202)

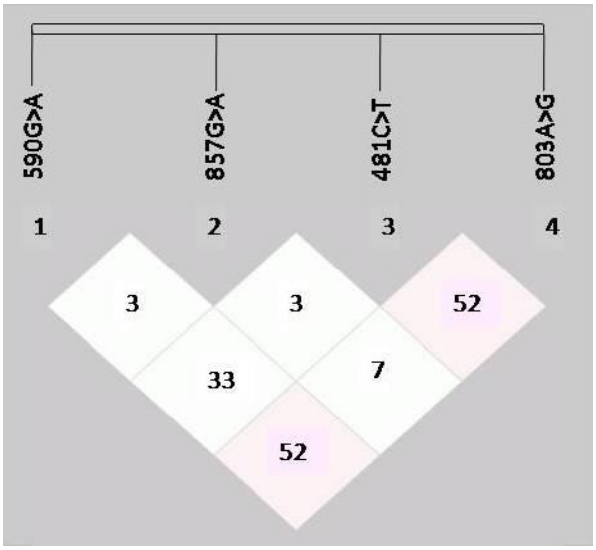


Figure 5.21: Showing pairwise linkage disequilibrium (*D'*) between the four variants of *NAT2* gene, wherein pink shows the average LD and white depicts the minimum LD between the polymorphic sites.

5.6.3.12. MDR analysis within *NAT2* gene

MDR uses a computational algorithm approach to investigate all the possible interactions between the different susceptibility contributing factors, including SNP's, other genes, and some environmental determinants. The results obtained from MDR analysis have been tabulated in Table 5.3746. The interpretation and selection of the best model are made based on the cross-validation consistency (CVC) and average prediction error (1-testing accuracy). The best model can be determined if it has maximum CVC and minimum prediction error or maximum testing accuracy. The MDR analysis was performed for both cases and controls, depicting the average CVC model and average prediction error (Table 5.37).

The best interaction model in this study (*857G>A(NAT2*7), 481C>T(NAT2*5B), 803A>G (NAT2*5C*)) has a CVC of 10/10 with a prediction error (PE) of 0.44 (p=0.22), but was not significant. Further, the interaction was also visualized using an interaction dendrogram

between *857G>A(NAT2*7), 481C>T(NAT2*5B), 803A>G(NAT2*5C*), and lung cancer risk (Figure 5.22 a)). The polymorphic variants (*857G>A(NAT2*7), 481C>T(NAT2*5B*) exhibited an intermediate interaction, and the variant *803A>G(NAT2*5C*) is linked with the green-colored line with other SNPs, thus depicting it has a weak interaction with other two SNPs.

Further, the MDR analysis was also carried out based on smoking status. The results depict that the best model for smokers was *481C>T(NAT2*5B*) with a CVC of 10/10 and prediction error of 0.43, but lacked statistical significance (p=0.17). The entropy dendrogram depicts that the variant *803A>G(NAT2*5C*) is linked with the green-colored line with other SNPs, thus suggesting that it is showing a weak interaction with these SNPs and an intermediate interaction was observed between (*857G>A(NAT2*7), 481C>T(NAT2*5B*)) (Figure 5.22 b)). Further, in the case of non-smokers, two models depict CVC of 10/10, but the one with the minor prediction error was chosen as the best model (*i.e.*, *857G>A(NAT2*7*), *803A>G(NAT2*5C*)), with PE of 0.40 (p=0.32), but here too this interaction is also not found to be significant (Table 5.37). The entropy dendrogram depicts that the SNPs *857G>A(NAT2*7), 803A>G(NAT2*5C*) were linked with the red-colored line, which depicts that these two are closely related and shows the synergy between two. Whereas the SNP *590G>A(NAT2*6*) exhibited a weak interaction with these two SNPs, as it is connected with a green colored line and is more evident from the length of the line connecting these SNPs (Figure 5.22 c)).

Table 5.37: Multifactor dimensionality reduction (MDR) analysis showing association of high-order interactions with lung cancer

No. of risk factors	Best interaction model	Cross-validation consistency (CVC)	Prediction error	*p*-value
	OVERALL			
1	481C>T(NAT2*5) (X3)	8/10	0.47	0.57
2	481C>T(NAT2*5), 803A>G(NAT2*5) (X3, X4))	9/10	0.48	0.74
3	857G>A(NAT2*7), 481C>T(NAT2*5), 803A>G(NAT2*5) (X2,X3, X4)	10/10	0.44	0.22
	SMOKERS			
1	481C>T(NAT2*5) (X3)	10/10	0.43	0.17
2	481C>T(NAT2*5), 803A>G(NAT2*5) (X3, X4))	9/10	0.47	0.54
3	857G>A(NAT2*7), 481C>T(NAT2*5), 803A>G(NAT2*5) (X2,X3, X4)	7/10	0.46	0.41
	NON-SMOKERS			
1	857G>A(NAT2*7) (X2)	10/10	0.41	0.40
2	857G>A(NAT2*7), 803A>G(NAT2*5) (X2, X4)	10/10	0.40	0.32
3	590G>A(NAT2*6), 857G>A(NAT2*7), 803A>G(NAT2*5) (X1, X2, X4)	7/10	0.49	0.93

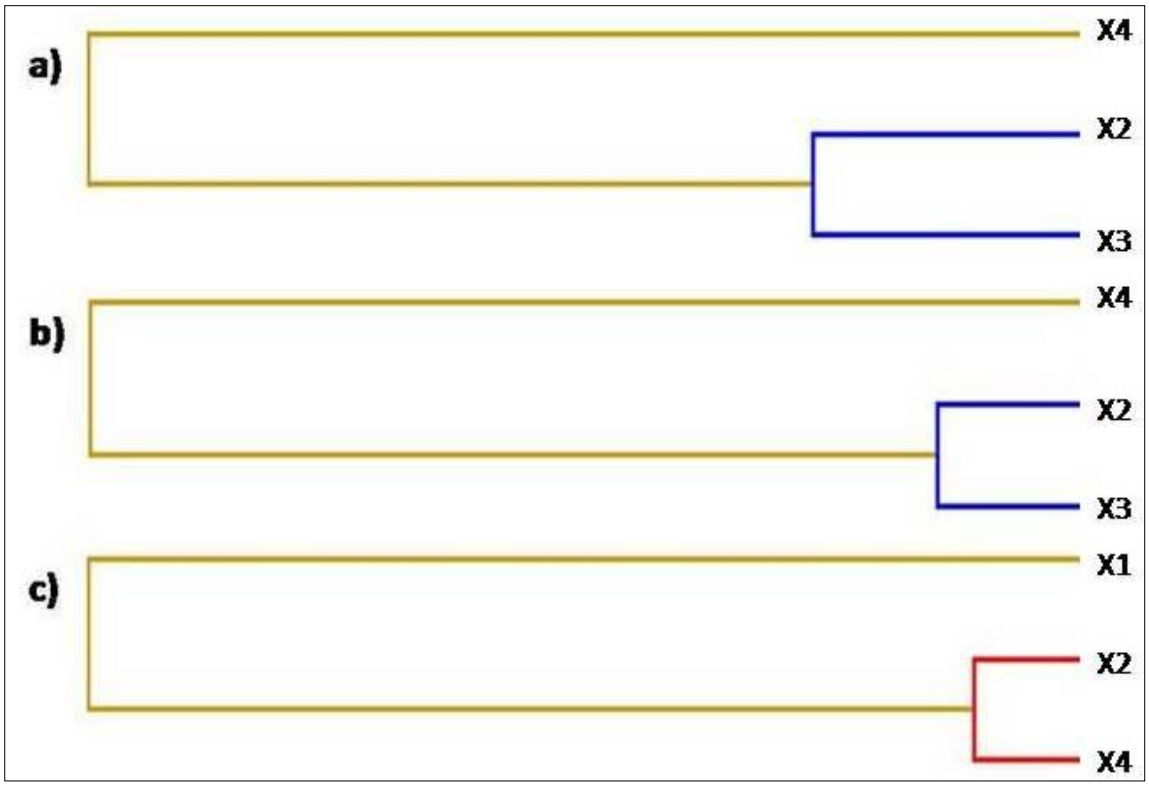


Figure 5.22: The multifactor dimensionality reduction (MDR) interaction dendrogram a) overall, b) smokers and c) non-smokers. (X1: NAT2*6 (590G>A), X2: NAT2*7 (857G>A), X3: NAT2*5 (481C>T), X4: NAT2*5 (803A>G))

5.6.3.13. Classification and Regression Tree (CART) analysis of *NAT2* variants

CART analysis is a tree-building technique that uses binary recursive partitioning to identify the different genotypic groups affecting the predisposition of lung cancer that cannot be derived from the traditional logistic regression analysis. Several variables are taken into account, and high-risk subgroups are identified. The study's outcome is a decision tree resulting from the data mining methodology used in the analysis. The most crucial factor is the root split, while the terminal nodes are the high-risk subgroups. The terminal node with the lowest case rate is considered a reference, and the OR was calculated respectively.

The initial split was formed by *481C>T* in the tree, suggesting it is the strongest risk factor for lung cancer. The terminal node 1 with *857G>A(M*) showed 2.20-fold lung cancer risk (OR=2.20; 95% C.I.=1.54-3.16; **p= 0.00002**). Terminal node 2 with genetic combination *803A>G(W), 857G>A(W), 481C>T(W*) also conferred a risk towards lung cancer susceptibility

(OR=1.68; 95% C.I.=1.14-2.49; **p**= **0.0087**). Subjects with genotypic combination of node 4 carrying *803A>G(M), 857G>A(M), 481C>T(M*) exhibited a marginal association towards lung cancer risk (OR=1.57; 95% C.I.=1.12-2.18; **p**=**0.008**) (Table 5.38, Figure 5.23). Further, subjects with the genotypic combination of terminal nodes 3 and 5 did not show any association with lung cancer, as shown in Table 5.47. We further divided the subjects based on histological subtypes. For exploring the association of *NAT2* variants with the risk of ADCC, the first split was observed at 857G>A, marking this polymorphism as the strongest risk factor for ADCC. Terminal node 1 *i.e.*, *481C>T(W*) had an almost 2-fold risk of having ADCC (OR=2.13; 95% C.I.=1.38-3.29; **p**=**0.0007**) (Table 5.38, Figure 5.24 a)). Terminal node 2 (*803A>G(M)/ 481C>T(M)/ 857G>A(M)*) also demonstrated an increased risk for lung cancer (OR=1.62; 95% C.I.=1.08-2.41; **p**=**0.018**). To observe the association of *NAT2* variants with risk of SQCC (Table 5.38, Figure 5.24 b)), the first split was observed at *481C>T* marking this polymorphism as the strongest risk factor SQCC. Terminal node 1 with genotypic combination *803A>G(W*) showed a 2-fold risk in SQCC subjects (OR=2.01; 95% C.I.=1.25-3.22; **p**=**0.004**). Furthermore, subjects with genotypic combination of node 5 carrying *590G>A(W)/ 857G>A(M)/ 481C>T(M*), exhibited an almost 2-fold increased risk in SQCC patients (OR=1.92; 95% C.I.=1.11-3.31; **p**=**0.02**). CART analysis was further applied on SCLC patients (Table 5.38); only a single split was observed at *481C>T*, marking it the strongest risk factor for SCLC. Only one terminal node was observed suggesting a 2-fold risk among SCLC patients (OR=2.16; 95% C.I.=1.41-3.32; **p**=**0.00043**) (Figure 5.24 c)).

5.6.3.14. CART analysis and its association with smoking status

The high order interaction was applied for smokers and non-smokers. In the case of smokers, there was only a single split at *481C>T* stating this polymorphism is the most substantial risk factor in smokers. Terminal node one, the only node observed, exhibited a marginal risk towards lung cancer (OR=1.28; 95% C.I.=0.98-1.67; p=0.07) but was not significant (Table 5.38, Figure 5.25 a)). Further, in the case of non-smokers, a single split was observed at *857G>A*, stating it to be the most decisive factor of risk in the case of non-smokers. Only a single terminal node was observed, revealing a 2-fold risk among subjects who were non-smokers with genotypic combination 857G>A(M) (OR=2.03; 95% C.I.=1.20-3.45; **p**=**0.008**) (Table 5.38, Figure 5.25 b)).

Table 5.38: Risk estimates based on classification and regression tree analysis (*NAT2* variants) overall, based on histology, and smoking status

Terminal Node	Genotype for each terminal node	Control (n)	Case (n)	Case rate[#]	OR (95% C.I.)	p-value[$]
Node 1	857G>A (M)	85	134	61.2	2.20 (1.54-3.16)	**0.00002**
Node 2	803A>G(W)/ 857G>A(W)/ 481C>T(W)	73	88	54.7	1.68 (1.14-2.49)	**0.0087**
Node 3	803A>G(M)/ 857G>A(W)/ 481C>T(W)	26	16	38.1	0.86 (0.44-1.68)	0.66
Node 4	803A>G(M)/ 857G>A(M)/ 481C>T(M)	134	150	52.8	1.57 (1.12-2.18)	**0.008**
Node 5	803A>G(W)/ 857G>A(M)/ 481C>T(M)	67	44	39.6	0.92 (0.59-1.44)	0.71
	ADCC					
Node 1	481C>T(W)	85	50	37	2.13 (1.38-3.29)	**0.0007**
Node 2	803A>G(M)/ 481C>T(M)/ 857G>A(M)	134	60	30.9	1.62 (1.08-2.41)	**0.018**
Node 3	803A>G(W)/ 481C>T(M)/ 857G>A(M)	67	17	20.2	0.92 (0.51-1.66)	0.78
	SQCC					
Node 1	803A>G(W)	136	77	36.2	2.01 (1.25-3.22)	**0.004**
Node 2	590G>A(M)/ 857G>A(M)/ 803A>G(M)/ 481C>T(W)	11	7	38.9	1.97 (0.55-7.09)	0.30
Node 3	590G>A(W)/ 857G>A(M)/ 803A>G(M)/ 481C>T(W)	11	3	21.4	1.15 (0.30-0.46)	0.84
Node 4	857G>A(W)/ 803A>G(M)/ 481C>T(W)	26	5	16.1	0.86 (0.30-2.45)	0.78
Node 5	590G>A(W)/ 857G>A(M)/ 481C>T(M)	86	41	32.3	1.92 (1.11-3.31)	**0.02**
Node 6	803A>G(M)/ 590G>A(M)/ 857G>A(M)/ 481C>T(M)	68	27	28.4	1.61 (0.86-3.02)	0.14
Node 7	803A>G(W)/ 590G>A(M)/ 857G>A(M)/ 481C>T(M)	47	10	17.5	0.88 (0.40-1.93)	0.76
	SCLC					
Node 1	481C>T(W)	184	61	24.9	2.16 (1.41-3.32)	**0.00043**
	SMOKERS					
Node 1	481C>T(W)	126	194	60.6	1.28 (0.98-1.67)	0.07
	NON-SMOKERS					
Node 1	857G>A(M)	56	66	54.1	2.03 (1.20-3.45)	**0.008**

The #Case rate is the percentage of LC cases among all individuals in each terminal node case/(case+ control) ×100.
$p<0.05 is considered statistically significant.
Note: The bold values in the Table are highly significant and are a positive indicator for any association or not.

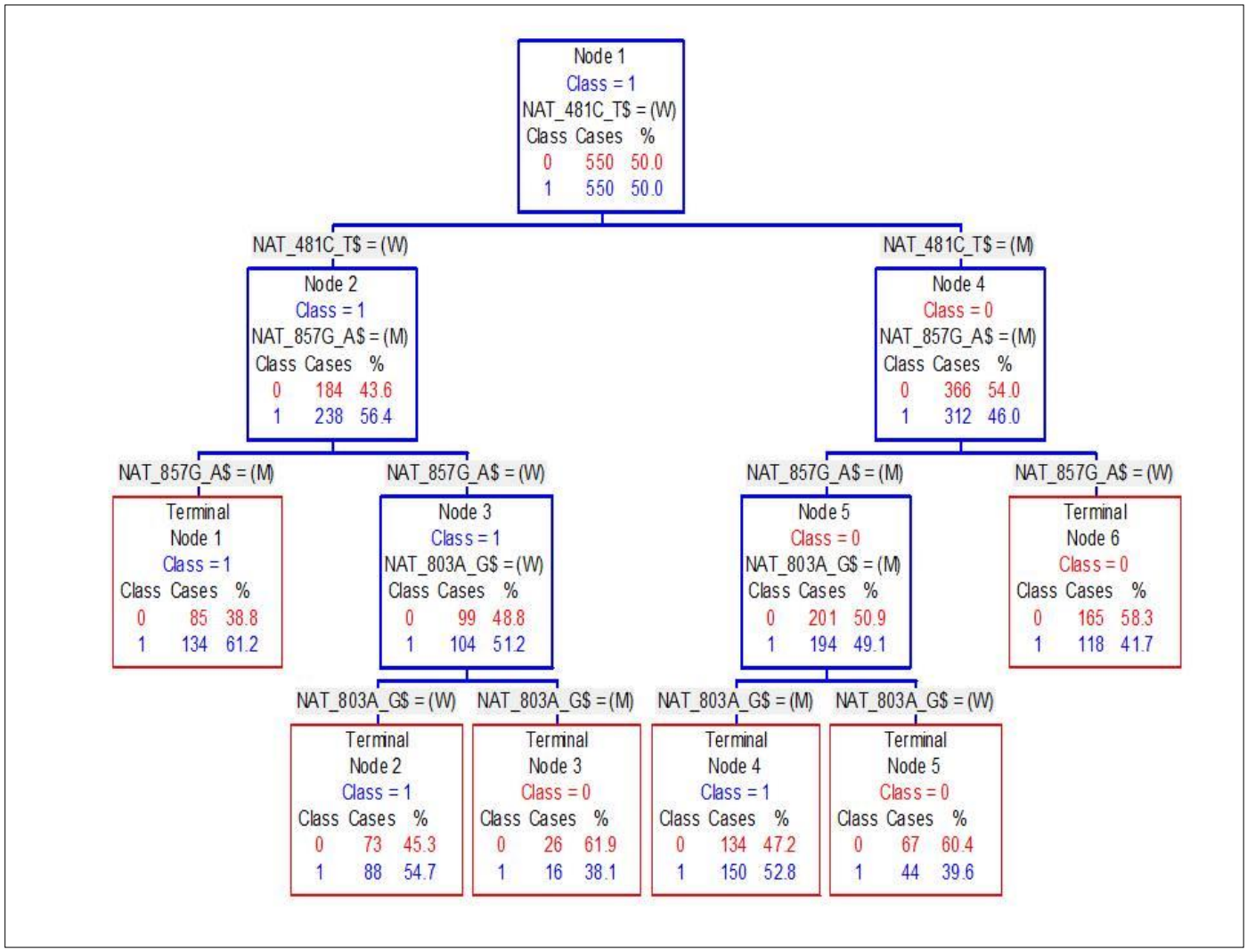


Figure 5.23: CART analysis in polymorphic NAT2.
W= homozygous wild type genotype, M = heterozygous + homozygous variant genotype

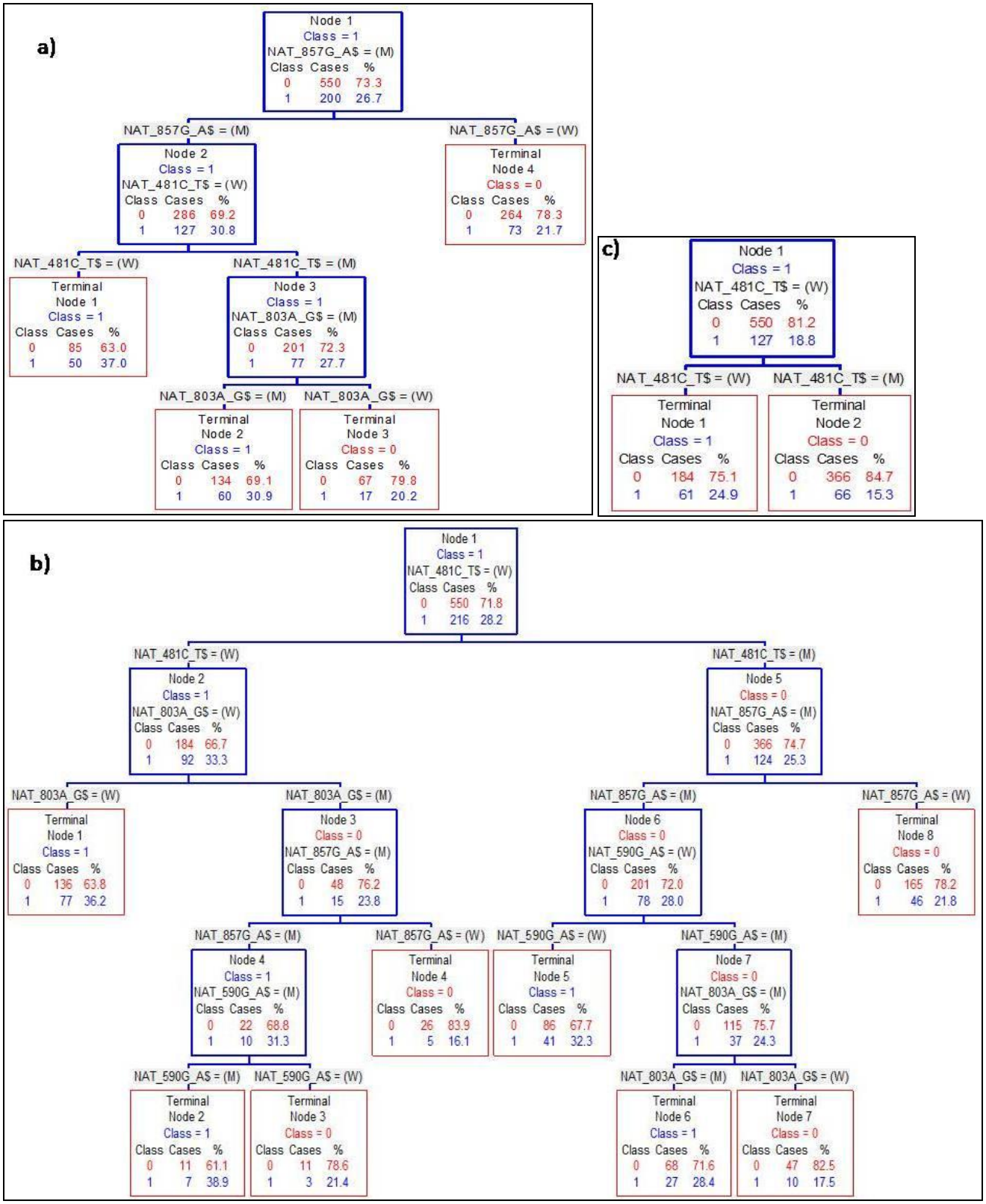


Figure 5.24: CART analysis in polymorphic *NAT2*based on histology, a) ADCC, b) SQCC, and c) SCLC. W= homozygous wild type genotype, M = heterozygous + homozygous variant genotype

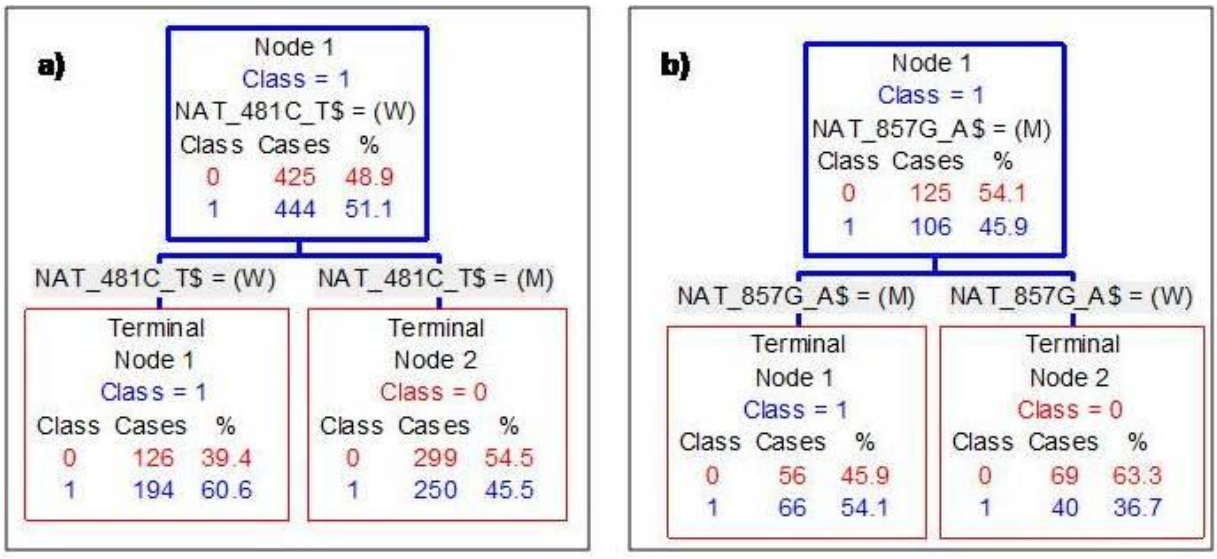


Figure 5.25: CART analysis in polymorphic *NAT2* based on smoking status, a) smokers, b) non-smokers. W= homozygous wild type genotype, M = heterozygous + homozygous variant genotype

5.7. Clinical characteristics of lung cancer patients evaluated for detecting the prognostic relevance of *NQO1, SULT1A1, EPHX1,* and *NAT2* polymorphism

Table 5.39 details the clinical features and demography of the study conducted for the overall survival of lung cancer patients to the polymorphic *NQO1, SULT1A1, EPHX1,* and *NAT2*. Out of 473 subjects, 83.09% were males, and the remaining (16.91%) were females. As smoking is an established cause of lung cancer, most of the sample group (81.40%) were smokers, whereas 18.60% were classified as non-smokers. Table 5.39 also includes clinical factors such as histology, TNM stage, performance status, and chemotherapeutic regimen. Based on the histological classification, there was a total of 37.63% adenocarcinomas, 40.59% squamous cell carcinomas, 20.72% small cell carcinomas, and 1.06% patients with unclassified histology. The TNM staging grouped the study subjects into 4.02% stage I and staged II patients, 42.07% stage III patients, and 46.30% stage IV patients, and the remaining 7.61% were unclassified. The study was analyzed for the performance status of ECOG and KPS. The classification according to performance status ECOG divided the study subjects into 43.76% subjects with ECOG score of 0-1, 38.90% with ECOG score of 2, 13.74% with ECOG score 3-4, and remaining 3.60% with unclassified ECOG score. Further, the KPS score for the study classified the subjects into 28.33% subjects with KPS scores between90-100, 52.85% subjects with KPS scores in the range of 70-80, and 15.22% with KPS score of below 60, and remaining 3.60% with unclassified KPS

score. There were 346 subjects whose chemotherapeutic data was available. The maximum number of patients were administered with pemetrexed alongwith cisplatin/carboplatin (31.21%), followed by 70 (20.23%) subjects administered with docetaxel along with cisplatin/carboplatin. Furthermore, 61 (17.63%) of the subjects received irinotecan along with cisplatin/carboplatin, 57 (16.47%) of subjects were given paclitaxel along with cisplatin/carboplatin, 19 (5.49%) of the subjects were given gemcitabine together with cisplatin/carboplatin. Further, 12 (3.47%) patients were given gefitinib, 11 (3.18%) subjects were administered etoposide, and only one subject was given ceritinib, and 7 (2.02%) subjects were given some other chemotherapeutic regimen apart from these.

Table 5.39: Patients characteristics and clinical features

Patients characteristics	**No. of patients, n(%) N=473**
Age (years)	
Mean ± SD (Range)	56.68 ± 10.16 (28-85)
Gender	
Male	393 (83.09)
Female	80 (16.91)
Smoking status	
Smokers	385 (81.40)
Non-smokers	88 (18.60)
Histology	
ADCC	178 (37.63)
SQCC	192 (40.59)
SCLC	98 (20.72)
Others	5 (1.06)
TNM stage	
I/II	19 (4.02%)
III	199 (42.07)
IV	219 (46.30)
Unclassified	36 (7.61)
ECOG performance status	
0-1	207 (43.76)
2	184 (38.90)
3-4	65 (13.74)
Unclassified	17 (3.60)
KPS performance status	
90-100	134 (28.33)
70-80	250 (52.85)
60 & below 60	72 (15.22)
Unclassified	17 (3.60)
Median Follow up (months)	7.3
Events, Deaths	N= 336 (71.04%)
Chemotherapy regimen	N= 346
Pemetrexed + cis/carboplatin	108 (31.21)
Irinotecan + cis/carboplatin	61 (17.63)
Docetaxel + cis/carboplatin	70 (20.23)
Paclitaxel + cis/carboplatin	57 (16.47)

Gemcitabine + cis/carboplatin	19 (5.49)
Gefitinib	12 (3.47)
Etoposide	11 (3.18)
Ceretinib	1 (0.29)
Others	7 (2.02)

Abbreviations: SD = Standard Deviation, n = total number of case-patients

5.8. The association of PhaseII detoxification genes towards outcome of lung cancer

5.8.1. Association of polymorphic *NQO1* towards lung cancer outcomes

5.8.1.2 Association of polymorphic *NQO1* genetic variant with overall survival of lung cancer patients

Table 5.40 summarizes the results of univariate and multivariate analysis for *NQO1* polymorphism. Kaplan-Meier statistics define univariate analysis, which provides median survival time (MST) and survival graphs for each genotype. The log-rank test in Kaplan-Meier statistics analyses the MST for different genotypes using the wild genotype as a reference. Kaplan-Meier statistics obtain unadjusted hazards ratios, but adjusted hazards ratios are provided by multivariate analysis using Cox regression. The hazards ratio was adjusted using the Cox-regression model for co-variates such as age, gender, smoking, histology, stage, ECOG, and KPS. All the 473 patients were followed from the day of diagnosis of lung cancer, and we found that 137 (28.96%) subjects were alive and 336 (71.04%) died due to disease.

The homozygous CC genotype was taken as a reference genotype. Although the median survival time (MST) of patients with mutant *TT* genotype was a bit higher in comparison with wild (*CC*) genotype (9.43 months *vs*. 8.23 months, Log-rank p= 0.28), the difference in the MST was not found to be significant. Better hazard rate was observed for subjects carrying the mutant alleles both in univariate (HR= 0.80, 95% C.I.= 0.55-1.17, Log-rank p=0.28) and multivariate(HR'= 1.03, 95% C.I.= 0.67-1.58, p= 0.89) analysis but lacks statistical significance.

Table 5.40: Association of *NQO1* SNP (rs1800566) on overall survival in lung cancer cases

Genotype rs1800566 C/T	CASES n(%), N=473	DEAD n(%), N=336	ALIVE n(%), N=137	Univariate analysis			Multivariate analysis	
				MST (months)	Log-rank *p*	Unadjusted HR[a]	Adjusted HR[b] (95% C.I.)	*p*
CC	231 (48.84)	165 (49.11)	66 (48.17)	8.23				
CT	196 (41.44)	143 (42.56)	53 (38.69)	8.13	0.65	0.95 (0.76-1.19)	0.95 (0.75-1.20)	0.65
TT	46 (9.72)	28 (8.33)	18 (13.14)	9.43	0.28	0.80 (0.55-1.17)	1.03 (0.67-1.58)	0.89
CT+TT	242 (51.16)	171 (50.89)	71 (51.83)	8.30	0.45	0.92 (0.74-1.14)	0.95 (0.76-1.19)	0.66

unadjusted Hazards ratio for Kaplan-Meier analysis, hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.8.1.3 Association of *NQO1* polymorphism and overall survival according to a histological subtype of lung cancer patients

To evaluate the relevance of the *NQO1 Pro*^{187}*Ser* polymorphic site in predicting survival of patients in different subgroups stratified based on histology, the association of the *NQO1* variant with the survival of patients was analyzed, and data was tabulated in Table 5.41. The percentage of heterozygous patients was 43.26%, 39.06%, and 41.84% among ADCC, SQCC, and SCLC subtypes, respectively. Similarly, the percentage of patients harboring mutant genotype was 11.80, 8.86, and 7.14%, respectively, among ADCC, SQCC, and SCLC subtypes. The ADCC patients carrying the mutant genotype (*TT*) showed a better OS as compared to wild genotype (*CC*) (9.43 vs. 8.23 months, log-rank p=0.60). Both the univariate (HR= 0.84, 95% C.I.= 0.46-1.55, Log-rank p=0.60) and multivariate (HR'= 1.01, 95% C.I.= 0.70-1.44, p= 0.98) analysis did not show any significant association. Similarly, the combined genotype (*CT+TT*) in the case of ADCC subjects also showed a better OS as compared to wild genotype (*CC*) (9.40 vs. 8.23 months, log-rank p= 0.42). Similarly, the SQCC patients did not show a significant correlation with OS. On the contrary, the SCLC subjects carrying mutant (*TT*) genotype showed a poor survival as compared to the subjects harboring wild (*CC*) genotype (5.57 months vs. 9.93 months, Log-rank p=0.75). The hazards ratio of SCLC subjects with mutant genotype was poor compared to wild genotype (HR'= 1.32, 95% C.I.= 0.48-3.67, p=0.59).

Table 5.41: Association of *NQO1* SNP (rs1800566) on survival in lung cancer cases according to tumor histology

ADCC								
Genotype rs1800566 C/T	**CASES n(%), N=178**	**DEAD n(%), N=116**	**ALIVE n(%), N=62**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	80 (44.94)	56 (48.28)	24 (38.71)	8.23				
CT	77 (43.26)	49 (42.24)	28 (45.16)	8.40	0.48	0.87 (0.59-1.28)	0.78 (0.52-1.18)	0.24
TT	21 (11.80)	11 (9.48)	10 (16.13)	9.43	0.60	0.84 (0.46-1.55)	1.01 (0.49-2.07)	0.98
CT+TT	98 (55.06)	60 (51.72)	38 (61.29)	9.40	0.42	0.86 (0.60-1.24)	0.79 (0.53-1.16)	0.23
SQCC								
Genotype rs1800566 C/T	**CASES n(%), N=192**	**DEAD n(%), N=139**	**ALIVE n(%), N=53**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	100 (52.08)	72 (51.80)	28 (52.83)	7.80				
CT	75 (39.06)	56 (40.29)	19 (35.85)	8.40	0.44	0.87 (0.62-1.24)	0.91 (0.63-1.34)	0.65
TT	17 (8.86)	11 (7.91)	6 (11.32)	8.60	0.46	0.79 (0.44-1.41)	1.04 (0.53-2.05)	0.90
CT+TT	92 (47.92)	67 (48.20)	25 (47.17)	8.40	0.37	0.86 (0.62-1.20)	0.92 (0.64-1.32)	0.65

SCLC

Genotype rs1800566 C/T	CASES n(%), N=98	DEAD n(%), N=78	ALIVE n(%), N=20	Univariate analysis			Multivariate analysis	
				MST (months)	Log rank *p*	Unadjusted HR[a]	Adjusted HR[b] (95% C.I.)	*P*
CC	50 (51.02)	37 (47.44)	13 (65.00)	9.93				
CT	41 (41.84)	35 (44.87)	6 (30.00)	6.87	0.36	1.24 (0.78-1.97)	1.50 (0.89-2.52)	0.12
TT	7 (7.14)	6 (7.69)	1 (5.00)	5.57	0.75	0.87 (0.38-1.98)	1.32 (0.48-3.67)	0.59
CT+TT	48 (48.98)	41 (52.56)	7 (35.00)	6.30	0.47	1.17 (0.75-1.83)	1.55 (0.94-2.55)	0.08

[a]Unadjusted Hazards ratio for Kaplan-Meier analysis, [b]hazards ratio for age, sex, smoking, stage, KPS, ECOG

5.8.1.4 Prognosis of LC patients treated with a different chemotherapeutic regimen

The survival for subjects treated with the different chemotherapeutic regimen and its association with *NQO1* polymorphism and OS has been studied and tabulated in Table 5.42. In the present study, different chemotherapeutic regimens were administered to treat lung cancer patients, but the analysis was carried on those regimens with which maximum numbers of patients were treated. Chemotherapeutic drugs were cyclically given to lung cancer patients. Patients were examined to determine whether drugs had an independent effect on survival. The OS of patients treated with four regimens *viz.* docetaxel + cis/carboplatin, irinotecan + cis/carboplatin, pemetrexed + cis/carboplatin, and paclitaxel + cis/carboplatin and their association with *NQO1* polymorphism has been illustrated in Table 5.42.

The lung cancer subjects frequency distribution with docetaxel and cis/carboplatin showed that 47.14% were single allelic carriers (*CT*) and 5.72% were double allelic carriers (*TT*) depicted in Table 5.42. The MST of subjects carrying the mutant (*TT*) genotype was lower as compared with the subjects having wild (*CC*) genotype (9.83 *vs.* 10.30 months, log-rank p=0.49). The univariate (HR=0.66, 95% C.I.= 0.24-1.85, Log-rank p=0.49) and multivariate (HR'=0.43, 95% C.I.= 0.11-1.73, p=0.24) analysis depicted no significant association between the subjects with mutant (*TT*) genotype for *NQO1* polymorphism (Table 5.42).

Furthermore, the frequency distribution of patients carrying wild, heterozygous, and mutant genotype was 52.63%, 35.09%, and 12.28%, respectively, especially in patients treated with paclitaxel-cis/carboplatin regimen. The subjects carrying mutant (*TT*) genotype depicted an inferior survival when compared with the subjects carrying the wild genotype (*CC*) (3.57 *vs.* 12.20 months; Log-rank p=0.13). In the multivariate cox proportional hazards model, a

significantly higher death rate (HR'= 7.95, 95% C.I.= 1.66-38.07, p=**0.0098**; Figure 5.26) was observed after adjusting for various patient factors, as shown in Table 5.42. The subjects treated with irinotecan cis/carboplatin and harboring the mutant (*TT*) genotype exhibited a poor survival of 5.57 months as compared to 9.13 months for the subjects with wild (*CC*) genotype (Log-rank p=0.93). The univariate and multivariate analysis revealed no significant relationship with survival outcomes. Furthermore, no significant association between the OS and *NQO1 609C>T* polymorphism was inferred in subjects treated with pemetrexed and cis/carboplatin (Table 5.42).

Table 5.42: Association of *NQO1* SNP (rs1800566) on overall survival according to the regimen

Docetaxel cisplatin/carboplatin								
Genotype rs1800566 C/T	**CASES n(%), N=70**	**DEAD n(%), N=56**	**ALIVE n(%), N=14**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	33 (47.14)	26 (46.43)	7 (50.00)	10.30				
CT	33 (47.14)	27 (48.21)	6 (42.86)	8.43	0.98	1.01 (0.59-1.73)	1.001 (0.52-1.92)	0.99
TT	4 (5.72)	3 (5.36)	1 (7.14)	9.83	0.49	0.66 (0.24-1.85)	0.43 (0.11-1.73)	0.24
CT+TT	37 (52.86)	30 (53.57)	7 (50.00)	8.63	0.85	0.95 (0.56-1.61)	0.90 (0.47-1.72)	0.76
Paclitaxel cisplatin/carboplatin								
Genotype rs1800566 C/T	**CASES n(%), N=57**	**DEAD n(%), N=34**	**ALIVE n(%), N=23**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	30 (52.63)	15 (44.12)	15 (65.22)	12.20				
CT	20 (35.09)	14 (41.18)	6 (26.09)	7.60	0.29	1.47 (0.69-3.11)	1.55 (0.71-3.40)	0.28
TT	7 (12.28)	5 (14.70)	2 (8.69)	3.57	0.13	2.12 (0.59-7.63)	7.95 (1.66-38.07)	**0.0098**
CT+TT	27 (47.37)	19 (55.88)	8 (34.78)	7.60	0.17	1.59 (0.81-3.13)	1.47 (0.73-2.97)	0.28
Irinotecan cisplatin/carboplatin								
Genotype rs1800566 C/T	**CASES n(%), N=61**	**DEAD n(%), N=49**	**ALIVE n(%), N=12**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	26 (42.62)	18 (36.73)	8 (66.67)	9.13				
CT	30 (49.18)	27 (55.10)	3 (25.00)	7.23	0.47	1.24 (0.69-2.24)	1.62 (0.81-3.27)	0.18
TT	5 (8.20)	4 (8.17)	1 (8.33)	5.57	0.93	0.95 (0.33-2.78)	0.90 (0.17-4.82)	0.89
CT+TT	35 (57.38)	31 (63.27)	4 (33.33)	7.23	0.52	1.21 (0.68-2.13)	1.59 (0.82-3.08)	0.17
Pemetrexed cisplatin/carboplatin								
Genotype rs1800566 C/T	**CASES n(%), N=108**	**DEAD n(%), N=73**	**ALIVE n(%), N=35**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	47 (43.52)	33 (45.21)	14 (40.00)	8.23				
CT	53 (49.07)	36 (49.31)	17 (48.57)	9.77	0.98	0.99 (0.62-1.57)	0.98 (0.59-1.62)	0.93
TT	8 (7.41)	4 (5.48)	4 (11.43)	9.43	0.69	0.81 (0.31-2.13)	1.49 (0.46-4.77)	0.50
CT+TT	61 (56.48)	40 (54.79)	21 (60.00)	9.77	0.92	0.98 (0.62-1.55)	0.99 (0.61-1.62)	0.97

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

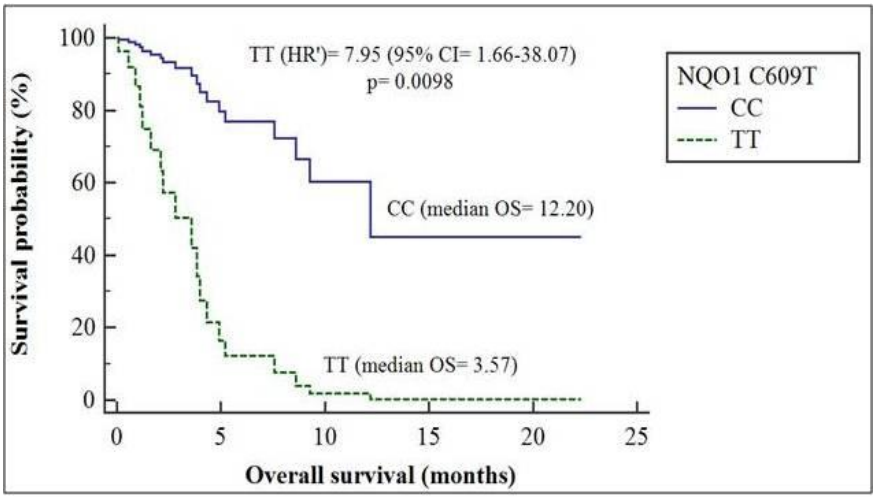


Figure 5.26: Cox proportional hazards-regression curve illustrating the association between lung cancer subjects with variant (*TT*) genotype treated with Paclitaxel cisplatin/carboplatin regimen showing *NQO1* 609C>T polymorphism

5.8.1.5 Association of *NQO1 Pro^{187}Ser* polymorphism with treatment response towards chemotherapeutic treatment and their correlation with objective response

The response variables were used to examine the treatment outcome for the *NQO1* polymorphism. The patients were classified as good or bad responders based on the RECIST criteria. Patients who had complete remission (CR) and partial remission (PR) were considered good responders, whereas those with stable disease (SD) and partial disease (PD) were considered poor responders (Table 5.43). The adjusted odds ratio was determined using wild-type genotype as a reference and controlling for covariates such as age, gender, smoking, stage, histology, KPS, ECOG, and chemotherapy regimens. The objective response outcome data were available for a total of 266 patients. The good responders with heterozygous genotype (47.46%) had a higher frequency representation than bad responders with an adjusted odds ratio of 1.22 (AOR= 1.22, 95% C.I.= 0.73-2.04, p= 0.45), but no significant difference was observed. Similarly, the subjects who were good responders and carrying the mutant genotype (11.01%) were over-represented than bad responders. An adjusted odds ratio of 2.2 fold was observed (AOR= 2.20, 95% C.I.= 0.86-5.64, p= 0.09) for the same (Table 5.43), signifying that CR+PR carrying mutant genotype had a good response towards platinum-based chemotherapy.

Table 5.43: Relationship of genotypic distribution with clinic-pathological parameter based on response to chemotherapy

Genotype rs1800566 C/T	Response of chemotherapy		AOR[a] (95% C.I.)	*p*[b]
	CR+PR (118), n (%)	SD+PD (148), n (%)		
CC	49 (41.53)	71 (47.97)	1.00 (Reference)	
CT	56 (47.46)	68 (45.95)	1.22 (0.73-2.04)	0.45
TT	13 (11.01)	9 (6.08)	2.20 (0.86-5.64)	0.09
CT + TT	69 (58.47)	77 (52.03)	1.35 (0.82-2.21)	0.24

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding p-values were calculated by logistic regression analysis after adjusting for age and gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table.

However, our data did not observe any significant association suggesting that lung cancer patients with variant genotype who underwent platinum-based chemotherapy were more likely to experience higher sensitivity and better response rates.

5.8.1.6 Association between *NQO1* polymorphism and toxicity

One of the current outcomes of interest in this study was to assess the development of toxicity in lung cancer patients having platinum-based doublet chemotherapy. To evaluate the risk of acquiring different toxicity grades after therapy in patients with each genotype, the odds ratio (OR) and 95% C.I. were calculated using the odd ratio test. We also utilized a logistic regression model to determine the adjusted odds ratio with the toxicity grades as the dependent variable. The covariates in this model were sex, age, performance status, and polymorphism, and they were assessed for possible interaction or confounding effects. We conducted two types of analysis to determine the effect of polymorphisms on different grades of toxicity. In the first analysis, we compared the presence of any toxicity grade (grades 1-4) versus the absence of toxicity (grade 0), and the second analysis compared subjects with intermediate/severe toxicity (grades 2-4) versus no/low-grade toxicity (grade 0-1) to determine whether the Phase II detoxification genetic variants were associated with severe toxicities. As a result, the patients were segregated into two different toxicity grade groups, and all patients were included in the analysis simultaneously. After each treatment cycle, the chemotherapy-related side effects were documented. In hematological toxicity, we evaluated the following parameters: leukopenia, anemia, thrombocytopenia, and neutropenia. The nephrotoxicity was evaluated using glomerular filtration rate (GFR). The gastrointestinal toxicity was evaluated using parameters like diarrhea, constipation, and other toxicities like anorexia and alopecia.

In the case of *NQO1 609C>T* polymorphism, the variant *TT* genotype was found to exhibit 2.7 folds increased risk of developing hematological toxicity, especially anemia (grade 1-4 *vs*. grade 0), but the data lacks statistical significance (Table 5.44 a)). Similarly, the *TT* genotype had a 2.2 fold risk of developing severe ANC (grade 2-4 *vs*. grade 0-1) but was insignificant. On the same lines, a trend of increased risk for neutropenia (grade 1-4 *vs*. grade 0) was observed using logistic regression analysis (AOR=2.57, 95% C.I.= 0.95-6.91, p= 0.06). In case of gastrointestinal toxicity, the variant *TT* genotype in case of constipation (OR=0.10, 95% C.I.=0.01-0.92, p=**0.04**) (χ^2= 7.83, p=**0.02**) and anorexia (OR=0.15, 95% C.I.=0.02-0.86, p=**0.03**) (χ^2= 7.84, p=**0.02**) were not associated with increased risk of these chemotherapeutic related side effects (Table 5.44 c)).

Apart from these, the *NQO1 609C>T* polymorphism was not found to be associated with any other hematological (Table 5.44 a)), nephrotoxicity (Table 5.44 b)), and gastrointestinal toxicity (Table 5.44c)).

Table 5.44 a): Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NQO1* polymorphisms in multivariate analysis

Leukopenia						
Genotype rs1800566 C/T	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	95(50.26)	42(50.00)	1.00 (Reference)		1.00 (Reference)	
CT	81(42.86)	33(39.29)	0.92(0.53-1.59)	0.77	0.90(0.52-1.56)	0.71
TT	13(6.88)	9(10.71)	1.57(0.62-3.95)	0.34	1.52(0.60-3.90)	0.38
CT+TT	94(49.74)	42(50.00)	1.01(0.60-1.69)	0.97	0.99(0.59-1.67)	0.97
Genotype rs1800566 C/T	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	123(49.80)	14(53.85)	1.00 (Reference)		1.0 (Reference)	
CT	106(42.91)	8(30.77)	0.66(0.27-1.64)	0.37	0.63(0.25-1.57)	0.32
TT	18(7.29)	4(15.38)	1.95(0.58-6.59)	0.28	2.24(0.64-7.88)	0.21
CT+TT	124(50.20)	12(46.15)	0.85(0.38-1.91)	0.69	0.82(0.36-1.86)	0.63
Anemia						
Genotype rs1800566 C/T	**Grade of toxicity 0, n (%)**	**Grade of toxicity 1-4, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	25(58.14)	112(48.70)	1.00 (Reference)		1.00 (Reference)	
CT	16(37.21)	98(42.61)	1.37(0.69-2.71)	0.37	1.47(0.73-2.97)	0.28
TT	2(4.65)	20(8.69)	2.23(0.49-10.17)	0.30	2.72(0.57-12.88)	0.21
CT+TT	18(41.86)	118(51.30)	1.46(0.76-2.83)	0.26	1.60(0.81-3.14)	0.17
Genotype rs1800566 C/T	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	72(53.73)	65(46.76)	1.00(Reference)		1.00(Reference)	
CT	51(38.06)	63(45.32)	1.37(0.83-2.25)	0.22	1.40(0.84-2.31)	0.19
TT	11(8.21)	11(7.91)	1.11(0.45-2.73)	0.22	1.19(0.48-2.99)	0.71
CT+TT	62(46.27)	74(53.24)	1.32(0.82-2.13)	0.25	1.36(0.84-2.19)	0.21
Thrombocytopenia						
Genotype rs1800566 C/T	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	103(50.74)	33(47.83)	1.00(Reference)		1.00(Reference)	
CT	85(41.87)	29(42.03)	1.06(0.60-1.89)	0.83	1.07(0.60-1.94)	0.81
TT	15(7.39)	7(10.14)	1.46(0.55-3.88)	0.45	1.58(0.57-4.39)	0.38
CT+TT	100(49.26)	36(52.17)	1.12(0.65-1.94)	0.68	1.14(0.65-2.00)	0.65

Genotype rs1800566 C/T	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	131(49.62)	5(62.50)	1.00(Reference)		1.00(Reference)	
CT	111(42.05)	3(37.50)	0.71(0.16-3.03)	0.64	0.77(0.18-3.37)	0.73
TT	22(8.33)	0(0.00)	0.53(0.03-9.94)	0.67	-	-
CT+TT	133(50.38)	3(37.50)	0.59(0.14-2.52)	0.48	0.64(0.15-2.79)	0.55

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

Table 5.44 b): Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NQO1* polymorphisms in multivariate analysis

Genotype rs1800566 C/T	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	39(51.32)	85(53.80)	1.00(Reference)		1.00(Reference)	
CT	28(36.84)	66(41.77)	1.08(0.60-1.94)	0.79	1.37(0.70-2.67)	0.36
TT	9(11.84)	7(4.43)	0.36(0.12-1.03)	0.05	0.36(0.10-1.32)	0.12
CT+TT	37(48.68)	73(46.20)	0.90(0.52-1.56)	0.72	1.17(0.60-2.13)	0.71
Genotype rs1800566 C/T	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	90(49.72)	34(64.15)	1.00(Reference)		1.00(Reference)	
CT	77(42.54)	17(32.08)	0.58(0.30-1.13)	0.11	0.56(0.26-1.19)	0.13
TT	14(7.73)	2(3.77)	0.38(0.08-1.75)	0.21	0.22(0.03-1.48)	0.12
CT+TT	91(50.28)	19(35.85)	0.55(0.29-1.04)	0.07	0.49(0.24-1.04)	0.06

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

Table 5.44 c): Gastrointestinal Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NQO1* polymorphisms in multivariate analysis

		Diarrohea				
Genotype rs1800566 C/T	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	25(58.14)	28(43.08)	1.00(Reference)		1.00(Reference)	
CT	14(32.56)	31(47.69)	1.98(0.86-4.53)	0.11	2.03(0.86-4.79)	0.10
TT	4(9.30)	6(9.23)	1.34(0.34-5.30)	0.68	1.33(0.32-5.54)	0.69
CT+TT	18(41.86)	37(56.92)	1.83(0.84-4.00)	0.13	1.85(0.83-4.12)	0.13
Genotype rs1800566 C/T	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	40(51.95)	13(41.94)	1.00(Reference)		1.00(Reference)	
CT	29(37.66)	16(51.61)	1.70(0.71-4.07)	0.23	1.66(0.67-4.11)	0.27
TT	8(10.39)	2(6.45)	0.77(0.14-4.09)	0.76	0.78(0.15-4.21)	0.77
CT+TT	37(48.05)	18(58.06)	1.50(0.64-3.47)	0.35	1.43(0.60-3.39)	0.42

Constipation						
Genotype rs1800566 C/T	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	15(48.39)	21(51.22)	1.00(Reference)		1.00(Reference)	
CT	9(29.03)	19(46.34)	1.51(0.54-4.24)	0.44	1.55(0.54-4.43)	0.41
TT	7(22.58)	1(2.44)	0.10(0.01-0.92)	**0.04**	0.13(0.01-1.27)	0.08
CT+TT	16(51.61)	20(48.78)	0.89(0.35-2.27)	0.81	0.99(0.38-2.58)	0.98
Anorexia						
Genotype rs1800566 C/T	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	4(40.00)	45(47.87)	1.00(Reference)		1.00(Reference)	
CT	3(30.00)	44(46.81)	1.30(0.27-6.16)	0.74	1.30(0.27-6.35)	0.74
TT	3(30.00)	5(5.32)	0.15(0.02-0.86)	**0.03**	0.22(0.04-1.31)	0.09
CT+TT	6(60.00)	49(52.13)	0.73(0.19-2.74)	0.64	0.78(0.20-3.02)	0.72
Alopecia						
Genotype rs1800566 C/T	**Grade of toxicity (No), n(%)**	**Grade of toxicity (Yes), n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	15(42.86)	36(46.15)	1.00(Reference)		1.00(Reference)	
CT	14(40.00)	37(47.44)	1.10(0.46-2.60)	0.83	1.07(0.45-2.54)	0.88
TT	6(17.14)	5(6.41)	0.35(0.09-1.31)	0.12	0.28(0.07-1.14)	0.08
CT+TT	20(57.14)	42(53.85)	0.87(0.39-1.95)	0.74	0.85(0.38-1.91)	0.69

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

5.8.2. Association of polymorphic *SULT1A1* towards lung cancer outcomes

5.8.2.1.Association of *SULT1A1* polymorphism with overall survival of LC patients

Overall survival analysis of 473 lung cancer subjects was conducted, as shown in Table 5.45. First, univariate analysis was performed using Kaplan-Meier curves and log-rank test. Then multivariate Cox hazard proportional analysis was performed to evaluate the association of *SULT1A1* polymorphism with the overall survival of patients. All the 473 patients were followed from the day of diagnosis of lung cancer, and we found that 137 (28.96%) subjects were alive and 336 (71.04%) died due to disease. The homozygous wild (*Arg/Arg*) genotype was taken as a reference. The median survival time (MST) of patients carrying the homozygous mutant (*His/His*) genotype was higher in comparison with patients carrying the homozygous wild (*Arg/Arg*) genotype (11.30 months *vs.* 8.07 months, Log-rank p= 0.19). There was no significant association observed for *Arg*^{213}*His* polymorphism and overall survival in lung cancer patients applying both univariate and multivariate analysis (Table 5.45).

Table 5.45: Univariate and Multivariate analysis for *SULT1A1*

Genotype rs9282861 *Arg*^{213}*His*	**CASES n(%), N=473**	**DEAD n(%), N=336**	**ALIVE n(%), N=137**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***p***
Arg/Arg	306 (64.70)	220 (65.48)	86 (62.77)	8.07				
Arg/His	143 (30.23)	103 (30.65)	40 (29.20)	8.60	0.38	0.90 (0.72-1.14)	0.95 (0.74-1.21)	0.67
His/His	24 (5.07)	13 (3.87)	11 (8.03)	11.3	0.19	0.69 (0.43-1.12)	0.64 (0.35-1.16)	0.14
Arg/His+His/His	167 (35.30)	116 (34.52)	51 (37.23)	9.03	0.23	0.87 (0.70-1.09)	0.89 (0.71-1.14)	0.37

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.8.2.2.Association of *SULT1A1 Arg*^{213}*His* with OS in LC histological subtypes

Overall survival in *SULT1A1* polymorphism and its association with different histological subtypes of lung cancer have been explained in Table 5.46. The percentage of patients carrying heterozygous genotype was 31.46%, 28.64%, and 29.59% among ADCC, SQCC, and SCLC subtypes. In comparison, the percentage of subjects carrying the homozygous mutant genotype *(His/His)* was 4.50%, 5.21%, and 6.12% among ADCC, SQCC, and SCLC subtypes, respectively. ADCC subjects carrying the heterozygous (*Arg/His*) genotype showed a bit lower OS than ADCC subjects carrying wild genotype (7.70 months *vs.* 8.30 months). The hazards

ratio obtained from the multivariate analysis was 1.02 with nominal p values of 0.92. Whereas, the ADCC subjects carrying double variant allele (*His/His*) or mutant alleles showed MST of 11.30 months pertaining to a decreased hazards ratio in both univariate analysis (HR=0.48, 95% C.I.= 0.21-1.10, log-rank p=0.19) and multivariate analysis (HR'= 0.63, 95% C.I.= 0.19-2.03, *p*=0.44) when compared with the subjects harboring wild genotype (*Arg/Arg*).

Further, the SQCC patients showed a favorable trend as the overall median survival time was higher among individuals carrying heterozygous genotype when compared to patients carrying the wild-type genotype (8.83 *vs*. 7.80 months; log-rank p=0.50). After applying multivariate analysis, no significant association was observed among patients carrying variant genotype and lung cancer risk (HR'= 0.78; 95% C.I.= 0.53-1.16). In SQCC patients showing $Arg^{123}His$ polymorphism, homozygous mutant genotype (*His/His*) (MST=7.33 months; log-rank p=0.62) showed lower survival than the reference genotype (*Arg/Arg*) (MST=7.80 months). There was no significant result observed between OS and $Arg^{123}His$ polymorphism in SQCC patients (Table 5.46).

The SCLC subjects carrying the heterozygous genotype showed better survival than the homozygous wild genotype (9.40 months *vs*. 8.27 months, log-rank p= 0.72). The subjects harboring homozygous mutant genotype showed lower survival than the wild genotype (7.23 months *vs*. 8.27 months, log-rank p= 0.73). No significant association was observed among patients carrying heterozygous genotype (HR'=1.07; 95% C.I.= 0.62-1.86) or homozygous mutant genotype (HR'= 0.46; 95% C.I.= 0.12-1.71) and lung cancer risk on applying multivariate analysis as shown in Table 5.46.

Table 5.46: Association of *SULT1A1* SNP (rs9282861) on overall survival according to tumor histology

ADCC								
Genotype rs9282861 *Arg^{213}His*	**CASES n(%), N=178**	**DEAD n(%), N=116**	**ALIVE n(%), N=62**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Arg/Arg	114 (64.04)	77 (66.38)	37 (59.68)	8.30				
Arg/His	56 (31.46)	36 (31.03)	20 (32.26)	7.70	0.94	0.98 (0.66-1.46)	1.02 (0.66-1.57)	0.92
His/His	8 (4.50)	3 (2.59)	5 (8.06)	11.30	0.19	0.48 (0.21-1.10)	0.63 (0.19-2.03)	0.44
Arg/His+His/His	64 (35.96)	39 (33.62)	25 (40.32)	9.13	0.64	0.91 (0.62-1.33)	0.96 (0.63-1.45)	0.83
SQCC								
Genotype rs9282861 *Arg^{213}His*	**CASES n(%), N=192**	**DEAD n(%), N=139**	**ALIVE n(%), N=53**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Arg/Arg	127 (66.15)	90 (64.75)	37 (69.81)	7.80				
Arg/His	55 (28.64)	43 (30.93)	12 (22.64)	8.83	0.50	0.88 (0.62-1.26)	0.78 (0.53-1.16)	0.22
His/His	10 (5.21)	6 (4.32)	4 (7.55)	7.33	0.62	0.81 (0.38-1.73)	0.56 (0.23-1.38)	0.21
Arg/His+His/His	65 (33.85)	49 (35.25)	16 (30.19)	8.83	0.44	0.87 (0.62-1.23)	0.75 (0.51-1.08)	0.13
SCLC								
Genotype rs9282861 *Arg^{213}His*	**CASES n(%), N=98**	**DEAD n(%), N=78**	**ALIVE n(%), N=20**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Arg/Arg	63 (64.29)	51 (65.38)	12 (60.00)	8.27				
Arg/His	29 (29.59)	23 (29.49)	6 (30.00)	9.40	0.72	0.92 (0.56-1.48)	1.07 (0.62-1.86)	0.81
His/His	6 (6.12)	4 (5.13)	2 (10.00)	7.23	0.73	0.84 (0.33-2.16)	0.46 (0.12-1.71)	0.25
Arg/His+His/His	35 (35.71)	27 (34.62)	8 (40.00)	9.40	0.65	0.89 (0.57-1.42)	0.94 (0.55-1.63)	0.84

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, stage, KPS, ECOG

5.8.2.3. Prognosis of LC patients treated with the different chemotherapeutic regimen and its relation with *SULT1A1 Arg^{213}His* polymorphism

Lung cancer patients were administered with different chemotherapeutic regimens and their association with OS, as shown in Table 5.47. Lung cancer subjects were administered with different chemotherapeutic regimens, but the analysis was only applied on those regimens with which the maximum numbers of patients were treated in the study. The OS of the patients treated with four regimens, *i.e.*, docetaxel + cis/carboplatin, irinotecan + cis/carboplatin, pemetrexed + cis/carboplatin, and paclitaxel + cis/carboplatin, and their association with *SULT1A1*polymorphism have been elaborated in Table 5.47.

In *SULT1A1* polymorphism, the patients treated with docetaxel + cis/carboplatin regimen for all the three genotypic variants (heterozygous, mutant, and combined) showed a poor survival (5.83, 2.30, and 5.83 months, respectively) when compared with wild type reference genotype (*Arg/Arg*) which have a median survival of 9.83 months (Table 5.47). No significant association can be stated between *SULT1A1* polymorphism and lung cancer survival rates.

Furthermore, the patients harboring mutant genotype (*His/His*) and treated with paclitaxel and cisplatin/carboplatin regimen depicted an inferior survival compared with the wild genotype (2.23 *vs.* 7.60 months) as shown in Table 5.47. The patients with heterozygous genotype showed better survival than the wild genotype (9.27 months *vs.* 7.60 months, log-rank p= 0.49) and expressed a low hazard ratio. However, results were not reported to be significant (HR'= 0.68, 95% C.I.= 0.27-1.74). In the case of lung cancer patients treated with irinotecan and cis/carboplatin, and harboring the heterozygous (MST= 6.73 months) or homozygous variant genotype (MST= 7.23 months) depicted a poor survival time as compared to subjects carrying the wild genotype (MST= 9.10 months). No significant association was seen between the patients treated with this regimen and lung cancer. The fourth regimen with which maximum patients in the present study have been evaluated is pemetrexed and cis/carboplatin. The combined genotype (*Arg/His* + *His/His*) showed better survival than patients carrying the wild genotype (MST= 12.60 *vs.* 9.43 months, log-rank p= 0.19).

Table 5.47: Association of *SULT1A1* SNP (rs9282861) on overall survival according to chemotherapeutic regimen

Docetaxel cisplatin/carboplatin								
Genotype rs9282861 *Arg^{213}His*	**CASES n(%), N=70**	**DEAD n(%), N=56**	**ALIVE n(%), N=14**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Arg/Arg	45 (64.29)	35 (62.50)	10 (71.43)	9.83				
Arg/His	21 (30.00)	19 (33.93)	2 (14.28)	5.83	0.77	0.92 (0.53-1.60)	0.68 (0.32-1.42)	0.30
His/His	4 (5.71)	2 (3.57)	2 (14.28)	2.30	0.67	0.73 (0.21-2.55)	0.56 (0.10-3.12)	0.51
Arg/His+His/His	25 (35.71)	21 (37.50)	4 (28.57)	5.83	0.68	0.89 (0.52-1.53)	0.67 (0.34-1.31)	0.24
Paclitaxel cisplatin/carboplatin								
Genotype rs9282861 *Arg^{213}His*	**CASES n(%), N=57**	**DEAD n(%), N=34**	**ALIVE n(%), N=23**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Arg/Arg	39 (68.42)	23 (67.65)	16 (69.57)	7.60				
Arg/His	16 (28.07)	9 (26.47)	7 (30.43)	9.27	0.49	0.76 (0.37-1.59)	0.68 (0.27-1.74)	0.68
His/His	2 (3.51)	2 (5.88)	0 (0.00)	2.23	0.67	1.36 (0.26-7.05)	0.90 (0.18-4.43)	0.90
Arg/His+His/His	18 (31.58)	11 (32.35)	7 (30.43)	9.27	0.60	0.82 (0.41-1.66)	0.71 (0.31-1.63)	0.42
Irinotecan cisplatin/carboplatin								
Genotype rs9282861 *Arg^{213}His*	**CASES n(%), N=61**	**DEAD n(%), N=49**	**ALIVE n(%), N=12**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Arg/Arg	42 (68.85)	34 (69.39)	8 (66.67)	9.10				
Arg/His	16 (26.23)	12 (24.49)	4 (33.33)	6.73	0.74	0.90 (0.47-1.71)	0.87 (0.39-1.94)	0.74
His/His	3 (4.92)	3 (6.12)	0 (0.00)	7.23	0.74	1.21 (0.34-4.37)	0.16 (0.02-1.22)	0.08
Arg/His+His/His	19 (31.15)	15 (30.61)	4 (33.33)	7.23	0.84	0.94 (0.51-1.71)	0.72 (0.32-1.62)	0.43
Pemetrexed cisplatin/carboplatin								
Genotype rs9282861 *Arg^{213}His*	**CASES n(%), N=108**	**DEAD n(%), N=73**	**ALIVE n(%), N=35**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Arg/Arg	71 (65.74)	51 (69.86)	20 (57.14)	9.13				
Arg/His	34 (31.48)	21 (28.77)	13 (37.15)	9.43	0.29	0.76 (0.47-1.24)	0.84 (0.48-1.45)	0.53
His/His	3 (2.78)	1 (1.37)	2 (5.71)	-	0.20	0.30 (0.10-0.95)	0.32 (0.04-2.45)	0.27
Arg/His+His/His	37 (34.26)	22 (30.14)	15 (42.86)	12.60	0.19	0.72 (0.45-1.26)	0.78 (0.45-1.35)	0.38

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

Thus, the combined genotype expressed a low hazard ratio (HR'= 0.78, 95% C.I.= 0.45-1.35, p= 0.38), but the results were insignificant. The lung cancer subjects did not exhibit any significant association between the OS and *SULT1A1 Arg*^{213}*His* polymorphism was treated with any afore mentioned chemotherapeutic regimens (Table 5.47).

5.8.2.4. *SULT1A1* polymorphism and its response towards chemotherapy

Table 5.48 shows the association between *SULT1A1* polymorphism and the response to chemotherapy in lung cancer. The patients who were given chemotherapy were divided into responders and non-responders based on RECIST criteria to analyze the response chemotherapeutic regimen concerning the *SULT1A1* genotype. Responders comprised of CR+PR and non-responders comprised of SD+PD patients. The percentage of mutant (*His/His*) genotype was higher among responders (5.09%) in comparison to non-responders (4.06%). Based on logistic regression analysis, the *SULT1A1* genotype did not exhibit any association towards lung cancer risk in subjects undergoing platinum-based chemotherapy (Table 5.48).

Table 5.48: Relationship of genotypic distribution with lung cancer based on response to chemotherapy

Genotype rs9282861 *Arg*^{213}*His*	**Response of chemotherapy**		**AOR[a] (95% C.I.)**	***P*[b]**
	CR+PR (118), n (%)	**SD+PD (148), n (%)**		
Co-dominant model				
Arg/Arg	78 (66.10)	96 (64.86)	1.00 (Reference)	
Arg/His	34 (28.81)	46 (31.08)	0.92 (0.54-1.59)	0.78
His/His	6 (5.09)	6 (4.06)	1.22 (0.37-3.97)	0.75
Dominant model				
Arg/Arg	78 (66.10)	96 (64.86)	1.00 (Reference)	
Arg/His+His/His	40 (33.90)	52 (35.14)	0.97 (0.58-1.62)	0.89
Recessive model				
Arg/His+Arg/Arg	112 (94.91)	142 (95.94)	1.00 (Reference)	
His/His	6 (5.09)	6 (4.06)	1.31 (0.41-4.24)	0.65

[a]Adjusted Odds ratios, [b] 95% confidence intervals, and corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender and smoking. Abbreviation: AOR, Adjusted Odds ratio; CI, confidence interval.

5.8.2.5.

5.8.2.5. Association between *SULT1A1* polymorphism and toxicity

For the SNP *SULT1A1 Arg*^{213}*His,* and the different genotypes of this SNP, were not found to be associated with any of the hematological, nephrotoxicity, and gastrointestinal toxicity studied as demonstrated in Table 5.49 a), Table 5.49 b), and Table 5.49 c), respectively. However, the nephrotoxicity *viz.* GFR in case of subjects harboring the heterozygous genotype

(*Arg/His*) (AOR=0.38, 95% C.I.=0.19-0.75, p=**0.006**) and combined genotype (*Arg/His* + *His/His*) (AOR=0.38, 95% C.I.=0.20-0.74, p= **0.004**) was found not to be associated with increased risk of nephrotoxicity when grade 1-4 was compared with grade 0 (*i.e.*, no toxicity) (Table 5.49 b)). The chi-square value for it was also calculated ((χ^2= 6.89, p=**0.03**)). Further, when intermediate/severe, *i.e.*, grade 2-4, as compared with low toxicity profile nephrotoxicity (grade 0-1), both the heterozygous and combined genotype for *SULT1A1 Arg*^{213}*His* polymorphism was found to be associated with decreased risk towards nephrotoxicity with AOR of 0.32 (95% C.I.= 0.13-0.79, p=**0.01**) and 0.44 (95% C.I.=0.20-0.96, p=**0.004**), respectively as shown in Table 5.49 b) and chi-square to be, χ^2= 6.13 (p=**0.05**).

Table 5.49 a): Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *SULT1A1* polymorphisms in multivariate analysis

Leukopenia						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Arg/Arg	116(61.38)	59(70.24)	1.00 (Reference)		1.00 (Reference)	
Arg/His	61(32.28)	22(26.19)	0.71(0.40-1.26)	0.24	0.69(0.38-1.24)	0.21
His/His	12(6.35)	3(3.57)	0.49(0.13-1.81)	0.28	0.55(0.15-2.08)	0.38
Arg/His+Arg/Arg	73(38.62)	25(29.76)	0.67(0.39-1.17)	0.16	0.67(0.38-1.16)	0.15
Genotype rs9282861 *Arg^{213}His*	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Arg/Arg	156(63.16)	19(73.08)	1.00 (Reference)		1.0 (Reference)	
Arg/His	78(31.58)	5(19.23)	0.53(0.19-1.46)	0.22	0.51(0.18-1.44)	0.20
His/His	13(5.26)	2(7.69)	1.26(0.26-6.03)	0.77	1.58(0.32-7.89)	0.57
Arg/His+Arg/Arg	91(36.84)	7(26.92)	0.63(0.25-1.56)	0.32	0.64(0.26-1.60)	0.34
Anemia						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 0,n (%)**	**Grade of toxicity 1-4, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Arg/Arg	24(55.81)	151(65.65)	1.00 (Reference)		1.00 (Reference)	
Arg/His	15(34.88)	68(29.57)	0.72(0.35-1.46)	0.36	0.62(0.32-1.22)	0.17
His/His	4(9.30)	11(4.78)	0.44(0.13-1.48)	0.18	0.66(0.32-1.35)	0.25
Arg/His+Arg/Arg	19(44.19)	79(34.35)	0.66(0.34-1.28)	0.22	0.52(0.15-1.78)	0.30
Genotype rs9282861 *Arg^{213}His*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Arg/Arg	79(58.96)	96(69.06)	1.00(Reference)		1.00(Reference)	
Arg/His	45(33.58)	38(27.34)	0.69(0.41-1.17)	0.17	0.68(0.40-1.16)	0.16
His/His	10(7.46)	5(3.60)	0.41(0.13-1.25)	0.12	0.41(0.13-1.27)	0.12
Arg/His+Arg/Arg	55(41.04)	43(30.94)	0.64(0.39-1.06)	0.08	0.64(0.38-1.05)	0.07
Thrombocytopenia						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Arg/Arg	132(65.02)	42(60.87)	1.00(Reference)		1.00(Reference)	
Arg/His	58(28.57)	25(36.23)	1.35(0.75-2.43)	0.31	1.48(0.81-2.71)	0.20
His/His	13(6.40)	2(2.90)	0.48(0.11-2.23)	0.35	0.38(0.08-1.78)	0.22
Arg/His+Arg/Arg	71(34.98)	27(39.13)	1.19(0.68-2.10)	0.53	1.24(0.69-2.20)	0.47

Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Arg/Arg	171(64.53)	3(42.86)	1.00(Reference)		1.00(Reference)	
Arg/His	79(29.81)	4(57.14)	2.89(0.63-13.20)	0.17	3.16(0.68-14.66)	0.14
His/His	15(5.66)	0(0.00)	1.58(0.08-32.02)	0.76	-	-
Arg/His+Arg/Arg	94(35.47)	4(57.14)	2.42(0.53-11.07)	0.25	2.55(0.55-11.80)	0.23
Neutropenia						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	128(62.75)	41(70.69)	1.00(Reference)		1.00(Reference)	
Arg/His	63(30.88)	15(25.86)	0.74(0.38-1.44)	0.38	0.54(0.11-2.53)	0.43
His/His	13(6.37)	2(3.45)	0.48(0.10-2.22)	0.35	0.73(0.37-1.42)	0.35
Arg/His+Arg/Arg	76(37.25)	17(29.31)	0.70(0.37-1.31)	0.27	0.70(0.37-1.32)	0.27
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	145(64.44)	24(64.86)	1.00(Reference)		1.00(Reference)	
Arg/His	66(29.33)	12(32.43)	1.10(0.52-2.33)	0.81	1.09(0.51-2.32)	0.82
His/His	14(6.22)	1(2.70)	0.43(0.05-3.43)	0.43	0.43(0.05-3.44)	0.42
Arg/His+Arg/Arg	80(35.56)	13(35.13)	0.98(0.47-2.03)	0.96	0.98(0.47-2.04)	0.97

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

Table 5.49 b): Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *SULT1A1* polymorphisms in multivariate analysis

Nephrotoxicity						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	39(51.32)	109(68.99)	1.00(Reference)		1.00(Reference)	
Arg/His	31(40.79)	41(25.95)	0.47(0.26-0.86)	**0.01**	0.38(0.19-0.75)	**0.006**
His/His	6(7.89)	8(5.06)	0.48(0.15-1.46)	0.19	0.35(0.09-1.39)	0.14
Arg/His+Arg/Arg	37(48.68)	49(31.01)	0.47(0.27-0.83)	**0.009**	0.38(0.20-0.74)	**0.004**
Genotype rs9282861 *Arg^{213}His*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	108(59.67)	40(75.47)	1.00(Reference)		1.00(Reference)	
Arg/His	63(34.81)	9(16.98)	0.38(0.17-0.85)	**0.02**	0.32(0.13-0.79)	**0.01**
His/His	10(5.52)	4(7.55)	1.08(0.32-3.64)	0.90	1.28(0.30-5.37)	0.74
Arg/His+Arg/Arg	73(40.33)	13(24.53)	0.48(0.24-0.96)	**0.04**	0.44(0.20-0.96)	**0.04**

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

Table 5.49 c): Gastrointestinal Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *SULT1A1* polymorphisms in multivariate analysis

Diarrohea						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	24(55.81)	47(72.31)	1.00(Reference)		1.00(Reference)	
Arg/His	16(37.21)	16(24.62)	0.51(0.22-1.19)	0.12	0.50(0.21-1.19)	0.12
His/His	3(6.98)	2(3.08)	0.34(0.05-2.18)	0.25	0.31(0.05-2.02)	0.22
Arg/His+Arg/Arg	19(44.19)	18(27.69)	0.48(0.21-1.09)	0.08	0.47(0.21-1.07)	0.07

Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	47(61.04)	24(77.42)	1.00(Reference)		1.00(Reference)	
Arg/His	26(33.77)	6(19.35)	0.45(0.16-1.25)	0.12	0.45(0.16-1.26)	0.13
His/His	4(5.19)	1(3.23)	0.49(0.05-4.62)	0.53	0.44(0.04-4.18)	0.47
Arg/His+Arg/Arg	30(38.96)	7(22.58)	0.46(0.17-1.19)	0.11	0.45(0.17-1.19)	0.11
Constipation						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	19(61.29)	22(53.66)	1.00(Reference)		1.00(Reference)	
Arg/His	10(32.26)	17(41.46)	1.47(0.54-3.96)	0.45	1.68(0.60-4.71)	0.32
His/His	2(6.45)	2(4.88)	0.86(0.11-6.73)	0.89	0.93(0.12-7.48)	0.94
Arg/His+Arg/Arg	12(38.71)	19(46.34)	1.37(0.53-3.53)	0.52	1.52(0.57-4.05)	0.39
Anorexia						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	7(70.00)	60(63.83)	1.00(Reference)		1.00(Reference)	
Arg/His	3(30.00)	29(30.85)	1.13(0.27-4.68)	0.87	1.09(0.25-4.71)	0.91
His/His	0(0.00)	5(5.32)	1.36(0.07-27.19)	0.84	-	-
Arg/His+Arg/Arg	3(30.00)	34(36.17)	1.32(0.32-5.45)	0.70	1.37(0.32-5.78)	0.67
Alopecia						
Genotype rs9282861, *Arg^{213}His*	**Grade of toxicity (No), n(%)**	**Grade of toxicity (Yes), n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Arg/Arg	26(74.29)	55(70.51)	1.00(Reference)		1.00(Reference)	
Arg/His	8(22.86)	21(26.92)	1.24(0.48-3.17)	0.65	1.22(0.47-3.15)	0.67
His/His	1(2.86)	2(2.56)	0.94(0.08-10.91)	0.96	0.88(0.07-10.47)	0.92
Arg/His+Arg/Arg	9(25.71)	23(29.49)	1.21(0.49-2.97)	0.68	1.18(0.48-2.93)	0.71

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

5.8.3. Association of polymorphic *EPHX1* towards lung cancer outcomes

5.8.3.1. Association of *EPHX1* SNPs with overall survival of lung cancer patients

Overall survival analysis of 473 lung cancer patients was evaluated as shown in Table 5.450. Univariate analysis using Kaplan-Meier curves and the log-rank test was performed to evaluate the association of the *EPHX1* polymorphism towards overall survival in lung cancer patients, as shown in Table 5.50. The frequency of genotypes of lung cancer patients showing $Tyr^{113}His$ polymorphism was as follows: *Tyr/Tyr* (33.83%), *Tyr/His* (48.41%), *His/His* (17.76%). In the *EPHX1* $Tyr^{113}His$ polymorphism survival time of mutant genotype was almost similar to that of patients with wild genotype (8.40 months *vs.* 8.60 months). In the heterozygous genotype (*Tyr/His*), the MST was less than the wild-type genotype (7.57 months *vs.* 8.60 months). There was no significant association observed between $Tyr^{113}His$ polymorphism and OS. As shown in Table 5.50, the lung patients carrying mutant allele (*Arg/Arg*) for $His^{139}Arg$ were associated with more prolonged survival as compared to subjects with wild-type genotype (*His/His*) (MST= 9.77 months *vs.* 7.60 months, log-rank p=0.49). In the multivariate Cox proportional hazards regression analysis, no significant association was observed between the $His^{139}Arg$ mutant genotype and overall lung cancer survival (HR'=0.89; 95% C.I.=0.36-2.17; p=0.80). The heterozygous genotype (*His/Arg*) showed better MST than the wild genotype (MST=9.80 months *vs.* 7.60 months, log-rank p=0.07), and the multivariate cox-proportional hazards model exhibit a significant lower hazards ratio (HR'=0.77, 95% C.I.=0.61-0.98, p=**0.04;** Figure 5.27). Similarly, the combined genotype (*His/Arg+Arg/Arg*) also exhibits a higher MST as compared to patients carrying the wild genotype (*His/His*) (MST= 9.80 months *vs.* 7.60 months, log-rank p= 0.06). Further, applying the multivariate cox-proportional hazards regression analysis, a significant association was observed between the $His^{139}Arg$ combined genotype and overall lung cancer survival (HR'=0.77, 95% C.I.=0.61-0.98, p=**0.04**) (Table 5.50).

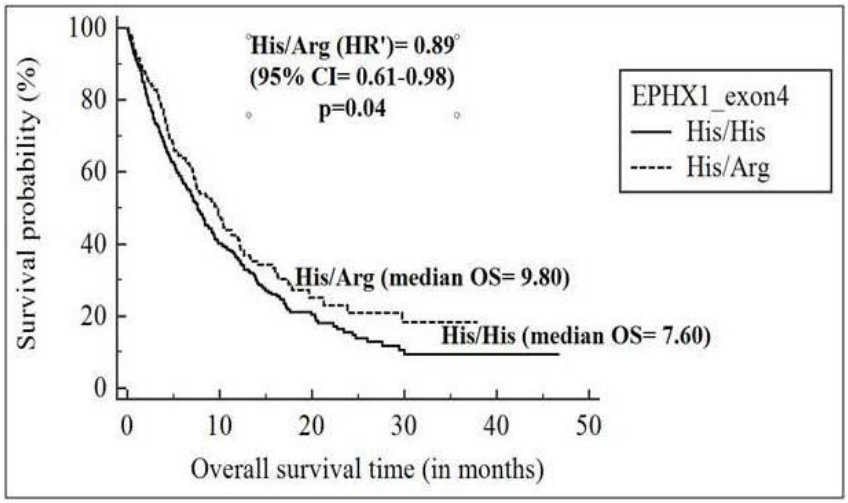


Figure 5.27: Cox hazards overall survival curve of heterozygous (*His/Arg*) genotype depicting *EPHX1 His*^{139}*Arg* polymorphism

Table 5.50: Association of *EPHX1* on overall survival in lung cancer cases and based on histology

OVERALL								
Genotype	**CASES**	**DEAD**	**ALIVE**	**Univariate analysis**			**Multivariate analysis**	
Exon3 ***Tyr^{113}His***	**n(%), N=473**	**n(%), N=362**	**n(%), N=111**	**MST (months)**	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b](95% C.I.)**	***p***
Tyr/Tyr	160 (33.83)	122 (33.70)	38 (34.23)	8.60				
Tyr/His	229 (48.41)	174 (48.07)	55 (49.55)	7.57	0.69	1.05 (0.83-1.32)	0.98 (0.77-1.26)	0.90
His/His	84 (17.76)	66 (18.23)	18 (16.22)	8.40	0.76	1.05 (0.77-1.42)	1.02 (0.74-1.40)	0.89
Tyr/His +His/His	313 (66.17)	240 (66.30)	73 (65.77)	8.07	0.67	1.05 (0.84-1.30)	0.99 (0.78-1.24)	0.91
Genotype	**CASES**	**DEAD**	**ALIVE**	**Univariate analysis**			**Multivariate analysis**	
Exon4 ***His^{139}Arg***	**n(%), N=473**	**n(%), N=362**	**n(%), N=111**	**MST (months)**	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
His/His	324 (34.32)	256 (33.51)	68 (37.14)	7.60				
His/Arg	141 (48.00)	101 (48.11)	40 (47.62)	9.80	0.07	0.81 (0.65-1.01)	0.77 (0.61-0.98)	**0.04**
Arg/Arg	8 (17.68)	5 (18.38)	3 (15.24)	9.77	0.49	0.73 (0.34-1.57)	0.89 (0.36-2.17)	0.80
His/Arg+Arg/Arg	149 (65.68)	106 (66.49)	43 (62.86)	9.80	0.06	0.81 (0.65-1.003)	0.77 (0.61-0.98)	**0.04**

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.8.3.2.Association of *EPHX1* SNPs with overall survival in lung cancer histological subtypes

To evaluate the relevance of these polymorphic sites in predicting the survival of patients in different subgroups stratified based on histology, the association of *EPHX1* variants with the survival of patients was analyzed, and data tabulated in Table 5.51. The data for each histological subtype has been summarized in the Table given below. In *Tyr*^{113}*His*, the percentage of heterozygous patients was 46.07%, 50.00%, and 50.00% among ADCC, SQCC, and SCLC subtypes. It was found that ADCC patients carrying the heterozygous (*Tyr/His*) genotype showed a better OS than the wild genotype (9.13 months *vs.* 7.60 months). Hazards ratio obtained from both univariate and multivariate analysis was 0.87 and 0.81 with non-significant *p* values of 0.48 and 0.33. On the contrary, the MST of SQCC subjects carrying mutant (*His/His*) genotype was 8.63 months, which was a bit lower than 9.13 months as in subjects having wild (*Tyr/Tyr*) genotype. However, the hazard ratio obtained after univariate and multivariate analysis was not significantly associated with the genotype.

The MST of heterozygous genotype for SQCC patients was the least (MST=6.47 months, log-rank p=0.11), and a 1.25 fold higher death rate was obtained using multivariate analysis (p=0.28). At the same time, SCLC patients with heterozygous genotype showed a slightly better OS than wild type (MST=9.10 months *vs.* 7.23 months, log-rank *p*=0.32) (Table 5.51). This polymorphic site did not predict survival in ADCC, SQCC, and SCLC tumor types.

In the case of *His*^{139}*Arg* polymorphism, the percentage of mutant genotype was 1.68%, 19.79%, and 3.06% among ADCC, SQCC, and SCLC subjects. ADCC subjects carrying the mutant genotype (*Arg/Arg*) had almost double MST as compared to patients carrying wild genotype (*His/His*) (MST= 16.17 months *vs*. 8.13 months, log-rank p= 0.24) pertaining to a non-significant but lower hazard ratio (HR'=0.55; 95% C.I.= 0.07-4.12). The survival of ADCC patients carrying heterozygous and combined genotypes was higher, *i.e.*, 10.30 months and 10.67 months, respectively. However, the hazards ratio obtained from the univariate and multivariate analysis was not significantly associated with these genotypes. SQCC subjects carrying the mutant genotype showed the least MST (2.27 months, log-rank p=0.34) (Table 5.51). The higher hazards rate was not significant (HR'=1.51; 95% C.I.= 0.36-6.23). It was seen that SCLC patients carrying the heterozygous genotype (*His/Arg*) had a longer OS than the patients carrying the homozygous wild-type genotype (*His/His*). Thus, better MST was observed for such

individuals than subjects with wild-type genotype (*His/His*) (MST=11.30 *vs*. 6.73 months; Log Rank **p=0.02,** Figure 5.28 a)). In Cox's proportional hazards model, after adjusting for various covariates like age, sex, smoking status, ECOG, KPS, and tumor stage, the *EPHX1 His*^{139}*Arg* polymorphism for OS still has the prognostic significance. Compared with the *His/His* genotype, the subjects harboring the *His/Arg+Arg/Arg* genotype had a significantly lower risk of death, as compared with the SCLC subjects with wild type genotype with HRs being (HR'=0.57, 95% C.I.=0.33-1.005, p=**0.05**; MST: 9.93 months; Figure 5.28 b)) (Table 5.51). Thus, this polymorphic site was found to act as an independent predictor of survival only in SCLC patients and not among ADCC and SQCC patients.

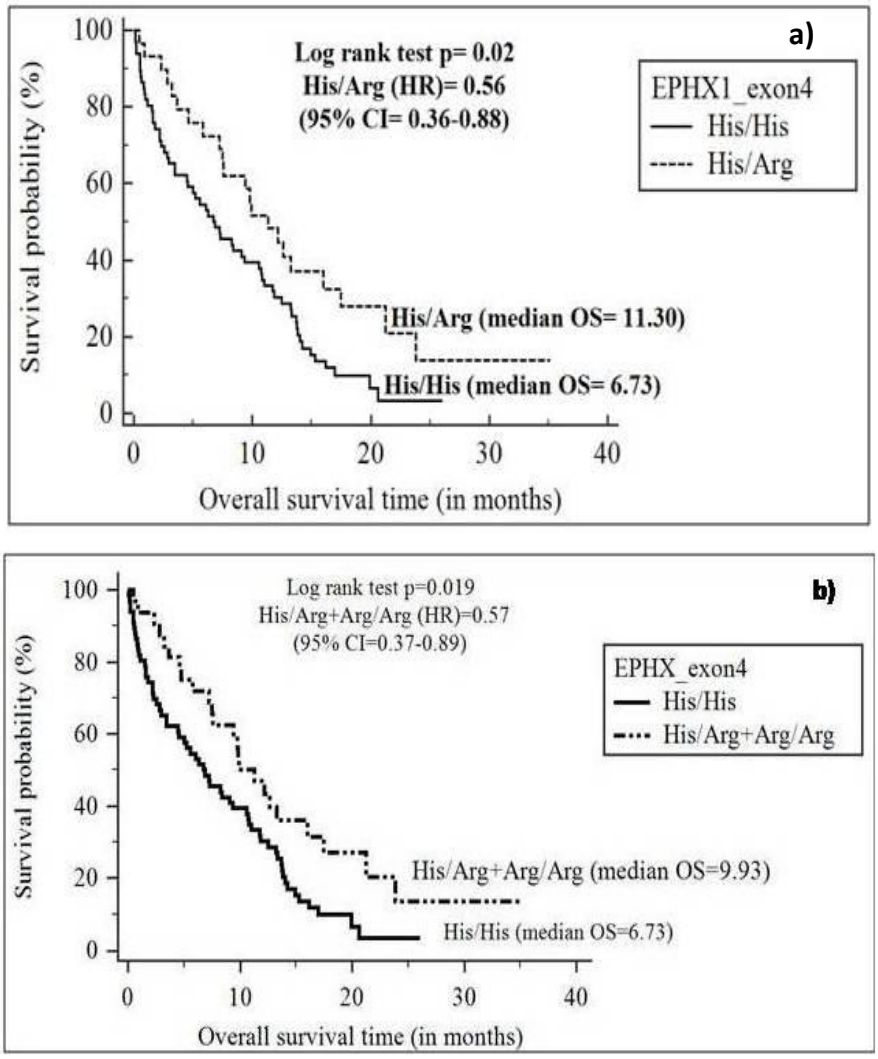


Figure 5.28: Kaplan Meier curves illustrating the association between overall survival in patients with a) heterozygous (*His/Arg*) genotype, b) combined genotype (*His/Arg+Arg/Arg)*, for SCLC showing *EPHX1 His*^{139}*Arg* polymorphism

Table 5.51: Association of *EPHX1* SNPs on overall survival according to tumor histology

ADCC								
Genotype Exon3	**CASES n(%), N=178**	**DEAD n(%), N=127**	**ALIVE n(%), N=51**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Tyr/Tyr	66 (37.08)	49 (38.58)	17 (33.33)	7.60				
Tyr/His	82 (46.07)	55 (43.31)	27 (52.94)	9.13	0.48	0.87 (0.59-1.28)	0.81 (0.54-1.23)	0.33
His/His	30 (16.85)	23 (18.11)	7 (13.73)	8.23	0.99	1.0006(0.61-1.64)	0.87 (0.51-1.50)	0.63
Tyr/His +His/His	112 (62.92)	78 (61.42)	34 (66.67)	8.30	0.59	0.91 (0.63-1.30)	0.82 (0.56-1.20)	0.30
Genotype Exon4	**CASES n(%), N=178**	**DEAD n(%), N=127**	**ALIVE n(%), N=51**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
His/His	120 (67.42)	89 (70.08)	31 (60.78)	8.13				
His/Arg	55 (30.90)	37 (29.13)	18 (35.30)	10.30	0.71	0.93 (0.64-1.36)	0.87 (0.57-1.33)	0.53
Arg/Arg	3 (1.68)	1 (0.79)	2 (3.92)	16.17	0.24	0.32 (0.10-1.03)	0.55 (0.07-4.12)	0.57
His/Arg+Arg/Arg	58 (32.58)	38 (29.92)	20 (39.22)	10.67	0.54	0.89 (0.61-1.29)	0.85 (0.56-1.28)	0.44
SQCC								
Genotype Exon3	**CASES n(%), N=192**	**DEAD n(%), N=148**	**ALIVE n(%), N=44**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Tyr/Tyr	58 (30.21)	41 (27.70)	17 (38.64)	9.13				
Tyr/His	96 (50.00)	76 (51.35)	20 (45.45)	6.47	0.11	1.36 (0.94-1.96)	1.25 (0.84-1.87)	0.28
His/His	38 (19.79)	31 (20.95)	7 (15.91)	8.63	0.30	1.28 (0.79-2.06)	1.45 (0.87-2.42)	0.15
Tyr/His +His/His	134 (69.79)	107 (72.30)	27 (61.36)	7.00	0.11	1.33 (0.95-1.88)	1.29 (0.88-1.87)	0.19
Genotype Exon4	**CASES n(%), N=192**	**DEAD n(%), N=148**	**ALIVE n(%), N=44**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
His/His	133 (69.27)	104 (70.27)	29 (65.91)	7.60				
His/Arg	57 (29.69)	42 (28.38)	15 (34.09)	8.50	0.34	0.84 (0.59-1.19)	0.74 (0.51-1.07)	0.11
Arg/Arg	2 (1.04)	2 (1.35)	0 (0.00)	2.27	0.34	1.96 (0.28-13.59)	1.51 (0.36-6.23)	0.57
His/Arg+Arg/Arg	59 (30.73)	44 (29.73)	15 (34.09)	8.37	0.41	0.86 (0.61-1.22)	0.75 (0.52-1.09)	0.13
SCLC								
Genotype Exon3	**CASES n(%), N=98**	**DEAD n(%), N=84**	**ALIVE n(%), N=14**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Tyr/Tyr	34 (34.69)	30 (35.71)	4 (28.57)	7.23				
Tyr/His	49 (50.00)	42 (50.00)	7 (50.00)	9.10	0.32	0.79 (0.49-1.29)	0.89 (0.53-1.53)	0.69
His/His	15 (15.31)	12 (14.29)	3 (21.43)	8.40	0.22	0.68 (0.36-1.28)	0.77 (0.34-1.74)	0.52

Tyr/His +His/His	64 (65.31)	54 (64.29)	10 (71.43)	8.40	0.25	0.78 (0.49-1.24)	0.83 (0.50-1.36)	0.45
Genotype	**CASES**	**DEAD**	**ALIVE**	**Univariate analysis**			**Multivariate analysis**	
Exon4	**n(%), N=98**	**n(%), N=84**	**n(%), N=14**	**MST (months)**	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
His/His	66 (67.35)	60 (71.43)	6 (42.86)	6.73				
His/Arg	29 (29.59)	22 (26.19)	7 (50.00)	11.30	**0.02**	0.56 (0.36-0.88)	0.55 (0.31-0.99)	**0.048**
Arg/Arg	3 (3.06)	2 (2.38)	1 (7.14)	9.77	0.71	0.77 (0.22-2.69)	1.12 (0.25-5.08)	0.88
His/Arg+Arg/Arg	32 (32.65)	24 (28.57)	8 (57.14)	9.93	**0.019**	0.57 (0.37-0.89)	0.57 (0.33-1.005)	**0.05**

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, stage, KPS, ECOG

5.8.3.3. Association of genetic determinants of *EPHX1* and overall survival of lung cancer patients treated with different chemotherapeutic regimens

Table 5.52 below show the effect of the two SNPs of *EPHX1* on survival after stratifying for the chemotherapeutic regimens used for the treatment of lung cancer patients. Chemotherapy included four primary regimens, which were administered to different sets of patients. The four regimens used included a platinum-based drug (cisplatin and carboplatin) common to all lung cancer patients. These patients also received a second-line chemotherapeutic agent like docetaxel paclitaxel, pemetrexed, or irinotecan. MST (median survival time) was evaluated for the two polymorphic variants using univariate analysis, while multivariate analysis evaluated the hazard rate.

As shown in Table 5.52, patients receiving cisplatin/ carboplatin along with docetaxel and carrying the mutant genotype *(His/His)* of *EPHX11 Tyr^{113}His* gene had a lower survival time as compared to subjects carrying the wild genotype (MST=8.63 months *vs*. 9.83 months, log-rank p=0.28). Single allelic carriers *(Tyr/His)* for the same gene also showed a lower survival time of 6.47 months and hazards rate of 1.2 fold (H.R=1.20, 95% C.I..=0.62-2.29). Whereas, for the *EPHX1 His^{139}Arg* polymorphism, the heterozygous genotype (*His/Arg*) showed a higher survival than the wild genotype (*His/His*) (9.83 months *vs.* 8.40 months, log-rank p=0.82) and a lower death hazards rate which was not significant (HR'=1.01, 95% C.I.= 0.49-2.08, p=0.97). The subjects carrying both the allelic carriers (*Arg/Arg*) for the same gene showed almost similar MST to wild genotype (8.37 months and 8.40 months, log-rank p=0.70). This genotype was not associated with lung cancer risk for subjects administered docetaxel+cisplatin/carboplatin.

For the patients administered with paclitaxel and cis/carboplatin, the single allelic carrier (*His/Arg*) for the *EPHX1 His^{139}Arg* polymorphism exhibits a higher survival than subjects carrying wild genotype (12.67 months *vs.* 5.83 months, log-rank p=0.09). The Cox regression model gave a shorter death rate (HR'=0.60, 95% C.I.=0.23-1.54) for heterozygous genotype but lacks statistical significance.

As shown in Table 5.52, patients carrying the heterozygous genotype *(Tyr/His)* for *EPHX11 Tyr^{113}His* gene and who were receiving cisplatin/carboplatin along with irinotecan had a longer survival time than the wild genotype (11.30 months *vs.* 7.23 months, log-rank p=**0.007**) (Figure 5.29) and a lower death rate (HR'=0.52, 95% C.I..=0.23-1.17, p=**0.01**) indicating a better survival. Similarly, the combined genotype for this polymorphism (*Tyr/His+His/His*) also had

higher survival of 10.80 months and a shorter death rate (HR'=0.46, 95% C.I.=0.22-0.94, p=**0.03**) (Table 5.52). Double allelic carriers *(His/His)* for the same gene showed a meager survival time of 2.8 months. However, the hazards ratio obtained from the univariate and multivariate analysis was not significantly associated with this genotype. The single allelic carrier (*His/Arg*) of *EPHX1 $His^{139}Arg$* also showed a more prolonged survival as compared to subjects carrying wild genotype (*His/His*) (11.30 months *vs.* 7.23 months, log-rank p=0.15). The lower death rate was seen using the multivariate cox proportional analysis, but it was not significant (HR'=0.71, 95% C.I.=0.33-1.50).

An MST of 14.23 months was observed for the lung cancer patients carrying the mutant genotype (*His/His*) for *EPHX11 $Tyr^{113}His$* polymorphism and treated with pemetrexed cisplatin/carboplatin regimen. This mutant genotype showed a lower probability of risk when compared with the wild-type genotype (*Tyr/Tyr*) (HR'= 0.71, 95% C.I.=0.35-1.44). The subjects possessing the heterozygous genotype (*His/Arg*) had lower survival in comparison to patients carrying wild genotype (*His/His*) for *EPHX1 $His^{139}Arg$* polymorphism (6.33 months *vs*. 9.43 months, log-rank p=0.43). However, the data was not found to be significant. However, no significant relationship was observed between *EPHX1 $His^{139}Arg$*polymorphism and OS in lung cancer treated with different chemotherapy regimens, as shown in Table 5.52. Whereas, in the case of *EPHX11 $Tyr^{113}His$* polymorphism, the heterozygous genotype was associated with higher survival for patients administered with irinotecan +cisplatin/carboplatin.

Table 5.52: Association of *EPHX1*genotype on survival in lung cancer cases based on the chemotherapeutic regimen

Docetaxel								
Genotype Exon3	**CASES n(%), N=70**	**DEAD n(%), N=57**	**ALIVE n(%), N=13**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Tyr/Tyr	15 (21.43)	12 (21.05)	3 (23.08)	9.83				
Tyr/His	37 (52.86)	29 (50.88)	8 (61.54)	6.47	0.59	1.20 (0.62-2.29)	0.61 (0.27-1.34)	0.22
His/His	18 (25.71)	16 (28.07)	2 (15.38)	8.63	0.28	1.48 (0.71-3.12)	1.12 (0.43-2.92)	0.81
Tyr/His+ His/His	55 (78.57)	45 (78.95)	10 (76.92)	8.43	0.45	1.27 (0.70-2.31)	0.71 (0.35-1.45)	0.35
Genotype Exon4	**CASES n(%), N=70**	**DEAD n(%), N=57**	**ALIVE n(%), N=13**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
His/His	55 (78.57)	43 (75.44)	12 (92.31)	8.40				
His/Arg	14 (20.00)	13 (22.81)	1 (7.69)	9.83	0.82	1.07 (0.57-2.02)	1.01 (0.49-2.08)	0.97
Arg/Arg	1 (1.43)	1 (1.75)	0 (0.00)	8.37	0.70	1.47 (0.13-16.11)	1.21 (0.16-9.15)	0.85
His/Arg+Arg/Arg	15 (21.43)	14 (24.56)	1 (7.69)	9.83	0.75	1.10 (0.59-2.04)	1.03 (0.51-2.07)	0.93
Paclitaxel								
Genotype Exon3	**CASES n(%), N=57**	**DEAD n(%), N=39**	**ALIVE n(%), N=18**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Tyr/Tyr	21 (36.84)	13 (33.33)	8 (44.44)	7.57				
Tyr/His	26 (45.62)	19 (48.72)	7 (38.89)	8.40	0.74	1.12 (0.56-2.26)	1.004(0.44-2.27)	0.99
His/His	10 (17.54)	7 (17.95)	3 (16.67)	8.07	0.86	1.09 (0.43-2.76)	0.92 (0.22-3.81)	0.91
Tyr/His + His/His	36 (63.16)	26 (66.67)	10 (55.56)	8.40	0.71	1.13 (0.59-2.18)	0.93 (0.44-1.95)	0.85
Genotype Exon4	**CASES n(%), N=57**	**DEAD n(%), N=39**	**ALIVE n(%), N=18**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
His/His	42 (73.68)	31 (79.49)	11 (61.11)	5.83				
His/Arg	15 (26.32)	8 (20.51)	7 (38.89)	12.67	0.09	0.52 (0.27-1.01)	0.60 (0.23-1.54)	0.29
Arg/Arg	0 (0.00)	0 (0.00)	0 (0.00)		-	-		
His/Arg+ Arg/Arg	15 (26.32)	8 (20.51)	7 (38.89)	12.67	0.09	0.52 (0.27-1.01)	0.60 (0.23-1.54)	0.29

Irinotecan								
Genotype Exon3	**CASES n(%), N=61**	**DEAD n(%), N=51**	**ALIVE n(%), N=10**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Tyr/Tyr	21 (34.43)	20 (39.22)	1 (10.00)	7.23				
Tyr/His	30 (49.18)	23 (45.10)	7 (70.00)	11.30	**0.007**	0.46 (0.24-0.90)	0.52 (0.23-1.17)	0.01
His/His	10 (16.39)	8 (15.68)	2 (20.00)	2.80	0.39	0.72 (0.33-1.57)	0.55 (0.17-1.78)	0.32
Tyr/His + His/His	40 (65.57)	31 (60.78)	9 (90.00)	10.80	**0.018**	0.53 (0.28-0.99)	0.46 (0.22-0.94)	**0.03**
Genotype Exon4	**CASES n(%), N=61**	**DEAD n(%), N=51**	**ALIVE n(%), N=10**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
His/His	42 (68.85)	36 (70.59)	6 (60.00)	7.23				
His/Arg	19 (31.15)	15 (29.41)	4 (40.00)	11.30	0.15	0.65 (0.37-1.14)	0.71 (0.33-1.50)	0.37
Arg/Arg	0 (0.00)	0 (0.00)	0 (0.00)	-	-	-		
His/Arg+ Arg/Arg	19 (31.15)	15 (29.41)	4 (40.00)	11.30	0.15	0.65 (0.37-1.14)	0.71 (0.33-1.50)	0.37
Pemetrexed								
Genotype Exon3	**CASES n(%), N=108**	**DEAD n(%), N=79**	**ALIVE n(%), N=29**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)		**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Tyr/Tyr	38 (35.19)	27 (34.18)	11 (37.93)	7.60				
Tyr/His	49 (45.37)	37 (46.83)	12 (41.38)	8.13	0.76	1.08 (0.66-1.77)	0.93 (0.55-1.58)	0.80
His/His	21 (19.44)	15 (18.99)	6 (20.69)	14.23	0.44	0.78 (0.42-1.44)	0.71 (0.35-1.44)	0.34
Tyr/His + His/His	70 (64.81)	52 (65.82)	18 (62.07)	9.43	0.93	0.98 (0.61-1.56)	0.87 (0.53-1.41)	0.56
Genotype Exon4	**CASES n(%), N=108**	**DEAD n(%), N=79**	**ALIVE n(%), N=29**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
His/His	73 (67.59)	54(68.35)	19 (65.52)	9.43				
His/Arg	34 (31.48)	25 (31.65)	9 (31.03)	6.33	0.43	1.20 (0.74-1.97)	1.05 (0.62-1.78)	0.85
Arg/Arg	1 (0.93)	0 (0.00)	1 (3.45)	-	0.35	-	0.00(3.06E-179-8.95E+168)	0.96
His/Arg+ Arg/Arg	35 (32.41)	25 (31.65)	10 (34.48)	9.13	0.54	1.16 (0.71-1.89)	1.01 (0.60-1.71)	0.96

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

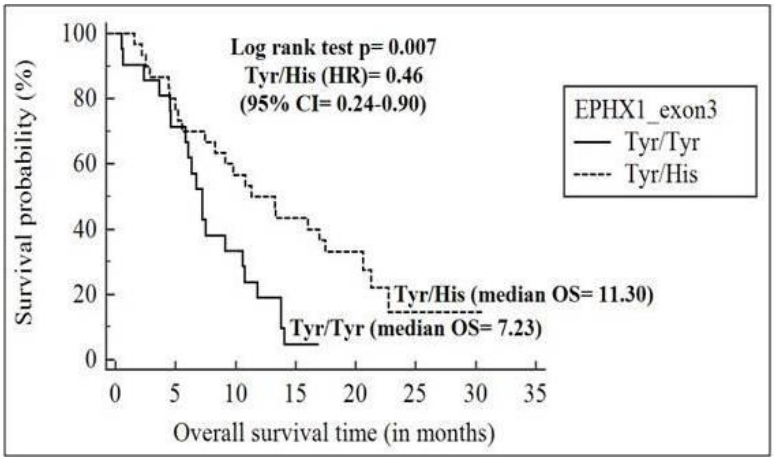


Figure 5.29: Kaplan Meier curves illustrating the association between overall survival in patients with heterozygous (*Tyr/His*) genotype for patients treated with cisplatin/carboplatin alongwith irinotecan depicting *EPHX1Tyr*^{113}*His* polymorphism

5.8.3.4. Association of *EPHX1* polymorphism and chemotherapy response

The patients were analyzed to respond to the chemotherapeutic regimen to their genotypes for different *EPHX1* variants, as shown in Table 5.53. The patients were divided into good and bad responders based on the RECIST criteria, where they are annotated into four groups, CR (complete response), PR (partial response), SD (stable disease), and PD (progressive disease). The proportion of good (CR+PR) and poor (SD+PD) responders was almost similar. However, the adjusted odds ratio was calculated for each polymorphic site, taking the wild genotype as the reference, and confounding factors were age, gender, smoking, stage, histology, KPS, ECOG, Chemotherapeutic regimen. The objective response data were available only for 266 patients in this case. None of the polymorphic sites of the *EPHX1* gene revealed a significant difference in the frequency of patients showing good and poor responses. In the case of *EPHX1 Tyr*^{113}*His*, the proportion of patients having *Tyr/His+His/His* genotype was slightly higher (66.10%) among good responders compared to poor responders (60.81%), but the difference was not found to be statistically significant.

In comparison, the representation of heterozygotes among good and bad responders was almost similar (43.22% and 41.22%, respectively). In *EPHX1 His*^{139}*Arg*, the percentage of subjects carrying combined (*His/Arg+Arg/Arg*) genotype was 28.81% among good responders, constituting 31.76% of good responders. Hence, no significant difference in the frequency of variant genotype was observed among the two groups, as shown in Table 5.53.

Table 5.53: Relationship of genotypic distribution with clinic-pathological parameter based on response to chemotherapy

	Response of chemotherapy		**AOR (95% C.I.)** [a]	***p*** [b]
Genotype ***Tyr*^{113}*His***	**CR+PR (118), n (%)**	**SD+PD (148), n (%)**		
Co-dominant model				
Tyr/Tyr	40 (33.90)	58 (39.19)	1.0(Reference)	
Tyr/His	51 (43.22)	61 (41.22)	1.15 (0.66-2.00)	0.63
His/His	27 (22.88)	29 (19.59)	1.32 (0.67-2.58)	0.42
Dominant model				
Tyr/Tyr	40 (33.90)	58 (39.19)	1.0(Reference)	
Tyr/His+His/His	78 (66.10)	90 (60.81)	1.21 (0.73-2.02)	0.46
Recessive model				
Tyr/Tyr+Tyr/His	91 (77.12)	119 (80.41)	1.0(Reference)	
His/His	27 (22.88)	29 (19.59)	1.20 (0.66-2.18)	0.54
Genotype ***His*^{139}*Arg***	**CR+PR (118), n (%)**	**SD+PD (148), n (%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Co-dominant model				
His/His	84 (71.19)	101 (68.24)	1.0(Reference)	
His/Arg	32 (27.12)	44 (29.73)	0.91 (0.52-1.58)	0.74
Arg/Arg	2 (1.69)	3 (2.03)	0.90 (0.14-5.71)	0.91
Dominant model				
His/His	84 (71.19)	101 (68.24)	1.0(Reference)	
His/Arg + Arg/Arg	34 (28.81)	47 (31.76)	0.89 (0.52-1.53)	0.67
Recessive model				
His/His + His/Arg	116 (98.31)	145 (97.97)	1.0(Reference)	
Arg/Arg	2 (1.69)	3 (2.03)	0.86 (0.14-5.31)	0.87

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding p-values were calculated by logistic regression analysis after adjusting for age and gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table

5.8.3.5. Association between *EPHX1* polymorphism and toxicity

In the case of *EPHX1*, two polymorphisms *viz.* *Tyr*^{113}*His* and *His*^{139}*Arg* were studied in association with hematological toxicity, nephrotoxicity, and gastrointestinal toxicity, as demonstrated in Table 5.54 a), Table 5.54 b), and Table 5.54 c), respectively. The subjects carrying heterozygous genotype (*His/Arg*) for *His*^{139}*Arg* polymorphism 2 fold increased risk of hematological toxicity, especially leukopenia, when compared grades 1-3 *vs*. grade 0 (AOR=2.03, 95% C.I.=0.37-10.98, p=0.41), but was not significant. Further the variant genotype His/His for *Tyr*^{113}*His* polymorphism (AOR=1.98, 95% C.I.=0.60-6.21, p= 0.27) and variant genotype (Arg/Arg) of *His*^{139}*Arg* polymorphism (AOR=2.19, 95% C.I.=0.12-40.04, p=0.59) was found to pose an approximately two-fold increased risk for hematological toxicity, basically anemia when comparison was made between grades 1-4 and no toxicity (grade 0) but was not statistically significant (Table 5.54 a)).

Table 5.54 a): Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *EPHX1* polymorphisms in multivariate analysis

Leukopenia

Genotype *Tyr^{113}His* (exon3)	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3,N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype *His^{139}Arg* (exon4)**	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Tyr/Tyr	69(36.51)	32(38.09)	1.00(Reference)		1.00(Reference)		*His/His*	122(64.55)	54(64.29)	1.00(Reference)		1.00(Reference)	
Tyr/His	89(47.09)	37(44.05)	0.89(0.51-1.58)	0.71	0.80(0.45-1.44)	0.46	*His/Arg*	64(33.86)	27(32.14)	0.95(0.55-1.65)	0.86	0.92(0.52-1.61)	0.76
His/His	31(16.40)	15(17.86)	1.04(0.49-2.20)	0.91	1.04(0.49-2.22)	0.92	*Arg/Arg*	3(1.59)	3(3.57)	2.26(0.44-11.55)	0.33	2.03(0.37-10.98)	0.41
Tyr/Tyr+Tyr/His	120(63.49)	52(61.90)	0.93(0.55-1.59)	0.80	0.87(0.51-1.50)	0.62	*His/Arg + Arg/Arg*	67(35.45)	30(35.71)	1.01(0.59-1.73)	0.97	0.95(0.55-1.65)	0.86
Genotype *Tyr^{113}His* (exon3)	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3,n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype *His^{139}Arg* (exon4)**	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Tyr/Tyr	90(36.44)	11(42.31)	1.00(Reference)		1.0 (Reference)		*His/His*	155(63.01)	21(77.77)	1.00(Reference)		1.00(Reference)	
Tyr/His	114(46.15)	12(46.15)	0.86(0.36-2.04)	0.73	0.67(0.27-1.67)	0.39	*His/Arg*	86(34.96)	5(18.52)	0.43(0.16-1.18)	0.10	0.40(0.14-1.12)	0.08
His/His	43(17.41)	3(11.54)	0.57(0.15-2.15)	0.41	0.52(0.13-2.02)	0.35	*Arg/Arg*	5(2.03)	1(3.70)	1.48(0.16-13.25)	0.73	1.45(0.14-15.19)	0.76
Tyr/Tyr+Tyr/His	157(63.56)	15(57.69)	0.78(0.34-1.77)	0.56	0.66(0.28-1.53)	0.33	*His/Arg + Arg/Arg*	91(36.99)	6(22.22)	0.48(0.19-1.25)	0.13	0.44(0.17-1.17)	0.10

Anemia

Genotype *Tyr^{113}His* (exon3)	**Grade of toxicity 0, n (%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype *His^{139}Arg* (exon4)**	**Grade of toxicity 0, n (%)**	**Grade of toxicity 1-4, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Tyr/Tyr	16(37.21)	85(36.96)	1.00(Reference)		1.00(Reference)		*His/His*	25(58.14)	151(65.65)	1.00(Reference)		1.00(Reference)	
Tyr/His	23(53.49)	103(44.78)	0.84(0.42-1.70)	0.63	0.83(0.40-1.71)	0.62	*His/Arg*	18(41.86)	73(31.74)	0.67(0.34-1.31)	0.24	0.59(0.30-1.18)	0.13
His/His	4(9.30)	42(18.26)	1.98(0.62-6.28)	0.25	1.98(0.60-6.21)	0.27	*Arg/Arg*	0(0.00)	6(2.61)	2.19(0.12-40.04)	0.59	-	-
Tyr/Tyr+Tyr/His	27(62.79)	145(63.04)	1.01(0.51-1.98)	0.97	1.02(0.51-2.02)	0.97	*His/Arg + Arg/Arg*	18(41.86)	79(34.35)	0.73(0.37-1.41)	0.35	0.64(0.32-1.26)	0.19
Genotype *Tyr^{113}His* (exon3)	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype *His^{139}Arg* (exon4)**	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**

Tyr/Tyr	54(40.30)	47(33.81)	1.00(Reference)		1.00(Reference)		*His/His*	88(65.67)	88(63.31)	1.00(Reference)		1.00(Reference)	
Tyr/His	60(44.77)	66(47.48)	1.26(0.75-2.13)	0.38	1.29(0.76-2.22)	0.34	*His/Arg*	43(32.09)	48(34.53)	1.12(0.67-1.85)	0.67	1.13(0.67-1.89)	0.65
His/His	20(14.93)	26(18.71)	1.49(0.74-3.01)	0.26	1.53(0.75-3.13)	0.24	*Arg/Arg*	3(2.24)	3(2.16)	1.00(0.19-5.09)	1.00	1.32(0.24-7.26)	0.75
Tyr/Tyr+Tyr/His	80(59.70)	92(66.19)	1.32(0.81-2.16)	0.27	1.36(0.83-2.26)	0.22	*His/Arg + Arg/Arg*	46(34.33)	51(36.69)	1.11(0.67-1.82)	0.68	1.13(0.68-1.88)	0.63
Thrombocytopenia													
Genotype *Tyr[113]His* (exon3)	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype *His[139]Arg* (exon4)**	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Tyr/Tyr	75(36.95)	26(37.68)	1.00(Reference)		1.00(Reference)		*His/His*	127(62.56)	48(69.57)	1.00(Reference)		1.00(Reference)	
Tyr/His	98(48.28)	27(39.13)	0.79(0.43-1.47)	0.46	0.88(0.47-1.67)	0.70	*His/Arg*	71(34.98)	20(28.99)	0.74(0.41-1.35)	0.33	0.86(0.46-1.59)	0.63
His/His	30(14.78)	16(23.19)	1.54(0.72-3.27)	0.26	1.75(0.80-3.83)	0.16	*Arg/Arg*	5(2.46)	1(1.45)	0.53(0.06-4.65)	0.56	0.71(0.07-6.67)	0.76
Tyr/Tyr+Tyr/His	128(63.05)	43(62.32)	0.97(0.55-1.70)	0.91	1.08(0.60-1.93)	0.80	*His/Arg + Arg/Arg*	76(37.44)	21(30.43)	0.73(0.41-1.31)	0.29	0.86(0.47-1.57)	0.62
Genotype *Tyr[113]His* (exon3)	**Grade of toxicity0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**		**Genotype *His[139]Arg* (exon4)**	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Tyr/Tyr	96(36.23)	5(71.43)	1.00(Reference)		1.00(Reference)		*His/His*	171(64.53)	4(57.14)	1.00(Reference)		1.00(Reference)	
Tyr/His	124(46.79)	1(14.29)	0.15(0.02-1.35)	0.09	0.19(0.02-1.70)	0.14	*His/Arg*	88(33.21)	3(42.86)	1.46(0.32-6.66)	0.63	1.65(0.35-7.80)	0.53
His/His	45(16.98)	1(14.29)	0.43(0.05-3.76)	0.44	0.49(0.05-4.41)	0.53	*Arg/Arg*	6(2.26)	0(0.00)	2.93(0.14-60.37)	0.48	-	-
Tyr/Tyr+Tyr/His	169(63.77)	2(28.57)	0.23(0.04-1.19)	0.08	0.27(0.05-1.42)	0.12	*His/Arg + Arg/Arg*	94(35.47)	3(42.86)	1.36(0.30-6.22)	0.69	1.56(0.33-7.37)	0.57
Neutropenia													
Genotype *Tyr[113]His* (exon3)	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His[139]Arg* (exon4)**	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Tyr/Tyr	74(36.27)	21(36.21)	1.00(Reference)		1.00(Reference)		*His/His*	130(63.73)	38(65.52)	1.00(Reference)		1.00(Reference)	
Tyr/His	93(45.59)	29(50.00)	1.10(0.58-2.08)	0.77	0.97(0.50-1.87)	0.92	*His/Arg*	71(34.80)	17(29.31)	0.82(0.43-1.55)	0.54	0.78(0.41-1.51)	0.47
His/His	37(18.13)	8(13.79)	0.76(0.31-1.88)	0.55	0.73(0.29-1.82)	0.50	*Arg/Arg*	3(1.47)	3(5.17)	3.42(0.66-17.65)	0.14	3.89(0.69-21.99)	0.12
Tyr/Tyr+T	130(63.72)	37(63.79)	1.00(0.55-1.84)	0.99	0.93(0.50-1.72)	0.81	*His/Arg +*	74(36.27)	20(34.48)	0.92(0.50-1.71)	0.80	0.86(0.46-1.62)	0.64

yr/His							*Arg/Arg*						
Genotype *Tyr[113]His* (exon3)	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His[139]Arg* (exon4)**	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Tyr/Tyr	80(35.56)	15(40.54)	1.00(Reference)		1.00(Reference)		*His/His*	142(63.11)	26(70.27)	1.00(Reference)		1.00(Reference)	
Tyr/His	105(46.67)	17(45.95)	0.86(0.41-1.83)	0.70	0.78(0.36-1.70)	0.54	*His/Arg*	79(35.11)	9(24.32)	0.62(0.28-1.39)	0.25	0.60(0.26-1.37)	0.23
His/His	40(17.78)	5(13.51)	0.67(0.23-1.96)	0.46	0.66(0.22-1.98)	0.46	*Arg/Arg*	4(1.78)	2(5.41)	2.73(0.47-15.68)	0.26	2.68(0.43-16.81)	0.29
Tyr/Tyr+Tyr/His	145(64.44)	22(59.46)	0.81(0.40-1.65)	0.56	0.77(0.37-1.59)	0.48	*His/Arg + Arg/Arg*	83(36.89)	11(29.73)	0.72(0.34-1.54)	0.40	0.69(0.32-1.50)	0.35

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratio, data were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

A trend of increased risk towards nephrotoxicity was observed in the case of subjects possessing heterozygous genotype (*Tyr/His*) for *Tyr*^{113}*His* polymorphism on comparing intermediate/severe GFR grade 2-4 with low GFR grade 0-1 (AOR=1.96, 95% C.I.=0.87-4.39, p=0.10) (Table 5.54 b)). Further, on comparison of anorexia with grade 2-3 *vs*. grade 1, two-fold increased risk of anorexia for subjects harboring heterozygous as well as combined genotype for *Tyr*^{113}*His* polymorphism was found with an odds of 2.32 (95% C.I.=0.48-11.07, p=0.29) and 2.01 (95% C.I.=0.52-7.66, p=0.31) (Table 5.54 c)). But overall, both the polymorphisms (*viz*. *Tyr*^{113}*His* and *His*^{139}*Arg*) of the *EPHX1* gene were not statistically associated with hematological toxicity, nephrotoxicity, or gastrointestinal toxicity.

Table 5.54 b): Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *EPHX1* polymorphisms in multivariate analysis

Nephrotoxicity													
Genotype *Tyr[113]His* (exon3)	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His[139]Arg* (exon4)**	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Tyr/Tyr	32(42.11)	56(35.44)	1.00(Reference)		1.00(Reference)		*His/His*	44(57.89)	102(64.56)	1.00(Reference)		1.00(Reference)	
Tyr/His	34(44.74)	73(46.20)	1.22(0.68-2.22)	0.50	1.62(0.80-3.28)	0.17	*His/Arg*	28(36.84)	54(34.18)	0.83(0.47-1.48)	0.53	0.88(0.46-1.69)	0.70
His/His	10(13.15)	29(18.35)	1.66(0.71-3.84)	0.24	1.83(0.72-4.70)	0.20	*Arg/Arg*	4(5.26)	2(1.27)	0.22(0.04-1.22)	0.08	0.32(0.04-2.70)	0.30
Tyr/Tyr+Tyr/His	44(57.89)	102(64.56)	1.32(0.76-2.32)	0.32	1.71(0.88-3.30)	0.11	*His/Arg + Arg/Arg*	32(42.10)	56(35.44)	0.75(0.43-1.32)	0.32	0.83(0.44-1.59)	0.58
Genotype *Tyr[113]His* (exon3)	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His[139]Arg* (exon4)**	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Tyr/Tyr	71(39.23)	17(32.08)	1.00(Reference)		1.00(Reference)		*His/His*	108(59.67)	38(71.70)	1.00(Reference)		1.00(Reference)	
Tyr/His	80(44.20)	27(50.94)	1.41(0.71-2.80)	0.33	1.96(0.87-4.39)	0.10	*His/Arg*	68(37.57)	14(26.42)	0.58(0.29-1.16)	0.12	0.52(0.23-1.14)	0.10
His/His	30(16.57)	9(16.98)	1.25(0.50-3.12)	0.63	1.37(0.46-4.09)	0.57	*Arg/Arg*	5(2.76)	1(1.89)	0.57(0.06-5.02)	0.61	1.19(0.10-14.26)	0.89
Tyr/Tyr+Tyr/His	110(60.77)	36(67.92)	1.37(0.71-2.62)	0.34	1.80(0.83-3.91)	0.14	*His/Arg + Arg/Arg*	73(40.33)	15(28.30)	0.58(0.30-1.14)	0.11	0.54(0.25-1.17)	0.12

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratio, data were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

Table 5.54 c): Gastrointestinal Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *EPHX1*polymorphisms in multivariate analysis

Diarrohea

Genotype *Tyr^{113}His* (exon3)	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His^{139}Arg* (exon4)**	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Tyr/Tyr	18(41.86)	21(32.31)	1.00(Reference)		1.00(Reference)		*His/His*	28(65.12)	47(72.31)	1.00(Reference)		1.00(Reference)	
Tyr/His	17(39.53)	34(52.31)	1.71(0.73-4.04)	0.22	1.67(0.70-3.96)	0.25	*His/Arg*	13(30.23)	17(26.15)	0.78(0.33-1.84)	0.57	0.76(0.31-1.89)	0.56
His/His	8(18.60)	10(15.38)	1.07(0.35-3.29)	0.90	1.15(0.36-3.66)	0.81	*Arg/Arg*	2(4.65)	1(1.54)	0.30(0.03-3.44)	0.33	0.25(0.01-5.29)	0.38
Tyr/Tyr+Tyr/His	25(58.14)	44(67.69)	1.51(0.68-3.35)	0.31	1.50(0.67-3.36)	0.32	*His/Arg + Arg/Arg*	15(34.88)	18(27.69)	0.71(0.31-1.64)	0.43	0.79(0.32-1.93)	0.60
Genotype *Tyr^{113}His* (exon3)	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His^{139}Arg* (exon4)**	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Tyr/Tyr	27(35.06)	12(38.71)	1.00(Reference)		1.00(Reference)		*His/His*	52(67.53)	23(74.19)	1.00(Reference)		1.00(Reference)	
Tyr/His	38(49.35)	13(41.94)	0.77(0.30-1.94)	0.58	0.79(0.31-2.03)	0.62	*His/Arg*	22(28.57)	8(25.81)	0.82(0.32-2.12)	0.68	0.75(0.28-2.05)	0.58
His/His	12(15.58)	6(19.35)	1.12(0.34-3.71)	0.85	1.18(0.35-3.94)	0.79	*Arg/Arg*	3(3.90)	0(0.00)	0.32(0.01-6.43)	0.46	0.81(0.30-2.20)	0.68
Tyr/Tyr+Tyr/His	50(64.94)	19(61.29)	0.85(0.36-2.02)	0.72	0.90(0.37-2.16)	0.81	*His/Arg + Arg/Arg*	25(32.47)	8(25.81)	0.72(0.28-1.84)	0.49	-	-

Constipation

Genotype *Tyr^{113}His* (exon3)	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His^{139}Arg* (exon4)**	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Tyr/Tyr	11(35.48)	11(26.83)	1.00(Reference)		1.00(Reference)		*His/His*	22(70.97)	26(63.41)	1.00(Reference)		1.00(Reference)	
Tyr/His	14(45.16)	21(51.22)	1.50(0.51-4.39)	0.46	1.58(0.52-4.78)	0.41	*His/Arg*	9(29.03)	15(36.59)	1.41(0.52-3.84)	0.50	1.42(0.51-3.99)	0.50
His/His	6(19.35)	9(21.95)	1.50(0.40-5.66)	0.55	1.61(0.38-6.79)	0.52	*Arg/Arg*	0(0.00)	0(0.00)	0.85(0.02-44.54)	0.93	-	-
Tyr/Tyr+Tyr/His	20(64.52)	30(73.17)	1.50(0.55-4.11)	0.43	1.52(0.54-4.34)	0.43	*His/Arg + Arg/Arg*	9(29.03)	15(36.59)	1.41(0.52-3.84)	0.50	1.42(0.51-3.99)	0.50

Anorexia

Genotype *Tyr^{113}His* (exon3)	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His^{139}Arg* (exon4)**	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**

Tyr/Tyr	5(50.00)	31(32.98)	1.00(Reference)		1.00(Reference)		*His/His*	8(80.00)	63(67.02)	1.00(Reference)		1.00(Reference)	
Tyr/His	3(30.00)	42(44.68)	2.26(0.50-10.17)	0.29	2.32(0.48-11.07)	0.29	*His/Arg*	2(20.00)	30(31.91)	1.90(0.38-9.52)	0.43	1.51(0.28-8.07)	0.63
His/His	2(20.00)	21(22.34)	1.69(0.30-9.56)	0.55	1.67(0.28-9.83)	0.57	*Arg/Arg*	0(0.00)	1(1.06)	0.40(0.02-10.67)	0.58	1.84	0.99
Tyr/Tyr+Tyr/His	5(50.00)	63(67.02)	2.03(0.55-7.55)	0.29	2.01(0.52-7.66)	0.31	*His/Arg + Arg/Arg*	2(20.00)	31(32.98)	1.97(0.39-9.83)	0.41	1.51(0.28-8.08)	0.63
Alopecia													
Genotype *Tyr[113]His* (exon3)	**Grade of toxicity(No), n(%)**	**Grade of toxicity (Yes), n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype *His[139]Arg* (exon4)**	**Grade of toxicity (No), n(%)**	**Grade of toxicity (Yes), n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Tyr/Tyr	11(31.43)	22(28.21)	1.00(Reference)		1.00(Reference)		*His/His*	22(62.86)	57(73.08)	1.00(Reference)		1.00(Reference)	
Tyr/His	16(45.71)	41(52.56)	1.28(0.51-3.23)	0.60	1.28(0.49-3.34)	0.61	*His/Arg*	13(37.14)	21(26.92)	0.62(0.27-1.46)	0.27	0.68(0.28-1.64)	0.39
His/His	8(22.86)	15(19.23)	0.94(0.30-2.88)	0.91 `	0.91(0.29-2.85)	0.88	*Arg/Arg*	0(0.00)	0(0.00)	0.39(0.007-20.33)	0.64		
Tyr/Tyr+Tyr/His	24(68.57)	56(71.79)	1.17(0.49-2.78)	0.73	1.15(0.47-2.79)	0.76	*His/Arg + Arg/Arg*	13(37.14)	21(26.92)	0.62(0.27-1.46)	0.27	0.68(0.28-1.64)	0.39

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratio, data were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

5.8.4. Association of *NAT2* polymorphisms with lung cancer outcomes

5.8.4.1. Association between the *NAT2* Polymorphisms and Prognosis in the Lung Cancer Patients

Overall survival analysis was carried out for the *NAT2* polymorphic sites, where 473 subjects were recruited. MST was estimated for each genotypic group and compared with the reference group using the Kaplan-Meier method and obtaining log-rank *p*. Also, the adjusted hazards ratio associated with the variant genotype was estimated using Cox proportional hazard after adjusting for age, gender, smoking status, ECOG, KPS, and tumor stage. None of the SNPs showed any contribution in modulating the MST of the particular genotypic group, as shown in Table 5.558.

In the case of *NAT2*6 (590G>A*), the percentage of subjects who were either dead or alive and carried heterozygous genotype was 48.07% and 43.24%, respectively (Table 5.55). The subjects carrying heterozygous genotype (*NAT2*4/*6*) had a better survival time (9.20 months *vs*. 7.53 months, log-rank p=0.79) as compared to the wild (*NAT2*4/*4*) genotype, which was not significant and hence, there was no association of this genotype with the survival of lung cancer patients. Whereas the MST of mutant and wild genotype (7.57 months and 7.53 months, log-rank p=0.57) was almost the same, no significant association was found with the survival of lung cancer subjects harboring mutant genotype. Further, in the case of *NAT2*7 (857G>A)*, the percentage of dead and alive patients who were mutant for this SNP was 11.21%. On comparing MST of subjects having different genotypes, it was evident that wild genotype patients showed a slightly higher MST of 8.40 months compared to 8.23 months in the case of mutant genotype. However, the hazards ratio obtained from both univariate (HR=0.86, 95% C.I.=0.61-1.23, log-rank p=0.43) and multivariate (HR=0.99, 95% C.I.=0.66-1.47) analysis lacked statistical significance. The proportion of dead and alive patients carrying mutant genotype for *NAT2*5B (481C>T)* was 11.60% and 12.61%, respectively, as shown in Table 5.55. The trend, in this case, showcased that the mutant genotype (*NAT2*5B/*5B*) had a better MST (9.13 months) as compared to the wild genotype (8.13 months). The difference in the MST was found to lack statistical significance, and also, the hazards ratio obtained showed that this polymorphic site could not act as an independent predictor of survival. For the *NAT2*5C (803A>G)* polymorphism, the percentage of patients who were either dead or alive and carried two copies of the variant allele was 12.71% and 14.41%, respectively, as shown in Table 5.55.

Table 5.55: Association of *NAT2* on overall survival in lung cancer cases and on the basis of histology

OVERALL								
Genotype rs1799930 *590G>A (NAT2*6)*	**CASES n(%), N=473**	**DEAD n(%), N=362**	**ALIVE n(%), N=111**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***p***
*NAT2 *4/*4*	192 (40.59)	147 (40.61)	45 (40.54)	7.53				
*NAT2 *4/*6*	222 (46.93)	174 (48.07)	48 (43.24)	9.20	0.79	0.97 (0.78-1.21)	1.02 (0.81-1.29)	0.86
*NAT2 *6/*6*	59 (12.48)	41 (11.32)	18 (16.22)	7.57	0.57	0.90 (0.64-1.27)	1.01 (0.69-1.46)	0.98
*NAT2 *4/*6 + *6/*6*	281 (59.41)	215 (59.39)	66 (59.46)	8.63	0.69	0.96 (0.78-1.18)	1.01 (0.81-1.26)	0.95
Genotype rs1799931 *857G>A (NAT2*7)*	**CASES n(%), N=473**	**DEAD n(%), N=362**	**ALIVE n(%), N=111**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	171 (36.15)	139 (38.40)	32 (28.83)	8.40				
*NAT2 *4/*7*	249 (52.64)	187 (51.66)	62 (55.86)	7.57	0.80	1.03 (0.82-1.28)	1.06 (0.84-1.34)	0.59
*NAT2 *7/*7*	53 (11.21)	36 (9.94)	17 (15.31)	8.23	0.43	0.86 (0.61-1.23)	0.99 (0.66-1.47)	0.95
*NAT2 *4/*7 + *7/*7*	302 (63.85)	223 (61.60)	79 (71.17)	7.60	0.95	0.99 (0.80-1.23)	1.04 (0.83-1.30)	0.72
Genotype rs1799929 *481C>T (NAT2*5B)*	**CASES n(%), N=473**	**DEAD n(%), N=362**	**ALIVE n(%), N=111**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	196 (41.44)	159 (43.92)	37 (33.33)	8.13				
*NAT2 *4/*5B*	221 (46.72)	161 (44.48)	60 (54.05)	7.60	0.39	0.91 (0.73-1.13)	0.89 (0.71-1.13)	0.35
*NAT2 *5B/*5B*	56 (11.84)	42 (11.60)	14 (12.61)	9.13	0.29	0.83 (0.60-1.15)	0.75 (0.52-1.08)	0.13
*NAT2 *4/*5B + *5B/*5B*	277 (58.56)	203 (56.08)	74 (66.67)	8.40	0.28	0.89 (0.72-1.10)	0.87 (0.70-1.08)	0.21
Genotype rs1208 *803A>G (NAT2*5C)*	**CASES n(%), N=473**	**DEAD n(%), N=362**	**ALIVE n(%), N=111**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	236 (49.89)	185 (51.10)	51 (45.95)	9.13				
*NAT2 *4/*5C*	175 (37.00)	131 (36.19)	44 (39.64)	6.90	0.34	1.11 (0.89-1.39)	1.15 (0.91-1.46)	0.23
*NAT2 *5C/*5C*	62 (13.11)	46 (12.71)	16 (14.41)	7.57	0.93	1.01 (0.73-1.40)	1.05 (0.74-1.48)	0.78
*NAT2 *4/*5C + *5C/*5C*	237 (50.11)	177 (48.90)	60 (54.05)	7.27	0.41	1.09 (0.89-1.34)	1.12 (0.90-1.39)	0.29

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio after adjusting for age, sex, smoking, histology, stage, KPS, ECOG

The subjects having wild genotype (9.13 months) were found to have better MST than the heterozygous (6.90 months) and mutant genotype (7.57 months) for this polymorphism. However, the difference in MST was not statistically significant, and hence the associated hazards ratio both obtained from univariate and multivariate analysis for these genotypes lacked statistical significance, as shown in Table 5.55.

5.8.4.2. Association of *NAT2* polymorphism and overall survival according to the histological subtype of lung cancer patients

The four *NAT2* variants were analyzed for finding their relevance in predicting the prognosis of lung cancer patients with different histological subtypes; the data for each histological subtype. For *NAT2*6 (590G>A*), the percentage of subjects with mutant genotype was 13.48%, 9.89%, and 16.33% in ADCC, SQCC, and SCLC subtype, respectively (Table 5.56 a)). ADCC patients having mutant genotype (*NAT2*6/*6*) showed poor survival as compared to ADCC patients having wild genotype (*NAT2*4/*4*) (5.80 months *vs.* 7.70 months). Hazards ratio obtained from both univariate (HR=1.14, 95% C.I.=0.65-2.00, log-rank p=0.64) and multivariate analysis (HR'=1.66, 95% C.I.=0.92-3.02) was found not to be statistically significant as shown in Table 5.56 a). However, in the case of SQCC patients, the MST of the mutant genotype (7.80 months) was more than the wild type (6.47 months) but less than the heterozygous genotype (8.40 months), but these genotypes did not show any significant association with the mortality rate. Further, in the case of SCLC patients, the MST of mutant and wild-type patients is the same (7.53 months), and no significant association was reported among this genetic variant and overall survival of SCLC patients (Table 5.56 a)).

In the case of *NAT2*7 (857G>A*), the proportion of mutants in ADCC, SQCC, and SCLC subtype was 11.80%, 9.89%, and 12.24%, respectively (Table 5.56 b)). ADCC patients with mutant genotype showed a lower MST of 7.80 months than wild subtype (9.70 months). The hazards ratio obtained by univariate and multivariate analysis was not found to be significant. Whereas, in the case of SQCC subjects, the MST of mutant genotype (9.53 months) was highest than wild (8.07 months) as well as heterozygous genotype (7.17 months). Whereas in the case of SCLC, the mutant genotype showed a much lower MST of 2.53 months compared to wild type (9.10 months). The adjusted hazards ratio obtained by Cox regression showed a two-fold increased rate of mortality (HR'=1.99, 95% C.I.=0.84-4.75) as shown in Table 5.56 b).

Table 5.56 a): Association of *NAT2*6 590G>A* polymorphism on overall survival according to tumor histology

ADCC								
Genotype rs1799930 *590G>A (NAT2*6)*	**CASES n(%), N=178**	**DEAD n(%), N=127**	**ALIVE n(%), N=51**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	75 (42.13)	54 (42.52)	21 (41.18)	7.70				
*NAT2 *4/*6*	79 (44.38)	56 (44.09)	23 (45.10)	9.77	0.50	0.88 (0.60-1.28)	0.98 (0.65-1.46)	0.92
*NAT2 *6/*6*	24 (13.48)	17 (13.39)	7 (13.72)	5.80	0.64	1.14 (0.65-2.00)	1.66 (0.92-3.02)	0.09
*NAT2 *4/*6 + *6/*6*	103 (57.87)	73 (57.48)	30 (58.82)	9.43	0.71	0.93 (0.66-1.33)	1.09 (0.75-1.59)	0.64
SQCC								
Genotype rs1799930 *590G>A (NAT2*6)*	**CASES n(%), N=192**	**DEAD n(%), N=148**	**ALIVE n(%), N=44**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	82 (42.71)	64 (43.24)	18 (40.91)	6.47				
*NAT2 *4/*6*	91 (47.40)	71 (47.98)	20 (45.45)	8.40	0.76	0.95 (0.68-1.33)	0.86 (0.60-1.22)	0.40
*NAT2 *6/*6*	19 (9.89)	13 (8.78)	6 (13.64)	7.80	0.65	0.87 (0.49-1.54)	0.83 (0.44-1.56)	0.56
*NAT2 *4/*6 + *6/*6*	110 (57.29)	84 (56.76)	26 (59.09)	8.20	0.70	0.94 (0.68-1.30)	0.85 (0.60-1.19)	0.34
SCLC								
Genotype rs1799930 *590G>A (NAT2*6)*	**CASES n(%), N=98**	**DEAD n(%), N=84**	**ALIVE n(%), N=14**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	31 (31.63)	26 (30.95)	5 (35.71)	7.53				
*NAT2 *4/*6*	51 (52.04)	47 (55.95)	4 (28.57)	8.27	0.74	1.08 (0.67-1.74)	1.21 (0.70-2.08)	0.50
*NAT2 *6/*6*	16 (16.33)	11 (13.10)	5 (35.71)	7.53	0.29	0.68 (0.35-1.33)	0.63 (0.26-1.54)	0.32
*NAT2 *4/*6 + *6/*6*	67 (68.37)	58 (69.05)	9 (64.29)	8.40	0.84	0.95 (0.60-1.52)	0.99 (0.59-1.67)	0.98

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, stage, KPS, ECOG

Table 5.56 b): Association of *NAT2*7 857G>A* polymorphism on overall survival according to tumor histology

ADCC								
Genotype rs1799931 *857G>A (NAT2*7)*	**CASES n(%), N=178**	**DEAD n(%), N=127**	**ALIVE n(%), N=51**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	59 (33.15)	42 (33.07)	17 (33.33)	9.70				
*NAT2 *4/*7*	98 (55.05)	71 (55.91)	27 (52.94)	7.60	0.45	1.15 (0.79-1.68)	1.30 (0.85-1.98)	0.23
*NAT2 *7/*7*	21 (11.80)	14 (11.02)	7 (13.73)	7.80	0.79	1.08 (0.58-2.01)	0.90 (0.43-1.88)	0.79
*NAT2 *4/*7 + *7/*7*	119 (66.85)	85 (66.93)	34 (66.67)	7.80	0.48	1.14 (0.79-1.64)	1.20 (0.80-1.81)	0.38
SQCC								
Genotype rs1799931 *857G>A (NAT2*7)*	**CASES n(%), N=192**	**DEAD n(%), N=148**	**ALIVE n(%), N=44**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	69 (35.94)	58 (39.19)	11 (25.00)	8.07				
*NAT2 *4/*7*	104 (54.17)	78 (52.70)	26 (59.09)	7.17	0.98	1.005(0.71-1.41)	1.14 (0.78-1.66)	0.51
*NAT2 *7/*7*	19 (9.89)	12 (8.11)	7 (15.91)	9.53	0.17	0.65 (0.37-1.12)	1.16 (0.59-2.26)	0.67
*NAT2 *4/*7 + *7/*7*	123 (64.06)	90 (60.81)	33 (75.00)	7.57	0.65	0.93 (0.66-1.29)	1.15 (0.80-1.66)	0.45
SCLC								
Genotype rs1799931 *857G>A (NAT2*7)*	**CASES n(%), N=98**	**DEAD n(%), N=84**	**ALIVE n(%), N=14**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b](95% C.I.)**	***P***
*NAT2 *4/*4*	42 (42.86)	38 (45.24)	4 (28.57)	9.10				
*NAT2 *4/*7*	44 (44.90)	36 (42.86)	8 (57.14)	7.47	0.89	1.03 (0.65-1.63)	1.02 (0.62-1.67)	0.95
*NAT2 *7/*7*	12 (12.24)	10 (11.90)	2 (14.29)	2.53	0.39	1.35 (0.63-2.90)	1.99 (0.84-4.75)	0.12
*NAT2 *4/*7 + *7/*7*	56 (57.14)	46 (54.76)	10 (71.43)	7.23	0.71	1.08 (0.71-1.66)	1.12 (0.70-1.79)	0.63

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, stage, KPS, ECOG

The following variant *NAT2*5B (481C>T*) was reported to lack the ability to act as an independent predictor of survival in any of the three histological subtypes (Table 5.56 c)). The proportion of mutant genotype (*NAT2*5B/*5B*) among ADCC, SQCC, and SCLC subtype was 14.05%, 10.42%, and 10.20%, respectively. ADCC patients having double susceptible allele for this genetic variant showed a higher MST than wild type (9.43 months *vs*. 7.80 months) and almost similar to that of heterozygous genotype (9.20 months). Nevertheless, the difference between mutant and heterozygous genotypes lacked statistical significance. Similarly, in SQCC subjects, the mutant genotype had the highest MST of 8.83 months compared to 7.93 months for the wild genotype. The univariate (HR=0.74, 95% C.I.=0.45-1.23, log-rank p=0.28) and multivariate analysis (HR'=0.73, 95% C.I.=0.39-1.36) did not show any statistical significance to predict the survival of lung cancer patients. The mutant genotype exhibits a slightly lower MST of 8.40 months than the wild genotype (9.10 months) in SCLC. However, the least survival in the SCLC subtype was found for the subjects carrying heterozygous genotype (6.07 months). However, there was no evident significant association as the log-rank p-value was 0.95 and 0.97 for mutant and heterozygous genotypes, respectively (Table 5.56 c)).

The genetic variant *NAT2*5C (803A>G)* showed a lack of association with lung cancer patients' overall survival in any histological subtypes. The MST of ADCC subjects with a mutant (*NAT2*5C/5C*) genotype was slightly shorter than wild-type subjects (7.57 months *vs*. 9.13 months), as shown in Table 5.56 d). The SQCC patients carrying mutant genotype showed a survival time almost similar to wild type (8.50 months and 8.60 months). Whereas, the heterozygous genotype exhibit the lowest survival of 5.27 months for SQCC subjects. The mutant genotype characterized with the SCLC subtype of lung cancer depicted a lower MST of 5.00 months, and the wild type showed the median survival time of 9.40 months. The log-rank test reported insignificant p-values in the case of each histological subtype, as shown in Table 5.56 d).

Table 5.56 c): Association of *NAT2*5B 481C>T* polymorphism on overall survival according to tumor histology

ADCC								
Genotype rs1799929 *481C>T (NAT2*5B)*	**CASES n(%), N=178**	**DEAD n(%), N=127**	**ALIVE n(%), N=51**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	69 (38.76)	56 (44.09)	13 (25.49)	7.80				
*NAT2 *4/*5B*	84 (47.19)	55 (43.31)	29 (56.86)	9.20	0.36	0.84 (0.58-1.22)	0.82 (0.55-1.23)	0.35
*NAT2 *5B/*5B*	25 (14.05)	16 (12.60)	9 (17.65)	9.43	0.52	0.83 (0.49-1.42)	0.71 (0.38-1.35)	0.30
*NAT2 *4/*5B + *5B/*5B*	109 (61.24)	71 (55.91)	38 (74.51)	9.43	0.32	0.84 (0.59-1.19)	0.80 (0.55-1.18)	0.26
SQCC								
Genotype rs1799929 *481C>T (NAT2*5B)*	**CASES n(%), N=192**	**DEAD n(%), N=148**	**ALIVE n(%), N=44**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	79 (41.14)	63 (42.57)	16 ()	7.93				
*NAT2 *4/*5B*	93 (48.44)	69 (46.62)	24 ()	7.17	0.53	0.90 (0.64-1.26)	0.91 (0.64-1.31)	0.63
*NAT2 *5B/*5B*	20 (10.42)	16 (10.81)	4 ()	8.83	0.28	0.74 (0.45-1.23)	0.73 (0.39-1.36)	0.32
*NAT2 *4/*5B + *5B/*5B*	113 (58.86)	85 (57.43)	28 ()	7.60	0.38	0.86 (0.62-1.20)	0.87 (0.61-1.23)	0.44
SCLC								
Genotype rs1799929 *481C>T (NAT2*5B)*	**CASES n(%), N=98**	**DEAD n(%), N=84**	**ALIVE n(%), N=14**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	46 (46.94)	39 (46.43)	7 (50.00)	9.10				
*NAT2 *4/*5B*	42 (42.86)	36 (42.86)	6 (42.86)	6.07	0.97	1.01 (0.64-1.59)	0.86 (0.50-1.45)	0.57
*NAT2 *5B/*5B*	10 (10.20)	9 (10.71)	1 (7.14)	8.40	0.95	1.02 (0.49-2.12)	0.72 (0.31-1.65)	0.44
*NAT2 *4/*5B + *5B/*5B*	52 (53.06)	45 (53.57)	7 (50.00)	6.87	0.95	1.01 (0.66-1.55)	0.85 (0.52-1.40)	0.53

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, stage, KPS, ECOG

Table 5.56 d): Association of *NAT2*5C 803A>G* polymorphism on overall survival according to tumor histology

ADCC								
Genotype	**CASES**	**DEAD**	**ALIVE**	**Univariate analysis**			**Multivariate analysis**	
rs1208	**n(%), N=178**	**n(%),**	**n(%), N=51**	**MST (months)**	**Log rank**	**Unadjusted HR[a]**	**Adjusted HR[b]**	***P***
803A>G (NAT2*5C)		**N=127**			***p***		**(95% C.I.)**	
*NAT2 *4/*4*	83 (46.63)	63 (49.61)	20 (39.22)	9.13				
*NAT2 *4/*5C*	68 (38.20)	44 (34.64)	24 (47.06)	8.23	0.74	0.94 (0.64-1.37)	0.91 (0.60-1.37)	0.66
*NAT2 *5C/*5C*	27 (15.17)	20 (15.75)	7 (13.72)	7.57	0.93	1.02 (0.61-1.69)	1.05 (0.62-1.79)	0.86
*NAT2 *4/*5C + *5C/*5C*	95 (53.37)	64 (50.39)	31 (60.78)	8.23	0.85	0.97 (0.68-1.37)	0.96 (0.66-1.39)	0.81
SQCC								
Genotype	**CASES**	**DEAD**	**ALIVE**	**Univariate analysis**			**Multivariate analysis**	
rs1208	**n(%), N=192**	**n(%),**	**n(%), N=44**	**MST (months)**	**Log rank**	**Unadjusted HR[a]**	**Adjusted HR[b]**	***P***
803A>G (NAT2*5C)		**N=148**			***p***		**(95% C.I.)**	
*NAT2 *4/*4*	101 (52.60)	77 (52.03)	24 (54.55)	8.60				
*NAT2 *4/*5C*	70 (36.46)	57 (38.51)	13 (29.55)	5.27	0.20	1.25 (0.88-1.78)	1.34 (0.94-1.91)	0.10
*NAT2 *5C/*5C*	21 (10.94)	14 (9.46)	7 (15.90)	8.50	0.66	0.88 (0.51-1.52)	1.09 (0.57-2.09)	0.78
*NAT2 *4/*5C + *5C/*5C*	91 (47.40)	71 (47.97)	20 (45.45)	6.73	0.36	1.16 (0.84-1.60)	1.25 (0.90-1.75)	0.18
SCLC								
Genotype	**CASES**	**DEAD**	**ALIVE**	**Univariate analysis**			**Multivariate analysis**	
rs1208	**n(%), N=98**	**n(%), N=84**	**n(%), N=14**	**MST (months)**	**Log rank**	**Unadjusted HR[a]**	**Adjusted HR[b]**	***P***
803A>G (NAT2*5C)					***p***		**(95% C.I.)**	
*NAT2 *4/*4*	50 (51.02)	43 (51.19)	7 (50.00)	9.40				
*NAT2 *4/*5C*	36 (36.73)	30 (35.71)	6 (42.86)	5.20	0.17	1.37 (0.84-2.23)	1.33 (0.81-2.19)	0.26
*NAT2 *5C/*5C*	12 (12.25)	11 (13.10)	1 (7.14)	5.00	0.31	1.40 (0.67-2.93)	1.41 (0.70-2.84)	0.34
*NAT2 *4/*5C + *5C/*5C*	48 (48.98)	41 (48.81)	7 (50.00)	5.20	0.13	1.38 (0.90-2.14)	1.36 (0.86-2.14)	0.19

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, stage, KPS, ECOG

5.8.4.3. Association of *NAT2* SNPs with chemotherapy regimens and OS

In the present study, since the majority of the lung cancer patients were treated with platinum-based doublet chemotherapy, we wanted to assess the modulating effects on the association between chemotherapy and overall patient survival, stratified by the *NAT2* polymorphisms. The effect of four SNPs on survival after stratifying according to the chemotherapy regimens. Chemotherapy included four primary regimens, which were administered to different sets of patients. The four regimens used included a platinum-based drug (cisplatin or carboplatin) common to all lung cancer patients, with these patients receiving a second-line chemotherapeutic agent like docetaxel, paclitaxel, pemetrexed, or irinotecan. It was observed that patients who were receiving cisplatin/carboplatin along with paclitaxel and were carrying both the variant alleles for the *857G>A (NAT2*7)* and *(NAT2*7/*7*) genotype had a shorter survival time with an MST of 2.23 months as compared to patients carrying the homozygous wild type (*NAT2*4/*4*) with MST of 8.40 months (MST=2.23 *vs*. 8.40 Log-rank p=0.21) (Table 5.57 a)). After applying the cox regression multivariate analysis after adjusting for various covariates, a significant but higher hazards ratio was obtained (HR'=23.38; 95% C.I.=3.11-175.42, **p=0.0023;** Figure 5.30 a)). Similarly, in the case of the *803A>G (NAT2*5C*) polymorphism of the *NAT2* gene, it was observed that lung cancer patients administered with cisplatin/carboplatin along with paclitaxel exhibit a lower survival time of 4.33 months, especially in those subjects who were carrying heterozygous genotype (*NAT2*4/*5C*) as compared to subjects harboring the wild type genotype (*NAT2*4/*4*) (MST=4.33 vs. 9.27 months; Log-rank **p= 0.04;** Figure 5.30 b)). In the multivariate Cox proportional hazards model, we found that the hazard ratio was significantly higher for these patients (HR'=2.55; 95% C.I.=1.05-6.18, **p=0.04**) (Table 5.57 a)). Since the number of patients carrying the mutant genotype was less, we combined both the mutant and heterozygous genotypes as a single group (*NAT2*4/*5C +*5C/*5C*). The analysis data obtained showcased a poor survival of the above group when compared with the wild-type genotype (MST=4.90 vs. 9.27 months), and cox regression model data reported a significantly higher death rate (HR'=2.29, 95% C.I.=1.00-5.24, **p=0.04**) (Table 5.57 a)).

Table 5.57 a): Association of *NAT2* genotype on survival in lung cancer cases based on the chemotherapeutic regimen (Paclitaxel + cis/carboplatin)

Paclitaxel								
Genotype rs1799930 *590G>A (NAT2*6)*	**CASES n(%), N=57**	**DEAD n(%), N=39**	**ALIVE n(%), N=18**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b](95% C.I.)**	***P***
*NAT2 *4/*4*	21 (36.84)	15 (38.46)	6 (33.33)	5.37				
*NAT2 *4/*6*	27 (47.37)	19 (48.72)	8 (44.45)	11.47	0.48	0.78 (0.39-1.57)	0.49 (0.22-1.09)	0.08
*NAT2 *6/*6*	9 (15.79)	5 (12.82)	4 (22.22)	9.27	0.43	0.66 (0.26-1.69)	0.32 (0.07-1.42)	0.14
*NAT2 *4/*6 + *6/*6*	36 (63.16)	24 (61.54)	12 (66.67)	9.27	0.41	0.76 (0.39-1.49)	0.50 (0.23-1.10)	0.09
Genotype rs1799931 *857G>A (NAT2*7)*	**CASES n(%), N=57**	**DEAD n(%), N=39**	**ALIVE n(%), N=18**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	19 (33.33)	15 (38.46)	4 (22.22)	8.40				
*NAT2 *4/*7*	34 (59.65)	21 (53.85)	13 (72.22)	7.57	0.21	0.66 (0.33-1.34)	1.02 (0.45-2.33)	0.96
*NAT2 *7/*7*	4 (7.02)	3 (7.69)	1 (5.56)	2.23	0.62	1.36 (0.34-5.44)	23.38 (3.11-175.42)	**0.0023**
*NAT2 *4/*7 + *7/*7*	38 (66.67)	24 (61.54)	14 (77.78)	6.90	0.30	0.72 (0.36-1.42)	1.21 (0.54-2.70)	0.64
Genotype rs1799929 *481C>T (NAT2*5B)*	**CASES n(%), N=57**	**DEAD n(%), N=39**	**ALIVE n(%), N=18**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	21 (36.84)	14 (35.90)	7 (38.89)	8.60				
*NAT2 *4/*5B*	30 (52.63)	20 (51.28)	10 (55.56)	8.40	0.80	0.92 (0.46-1.83)	0.82 (0.41-1.66)	0.59
*NAT2 *5B/*5B*	6 (10.53)	5 (12.82)	1 (5.55)	3.27	0.52	1.39 (0.45-4.24)	0.93 (0.26-3.26)	0.91
*NAT2 *4/*5B + *5B/*5B*	36 (63.16)	25 (64.10)	11 (61.11)	8.07	0.99	0.99 (0.52-1.91)	0.85 (0.44-1.66)	0.64
Genotype	**CASES**	**DEAD**	**ALIVE**	**Univariate**	**Multivariate**			

rs1208 ***803A>G (NAT2*5C)***	**n(%), N=57**	**n(%), N=39**	**n(%), N=18**	**analysis** **MST (months)**	**analysis** **Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	31 (54.39)	19 (48.72)	12 (66.67)	9.27				
*NAT2 *4/*5C*	19 (33.33)	16 (41.02)	3 (16.67)	4.33	**0.04**	1.94 (0.94-4.00)	2.55 (1.05-6.18)	**0.04**
*NAT2 *5C/*5C*	7 (12.28)	4 (10.26)	3 (16.66)	8.50	0.89	0.92 (0.32-2.65)	1.01 (0.27-3.74)	0.99
*NAT2 *4/*5C + *5C/*5C*	26 (45.61)	20 (51.28)	6 (33.33)	4.90	0.13	1.61 (0.85-3.06)	2.29 (1.00-5.24)	**0.05**

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

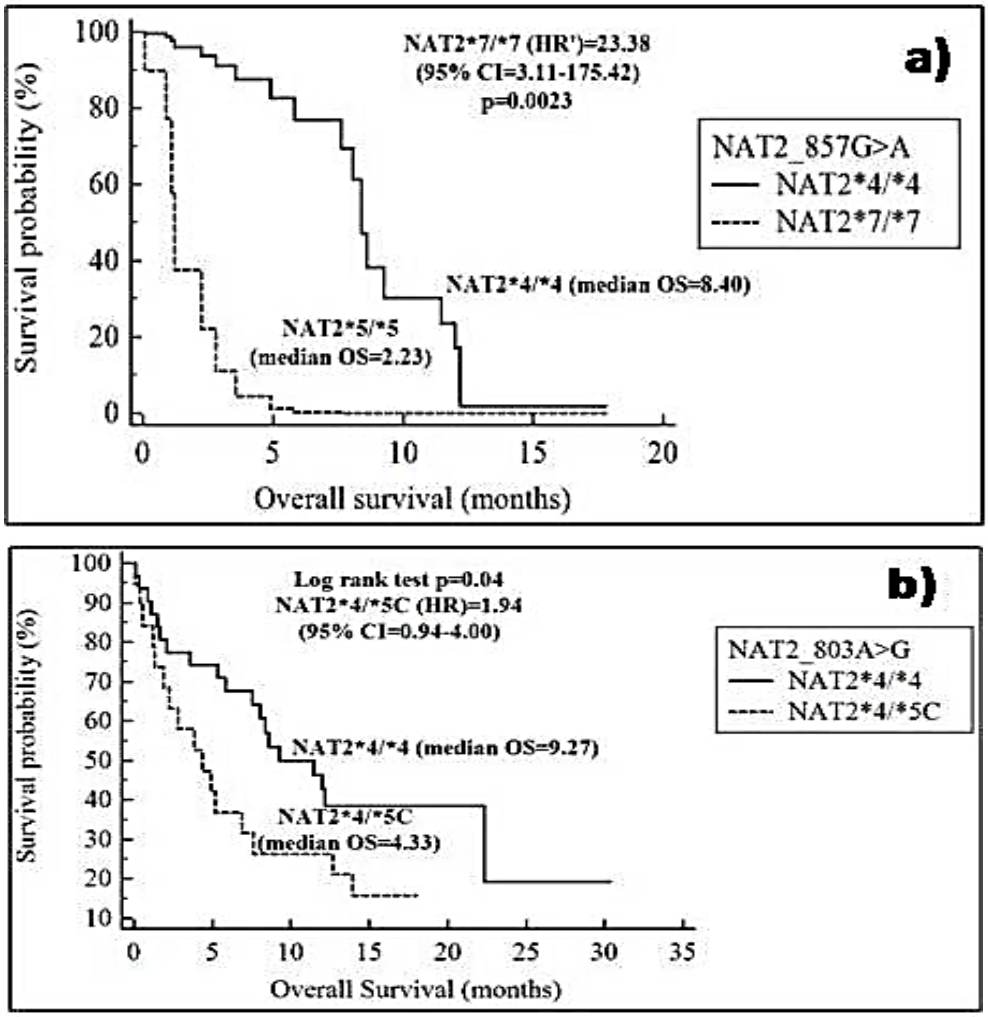


Figure 5.30: a) Cox hazards curve of mutant (NAT2*7/*7) genotype of 857G>A for subjects treated with cisplatin/carboplatin along with paclitaxel; **b)** Kaplan Meier curves illustrating the association between overall survival in patients with heterozygous (NAT2*4/*5C) genotype for patients treated with cisplatin/carboplatin along with paclitaxel.

Furthermore, among another set of lung cancer patients who were administered with cisplatin/carboplatin and irinotecan and carrying the mutant genotype (*NAT2 *6/*6*) for *590G>A (NAT2*6*) revealed a higher survival time (MST=11.30 *vs.* 9.13 months, Log-rank p=0.11) and a significantly lower death rate obtained using the Cox-regression model (HR'=0.15, 95% C.I.=0.04-0.64, **p=0.01;** Figure 5.31 a)) as compared to those patients carrying the wild type genotype (*NAT2 *4/*4*) (Table 5.57 b)). Furthermore, the patients administered with the same chemotherapeutic regimen (*i.e.*, cisplatin/carboplatin and irinotecan) but carrying the mutant genotype (*NAT2*4/*7*) for *857G>A (NAT2*7*) polymorphism showed a poor overall survival as compared to subjects with the wild type genotype (*NAT2*4/*4*) (MST=5.80 vs.10.80 months). Using the multivariate Cox-regression model after adjusting for various covariates, it was visible that the above patients had a higher death rate probability (HR'=5.14, 95% C.I.=1.32-19.99, p=**0.02;** Figure 5.31 b)). Taken together, the *NAT2 590G>A* (*NAT2*6*) appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy as the median survival time is better than the rest of the chemotherapeutic regimens wherein the poor prognosis was observed as compared to the wild type genotype for all the SNPs of *NAT2*.

Table 5.57 b): Association of *NAT2* genotype on survival in lung cancer cases based on the chemotherapeutic regimen (Irinotecan + cis/carboplatin)

Irinotecan								
Genotype rs1799930 *590G>A (NAT2*6)*	**CASES n(%), N=61**	**DEAD n(%), N=51**	**ALIVE n(%), N=10**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	21 (34.43)	18 (35.30)	3 (30.00)	9.13				
*NAT2 *4/*6*	29 (47.54)	27 (52.94)	2 (20.00)	6.73	0.76	1.09 (0.61-1.98)	1.11 (0.55-2.23)	0.77
*NAT2 *6/*6*	11 (18.03)	6 (11.76)	5 (50.00)	11.30	0.11	0.48 (0.21-1.08)	0.15 (0.04-0.64)	**0.01**
*NAT2 *4/*6 + *6/*6*	40 (65.57)	33 ()	7 (70.00)	8.27	0.66	0.88 (0.49-1.58)	0.83 (0.41-1.68)	0.61
Genotype rs1799931 *857G>A (NAT2*7)*	**CASES n(%), N=61**	**DEAD n(%), N=51**	**ALIVE n(%), N=10**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	24 (39.35)	21 (41.18)	3 (30.00)	10.80				
*NAT2 *4/*7*	29 (47.54)	24 (47.06)	5 (50.00)	6.07	0.32	1.34 (0.75-2.41)	1.45 (0.73-2.87)	0.29
*NAT2 *7/*7*	8 (13.11)	6 (11.76)	2 (20.00)	5.80	0.78	1.13 (0.44-2.91)	5.14 (1.32-19.99)	**0.02**
*NAT2 *4/*7 + *7/*7*	37 ()	30 (58.82)	7 (70.00)	6.07	0.37	1.29 (0.74-2.23)	1.50 (0.80-2.81)	0.21
Genotype rs1799929 *481C>T (NAT2*5B)*	**CASES n(%), N=61**	**DEAD n(%), N=51**	**ALIVE n(%), N=10**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	32 (52.46)	27 (52.94)	5 (50.00)	9.10				
*NAT2 *4/*5B*	25 (40.98)	20 (39.22)	5 (50.00)	9.80	0.78	0.92 (0.52-1.64)	0.70 (0.35-1.40)	0.31
*NAT2 *5B/*5B*	4 (6.56)	4 (7.84)	0 (0.00)	4.43	0.06	2.57 (0.54-12.13)	3.01 (0.65-13.94)	0.16
*NAT2 *4/*5B + *5B/*5B*	29 (47.54)	24 (47.06)	5 (50.00)	6.73	0.91	1.03 (0.59-1.79)	0.78 (0.40-1.51)	0.46
Genotype rs1208 *803A>G (NAT2*5C)*	**CASES n(%), N=61**	**DEAD n(%), N=51**	**ALIVE n(%), N=10**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	34 (55.74)	28 (54.90)	6 (60.00)	10.57				
*NAT2 *4/*5C*	19 (31.15)	16 (31.37)	3 (30.00)	6.07	0.23	1.45 (0.75-2.79)	1.73 (0.85-3.52)	0.13
*NAT2 *5C/*5C*	8 (13.11)	7 (13.73)	1 (10.00)	5.00	0.19	1.71 (0.63-4.62)	2.12 (0.73-6.19)	0.17
*NAT2 *4/*5C + *5C/*5C*	27 (65.76)	23 (45.10)	4 (40.00)	6.07	0.13	1.51 (0.85-2.68)	1.72 (0.91-3.26)	0.10

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

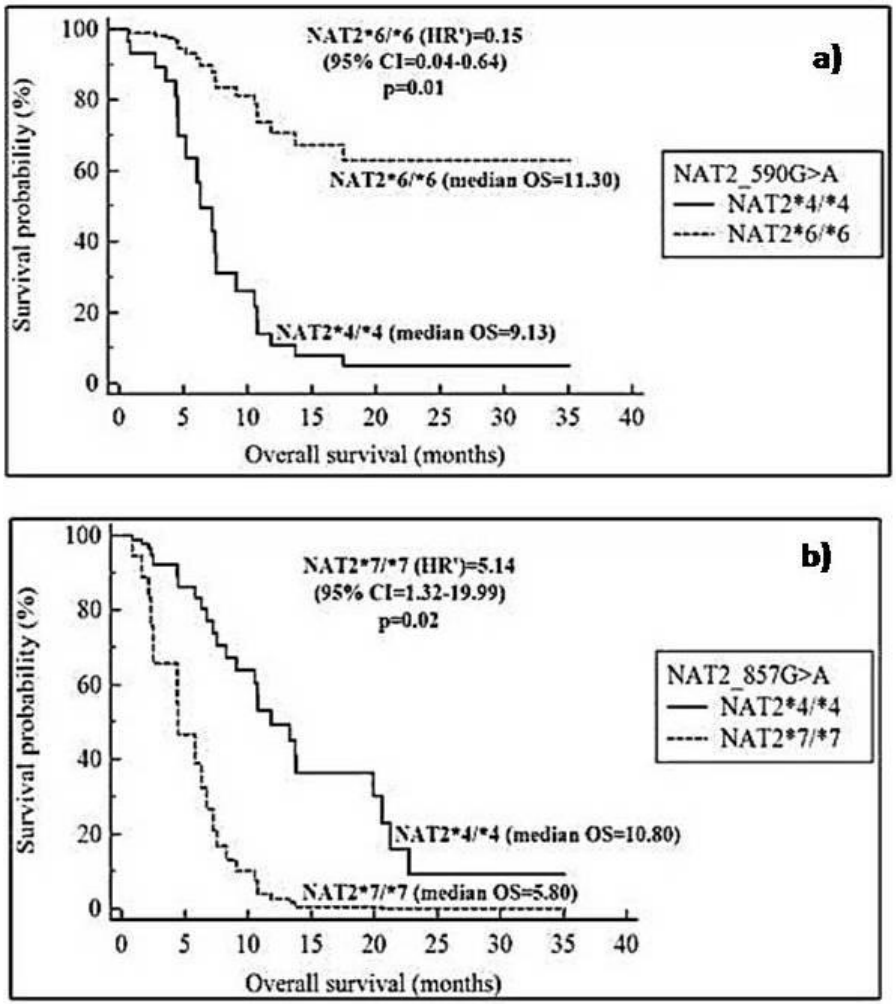


Figure 5.31: a) Cox proportional hazards curve illustrating mutant (NAT2*6/*6) genotype for patients treated with cisplatin/carboplatin alongwith irinotecan depicting 590G>Apolymorphism; **b)** Cox proportional hazards curve illustrating mutant (NAT2*7/*7) genotype of 857G>A polymorphism for patients treated with cisplatin/carboplatin alongwith irinotecan.

Further, the *NAT2* variants were also evaluated for survival association with other two chemotherapeutic regimens, *viz*. docetaxel + cis/carboplatin and pemetrexed + cis/ carboplatin. But, none of the genotypes of *NAT2* were found to have any profound effect on the prognosis of lung cancer patients after administration with docetaxel + cis/carboplatin or pemetrexed + cis/ carboplatin as demonstrated in Table 5.57 c) and Table 5.57 d), respectively.

Table 5.57 c): Association of *NAT2* genotype on survival in lung cancer cases based on the chemotherapeutic regimen (Docetaxel + cis/carboplatin)

Docetaxel								
Genotype rs1799930 *590G>A (NAT2*6)*	**CASES n(%), N=70**	**DEAD n(%), N=57**	**ALIVE n(%), N=13**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	32 (45.71)	24 (42.11)	8 (61.54)	9.83				
*NAT2 *4/*6*	33 (47.14)	29 (50.88)	4 (30.77)	8.63	0.22	1.39 (0.81-2.39)	1.34 (0.72-2.53)	0.36
*NAT2 *6/*6*	5 (7.14)	4 (7.01)	1 (7.69)	4.67	0.29	1.75 (0.47-6.52)	4.56 (0.82-25.15)	0.08
*NAT2 *4/*6 + *6/*6*	38 (54.29)	33 (57.89)	5 (38.46)	8.40	0.17	1.44 (0.85-2.41)	1.38 (0.74-2.57)	0.31
Genotype rs1799931 *857G>A (NAT2*7)*	**CASES n(%), N=70**	**DEAD n(%), N=57**	**ALIVE n(%), N=13**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	30 (42.86)	27 (47.37)	3 (23.08)	8.43				
*NAT2 *4/*7*	33 (47.14)	27 (47.37)	6 (46.15)	8.37	0.57	1.16 (0.68-1.98)	1.14 (0.59-2.23)	0.69
*NAT2 *7/*7*	7 (10.00)	3 (5.26)	4 (30.77)	-	0.037	0.31 (0.14-0.69)	0.41 (0.12-1.49)	0.18
*NAT2 *4/*7 + *7/*7*	40 (57.14)	30 (52.63)	10 (76.92)	8.63	0.66	0.89 (0.53-1.50)	1.01 (0.54-1.91)	0.97
Genotype rs1799929 *481C>T (NAT2*5B)*	**CASES n(%), N=70**	**DEAD n(%), N=57**	**ALIVE n(%), N=13**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	28 (40.00)	22 (38.60)	6 (46.15)	8.43				
*NAT2 *4/*5B*	31 (44.29)	27 (47.37)	4 (30.77)	8.63	0.45	1.24 (0.71-2.17)	1.44 (0.77-2.69)	0.25
*NAT2 *5B/*5B*	11 (15.71)	8 (14.03)	3 (23.08)	10.13	0.57	0.79 (0.37-1.71)	0.46 (0.16-1.29)	0.14
*NAT2 *4/*5B + *5B/*5B*	42 (60.00)	35 (61.40)	7 (53.85)	8.63	0.72	1.10 (0.65-1.87)	0.98 (0.54-1.78)	0.96
Genotype rs1208 *803A>G (NAT2*5C)*	**CASES n(%), N=70**	**DEAD n(%), N=57**	**ALIVE n(%), N=13**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	35 (50.00)	29 (50.88)	6 (46.15)	8.43				
*NAT2 *4/*5C*	24 (34.29)	20 (35.09)	4 (30.77)	8.63	0.87	1.05 (0.59-1.86)	1.29 (0.66-2.52)	0.46
*NAT2 *5C/*5C*	11 (15.71)	8 (14.03)	3 (23.08)	10.30	0.87	0.94 (0.43-2.02)	0.53 (0.19-1.49)	0.23
*NAT2 *4/*5C + *5C/*5C*	35 (50.00)	28 (49.12)	7 (53.85)	9.13	0.94	1.02 (0.61-1.71)	0.96 (0.51-1.82)	0.91

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

Table 5.57 d): Association of *NAT2* genotype on survival in lung cancer cases based on the chemotherapeutic regimen (Pemetrexed + cis/carboplatin)

Pemetrexed								
Genotype rs1799930 *590G>A (NAT2*6)*	**CASES n(%), N=108**	**DEAD n(%), N=79**	**ALIVE n(%), N=29**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	48 (44.44)	38 (48.10)	10 (34.48)	7.60				
*NAT2 *4/*6*	47 (43.52)	33 (41.77)	14 (48.28)	12.23	0.13	0.70 (0.44-1.12)	0.86 (0.52-1.42)	0.56
*NAT2 *6/*6*	13 (12.04)	8 (10.13)	5 (17.24)	7.57	0.76	0.89 (0.43-1.86)	0.94 (0.40-2.21)	0.88
*NAT2 *4/*6 + *6/*6*	60 (55.56)	41 (51.90)	19 (65.52)	12.07	0.17	0.74 (0.47-1.15)	0.88 (0.55-1.40)	0.59
Genotype rs1799931 *857G>A (NAT2*7)*	**CASES n(%), N=108**	**DEAD n(%), N=79**	**ALIVE n(%), N=29**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	34 (31.48)	27 (34.18)	7 (24.14)	9.43				
*NAT2 *4/*7*	62 (57.41)	43 (54.43)	19 (65.52)	11.63	0.87	0.96 (0.59-1.56)	1.06 (0.63-1.79)	0.82
*NAT2 *7/*7*	12 (11.11)	9 (11.39)	3 (10.34)	7.10	0.59	1.23 (0.55-2.72)	1.45 (0.60-3.53)	0.41
*NAT2 *4/*7 + *7/*7*	74 (68.52)	52 (65.82)	22 (75.86)	8.23	0.97	0.99 (0.62-1.58)	1.07 (0.65-1.77)	0.77
Genotype rs1799929 *481C>T (NAT2*5B)*	**CASES n(%), N=108**	**DEAD n(%), N=79**	**ALIVE n(%), N=29**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	47 (43.52)	38 (48.10)	9 (31.03)	7.90				
*NAT2 *4/*5B*	45 (41.67)	29 (36.71)	16 (55.17)	12.07	0.18	0.72 (0.45-1.17)	0.74 (0.44-1.25)	0.27
*NAT2 *5B/*5B*	16 (14.81)	12 (15.19)	4 (13.80)	6.97	0.84	1.07 (0.55-2.06)	1.35 (0.64-2.88)	0.43
*NAT2 *4/*5B + *5B/*5B*	61 (56.48)	41 (51.90)	20 (68.97)	10.30	0.32	0.80 (0.51-1.25)	0.78 (0.49-1.23)	0.28
Genotype rs1208 *803A>G (NAT2*5C)*	**CASES n(%), N=108**	**DEAD n(%), N=79**	**ALIVE n(%), N=29**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
*NAT2 *4/*4*	50 (46.30)	39 (49.37)	11 (37.93)	7.90				
*NAT2 *4/*5C*	42 (38.89)	28 (35.44)	14 (48.28)	9.77	0.39	0.81 (0.50-1.31)	0.81 (0.48-1.34)	0.41
*NAT2 *5C/*5C*	16 (14.81)	12 (15.19)	4 (13.79)	7.57	0.79	0.91 (0.49-1.72)	0.86 (0.43-1.71)	0.67
*NAT2 *4/*5C + *5C/*5C*	58 (53.70)	40 (50.63)	18 (62.07)	9.77	0.44	0.84 (0.54-1.31)	0.83 (0.52-1.32)	0.43

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.8.4.4. Association of *NAT2* polymorphism and chemotherapy response

Association between the four *NAT2* polymorphisms and the efficacy outcome of response rate and clinical benefits were assessed using univariate logistic regression. Patients were divided into groups based on treatment response as responders (PR+CR) or non-responders (PD+SD). The association of *NAT2* polymorphism with chemotherapy response was summarized in Table 5.58. No significant correlation was observed between the *NAT2* polymorphic variants and objective response. Therefore, it seems plausible to suggest that lung cancer patients with variant genotypes for both the *NAT2* polymorphisms and who underwent platinum-based chemotherapy were unlikely to experience higher sensitivity and better response rates. As a result, none of the *NAT2* genotypes could significantly predict chemotherapeutic response rates.

Table 5.58: Relationship of genotypic distribution with clinic-pathological parameter based on response to chemotherapy

Genotype rs1799930 ***590G>A (NAT2*6)***	**CR+PR (118), n (%)**	**SD+PD (148), n (%)**	**AOR (95% C.I.)** [a]	***p*** [b]	**Genotype rs1799929** ***481C>T (NAT2*5B)***	**CR+PR (118), n (%)**	**SD+PD (148), n (%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model					**Codominant model**				
*NAT2 *4/*4*	41 (34.75)	60 (40.54)	1.00 (Reference)		*NAT2 *4/*4*	54 (45.76)	63 (42.57)	1.00 (Reference)	
*NAT2 *4/*6*	62 (52.54)	75 (50.68)	1.22 (0.72-2.07)	0.47	*NAT2 *4/*5B*	49 (41.53)	65 (43.92)	0.88 (0.52-1.48)	0.63
*NAT2 *6/*6*	15 (12.71)	13 (8.78)	1.90 (0.79-4.58)	0.15	*NAT2 *5B/*5B*	15 (12.71)	20 (13.51)	0.83 (0.38-1.81)	0.65
Dominant model					**Dominant model**				
*NAT2 *4/*4*	41 (34.75)	60 (40.54)	1.00 (Reference)		*NAT2 *4/*4*	54 (45.76)	63 (42.57)	1.00 (Reference)	
*NAT2 *4/*6 + *6/*6*	77 (65.25)	88 (59.46)	1.31 (0.79-2.17)	0.30	*NAT2*4/*5B +*5B/*5B*	64 (54.24)	85 (57.43)	0.87 (0.53-1.42)	0.58
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*6*	103 (87.29)	135 (91.22)	1.00 (Reference)		*NAT2*4/*4 +*4/*5B*	103 (87.29)	128 (86.49)	1.00 (Reference)	
*NAT2 *6/*6*	15 (12.71)	13 (8.78)	1.61 (0.72-3.59)	0.24	*NAT2 *5B/*5B*	15 (12.71)	20 (13.51)	0.89 (0.43-1.84)	0.76
Genotype rs1799931 ***857G>A (NAT2*7)***	**CR+PR (118), n (%)**	**SD+PD (148), n (%)**	**AOR (95% C.I.)** [a]	***p*** [b]	**Genotype rs1208** ***803A>G (NAT2*5C)***	**CR+PR (118), n (%)**	**SD+PD (148), n (%)**	**AOR (95% C.I.)** [a]	***p*** [b]
Codominant model					**Codominant model**				
*NAT2 *4/*4*	45 (38.14)	55 (37.16)	1.00 (Reference)		*NAT2 *4/*4*	63 (53.39)	66 (44.60)	1.00 (Reference)	
*NAT2 *4/*7*	66 (55.93)	77 (52.03)	1.03 (0.61-1.73)	0.91	*NAT2 *4/*5C*	39 (33.05)	64 (43.24)	0.63 (0.37-1.08)	0.09
*NAT2 *7/*7*	7 (5.93)	16 (10.81)	0.52 (0.19-1.40)	0.20	*NAT2 *5C/*5C*	16 (13.56)	18 (12.16)	0.94 (0.43-2.02)	0.87
Dominant model					**Dominant model**				
*NAT2 *4/*4*	45 (38.14)	55 (37.16)	1.00 (Reference)		*NAT2 *4/*4*	63 (53.39)	66 (44.60)	1.00 (Reference)	
*NAT2 *4/*7 + *7/*7*	73 (61.86)	93 (62.84)	0.94 (0.57-1.56)	0.81	*NAT2 *4/*5C + *5C/*5C*	55 (46.61)	82 (55.40)	0.70 (0.42-1.14)	0.15
Recessive model					**Recessive model**				
*NAT2 *4/*4 + *4/*7*	111 (94.07)	132 (89.19)	1.00 (Reference)		*NAT2 *4/*4 + *4/*5C*	102 (86.44)	130 (87.84)	1.00 (Reference)	
*NAT2 *7/*7*	7 (5.93)	16 (10.81)	0.52 (0.20-1.31)	0.16	*NAT2 *5C/*5C*	16 (13.56)	18 (12.16)	1.13 (0.54-2.33)	0.75

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table

5.8.4.5. Association between *NAT2* polymorphisms and toxicity

The *NAT2* genotype was studied concerning toxicity after chemotherapeutic treatment, of which four polymorphisms were studied *viz*. *NAT2*5B* (rs1799929, *481C>T*), *NAT2*5C* (rs1208, *803A>G*), *NAT2*6* (rs1799930, *590G>A*), and *NAT2*7* (rs1799931, *857G>A*) (Table 5.59 a), b), c)). The results depict that the subjects harboring heterozygous (*NAT2*4/*6*) and combined genotypes (*NAT2*4/*4 + *4/*6*) for *NAT2*6* polymorphism was found to exhibit 2.4 folds (p=**0.005**), and two-folds (p=**0.01**) increased risk of hematological toxicity especially leukopenia (grade 1-3) when compared with subjects with no toxicity (grade 0) (Table 5.59 a)) (χ^2= 8.20, p=**0.02**). Further, the subjects carrying heterozygous genotype (*NAT*4/*5C*) for *NAT2*5C* polymorphism did not have any pronounced risk for leukopenia (AOR=0.23, 95% C.I.=0.06-0.85, p=**0.03**) (χ^2= 9.72, p=**0.008**). Further, on comparing hematological toxicity, especially anemia grade 1-4 *vs*. grade 0 (no toxicity), the subjects carrying heterozygous genotype for *NAT2*6* polymorphism exhibit three-fold (AOR=3.43, 95% C.I.=1.52-7.74, p=**0.003**) increased risk of developing anemia (χ^2= 12.35, p=**0.002**). In the case of *NAT2*7* polymorphism of *NAT2* gene, the heterozygous as well as the combined genotype was found to decrease the risk of developing hematological toxicity, basically intermediate/severe anemia (grade 2-4) *vs*. no/low anemia (grade 0-1) with borderline significance (Table 5.59 a)). Further on same lines the heterozygous (AOR=0.45, 95% C.I.=0.26-0.77, p=**0.004**), the variant genotype (AOR=0.39, 95% C.I.=0.17-0.89, p=**0.02**), as well as the combined genotype (AOR=0.45, 95% C.I.=0.27-0.74, p=**0.002**) for *NAT2*5C* polymorphism was found to have decreased risk for hematological toxicity particularly anemia (χ^2= 9.46, p=**0.009**).

The heterozygous genotype for *NAT2*7* polymorphism was found to pose an approximately two-fold risk for developing thrombocytopenia when the grade 1-3 thrombocytopenia was compared with patients without thrombocytopenia (grade 0) (AOR=1.96, 95% C.I.=1.02-3.79, p=**0.04**) (χ^2= 5.36, p=0.07). The neutropenia was found to be affected in the case of subjects carrying heterozygous genotype (*NAT2*4/*6*) for *NAT2*6* polymorphism wherein odds of 2.28 (95% C.I.=1.17-4.44, p=**0.02**) (χ^2= 10.84, p=**0.004**) was found as compared to wild type genotype when comparison was made between grade 1-4 and grade 0 neutropenia (Table 5.59 a)).

Table 5.59 a): Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NAT2* polymorphisms in multivariate analysis

Leukopenia													
Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	85(44.97)	24(28.57)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	84(44.44)	41(48.81)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	78(41.27)	50(59.52)	2.27(1.27-4.04)	**0.005**	2.38(1.30-4.35)	**0.005**	*NAT2*4/*5B*	77(40.74)	34(40.48)	0.90(0.52-1.57)	0.72	0.92(0.52-1.64)	0.78
*NAT2*6/*6*	26(13.76)	10(11.90)	1.36(0.58-3.21)	0.48	1.24(0.49-3.13)	0.65	*NAT2*5B/*5B*	28(14.81)	9(10.71)	0.66(0.28-1.52)	0.33	0.58(0.23-1.44)	0.24
*NAT2*4/*4 + *4/*6*	104(55.03)	60(71.43)	2.04(1.17-3.55)	**0.01**	2.07(1.17-3.68)	**0.01**	*NAT2*4/*5B + *5B/*5B*	105(55.55)	43(51.19)	0.84(0.50-1.40)	0.50	0.86(0.49-1.48)	0.58
Genotype rs1799931 *857G>A (NAT2*7)*	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	56(29.63)	34(40.48)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	96(50.79)	48(57.14)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	110(58.20)	43(51.19)	0.64(0.37-1.12)	0.12	0.76(0.42-1.36)	0.35	*NAT2*4/*5C*	72(38.10)	23(27.38)	0.64(0.36-1.14)	0.13	0.58(0.32-1.07)	0.08
*NAT2*7/*7*	23(12.17)	7(8.33)	0.50(0.19-1.29)	0.15	0.54(0.20-1.46)	0.22	*NAT2*5/*5C*	21(11.11)	13(15.48)	1.24(0.57-2.68)	0.59	1.10(0.49-2.49)	0.81
*NAT2*4/*7 + *7/*7*	133(70.37)	50(59.52)	0.62(0.36-1.06)	0.08	0.71(0.41-1.25)	0.24	*NAT2*4/*5C + *5C/*5C*	93(49.21)	36(42.86)	0.77(0.46-1.30)	0.33	0.69(0.40-1.19)	0.18
Genotype rs1799930 *590G>A*	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1799929 *481C>T***	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**

(NAT2*6)							***(NAT2*5B)***						
*NAT2*4/*4*	98(39.68)	11(42.31)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	112()45.34	13(50.00)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	114(46.15)	14(53.85)	1.09(0.47-2.52)	0.83	1.05(0.44-2.50)	0.91	*NAT2*4/*5B*	102(41.30)	9(34.62)	0.76(0.31-1.85)	0.55	0.79(0.31-2.00)	0.62
*NAT2*6/*6*	35(14.17)	1(3.85)	0.25(0.03-2.04)	0.20	0.17(0.02-1.66)	0.13	*NAT2*5/*5B*	33(13.36)	4(15.38)	1.04(0.32-3.42)	0.94	0.88(0.23-3.40)	0.85
*NAT2*4/*4 + *4/*6*	149(60.32)	15(57.69)	0.90(0.39-2.03)	0.79	0.85(0.36-2.01)	0.72	*NAT2*4/*5B + *5B/*5B*	135(54.66)	13(50.00)	0.83(0.37-1.86)	0.65	0.81(0.34-1.92)	0.64
Genotype rs1799931 *857G>A* *(NAT2*7)*	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1208 *803A>G* *(NAT2*5C)***	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	78(31.58)	12(46.15)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	128(51.82)	16(61.54)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	141(57.09)	12(46.15)	0.55(0.24-1.29)	0.17	0.76(0.31-1.90)	0.56	*NAT2*4/*5C*	92(37.25)	3(11.54)	0.26(0.07-0.92)	**0.04**	0.23(0.06-0.85)	**0.03**
*NAT2*7/*7*	28(11.34)	2(7.69)	0.46(0.10-2.20)	0.33	0.70(0.13-3.80)	0.68	*NAT2*5/*5C*	27(10.93)	7(26.92)	2.07(0.78-5.53)	0.14	1.79(0.61-5.24)	0.28
*NAT2*4/*7 + *7/*7*	169(68.42)	14(53.85)	0.54(0.24-1.22)	0.14	0.73(0.30-1.76)	0.49	*NAT2*4/*5C + *5C/*5C*	119(48.18)	10(38.46)	0.67(0.29-1.54)	0.35	0.56(0.23-1.34)	0.19
Anemia													
Genotype rs1799930 *590G>A* *(NAT2*6)*	**Grade of toxicity 0, n (%)**	**Grade of toxicity 1-4, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1799929 *481C>T* *(NAT2*5B)***	**Grade of toxicity 0, n (%)**	**Grade of toxicity 1-4, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	23(53.49)	86(37.39)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	19(44.19)	106(46.09)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	10(23.26)	118(51.30)	3.16(1.43-6.97)	**0.005**	3.43(1.52-7.74)	**0.003**	*NAT2*4/*5B*	19(44.19)	92(40.00)	0.87(0.43-1.74)	0.69	0.74(0.36-1.54)	0.43

*NAT2*6/*6*	10(23.26)	26(11.30)	0.69(0.29-1.65)	0.41	0.75(0.31-1.82)	0.53	*NAT2*5B/*5B*	5(11.63)	32(13.91)	1.15(0.40-3.32)	0.80	0.86(0.28-2.62)	0.79
*NAT2*4/*4 + *4/*6*	20(46.51)	144(62.61)	1.92(1.00-3.71)	0.05	2.07(1.06-4.05)	**0.03**	*NAT2*4/*5B + *5B/*5B*	24(55.81)	124(53.91)	0.93(0.48-1.78)	0.82	0.78(0.39-1.55)	0.48
Genotype rs1799931 *857G>A* *(NAT2*7)*	**Grade of toxicity 0, n (%)**	**Grade of toxicity 1-4, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1208 *803A>G* *(NAT2*5C)***	**Grade of toxicity 0, n (%)**	**Grade of toxicity 1-4, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	10(23.25)	80(34.78)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	19(44.19)	125(54.35)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	28(65.12)	125(54.35)	0.56(0.26-1.21)	0.14	0.52(0.23-1.17)	0.12	*NAT2 *4/*5C*	17(39.53)	78(33.91)	0.70(0.34-1.42)	0.32	0.60(0.29-1.25)	0.17
*NAT2*7/*7*	5(11.63)	25(10.87)	0.62(0.19-2.00)	0.43	0.60(0.18-1.98)	0.40	*NAT2 *5C/*5C*	7(16.28)	27(11.74)	0.59(0.22-1.53)	0.28	0.43(0.15-1.21)	0.11
*NAT2*4/*7 + *7/*7*	33(76.74)	150(65.22)	0.57(0.27-1.21)	0.14	0.52(0.24-1.14)	0.10	*NAT2*4/*5C + *5C/*5C*	24(55.81)	105(45.65)	0.66(0.34-1.28)	0.22	0.56(0.28-1.10)	0.09
Genotype rs1799930 *590G>A* *(NAT2*6)*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1799929 *481C>T* *(NAT2*5B)***	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	55(41.04)	54(38.85)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	62(46.27)	63(45.32)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	60(44.78)	68(48.92)	1.15(0.69-1.92)	0.58	1.17(0.69-1.98)	0.55	*NAT2*4/*5B*	55(41.04)	56(40.29)	1.00(0.60-1.67)	0.99	0.92(0.54-1.56)	0.75
*NAT2*6/*6*	19(14.18)	17(12.23)	0.91(0.43-1.94)	0.81	1.08(0.49-2.37)	0.84	*NAT2*5B/*5B*	17(12.69)	20(14.39)	1.16(0.55-2.41)	0.70	1.03(0.48-2.24)	0.93
*NAT2*4/*4*	79(58.96)	85(61.15)	1.09(0.67-1.78)	0.71	1.14(0.69-1.86)	0.61	*NAT2*4/*5B*	72(53.73)	76(54.68)	1.04(0.64-1.67)	0.87	0.96(0.58-1.57)	0.86

*+ *4/*6*							*+ *5B/*5B*						
Genotype rs1799931 *857G>A (NAT2*7)*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	38(28.36)	52(37.41)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	58(43.28)	86(61.87)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	81(60.45)	72(51.80)	0.65(0.38-1.10)	0.11	0.58(0.34-1.00)	**0.05**	*NAT2 *4/*5C*	56(41.79)	39(28.06)	0.47(0.28-0.79)	**0.005**	0.45(0.26-0.77)	**0.004**
*NAT2*7/*7*	15(11.19)	15(10.79)	0.73(0.32-1.67)	0.46	0.67(0.29-1.59)	0.37	*NAT2 *5C/*5C*	20(14.93)	14(10.07)	0.47(0.22-1.01)	0.05	0.39(0.17-0.89)	**0.02**
*NAT2*4/*7 + *7/*7*	96(71.64)	87(62.59)	0.66(0.40-1.10)	0.11	0.59(0.35-1.00)	**0.05**	*NAT2*4/*5C + *5C/*5C*	76(56.72)	53(38.13)	0.47(0.29-0.76)	**0.002**	0.45(0.27-0.74)	**0.002**
Thrombocytopenia													
Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	84(41.38)	24(34.78)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	97(47.78)	27(39.13)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	87(42.86)	41(59.42)	1.65(0.92-2.96)	0.09	1.77(0.97-3.24)	0.06	*NAT2*4/*5B*	77(37.93)	34(49.28)	1.59(0.88-2.85)	0.12	1.56(0.85-2.87)	0.15
*NAT2*6/*6*	32(15.76)	4(5.80)	0.44(0.14-1.36)	0.15	0.40(0.12-1.32)	0.13	*NAT2*5B/*5B*	29(14.29)	8(11.59)	0.99(0.41-2.42)	0.98	0.79(0.30-2.07)	0.64
*NAT2*4/*4 + *4/*6*	119(58.62)	45(65.22)	1.32(0.75-2.34)	0.33	1.39(0.77-2.50)	0.27	*NAT2*4/*5B + *5B/*5B*	106(52.22)	42(60.87)	1.42(0.82-2.48)	0.21	1.39(0.78-2.49)	0.26
Genotype rs1799931	**Grade of toxicity 0,**	**Grade of toxicity 1-3,**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1208**	**Grade of toxicity 0,**	**Grade of toxicity 1-**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**

857G>A (NAT2*7)	**n(%)**	**n(%)**					***803A>G (NAT2*5C)***	**n(%)**	**3, n(%)**				
*NAT2*4/*4*	72(35.47)	17(24.64)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	110(54.19)	34(49.28)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	106(52.22)	47(68.12)	1.88(1.00-3.53)	0.05	1.96(1.02-3.79)	**0.04**	*NAT2 *4/*5C*	69(33.99)	25(36.23)	1.17(0.64-2.13)	0.60	1.23(0.67-2.28)	0.50
*NAT2*7/*7*	25(12.32)	5(7.25)	0.85(0.28-2.53)	0.77	0.97(0.31-3.01)	0.96	*NAT2 *5C/*5C*	24(11.82)	10(14.49)	1.35(0.59-3.10)	0.48	1.39(0.57-3.42)	0.47
*NAT2*4/*7 + *7/*7*	131(64.53)	52(75.36)	1.68(0.90-3.12)	0.10	1.73(0.91-3.29)	0.09	*NAT2*4/*5C + *5C/*5C*	93(45.81)	35(50.72)	1.22(0.70-2.10)	0.48	1.28(0.73-2.25)	0.39
Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
*NAT2*4/*4*	104(39.25)	4(57.14)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	123(46.42)	1(14.29)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	125(47.17)	3(42.86)	0.62(0.14-2.85)	0.54	0.69(0.15-3.25)	0.64	*NAT2*4/*5B*	107(40.38)	4(57.14)	4.60(0.51-41.77)	0.17	4.14(0.44-38.42)	0.21
*NAT2*6/*6*	36(13.58)	0(0.00)	0.32(0.02-6.05)	0.45	-	-	*NAT2*5B/*5B*	35(13.21)	2(28.57)	7.03(0.62-79.81)	0.12	8.92(0.61-130.27)	0.11
*NAT2*4/*4 + *4/*6*	161(60.75)	3(42.86)	0.48(0.11-2.21)	0.35	0.51(0.11-2.42)	0.40	*NAT2*4/*5B + *5B/*5B*	142(53.58)	6(85.71)	5.20(0.62-43.77)	0.13	4.83(0.56-41.43)	0.15
Genotype rs1799931 *857G>A (NAT2*7)*	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**

*NAT2*4/*4*	87(32.83)	2(28.57)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	140(52.83)	4(57.14)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	148(55.85)	5(71.43)	1.47(0.28-7.74)	0.65	1.28(0.24-6.91)	0.77	*NAT2 *4/*5C*	92(34.72)	2(28.57)	0.76(0.14-4.24)	0.75	0.78(0.14-4.40)	0.78
*NAT2*7/*7*	30(11.32)	0(0.00)	0.57(0.03-12.29)	0.72	-	-	*NAT2 *5C/*5C*	33(12.45)	1(14.28)	1.06(0.11-9.80)	0.96	0.96(0.10-9.36)	0.97
*NAT2*4/*7 + *7/*7*	178(67.17)	5(71.43)	1.22(0.23-6.42)	0.81	1.08(0.20-5.79)	0.93	*NAT2*4/*5C + *5C/*5C*	125(47.17)	3(42.86)	0.84(0.18-3.83)	0.82	0.86(0.18-3.96)	0.84
Neutropenia													
Genotype rs1799930 *590G>A* *(NAT2*6)*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T* *(NAT2*5B)***	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	86(42.16)	17(29.31)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	90(44.12)	30(51.72)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	86(42.16)	38(65.52)	2.23(1.17-4.26)	**0.01**	2.28(1.17-4.44)	**0.02**	*NAT2*4/*5B*	83(40.69)	23(39.66)	0.83(0.45-1.54)	0.56	0.84(0.44-1.61)	0.61
*NAT2*6/*6*	32(15.69)	3(5.17)	0.47(0.13-1.73)	0.26	0.39(0.09-1.60)	0.19	*NAT2*5B/*5B*	31(15.20)	5(8.62)	0.48(0.17-1.36)	0.17	0.39(0.13-1.18)	0.09
*NAT2*4/*4 + *4/*6*	118(57.84)	41(70.69)	1.76(0.94-3.30)	0.08	1.75(0.91-3.36)	0.09	*NAT2*4/*5B + *5B/*5B*	114(55.88)	28(48.28)	0.74(0.41-1.32)	0.30	0.71(0.38-1.32)	0.28
Genotype rs1799931 *857G>A* *(NAT2*7)*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208 *803A>G* *(NAT2*5C)***	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	69(33.82)	21(36.21)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	106(51.96)	34(58.62)	1.00(Reference)		1.00(Reference)	

*NAT2*4/*7*	114(55.88)	32(55.17)	0.92(0.49-1.72)	0.80	1.19(0.61-2.34)	0.61	*NAT2 *4/*5C*	73(35.78)	15(25.86)	0.64(0.32-1.26)	0.20	0.59(0.29-1.20)	0.14
*NAT2*7/*7*	21(10.29)	5(8.62)	0.78(0.26-2.33)	0.66	0.94(0.29-2.97)	0.91	*NAT2 *5C/ *5C*	25(12.25)	9(15.52)	1.12(0.48-2.64)	0.79	0.95(0.39-2.33)	0.92
*NAT2*4/*7 + *7/*7*	135(66.18)	37(63.78)	0.90(0.49-1.65)	0.74	1.12(0.58-2.14)	0.73	*NAT2*4/*5C + *5C/*5C*	98(48.04)	24(41.38)	0.76(0.42-1.38)	0.37	0.67(0.36-1.25)	0.21
Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	90(40.00)	13(35.14)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	101(44.89)	19(51.35)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	102(45.33)	22(59.46)	1.49(0.71-3.13)	0.29	1.47(0.68-3.16)	0.32	*NAT2*4/*5B*	93(41.33)	13(35.14)	0.74(0.35-1.59)	0.44	0.74(0.34-1.64)	0.46
*NAT2*6/*6*	33(14.67)	2(5.40)	0.42(0.09-1.96)	0.27	0.35(0.06-1.87)	0.22	*NAT2*5B/*5 B*	31(13.78)	5(13.51)	0.86(0.29-2.48)	0.78	0.71(0.22-2.23)	0.55
*NAT2*4/*4 + *4/*6*	135(60.00)	24(64.86)	1.23(0.59-2.54)	0.57	1.18(0.56-2.50)	0.66	*NAT2*4/*5B + *5B/*5B*	124(55.11)	18(48.65)	0.77(0.38-1.55)	0.46	0.74(0.35-1.53)	0.41
Genotype rs1799931 *857G>A (NAT2*7)*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	76(33.78)	14(37.84)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	117(52.00)	23(62.16)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	127(56.44)	19(51.35)	0.81(0.38-1.71)	0.58	1.07(0.48-2.39)	0.87	*NAT2 *4/*5C*	80(35.56)	8(21.62)	0.51(0.22-1.19)	0.12	0.46(0.19-1.10)	0.08
*NAT2*7/*7*	22(9.78)	4(10.81)	0.99(0.29-3.30)	0.98	1.26(0.35-4.55)	0.73	*NAT2 *5C/*	28(12.44)	6(16.22)	1.09(0.40-2.93)	0.86	1.01(0.35-2.84)	0.99

							5C					
*NAT2*4/*7 + *7/*7*	149(66.22)	23(62.16)	0.84(0.41-1.72)	0.63	1.05(0.49-2.25)	0.90	*NAT2*4/*5C + *5C/*5C*	108(48.00)	14(37.84)	0.66(0.32-1.35)	0.25	0.57(0.27-1.21)	0.15

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR**[2]Adjusted Odds ratio data were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

Nephrotoxicity was associated with those lung cancer patients carrying the variant genotype for *NAT2*5C* polymorphism (Table 5.59 b)). It was observed that such patients exhibited 3.20 folds increased risk for nephrotoxicity, especially intermediate/ severe nephrotoxicity (grade 2-4) compared to no/low nephrotoxicity (grade 0-1) (χ^2= 8.28, p=**0.01**), as shown in Table 5.59 b). Further, the heterozygous genotype for *NAT2*7* polymorphism decreased risk towards anorexia and alopecia with odds of 0.10 (95% C.I.=0.01-0.96, p=0.04) 0.37 (95% C.I.=0.14-0.97, p=**0.04**), respectively (Table 5.59 c)) (χ^2= 2.90, p=0.23).

Table 5.59 b): Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NAT2* polymorphisms in multivariate analysis

Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	27(35.53)	62(39.24)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	39(51.32)	66(41.77)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	32(42.10)	79(50.00)	1.07(0.58-1.98)	0.82	1.27(0.62-2.59)	0.51	*NAT2*4/*5B*	31(40.79)	66(41.77)	1.26(0.70-2.25)	0.44	0.92(0.46-1.81)	0.80
*NAT2*6/*6*	17(22.37)	17(10.76)	0.43(0.19-0.98)	**0.04**	0.49(0.19-1.27)	0.14	*NAT2*5/*5B*	6(7.89)	26(16.46)	2.56(0.97-6.77)	0.06	1.42(0.46-4.41)	0.54
*NAT2*4/*4 + *4/*6*	49(64.47)	96(60.76)	0.85(0.48-1.50)	0.58	0.98(0.51-1.89)	0.95	*NAT2*4/*5B+ *5B/*5B*	37(48.68)	92(58.23)	1.47(0.85-2.55)	0.17	1.03(0.54-1.95)	0.93
Genotype rs1799931 *857G>A (NAT2*7)*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	21(127.63)	54(34.18)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	42(55.26)	83(52.53)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	43(56.58)	90(56.96)	0.81(0.44-1.51)	0.52	0.72(0.35-1.47)	0.36	*NAT2 *4/*5C*	29(38.16)	53(33.54)	0.92(0.51-1.66)	0.79	0.78(0.40-1.52)	0.46
*NAT2*7/*7*	12(15.79)	14(8.86)	0.33(0.13-0.84)	**0.02**	0.42(0.15-1.22)	0.11	*NAT2*5C/*5C*	5(6.58)	22(13.92)	2.23(0.79-6.30)	0.13	1.67(0.47-5.89)	0.43
*NAT2*4/*7 + *7/*7*	55(72.37)	104(65.82)	0.73(0.40-1.34)	0.31	0.64(0.32-1.29)	0.22	*NAT2*4/*5C + *5C/*5C*	34(44.74)	75(47.47)	1.12(0.64-1.93)	0.69	0.91(0.48-1.72)	0.77
Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**

*NAT2*4/*4*	65(35.91)	24(45.28)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	84(46.41)	21(39.62)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	89(49.17)	22(41.51)	0.67(0.34-1.30)	0.23	0.62(0.29-1.35)	0.23	*NAT2*4/*5B*	76(41.99)	21(39.62)	1.11(0.56-2.18)	0.77	0.67(0.30-1.49)	0.32
*NAT2*6/*6*	27(14.92)	7(13.21)	0.70(0.27-1.82)	0.47	1.38(0.45-4.26)	0.58	*NAT2*5B/*5B*	21(11.60)	11(20.75)	2.09(0.87-5.01)	0.10	1.16(0.41-3.27)	0.77
*NAT2*4/*4 + *4/*6*	116(64.09)	29(54.72)	0.68(0.36-1.26)	0.22	0.75(0.36-1.54)	0.43	*NAT2*4/*5B+ *5B/*5B*	97(53.59)	32(60.38)	1.32(0.71-2.46)	0.38	0.73(0.34-1.56)	0.41
Genotype rs1799931 *857G>A* *(NAT2*7)*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208 *803A>G* *(NAT2*5C)***	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	58(32.04)	17(32.08)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	100(55.25)	25(47.17)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	102(56.35)	31(58.49)	1.04(0.53-2.03)	0.92	0.83(0.37-1.86)	0.64	*NAT2 *4/*5C*	66(36.46)	16(30.19)	0.97(0.48-1.95)	0.93	0.89(0.40-1.99)	0.78
*NAT2*7/*7*	21(11.60)	5(9.43)	0.81(0.27-2.48)	0.71	0.99(0.27-3.57)	0.99	*NAT2*5C/*5C*	15(8.29)	12(22.64)	3.20(1.33-7.69)	**0.01**	2.42(0.84-6.99)	0.10
*NAT2*4/*7 + *7/*7*	123(67.96)	36(67.92)	0.99(0.52-1.92)	0.99	0.83(0.37-1.83)	0.64	*NAT2*4/*5C + *5C/*5C*	81(44.75)	28(52.83)	1.38(0.75-2.55)	0.30	1.19(0.58-2.43)	0.63

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratio data were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

Table 5.55 c): Gastrointestinal **t**oxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4) associated with *NAT2* polymorphisms in multivariate analysis

Diarrohea													
Genotype rs1799930 *590G>A* *(NAT2*6)*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T* *(NAT2*5B)***	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**

*NAT2*4/*4*	16(37.21)	22(33.85)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	23(53.49)	30(46.15)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	23(53.49)	35(53.85)	1.11(0.48-2.54)	0.81	1.10(0.47-2.58)	0.83	*NAT2*4/*5B*	14(32.56)	28(43.08)	1.53(0.66-3.55)	0.32	1.66(0.69-3.96)	0.25
*NAT2*6/*6*	4(9.30)	8(12.30)	1.45(0.37-5.68)	0.59	1.43(0.33-6.07)	0.63	*NAT2*5/*5B*	6(13.95)	7(10.77)	0.89(0.26-3.02)	0.86	0.96(0.27-3.40)	0.95
*NAT2*4/*4 + *4/*6*	27(62.79)	43(66.15)	1.16(0.52-2.59)	0.72	1.16(0.51-2.62)	0.72	*NAT2*4/*5B+ *5B/*5B*	20(46.51)	35(53.85)	1.34(0.62-2.90)	0.45	1.43(0.65-3.16)	0.38
Genotype rs1799931 *857G>A* *(NAT2*7)*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208 *803A>G* *(NAT2*5C)***	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	17(39.53)	25(41.67)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	23(53.49)	34(52.31)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	22(51.16)	35(53.85)	1.08(0.48-2.44)	0.85	1.04(0.45-2.37)	0.93	*NAT2 *4/*5C*	15(34.88)	25(38.46)	1.13(0.49-2.59)	0.78	1.26(0.53-2.97)	0.60
*NAT2*7/*7*	4(9.30)	5(7.69)	0.85(0.20-3.63)	0.83	0.91(0.20-4.41)	0.91	*NAT2*5/*5C*	5(11.63)	6(9.23)	0.81(0.22-2.98)	0.75	0.96(0.24-3.80)	0.95
*NAT2*4/*7 + *7/*7*	26(60.46)	40(61.53)	1.05(0.47-2.30)	0.91	0.99(0.44-2.22)	0.99	*NAT2*4/*5C + *5C/*5C*	20(46.51)	31(47.69)	1.05(0.48-2.27)	0.90	1.16(0.52-2.60)	0.71
Genotype rs1799930 *590G>A* *(NAT2*6)*	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T* *(NAT2*5B)***	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	29(37.66)	9(29.03)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	40(51.95)	13(41.94)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	41(53.25)	17(54.84)	1.34(0.52-3.41)	0.54	1.49(0.56-3.96)	0.42	*NAT2*4/*5B*	28(36.36)	14(45.16)	1.53(0.63-3.77)	0.35	1.48(0.59-3.71)	0.40
*NAT2*6/*6*	7(9.09)	5(16.13)	2.30(0.58-9.05)	0.23	2.72(0.55-13.41)	0.22	*NAT2*5/*5B*	9(11.69)	4(12.90)	1.37(0.36-5.19)	0.64	1.48(0.37-5.89)	0.58
*NAT2*4/*4*	48(62.34)	22(70.97)	1.48(0.60-3.64)	0.40	1.63(0.65-4.12)	0.30	*NAT2*4/*5B+*	37(48.05)	18(58.06)	1.50(0.64-3.47)	0.35	1.43(0.60-3.37)	0.41

*+ *4/*6*							**5B/*5B*						
Genotype rs1799931 *857G>A (NAT2*7)*	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	30(38.96)	12(38.71)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	42(54.54)	15(48.39)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	42(54.54)	15(48.39)	0.89(0.36-2.18)	0.80	0.85(0.34-2.13)	0.74	*NAT2 *4/*5C*	25(32.47)	15(48.39)	1.68(0.70-4.01)	0.24	1.64(0.67-3.99)	0.28
*NAT2*7/*7*	5(6.49)	4(12.90)	2.00(0.46-8.74)	0.36	2.13(0.46-9.99)	0.33	*NAT2 *5C/*5C*	10(12.99)	1(3.22)	0.28(0.03-2.38)	0.24	0.29(0.03-2.84)	0.29
*NAT2*4/*7 + *7/*7*	47(61.04)	19(61.29)	1.01(0.43-2.38)	0.98	0.96(0.40-2.30)	0.93	*NAT2*4/*5C + *5C/*5C*	35(45.45)	16(51.61)	1.28(0.55-2.95)	0.56	1.26(0.53-2.96)	0.60
Constipation													
Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	8(25.81)	14(34.15)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	15(48.39)	22(53.66)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	18(58.06)	21(51.22)	0.67(0.23-1.95)	0.46	0.62(0.20-1.93)	0.41	*NAT2*4/*5B*	15(48.39)	12(29.27)	0.54(0.20-1.49)	0.24	0.40(0.13-1.23)	0.11
*NAT2*6/*6*	5(16.13)	6(14.63)	0.68(0.16-2.98)	0.61	0.46(0.08-2.74)	0.39	*NAT2*5/*5B*	1(3.22)	7(17.07)	4.77(0.53-42.89)	0.16	6.47(0.68-61.68)	0.10
*NAT2*4/*4 + *4/*6*	23(74.19)	27(65.85)	0.67(0.24-1.88)	0.45	0.60(0.20-1.77)	0.35	*NAT2*4/*5B+ *5B/*5B*	16(51.61)	19(46.34)	0.81(0.32-2.06)	0.66	0.66(0.24-1.81)	0.42
Genotype rs1799931	**Grade of toxicity 1,**	**Grade of toxicity 2-**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208**	**Grade of toxicity 1,**	**Grade of toxicity 2-**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**

857G>A (NAT2*7)	**n(%)**	**3, n(%)**					***803A>G (NAT2*5C)***	**n(%)**	**3, n(%)**				
*NAT2*4/*4*	13(41.94)	16(39.02)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	16(51.61)	24(58.54)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	18(58.06)	22(53.66)	0.99(0.38-2.59)	0.99	0.92(0.34-2.51)	0.88	*NAT2 *4/*5C*	9(29.03)	8(19.51)	0.59(0.19-1.86)	0.37	0.47(0.13-1.61)	0.23
*NAT2*7/*7*	0(0.00)	3(7.32)	5.73(0.27-120.85)	0.26	-	-	*NAT2 *5C/*5C*	6(19.35)	9(21.95)	1.00(0.30-3.36)	1.00	0.74(0.20-2.72)	0.66
*NAT2*4/*7 + *7/*7*	18(58.06)	25(60.98)	1.13(0.44-2.92)	0.80	1.06(0.40-2.84)	0.90	*NAT2*4/*5C + *5C/*5C*	15(48.39)	17(41.46)	0.76(0.29-1.93)	0.56	0.61(0.23-1.66)	0.33

Anorexia

Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	2(20.00)	34(36.17)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	5(50.00)	44(46.81)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	6(60.00)	49(52.13)	0.48(0.09-2.52)	0.39	0.48(0.09-2.66)	0.40	*NAT2*4/*5B*	5(50.00)	37(39.36)	0.84(0.22-3.13)	0.79	0.68(0.17-2.77)	0.59
*NAT2*6/*6*	2(20.00)	11(11.70)	0.32(0.04-2.57)	0.29	0.82(0.08-8.53)	0.87	*NAT2*5B/*5B*	0(0.00)	13(13.83)	3.34(0.17-64.30)	0.42	-	-
*NAT2*4/*4 + *4/*6*	8(80.00)	60(63.83)	0.44(0.09-2.20)	0.32	0.55(0.10-2.90)	0.48	*NAT2*4/*5B+ *5B/*5B*	5(50.00)	50(53.19)	1.14(0.31-4.19)	0.85	0.92(0.23-3.68)	0.90
Genotype rs1799931 *857G>A (NAT2*7)*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**

*NAT2*4/*4*	1(10.00)	33(35.11)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	6(60.00)	53(56.38)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	9(90.00)	53(56.38)	0.18(0.02-1.47)	0.11	0.10(0.01-0.96)	**0.04**	*NAT2 *4/*5C*	2(20.00)	32(34.04)	1.81(0.34-9.52)	0.48	1.19(0.21-6.80)	0.85
*NAT2*7/*7*	0(0.00)	8(8.51)	0.76(0.03-20.39)	0.87	-	-	*NAT2 *5C/*5C*	2(20.00)	9(9.57)	0.51(0.09-2.93)	0.45	0.12(0.01-1.25)	0.07
*NAT2*4/*7 + *7/*7*	9(90.00)	61(64.89)	0.20(0.02-1.69)	0.14	0.11(0.01-1.03)	0.05	*NAT2*4/*5C + *5C/*5C*	4(40.00)	41(43.62)	1.16(0.31-4.38)	0.83	0.77(0.18-3.19)	0.71
Alopecia													
Genotype rs1799930 *590G>A (NAT2*6)*	**Grade of toxicity (No), n(%)**	**Grade of toxicity (Yes), n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1799929 *481C>T (NAT2*5B)***	**Grade of toxicity (No), n(%)**	**Grade of toxicity (Yes), n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	14(40.00)	26(33.33)	1.00(Reference)		1.00(Reference)		*NAT2*4/*4*	18(51.43)	39(50.00)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*6*	19(54.29)	46(58.97)	1.30(0.56-3.02)	0.54	1.42(0.60-3.37)	0.43	*NAT2*4/*5B*	12(34.28)	31(39.74)	1.19(0.50-2.84)	0.69	1.17(0.48-2.86)	0.72
*NAT2*6/*6*	2(5.71)	6(7.69)	1.62(0.29-9.09)	0.59	1.90(0.27-13.14)	0.51	*NAT2*5/*5B*	5(14.29)	8(10.26)	0.74(0.21-2.57)	0.63	0.76(0.21-2.73)	0.68
*NAT2*4/*4 + *4/*6*	21(60.00)	52(66.67)	1.33(0.58-3.04)	0.49	1.45(0.62-3.37)	0.39	*NAT2*4/*5B+ *5B/*5B*	17(48.57)	39(50.00)	1.06(0.48-2.35)	0.89	1.02(0.45-2.32)	0.95
Genotype rs1799931 *857G>A (NAT2*7)*	**Grade of toxicity (No), n(%)**	**Grade of toxicity (Yes), n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**	**Genotype rs1208 *803A>G (NAT2*5C)***	**Grade of toxicity (No), n(%)**	**Grade of toxicity (Yes), n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
*NAT2*4/*4*	9(25.71)	34(43.59)	1.00(Reference)		1.00(Reference)		*NAT2 *4/*4*	16(45.71)	45(57.69)	1.00(Reference)		1.00(Reference)	
*NAT2*4/*7*	23(65.71)	38(48.72)	0.44(0.18-1.07)	0.07	0.37(0.14-0.97)	**0.04**	*NAT2 *4/*5C*	14(40.00)	24(30.77)	0.61(0.25-1.46)	0.26	0.64(0.26-1.56)	0.32

*NAT2*7/*7*	3(8.57)	6(7.69)	0.53(0.11-2.54)	0.43	0.54(0.10-2.82)	0.47	*NAT2*5/*5C*	5(14.29)	9(11.54)	0.64(0.19-2.20)	0.48	0.73(0.20-2.70)	0.64
*NAT2*4/*7 + *7/*7*	26(74.29)	44(56.41)	0.45(0.18-1.08)	0.07	0.41(0.16-1.03)	0.06	*NAT2*4/*5C + *5C/*5C*	19(54.29)	33(42.31)	0.62(0.28-1.38)	0.24	0.63(0.27-1.43)	0.26

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR**[2]Adjusted Odds ratio data were calculated by unconditional logistic regression and adjusted for age, gender, performance status, and type of treatment regimen. A p-value < 0.05 was considered statistically significant

5.9. Statistical analysis of polymorphic sites of *MTHFR* gene

5.9.1. Patients characteristics and clinical predictors

Demographic characteristics of the study groups included age, gender, smoking status, pack-years, TNM staging, and other clinical parameters are depicted in Table 5.60. The study conducted comprised 253 adenocarcinoma cases. The adenocarcinoma patients were only studied because our main goal was to access association of *MTHFR* polymorphism towards lung cancer outcome. The mean age of the subjects was 58.17 ± 10.69 (range 28–85). There were 173 (68.38%) males and 80 (31.62%) females in the case group. In the present study, 58.50% of the cases were smokers, and 41.50% were non-smokers. Among the smokers, the mean pack-year was 27.47±26.55. TNM stage data were as follows: [stage I: 2 (0.79%), II: 2 (0.79%), III: 74 (29.25%), IV: 155 (61.26%)]. Tumour size T1 and T2 had low frequencies of 7.11% and 12.65%, whereas T3 and T4 had 21.74 and 46.64%, respectively. Among all the cases for which TNM data were available, 30.83% had no metastasis involvement (M0), whereas 65.22% had distant metastasis (M1). For lymph-node involvement, N0 had 14.62%, whereas N1, N2, N3, and N4 had frequencies of 7.11, 41.11, 32.01, and 0.40%, respectively. Furthermore, Table 5.60 also summarizes the clinical parameters of the subjects (cases) considered for survival analysis. Therefore, patients have also differentiated accordingly to KPS and ECOG, representing the subjects' performance status. In our study, there were 73 subjects (28.85%) with KPS scores of 100-90, 121 (47.83%) with KPS scores of 80-70, and 53 cases (20.95%) who had a KPS score of less than 60. Based upon the ECOG score, there were 55.73% (141) subjects with ECOG scores between 0-1, 29.64% (75) were ECOG 2, and 12.25% (31) had an ECOG score of 3-4. Likewise, as shown in Table 5.60, out of 186 subjects who had undergone chemotherapy, most were administered with pemetrexed + cis/carboplatin (144), as subjects diagnosed with adenocarcinoma were mostly administered with pemetrexed.

5.9.2. Genotype and genetic equilibrium test for *MTHFR* variants

The two polymorphic variants of the *MTHFR* gene were studied in this study *viz. 677C>T* and *1298A>C* and the genotypic frequencies have been tabulated in Table 5.61. In the case of *MTHFR 677C>T* polymorphism, 200 (79.05%) patients were homozygous for the wild (*CC*) genotype, whereas 49 (19.37%) were heterozygous (*CT*) carriers and 4 (1.58%) were variant homozygotes (*TT*) (Table 5.61).

Table 5.60: Patient characteristics

Variable	**MTHFR SNPs**			
	Total (N)	**Cases, n (%)**	**Cases, n (%)**	**Total (N)**	**Cases, n (%)**
Age (years)	253		**Lymph Node**	253	
Mean ± SD		58.17 ± 10.69	Nx		1 (0.40)
Range		28-85	N0		37 (14.62)
Gender	253		N1		18 (7.11)
Male		173 (68.38)	N2		104 (41.11)
Female		80 (31.62)	N3		81 (32.01)
Smoking status	253		N4		1 (0.40)
Smokers		148 (58.50)	Unknown		11 (4.35)
Non-smokers		105 (41.50)	**Performance status**		
Pack years	148		KPS (below 60)	253	53 (20.95)
Mean ± SD		27.47 ± 26.55	KPS (70-80)		121 (47.83)
Overall survival	253		KPS (90-100)		73 (28.85)
Dead		172 (67.98)	Unknown		6 (2.37)
Alive		81 (32.02)	ECOG (0-1)	253	141 (55.73)
TNM staging	253		ECOG (2)		75 (29.64)
I		2 (0.79)	ECOG (3-4)		31 (12.25)
II		2 (0.79)	Unknown		6 (2.37)
III		74 (29.25)	**Chemotherapy regimen**	186	
IV		155 (61.26)	Pemetrexed cis/carboplatin		144 (77.42)
Unclassified		20 (7.91)	Irinotecan cis/carboplatin		4 (2.15)
Tumuor Size	253		Docetaxel cis/carboplatin		5 (2.69)
Tx		19 (7.51)	Paclitaxel cis/carboplatin		12 (6.45)
T1		18 (7.11)	Gemcitabine cis/carboplatin		2 (1.08)
T2		32 (12.65)	Gefitinib		15 (8.06)
T3		55 (21.74)	Etoposide		1 (0.54)
T4		118 (46.64)	Others		3 (1.61)
Unknown		11 (4.35)			
Metastasis	253				
M0		78 (30.83)			
M1		165 (65.22)			
Unknown		10 (3.95)			

Abbreviations: SD = Standard Deviation, n = total number of lung cancer cases or control subjects, a) *p*-values were derived from Pearson Chi – square test except age and pack-years; Student t-test was used for age and pack-years. All *p*-values are two – sided. $p < 0.05$ was considered statistically significant

Whereas for the *MTHFR 1298A>C* polymorphism, 121 (47.83%) patients were harboring homozygous wild (*AA*) genotype, 106 (41.89%) possessed heterozygous (*AC*) genotype, and 26 (10.28%) were homozygote variant (*CC*). The genotypic frequencies of *MTHFR 1298A>C* ($\chi 2$ = 4.13, df = 1; p=0.99) and *MTHFR 677C>T* polymorphism was ($\chi 2$ = 0.15, df = 1; p=0.69). Thus, genotypic distributions for both the polymorphic sites followed the Hardy-Weinberg equilibrium with a p-value >0.05 (Table 5.61). Thus, these samples could represent a Mendelian population with a genetic equilibrium.

Table 5.61: Results of genetic equilibrium test

MTHFR Genotype rs1801133 *Ala^{222}Val (677C>T)*	**No. patients (%)**	**$\chi 2$**	**P-value**
CC	200 (79.05)		
CT	49 (19.37)	0.15	0.69
TT	4 (1.58)		
CT+TT	53 (20.95)		
MTHFR Genotype rs1801131 *Glu^{429}Ala (1298A>C)*	**No. patients (%)**	**$\chi 2$**	**P-value**
AA	121 (47.83)		
AC	106 (41.89)	4.13	0.99
CC	26 (10.28)		
AC+CC	132 (52.17)		

5.9.3. Overall survival analysis of *MTHFR 677C>T* and *1298A>C* genotypes

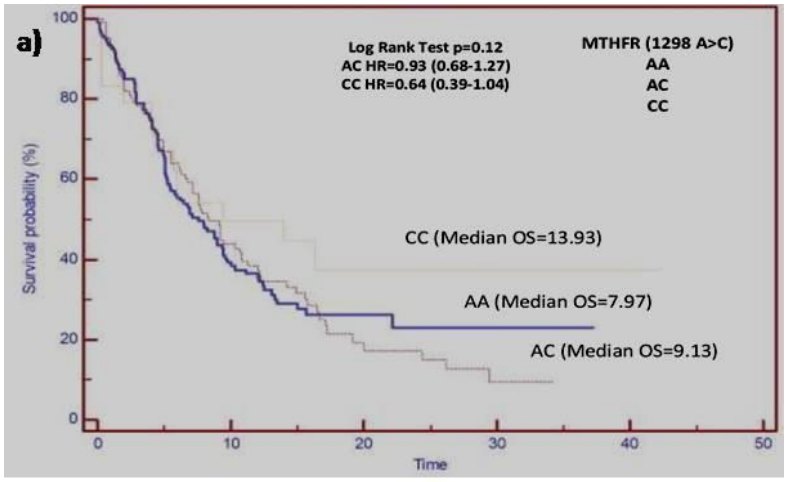
The survival analysis for 253 lung cancer subjects and their association with the *MTHFR* polymorphisms was carried out using the Kaplan-Meier and Cox regression method for the two genetic variants of the *MTHFR* gene and depicted in Table 5.62. The associations of the *MTHFR* SNPs with lung cancer survival were analyzed using univariate and multivariate analyses after adjusting for age, gender, smoking status, ECOG, KPS, and tumor stage. Lung cancer patients who were carrying both the mutant alleles (*CC*) for the *MTHFR 1298A>C* polymorphism had higher MST as compared to the patients bearing the wild-type genotype (*AA*) (MST=13.93 *vs.* 7.97, Log Rank p=0.12, Figure 5.32 a)). In the case of *677C>T* polymorphism, the carriers of the single variant allele (*CT*) showed a poor survival as compared to the wild genotype (5.20 months *vs.* 9.00 months). No significant difference in survival was observed for the *MTHFR 677C>T* polymorphism (Figure 5.32 b)). In the Cox proportional hazards model, after adjusting for various covariates like age, sex, smoking status, ECOG, KPS, and tumor stage, no prognostic

significance was observed for *MTHFR 677C>T* and *1298A>C* polymorphisms for OS in lung cancer subjects

Table 5.62: Univariate and Multivariate analysis for *MTHFR* variants

Genotype rs1801133 *Ala^{222}Val* (*677C>T*)	CASES n(%), N=253	DEAD n(%), N=172	ALIVE n(%), N=81	Univariate analysis			Multivariate analysis	
				MST (months)	Log-rank *p*	Unadjusted HR[a]	Adjusted HR[b] (95% C.I.)	*p*
CC	200 (79.05)	139 (80.81)	61 (75.31)	9.00				
CT	49 (19.37)	32 (18.61)	17 (20.99)	5.20	0.38	1.20 (0.79-1.78)	1.27 (0.85-1.88)	0.24
TT	4 (1.58)	1 (0.58)	3 (3.70)	-	0.16	0.27 (0.09-0.76)	0.21 (0.03-1.53)	0.13
CT+TT	53 (20.95)	33 (19.19)	20 (24.69)	5.93	0.70	1.08 (0.73-1.59)	0.95 (0.63-1.44)	0.83
Genotype rs1801131 *Glu^{429}Ala (1298A>C)*	**CASES n(%), N=253**	**DEAD n(%), N= 172**	**ALIVE n(%), N= 81**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	121 (47.83)	86 (50.00)	35 (43.21)	7.97				
AC	106 (41.89)	72 (41.86)	34 (41.98)	9.13	0.63	0.93 (0.68-1.27)	0.93 (0.66-1.31)	0.69
CC	26 (10.28)	14 (8.14)	12 (14.81)	13.93	0.12	0.64 (0.39-1.04)	0.76 (0.42-1.38)	0.37
AC+CC	132 (52.17)	86 (50.00)	46 (56.79)	9.23	0.32	0.86 (0.64-1.16)	0.88 (0.64-1.22)	0.45

[a]UnadjustedHazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG



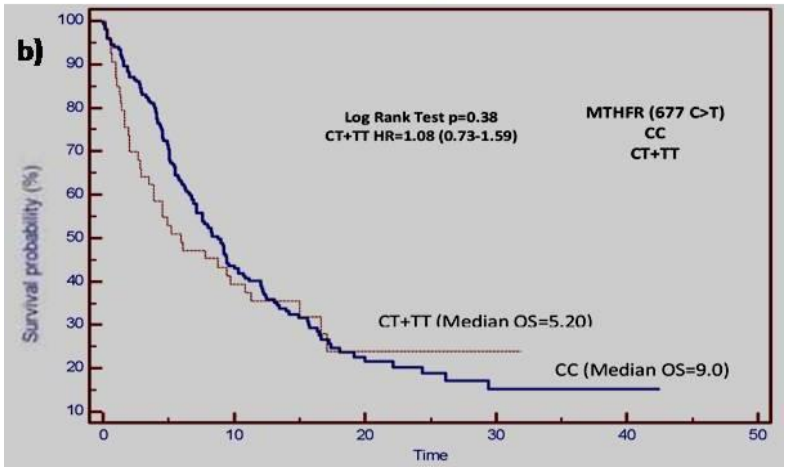


Figure 5.32: a) Kaplan Meier curves illustrating the association between overall survival in different genotypes of *MTHFR A1298C* polymorphism in ADCC subjects. (Note: *AA* represents wild-type genotype, *AC* represents heterozygous genotype, and *CC* represents homozygous for mutant genotype).

b) Kaplan Meier curves illustrate the association between overall survival in different genotypes of *MTHFR C677T* polymorphism in ADCC subjects. (Note: *CC* represents wild-type genotype, *CT* represents heterozygous genotype, and mutant genotype).

5.9.4. Association of *MTHFR* polymorphism with chemotherapy regimens and OS

Most lung cancer patients were treated with platinum-based doublet chemotherapy and pemetrexed as second-line therapy in the present study. Thus, we want to assess the modulating effects of the *MTHFR 677C>T* and *1298A>C* polymorphism and its association between chemotherapy and overall patient survival. The outcome of the *MTHFR* polymorphisms on survival according to the chemotherapy regimen *viz.* cisplatin/carboplatin and pemetrexed used are as shown in Table 5.63. For the *MTHFR 677C>T* polymorphism, the frequency of patients carrying mutant genotype (*TT*) was significantly less as compared to wild genotype (1.39% *vs.* 79.86%). Thus, subjects with mutant genotype exhibit inferior survival than the wild (0.60 months *vs*. 9.13 months, log-rank p=0.98) but did not exhibit any significance in survival between different genotypes of *MTHFR 677C>T* polymorphism as compared to the wild genotype. At the same time, the subjects carrying heterozygous for *MTHFR 1298A>C* genotype exhibited a better MST than those carrying the homozygous wild type genotype (MST=10.67 vs. 6.93 months; Log-rank p=0.23). In the multivariate Cox proportional hazards model, the hazards ratio was lower for such patients (HR'=0.69, 95% C.I=0.45-1.08, p=0.11). The survival of subjects carrying mutant genotype was also higher than the wild (9.43 months, log-rank p=0.46) but lacked statistical significance (Table 5.63).

Table 5.63: Association of *MTHFR* SNPs on overall survival according to Chemotherapeutic regimen

REGIMEN- Pemetrexed cisplatin/carboplatin								
Genotype rs1801133 *Ala[222]Val (677C>T)*	**CASES n(%), N=144**	**DEAD n(%), N=97**	**ALIVE n(%), N=47**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	115 (79.86)	81 (83.51)	34 (72.34)	9.13				
CT	27 (18.75)	15 (15.46)	12 (25.53)	6.10	0.87	0.96 (0.55-1.64)	0.96 (0.56-1.74)	0.96
TT	2 (1.39)	1 (1.03)	1 (2.13)	0.60	0.98	0.97 (0.14-6.83)	0.72 (0.10-5.34)	0.75
CT+TT	29 (20.14)	16 (16.49)	13 (27.66)	6.10	0.87	0.96 (0.56-1.62)	0.96 (0.56-1.67)	0.89
Genotype rs1801131 *Glu[429]Ala (1298A>C)*	**CASES n(%), N=144**	**DEAD n(%), N=97**	**ALIVE n(%), N=45**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	68 (47.22)	49 (50.52)	19 (42.22)	6.93				
AC	65 (45.14)	41 (42.27)	24 (53.33)	10.67	0.23	0.78 (0.51-1.17)	0.69 (0.45-1.08)	0.11
CC	11 (7.64)	7 (7.22)	4 (8.89)	9.43	0.46	0.74 (0.36-1.51)	0.61 (0.26-1.40)	0.24
AC+CC	76 (52.78)	48 (49.48)	28 (62.22)	10.67	0.19	0.77 (0.51-1.15)	0.69 (0.45-1.06)	0.09

[a]UnadjustedHazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.9.5. Association of *MTHFR* polymorphisms with gender and OS

Further, the analysis was carried out by segregating the data based on gender, as described in Table 5.64. The subjects harboring mutant genotype for the *MTHFR 677C>T* polymorphism in the case of females was limited as compared to subjects carrying the wild genotype (MST= 0.60 *vs.* 12.07, Log-rank p= **<0.0001**), the reason behind this could be that only one patient was carrying the mutant genotype for *MTHFR 677C>T* polymorphism. The hazards ratio obtained after multivariate analysis was very high (HR'=68.43, 95% C.I.=4.99-939.13, p=**0.0017**) as the sample size is too small to say anything with surety. In the male's case, the survival of heterozygotes was almost half that of the subjects with wild genotype (4.53 months *vs.* 8.30 months). Further, applying the univariate analysis, the subjects possessing the mutant genotype of *MTHFR 1298A>C* polymorphism in the case of both males (MST=9.43 *vs.* 7.60 months) and females (MST=13.93 *vs.* 7.97) correlated with a higher overall survival but was not significant (Table 5.64). Thus, both the polymorphisms of the *MTHFR* gene were not correlated with OS of lung cancer patients even after stratification based on gender.

5.9.6. Association of the *MTHFR* polymorphism & clinic-pathological characteristics with OS

The effect of *MTHFR* polymorphisms on the various clinic-pathological features such as tumor stage and performance status (including KPS, and ECOG) was also calculated in terms of OS. The patients were stratified into stage III and stage IV patients (Table 5.65 a)). The heterozygous genotype (*CT*) showed poor survival of 4.90 months, and the mutant genotype (*TT*) exhibited the least survival of 0.60 months as compared to wild genotype (*CC*) (9.43 months) for *MTHFR 677C>T* polymorphism for stage III patients. In stage IV, the heterozygous genotype exhibits poor survival than the wild genotype (5.20 months *vs.* 7.97 months, log-rank p=0.84). In stage III subjects for *MTHFR 1298A>C* polymorphism, the heterozygotes (*AC*) had lower survival than the wild genotype (*AA*) (7.13 months *vs.* 9.57 months). The heterozygotes and mutant genotype (*CC*) in stage IV subjects exhibit higher survival rates of 9.13 months and 9.43 months than the wild type (6.43 months). Both the *MTHFR* variants did not show any significant association with the prognosis of the patients, neither in stage III subjects nor stage IV subjects (Table 5.65 a)).

Table 5.64: Association of *MTHFR* SNPs on overall survival based on gender

MALE								
Genotype rs1801133 *Ala[222]Val (677C>T)*	**CASES n(%), N=173**	**DEAD n(%), N=120**	**ALIVE n(%), N=53**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	135 (78.03)	95 (79.17)	40 (75.47)	8.30				
CT	35 (20.24)	25 (20.83)	10 (18.87)	4.53	0.13	1.40 (0.86-2.28)	1.45 (0.90-2.33)	0.12
TT	3 (1.73)	0 (0.00)	3 (5.66)	-	0.05	-	0.00	0.96
CT+TT	38 (21.97)	25 (20.83)	13 (24.53)	4.53	0.50	1.16 (0.73-1.84)	1.11 (0.69-1.77)	0.66
Genotype rs1801131 *Glu[429]Ala (1298A>C)*	**CASES n(%), N=173**	**DEAD n(%), N=120**	**ALIVE n(%), N=53**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	73 (42.20)	52 (43.33)	21 (39.62)	7.60				
AC	82 (47.40)	59 (49.17)	23 (43.40)	7.57	0.91	1.02 (0.70-1.48)	1.03 (0.69-1.54)	0.86
CC	18 (10.40)	9 (7.50)	9 (16.98)	9.43	0.23	0.65 (0.35-1.21)	0.75 (0.36-1.57)	0.45
AC+CC	100 (57.80)	68 (56.67)	32 (60.38)	7.70	0.74	0.94 (0.65-1.35)	0.96 (0.65-1.40)	0.83
FEMALE								
Genotype rs1801133 *Ala[222]Val (677C>T)*	**CASES n(%), N=80**	**DEAD n(%), N=52**	**ALIVE n(%), N=28**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	65 (81.25)	44 (84.62)	21 (75.00)	12.07				
CT	14 (17.50)	7 (13.46)	7 (25.00)	7.80	0.43	0.73 (0.35-1.49)	0.55 (0.21-1.42)	0.22
TT	1 (1.25)	1 (1.92)	0 (0.00)	0.60	**<0.0001**	22.13 (0.002-203684.99)	68.43 (4.99-939.13)	**0.0017**
CT+TT	15 (18.75)	8 (15.38)	7 (25.00)	7.80	0.62	0.83 (0.41-1.68)	0.67 (0.28-1.61)	0.37
Genotype rs1801131 *Glu[429]Ala (1298A>C)*	**CASES n(%), N=80**	**DEAD n(%), N=52**	**ALIVE n(%), N=28**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	48 (60.00)	34 (65.38)	14 (50.00)	7.97				
AC	24 (30.00)	13 (25.00)	11 (39.29)	15.63	0.16	0.64 (0.35-1.15)	0.57 (0.26-1.25)	0.16
CC	8 (10.00)	5 (9.62)	3 (10.71)	13.93	0.31	0.62 (0.28-1.38)	0.58 (0.19-1.75)	0.34
AC+CC	32 (40.00)	18 (34.62)	14 (50.00)	15.63	0.11	0.63 (0.37-1.09)	0.60 (0.30-1.19)	0.14

[a]UnadjustedHazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

Table 5.65 a): Association of *MTHFR* SNPs on overall survival according to clinic-pathological features

Tumor stage- III								
Genotype rs1801133 *Ala^{222}Val (677C>T)*	**CASES n(%), N=74**	**DEAD n(%), N=54**	**ALIVE n(%), N=20**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	62 (83.78)	45 (83.33)	17 (85.00)	9.43				
CT	10 (13.51)	8 (14.82)	2 (10.00)	4.90	0.35	1.42 (0.60-3.35)	1.86 (0.83-4.14)	0.13
TT	2 (2.70)	1 (1.85)	1 (5.00)	0.60	0.95	1.06 (0.14-8.20)	1.41 (0.18-10.88)	0.74
CT+TT	12 (16.22)	9 (16.67)	3 (15.00)	4.90	0.38	1.37 (0.61-3.04)	1.76 (0.83-3.76)	0.14
Genotype rs1801131 *Glu^{429}Ala (1298A>C)*	**CASES n(%), N=74**	**DEAD n(%), N=54**	**ALIVE n(%), N=20**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	39 (52.70)	29 (53.70)	10 (50.00)	9.57				
AC	26 (35.14)	21 (38.89)	5 (25.00)	7.13	0.41	1.25 (0.71-2.23)	1.32 (0.71-2.43)	0.38
CC	9 (12.16)	4 (7.41)	5 (25.00)	-	0.14	0.47 (0.21-1.06)	0.55 (0.18-1.63)	0.28
AC+CC	35 (47.30)	25 (46.30)	10 (50.00)	7.57	0.96	0.99 (0.58-1.68)	1.07 (0.60-1.94)	0.81
Tumor stage- IV								
Genotype rs1801133 *Ala^{222}Val (677C>T)*	**CASES n(%), N=155**	**DEAD n(%), N=105**	**ALIVE n(%), N=50**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	120 (77.42)	86 (81.90)	34 (68.00)	7.97				
CT	33 (21.29)	19 (18.10)	14 (28.00)	5.20	0.84	0.95 (0.58-1.55)	0.98 (0.59-1.61)	0.93
TT	2 (1.29)	0 (0.00)	2 (4.00)	-	0.08	-	0.00	0.95
CT+TT	35 (22.58)	19 (18.10)	16 (32.00)	6.10	0.48	0.84 (0.52-1.34)	0.83 (0.50-1.36)	0.46
Genotype rs1801131 *Glu^{429}Ala (1298A>C)*	**CASES n(%), N=155**	**DEAD n(%), N=105**	**ALIVE n(%), N=50**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	73 (47.10)	51 (48.57)	22 (44.00)	6.43				
AC	68 (43.87)	45 (42.86)	23 (46.00)	9.13	0.36	0.83 (0.56-1.24)	0.81 (0.53-1.23)	0.32
CC	14 (9.03)	9 (8.57)	5 (10.00)	9.43	0.58	0.82 (0.42-1.60)	0.95 (0.46-1.96)	0.90
AC+CC	82 (52.90)	54 (51.43)	28 (56.00)	9.20	0.34	0.83 (0.56-1.22)	0.83 (0.56-1.23)	0.35

[a]UnadjustedHazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, KPS, ECOG

The subjects were categorized based on KPS with KPS scores of 90-100, 70-80, and those with KPS score 60 and below 60 (Table 5.65 b)). In the case of *MTHFR 677C>T* polymorphism, the subjects carrying heterozygous genotype exhibit poor survival, which is almost half the survival of subjects with wild genotype in both the subjects with KPS score of 90-100(6.10 months *vs*. 12.07 months, log-rank p=0.53) and KPS score of 70-80 (4.90 months *vs*. 8.23 months, log-rank p=0.76). In the case of *1298A>C*, the subjects with KPS score of 90-100 and carrying mutant genotype (*CC*) exhibit the least survival of 5.93 months as compared to subjects with wild genotype (*AA*) (5.93 months *vs*. 12.07 months, log-rank p=0.99), for subjects with KPS score 70-80 the carriers of a double allele of variant genotype (CC) exhibit a better survival than the wild genotype (*AA*) (13.93 months *vs.* 5.80 months, log-rank p=0.08). Thus, both the polymorphisms of the *MTHFR* gene did not play any significant association in the survival of patients (Table 5.65 b)).

Based on ECOG, patients were split into ECOG scores (0,1), (2), and (3,4) groups (Table 5.65 c)). For the *677C>T* polymorphism, the survival of heterozygous genotype was 5.20, 4.90, and 3.47 months in subjects with ECOG scores of (0,1), (2), and (3,4), which is comparatively lower than the carriers of wild genotype. For both these polymorphisms of *MTHFR*, no significant association towards survival was observed. Regarding clinical stage and performance status, no significant association was found towards the prognosis of patients based on these clinic-pathological features (Table 5.65 c)).

Table 5.65 b): Association of *MTHFR* SNPs on overall survival according to clinic-pathological features

KPS (90-100)								
Genotype rs1801133 *Ala^{222}Val (677C>T)*	**CASES n(%), N=73**	**DEAD n(%), N=42**	**ALIVE n(%), N=31**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	60 (82.19)	35 (83.33)	25 (80.65)	12.07				
CT	13 (17.81)	7 (16.67)	6 (19.35)	6.10	0.53	1.29 (0.53-3.14)	1.04 (0.43-2.55)	0.93
TT	-	-	-	-	-	-	-	-
CT+TT	13 (17.81)	7 (16.67)	6 (19.35)	6.10	0.53	1.29 (0.53-3.14)	1.04 (0.43-2.55)	0.93
Genotype rs1801131 *Glu^{429}Ala (1298A>C)*	**CASES n(%), N=73**	**DEAD n(%), N=42**	**ALIVE n(%), N=31**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	35 (47.95)	20 (47.62)	15 (48.39)	12.07				
AC	28 (38.36)	17 (40.48)	11 (35.48)	9.20	0.63	1.17 (0.61-2.25)	1.36 (0.63-2.94)	0.43
CC	10 (13.69)	5 (11.90)	5 (16.13)	5.93	0.99	1.001 (0.37-2.67)	1.07 (0.34-3.33)	0.91
AC+CC	38 (52.05)	22 (52.38)	16 (51.61)	9.20	0.71	1.12 (0.61-2.05)	1.29 (0.64-2.61)	0.48
KPS (70-80)								
Genotype rs1801133 *Ala^{222}Val (677C>T)*	**CASES n(%), N=121**	**DEAD n(%), N=87**	**ALIVE n(%), N=34**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	92 (76.03)	68 (78.16)	24 (70.59)	8.23				
CT	26 (21.49)	18 (20.69)	8 (23.53)	4.90	0.76	1.09 (0.64-1.85)	1.01 (0.59-1.74)	0.96
TT	3 (2.48)	1 (1.15)	2 (5.88)	-	0.44	0.47 (0.12-1.84)	0.38 (0.05-2.88)	0.35
CT+TT	29 (23.97)	19 (21.84)	10 (29.41)	5.20	0.96	1.01 (0.61-1.69)	0.92 (0.54-1.56)	0.75
Genotype rs1801131 *Glu^{429}Ala (1298A>C)*	**CASES n(%), N=121**	**DEAD n(%), N=87**	**ALIVE n(%), N=34**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	53 (43.80)	41 (47.13)	12 (35.29)	5.80				
AC	58 (47.93)	41 (47.13)	17 (50.00)	9.23	0.25	0.78 (0.50-1.20)	0.78 (0.48-1.25)	0.31
CC	10 (8.26)	5 (5.74)	5 (14.71)	13.93	0.08	0.45 (0.22-0.92)	0.55 (0.21-1.43)	0.22
AC+CC	68 (56.20)	46 (52.87)	22 (64.71)	9.40	0.12	0.72 (0.47-1.10)	0.75 (0.47-1.18)	0.21

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

Table 5.65 c): Association of *MTHFR* SNPs on overall survival according to clinic-pathological features

ECOG (0,1)								
Genotype rs1801133 *Ala^{222}Val (677C>T)*	**CASES n(%), N=141**	**DEAD n(%), N=92**	**ALIVE n(%), N=49**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	111 (78.72)	74 (80.43)	37 (75.51)	10.33				
CT	28 (19.86)	17 (18.48)	11 (22.45)	5.20	0.38	1.26 (0.71-2.23)	1.21 (0.70-2.10)	0.49
TT	2 (1.42)	1 (1.09)	1 (2.04)	0.60	0.66	0.65 (0.13-3.23)	0.51 (0.06-4.45)	0.54
CT+TT	30 (21.28)	18 (19.57)	12 (24.49)	5.20	0.49	1.20 (0.69-2.06)	1.13 (0.66-1.94)	0.66
Genotype rs1801131 *Glu^{429}Ala (1298A>C)*	**CASES n(%), N=141**	**DEAD n(%), N=92**	**ALIVE n(%), N=49**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	59 (41.84)	38 (41.30)	21 (42.86)	9.47				
AC	65 (46.10)	46 (50.00)	19 (38.77)	10.33	0.68	1.09 (0.71-1.68)	1.04 (0.65-1.65)	0.88
CC	17 (12.06)	8 (8.70)	9 (18.37)	-	0.36	0.70 (0.35-1.39)	0.55 (0.24-1.27)	0.16
AC+CC	82 (58.16)	54 (58.70)	28 (57.14)	10.67	0.99	0.99 (0.66-1.51)	0.93 (0.59-1.46)	0.75
ECOG (2)								
Genotype rs1801133 Ala^{222}Val (*677C>T*)	**CASES n(%), N=75**	**DEAD n(%), N=53**	**ALIVE n(%), N=22**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	60 (80.00)	43 (81.13)	17 (77.27)	6.77				
CT	13 (17.33)	10 (18.87)	3 (13.64)	4.90	0.41	1.33 (0.63-2.84)	1.17 (0.54-2.56)	0.69
TT	2 (2.67)	0 (0.00)	2 (9.09)	-	0.13	-	0.00	0.95
CT+TT	15 (20.00)	10 (18.87)	5 (22.73)	6.10	0.95	1.02 (0.51-2.05)	0.86 (0.40-1.83)	0.70
Genotype rs1801131 Glu^{429}Ala (*1298A>C*)	**CASES n(%), N=75**	**DEAD n(%), N=53**	**ALIVE n(%), N=22**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	43 (57.33)	34 (64.15)	9 (40.91)	5.43				
AC	27 (36.00)	16 (30.19)	11 (50.00)	7.80	0.23	0.70 (0.40-1.23)	0.67 (0.36-1.26)	0.22
CC	5 (6.67)	3 (5.66)	2 (9.09)	13.93	0.19	0.47 (0.19-1.13)	0.71 (0.21-2.37)	0.58
AC+CC	32 (42.67)	19 (35.85)	13 (59.09)	7.80	0.13	0.65 (0.38-1.11)	0.66 (0.37-1.19)	0.17
ECOG (3,4)								
Genotype rs1801133 Ala^{222}Val (*677C>T*)	**CASES n(%), N=31**	**DEAD n(%), N=23**	**ALIVE n(%), N=8**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
CC	26 (83.87)	20 (86.96)	6 (75.00)	5.53				
CT	5 (16.13)	3 (13.04)	2 (25.00)	3.47	0.57	0.71 (0.24-2.09)	1.01 (0.20-5.19)	0.99
TT	-	-	-	-	-	-	-	-
CT+TT	5(16.13)	3 (13.04)	2 (25.00)	3.47	0.57	0.71 (0.24-2.09)	1.01 (0.20-5.19)	0.99

Genotype rs1801131 Glu[429]Ala (*1298A>C*)	**CASES n(%), N=31**	**DEAD n(%), N=23**	**ALIVE n(%), N=8**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
AA	17 (54.84)	12 (52.17)	5 (62.50)	6.90				
AC	11 (35.48)	8 (34.78)	3 (37.50)	5.53	0.90	1.06 (0.43-2.60)	0.56 (0.17-1.84)	0.34
CC	3 (9.68)	3 (13.04)	0 (0.00)	1.97	0.97	0.98 (0.28-3.47)	0.28 (0.03-2.77)	0.28
AC+CC	14 (45.16)	11 (47.83)	3 (37.50)	4.47	0.88	1.06 (0.47-2.41)	0.58 (0.19-1.78)	0.34

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.9.7. Association of *MTHFR* polymorphism and chemotherapy response

Association between *MTHFR* polymorphisms and the efficacy outcome of response rate and clinical benefits were estimated using univariate logistic regression. The association of *MTHFR* polymorphism with chemotherapy response is displayed in Table 5.66. The patients exhibiting complete and partial responses were categorized as good responders, and those exhibiting sTable and progressive disease were restricted as inferior responders. For patients receiving chemotherapy, there was no statistically significant difference in response to chemotherapy. No significant relationship was illustrated between the *MTHFR* polymorphism according to response to platinum-based chemotherapy. Therefore, we conclude that none of the MTHFR polymorphisms could significantly predict chemotherapeutic response rate or clinical benefit.

Table 5.66: Association of *MTHFR* SNPs and chemotherapeutic response

	Response of chemotherapy		**AOR (95% C.I.)** [a]	***p*** [b]
Genotype rs1801133 Ala[222]Val (*677C>T*)	**CR+PR (40), n (%)**	**SD+PD (81), n (%)**		
CC	35 (87.50)	69 (85.19)	1.0(Reference)	
CT	5 (12.50)	10 (12.34)	0.99 (0.31-3.16)	0.98
TT	0 (0.00)	2 (2.47)	-	0.99
CT+TT	5 (12.50)	12 (14.81)	1.17 (0.37-3.63)	0.79
Genotype rs1801131 Glu[429]Ala (*1298A>C*)	**CR+PR (40), n (%)**	**SD+PD (81), n (%)**	**AOR (95% C.I.)** [a]	***p*** [b]
AA	20 (50.00)	42 (51.85)		
AC	15 (37.50)	32 (39.51)	0.87 (0.36-2.09)	0.76
CC	5 (12.50)	7 (8.64)	0.66 (0.18-2.43)	0.53
AC+CC	20 (50.00)	39 (48.15)	0.76 (0.33-1.71)	0.50

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding p-values were calculated by logistic regression analysis after adjusting for age and gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table

5.9.8. Cox proportional hazards regression model analysis of overall survival in adenocarcinoma

Table 5.67 shows the multivariate analysis using Cox regression to assess the significance for survival according to *MTHFR A1298C* polymorphism, age, gender, smoking status, tumor stage, and ECOG. Data indicated that ECOG status was the only independent prognostic factor; thus, there might be a potential association of ECOG status with the prognosis of ADCC patients (OR=1.55, 95% C.I.=1.26–1.91, p=**<0.0001**). Other clinical features were all seen to have no

apparent effect in influencing the prognostic outcomes of lung cancer patients receiving chemotherapy.

Table 5.67: Cox proportional hazards regression model analysis of overall survival in ADCC subjects

N=253 **Censored=23** **Uncensored=60** **Covariates**	**b**	**SE**	**P-value**	**Adjusted OR (Exp b)**	**95% C.I. of Exp (b)**
Age	0.002169	0.007784	0.7805	1.0022	0.9871 to 1.0175
Sex	-0.1302	0.2075	0.5302	0.8779	0.5858 to 1.3157
Smoking	0.2311	0.1975	0.2420	1.2600	0.8572 to 1.8518
Stage	0.1560	0.1679	0.3530	1.1688	0.8424 to 1.6217
ECOG	0.4370	0.1072	<0.0001	1.5481	1.2561 to 1.9080
MTHFR_A1298C	-0.07625	0.1252	0.5424	0.9266	0.7259 to 1.1827

5.9.9. Association between *MTHFR* Genotypes and Toxicity

In this study, one of the current outcomes of interest was to evaluate the development of toxicity in adenocarcinoma patients administered with pemetrexed-based platinum doublet chemotherapy. The univariate logistic regression analysis was employed to calculate the odds ratios (OR) and 95% confidence intervals (95% C.I.) to estimate the risk of developing different toxicity grades after therapy in patients with each specific genotype. We also used the multivariate regression model to calculate the adjusted ORs, with the dependent variable being the toxicity grades. The multivariate model included sex, age, performance status, and polymorphisms as covariates and was evaluated for possible interaction or confounding effects. To observe the impact of the polymorphisms on different grades of toxicity grades, we performed two types of analysis. In the first analysis, we compared the presence of any toxicity grade (grades 1-4) versus absence of toxicity (grade 0), in the second analysis, we compared the subjects having intermediate/ severe toxicity (grades 2-4) versus no /low-grade toxicity (0-1) to comprehensively evaluate whether the *MTHFR* polymorphisms were associated with severe toxicities. In this way, the patients were grouped into two different toxicity grades.

All chemotherapy-related toxicities were recorded for each treatment cycle. A total of 123 patients' toxicity data was available. Table 5.68 a) demonstrates the hematological toxicity, where 7 (46.67%) patients were classified as grade 1-3 neutropenia and were carrying a single copy of the mutant allele (*CT*) of the *MTHFR C677T* polymorphism. For *MTHFR A1298T*

polymorphism 6 (40.00%) patients were harboring the heterozygous genotype (*AT*). Table 5.60 a) reveals that patients with the heterozygous genotype (*CT*) for the *MTHFR C677T* polymorphism had a 5.34-fold increased risk of developing neutropenia toxicity (AOR=5.34, 95% C.I.= 1.49-19.06, **p=0.009**) when compared to patients with the wild type genotype (CC). A comparison was made between patients with any toxicity grade (grades 1-4) versus the absence of toxicity (grade 0). Univariate and multivariate analysis found no statistically significant association between the *MTHFR* polymorphism and other hematological toxicities such as thrombocytopenia (Table 5.68 a)).

We also analyzed gastrointestinal toxicity by comparing patients with low-grade toxicity (WHO grades 1) with intermediate-high grades (WHO grades 2-3). As shown in Table 5.68 b), lung cancer patients with the *AT* (heterozygous) genotype (AOR=12.54, 95% C.I.=1.54-101.86, **p= 0.018**) for *MTHFR A1298C* polymorphism had a 12-fold risk of diarrhea as compared to subjects carrying both copies of the wild alleles (*AA*) after adjusting for confounding factors like age, gender, and performance status. Further, the combined genotype (*AT+TT*) for *A1298C* polymorphism showed a more pronounced diarrhea risk (AOR=14.48, 95% C.I.= 1.78-117.89, **p=0.01**). The patients harboring heterozygous genotype for *A1298C* polymorphism the heterozygous genotype (AOR= 7.23, 95% C.I.= 0.99-52.99, **p=0.04**) exhibit a 7-fold risk for diarrhea when patients with low-intermediate grade toxicity (WHO grades 1-2) were compared to those with high grades (WHO grades 2-3). We did not observe any association between the *MTHFR* polymorphism nephrotoxicity toxicity and other adverse events due to chemotherapy (Table 5.68 c)).

Table 5.68 a): Hematological Toxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4: grades 0-2 vs. 3-4) associated with *MTHFR* polymorphisms in multivariate analysis

Neutropenia						
Genotype MTHFR *C677T*	**Grade of toxicity 0,n(%)**	**Grade of toxicity 1-3,n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	81(82.65)	8(53.33)	1.00 (Reference)		1.00 (Reference)	
CT	15(15.31)	7(46.67)	4.72 (1.49-14.99)	**0.008**	5.34 (1.49-19.06)	**0.009**
TT	2(2.04)	0(0.00)	1.92 (0.08-43.32)	0.68	-	0.99
CT+TT	17(17.35)	7(46.67)	4.17 (1.33-13.05)	**0.01**	4.45 (1.28-15.43)	**0.019**
Genotype MTHFR *A1298T*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	47 (47.96)	9(60.00)	1.00 (Reference)		1.00 (Reference)	
AT	43(43.88)	6(40.00)	0.73 (0.24-2.22)	0.58	0.98 (0.30-3.23)	0.97
TT	8(8.16)	0(0.00)	0.29 (0.01-5.54)	0.41	-	0.99
AT+TT	51(52.04)	6(40.00)	0.61 (0.20-1.86)	0.39	0.79 (0.24-2.58)	0.70
Genotype MTHFR *C677T*	**Grade of toxicity 0-1,n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	82(79.61)	7(70.00)	1.00 (Reference)		1.00 (Reference)	
CT	19(18.45)	3(30.00)	1.85 (0.44-7.82)	0.40	1.96 (0.40-9.62)	0.40
TT	2(1.94)	0(0.00)	2.20 (0.10-50.17)	0.62	-	0.99
CT+TT	21(20.39)	3(30.00)	1.67 (0.40-7.03)	0.48	1.71 (0.36-8.21)	0.50
Genotype MTHFR *A1298T*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	49(47.57)	7(70.00)	1.00 (Reference)		1.00 (Reference)	
AT	46(44.66)	3(30.00)	0.46 (0.11-1.87)	0.28	0.65 (0.14-2.97)	0.58
TT	8(7.77)	0(0.00)	0.39 (0.02-7.44)	0.53	-	0.99
AT+TT	54(52.43)	3(30.00)	0.39 (0.09-1.59)	0.19	0.54 (0.12-2.45)	0.43
Leukopenia						
Genotype MTHFR *C677T*	**Grade of toxicity 0, N (%)**	**Grade of toxicity 1-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	72 (80.90)	25 (73.53)	1.00 (Reference)		1.00 (Reference)	
CT	15 (16.85)	9 (26.47)	1.73 (0.67-4.44)	0.25	1.87 (0.69-5.02)	0.21
TT	2 (2.25)	0 (0.00)	0.57 (0.03-12.25)	0.72	-	0.99
CT+TT	17 (19.10)	9 (26.47)	1.52 (0.60-3.85)	0.37	1.60 (0.61-4.20)	0.34

Genotype MTHFR *A1298T*	**Grade of toxicity 0,N(%)**	**Grade of toxicity 1-3,N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
AA	45 (50.56)	18 (52.94)	1.00 (Reference)		1.00 (Reference)	
AT	37 (41.57)	14 (41.18)	0.95 (0.41-2.15)	0.89	1.19 (0.49-2.84)	0.70
TT	7 (7.87)	2 (5.88)	0.71 (0.14-3.77)	0.69	0.91 (0.16-5.30)	0.91
AT+TT	44 (49.44)	16 (47.06)	0.91 (0.41-2.01)	0.81	1.07 (0.46-2.48)	0.87
Genotype MTHFR *C677T*	**Grade of toxicity 0-1, N (%)**	**Grade of toxicity 2-3, N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	91 (78.45)	6 (85.71)	1.00 (Reference)		1.00 (Reference)	
CT	23 (19.83)	1 (14.29)	0.66 (0.07-5.75)	0.71	0.71 (0.07-6.79)	0.77
TT	2 (1.72)	0 (0.00)	2.81 (0.12-64.97)	0.52	-	0.99
CT+TT	25 (21.55)	1 (14.29)	0.61 (0.07-5.27)	0.65	0.64 (0.07-6.06)	0.70
Genotype MTHFR *A1298T*	**Grade of toxicity 0-1,N (%)**	**Grade of toxicity 2-3,N (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
AA	59 (50.86)	4 (57.14)	1.00 (Reference)		1.00 (Reference)	
AT	48 (41.38)	3 (42.86)	0.92 (0.20-4.32)	0.92	1.58 (0.29-8.63)	0.59
TT	9 (7.76)	0 (0.00)	0.69 (0.03-13.99)	0.81	-	0.99
AT+TT	57 (49.14)	3 (42.86)	0.78 (0.17-3.62)	0.75	1.32 (0.24-7.12)	0.75
Anemia						
Genotype MTHFR *C677T*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	14 (70.00)	83 (80.58)	1.00 (Reference)		1.00 (Reference)	
CT	5 (25.00)	19 (18.45)	0.64 (0.20-1.99)	0.44	0.58 (0.17-1.95)	0.38
TT	1 (5.00)	1 (0.97)	0.17 (0.01-2.85)	0.22	0.08 (0.004-1.59)	0.10
CT+TT	6 (30.00)	20 (19.42)	0.56 (0.19-1.64)	0.29	0.49 (0.15-1.55)	0.22
Genotype MTHFR *A1298T*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
AA	8 (40.00)	55 (53.40)	1.00 (Reference)		1.00 (Reference)	
AT	10 (50.00)	41 (39.81)	0.60 (0.22-1.64)	0.32	0.80 (0.27-2.38)	0.69
TT	2 (10.00)	7 (6.79)	0.51 (0.09-2.89)	0.45	0.64 (0.09-4.41)	0.65
AT+TT	12 (60.00)	48 (46.60)	0.58 (0.22-1.54)	0.28	0.61 (0.21-1.75)	0.36
Genotype MTHFR *C677T*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
CC	51 (77.27)	46 (80.70)	1.00 (Reference)		1.00 (Reference)	
CT	14 (21.21)	10 (17.54)	0.79 (0.32-1.95)	0.61	0.82 (0.31-2.18)	0.69
TT	1 (1.52)	1 (1.76)	1.11 (0.07-18.24)	0.94	0.88 (0.05-15.45)	0.93

CT+TT	15 (22.73)	11 (19.30)	0.81 (0.34-1.95)	0.64	0.84 (0.33-2.14)	0.71
Genotype MTHFR *A1298T*	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-4, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
AA	30 (45.45)	33 (57.89)	1.00 (Reference)		1.00 (Reference)	
AT	31 (46.97)	20 (35.09)	0.59 (0.28-1.24)	0.16	0.73 (0.32-1.64)	0.44
TT	5 (7.58)	4 (7.02)	0.73 (0.18-2.96)	0.66	1.29 (0.24-6.87)	0.77
AT+TT	36 (54.55)	24 (42.11)	0.61 (0.29-1.24)	0.17	0.74 (0.34-1.59)	0.44
			Thrombocytopenia			
Genotype MTHFR *C677T*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-2, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	77 (77.78)	20 (83.33)	1.00 (Reference)		1.00 (Reference)	
CT	20 (20.20)	4 (16.67)	0.77 (0.24-2.51)	0.66	0.75 (0.22-2.55)	0.64
TT	2 (2.02)	0 (0.00)	0.76 (0.03-16.37)	0.86	-	0.99
CT+TT	22 (22.22)	4 (16.67)	0.70 (0.22-2.26)	0.55	0.68 (0.20-2.30)	0.54
Genotype MTHFR *A1298T*	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-2, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	51 (51.52)	12 (50.00)	1.00 (Reference)		1.00 (Reference)	
AT	42 (42.42)	9 (37.50)	0.91 (0.35-2.37)	0.85	1.27 (0.45-3.59)	0.65
TT	6 (6.06)	3 (12.50)	2.12 (0.46-9.73)	0.33	4.54 (0.78-26.34)	0.09
AT+TT	48 (48.48)	12 (50.00)	1.06 (0.43-2.59)	0.89	1.48 (0.56-3.95)	0.43
Genotype MTHFR *C677T*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	96 (78.69)	1 (100.00)	1.00 (Reference)		1.00 (Reference)	
CT	24 (19.67)	0 (0.00)	1.31 (0.05-3323)	0.87	-	0.99
TT	2 (1.64)	0 (0.00)	12.87 (0.41-400.52)	0.14	0.63 (-)	1.00
CT+TT	26 (21.31)	0 (0.00)			-	0.99
Genotype MTHFR *A1298T*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	63 (51.64)	0 (0.00)	1.00 (Reference)		1.00 (Reference)	
AT	50 (40.98)	1 (100.00)	3.77 (0.15-94.59)	0.42	-	0.99
TT	9 (7.38)	0 (0.00)	6.68 (0.12-357.31)	0.35	-	-
AT+TT	59 (48.36)	1 (100.00)	3.20 (0.13-80.14)	0.48	1.88 (0.03-109.66)	0.76

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratio data were calculated by unconditional logistic regression and adjusted for age, gender, and performance status. A p value < 0.05 was considered statistically significant

We also analyzed gastrointestinal toxicity by comparing patients with low-grade toxicity (WHO grades 1) with intermediate-high grades (WHO grades 2-3). As shown in Table 5.68 b), lung cancer patients with the *AT* (heterozygous) genotype (AOR=12.54, 95% C.I.=1.54-101.86, **p= 0.018**) for *MTHFR A1298C* polymorphism had a 12-fold risk of diarrhea as compared to subjects carrying both copies of the wild alleles (*AA*) after adjusting for confounding factors like age, gender, and performance status. Further, the combined genotype (*AT+TT*) for *A1298C* polymorphism showed a more pronounced diarrhea risk (AOR=14.48, 95% C.I.= 1.78-117.89, **p=0.01**). The patients harboring heterozygous genotype for *A1298C* polymorphism the heterozygous genotype (AOR= 7.23, 95% C.I.= 0.99-52.99, **p=0.04**) exhibit a 7-fold risk for diarrhea when patients with low-intermediate grade toxicity (WHO grades 1-2) were compared to those with high grades (WHO grades 2-3). We did not observe any association between the *MTHFR* polymorphism nephrotoxicity toxicity and other adverse events due to chemotherapy (Table 5.68 c)).

Table 5.68 b): GastrointestinalToxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4: grades 0-2 vs. 3-4) associated with *MTHFR* polymorphisms in multivariate analysis

Diarrhoea						
Genotype MTHFR *C677T*	**Grade of toxicity 0-1,n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	14(77.78)	14(70.00)	1.00 (Reference)		1.00 (Reference)	
CT	4 (22.22)	6(30.00)	1.50 (0.35-6.50)	0.59	1.38 (0.29-6.52)	0.68
TT	0(0.00)	0(0.00)	1.00 (0.02-53.89)	1.00	-	0.99
CT+TT	4(22.22)	6(30.00)	1.50 (0.35-6.50)	0.59	1.38 (0.29-6.52)	0.68
Genotype MTHFR *A1298T*	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	15(83.33)	9(45.00)	1.00 (Reference)		1.00 (Reference)	
AT	2(11.11)	10(50.00)	8.33 (1.48-46.94)	**0.016**	12.54 (1.54-101.86)	**0.018**
TT	1(5.56)	1(5.00)	1.67 (0.09-30.06)	0.73	11.71 (0.13-1038.28)	0.28
AT+TT	3(16.67)	11(55.00)	6.11 (1.33-27.96)	**0.019**	14.48 (1.78-117.89)	**0.012**
Genotype MTHFR *C677T*	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	22(73.33)	6(75.00)	1.00 (Reference)		1.00 (Reference)	
CT	8(26.67)	2(25.00)	0.92 (0.15-5.51)	0.92	0.71 (0.10-4.92)	0.73
TT	0(0.00)	0(0.00)	3.46 (0.06-192.04)	0.54	-	0.99
CT+TT	8(26.67)	2(25.00)	0.92 (0.15-5.51)	0.92	0.71 (0.10-4.92)	0.73
Genotype MTHFR *A1298T*	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	21(70.00)	3(37.50)	1.00 (Reference)		1.00 (Reference)	
AT	7(23.33)	5(62.50)	5.00 (0.94-26.49)	0.05	7.23 (0.99-52.99)	**0.04**
TT	2(6.67)	0(0.00)	1.23 (0.05-31.36)	0.90	-	0.99
AT+TT	9(30.00)	5(62.50)	3.89 (0.76-19.86)	0.10	5.90 (0.84-41.57)	0.07
Constipation						
Genotype MTHFR *C677T*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	9(90.00)	20 (83.33)	1.00 (Reference)		1.00 (Reference)	
CT	1(10.00)	3(12.50)	1.35 (0.12-14.82)	0.81	1.74 (0.14-21.95)	0.67
TT	0(0.00)	1(4.17)	1.39 (0.05-37.38)	0.84	-	0.99
CT+TT	1(10.00)	4(16.67)	1.80 (0.17-18.47)	0.62	2.49 (0.20-30.32)	0.47

Genotype MTHFR *A1298T*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	6(60.00)	12(50.00)	1.00 (Reference)		1.00 (Reference)	
AT	2(20.00)	11(45.83)	2.75 (0.45-16.59)	0.27	3.42 (0.39-29.84)	0.27
TT	2(20.00)	1(4.17)	0.25 (0.02-3.34)	0.29	0.25 (0.01-4.26)	0.34
AT+TT	4(40.00)	12(50.00)	1.50 (0.33-6.70)	0.59	1.84 (0.30-11.28)	0.51

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratio data were calculated by unconditional logistic regression and adjusted for age, gender, and performance status. A p-value < 0.05 was considered statistically significant

Table 5.68 c): Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4: grades 0-2 vs. 3-4) associated with *MTHFR* polymorphisms in multivariate analysis

		Nephrotoxicity				
Genotype MTHFR *C677T*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	36 (78.26)	53 (79.10)	1.00 (Reference)		1.00 (Reference)	
CT	10 (21.74)	12 (17.91)	0.81 (0.32-2.09)	0.67	0.50 (0.15-1.71)	0.27
TT	0 (0.00)	2 (2.99)	3.41 (0.16-73.15)	0.43	-	0.99
CT+TT	10 (21.74)	14 (20.90)	0.95 (0.38-2.37)	0.91	0.64 (0.19-2.09)	0.46
Genotype MTHFR *A1298T*	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	22(47.83)	33(49.25)	1.00 (Reference)		1.00 (Reference)	
AT	20(43.48)	30(44.78)	1.00 (0.46-2.18)	1.00	1.89 (0.67-5.33)	0.23
TT	4(8.69)	4(5.97)	0.67 (0.15-2.95)	0.59	1.03 (0.17-6.28)	0.98
AT+TT	24(52.17)	34(50.75)	0.94 (0.44-2.00)	0.88	1.64 (0.62-4.30)	0.32
Genotype MTHFR *C677T*	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
CC	73(76.84)	15(88.24)	1.00 (Reference)		1.00 (Reference)	
CT	20(21.05)	2(11.76)	0.49 (0.10-2.31)	0.36	0.31 (0.05-1.90)	0.21
TT	2(2.11)	0(0.00)	0.95 (0.04-20.75)	0.97	-	0.99
CT+TT	22(23.16)	2(11.76)	0.44 (0.09-2.08)	0.30	0.29 (0.05-1.57)	0.15
Genotype MTHFR *A1298T*	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
AA	43(45.26)	12(66.67)	1.00 (Reference)		1.00 (Reference)	
AT	44(46.32)	6(33.33)	0.49 (0.17-1.42)	0.19	0.49 (0.14-1.72)	0.27
TT	8(8.42)	0(0.00)	0.20 (0.01-3.80)	0.29	-	0.99
AT+TT	52(54.74)	6(33.33)	0.41 (0.14-1.19)	0.10	0.39 (0.11-1.33)	0.13

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratio data were calculated by unconditional logistic regression and adjusted for age, gender, and performance status. A p-value < 0.05 was considered statistically significant

5.10. Statistical analysis of *GSTP1* gene

5.10.1. Patient characteristics and clinical predictors

The study group's demographic characteristics, including age, gender, smoking status, pack-years, histological subtypes, TNM staging, and other clinical parameters, are shown in Table 5.69. The study conducted comprised 682 lung cancer cases. The mean age was 59.76 ± 10.05 (range 28–86). There were 565 (82.84%) males and 117 (17.16%) females in the case group. In the present study, 79.33% of the cases were smokers, and 20.67% were non-smokers. Among the smokers, the mean pack-year was 27.79±28.40. Of the lung cancer cases, 40.62% were diagnosed with squamous-cell carcinoma (SQCC), 37.68% had adenocarcinoma (ADCC), and 19.35% had SCLC. TNM stage data were [stage I: 6 (0.88%), II: 22 (3.23%), III: 284 (41.64%), IV: 311 (45.60%)]. Tumor size T1 and T2 had very low frequencies of 5.42% and 9.53%, whereas T3 and T4 had 19.35 and 54.55%, respectively. For lymph-node involvement, N0 had 12.17%, whereas N1, N2, N3, and N4 had frequencies of 8.36, 42.37, 28.45, and 0.44%, respectively. Among all the cases for which TNM data were available, 45.46% had no metastasis involvement (M0), whereas 46.63% had distant metastasis (M1). Furthermore, the cases were also distributed accordingly to KPS and ECOG. In our study, there were 199 subjects (29.18%) with KPS scores of 100-90, 334 (48.97%) with KPS scores of 80-70 ($p<0.0001$), and 116 cases (17.01%) had a KPS score of less than 60. Based upon the ECOG score, there were 43.11% (294) subjects with ECOG scores between 0-1, 37.09% (253) were ECOG 2, and 14.96% (102) had an ECOG score of 3-4. Likewise, as shown in Table 5.69 out of 489 subjects who had undergone chemotherapy, 31.90% received pemetrexed+cis/carboplatin, 18.20% were administered irinotecan+cis/carboplatin, 18.61% received docetaxel+cisplatin/carboplatin, 17.79% received paclitaxel+cisplatin/carboplatin and 5.11% were administered with cis/carboplatin along with gemcitabine, respectively.

5.10.2. Genotypic distribution of *GSTP1*

For the *GSTP1 Ile^{105}Val* polymorphism, 428 (62.76%) patients were homozygous for the *A/A* genotype, whereas 222 (32.55%) were heterozygous *A/G* and 32 (4.69%) were variant homozygotes *G/G* (Table 5.70). The genotypic frequencies of *GSTP1Ile^{105}Val* did not exhibit any significant difference (χ^2=0.21; df=2; p=0.64).

Table 5.69: Clinical predictor among cases

Variable	Total (N)	Cases, n (%)	Variable	Total (N)	Cases, n (%)
Age (years)	682		**Lymph Node**	682	
Mean ± SD		59.76 ± 10.05	Nx		2 (0.29)
Range		28-86	N0		83 (12.17)
Gender	682		N1		57 (8.36)
Male		565 (82.84)	N2		289 (42.37)
Female		117 (17.16)	N3		194 (28.45)
Smoking status	682		N4		3 (0.44)
Smokers		541 (79.33)	Unknown		54 (7.92)
Non-smokers		141 (20.67)	**Metastasis**	682	
Pack years	541		Mx		2 (0.29)
Mean ± SD		27.79 ± 28.40	M0		310 (45.46)
Histological types	682		M1		318 (46.63)
SQCC		277 (40.62)	Unknown		52 (7.62)
ADCC		257 (37.68)	**Performance status**		
SCLC		132 (19.35)	KPS (below 60)	682	116 (17.01)
Others		4 (0.59)	KPS (70-80)		334 (48.97)
Unknown		12 (1.76)	KPS (90-100)		199 (29.18)
Overall survival	682		Unknown		33 (4.84)
Dead		502 (73.61)	ECOG (0-1)	682	294 (43.11)
Alive		180 (26.39)	ECOG (2)		253 (37.09)
TNM staging	682		ECOG (3-4)		102 (14.96)
I		6 (0.88)	Unknown		33 (4.84)
II		22 (3.23)	**Chemotherapy regimen**	489	
III		284 (41.64)	Pemetrexed cis/carboplatin		156 (31.90)
IV		311 (45.60)	Irinotecan cis/carboplatin		89 (18.20)
Unclassified		59 (8.65)	Docetaxel cis/carboplatin		91 (18.61)
Tumor Size	682		Paclitaxel cis/carboplatin		87 (17.79)
Tx		22 (3.23)	Gemcitabine cis/carboplatin		25 (5.11)
T1		37 (5.42)	Gefitinib		15 (3.07)
T2		65 (9.53)	Etoposide		17 (3.48)
T3		132 (19.35)	Ceretinib		1 (0.20)
T4		372 (54.55)	Others		8 (1.64)
Unknown		54 (7.92)			

Abbreviations: SD =Standard Deviation, n =total number of lung cancer cases or control subjects, a) *p*-values were derived from Pearson Chi-square test except for age and pack-years; Student t-test was used for age and pack-years. All *p*-values are two-sided. $p<0.05$ was considered statistically significant

Table 5.70: Genotypic distribution of *GSTP1* gene

Genotype rs1695 *Ile^{105}Val (313A>G)*	CASES n(%), N=682	DEAD n(%), N=502	ALIVE n(%), N=180
Ile/Ile	428 (62.76)	324 (64.54)	104 (57.78)
Ile/Val	222 (32.55)	156 (31.08)	66 (36.67)
Val/Val	32 (4.69)	22 (4.38)	10 (5.55)
Ile/Val+Val/Val	254 (37.24)	178 (35.46)	76 42.22)

5.10.3. Association of the *GSTP1 Ile^{105}Val* polymorphism & clinic-pathological characteristics

The effect of *GSTP1* variant SNPs on the various clinic-pathological features such as tumor stage, primary tumor extension, lymph node invasion, and distant metastasis was also calculated (Table 5.71). The patients were stratified into stage III and stage IV patients. The percentage of subjects carrying heterozygous genotype (*Ile/Val*) as well as mutant genotype (*Val/Val*) was almost similar among stage III and stage IV patients (*Ile/Val*: 29.22% *vs.* 31.19%; *Val/Val*: 4.93% *vs.* 5.79%, respectively). The odds ratio calculated using logistic regression was not found to be significant. The subjects were also categorized based on tumor extension into T3 and T4. It was found that the proportion of subjects carrying heterozygous genotype was represented more in T3 extension as compared to T4 tumor extension (37.12% *vs.* 29.30%). The wild genotype was presented more in the T4 subgroup than the T3 tumor extension group (64.78% *vs.* 59.09%). However, this genotype did not confer any effect on the tumor extension.

Based on lymph node invasion, patients were split into Nx+N0+N1 and N2+N3+N4 groups. The heterozygous patients were found at a slightly higher frequency among the Nx+N0+N1 (36.88%) than the N2+N3+N4 (30.04%) group. At the same time, the carriers of mutant genotype were found at a slightly higher frequency among the N2+N3+N4 (5.76%) than Nx+N0+N1 (36.88%) group. However, the difference was not statistically significant. Further, to classify the patients on the distant metastatic status, they were stratified into M0 and M1 groups. The frequency distribution of patients among M0 and M1 was almost similar in patients having heterozygous genotype (31.73% and 31.45%). Whereas, subjects carrying mutant genotype a bit higher percentage Regarding clinical stage, tumor size, lymph node invasion, and distant metastasis, no relationship between these clinic-pathological features and *GSTP1* polymorphism was observed in patients (Table 5.71)

Table 5.71: Relationship of different genotypes with the clinic-pathological parameters

Genotype rs1695 *Ile^{105}Val (313A>G)*	CLINICAL STAGE			
	III	**IV**	**AOR (95% C.I.) [a]**	***p*[b]**
	284, n(%)	**311, n(%)**		
Ile/Ile	187 (65.85)	196 (63.02)	1.0(Reference)	
Ile/Val	83 (29.22)	97 (31.19)	1.10 (0.76-1.60)	0.60
Val/Val	14 (4.93)	18 (5.79)	1.14 (0.54-2.40)	0.74
Ile/Val+Val/Val	97 (34.15)	115 (36.98)	1.11 (0.78-1.57)	0.56
Genotype rs1695 *Ile^{105}Val (313A>G)*	**LYMPH NODE INVASION**			
	Nx+N0+N1	**N2+N3+N4**	**AOR (95% C.I.) [a]**	***p*[b]**
	141, n(%)	**486, n(%)**		
Ile/Ile	85 (60.28)	312 (64.20)	1.0(Reference)	
Ile/Val	52 (36.88)	146 (30.04)	0.77 (0.51-1.14)	0.19
Val/Val	4 (2.84)	28 (5.76)	1.89 (0.64-5.57)	0.25
Ile/Val+Val/Val	56 (39.72)	174 (35.80)	0.85 (0.58-1.25)	0.41
Genotype rs1695 *Ile^{105}Val (313A>G)*				
	T3	**T4**	**AOR (95% C.I.) [a]**	***p*[b]**
	132, n(%)	**372, n(%)**		
Ile/Ile	78 (59.09)	241 (64.78)	1.0(Reference)	
Ile/Val	49 (37.12)	109 (29.30)	0.73 (0.48-1.13)	0.16
Val/Val	5 (3.79)	22 (5.91)	1.45 (0.53-3.99)	0.47
Ile/Val+Val/Val	54 (40.91)	131 (35.22)	0.80 (0.53-1.21)	0.29
Genotype rs1695 *Ile^{105}Val (313A>G)*	**METASTASIS**			
	No	**Yes**	**AOR (95% C.I.) [a]**	***p*[b]**
	312, n(%)	**318, n(%)**		
Ile/Ile	200 (64.10)	199 (62.58)	1.0(Reference)	
Ile/Val	99 (31.73)	100 (31.45)	1.01 (0.70-1.44)	0.97
Val/Val	13 (4.17)	19 (5.97)	1.38 (0.65-2.95)	0.40
Ile/Val+Val/Val	112 (35.90)	119 (37.42)	1.05 (0.75-1.48)	0.76

a)-Adjusted Odds ratios, 95% confidence intervals and corresponding *p*-values were calculated by logistic regression analysis after adjusting for age, gender, smoking status. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls.

5.10.4. Association of *GSTP1 IIe^{105}Val* polymorphism and chemotherapy response

Association between *GSTP1 IIe^{105}Val* polymorphisms and the efficacy outcome of response rate and clinical benefit were estimated using univariate logistic regression. Association of *GSTP1 IIe^{105}Val* polymorphism with chemotherapy response has been summed in Table 5.72. The patients exhibiting complete and partial responses were categorized as good responders, and those exhibiting stable and progressive disease were classified as bad responders. The frequency of patients carrying the mutant genotype was almost similar in both the group of responders (4.83% and 5.26%). For patients receiving chemotherapy, there was no statistically significant

difference in response to chemotherapy. No significant relationships were demonstrated between the *GSTP1 IIe*^{105}*Val* allele response to platinum-based chemotherapy (p=0.33). Thus, we can conclude that none of the *GSTP1 IIe*^{105}*Val* genotypes could significantly predict chemotherapeutic response rate or clinical benefit.

Table 5.72: Relationship of *GSTP1 Ile*^{105}*Val* polymorphism based on response tochemotherapy

	Response of chemotherapy		**AOR (95% C.I.)** [a]	***p*** [b]
Genotype rs1695 *Ile*^{105}*Val (313A>G)*	**CR+PR (145), n (%)**	**SD+PD (171), n (%)**		
Ile/Ile	98 (67.59)	106 (61.99)	1.0(Reference)	
Ile/Val	40 (27.58)	56 (32.75)	0.78 (0.48-1.29)	0.34
Val/Val	7 (4.83)	9 (5.26)	0.83 (0.29-2.33)	0.72
Ile/Val+Val/Val	47 (32.41)	65 (38.01)	0.79 (0.49-1.27)	0.33

a)-Adjusted Odds ratios, 95% confidence intervals, and corresponding p-values were calculated by logistic regression analysis after adjusting for age and gender. (b)-Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. The number in bold indicates the significant values in the Table

5.10.5. Survival analysis of *GSTP1* genotype

The survival analysis for 682 lung cancer subjects and their association with the *GSTP1* polymorphisms has been depicted in Table 5.73. The associations of the *GSTP1* SNP with lung cancer survival were analyzed using univariate and multivariate analyses after adjusting for age, gender, smoking status, ECOG, KPS, and tumor stage. Lung cancer patients who were carrying both the mutant alleles (*Val/Val*) for the *GSTP1 Ile*^{105}*Val* polymorphism had higher MST as compared to the patients bearing the wild type genotype (*Ile/Ile*) (MST=8.30 months *vs.* 7.47 months, Log-Rank p=0.56); however, no statistically significant difference in survival was observed between the chemotherapy and survival for the *GSTP1 Ile*^{105}*Val* polymorphism. After adjusting for various covariates like age, sex, smoking status, ECOG, KPS, and tumor stage in the Cox proportional hazards model, the prognostic significance was not observed for *GSTP1 Ile*^{105}*Val* polymorphism for OS in lung cancer subjects.

Table 5.73: Survival among patients of lung cancer categorized with *GSTP1 Ile^{105}Val* genotype

Genotype rs1695 *Ile^{105}Val* (*313A>G*)	CASES n(%), N=682	DEAD n(%), N=502	ALIVE n(%), N=180	Univariate analysis			Multivariate analysis	
				MST (months)	Log-rank *p*	Unadjusted HR[a]	Adjusted HR[b] (95% C.I.)	*P*
Ile/Ile	428 (62.76)	324 (64.54)	104 (57.78)	7.47		1.0(Reference)	1.0(Reference)	
Ile/Val	222 (32.55)	156 (31.08)	66 (36.67)	9.23	0.71	0.96 (0.80-1.17)	0.95 (0.78-1.17)	0.66
Val/Val	32 (4.69)	22 (4.38)	10 (5.55)	8.30	0.56	0.88 (0.58-1.32)	0.86 (0.56-1.33)	0.51
Ile/Val+Val/Val	254 (37.24)	178 (35.46)	76 (42.22)	9.23	0.60	0.95 (0.79-1.14)	0.94 (0.78-1.14)	0.54

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.10.6. Survival analysis of *GSTP1* polymorphism based on histological subtype

Furthermore, we evaluated the prognosis of lung cancer patients to the genotypes of *GSTP1* when stratified based on histological subtypes (Table 5.74). It was perceived that the ADCC subjects harboring mutant genotype (*Val/Val*) for *GSTP1 Ile^{105}Val* polymorphism had a higher median survival time than the subjects with wild-type genotype (*Ile/Ile*) (MST=12.07 *vs.* 7.13 months, Log-rank p=0.62, Figure 5.33). However, the data lacked significant values. When we evaluated our data in SQCC patients, we observed that SQCC patients carrying the heterozygous genotype (*Ile/Val*) had a higher OS as compared to the patients carrying the homozygous wild type genotype (*Ile/Ile*) (MST=8.63 *vs.* 6.23 months, Log-rank p=0.56), but lacked statistical significance (Table 5.74). A better MST was observed for the SCLC subjects harboring the mutant genotype than subjects with wild-type genotype (MST=11.00 *vs.* 8.27 months), and multivariate Cox proportional analysis with no prognostic significance was obtained for this genotype of *GSTP1 IIe^{105}Val* polymorphism (Table 5.74).

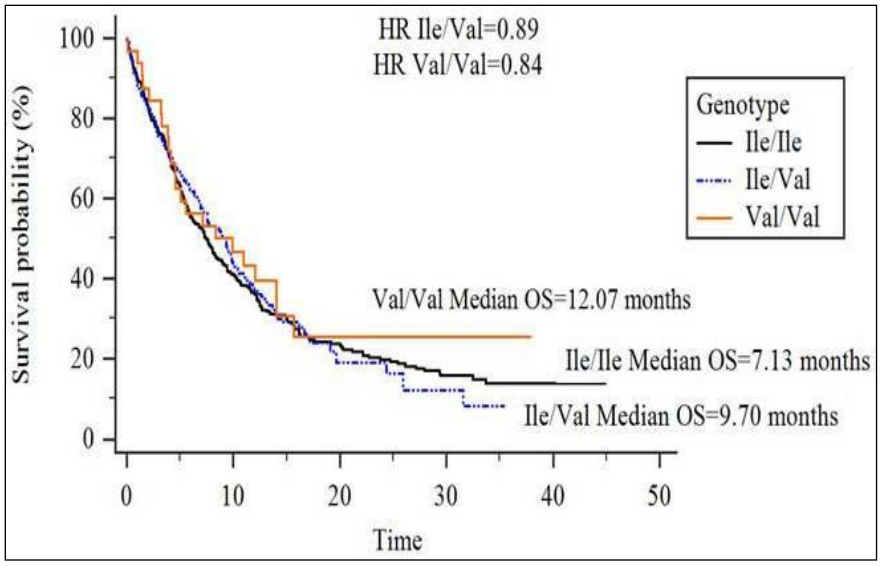


Figure 5.33: Kaplan Meier curves illustrating the association between overall survival in different genotypes of *GSTP1 Ile^{105}Val* polymorphism in ADCC subjects. (Note: *Ile/Ile* represents wild type genotype, *Ile/Val* represents heterozygous genotype, and *Val/Val* represents homozygous variant genotype)

Table 5.74: Association of *GSTP1 IIe^{105}Val* polymorphism on overall survival according to tumor histology

ADCC								
Genotype rs1695 *Ile^{105}Val* (313A>G)	**CASES n(%), N=257**	**DEAD n(%), N=174**	**ALIVE n(%), N=83**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	160 (62.26)	113 (64.94)	47 (56.63)	7.13		1.0(Reference)	1.0(Reference)	
Ile/Val	82 (31.91)	52 (29.89)	30 (36.14)	9.70	0.47	0.89 (0.64-1.22)	0.87 (0.62-1.23)	0.44
Val/Val	15 (5.83)	9 (5.17)	6 (7.23)	12.07	0.62	0.84 (0.45-1.58)	0.84 (0.43-1.67)	0.63
Ile/Val+Val/Val	97 (37.74)	61 (35.06)	36 (43.37)	9.70	0.42	0.88 (0.65-1.19)	0.87 (0.62-1.20)	0.39
SQCC								
Genotype rs1695 *Ile^{105}Val* (313A>G)	**CASES n(%), N=277**	**DEAD n(%), N=209**	**ALIVE n(%), N=68**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	179 (64.62)	140 (66.99)	39 (57.35)	6.23		1.0(Reference)	1.0(Reference)	
Ile/Val	89 (32.13)	63 (30.14)	26 (38.24)	8.63	0.56	0.91 (0.68-1.23)	0.99 (0.73-1.36)	0.99
Val/Val	9 (3.25)	6 (2.87)	3 (4.41)	7.10	0.56	0.79 (0.38-1.63)	0.75 (0.33-1.70)	0.49
Ile/Val+Val/Val	98 (35.38)	69 (33.01)	29 (42.65)	8.37	0.47	0.90 (0.68-1.19)	0.96 (0.71-1.29)	0.77
SCLC								
Genotype rs1695 *Ile^{105}Val* (313A>G)	**CASES n(%), N=132**	**DEAD n(%), N=112**	**ALIVE n(%), N=20**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	83 (62.88)	69 (61.61)	14 (70.00)	8.27		1.0(Reference)	1.0(Reference)	
Ile/Val	42 (31.82)	37 (33.03)	5 (25.00)	6.73	0.46	1.16 (0.77-1.75)	1.06 (0.68-1.67)	0.78
Val/Val	7 (5.30)	6 (5.36)	1 (5.00)	11.00	0.97	0.98 (0.43-1.25)	0.90 (0.36-2.24)	0.83
Ile/Val+Val/Val	49 (37.12)	43 (38.39)	6 (30.00)	9.00	0.52	1.13 (0.77-1.67)	1.02 (0.66-1.55)	0.94

[a]Unadjusted Hazards ratio for Kaplan meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.10.7. Association of *GSTP1 IIe^{105}Val* polymorphism with chemotherapy regimens and survival

In the present study, most lung cancer patients were subjected to treatment with platinum-based doublet chemotherapy. Therefore, we want to assess the modulating effects of the *GSTP1 IIe^{105}Val* polymorphism and its association between chemotherapy and overall patient survival. The outcome of the *GSTP1 IIe^{105}Val* polymorphism on survival after stratifying according to the chemotherapy regimens used is demonstrated in Table 5.75. Chemotherapy included four primary regimens, which were administered to different sets of patients. The four regimens used included a platinum-based drug (cisplatin and carboplatin), standard and administered to all lung cancer patients. The patients also received a second-line chemotherapeutic agent like docetaxel,

paclitaxel, pemetrexed, or irinotecan. The subjects administered with cisplatin/carboplatin and docetaxel were not found to exhibit any statistical significance in survival between different genotypes of *Ile^{105}Val* polymorphism compared to the wild genotype. It was observed that patients who were receiving cisplatin/carboplatin along with paclitaxel and who were carrying combined genotype for the *GSTP1 IIe^{105}Val* genotype (*Ile/Val+Val/Val*) had a better MST compared to patients carrying the homozygous wild type genotype (*Ile/Ile*) (MST=9.27 *vs.*5.63; Log-rank p=0.04). In the multivariate Cox proportional hazards model, the hazards ratio was lower for such patients (HR'=0.65, 95% C.I.=0.37-1.14, p=0.13). In lung cancer, patients harboring mutant genotype and receiving cisplatin/carboplatin along with irinotecan had a shorter MST-OS than the patients who have the wild type genotype (*Ile/Ile*) (MST=4.53 *vs.* 9.00, Log-Rank p=0.56). Further, the subjects administered cisplatin/carboplatin along with pemetrexed and carrying mutant genotype (*Val/Va*l) had a longer survival time (MST=12.07 *vs.* 6.73) compared to patients who carried the wild type genotype. However, data lacked statistical significance (Table 5.75).

Table 5.75: Association of *GSTP1 IIe^{105}Val* on overall survival according to chemotherapeutic regimen

REGIMEN- Docetaxel cisplatin/carboplatin								
Genotype rs1695 *Ile^{105}Val* *(313A>G)*	**CASES n(%), N=91**	**DEAD n(%), N=72**	**ALIVE n(%), N=19**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	65 (71.43)	53 (73.61)	12 (63.16)	9.43		1.0(Reference)	1.0(Reference)	
Ile/Val	25 (27.47)	18 (25.00)	7 (36.84)	9.73	0.86	0.95 (0.56-1.61)	1.01 (0.55-1.86)	0.98
Val/Val	1 (1.10)	1 (1.39)	0 (0.00)	9.87	0.77	1.34 (0.14-13.20)	1.25 (0.15-10.22)	0.84
Ile/Val+Val/Val	26 (28.57)	19 (26.39)	7 (36.84)	9.73	0.89	0.96 (0.57-1.62)	1.02 (0.56-1.87)	0.94
REGIMEN- Paclitaxel cisplatin/carboplatin								
Genotype rs1695 *Ile^{105}Val* *(313A>G)*	**CASES n(%), N=87**	**DEAD n(%), N=64**	**ALIVE n(%), N=23**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	55 (63.22)	45 (70.31)	10 (43.48)	5.63		1.0(Reference)	1.0(Reference)	
Ile/Val	31 (35.63)	19 (29.69)	12 (52.17)	9.27	0.06	0.60 (0.37-0.99)	0.69 (0.39-1.20)	0.19
Val/Val	1 (1.15)	0 (0.00)	1 (4.35)	-	0.23	-	-	0.96
Ile/Val+Val/Val	32 (36.78)	19 (29.69)	13 (56.52)	9.27	**0.04**	0.58 (0.35-0.95)	0.65 (0.37-1.14)	0.13
REGIMEN- Irinotecan cisplatin/carboplatin								
Genotype rs1695 *Ile^{105}Val* *(313A>G)*	**CASES n(%), N=89**	**DEAD n(%), N=73**	**ALIVE n(%), N=16**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	58 (65.17)	46 (63.01)	12 (75.00)	9.00		1.0(Reference)	1.0(Reference)	
Ile/Val	27 (30.34)	24 (32.88)	3 (18.75)	9.00	0.26	1.32 (0.78-2.22)	0.94 (0.51-1.74)	0.85
Val/Val	4 (4.49)	3 (4.11)	1 (6.25)	4.53	0.56	1.40 (0.36-5.47)	1.12 (0.32-3.84)	0.86

Ile/Val+Val/Val	31 (34.83)	27 (36.99)	4 (25.00)	6.73	0.23	1.33 (0.81-2.19)	0.96 (0.54-1.72)	0.90

REGIMEN- Pemetrexed cisplatin/carboplatin

Genotype rs1695 *Ile^{105}Val (313A>G)*	**CASES n(%), N=156**	**DEAD n(%), N=106**	**ALIVE n(%), N=50**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	97 (62.18)	67 (63.21)	30 (60.00)	6.73		1.0(Reference)	1.0(Reference)	
Ile/Val	47 (30.13)	31 (29.24)	16 (32.00)	9.57	0.65	0.91 (0.59-1.38)	0.80 (0.51-1.25)	0.33
Val/Val	12 (7.69)	8 (7.55)	4 (8.00)	12.07	0.66	0.85 (0.43-1.69)	0.73 (0.34-1.57)	0.42
Ile/Val+Val/Val	59 (37.82)	39 (36.79)	20 (40.00)	10.00	0.58	0.89 (0.61-1.32)	0.79 (0.52-1.20)	0.27

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.10.8. Association of *GSTP1 IIe^{105}Val* polymorphism with gender and OS

Further, the survival of lung cancer patients concerning *IIe^{105}Val* polymorphism of *GSTP1* was carried out based on gender as tabulated in Table 5.76. Applying the univariate analysis, the subjects possessing the heterozygous and combined genotype of *GSTP1 IIe^{105}Val* polymorphism in case of both males (MST=9.00 *vs.* 7.33 and MST=8.30 vs. 7.33, respectively) and females (MST=13.70 vs. 7.80 and MST=12.07 vs. 7.80, respectively) correlated with a higher overall survival but was not significant (Table 5.76). Thus, neither males nor females were significantly associated with the prognosis of lung cancer.

Table 5.76: Association of *GSTP1 IIe^{105}Val* on overall survival according to gender

Male

Genotype rs1695 *Ile^{105}Val (313A>G)*	**CASES n(%), N=565**	**DEAD n(%), N=420**	**ALIVE n(%), N=145**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	357 (63.19)	270 (64.28)	87 (60.00)	7.33		1.0(Reference)	1.0(Reference)	
Ile/Val	182 (32.21)	133 (31.67)	49 (33.79)	9.00	0.98	1.00 (0.81-1.23)	0.98 (0.79-1.22)	0.88
Val/Val	26 (4.60)	17 (4.05)	9 (6.21)	8.30	0.48	0.84 (0.53-1.32)	0.78 (0.48-1.28)	0.33
Ile/Val+Val/Val	208 (36.81)	150 (35.72)	58 (40.00)	9.00	0.84	0.98 (0.80-1.19)	0.95 (0.77-1.17)	0.64

Female

Genotype rs1695 *Ile^{105}Val (313A>G)*	**CASES n(%), N=117**	**DEAD n(%), N=82**	**ALIVE n(%), N=35**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log-rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	71 (60.68)	54 (65.85)	17 (48.57)	7.80		1.0(Reference)	1.0(Reference)	
Ile/Val	40 (34.19)	23 (28.05)	17 (48.57)	13.70	0.35	0.79 (0.50-1.27)	0.76 (0.44-1.30)	0.32
Val/Val	6 (5.13)	5 (6.10)	1 (2.86)	7.10	0.86	1.08 (0.42-2.80)	1.42 (0.50-4.09)	0.51
Ile/Val+Val/Val	46 (39.32)	28 (34.15)	18 (51.43)	12.07	0.43	0.83 (0.53-1.30)	0.84 (0.51-1.38)	0.49

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

5.10.9. Association of *GSTP1 IIe105Val* on overall survival according to clinic-pathological features

The effect of *GSTP1* polymorphisms on the various clinic-pathological features such as tumor stage, performance status (including KPS and ECOG) was also calculated about OS (Table 5.77). The patients were stratified into stage III and stage IV patients. The heterozygous genotype showed better survival of 9.27 months, and the mutant genotype exhibited the least survival of 7.10 months compared to the wild genotype (8.03 months) for *GSTP1 IIe^{105}Val* polymorphism for stage III patients. In stage IV, the mutant genotype exhibits slightly better survival thanthe wild genotype (8.30 months *vs*. 6.10 months, log-rank p=0.91). The *GSTP1 IIe^{105}Val* polymorphism did not show any significant association with the prognosis of the patients in stage III subjects or stage IV subjects.

The subjects were also categorized based on KPS with 90-100 and 70-80. The subjects carrying mutant genotype exhibit better survival than subjects with wild genotype in subjects with KPS score of 90-100 (11.00 months *vs*. 9.37 months, log-rank p=0.95). The subjects with a KPS score of 70-80 and carrying a mutant genotype exhibit the least survival of 5.57 months compared to subjects with a wild genotype of 7.23 months (log-rank p=0.54). Thus, the polymorphisms of the *GSTP1* gene did not play any significant association in the survival of patients.

Based on ECOG, patients were split into ECOG scores (0,1), (2), and (3,4) groups. The survival of subjects carrying mutant genotype was slightly higher as compared to wild genotype in subjects with ECOG score of (0,1) (11.00 months *vs*. 10.60 months, log-rank p=0.76) and ECOG (3,4) (5.07 months *vs*. 4.60 months, log-rank p=0.53). In these genotypes of *GSTP1,* no significant association towards survival was observed. Regarding clinical stage and performance status, no significant association was found towards the prognosis of patients based on these clinic-pathological features (Table 5.77).

Table 5.77: Association of *GSTP1* SNP (rs1695) on overall survival according to clinic-pathological features

Genotype rs1695 *Ile^{105}Val (313A>G)*	CASES n(%), N=284	DEAD n(%), N=214	ALIVE n(%), N=70	Univariate analysis			Multivariate analysis	
				MST (months)	Log-rank *p*	Unadjusted HR[a]	Adjusted HR[b] (95% C.I.)	*P*
Tumor stage- III								
Ile/Ile	187 (65.85)	146 (68.22)	41 (58.57)	8.03		1.0(Reference)	1.0(Reference)	
Ile/Val	83 (29.22)	61 (28.51)	22 (31.43)	9.27	0.96	1.01 (0.75-1.36)	1.01 (0.74-1.37)	0.96
Val/Val	14 (4.93)	7 (3.27)	7 (10.00)	7.10	0.15	0.58 (0.32-1.05)	0.60 (0.28-1.29)	0.19
Ile/Val+Val/Val	97 (34.15)	68 (31.78)	29 (41.43)	9.27	0.62	0.93 (0.70-1.24)	0.93 (0.69-1.25)	0.63
Tumor stage- IV								
Ile/Ile	196 (63.02)	148 (62.71)	48 (64.00)	6.10		1.0(Reference)	1.0(Reference)	
Ile/Val	97 (31.19)	73 (30.93)	24 (32.00)	6.73	0.92	1.01 (0.76-1.34)	0.90 (0.67-1.19)	0.46
Val/Val	18 (5.79)	15 (5.77)	3 (4.00)	8.30	0.91	1.03 (0.60-1.77)	1.18 (0.69-2.04)	0.54
Ile/Val+Val/Val	115 (36.98)	88 (37.29)	27 (36.00)	6.77	0.88	1.02 (0.78-1.33)	0.94 (0.72-1.23)	0.66
KPS (90-100)								
Ile/Ile	126 (63.32)	91 (65.47)	35 (58.33)	9.37		1.0(Reference)	1.0(Reference)	
Ile/Val	59 (29.65)	38 (27.34)	21 (35.00)	10.00	0.95	0.99 (0.68-1.44)	0.99 (0.67-1.48)	0.98
Val/Val	14 (7.03)	10 (7.19)	4 (6.67)	11.00	0.95	0.98 (0.51-1.87)	0.76 (0.38-1.51)	0.44
Ile/Val+Val/Val	73 (36.68)	48 (34.53)	25 (41.67)	10.33	0.92	0.98 (0.69-1.39)	0.95 (0.66-1.36)	0.77
KPS (80-70)								
Ile/Ile	219 (65.57)	170 (67.19)	49 (60.49)	7.23		1.0(Reference)	1.0(Reference)	
Ile/Val	103 (30.84)	75 (29.64)	28 (34.57)	8.73	0.65	0.94 (0.72-1.23)	0.94 (0.71-1.24)	0.66
Val/Val	12 (3.59)	8 (3.16)	4 (4.94)	5.57	0.54	0.80 (0.42-1.52)	0.81 (0.39-1.66)	0.57
Ile/Val+Val/Val	115 (34.43)	83 (32.81)	32 (39.51)	8.63	0.54	0.92 (0.71-1.19)	0.92 (0.71-1.21)	0.56
ECOG (0,1)								
Ile/Ile	193 (65.65)	131 (67.88)	62 (61.39)	10.60		1.0(Reference)	1.0(Reference)	
Ile/Val	82 (27.89)	49 (25.39)	33 (32.67)	11.13	0.75	0.95 (0.68-1.31)	1.03 (0.73-1.46)	0.84
Val/Val	19 (6.46)	13 (6.73)	6 (5.94)	11.00	0.76	1.09 (0.60-1.97)	1.14 (0.63-2.07)	0.67
Ile/Val+Val/Val	101 (34.35)	62 (32.12)	39 (38.61)	11.13	0.87	0.97 (0.72-1.32)	1.04 (0.76-1.43)	0.79
ECOG (2)								
Ile/Ile	187 (65.85)	146 (68.22)	41 (58.57)	8.03		1.0(Reference)	1.0(Reference)	
Ile/Val	83 (29.22)	61 (28.51)	22 (31.43)	9.27	0.96	1.01 (0.75-1.36)	1.01 (0.74-1.37)	0.96
Val/Val	14 (4.93)	7 (3.27)	7 (10.00)	7.10	0.15	0.58 (0.32-1.05)	0.60 (0.28-1.29)	0.19
Ile/Val+Val/Val	97 (34.15)	68 (31.78)	29 (41.43)	9.27	0.62	0.93 (0.70-1.24)	0.93 (0.69-1.25)	0.63

(313A>G)				**ths)**	***p***			
Ile/Ile	161 (63.64)	132 (64.71)	29 (59.18)	5.80		1.0(Reference)	1.0(Reference)	
Ile/Val	84 (33.20)	66 (32.35)	18 (36.73)	8.37	0.44	0.89 (0.67-1.19)	0.87 (0.64-1.18)	0.39
Val/Val	8 (3.16)	6 (2.94)	2 (4.08)	4.53	0.65	0.83 (0.39-1.76)	0.76 (0.33-1.75)	0.52
Ile/Val+Val/Val	92 (36.36)	72 (35.29)	20 (40.82)	7.80	0.40	0.88 (0.67-1.17)	0.86 (0.64-1.16)	0.32
ECOG (3,4)								
Genotype rs1695 *Ile^{105}Val (313A>G)*	**CASES n(%), N=102**	**DEAD n(%), N=87**	**ALIVE n(%), N=15**	**Univariate analysis**			**Multivariate analysis**	
				MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
Ile/Ile	59 (57.84)	53 (60.92)	6 (40.00)	4.60		1.0(Reference)	1.0(Reference)	
Ile/Val	38 (37.25)	31 (35.63)	7 (46.67)	4.47	0.87	1.04 (0.66-1.62)	1.01 (0.63-1.63)	0.96
Val/Val	5 (4.90)	3 (3.45)	2 (13.33)	5.07	0.53	0.69 (0.26-1.87)	0.93 (0.27-3.19)	0.91
Ile/Val+Val/Val	43 (42.16)	34 (39.08)	9 (60.00)	4.47	0.97	0.99 (0.64-1.53)	0.97 (0.62-1.54)	0.91

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, KPS, ECOG

5.10.10. Association Between *GSTP1 IIe^{105}Val* Genotypes and toxicity

One of the current outcomes of interest in this study was evaluating the toxicity development in lung cancer patients undergoing platinum-based doublet chemotherapy. The univariate logistic regression analysis was used to calculate the odds ratios (OR) and 95% confidence intervals (95% C.I.) to estimate the risk of developing different toxicity grades after therapy in patients with each specific genotype. We also used the multivariate regression model to calculate the adjusted ORs, with the dependent variable being the toxicity grades. The multivariate model included sex, age, performance status, and polymorphisms as covariates, and they were evaluated for possible interaction or confounding effects. To observe the impact of the polymorphisms on different grades of toxicity grades, we performed two types of analysis. In the first analysis, we compared the presence of any toxicity grade (grades 1-4) versus absence of toxicity (grade 0), and in the second analysis, we compared the subjects having intermediate/ severe toxicity (grades 2-4) versus no /low-grade toxicity (0-1) to comprehensively evaluate whether the *GSTP1* polymorphisms were associated with severe toxicities. In this way, the patients were subdivided into two different toxicity grade groups, and all patients were simultaneously included in the analyses.

All chemotherapy-related toxicities were recorded for each treatment cycle. 317 patients were enrolled for the study; hematologic toxicity was evaluated in 317 patients (Table 5.78 a)). As shown in Table 5.78 a), 33 (10.41%) of the patients were classified as grade 3-4 anemia and carried both copies of the mutant allele (*Val/Val*) of the *GSTP1* polymorphism. 182 (57.4%)

patients were classified as one is suffering from grade 0-2 anemia and carrying the wild type genotype (*Ile/Ile*), whereas 85 (26.81%) of the patients with grade 0-2 anemia was harboring the heterozygous genotype (*Ile/Val*) of the *GSTP1* polymorphism. Table 5.78 a) reveals that patients with the heterozygous genotype (*Ile/Val*) had a 2.12-fold increased risk of developing hematological toxicity, especially severe anemia (OR=2.14, 95% C.I.=0.97-4.62, **p=0.03**) when compared to patients with the wild-type genotype (*Ile/Ile*). When we compared the impact of the *GSTP1* polymorphism between grade 0 and grade 1-4 toxicity, it was seen that 5.67% of the patients were showing grade 1-4 anemia, and the risk of anemia toxicity was slightly increased for patients carrying the mutant genotype of the *GSTP1*, but not statistically significantly (OR=2.70, 95% C.I.=0.58-12.49).

We further analyzed hematological toxicity, especially leukopenia, by comparing patients with low or low-grade toxicity (WHO grades 0-1) with those with intermediate-high grades (WHO grades 2-4). As shown in Table 5.78 a), lung cancer patients with the *Val/Val* (mutant) genotype had an increased risk of leukopenia (OR=2.41; 95% C.I.=1.39-4.18, p=**0.001**) as compared to subjects carrying both copies of the wild alleles (*Ile/Ile*) after adjusting for confounding factors like age, gender, performance status, and chemotherapeutic regimen. No such association was found in the grade 0 *vs*. 1-4 category both in univariate and multivariate groups. The univariate and multivariate analysis found no statistically significant association between the *GSTP1* polymorphism and other hematological toxicities such as thrombocytopenia and neutropenia (Table 5.78 a)).

Table 5.78 a): Hematological Toxicities (grades 0 vs. 1-4: grades 0-1 vs. 2-4 : grades 0-2 vs 3-4) associated with *GSTP1 IIe^{105}Val* polymorphisms in multivariate analysis

Anaemia- Lung Cancer

GSTP1 IIe^{105}Val Genotype	**Grade of toxicity 0-2, N (%)**	**Grade of toxicity 3-4, N (%)**	**OR^{1} (95% C.I.)**	**P**	**AOR^{2} (95% C.I.)**	**P**
Ile/Ile	182 (57.4)	15 (4.73)	1.00 (Reference)		1.00 (Reference)	
Ile/Val	85 (26.81)	15 (4.73)	2.14 (1.00-4.58)	**0.03**	2.12 (0.97-4.62)	**0.04**
Val/Val	17 (5.36)	3 (0.94)	2.14 (0.56-8.14)	0.26	2.26 (0.58-8.71)	0.23
GSTP1 *IIe^{105}Val* Genotype	**Grade of toxicity 0 n (%)**	**Grade of toxicity 1-4, n (%)**	**OR^{1} (95% C.I.)**	**P**	**AOR^{2} (95% C.I.)**	**p**
Ile/Ile	39 (12.3)	158 (49.8)	1.00 (Reference)		1.0 (Reference)	
Ile/Val	17 (5.36)	83 (26.1)	1.20 (0.64-2.25)	0.56	1.11 (0.57-2.12)	0.75
Val/Val	2 (0.63)	18 (5.67)	2.22 (0.49-9.97)	0.29	2.70 (0.58-12.49)	0.19
Ile/Val + Val/Val	19	101	1.31 (0.71-2.39)	0.37	1.28 (0.69-2.38)	0.42
GSTP1 *IIe^{105}Val* Genotype	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-4, n (%)**	**OR^{1} (95% C.I.)**	**P**	**AOR^{2} (95% C.I.)**	**P**
Ile/Ile	106 (33.4)	91 (28.7)	1.00 (Reference)		1.00 (Reference)	
Ile/Val	53 (16.7)	47 (14.8)	1.03 (0.63-1.67)	0.89	0.98 (0.60-1.61)	0.96
Val/Val	11 (3.4)	9 (2.8)	0.95 (0.38-2.40)	0.92	1.00 (0.39-2.55)	0.99
Ile/Val + Val/Val	64 (20.1)	56 (17.6)	1.01 (0.64-1.60)	0.93	0.98 (0.62-1.57)	0.96

LEUKOPENIA- Lung cancer

GSTP1 *IIe^{105}Val* Genotype	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR^{1} (95% C.I.)**	**P**	**AOR^{2} (95% C.I.)**	**P**
Ile/Ile	145 (45.7)	52 (16.4)	1.00(Reference)		1.00(Reference)	
Ile/Val	72 (22.7)	28 (8.8)	1.08 (0.63-1.85)	0.76	1.05 (0.72-1.97)	0.81
Val/Val	12 (3.78)	8 (2.16)	1.36 (0.84-2.19)	0.20	1.31 (0.80-2.14)	0.27
Ile/Val + Val/Val	84 (26.4)	36 (11.3)	1.19 (0.72-1.97)	0.48	1.25 (0.75-2.09)	0.39
GSTP1 *IIe^{105}Val* Genotype	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR^{1} (95% C.I.)**	**P**	**AOR^{2} (95% C.I.)**	**P**
Ile/Ile	183 (57.7)	14 (4.41)	1.00(Reference)		1.00(Reference)	
Ile/Val	93 (29.3)	7 (2.2)	0.98 (0.38-2.52)	0.97	1.07 (0.41-2.80)	0.87
Val/Val	17 (5.3)	3 (0.94)	2.39 (0.92-3.76)	0.22	2.41 (1.39-4.18)	**0.001**
Ile/Val + Val/Val	110 (34.7)	10 (3.15)	1.18 (0.51-2.76)	0.68	1.24 (0.52-2.92)	0.61

THROMBOCYTOPENIA - Lung cancer

GSTP1 *IIe^{105}Val* Genotype	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR^{1} (95% C.I.)**	**P**	**AOR^{2} (95% C.I.)**	**P**
Ile/Ile	150 (63.03)	46 (58.97)	1.00 (Reference)		1.00 (Reference)	
Ile/Val	74 (31.09)	26 (33.33)	1.14 (0.65-1.99)	0.63	1.11 (0.62-1.96)	0.71
Val/Val	14 (5.88)	6 (7.69)	1.39 (0.50-3.84)	0.51	1.38 (0.49-3.87)	0.53
Ile/Val + Val/Val	88 (36.97)	32 (41.03)	1.18 (0.70-1.99)	0.52	1.14 (0.75-1.74)	0.51
GSTP1 *IIe^{105}Val* Genotype	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR^{1} (95% C.I.)**	**P**	**AOR^{2} (95% C.I.)**	**P**

Ile/Ile	192 (63.34)	4 (50.00)	1.00 (Reference)		1.00 (Reference)	
Ile/Val	96 (31.17)	4 (50.00)	2.00 (0.49-8.17)	0.33	2.02 (0.47-8.56)	0.33
Val/Val	20 (6.49)	0 (0.00)	1.04 (0.05-20.07)	0.98	-	
Ile/Val + Val/Val	116 (37.66)	4 (50.00)	1.65 (0.41-6.74)	0.48	2.02 (0.47-8.56)	0.33
NEUTROPENIA - Lung cancer						
Genotype	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR (95% C.I.)**	**P**	**AOR (95% C.I.)**	**P**
Ile/Ile	152 (62.55)	36 (62.07)	1.00 (Reference)		1.00 (Reference)	
Ile/Val	78 (32.10)	15 (25.86)	0.81 (0.42-1.57)	0.53	0.82 (1.42-1.63)	0.58
Val/Val	13 (5.35)	7 (12.07)	2.27 (0.84-6.11)	0.10	2.19 (0.79-6.01)	0.12
Ile/Val + Val/Val	91 (37.45)	22 (37.93)	1.02 (0.56-1.84)	0.94	1.20 (0.76-1.89)	0.42
Genotype	**Grade of toxicity 0-1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR (95% C.I.)**	**P**	**AOR (95% C.I.)**	**P**
Ile/Ile	164 (62.60)	24 (61.54)	1.00 (Reference)		1.00 (Reference)	
Ile/Val	81 (30.92)	12 (30.77)	1.01 (0.48-2.12)	0.97	1.01 (0.47-2.16)	0.96
Val/Val	17 (6.49)	3 (7.69)	1.20 (0.32-4.40)	0.77	1.16 (0.31-4.31)	0.82
Ile/Val + Val/Val	98 (37.40)	15 (38.46)	1.06 (0.61-1.81)	0.83	1.05 (0.61-1.80)	0.85

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status and type of treatment regimen.A P value < 0.05 was considered statistically significant

We did not observe any association between the *GSTP1* polymorphism nephrotoxicity (Table 5.78 b)) and gastrointestinal toxicity (Table 5.78 c)).

Table 5.78 b): Nephrotoxicity (grades 0 vs. 1-4: grades 0-1 vs. 2-4 : grades 0-2 vs 3-4) associated with *GSTP1 Ile^{105}Val* polymorphisms in multivariate analysis

Genotype	**Grade of toxicity 1, n(%)**	**Grade of toxicity 2-4, n(%)**	**OR (95% C.I.)**	**P**	**AOR (95% C.I.)**	**P**
Ile/Ile	64 (64.00)	127 (61.95)				
Ile/Val	29 (29.00)	65 (31.71)	1.13 (0.66-1.92)	0.65	0.63 (0.31-1.27)	0.19
Val/Val	7 (7.00)	13 (6.34)	0.94 (0.36-2.46)	0.89	1.72 (0.38-7.85)	0.48
Ile/Val + Val/Val	36 (36.00)	78 (38.05)	1.09 (0.66-1.79)	0.73	0.72 (0.37-1.40)	0.34

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status and type of treatment regimen.A P value < 0.05 was considered statistically significant

Table 5.78c): Gastrointestinal toxicity (grades 0 *vs*. 1-4: grades 0-1 *vs*. 2-4 : grades 0-2 *vs* 3-4) associated with *GSTP1 Ile*^{105}*Val* polymorphisms in multivariate analysis

ALOPECIA - Lung cancer						
Genotype	**Grade of toxicity No, n(%)**	**Grade of toxicity Yes, n(%)**	**OR (95% C.I.)**	**P**	**AOR (95% C.I.)**	**P**
Ile/Ile	13 (52.00)	6 (54.55)	1.00 (Reference)		1.00 (Reference)	
Ile/Val	8 (32.00)	5 (45.45)	1.35 (0.31-5.94)	0.69	1.41 (0.38-6.03)	0.61
Val/Val	4 (16.00)	0 (0.00)	0.23 (0.01-4.96)	0.35	0.28 (0.08-5.13)	0.41
Ile/Val + Val/Val	12 (48.00)	5 (45.45)	0.90 (0.22-3.74)	0.89	0.98 (0.28-3.83)	0.92
ANOREXIA - Lung cancer						
Genotype	**Grade of toxicity 1-2, n(%)**	**Grade of toxicity 3, n(%)**	**OR (95% C.I.)**	**P**	**AOR (95% C.I.)**	**P**
Ile/Ile	22 (56.41)	5 (83.33)	1.00 (Reference)		1.00 (Reference)	
Ile/Val	15 (38.46)	1 (16.67)	0.29 (0.03-2.77)	0.28	0.32 (0.06-3.23)	0.34
Val/Val	2 (5.13)	0 (0.00)	0.82 (0.03-19.60)	0.90	0.89 (0.06-20.23)	0.94
Ile/Val + Val/Val	17 (43.59)	1 (16.67)	0.26 (0.03-2.43)	0.24	0.31 (0.05-3.12)	0.28

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR**[2]Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status and type of treatment regimen. A P value < 0.05 was considered statistically significant

We also investigated the role of the *GSTP1* polymorphism and hematological toxicities after stratification based upon different histological subtypes of lung cancer. As shown in Table 5.78 d), in the case of SCLC patients, the polymorphism of *GSTP1* was associated with thrombocytopenia ($\chi 2$ test=7.32, p=**0.02**) where the minor allele G was predictive for more toxicity. The univariate analysis revealed that SCLC patients with a mutant genotype had a 9-fold risk of developing thrombocytopenia (OR=9.14, 95% C.I.=1.39-60.17, p=**0.02**) and adjusting for confounding factors, there was still a slight increase in toxicity was observed (OR=9.75, 95% C.I.=1.33-71.52, p=**0.01**). Further, when we combined both the heterozygous and mutant genotypes and compared it with the wild-type genotype, the polymorphism of *GSTP1* was found to be associated with thrombocytopenia in SCLC subjects both in univariate and multivariate models (OR=4.98, 95% C.I.=1.50-16.50 p=**0.008**) (Table 5.78 d)).

Table 5.78 d): Hematological Toxicities (grades 0 vs. 1-4: grades 0-1 vs. 2-4 : grades 0-2 vs 3-4) associated with *GSTP1 Ile*^{105}*Val* polymorphisms in multivariate analysis for SCLC patients

THROMBOCYTOPENIA-SCLC						
Genotype	**Grade of toxicity 0-1, n (%)**	**Grade of toxicity 2-3, n (%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2] (95% C.I.)**	**P**
Ile/Ile	38 (57.5)	1 (1.51)	1.00(Reference)		1.00(Reference)	
Ile/Val	20 (30.3)	1 (1.51)	1.90 (0.11-32.01)	0.66	3.15 (0.16-63.49)	0.45
Val/Val	6 (9.1)	0 (0.00)	1.97 (0.07-53.90)	0.69	-	0.99
Ile/Val + Val/Val	26 (39.4)	1 (1.51)	1.46 (0.09-24.43)	0.79	1.37 (0.15-12.51)	0.77
Genotype	**Grade of toxicity 0, n(%)**	**Grade of toxicity 1-4, n(%)**	**OR[1] (95% C.I.)**	**P**	**AOR[2](95% C.I.)**	**P**
Ile/Ile	32 (48.5)	7 (10.6)	1.00(Reference)		1.00(Reference)	
Ile/Val	13 (19.7)	8 (12.1)	2.81 (0.84-9.36)	0.09	4.19 (1.13-15.59)	**0.03**
Val/Val	2 (3.03)	4 (6.06)	9.14 (1.39-60.17)	**0.02**	9.75 (1.33-71.52)	**0.01**
Ile/Val + Val/Val	15(22.7)	12 (18.1)	3.66 (1.19-11.16)	**0.02**	4.98 (1.50-16.50)	**0.008**

OR[1] Crude odds ratio, 95% C.I.: 95% confidence interval, **AOR[2]**Adjusted Odds ratiodata were calculated by unconditional logistic regression and adjusted for age, gender, performance status and type of treatment regimen.A P value < 0.05 was considered statistically significant

5.11. High order gene-gene interaction

5.11.1. Demographic characteristics

The current study was a hospital-based case-control study that involved 550 cases and 550 healthy controls. The subjects in this study were evaluated based upon demographic characteristics such as age, gender, smoking status, pack-years, and histological subtypes, as summarized in Table 5.79. The mean age of cases was 59.23 ± 10.28 (range 28-86), whereas, among controls, it was 58.49 ± 11.67 (range 27-89). Among cases, there were 458 (83.27%) males and 92 (16.73%) females, whereas, among controls, there were 459 (83.45%) males and 91 (16.55%) females; thus, there was no difference observed in the distribution of males and females, suggesting significant matching (p=0.94). In the present study, 80.73% were smokers, and 19.27% were non-smokers in cases compared to 77.27% smokers and 22.73% non-smokers in controls. The smokers were elaborated in terms of pack-years. Pack years were significantly higher in cases than controls (31.16 ± 28.28 *vs.* 20.76 ± 15.95, p<0.0001). Based on the pathological classification, 216 (39.27%) patients were diagnosed with squamous cell carcinoma (SQCC), 200 (36.36%) with adenocarcinoma (ADCC), and 127 (23.1%) with small cell lung carcinoma (SCLC). There were 6 (1.09%) patients who were diagnosed with other histology as well.

Table 5.79: Demographic characteristics among cases and controls

Variable	Total (N)	Cases, n (%)	Controls, n (%)	[a]*p*-value
Age (years)	550			
Mean ± SD		59.23 ± 10.28	58.49 ± 11.67	0.27
Range		28-86	27-89	
Gender	550			
Male		458 (83.27)	459 (83.45)	
Female		92 (16.73)	91 (16.55)	0.94
Smoking status	550			
Smokers		444 (80.73)	425 (77.27)	
Non-smokers		106 (19.27)	125 (22.73)	0.18
Pack years	550			
Mean ± SD		31.16 ± 28.28	20.76 ± 15.95	***p*<0.0001**
Histological types	550			
SQCC		216 (39.27)		
ADCC		200 (36.36)		
SCLC		127 (23.1)		
Others		6 (1.09)		
Unknown		1 (0.18)		

Abbreviations: SD = Standard Deviation, n = total number of lung cancer cases or control subjects, a) *p*-values were derived from Pearson Chi-square test except for age and pack-years; Student t-test was used for age and pack-years. All *p*-values are two-sided. $p < 0.05$ was considered statistically significant

5.11.2. Genotypic combination of Phase II detoxification genes and their association with Lung cancer

Table 5.80 summarizes the minor allele frequency (MAF) in both cases and controls, and the risk of the three genotypes for the different SNPs of phase II detoxifying genesstudied (*NQO1, SULT1A1, EPHX1, NAT2*). As shown in Table 5.80, the MAF for Phase II detoxification genes was higher in controls than cases, except for *NAT2*7 G286E* and *NQO1 P187S,* wherein the MAF was higher in cases than controls (0.35 vs.0.32; 0.31 vs. 0.28, respectively). The adjusted odds ratio for individuals with homozygous variant genotype for *NAT2*5B L161L* showed a protective effect towards susceptibility for lung cancer (AOR=0.50, 95% C.I.= 0.34-0.73, **p=0.0003,** FDR=**0.002**). Furthermore, it was also evident that subjects who were heterozygous for the *NAT2*5B L161L* genotype also exhibited a protective effect towards LC (AOR= 0.69, 95% C.I.= 0.53-0.90, **p= 0.005**, FDR=**0.02**) and on combining the heterozygous genotype with the homozygous variant genotype a protective effect towards LC risk was also observed (AOR=

0.64, 95% C.I.= 0.50-0.83, p = **0.0005**, FDR=**0.004**)). On the contrary, it was also demonstrated that subjects who were carrying a single copy of the variant allele (heterozygous genotype) for the *NAT2*7 G286E* showed a risk towards susceptibly for LC (AOR =1.43, 95% C.I. = 1.11-1.85, p = **0.005**, FDR p=**0.002**; AOR = 1.36, 95% C.I. = 1.07-1.73, p= **0.01**). Further, the remaining polymorphic variants of Phase II detoxification genes studied did not show any prominent and underlying effect towards LC susceptibility and risk (Table 5.80).

The present study also examined the various genotypic combinations of phase II polymorphic variants and their possible association towards modulating risk for lung cancer, as shown in Table 5.81. For the current work, the double and triple combinations and the combinations of four SNPs were considered. The first segment of the study gives a clear account of the eight SNPs' double genotypic combinations and their susceptibility to lung cancer, as shown in Table 5.81. The dual combination data depicting their association with histological subtypes have been illustrated in Table 5.82. For all the combination analyses, the wild-type genotype was taken as the reference.

The genetic combination of *SULT1A1* $R^{213}H$ and *NAT2*5B* $L^{161}L$ had a higher representation of wild genotype (*GG+CC*) in cases than controls, respectively, (68.56 *vs.* 55.91%). In contrast, patients with heterozygous genotypes for both alleles (*GACT*) were more illustrated in controls than cases (41.36 *vs.*31.44%). Further, the combination of heterozygous and mutant subjects as a single genotype (*GA + AA + CT + TT*) depicted a higher representation in controls than cases (44.09% vs. 31.44%). Logistic regression analysis revealed a strong protective effect towards lung cancer risk (*GA + CT*) when compared with the reference subjects (AOR=0.59; 95% C.I.=0.39-0.87; **p=0.009**; FDRp=0.01). Similarly, the combination of heterozygous and mutant genotypes together as a single genotype showed a protective effect towards lung cancer risk (AOR=0.55; 95% C.I.=0.37-0.82; **p=0.003**; FDRp=0.03) (Table 5.81).

Table 5.80: Minor allele frequencies and risk associated with single-locus sites among cases and controls.

Gene	Polymorphism	Amino acid change	MAF controls	MAF cases	AOR_{het}[a]	p_{het}[b]	FDR p_{het}[b]	AOR_{homo}[a]	p_{homo}[b]	FDR p_{homo}[b]	AOR_{dom}[a]	p_{dom}[b]	FDR p_{dom}[b]
NQO1	*C609T* (rs1800566)	P187S	0.28	0.31	1.06 (0.83-1.36)	0.65	0.74	1.42 (0.91-2.22)	0.12	0.32	1.11 (0.88-1.41)	0.38	0.51
SULT1A1	*G638A* (rs9282861)	R13H	0.23	0.19	0.82 (0.63-1.07)	0.15	0.29	0.66 (0.38-1.12)	0.12	0.32	0.80 (0.62-1.02)	0.07	0.19
EPHX1	*T337C* (rs1051740)	Y113H	0.38	0.31	1.23 (0.95-1.60)	0.12	0.29	1.23 (0.87-1.75)	0.24	0.38	1.23 (0.96-1.57)	0.10	0.2
EPHX1	*A416G* (rs2234922)	H139R	0.17	0.17	1.07 (0.82-1.39)	0.61	0.74	0.77 (0.34-1.77)	0.54	0.71	1.05 (0.81-1.35)	0.73	0.83
NAT2*5B	*C481T* (rs1799929)	L161L	0.42	0.34	0.69 (0.53-0.90)	**0.005**	**0.02**	0.50 (0.34-0.73)	**0.0003**	**0.002**	0.64 (0.50-0.83)	**0.0005**	**0.004**
NAT2*6	*G590A* (rs1799930)	R197Q	0.37	0.35	1.04 (0.81-1.35)	0.74	0.74	0.78 (0.54-1.14)	0.20	0.38	0.98 (0.77-1.25)	0.86	0.86
NAT2*5C	*A803G* (rs1208)	K268R	0.32	0.31	0.84 (0.65-1.08)	0.18	0.29	1.04 (0.86-1.25)	0.71	0.71	0.89 (0.70-1.13)	0.34	0.51
NAT2*7	*G857A* (rs1799931)	G286E	0.32	0.35	1.43 (1.11-1.85)	**0.005**	**0.02**	1.09 (0.73-1.62)	0.67	0.71	1.36 (1.07-1.73)	**0.01**	**0.04**

Abbreviations: MAF controls and MAF cases-Minor allele frequency; AOR_{het}[a] - Adjusted Odds ratio for heterozygous; AOR_{homo}[a]-Adjusted Odds ratio for homozygous mutant; AOR_{dom}[a]- Adjusted Odds ratio for combined genotype.

[a]Adjusted odds ratios, 95% confidence intervals, and their corresponding p-values were calculated using logistic regression analysis after adjusting for age, gender, and smoking status. Highlighted p-value marks the significant value. [b] Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls.

Lung cancer subjects who were carrying a single copy of the variant allele for both the *EPHX1* $Y^{113}H$ and *NAT2*7* $G^{286}E$ polymorphism (heterozygous genotype, *TC+GA*) were more presented in the cases as compared with the controls (61.67% *vs*. 45.58%), whereas subjects with the wild genotype (*TT+GG*) were higher in controls than cases (51.16% *vs*. 33.33%). Logistic regression analysis revealed a 2-fold risk for lung cancer for individuals who were common single allelic carriers (*TC + GA*) compared with the reference subjects (AOR=2.09; 95% C.I. =1.38-3.17; **p=0.0005**; FDR p= **0.001**). Similarly, on combining both heterozygous and mutant genotypes, as a single group (*TC+ CC + GA+AA*), a 2-fold risk of lung cancer was also duly observed (AOR=2.10; 95% C.I.=1.39-3.17; **p=0.0004;** FDR p= **0.001**).

Table 5.81 also shows that subjects who were carrying a single copy of variant allele (*GA+AG*) for the genotypic combination of *SULT1A1* $R^{213}H$ and *NAT2*5C* $K^{268}R$ showed a protective effect towards lung cancer or no propensity towards lung cancer predisposition (AOR=0.66; 95% C.I.: 0.44-0.99; p=**0.004**; FDRp=0.21) as the frequency of subjects was less in cases as compared to controls (24.27 *vs*. 31.17%). Likewise, for the combination of heterozygous and mutant genotype as a single genotype (*GA + AA + AG + GG*), the frequency distribution suggested a higher representation of subjects in controls than cases (34.00 *vs*. 25.10%). A significant, small odds ratio was obtained on risk analysis (AOR= 0.63; 95% C.I.= 0.43-0.94; p = **0.02**; Fdr p = 0.16).

Subjects harboring both single allelic variant (*AG + CT*) genotype (29.57 *vs*. 38.71%) and combined variant (*AG + GG + CT + TT*) genotype (30.00 *vs*. 38.71%) had lower representation in cases as compared to controls for the combination of *EPHX1* $H^{139}R$ and *NAT2*5B* $L^{161}L$. A protective effect was observed for the heterozygous (*AG + CT*) (AOR= 0.66; 95% C.I.: 0.44-0.98; p = **0.04**; Fdr p = 0.21) genotype and subjects carrying heterozygous and mutant genotype (*AG + GG + CT + TT*) (AOR= 0.67; 95% C.I.=0.45-1.00; p= **0.049**; Fdr p =0.24) as a single genotype (Table 5.81).

The next genotypic combination considered was *EPHX1* $H^{139}R$ and *NAT2*7* $G^{286}E$. In this case the subjects with both heterozygous (*AG + GA*) (AOR= 1.54;95% C.I.: 1.04–2.29; p = **0.03**; Fdr p=0.19) and combined variant genotype (*AG + GG + GA + AA*) (AOR= 1.53; 95% C.I.= 1.03-2.26;p = **0.03**; Fdr p = 0.19) showed an increased risk for the disease, respectively.

Table 5.81: Distribution of genotypes depicting dual combination and their association with lung cancer risk

NQO1 P187S & SULT1A1 R213H	**Controls N= 256, n(%)**	**Cases N= 250, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC + GG)	172 (67.19)	179 (71.60)	1.00 (Reference)		
1 (CT + GA)	80 (31.25)	67 (26.80)	0.80 (0.54-1.19)	0.27	0.67
2 (TT + AA)	4 (1.56)	4 (1.60)	1.04 (0.25-4.34)	0.96	0.96
1-2	84 (32.81)	71 (28.40)	0.82 (0.56-1.20)	0.30	0.67
NQO1 P187S & EPHX1 Y113H	**Controls N= 209, n(%)**	**Cases N= 214, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC + TT)	103 (49.28)	94 (43.93)	1.00 (Reference)		
1 (CT + TC)	99 (47.37)	113 (52.80)	1.25 (0.84-1.84)	0.27	0.67
2 (TT + CC)	7 (3.35)	7 (3.27)	1.12 (0.37-3.36)	0.84	0.93
1-2	106 (50.72)	120 (56.07)	1.24 (0.84-1.82)	0.28	0.67
NQO1 P187S & EPHX1 H139R	**Controls N= 259, n(%)**	**Cases N= 248, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +AA)	192 (74.13)	178 (71.77)	1.00 (Reference)		
1 (CT + AG)	65 (25.10)	69 (27.82)	1.15 (0.78-1.72)	0.48	0.82
2 (TT + GG)	2 (0.77)	1 (0.41)	0.66 (0.06-7.62)	0.74	0.92
1-2	67 (25.87)	70 (28.23)	1.14 (0.77-1.69)	0.51	0.82
NQO1 P187S & NAT2*5B L161L	**Controls N= 219, n(%)**	**Cases N= 223, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +CC)	97 (44.29)	111 (49.78)	1.00 (Reference)		
1 (CT + CT)	116 (52.97)	107 (47.98)	0.78 (0.53-1.15)	0.21	0.60
2 (TT + TT)	6 (2.74)	5 (2.24)	0.84 (0.24-2.91)	0.78	0.93
1-2	122 (47.11)	112 (50.22)	0.78 (0.54-1.14)	0.20	0.60
NQO1 P187S & NAT2*6 R197Q	**Controls N= 216, n(%)**	**Cases N= 245, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +GG)	109 (50.46)	116 (47.35)	1.00 (Reference)		
1 (CT + GA)	100 (46.30)	120 (48.98)	1.17 (0.80-1.71)	0.41	0.74
2 (TT + AA)	7 (3.24)	9 (3.67)	1.30 (0.46-3.68)	0.62	0.89
1-2	107 (49.54)	129 (52.65)	1.18 (0.81-1.70)	0.39	0.74
NQO1 P187S & NAT2*5C K268R	**Controls N= 242, n(%)**	**Cases N= 212, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +AA)	136 (56.20)	126 (59.43)	1.00 (Reference)		
1 (CT + AG)	100 (41.32)	79 (37.26)	0.84 (0.57-1.23)	0.37	0.74
2 (TT + GG)	6 (2.48)	7 (3.31)	1.25 (0.40-3.85)	0.70	0.90
1-2	106 (43.80)	86 (40.57)	0.86 (0.59-1.26)	0.44	0.77
NQO1 P187S & NAT2*7 G286E	**Controls N= 236, n(%)**	**Cases N= 232, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +GG)	130 (55.08)	111 (47.84)	1.00 (Reference)		
1 (CT + GA)	103 (43.64)	114 (49.14)	1.31 (0.90-1.89)	0.16	0.53
2 (TT + AA)	3 (1.27)	7 (3.02)	2.62 (0.66-10.41)	0.17	0.53
1-2	106 (44.92)	121 (52.16)	1.35 (0.93-1.94)	0.11	0.43
SULT1A1 R213H & EPHX1 Y113H	**Controls N= 229, n(%)**	**Cases N= 210, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ TT)	137 (59.83)	127 (60.47)	1.00 (Reference)		
1 (GA + TC)	87 (37.99)	77 (36.67)	0.93 (0.63-1.38)	0.93	0.96
2 (AA + CC)	5 (2.18)	6 (2.86)	1.33 (0.39-4.48)	0.65	0.89
1-2	92 (40.17)	83 (39.52)	0.96 (0.65-1.41)	0.83	0.93
SULT1A1 R213H & EPHX1 H139R	**Controls N= 274, n(%)**	**Cases N= 321, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ AA)	228 (83.21)	261 (81.31)	1.00 (Reference)		
1 (GA + AG)	44 (16.06)	60 (18.69)	1.20 (0.78-1.85)	0.41	0.74
2 (AA + GG)	2 (0.73)	-	-	-	-

1-2	46 (16.79)	60 (18.69)	1.15 (0.75-1.76)	0.52	0.82
SULT1A1 R213H & NAT2*5B L161L	**Controls N= 220, n(%)**	**Cases N= 229, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ CC)	123 (55.91)	157 (68.56)	1.00 (Reference)		
1 (GA + CT)	91 (41.36)	72 (31.44)	0.59 (0.39-0.87)	**0.009**	**0.08**
2 (AA + TT)	6 (2.73)	-	-	-	-
1-2	97 (44.09)	72 (31.44)	0.55 (0.37-0.82)	**0.003**	**0.03**
SULT1A1 R213H & NAT2*6 R197Q	**Controls N= 213, n(%)**	**Cases N= 243, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ CC)	132 (61.97)	155 (63.79)	1.00 (Reference)		
1 (GA + CT)	74 (34.74)	82 (33.74)	0.95 (0.64-1.41)	0.81	0.93
2 (AA + TT)	7 (3.29)	6 (2.47)	0.80 (0.26-2.47)	0.70	0.90
1-2	81 (38.03)	88 (36.21)	0.94 (0.64-1.38)	0.75	0.92
SULT1A1 R213H & NAT2*5C K268R	**Controls N= 247, n(%)**	**Cases N= 239, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ AA)	163 (66.00)	179 (74.90)	1.00 (Reference)		
1 (GA + AG)	77 (31.17)	58 (24.27)	0.66 (0.44-0.99)	**0.04**	0.21
2 (AA + GG)	7 (2.83)	2 (0.83)	0.31 (0.06-1.54)	0.15	0.525
1-2	84 (34.00)	60 (25.10)	0.63 (0.43-0.94)	**0.02**	0.16
SULT1A1 R213H & NAT2*7 G286E	**Controls N=240 , n(%)**	**Cases N= 237, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ GG)	162 (67.50)	150 (63.29)	1.00 (Reference)		
1 (GA + GA)	72 (30.00)	83 (35.02)	1.21 (0.82-1.79)	0.34	0.71
2 (AA + AA)	6 (2.50)	4 (1.69)	0.72 (0.20-2.64)	0.62	0.89
1-2	78 (32.50)	87 (36.71)	1.18 (0.81-1.72)	0.40	0.74
EPHX1 Y113H & NAT2*5B L161L	**Controls N= 199, n(%)**	**Cases N= 218, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (TT + CC)	70 (35.18)	81 (37.16)	1.00 (Reference)		
1 (TC + CT)	118 (59.29)	126 (57.80)	0.90 (0.60-1.36)	0.63	0.89
2 (CC + TT)	11 (5.53)	11 (5.04)	0.89 (0.36-2.20)	0.80	0.93
1-2	129 (64.82)	137 (62.84)	0.90 (0.60-1.34)	0.60	0.89
EPHX1 Y113H & NAT2*6 R197Q	**Controls N= 209, n(%)**	**Cases N= 230, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (TT + GG)	88 (42.11)	83 (36.09)	1.00 (Reference)		
1 (TC + GA)	107 (51.20)	134 (58.26)	1.38 (0.93-2.06)	0.11	0.43
2 (CC + AA)	14 (6.69)	13 (5.65)	1.02 (0.45-2.32)	0.96	0.96
1-2	121 (57.89)	147 (63.91)	1.34 (0.91-1.98)	0.14	0.52
EPHX1 Y113H & NAT2*5C K268R	**Controls N= 219, n(%)**	**Cases N= 212, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (TT + AA)	102 (46.58)	101 (47.64)	1.00 (Reference)		
1 (TC + AG)	107 (48.86)	95 (44.81)	0.88 (0.59-1.30)	0.52	0.82
2 (CC + GG)	10 (4.56)	16 (7.55)	1.62 (0.70-3.76)	0.26	0.67
1-2	117 (53.42)	111 (52.36)	0.94 (0.64-1.38)	0.76	0.92
EPHX1 Y113H & NAT2*7 G286E	**Controls N= 215, n(%)**	**Cases N= 180, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (TT + GG)	110 (51.16)	60 (33.33)	1.00 (Reference)		
1 (TC + GA)	98 (45.58)	111 (61.67)	2.09 (1.38-3.17)	**0.0005**	**0.006**
2 (CC + AA)	7 (3.26)	9 (5.00)	2.41 (0.85-6.85)	0.10	0.43
1-2	105 (48.84)	120 (66.67)	2.10 (1.39-3.17)	**0.0004**	**0.006**
EPHX1 H139R & NAT2*5B L161L	**Controls N= 217, n(%)**	**Cases N= 230, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (AA + CC)	133 (61.29)	161 (70.00)	1.00 (Reference)		
1 (AG + CT)	84 (38.71)	68 (29.57)	0.66 (0.44-0.98)	**0.04**	0.21
2 (GG + TT)	-	1 (0.43)	-	-	-
1-2	84 (38.71)	69 (30.00)	0.67 (0.45-1.00)	**0.049**	0.24

EPHX1 H139R & NAT2*6 R197Q	**Controls N= 243, n(%)**	**Cases N= 226, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (AA + GG)	161 (66.26)	154 (68.14)	1.00 (Reference)		
1 (AG + GA)	79 (32.51)	69 (30.53)	0.91 (0.62-1.36)	0.65	0.89
2 (GG + AA)	3 (1.23)	3 (1.33)	1.12 (0.22-5.71)	0.89	0.96
1-2	82 (33.74)	72 (31.86)	0.92 (0.63-1.36)	0.69	0.90
EPHX1 H139R & NAT2*5C K268R	**Controls N= 260, n(%)**	**Cases N= 246, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (AA + AA)	191 (73.46)	190 (77.23)	1.00 (Reference)		
1 (AG + AG)	68 (26.15)	55 (22.36)	0.81 (0.54-1.22)	0.31	0.67
2 (GG + GG)	1 (0.38)	1 (0.41)	0.88 (0.05-14.31)	0.93	0.96
1-2	69 (26.54)	56 (22.77)	0.81 (0.54-1.22)	0.31	0.67
EPHX1 H139R & NAT2*7 G286E	**Controls N= 238, n(%)**	**Cases N=248 , n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (AA + GG)	177 (74.37)	162 (65.32)	1.00 (Reference)		
1 (AG + GA)	59 (24.79)	84 (33.87)	1.54 (1.04-2.29)	**0.03**	0.19
2 (GG + AA)	2 (0.84)	2 (0.81)	1.05 (0.14-7.80)	0.96	0.96
1-2	61 (25.63)	86 (53.08)	1.53 (1.03-2.26)	**0.03**	0.19

Abbreviations: 0- a combination of wild genotypes of any two polymorphism, 1- heterozygous genotype of two polymorphic variants in combination, 2: homozygous variant genotype of two polymorphic variants; 1–2: Heterozygous + mutant genotype of two polymorphic variants.

Adjusted Odds ratios, 95% confidence intervals, and corresponding p-values were calculated logistic regression analysis after adjusting for age, gender, and smoking status.

[b] Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. All p-values are two-sided.p <0.05 was considered statistically significant. "-" denotes the blanks where the number of subjects was nil.

5.11.3. Distribution of polymorphic phase II detoxification genes and their association with histological subtypes

The combinations of the different phase II genes were subsequently sorted based upon histological subtypes to identify high-risk posing subgroups, as shown in Table 5.82. Subjects diagnosed with ADCC for the group comprising both *NQO1* $P^{187}S$ and *NAT2*7* $G^{286}E$ had a significant difference in the distribution of heterozygote between cases and controls (57.14 *vs.* 43.64%). Logistic regression analysis revealed an approximately 2-fold increased risk of lung cancer development for subjects carrying both the heterozygous alleles (*CT+GA*) (AOR= 1.92; 95% C.I.=1.14-3.24; p=**0.01**; FDRp= 0.02). Combination of *SULT1A1* $R^{213}H$ and *NAT2*5B* $L^{161}L$, the patients who were heterozygous for both the SNPs, the frequency of heterozygotes was almost similar both in SCLC (29.69%) and SQCC (29.63%) subtypes, and this was less represented than the controls (41.36%). The subjects with heterozygous (*GA+CT*) (AOR= 0.49; 95% C.I.= 0.26-0.92; p= **0.025**; FDR p= 0.17) and combined genotype (*GA + AA + CT + TT*) (AOR= 0.46; 95% C.I.= 0.25-0.87; p= **0.017**; FDR p= 0.13) for SCLC subtype exhibited a protective effect towards lung cancer. Furthermore, we also observed that, in the case of *SULT1A1* $R^{213}H$ and *NAT2*5B* $L^{161}L$ combination, the lung cancer patients who were carrying a

single copy of the variant allele for both the polymorphisms (heterozygous; *GA+AG*) had a lower risk for developing SQCC (AOR= 0.45; 95% C.I.= 0.25-0.79; p= **0.006**; FDR **p=0.03**) and therefore was found to exhibit a protective effect towards lung cancer initiation. On the same lines, SQCC subtype for *SULT1A1* $R^{213}H$ and *NAT2*5C* $K^{268}R$ together revealed a protective effect for lung cancer patients carrying the heterozygous (GA + AG) (AOR= 0.44; 95% C.I.= 0.24-0.82; p= **0.009**; FDR p= 0.04) and combined genotype (GA + AA + AG + GG) (AOR= 0.42; 95% C.I.= 0.23-0.76; p= **0.005**; FDR p= 0.03).

On the contrary, when genotypic variants *SULT1A1* $R^{213}H$ and *NAT2*7* $G^{286}E$ were analyzed in combination, the subjects carrying (*GA+GA*) genotypes for both SNPs revealed a trend of increased risk towards ADCC subtype (AOR= 1.74; 95% C.I.= 1.02-2.97; p= **0.04**; FDR p= 0.22). Similarly, the combined variant genotype also revealed a trend of increased risk towards ADCC subtype (AOR= 1.70; 95% C.I.= 1.0-2.87; p= **0.049**; FDR p= 0.26) as shown in Table 5.82.

Among the other different genotypic combinations analyzed, when the group *EPHX1 Y113H* and *NAT2*6* $R^{197}Q$ were combined as a single group, the patients with single allelic variant genotype (*TC + GA*) showed a 2-fold risk of having SQCC (AOR= 1.91; 95% C.I.= 1.10-3.32; p = **0.02**; FDRp = **0.08**), which is a subtype of Lung cancer. Further, *EPHX1* $Y^{113}H$ in combination with *NAT2*7* $G^{286}E$ polymorphism suggested an association towards modulating the risk for ADCC development, especially in subjects who were heterozygotes (TC+GA) for both above SNPs. Logistic regression analysis revealed that such subjects had a high propensity towards risk for lung cancer (AOR= 2.41; 95% C.I.= 1.33-4.38; p= **0.004**; FDRp = **0.04**) especially ADCC. Our data also shows that patients diagnosed with SQCC and carrying one allele of variant genotype for both SNPs (*TC+GA*) revealed a high risk (AOR= 2.37; 95% C.I.= 1.27-4.41; p= **0.006**; FDRp= **0.03**). Also, the combined heterozygous and variant genotype as a single genotype (*TC + CC + GA + AA*) possessed a 2-fold risk towards lung cancer (AOR= 2.36; 95% C.I.= 1.28-4.35; p= **0.006**; FDRp= **0.03**) development. On the same lines, ADCC subtype for *EPHX1* $H^{139}R$ and *NAT2*7* $G^{286}E$ together revealed a high risk for lung cancer patients carrying one copy of variant allele (*AG + GA*) (AOR=1.85; 95% C.I.=1.08-3.16; p= **0.02**; FDRp= 0.14) as well as heterozygous + mutant genotype (*AG + GG + GA + AA*) (AOR= 1.83; 95% C.I.= 1.07-3.10; p= **0.03**; FDRp= 0.19) (Table 5.82).

Table 5.82: Distribution of genotypes depicting dual combination and their association with histological subtypes of lung cancer

NQO1 P187S & SULT1A1 R213H	**Controls N= 256, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 59, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 86, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 102 , n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +GG)	172 (67.19)	44 (74.58)	1.00 (Reference)			58 (67.44)	1.00 (Reference)			75 (73.53)	1.00 (Reference)		
1 (CT + GA)	80 (31.25)	15 (25.42)	0.64 (0.31-1.31)	0.22	0.69	26 (30.23)	1.04 (0.60-1.81)	0.88	0.97	25 (24.51)	0.67 (0.39-1.16)	0.15	0.45
2 (TT + AA)	4 (1.56)	-	-	-	-	2 (2.33)	1.25 (0.21-7.47)	0.81	0.97	2 (1.96)	1.01 (0.16-6.26)	0.99	0.99
1-2	84 (32.81)	15 (25.42)	0.61 (0.30-1.25)	0.18	0.69	28 (32.56)	1.06 (0.62-1.82)	0.83	0.97	27 (26.47)	0.69 (0.40-1.18)	0.17	0.46
NQO1 P187S & EPHX1 Y113H	**Controls N= 209, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 53 , n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 78, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 81, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC + TT)	103 (49.28)	21 (39.62)	1.00 (Reference)			37 (47.44)	1.00 (Reference)			35 (43.21)	1.00(Reference)		
1 (CT + TC)	99 (47.37)	31 (58.49)	1.56 (0.82-2.97)	0.18	0.69	39 (50.00)	1.11 (0.63-1.93)	0.72	0.97	42 (51.85)	1.30 (0.76-2.24)	0.33	0.59
2 (TT + CC)	7 (3.35)	1 (1.89)	0.83 (0.09-7.28)	0.87	0.96	2 (2.56)	0.70 (0.12-3.93)	0.68	0.97	4 (4.94)	1.93 (0.50-7.49)	0.34	0.59
1-2	106 (50.72)	32 (60.38)	1.52 (0.80-2.87)	0.20	0.69	41 (52.56)	1.07 (0.62-1.85)	0.80	0.97	46 (56.79)	1.35 (0.80-2.29)	0.26	0.52
NQO1 P187S & EPHX1 H139R	**Controls N= 259, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 59, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 89 , n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 99, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +AA)	192 (74.13)	43 (72.88)	1.00 (Reference)			60 (67.42)	1.00 (Reference)			74 (74.75)	1.00(Reference)		
1 (CT + AG)	65 (25.10)	15 (25.43)	1.17 (0.59-2.33)	0.65	0.87	29 (32.58)	1.39 (0.81-2.38)	0.23	0.80	25 (25.25)	1.05 (0.60-1.82)	0.87	0.93
2 (TT + GG)	2 (0.77)	1 (1.69)	4.12(0.26-65.92)	0.32	0.84	-	-	-	-	-	-	-	-
1-2	67 (25.87)	16 (27.12)	1.23 (0.63-2.42)	0.55	0.87	29 (32.58)	1.34 (0.79-2.29)	0.28	0.84	25 (25.25)	1.03 (0.59-1.78)	0.93	0.97
NQO1 P187S & NAT2*5B L161L	**Controls N= 219, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 50 , n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 84 , n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 86, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +CC)	97 (44.29)	28 (56.00)	1.00 (Reference)			37 (44.05)	1.00 (Reference)			44 (51.16)	1.00(Reference)		
1 (CT + CT)	116 (52.97)	20 (40.00)	0.54 (0.27-1.06)	0.07	0.44	44 (52.38)	0.98 (0.58-1.66)	0.95	0.97	42 (48.84)	0.71 (0.42-1.20)	0.20	0.48
2 (TT + TT)	6 (2.74)	2 (4.00)	1.53 (0.27-8.64)	0.63	0.87	3 (3.57)	1.28 (0.30-5.45)	0.74	0.97	-	-	-	-
1-2	122 (55.71)	22 (44.00)	0.59 (0.30-1.13)	0.11	0.53	47 (55.95)	0.99 (0.59-1.66)	0.96	0.97	42 (48.84)	0.68 (0.41-1.15)	0.15	0.45
NQO1 P187S & NAT2*6 R197Q	**Controls N= 216, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N=54 ,	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-**	**Cases N= 88, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-**	**Cases N= 102, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-**

		n(%)			value				value				value
0 (CC +GG)	109 (50.46)	27 (50.00)	1.00 (Reference)			41 (46.59)	1.00 (Reference)			47 (46.08)	1.00(Reference)		
1 (CT + GA)	100 (46.30)	26 (48.15)	1.25 (0.65-2.40)	0.51	0.87	44 (50.00)	1.15 (0.69-1.93)	0.59	0.97	50 (49.02)	1.30 (0.78-2.16)	0.31	0.57
2 (TT + AA)	7 (3.24)	1 (1.85)	0.86 (0.09-7.82)	0.89	0.96	3 (3.41)	1.10 (0.26-4.69)	0.90	0.97	5 (4.90)	2.33 (0.62-8.83)	0.21	0.49
1-2	107 (49.54)	27 (50.00)	1.21 (0.64-2.31)	0.56	0.87	47 (53.41)	1.15 (0.69-1.92)	0.59	0.97	55 (53.92)	1.33 (0.81-2.18)	0.26	0.53
NQO1 P187S & NAT2*5C K268R	**Controls N= 242, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 48, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 73, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 89, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +AA)	136 (56.20)	28 (58.33)	1.00 (Reference)			40 (54.79)	1.00 (Reference)			56 (62.92)	1.00(Reference)		
1 (CT + AG)	100 (41.32)	18 (37.50)	0.82 (0.41-1.63)	0.57	0.87	29 (39.73)	1.02 (0.59-1.77)	0.93	0.97	32 (35.96)	0.76 (0.45-1.26)	0.29	0.55
2 (TT + GG)	6 (2.48)	2 (4.17)	1.93 (0.35-10.67)	0.45	0.87	4 (5.48)	2.43 (0.64-9.16)	0.19	0.80	1 (1.12)	0.35 (0.04-3.10)	0.35	0.59
1-2	106 (43.80)	20 (41.67)	0.88 (0.45-1.71)	0.71	0.92	33 (45.21)	1.10 (0.65-1.89)	0.72	0.97	33 (37.08)	0.73 (0.44-1.22)	0.23	0.50
NQO1 P187S & NAT2*7 G286E	**Controls N= 236, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 49, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 84, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 95, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (CC +GG)	130 (55.08)	30 (61.22)	1.00 (Reference)			35 (41.67)	1.00 (Reference)			45 (47.37)	1.00(Reference)		
1 (CT + GA)	103 (43.64)	19 (38.78)	0.81 (0.41-1.61)	0.56	0.87	48 (57.14)	1.92 (1.14-3.24)	**0.01**	**0.08**	45 (47.37)	1.22 (0.74-2.01)	0.44	0.71
2 (TT + AA)	3 (1.27)	0 (0.00)	-	-	-	1 (1.19)	1.53 (0.15-15.73)	0.72	0.97	5 (5.26)	3.81(0.86-16.78)	0.07	0.26
1-2	106 (44.92)	19 (38.78)	0.79 (0.40-1.56)	0.49	0.87	49 (58.33)	1.92 (1.14-3.24)	**0.01**	**0.08**	50 (52.63)	1.31 (0.80-2.14)	0.28	0.55
SULT1A1 R213H & EPHX1 Y113H	**Controls N= 229, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 47, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 79, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 81, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ TT)	137 (59.83)	27 (57.45)	1.00 (Reference)			50 (63.29)	1.00 (Reference)			48 (59.26)	1.00(Reference)		
1 (GA + TC)	87 (37.99)	17 (36.17)	0.92 (0.46-1.86)	0.82	0.95	28 (35.44)	0.87 (0.49-1.54)	0.62	0.97	31 (38.27)	0.98 (0.57-1.67)	0.94	0.97
2 (AA +CC)	5 (2.18)	3 (6.38)	3.20 (0.71-14.39)	0.13	0.58	1 (1.27)	0.50 (0.05-4.82)	0.55	0.97	2 (2.47)	1.20 (0.22-6.63)	0.83	0.92
1-2	92 (40.17)	20 (42.55)	1.04 (0.53-2.04)	0.90	0.96	29 (36.71)	0.86 (0.49-1.51)	0.59	0.97	33 (40.74)	0.99 (0.59-1.67)	0.97	0.98
SULT1A1 R213H & EPHX1 H139R	**Controls N= 274, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 71, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 117, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 131, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ AA)	228 (83.21)	60 (84.51)	1.00 (Reference)			94 (80.34)	1.00 (Reference)			105(80.15)	1.00(Reference)		

1 (GA +AG)	44 (16.06)	11 (15.49)	1.02 (0.47-2.19)	0.97	0.97	23 (19.66)	1.25 (0.71-2.21)	0.44	0.97	26 (19.85)	1.21 (0.69-2.13)	0.51	0.80
2 (AA +GG)	2 (0.73)	-	-	-	-	-	-	-	-	-	-	-	-
1-2	46 (16.79)	11 (15.49)	0.98 (0.46-2.12)	0.97	0.97	23 (19.66)	1.19 (0.68-2.08)	0.55	0.97	26 (19.85)	1.17 (0.67-2.06)	0.57	0.85
SULT1A1 R213H & NAT2*5B L161L	**Controls N= 220, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 64, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 79, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 81, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ CC)	123 (55.91)	45 (70.31)	1.00 (Reference)			52 (65.82)	1.00 (Reference)			57 (70.37)	1.00(Reference)		
1 (GA + CT)	91 (41.36)	19 (29.69)	0.49 (0.26-0.92)	**0.025**	0.17	27 (34.18)	0.76 (0.44-1.32)	0.33	0.94	24 (29.63)	0.45 (0.25-0.79)	**0.006**	**0.03**
2 (AA + TT)	6 (2.73)	-	-	-	-	-	-	-	-	-	-	-	-
1-2	97 (44.09)	19 (29.69)	0.46 (0.25-0.87)	**0.017**	0.13	27 (34.18)	0.70 (0.41-1.21)	0.21	0.80	24 (29.63)	0.43 (0.24-0.76)	**0.003**	**0.02**
SULT1A1 R213H & NAT2*6 R197Q	**Controls N= 213, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 56, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 89, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 95, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ CC)	132 (61.97)	36 (64.29)	1.00 (Reference)			55 (61.80)	1.00 (Reference)			62 (65.26)	1.00(Reference)		
1 (GA + CT)	74 (34.74)	18 (32.14)	0.89 (0.44-1.78)	0.74	0.92	32 (35.95)	1.01 (0.59-1.73)	0.97	0.97	31 (32.63)	0.91 (0.54-1.54)	0.73	0.85
2 (AA + TT)	7 (3.29)	2 (3.57)	1.29 (0.25-6.69)	0.76	0.92	2 (2.25)	0.52 (0.10-2.78)	0.45	0.97	2 (2.11)	0.65 (0.13-3.35)	0.60	0.85
1-2	81 (38.03)	20 (35.71)	0.92 (0.47-1.80)	0.82	0.95	34 (38.20)	0.97 (0.58-1.64)	0.92	0.97	33 (34.74)	0.89 (0.53-1.49)	0.66	0.85
SULT1A1 R213H & NAT2*5C K268R	**Controls N= 247, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 62, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 87, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 85, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ AA)	163 (65.99)	45 (72.58)	1.00 (Reference)			62 (71.26)	1.00 (Reference)			68 (80.00)	1.00(Reference)		
1 (GA +AG)	77 (31.17)	17 (27.42)	0.69 (0.36-1.33)	0.27	0.77	23 (26.44)	0.84 (0.48-1.48)	0.55	0.97	17 (20.00)	0.44 (0.24-0.82)	**0.009**	**0.04**
2 (AA +GG)	7 (2.83)	-	-	-	-	2 (2.30)	0.70 (0.14-3.55)	0.67	0.97	-	-	-	-
1-2	84 (34.01)	17 (27.42)	0.65 (0.34-1.25)	0.20	0.69	25 (28.74)	0.83 (0.48-1.43)	0.50	0.97	17 (20.00)	0.42 (0.23-0.76)	**0.005**	**0.03**
SULT1A1 R213H & NAT2*7 G286E	**Controls N= 240, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 61, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 84, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 88, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (GG+ GG)	162 (67.50)	42 (68.85)	1.00 (Reference)			48 (57.14)	1.00 (Reference)			58 (65.91)	1.00(Reference)		

1 (GA +GA)	72 (30.00)	18 (29.51)	0.77 (0.39-1.52)	0.46	0.87	34 (40.48)	1.74 (1.02-2.97)	**0.04**	0.22	29 (32.95)	0.94 (0.55-1.62)	0.83	0.92
2 (AA +AA)	6 (2.50)	1 (1.64)	0.82 (0.09-7.41)	0.86	0.96	2 (2.38)	1.06 (0.20-5.58)	0.95	0.97	1 (1.14)	0.42 (0.05-3.68)	0.43	0.71
1-2	78 (32.50)	19 (31.15)	0.77 (0.40-1.50)	0.45	0.87	36 (42.86)	1.70 (1.00-2.87)	**0.04 9**	0.26	30 (34.09)	0.90 (0.53-1.54)	0.71	0.85
EPHX1 Y113H & NAT2*5B L161L	**Controls N= 199, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 53, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 79, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 84, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (TT + CC)	70 (35.18)	20 (37.74)	1.00 (Reference)			28 (35.44)	1.00 (Reference)			31 (36.90)	1.00(Reference)		
1 (TC + CT)	118 (59.29)	28 (52.83)	0.84 (0.42-1.66)	0.61	0.87	47 (59.50)	1.04 (0.58-1.85)	0.89	0.97	51 (60.72)	0.89 (0.52-1.55)	0.69	0.85
2 (CC + TT)	11 (5.53)	5 (9.43)	1.62 (0.48-5.40)	0.43	0.87	4 (5.06)	1.13 (0.31-4.07)	0.86	0.97	2 (2.38)	0.38 (0.07-1.88)	0.23	0.50
1-2	129 (64.82)	33 (62.26)	0.90 (0.47-1.74)	0.76	0.92	51 (64.56)	1.04 (0.58-1.83)	0.91	0.97	53 (63.10)	0.84 (0.49-1.45)	0.53	0.81
EPHX1 Y113H & NAT2*6 R197Q	**Controls N= 209, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 61, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 79, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 86, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (TT + GG)	88 (42.10)	22 (36.06)	1.00 (Reference)			31 (39.24)	1.00 (Reference)			28 (32.56)	1.00(Reference)		
1 (TC + GA)	107 (51.20)	36 (59.02)	1.48 (0.79-2.79)	0.22	0.69	41 (51.90)	0.98 (0.55-1.75)	0.96	0.97	55 (63.95)	1.91 (1.10-3.32)	**0.02**	**0.08**
2 (CC +AA)	14 (6.70)	3 (4.92)	0.96 (0.25-3.69)	0.95	0.97	7 (8.86)	1.53 (0.55-4.25)	0.41	0.97	3 (3.49)	0.74 (0.19-2.85)	0.66	0.85
1-2	121 (57.90)	39 (63.94)	1.43 (0.77-2.65)	0.26	0.77	48 (60.76)	1.07 (0.61-1.85)	0.81	0.97	58 (67.44)	1.76 (1.02-3.02)	**0.04**	0.16
EPHX1 Y113H & NAT2*5C K268R	**Controls N= 219, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 45, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 74, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 89, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (TT + AA)	102 (46.58)	22 (48.89)	1.00 (Reference)			35 (47.30)	1.00 (Reference)			42 (47.19)	1.00(Reference)		
1 (TC + AG)	107 (48.86)	20 (44.44)	0.85 (0.42-1.70)	0.64	0.87	33 (44.59)	0.92 (0.51-1.64)	0.77	0.97	41 (46.07)	0.93 (0.55-1.56)	0.77	0.88
2 (CC +GG)	10 (4.56)	3 (6.67)	1.45 (0.36-5.90)	0.60	0.87	6 (8.11)	1.61 (0.49-5.29)	0.43	0.97	6 (6.74)	1.29 (0.43-3.85)	0.65	0.85
1-2	117 (53.42)	23 (51.11)	0.90 (0.46-1.76)	0.76	0.92	39 (52.70)	0.98 (0.56-1.72)	0.94	0.97	47 (52.81)	0.96 (0.57-1.59)	0.87	0.93
EPHX1 Y113H & NAT2*7 G286E	**Controls N= 215, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 46, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 70, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 64, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (TT + GG)	110 (51.16)	19 (41.30)	1.00 (Reference)			22 (31.43)	1.00 (Reference)			19 (29.69)	1.00(Reference)		
1 (TC + GA)	98 (45.58)	25 (54.35)	1.37 (0.70-2.71)	0.36	0.87	45 (64.29)	2.41 (1.33-4.38)	**0.004**	**0.04**	41 (64.06)	2.37 (1.27-4.41)	**0.006**	**0.03**
2 (CC +AA)	7 (3.26)	2 (4.35)	1.55 (0.29-8.17)	0.61	0.87	3 (4.28)	2.66 (0.60-11.76)	0.20	0.80	4 (6.25)	2.81 (0.71-11.08)	0.14	0.45
1-2	105 (48.84)	27 (58.70)	1.39 (0.71-2.69)	0.34	0.86	48 (68.57)	2.41 (1.34-4.36)	**0.003**	**0.04**	45 (70.31)	2.36 (1.28-4.35)	**0.006**	**0.03**

EPHX1 H139R & NAT2*5B L161L	**Controls N= 217, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 54, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 78 , n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 96, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (AA +CC)	133 (61.29)	40 (74.07)	1.00 (Reference)			55 (70.51)	1.00 (Reference)			64 (66.67)	1.00(Reference)		
1 (AG + CT)	84 (38.71)	14 (25.93)	0.55 (0.27-1.13)	0.10	0.52	22 (28.21)	0.66 (0.37-1.18)	0.16	0.77	32 (33.33)	0.70 (0.42-1.19)	0.19	0.48
2 (GG + TT)	-	-	-	-	-	1 (1.28)	-	-	-	-	-	-	-
1-2	84 (38.71)	14 (25.93)	0.55 (0.27-1.13)	0.10	0.52	23 (29.49)	0.70 (0.40-1.23)	0.22	0.80	32 (33.33)	0.70 (0.42-1.19)	0.19	0.48
EPHX1 H139R & NAT2*6 R197Q	**Controls N= 243, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 50, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 75, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 96, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (AA +GG)	161 (66.26)	35 (70.00)	1.00 (Reference)			53 (70.67)	1.00 (Reference)			62 (64.58)	1.00(Reference)		
1 (AG +GA)	79 (32.51)	14 (28.00)	0.93 (0.46-1.87)	0.83	0.95	21 (28.00)	0.70 (0.39-1.28)	0.25	0.82	33 (34.38)	1.14 (0.68-1.91)	0.61	0.85
2 (GG +AA)	3 (1.23)	1 (2.00)	3.06 (0.26-36.50)	0.38	0.87	1 (1.33)	0.74 (0.07-7.79)	0.80	0.97	1 (1.04)	1.54 (0.13-18.52)	0.73	0.85
1-2	82 (33.74)	15 (30.00)	0.97 (0.49-1.94)	0.94	0.97	22 (29.33)	0.71 (0.39-1.28)	0.26	0.82	34 (35.42)	1.15 (0.69-1.92)	0.58	0.85
EPHX1 H139R & NAT2*5C K268R	**Controls N= 260, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 57, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 87, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 100, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (AA +AA)	191 (73.46)	44 (77.19)	1.00 (Reference)			68 (78.16)	1.00 (Reference)			76 (76.00)	1.00(Reference)		
1 (AG +AG)	68 (26.15)	13 (22.81)	0.84 (0.39-1.78)	0.65	0.87	18 (20.69)	0.75 (0.42-1.37)	0.35	0.96	24 (24.00)	0.89 (0.51-1.54)	0.68	0.85
2 (GG +GG)	1 (0.38)	-	-	-	-	1 (1.15)	3.28 (0.20-53.95)	0.40	0.97	-	-	-	-
1-2	69 (26.54)	13 (22.81)	0.82 (0.39-1.74)	0.61	0.87	19 (21.84)	0.79 (0.44-1.42)	0.43	0.97	24 (24.00)	0.88 (0.51-1.52)	0.64	0.85
EPHX1 H139R & NAT2*7 G286E	**Controls N= 238, n(%)**	**SCLC**				**ADCC**				**SQCC**			
		Cases N= 60, n(%)	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 88, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**	**Cases N= 99, n(%)**	**AOR (95% C.I.)[a]**	**p [b]**	**FDR p-value**
0 (AA +GG)	177 (74.37)	42 (70.00)	1.00 (Reference)			55 (62.50)	1.00 (Reference)			64 (64.65)	1.00(Reference)		
1 (AG +GA)	59 (24.79)	17 (28.33)	1.23 (0.62-2.45)	0.55	0.87	32 (36.36)	1.85 (1.08-3.16)	**0.02**	0.14	35 (35.35)	1.48 (0.88-2.50)	0.14	0.45
2 (GG +AA)	2 (0.84)	1 (1.67)	4.32 (0.26-71.16)	0.30	0.82	1 (1.14)	1.32 (0.11-15.73)	0.83	0.97	-	-	-	-
1-2	61 (25.63)	18 (30.00)	1.29 (0.66-2.54)	0.46	0.87	33 (37.50)	1.83 (1.07-3.10)	**0.03**	0.19	35 (35.35)	1.45 (0.86-2.44)	0.16	0.45

Abbreviations: 0- the combination of wild genotypes of any two polymorphism, 1- heterozygous genotype of two polymorphic variants in combination, 2: homozygous variant genotype of two polymorphic variants; 1–2: Heterozygous + mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

[a] Adjusted Odds ratios, 95% confidence intervals, and their corresponding p-values were calculated using logistic regression analysis after adjusting for age, gender, and smoking status.

[b] Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. All p-values are two-sided. $p < 0.05$ was considered statistically significant. "-" denotes the blanks where the number of subjects was nil.

5.11.4. Genotypic distribution of triple combinations and combination of four SNPs between Phase II detoxification genes (*NQO1, SULT1A1, EPHXI,* and *NAT2*) and their association with lung cancer predisposition

The present study also evaluated the triple polymorphic combinations, and the combination of four SNPs polymorphic combinations of 8 SNPs studied as depicted in Table 5.83 and Table 5.84, respectively. The combined genotypes (heterozygous + mutant genotype) were evaluated as a single group since the number of subjects with a mutant genotype was less or not present. This combined genotype was then compared with the subjects with a wild-type genotype (reference) for all the combinations studied.

As shown in Table 5.83, lung cancer patients who were carrying a single copy of the variant alleles for the *NQO1* $P^{187}S$, *EPHX1* $Y^{113}H$, and *EPHX1* $H^{139}R$ (*CT + TT + TC + CC + AG + GG*) polymorphisms expressed a 2-fold risk for lung cancer as compared to subjects who were harboring the wild genotype for the above three polymorphic variants (AOR= 2.20; 95% C.I.= 1.11-4.36; p= **0.02**; FDR p= **0.04**). Similarly, a higher risk was also observed for those patients who had the combined variant genotype (*CT + TT +TC +CC + GA + AA*) for *NQO1* $P^{187}S$, *EPHX1* $Y^{113}H$, and *NAT2*6* $R^{197}Q$ polymorphism (AOR= 2.13; 95% C.I.= 1.15-3.92; p= **0.016**; FDR p= **0.04**). As shown in Table 5.83, similar trends were visible for other triple combinations where subjects having a single copy of the variant alleles for *NQO1* $P^{187}S$, *EPHX1* $Y^{113}H$, and *NAT2*7* $G^{286}E$ polymorphisms depicted a two-fold risk of developing lung cancer (AOR= 2.00; 95% C.I.= 1.07-3.73; p= **0.03**; FDR p= **0.04**). Among the different triple combinations we studied, our data, as shown in Table 5.83, inferred that the lung cancer subjects harboring a single copy of the variant alleles for *EPHX1* $Y^{113}H$, *EPHX1* $H^{139}R$, and *NAT2*7* $G^{286}E$ polymorphisms expressed a 3-fold risk for the disease in patients as compared to subjects who were carrying the wild genotype for the above polymorphisms (AOR= 2.94; 95% C.I.= 1.46-5.89; p=**0.002**; FDRp= **0.016**). The other polymorphic combinations which suggested a significant association of heterozygous + mutant genotype (GA +AA + AG + GG + GA + AA) was *SULT1A1* $R^{213}H$, *EPHX1* $H^{139}R$, and *NAT2*7* $G^{286}E$ depicting a 2-fold risk for lung cancer (AOR= 2.17; 95% C.I.= 1.06-4.44; p = **0.03**; FDRp= **0.04**). On the contrary the genotypic combinations of *SULT1A1* $R^{213}H$, *NAT2*5B* $L^{161}L$, and *NAT2*5C* $K^{268}R$ showed a strong protective effect for lung cancer in subjects carrying heterozygous and mutant genotype as a

single genotype (GA + AA + AG + GG + AG + GG) (AOR= 0.58; 95% C.I.=0.35-0.95; p=**0.03**; FDRp= **0.04**).

Table 5.83: Distribution of genotypes depicting triple combination and their association with lung cancer risk

NQO1 P187S, EPHX1 Y113H, & EPHX1 H139R	**Controls N= 80, n(%)**	**Cases N= 89, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (CC + TT + AA)	62 (77.50)	54 (60.67)	1.00 (Reference)		
1-2 (CT + TT + TC + CC + AG + GG)	18 (22.50)	35 (39.33)	2.20 (1.11-4.36)	**0.02**	**0.04**
NQO1 P187S, EPHX1 Y113H, & NAT2*6 R197Q	**Controls N= 81, n(%)**	**Cases N= 101, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (CC + TT + GG)	42 (51.85)	36 (35.64)	1.00 (Reference)		
1-2 (CT + TT +TC +CC + GA + AA)	39 (48.15)	65 (64.36)	2.13 (1.15-3.92)	**0.016**	**0.04**
NQO1 P187S, EPHX1 Y113H, & NAT2*7 G286E	**Controls N= 93, n(%)**	**Cases N= 77, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (CC + TT + GG)	51 (54.84)	29 (37.66)	1.00 (Reference)		
1-2 (CT + TT +TC +CC + GA + AA)	42 (45.16)	48 (62.34)	2.00 (1.07-3.73)	**0.03**	**0.04**
SULT1A1 R213H, EPHX1 Y113H, &EPHX1 H139R	**Controls N= 100, n(%)**	**Cases N=111,n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (GG + TT + AA)	85 (85.00)	84 (75.68)	1.00 (Reference)		
1-2 (GA + AA + TC + CC + AG + GG)	15 (15.00)	27 (24.32)	1.79 (0.89-3.62)	0.10	0.11
SULT1A1 R213H, EPHX1 Y113H, & NAT2*7 G286E	**Controls N= 110, n(%)**	**Cases N=75 , n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (GG + TT + GG)	73 (66.36)	40 (53.33)	1.00 (Reference)		
1-2 (GA + AA + TC + CC + GA + AA)	37 (33.64)	35 (46.67)	1.63 (0.88-2.99)	0.12	0.12
SULT1A1 R213H, EPHX1 H139R, & NAT2*7 G286E	**Controls N=124 , n(%)**	**Cases N= 143, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (GG + AA + GG)	111 (89.52)	114 (79.72)	1.00 (Reference)		
1-2 (GA +AA + AG + GG + GA + AA)	13 (10.48)	29 (20.28)	2.17 (1.06-4.44)	**0.03**	**0.04**
SULT1A1 R213H, NAT2*5B L161L, & NAT2*5C K268R	**Controls N=143 , n(%)**	**Cases N= 168, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (GG + GG + GG)	90 (62.94)	123 (73.21)	1.00 (Reference)		
1-2 (GA + AA + AG + GG + AG + GG)	53 (37.06)	45 (26.79)	0.58 (0.35-0.95)	**0.03**	**0.04**
EPHX1 Y113H, EPHX1 H139R, & NAT2*7 G286E	**Controls N= 81, n(%)**	**Cases N=76 , n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (TT + AA + GG)	63 (77.78)	41 (53.95)	1.00 (Reference)		
1-2 (TC + CC + AG + GG + GA + AA)	18 (22.22)	35 (46.05)	2.94 (1.46-5.89)	**0.002**	**0.016**

Abbreviations: 0- the combination of wild genotypes of any two polymorphism; 1–2: Heterozygous + mutant genotype of two polymorphic variants. [a] Adjusted odds ratios, 95% confidence intervals, and corresponding p-values were calculated logistic regression analysis after adjusting for age, gender, and smoking status.

[b] Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. All p-values are two-sided. $p < 0.05$ was considered statistically significant.

Further, the interaction of four SNPs was also carried out, as shown in Table 5.84. The genotypic combination of *NQO1 $P^{187}S$, EPHX1 $Y^{113}H$, EPHX1 $H^{139}R$, and NAT2*7 $G^{286}E$* when considered as a single group depicted a 5 fold risk for lung cancer in subjects with heterozygous + mutant genotype (*CT + TT + TC + CC + AG + GG + GA + AA*) (AOR= 5.32; 95% C.I.=1.66-16.99; p=**0.005**; FDRp=**0.035**). None of any other quadruple SNP combinations of the phase II genes showed any association towards risk for lung cancer. A 2.7-fold increased lung cancer risk

was observed for the combined variant genotype (*CT +TT + TC + CC + GA + AA + GA + AA*) for the polymorphic group of *NQO1 $P^{187}S$, EPHX1 $Y^{113}H$, NAT2*6 $R^{197}Q$,* and *NAT2*7 $G^{286}E$* (AOR=2.78; 95% C.I.= 0.99-7.83; p=**0.05**; FDRp=**0.08**). Further, the other polymorphic combinations which suggested a significant association of heterozygous + mutant genotype was *SULT1A1 $R^{213}H$, EPHX1 $Y^{113}H$, EPHX1 $H^{139}R$,* and *NAT2*7 $G^{286}E$* depicting a 3.5-fold risk for lung cancer (AOR= 3.59; 95% C.I.= 1.08-11.94; p = **0.04**; FDRp = **0.08**) (Table 5.84).

Table 5.84: Genotypic distribution based on different genotypic combinations and their association with lung cancer risk

NQO1 P187S, EPHX1 Y113H, EPHX1 H139R,& NAT2*6 R197Q	**Controls N= 35, n(%)**	**Cases N= 39, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	**FDR p-value**
0 (CC + TT + AA + GG) *1-2 (CT + TT + TC + CC + AG + GG + GA + AA)*	28 (80.00) 7 (20.00)	20 (51.28) 19 (48.72)	1.00 (Reference) 3.61 (1.24-10.54)	**0.02**	**0.07**
NQO1 P187S, EPHX1 Y113H, EPHX1 H139R,& NAT2*7 G286E	**Controls N=35 , n(%)**	**Cases N=38,n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	
0 (CC + TT + AA + GG) *1-2 (CT + TT + TC + CC + AG + GG + GA + AA)*	29 (82.86) 6 (17.14)	20 (52.63) 18 (47.37)	1.00 (Reference) 5.32 (1.66-16.99)	**0.005**	**0.035**
NQO1 P187S, EPHX1 Y113H, EPHX1 H139R,& NAT2*5B L161L	**Controls N= 32, n(%)**	**Cases N= 22, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	
0 (CC + TT + AA + CC) *1-2 (CT + TT + TC + CC + AG + GG + CT + TT)*	22 (68.75) 10 (31.25)	21 (95.45) 1 (4.55)	1.00 (Reference) 2.14 (0.80-5.66)	0.13	0.13
NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, &NAT2*7 G286E	**Controls N= 37, n(%)**	**Cases N= 38, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	
0 (CC + TT + GG +GG) *1-2 (CT +TT + TC + CC + GA + AA + GA + AA)*	18 (48.65) 19 (51.35)	11 (28.95) 27 (71.05)	1.00 (Reference) 2.78 (0.99-7.83)	**0.05**	**0.08**
NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, & NAT2*5B L161L	**Controls N=36 , n(%)**	**Cases N=46,n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	
0 (CC + TT + GG + CC)	15 (41.67)	11 (23.91)	1.00 (Reference)		
1-2 (CT +TT + TC + CC + GA + AA + CT + TT)	21 (58.33)	35 (76.09)	2.55 (0.95-6.80)	0.06	0.08
SULT1A1 R213H, EPHX1 Y113H, EPHX1 H139R, & NAT2*7 G286E	**Controls N=53 , n(%)**	**Cases N= 38, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	
0 (GG + TT + AA + GG) *1-2 (GA + AA + TC + CC + AG + GG + GA + AA)*	48 (90.57) 5 (9.43)	27 (71.05) 11 (28.95)	1.00 (Reference) 3.59 (1.08-11.94)	**0.04**	**0.08**
EPHX1 Y113H, EPHX1 H139R, NAT2*6 R197Q, & NAT2*7 G286E	**Controls N= 37, n(%)**	**Cases N= 36, n(%)**	**AOR (95% C.I.)[a]**	**p[b]**	
0 (TT + AA + GG + GG) *1-2 (GA + AA + TC + CC + GA + AA + GA + AA)*	26 (70.27) 11 (29.73)	18 (50.00) 18 (50.00)	1.00 (Reference) 2.19 (0.81-5.89)	0.12	0.13

Abbreviations: 0- the combination of wild genotypes of any two polymorphism; 1–2: Heterozygous + mutant genotype of two polymorphic variants.

[a] Adjusted odds ratios, 95% confidence intervals, and corresponding p-values were calculated logistic regression analysis after adjusting for age, gender, and smoking status.

[b] Two-sided $\chi 2$ test for either genotype distribution or allelic frequencies between the cases and controls. All p-values are two-sided. $p < 0.05$ was considered statistically significant.

5.11.5. GMDR analysis and risk of lung cancer

This study, employed the GMDR model, a data mining technique, to identify any potential synergistic effects among the various SNPs. This approach helps to enhance and validate the results to some extent, and it also assists in statistically overcoming the small sample size since GMDR has no dimension limitations in the interaction analysis. The GMDR approach was used to analyze the gene-gene interactions for lung cancer to identify which characteristics were the best model for predicting high-risk subgroups. A total of eight SNPs were included in the study. Among the best multifactor models, a higher CVC and a minimum prediction error are preferred for the best results.

As shown in figure 5.34, the entropy dendrogram has been utilized to determine the visible interaction between the genes involved and the risk of lung cancer. The length of the lines linking two risk factors in the dendrogram determines their interaction. The longer the distance between two risk factors, the weaker the interaction. The color of the lines connecting the two polymorphisms depicts the degree of interaction. Red and orange lines depict synergy or a non-additive relationship between the two SNPs, and the yellow line indicates independence or additivity, the brown line indicates weak interaction and green and blue lines denote loss of information, which can be interpreted as redundancy or correlation (for example, linkage disequilibrium).

Table 5.85 displays the CVC, prediction error, and p-value calculated by the GMDR software for each factor evaluated (N = 1 to 8). As a result, in the one-factor model, it was deduced that *NAT2*5B $L^{161}L$* was the best one-factor model with CVC 8/10 and prediction error of 0.47 among all the eight factors, as it emerged as the best attribute in lung cancer risk prediction (**p=0.0008**). In our study, the five-factor model *(NQO1 $P^{187}S$, EPHX1 $Y^{113}H$, NAT2*6 $R^{197}Q$, NAT2*5B $L^{161}L$, NAT2*5C $K^{268}R$)* with a CVC of 10/10 and prediction error of 0.4345 and p-value **<0.0001** was the overall best model for the interaction of selected Phase II polymorphisms. This indicates that this model occurs once in 1000 permutations, implausible under the null association hypothesis. While *NQO1 $P^{187}S$, EPHX1 $Y^{113}H$, NAT2*6 $R^{197}Q$, NAT2*7 $G^{286}E$, NAT2*5B $L^{161}L$, NAT2*5C $K^{268}R$* was the best six-factor model, with a maximum CVC of 10 and prediction error of 0.47 with p-value **<0.0001**. The dendrogram interaction analysis of the data revealed that, for lung cancer, X6 and X7 are on the same branch (green)

bifurcating into blue hue, indicating a redundant effect or association with disease phenotype shown in figure 5.34.

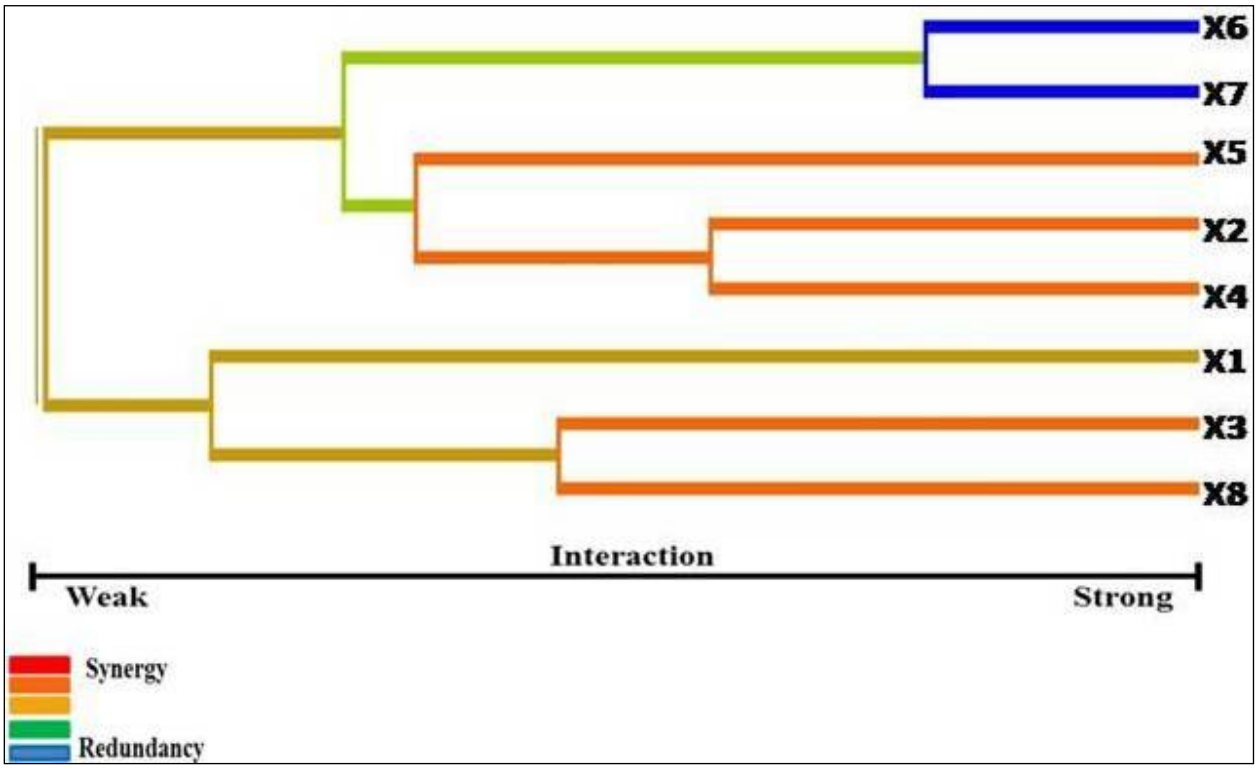


Figure 5.34: GMDR interaction dendrogram between Phase II detoxification genes (X1: *NQO1*, X2: *SULT1A1*, X3: *EPHX1Tyr*^{113}*His*, X4: *EPHX1 His*^{139}*Arg*, X5: *NAT2*6*, X6: *NAT2*7*, X7: *NAT2*5B*, X8: *NAT2*5C*)

Table 5.85: Multifactor Dimensionality Reduction (MDR) analysis showing interactions of Phase II detoxification variants with lung cancer risk

No. of risk factors	Best interaction model	CV consistency	Prediction error	P
1	*NAT2*5B L161L*	8/10	0.4736	**0.0008**
2	*NAT2*5B L161L, NAT2*5C K268R*	8/10	0.5027	**<0.0001**
3	*NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	6/10	0.4645	**<0.0001**
4	*EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L*	3/10	0.5182	**<0.0001**
5	*NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.4345	**<0.0001**
6	*NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.4745	**<0.0001**
7	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	7/10	0.5027	**<0.0001**
8	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, EPHX1 H139R, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.5073	**<0.0001**

Figure 5.35 shows the interaction map of the polymorphisms studied based on entropy measures among individual variables. A strong interaction effect was observed between X5 & X4 (*NAT2*6* and *EPHX1 His*^{139}*Arg*) polymorphisms, followed by X8 & X3 (*NAT2*5C&EPHX1 Tyr*^{113}*His*), and X8 & X1 (*NAT2*5C* and *NQO1*) and X6 & X3 (*NAT2*7&EPHX1 Tyr*^{113}*His*) polymorphisms, with information gain values of 0.19%. 0.19%, 0.17%, and 0.17%, respectively.

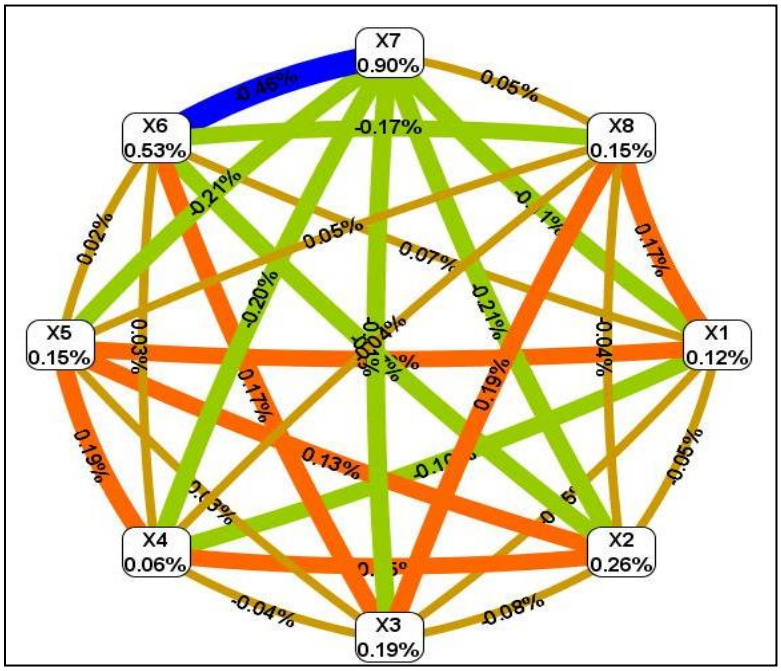


Figure 5.35: Entropy interaction map of lung cancer risk between Phase II detoxification genes. (*X1: NQO1, X2: SULT1A1, X3: EPHX1Tyr113His, X4: EPHX1 His139Arg, X5: NAT2*6, X6: NAT2*7, X7: NAT2*5B, X8: NAT2*5C*)

Figure 5.36 shows the epistasis interactions between lung cancer and healthy subjects between *NQO1 P*^{187}*S, EPHX1 Y*^{113}*H,* and *NAT2*6 R*^{197}*Q* polymorphisms. In three-factor model the bar having light grey in each of the cells shows cases of individuals having lung cancer, and the white bar represents the frequency of healthy individuals. The high-risk genotype combinations are shown as dark grey cells, whereas the white light grey cells represent genotype combinations of the low-risk genotypes. Cells with no shading or white cells represent genotype combinations for which no data is observed.

Figure 5.36 shows the distribution of high-risk and low-risk genotypes in the best three-locus model. The distribution shows high-risk (dark shading) and low-risk (light shading) genotypes associated with lung cancer in the three-locus interaction detected by MDR analysis. The percentage of lung cancer subjects (left black bar in boxes) and control subjects (right hatched bar in boxes) is shown for each three-locus genotype combination. Boxes were labeled

as high-risk if ratio of the percentage of cases to controls met or exceeded the threshold of 1.0. Boxes were labeled as low-risk if the threshold was not exceeded. Based on the pattern of high-risk and low-risk genotypes, this three-locus model is evidence of gene-to-gene interaction.

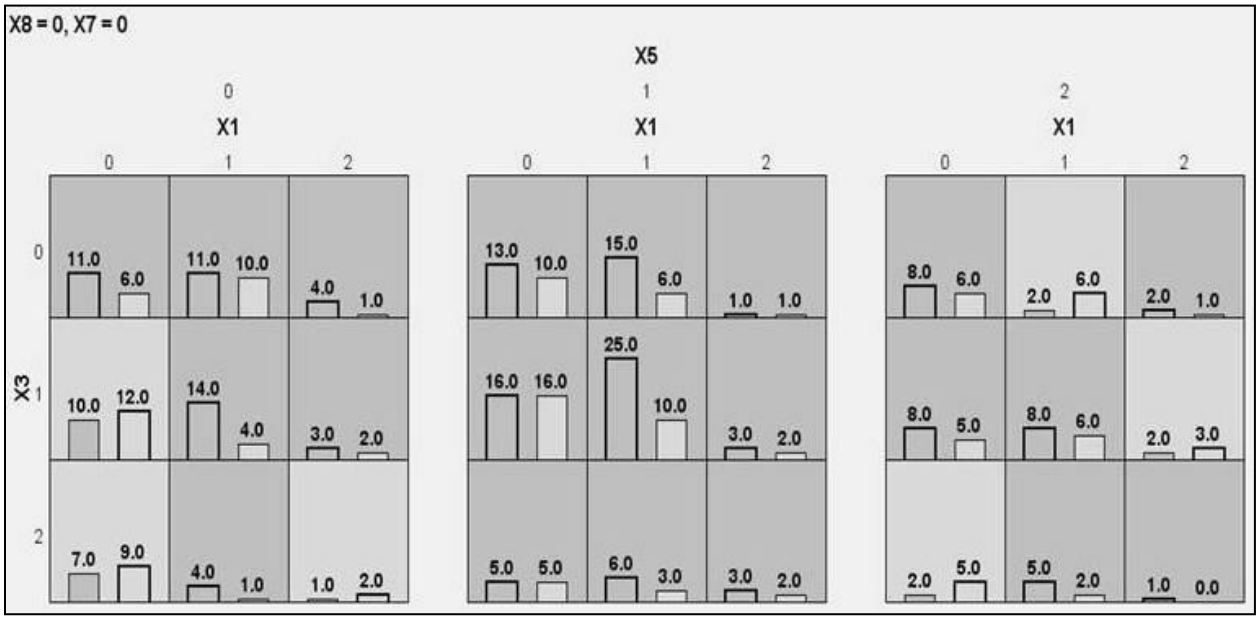


Figure 5.36: Distribution of high-risk and low-risk genotypes in the best three-locus model. (*X1: NQO1, X3: EPHX1Tyr113His, X5: NAT2*6, X7: NAT2*5B, X8: NAT2*5C)*

5.11.6. GMDR and association with histological subtypes of lung cancer

Furthermore, the study subjects were comprehended for the best interaction model posing a higher risk of lung cancer based on histology. As shown in Table 5.86, the *NAT2*5B* $L^{161}L$ was the highest risk posing single-factor model for SCLC patients, with a CVC of 10 and minimum prediction error of 0.43 (p= **0.0021**).

In the case of SCLC, the dendrogram demonstrates that *NQO1* $P^{187}S$ and *EPHX1* $Y^{113}H$ (X1 & X3) show the maximum synergy between them as they had the shortest line connecting them. The blue line that connects *SULTIA1* $R^{213}H$ and *NAT2*5B* $L^{161}L$ are on the same branch (orange) that bifurcates into blue color, indicating a redundant effect or association with disease phenotype (Figure 5.37 b)).

Table 5.86: Multifactor Dimensionality Reduction (MDR) analysis showing interactions of Phase II detoxification variants with SCLC risk

No. of risk factors	Best interaction model	CV consistency	Prediction error	p
1	*NAT2*5B L161L*	10/10	0.4271	**0.0021**
2	*NQO1 P187S, NAT2*5B L161L*	7/10	0.4496	**0.0002**
3	*NQO1 P187S, EPHX1 Y113H, NAT2*5B L161L*	7/10	0.4753	**<0.0001**
4	*NQO1 P187S, EPHX1 Y113H, NAT2*5B L161L, NAT2*5C K268R*	7/10	0.4592	**<0.0001**
5	*NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, NAT2*5B L161L, NAT2*5C K268R*	8/10	0.4068	**<0.0001**
6	*NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.4502	**<0.0001**
7	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.4596	**<0.0001**
8	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, EPHX1 H139R, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.4696	**<0.0001**

As shown in Table 5.87, the *NAT2*7 $G^{286}E$* was the highest risk posing single factor model among ADCC subjects, with a CVC of 10 and minimum prediction error of 0.46 with a p-value of **0.005**. The six-factor model (*NQO1 $P^{187}S$, EPHX1 $Y^{113}H$, NAT2*6 $R^{197}Q$, NAT2*7 $G^{286}E$, NAT2*5B $L^{161}L2$,NAT2*5C $K^{268}R$*) with a CVC of 10/10 and prediction error of 0.4502, 0.4611 and 0.4386 was the best interacting model for SCLC , ADCC (Table 5.87) and SQCC (Table 5.88) (p **<0.0001**) respectively, thus suggesting that contribution to the risk of the three histology's of lung cancer was due to the joint action of the six SNPs located in the three genes. The dendrogram analysis for ADCC subjects reveals synergy between *EPHX1 $Y^{113}H$* and *NAT2*5B $L^{161}L$* (X4 & X7) since they have the shortest line connecting them. *SULTIA1 $R^{213}H$, NQO1 $P^{187}S$,* and *NAT2*5C $K^{268}R$* are all on the same line, showing a significant interaction (Figure 5.37 a)).

Table 5.87: Multifactor Dimensionality Reduction (MDR) analysis showing interactions of Phase II detoxification variants with ADCC risk

No. of risk factors	Best interaction model	CV consistency	Prediction error	p
1	*NAT2*7 G286E*	10/10	0.4593	**0.0051**
2	*NQO1 P187S, NAT2*7 G286E*	3/10	0.5	**0.0009**
3	*EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E*	5/10	0.4934	**<0.0001**
4	*NQO1 P187S, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5C K268R*	3/10	0.5327	**<0.0001**
5	*NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L*	9/10	0.4827	**<0.0001**
6	*NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.4611	**<0.0001**
7	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, EPHX1 H139R, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L*	9/10	0.49	**<0.0001**
8	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, EPHX1 H139R, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.4893	**<0.0001**

Table 5.88: Multifactor Dimensionality Reduction (MDR) analysis showing interactions of Phase II detoxification variants with SQCC risk

No. of risk factors	Best interaction model	CV consistency	Prediction error	p
1	*NAT2*7 G286E*	6/10	0.4893	**0.0127**
2	*EPHX1 Y113H, NAT2*7 G286E*	6/10	0.5079	**<0.0001**
3	*EPHX1 Y113H, NAT2*7 G286E, NAT2*5B L161L*	4/10	0.4687	**<0.0001**
4	*EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L*	5/10	0.4597	**<0.0001**
5	*SULT1A1 R213H, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L*	5/10	0.479	**<0.0001**
6	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L*	10/10	0.4386	**<0.0001**
7	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.4782	**<0.0001**
8	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, EPHX1 H139R, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.5006	**<0.0001**

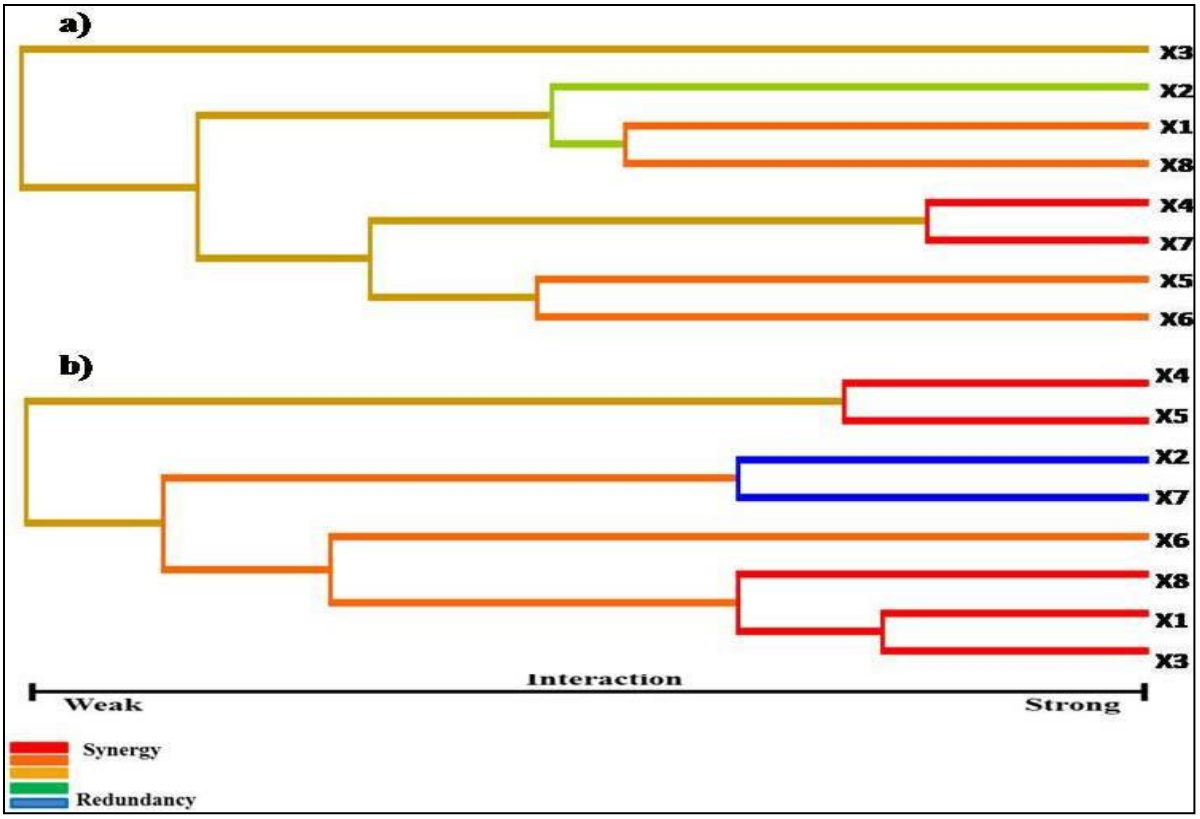


Figure 5.37: GMDR interaction dendrogram between Phase II detoxification genes (a) in ADCC subjects, (b) in SCLC subjects *(X1: NQO1, X2: SULT1A1, X3: EPHX1Tyr113His, X4: EPHX1 His139Arg, X5: NAT2*6, X6: NAT2*7, X7: NAT2*5B, X8: NAT2*5C)*

5.11.7. GMDR analysis and role of gene-environment interaction in lung cancer risk

In this study, GMDR analysis was employed to evaluate Phase II detoxifying gene and smoking relationships. The findings of GMDR models are shown in Table 5.89 and 5.90. The best interaction models for predicting lung cancer risk in smokers and nonsmokers were also examined. As shown in Table 5.89, *NAT2*5B* $L^{161}L$ was the best model, with a CVC of 10 and a prediction error of 0.43 (p<**0.0001**), demonstrating a possible gene-environment interaction between *NAT2*5B* $L^{161}L$ and smoking. The second best interaction model among smokers was a five-factor model that included *NQO1* $P^{187}S$, *EPHX1* $Y^{113}H$, *NAT2*6* $R^{197}Q$, *NAT2*7* $G^{286}E$, *NAT2*5B* $L^{161}L$, with CVC of 10 and prediction error of 0.45 (p<**0.0001**) (Table 5.90). *NAT2*7* $G^{286}E$, on the other hand, was the risk posing a single-factor model for non-smokers, yielding a maximum CVC of 10 and prediction error of 0.41. However, the two-factor model, consisting of *NAT2*7* $G^{286}E$, *NAT2*5C* $K^{268}R$, was the best interaction model, with a maximum CVC of 10 and minimum prediction error of 0.40 (**p=0.0001**). In the case of the smokers, the dendrogram

shows a powerful interaction and synergy between *SULTIA1* $R^{213}H$ and *EPHX1* $H^{139}R$ (X1 & X4) as they are joined together by the red line, but the blue line linking *NAT2*7* $G^{286}E$ and *NAT2*5B* $L^{161}L$(X6 &X7) indicates redundancy or no interaction (Figure 5.38).

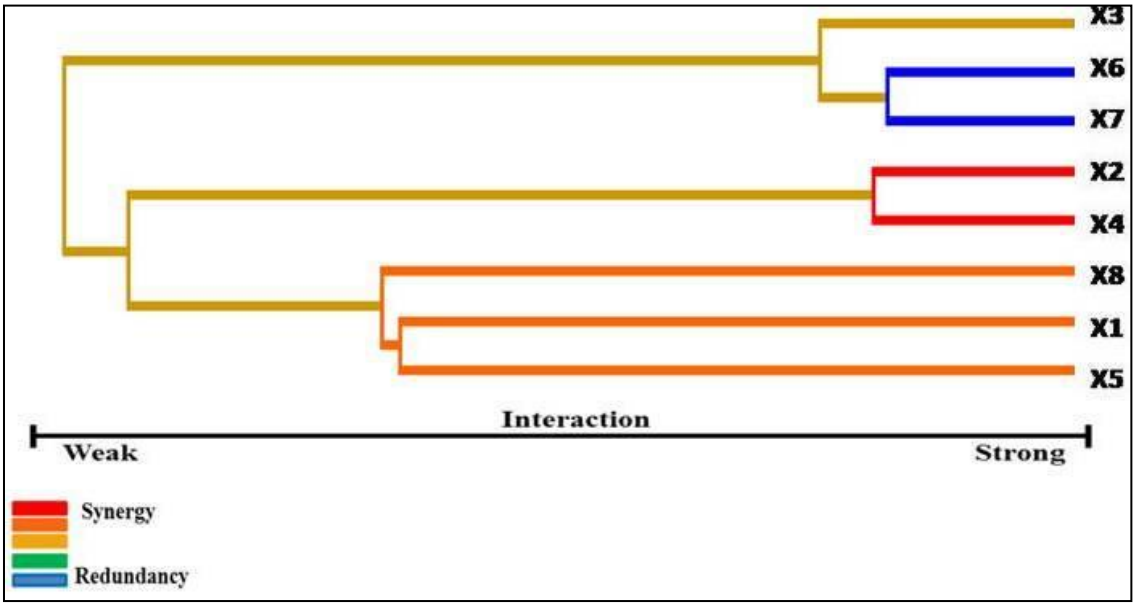


Figure 5.38: GMDR interaction dendrogram between Phase II detoxification genes in smokers (*X1: NQO1, X2: SULT1A1, X3: EPHX1Tyr113His, X4: EPHX1 His139Arg, X5: NAT2*6, X6: NAT2*7, X7: NAT2*5B, X8: NAT2*5C)*

Table 5.89: Multifactor Dimensionality Reduction (MDR) analysis showing gene-environment interactions (smokers) of Phase II detoxification variants with lung cancer risk

No. of risk factors	Best interaction model	CV consistency	Prediction error	P
1	*NAT2*5B L161L*	10/10	0.4298	**<0.0001**
2	*NAT2*5B L161L, NAT2*5C K268R*	7/10	0.4807	**<0.0001**
3	*EPHX1 Y113H, NAT2*7 G286E, BAT2*5B L161L*	7/10	0.4569	**<0.0001**
4	*EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, BAT2*5B L161L*	9/10	0.4261	**<0.0001**
5	*NQO1 P187S, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, BAT2*5B L161L*	10/10	0.4477	**<0.0001**
6	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, BAT2*5B L161L*	5/10	0.4904	**<0.0001**
7	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, BAT2*5B L161L, NAT2*5C K268R*	9/10	0.5138	**<0.0001**
8	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, EPHX1 H139R, NAT2*6 R197Q, NAT2*7 G286E, BAT2*5B L161L, NAT2*5C K268R*	10/10	0.5087	**<0.0001**

Table 5.90: Multifactor Dimensionality Reduction (MDR) analysis showing gene-environment interactions (non-smokers) of Phase II detoxification variants with lung cancer risk

No. of risk factors	Best interaction model	CV consistency	Prediction error	P
1	*NAT2*7 G286E*	10/10	0.4127	**0.0081**
2	*NAT2*7 G286E, NAT2*5C K268R*	10/10	0.4009	**0.0001**
3	*NQO1 P187S, NAT2*7 G286E, NAT2*5B L161L*	3/10	0.536	**<0.0001**
4	*NQO1 P187S, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5C K268R*	5/10	0.5053	**<0.0001**
5	*NQO1 P187S, SULT1A1 R213H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5C K268R*	8/10	0.4668	**<0.0001**
6	*NQO1 P187S, SULT1A1 R213H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	5/10	0.4768	**<0.0001**
7	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	5/10	0.5165	**<0.0001**
8	*NQO1 P187S, SULT1A1 R213H, EPHX1 Y113H, EPHX1 H139R, NAT2*6 R197Q, NAT2*7 G286E, NAT2*5B L161L, NAT2*5C K268R*	10/10	0.5108	**<0.0001**

5.11.8. CART analysis of Phase II detoxification genes

CART analysis is a tree-building technique that uses binary recursive partitioning to identify the different genotypic groups influencing lung cancer risk, which the logistic regression analysis cannot. A vast number of variables are considered, and high-risk categories are found. The analysis yielded a decision tree as a result of the data mining approach. The decision tree displayed many elements and their interactions and also the risk associated with the combinations. The root split is the most critical aspect.

CART analysis was performed on the genes involved in the DNA repair process. The tree demonstrates that *NAT481C>T* caused the initial split. As seen in Table 5.91, the SNP is the most important interaction element. Table 5.91 displays the odds ratio and p-values of interaction subgroups. Individuals carrying the genotypic combination of *NAT803A>G(M)/NAT857G>A(M)/EPHX1337T>C(W)/NQO1609C>T(M)/EPHX1415A>G (W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M) i.e.* Node 33 conferred the highest risk towards lung cancer (OR=24.35, 95% C.I..=2.94-201.26, p= **0.003**). Terminal node 26 with the genetic combination of *EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT803A>G(W)/ EPHX1337T>C(W)/NAT857G>A(W)/NAT481C>T(W)* also possessed a significantly high risk towards lung cancer (OR=14.61,95% C.I..=1.66-128.19, p= **0.01**). Subjects with genotypic combination of node 10 carrying *NAT803A>G(M)/NAT590G>A(W)/EPHX1337T>C(M)/ NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)* exhibited an odds of twelve folds (OR=12.17, 95% C.I..=1.35-110.00, p= **0.03**).

Table 5.91: Overall risk estimates based on CART analysis of Phase II detoxification genes

Terminal Node	*Genotype for each terminal node*	Control (n)	Case (n)	Case rate	OR (95% C.I.)	p-value
Node 1	*NAT590G>A(M)/NQO1609C>T(M)/NAT857G>A(M)/NAT481C>T(W)*	20	49	71.00	5.96 (2.93-12.15)	**0.000001**
Node 2	*NAT803A>G(W)/NAT590G>A(W)/NQO1609C>T(M)/NAT857G>A(M)/NAT481C>T(W)*	10	22	68.80	5.36 (2.20-13.06)	**0.0002**
Node 3	*NAT803A>G(M)/NAT590G>A(W)/NQO1609C>T(M)/NAT857G>A(M)/NAT481C>T(W)*	5	4	44.40	1.95 (0.48-7.91)	0.35
	NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(W)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)	0	8	100.00	-	
Node 5	*SULT1A1638G>A(M)/NAT803A>G(W)/NAT590G>A(M)/EPHX1337T>C(W)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	2	6	75.00	7.30 (1.37-38.89)	**0.02**
Node 6	*EPHX1415A>G(M)/SULT1A1638G>A(W)/NAT803A>G(W)/NAT590G>A(M)/EPHX1337T>C(W)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	2	3	60.00	3.65 (0.57-23.32)	0.17
Node 7	*EPHX1415A>G(W)/SULT1A1638G>A(W)/NAT803A>G(W)/NAT590G>A(M)/EPHX1337T>C(W)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	4	3	42.9	1.83 (0.38-8.81)	0.45
Node 8	*NAT803A>G(W)/NAT590G>A(W)/EPHX1337T>C(W)/NQO609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	5	8	61.50	3.89 (1.15-13.17)	**0.03**
Node 9	*NAT803A>G(M)/NAT590G>A(W)/EPHX1337T>C(W)/NQO609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	5	0	0.00	-	
Node 10	*NAT803A>G(M)/NAT590G>A(W)/EPHX1337T>C(M)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	1	5	83.30	12.17 (1.35-110.00)	**0.03**
Node 11	*EPHX1415A>G(W)/NAT803A>G(W)/NAT590G>A(W)/EPHX1337T>C(M)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	4	6	60.00	3.65 (0.94-14.16)	0.06
Node 12	*EPHX1415A>G(M)/NAT803A>G(W)/NAT590G>A(W)/EPHX1337T>C(M)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	4	2	33.30	1.22 (0.21-7.11)	0.83
Node 13	*SULT1A1638G>A(M)/NAT803A>G(W)/NAT590G>A(M)/EPHX1337T>C(M)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	3	5	62.50	4.06 (0.89-18.39)	0.07
Node 14	*SULT1A1638G>A(W)/NAT803A>G(W)/NAT590G>A(M)/EPHX1337T>C(M)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	13	10	43.50	1.87 (0.72-4.87)	0.19
Node 15	*NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(M)/NQO1609C>T(W)/NAT857G>A(M)/NAT481C>T(W)*	7	3	30.00	1.04 (0.25-4.39)	0.95
Node 16	*NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	12	25	67.60	5.07 (2.18-11.78)	**0.0002**
Node 17	*NAT803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	3	1	25.00	0.81 (0.08-8.21)	0.86
Node 18	*EPHX1415A>G(W)/NAT803A>G(M)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	0	8	100.00	-	
Node 19	*EPHX1415A>G(M)/NAT803A>G(M)/NQO1609C>T(W)/SULT1A1638G>A(W)/EP*	2	1	33.30	1.22 (0.11-14.09)	0.87

	HX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)					
Node 20	*NAT590G>A(M)/NAT803A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	8	10	55.60	3.04 (1.07-8.69)	**0.04**
Node 21	*NAT590G>A(W)/NAT803A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	9	6	40.00	1.62 (0.52-5.08)	0.40
Node 22	*NAT803A>G(W)/EPHX1415A>G(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	3	8	72.70	6.49 (1.58-26.67)	**0.009**
Node 23	*NAT803A>G(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	1	0	0.00	-	
Node 24	*NAT803A>G(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	3	5	62.50	4.06 (0.89-18.39)	0.07
Node 25	*NAT803A>G(W)/EPHX1415A>G(W)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT857G>A(W)/NAT481C>T(W)*	19	13	40.60	1.66 (0.71-3.92)	0.24
Node 26	*EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT803A>G(W)/EPHX1337T>C(W)/NAT857G>A(W)/NAT481C>T(W)*	1	6	85.70	14.61 (1.66-128.19)	**0.01**
Node 27	*EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT803A>G(W)/EPHX1337T>C(W)/NAT857G>A(W)/NAT481C>T(W)*	5	3	37.50	1.46 (0.32-6.62)	0.62
Node 28	*EPHX1415A>G(M)/SULT1A1638G>A(W)/NAT803A>G(W)/EPHX1337T>C(W)/NAT857G>A(W)/NAT481C>T(W)*	3	6	66.70	4.87 (1.12-21.15)	**0.03**
Node 29	*EPHX1415A>G(W)/SULT1A1638G>A(W)/NAT803A>G(W)/EPHX1337T>C(W)/NAT857G>A(W)/NAT481C>T(W)*	13	11	45.80	2.06 (0.81-5.26)	0.13
Node 30	*NAT803A>G(M)/EPHX1337T>C(W)/NAT857G>A(W)/NAT481C>T(W)*	17	1	5.60	0.14 (0.02-1.14)	0.07
Node 31	*NAT803A>G(M)/NQO1609C>T(W)/EPHX1415A>G(W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	18	38	67.90	5.14 (2.45-10.79)	**0.00002**
Node 32	*NAT803A>G(W)/NQO1609C>T(W)/EPHX1415A>G(W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	5	3	37.50	1.46 (0.32-6.62)	0.62
Node 33	*NAT803A>G(M)/NAT857G>A(M)/EPHX1337T>C(W)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	1	10	90.90	24.35 (2.94-201.26)	**0.003**
Node 34	*NAT803A>G(W)/NAT857G>A(M)/EPHX1337T>C(W)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	1	0	0.00	-	
Node 35	*NAT857G>A(W)/EPHX1337T>C(W)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	5	4	44.40	1.95 (0.48-7.91)	0.35
Node 36	*NAT857G>A(W)/EPHX1337T>C(M)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	7	10	58.80	3.48 (1.18-10.25)	**0.02**
Node 37	*NAT857G>A(M)/EPHX1337T>C(M)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	13	8	38.10	1.49 (0.55-4.09)	0.43
Node 38	*EPHX1337T>C(M)/NAT857G>A(M)/EPHX1415A>G(M)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	4	12	75.00	7.30 (2.13-25.02)	**0.001**
Node 39	*EPHX1337T>C(W)/NAT857G>A(M)/EPHX1415A>G(M)/NAT590G>A(W)/SULT1*	10	5	38.30	1.22 (0.41-3.63)	0.74

	A1638G>A(W)/NAT481C>T(M)					
Node 40	*NQO1609C>T(M)/NAT857G>A(W)/EPHX1415A>G(M)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	3	4	57.10	3.24 (0.67-15.66)	0.14
Node 41	*NQO1609C>T(W)/NAT857G>A(W)/EPHX1415A>G(M)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)*	12	6	33.30	1.22 (0.41-3.63)	0.72
Node 42	*NAT857G>A(W)/NQO1609C>T(M)/NAT803A>G(M)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	5	10	66.70	4.87 (1.50-15.82)	**0.008**
Node 43	*EPHX1415A>G(W)/NAT857G>A(M)/NQO1609C>T(M)/NAT803A>G(M)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	8	12	60.00	3.65 (1.32-10.11)	**0.01**
Node 44	*EPHX1415A>G(M)/NAT857G>A(M)/NQO1609C>T(M)/NAT803A>G(M)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	22	19	46.30	2.03 (0.56-7.31)	0.28
Node 45	*NQO1609C>T/NAT803A>G(M)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	22	19	46.30	2.10 (0.96-4.60)	0.06
Node 46	*EPHX1415A>G(W)/NQO1609C>T(W)/NAT857G>A(W)/NAT803A>G(W)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	5	8	61.50	3.89 (1.15-13.17)	**0.03**
Node 47	*EPHX1415A>G(M)/NQO1609C>T(W)/NAT857G>A(W)/NAT803A>G(W)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	1	1	50.00	2.43 (0.15-40.60)	0.53
Node 48	*NQO1609C>T/NAT857G>A(W)/NAT803A>G(W)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	12	9	42.90	1.83 (0.68-4.92)	0.23
Node 49	*NQO1609C>T(W)/NAT857G>A(M)/NAT803A>G(W)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	9	5	35.70	1.35 (0.41-4.47)	0.62
Node 50	*NQO1609C>T(M)/NAT857G>A(M)/NAT803A>G(W)/EPHX1337T>C(M)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	8	2	20.00	0.61 (0.12-3.09)	0.55
Node 51	*EPHX1415A>G(W)/NQO1609C>T(W)/NAT857G>A(M)/EPHX1337T>C(W)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	5	9	64.30	4.38 (1.32-14.49)	**0.01**
Node 52	*NAT803A>G(M)/EPHX1415A>G(M)/NQO1609C>T(W)/NAT857G>A(M)/EPHX1337T>C(W)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	3	5	62.50	4.06 (0.89-18.39)	0.07
Node 53	*NAT803A>G(W)/EPHX1415A>G(M)/NQO1609C>T(W)/NAT857G>A(M)/EPHX1337T>C(W)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	3	1	25.00	0.81 (0.08-8.21)	0.86
Node 54	*EPHX1415A>G(M)/NQO1609C>T(M)/NAT857G>A(M)/EPHX1337T>C(W)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	5	6	54.50	2.92 (0.81-10.53)	0.10
Node 55	*EPHX1415A>G(W)/NQO1609C>T(M)/NAT857G>A(M)/EPHX1337T>C(W)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	9	6	40.00	1.62 (0.52-5.08)	0.40
Node 56	*NAT857G>A(W)/EPHX1337T>C(W)/NAT590G>A(M)/SULT1A1638G>A(W)/NAT481C>T(M)*	33	9	21.40	0.66 (0.27-1.60)	0.36
Node 57	*NAT803A>G(W)/NAT590G>A(W)/NAT857G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	0	4	100.00	-	0.98
Node 58	*NQO1609C>T(W)/NAT803A>G(M)/NAT590G>A(W)/NAT857G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	2	4	66.70	4.87 (0.83-28.46)	0.08

Node 59	*NQO1609C>T(M)/NAT803A>G(M)/NAT590G>A(W)/NAT857G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	5	7	58.30	3.41 (0.98-11.85)	0.05
Node 60	*EPHX1337T>C(W)/NAT590G>A(M)/NAT857G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	3	5	62.50	4.06 (0.89-18.39)	0.07
Node 61	*NAT803A>G(W)/EPHX1337T>C(M)/NAT590G>A(M)/NAT857G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	3	3	50.00	2.43 (0.46-12.96)	0.29
Node 62	*NQO1609C>T(M)/NAT803A>G(M)/EPHX1337T>C(M)/NAT590G>A(M)/NAT857G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	1	3	75.00	7.30 (0.72-73.93)	0.09
Node 63	*NQO1609C>T(W)/NAT803A>G(M)/EPHX1337T>C(M)/NAT590G>A(M)/NAT857G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	6	2	25.00	0.81 (0.15-4.32)	0.81
Node 64	*NQO1609C>T(W)/EPHX1337T>C(M)/NAT857G>A(W)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	2	6	75.00	7.30 (1.37-38.89)	**0.02**
Node 65	*NQO1609C>T(M)/EPHX1337T>C(M)/NAT857G>A(W)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	8	4	33.30	1.22 (0.33-4.44)	0.76
Node 66	*EPHX1337T>C(W)/NAT857G>A(W)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)*	15	3	16.70	0.49 (0.13-1.84)	0.29
Node 67	*NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)*	11	17	60.70	3.76 (1.53-9.26)	**0.004**
Node 68	*NAT857G>A(W)/NAT803A>G(W)/NQO1609C>T(M)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)*	5	7	58.30	3.41 (0.98-11.85)	0.05
Node 69	*NAT857G>A(M)/NAT803A>G(W)/NQO1609C>T(M)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)*	5	2	28.60	0.97 (0.17-5.38)	0.97
Node 70	*NAT803A>G(W)/NAT857G>A(W)/NQO1609C>T(W)/ NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)*	4	5	55.60	3.04 (0.75-12.36)	0.12
Node 71	*NAT803A>G(M)/NAT857G>A(W)/NQO1609C>T(W)/ NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)*	6	3	33.30	1.22 (0.28-5.29)	0.79
Node 72	*EPHX1337T>C(M)/NAT857G>A(M)/NQO1609C>T(W)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)*	14	5	26.30	0.87 (0.28-2.69)	0.81
Node 73	*EPHX1337T>C(W)/NAT857G>A(M)/NQO1609C>T(W)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)*	7	2	22.20	0.69 (0.13-3.60)	0.66

Subjects with the genotypic combination of nodes 5, 22, 38, and 64 exhibited an odd of seven fold as shown in Table 5.91. Further, nodes 1, 2, 16, 31 showed to exhibit a five-fold risk of lung cancer, and nodes 8, 28, 42,51 and nodes 6, 20, 36, 43, 46, 67 were found to exhibit an odds of four-fold and three-fold, respectively towards lung cancer risk (Table 5.91).

Further, to explore the risk based on different histological subtypes, interaction CART analysis was also performed. Table 5.92 shows the data of all the nodes for SCLC, and Table 5.93 for ADCC and show all nodes data for SQCC.

Subgroups under SCLC category with the highest risk was of node 18, 39, 45 and the genotypic combination of these nodes were *NAT803A>G(M)/NAT857G>A(W)/EPHX1337T>C(M)/NQO1609C>T(W)/SULT1A1638G>A(M)/NAT481C>T(W),EPHX1415A>G(M)/NAT590G>A(W)/NAT857G>A(M)/NAT803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)* and *EPHX1415A>G(M)/NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)* with an odds of 78 (Table 5.92). The node 16 with genotypic combination of *NAT857G>A(M)/EPHX1337T>C(W)/NQO1609C>T(W)/SULT1A1638G>A(M)/NAT481C>T (W)* showed a 52 folds increased risk of SCLC subtype of lung cancer. Terminal node 7 with the genetic combination of *EPHX1337T>C(M)/NAT803A>G(M)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT481C>T(W)* also possessed a significantly high risk towards SCLC subtype (OR=44.57, 95% C.I..=4.80-414.14, p= **0.0008**). The subjects with nodes 8, 10, 19, 23, and 65 exhibited an odds of thirty-nine folds (OR=39.00, 95% C.I..=2.39-634.65, p= **0.01**). Further, the node 1 with genotypic combination*NAT857G>A(M)/EPHX1337T>C(M)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)* and node 46 with genotypic combination *NAT857G>A(M)/EPHX1415A>G(W)/NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)* also had a higher risk for development of SCLC subtype with an odds of 39 and 32 fold, respectively. further the node 2 with genotypic combination of *NAT590G>A(M)/NAT857G>A(W)/EPHX1337T>C(M)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)* exhibited an odds of 24 folds (OR=24.82, 95% C.I..=2.75-223.88, p= **0.004**) and node 4 (*EPHX1337T>C(W)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)*) (OR=21.67, 95%

C.I.= 2.57-223.88, p= **0.004**) and node 37 (*NAT590G>A(M)/NAT857G>A(W)/NAT 803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*) (OR=23.40, 95% C.I..=2.02-270.41, p= **0.01**) also exhibit high risk towards the development of SCLC subtype. There other genotypic combinations as tabulated in Table 5.92 wherein a high risk towards development of SCLC subtype is found. In the ADCC subtype, Table 5.93 and figure 5.39, shows the genotypic combinations of nodes. The initial split was formed by *NAT857G>A*. Subjects with genotypic combination of node 1 carrying *NAT481C>T(W)/NAT857G>A(M)* exhibited an odds of two-folds (OR=2.12, 95% C.I.= 1.30-3.45, p=**0.002**).

Table 5.92: Overall risk estimates based on CART analysis of Phase II detoxification genes for SCLC subgroup

Terminal Node	Genotype for each terminal node	Control (n)	Case (n)	Case rate	OR (95% C.I.)	p-value
Node 1	*NAT857G>A(M)/EPHX1337T>C(M)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)*	9	8	47.10	34.67(3.84-313.36)	**0.002**
Node 2	*NAT590G>A(M)/NAT857G>A(W)/EPHX1337T>C(M)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)*	11	7	38.90	24.82(2.75-223.88)	**0.004**
Node 3	*NAT590G>A(W)/NAT857G>A(W)/EPHX1337T>C(M)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)*	1	0	0.00	-	0.99
Node 4	*EPHX1337T>C(W)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)*	18	10	35.70	21.67(2.57-182.5)	**0.005**
Node 5	*NAT590G>A(M)/NAT803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)*	6	2	25.00	13.00(1.02-166.44)	**0.048**
Node 6	*NAT590G>A(W)/NAT803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/NAT481C>T(W)*	10	0	0.00	-	0.98
Node 7	*EPHX1337T>C(M)/NAT803A>G(M)/NQO1609C>T(W)/ SULT1A1638G>A(W)/NAT481C>T(W)*	7	8	53.30	44.57(4.80-414.14)	**0.0008**
Node 8	*NAT590G>A(M)/EPHX1337T>C(W)/NAT803A>G(M)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT481C>T(W)*	2	2	50.00	39.00(2.40-634.65)	**0.01**
Node 9	*NAT590G>A(W)/EPHX1337T>C(W)/NAT803A>G(M)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT481C>T(W)*	8	0	0.00	-	0.98
Node 10	*EPHX1415A>G(M)/NAT857G>A(W)/EPHX1337T>C(M)/NAT590G>A(M)/NAT803A>G(W)/NQO1609C>T(W)/ SULT1A1638G>A(W)/NAT481C>T(W)*	2	2	50.00	39.00(2.40-634.65)	**0.01**
Node 11	*EPHX1415A>G(W)/NAT857G>A(W)/EPHX1337T>C(M)/NAT590G>A(M)/NAT803A>G(W)/NQO1609C>T(W)/ SULT1A1638G>A(W)/NAT481C>T(W)*	6	1	14.30	6.50(0.36-118.37)	0.21
Node 12	*NAT857G>A(M)/EPHX1337T>C(M)/NAT590G>A(M)/NAT803A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT481C>T(W)*	13	2	13.30	6.00(0.50-71.73)	0.16
Node 13	*EPHX1337T>C(W)/NAT590G>A(M)/NAT803A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT481C>T(W)*	11	1	8.30	3.55(0.20-61.38)	0.38
Node 14	*EPHX1337T>C(W)/NAT590G>A(W)/NAT803A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT481C>T(W)*	5	2	28.60	15.60(1.19-204.78)	**0.04**
Node 15	*EPHX1337T>C(M)/NAT590G>A(W)/NAT803A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT481C>T(W)*	14	0	0.00	-	0.98
Node 16	*NAT857G>A(M)/EPHX1337T>C(W)/NQO1609C>T(W)/SULT1A1638G>A(M)/ NAT481C>T(W)*	3	4	57.10	52.00(4.32-624.76)	**0.002**
Node 17	*NAT857G>A(W)/EPHX1337T>C(W)/NQO1609C>T(W)/SULT1A1638G>A(M)/ NAT481C>T(W)*	5	0	0.00	-	0.98
Node 18	*NAT803A>G(M)/NAT857G>A(W)/EPHX1337T>C(M)/ NQO1609C>T(W)/SULT1A1638G>A(M)/ NAT481C>T(W)*	1	2	66.70	78.00(3.46-1757.30)	**0.006**

Node 19	*EPHX1415A>G(M)/NAT803A>G(W)/NAT857G>A(W)/EPHX1337T>C(M)/NQO1609C>T(W)/SULT1A1638G>A(M)/ NAT481C>T(W)*	2	2	50.00	39.00(2.40-634.65)	**0.01**
Node 20	*EPHX1415A>G(W)/NAT803A>G(W)/NAT857G>A(W)/EPHX1337T>C(M)/NQO1609C>T(W)/SULT1A1638G>A(M)/ NAT481C>T(W)*	9	1	10.00	4.33(0.25-76.05)	0.31
Node 21	*NAT590G>A(W)/NAT857G>A(M)/EPHX1337T>C(M)/NQO1609C>T(W)/SULT1A1638G>A(M)/NAT481C>T(W)*	3	1	25.00	13.00(0.64-263.82)	0.09
Node 22	*NAT590G>A(M)/NAT857G>A(M)/EPHX1337T>C(M)/NQO1609C>T(W)/SULT1A1638G>A(M)/NAT481C>T(W)*	6	0	0.00	-	0.98
Node 23	*EPHX1415A>G(M)/NAT590G>A(M)/EPHX1337T>C(M)/NQO1609C>T(M)/SULT1A1638G>A(M)/NAT481C>T(W)*	2	2	50.00	39.00(2.40-634.65)	**0.01**
Node 24	*NAT857G>A(W)/EPHX1415A>G(W)/NAT590G>A(M)/EPHX1337T>C(M)/NQO1609C>T(M)/SULT1A1638G>A(M)/NAT481C>T(W)*	8	3	27.30	14.63(1.34-159.23)	**0.03**
Node 25	*NAT857G>A(M)/EPHX1415A>G(W)/NAT590G>A(M)/EPHX1337T>C(M)/NQO1609C>T(M)/SULT1A1638G>A(M)/NAT481C>T(W)*	4	0	0.00	-	0.99
Node 26	*NAT590G>A(W)/EPHX1337T>C(M)/NQO1609C>T(M)/SULT1A1638G>A(M)/NAT481C>T(W)*	6	0	0.00	-	0.98
Node 27	*NAT590G>A(W)/EPHX1337T>C(W)/NQO1609C>T(M)/SULT1A1638G>A(M)/NAT481C>T(W)*	2	1	33.33	19.50(0.87-439.33)	0.06
Node 28	*NAT590G>A(M)/EPHX1337T>C(W)/NQO1609C>T(M)/SULT1A1638G>A(M)/NAT481C>T(W)*	10	0	0.00	-	0.98
Node 29	*NAT590G>A(M)/NAT803A>G(W)/EPHX1415A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	11	6	35.30	21.27(2.31-140.67)	0.007
Node 30	*NAT590G>A(W)/NAT803A>G(W)/EPHX1415A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	3	0	0.00	-	0.99
Node 31	*NAT857G>A(M)/NAT803A>G(M)/EPHX1415A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	15	6	28.60	15.60(1.73-140.67_	**0.01**
Node 32	*NAT590(W)/NAT857G>A(W)/NAT803A>G(M)/EPHX1415A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	5	2	28.60	15.60(1.19-204.78)	**0.036**
Node 33	*NAT590G>A(M)/NAT857G>A(W)/NAT803A>G(M)/EPHX1415A>G(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	8	0	0.00	-	0.98
Node 34	*NAT857G>A(M)/NAT803>G(M)/EPHX1415A>G(M)/ NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	7	3	30.00	16.71(1.51-184.60)	**0.02**
Node 35	*NAT857G>A(W)/NAT803>G(M)/EPHX1415A>G(M)/ NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	9	1	10.00	4.33(0.25-76.05)	0.31
Node 36	*NAT803>G(W)/EPHX1415A>G(M)/NQO1609C>T(W)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	5	0	0.00	-	0.98
Node 37	*NAT590G>A(M)/NAT857G>A(W)/NAT803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	5	3	37.50	23.40(2.02-270.41)	**0.01**
Node 38	*NAT590G>A(W)/NAT857G>A(W)/NAT803A>G(M)/NQO1609C>T(M)/SULT1A1*	8	1	11.10	4.88(0.28-86.35)	0.28

	638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)					
Node 39	*EPHX1415A>G(M)/NAT590G>A(W)/NAT857G>A(M)/NAT803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	1	2	66.70	78.00(3.46-1757.30)	**0.006**
Node 40	*EPHX1415A>G(W)/NAT590G>A(W)/NAT857G>A(M)/NAT803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	12	2	14.30	6.50(0.54-78.09)	0.14
Node 41	*NAT590G>A(M)/NAT857G>A(M)/NAT803A>G(M)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	14	1	6.70	2.79(0.16-47.60)	0.48
Node 42	*NAT590G>A(W)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	3	1	25.00	13.00(0.64-263.82)	0.09
Node 43	*EPHX1415A>G(M)/NAT590G>A(M)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	4	1	20.00	9.75(0.51-187.53)	0.13
Node 44	*EPHX1415A>G(W)/NAT590G>A(M)/NAT803A>G(W)/NQO1609C>T(M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)*	16	1	5.90	2.44(0.14-41.40)	0.54
Node 45	*EPHX1415A>G(M)/NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	1	2	66.70	78.00(3.46-1757.30)	**0.006**
Node 46	*NAT857G>A(M)/EPHX1415A>G(W)/NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	6	5	45.50	32.50(3.22-328.30)	**0.003**
Node 47	*NAT857G>A(W)/EPHX1415A>G(W)/NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	2	0	0.00	-	0.99
Node 48	*NAT803A>G(W)/NQO1609C>T(M)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	12	0	0.00	-	0.98
Node 49	*EPHX1415A>G(W)/NAT803A>G(W)/NQO1609C>T(W)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	9	4	30.80	17.33(1.72-174.29)	**0.01**
Node 50	*EPHX1415A>G(M)/NAT803A>G(W)/NQO1609C>T(W)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	2	0	0.00	-	0.99
Node 51	*NAT803A>G(M)/NQO1609C>T(W)/NAT590G>A(M)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	20	0	0.00	-	0.98
Node 52	*EPHX1415A>G(M)/NAT590G>A(W)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	7	2	22.20	11.14(0.89-140.12)	0.06
Node 53	*EPHX1415A>G(W)/NAT590G>A(W)/SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT481C>T(M)*	36	2	5.30	2.17(0.19-24.93)	0.53
Node 54	*NAT857G>A(W)/SULT1A1638G>A(W)/NQO1609C>T(W)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	7	3	30.00	16.71(1.51-184.60)	**0.02**
Node 55	*EPHX1415A>G(W)/NAT857G>A(M)/SULT1A1638G>A(W)/NQO1609C>T(W)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	4	2	33.30	19.50(1.43-265.64)	**0.02**
Node 56	*EPHX1415A>G(M)/NAT857G>A(M)/SULT1A1638G>A(W)/NQO1609C>T(W)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	4	0	0.00	-	0.99
Node 57	*EPHX1415A>G(M)/SULT1A1638G>A(M)/NQO1609C>T(W)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	4	2	33.30	19.50(1.43-265.64)	**0.02**

Node 58	*EPHX1415A>G(W)/SULT1A1638G>A(M)/NQO1609C>T(W)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	8	0	0.00	-	0.98
Node 59	*NAT803A>G(W)/NAT857G>A(W)/NQO1609C>T(M)/ NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	1	2	66.70	78.00(3.46-1757.30)	**0.006**
Node 60	*NAT803A>G(M)/NAT857G>A(W)/NQO1609C>T(M)/ NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	9	1	10.00	4.33(0.25-76.05)	0.31
Node 61	*EPHX1415A>G(M)/NAT803A>G(M)/SULT1A1638G>A(M)/NAT857G>A(M)/NQO1609C>T(M)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	3	1	25.00	13.00(0.64-263.82)	0.09
Node 62	*EPHX1415A>G(W)/NAT803A>G(M)/SULT1A1638G>A(M)/NAT857G>A(M)/NQO1609C>T(M)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	6	1	14.30	6.50(0.36-118.37)	0.21
Node 63	*NAT803A>G(W)/SULT1A1638G>A(M)/NAT857G>A(M)/NQO1609C>T(M)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	4	0	0.00	-	0.99
Node 64	*SULT1A1638G>A(W)/NAT857G>A(M)/NQO1609C>T(M)/NAT590G>A(W)/EPHX1377T>C(W)/NAT481C>T(M)*	8	0	0.00	-	0.98
Node 65	*EPHX1415A>G(W)/NAT857G>A(W)/SULT1A1638G>A(M)/NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(W)/NAT481C>T(M)*	2	2	50.00	39.00(2.39-634.65)	**0.01**
Node 66	*EPHX1415A>G(M)/NAT857G>A(W)/SULT1A1638G>A(M)/NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(W)/NAT481C>T(M)*	5	1	16.70	7.80(0.42-145.20)	0.17
Node 67	*NAT857G>A(M)/SULT1A1638G>A(M)/NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(W)/NAT481C>T(M)*	8	1	11.10	4.88(0.28-86.35)	0.28
Node 68	*EPHX1415A>G(W)/NAT857G>A(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(W)/NAT481C>T(M)*	7	2	22.20	11.14(0.89-140.12)	0.06
Node 69	*EPHX1415A>G(M)/NAT857G>A(W)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(W)/NAT481C>T(M)*	3	0	0.00	-	0.99
Node 70	*NAT857G>A(M)/NQO1609C>T(W)/SULT1A1638G>A(W)/NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(W)/NAT481C>T(M)*	7	1	12.50	5.57(0.31-99.87)	0.24
Node 71	*NQO1609C>T(M)/SULT1A1638G>A(W)/NAT803A>G(M)/NAT590G>A(M)/EPHX1337T>C(W)/NAT481C>T(M)*	16	1	5.90	2.44(0.14-41.40)	0.54

Table 5.93: Overall risk estimates based on CART analysis of Phase II detoxification genes for ADCC subgroup

Terminal Node	Genotype for each terminal node	Control (n)	Case (n)	Case rate	OR (95% C.I.)	p-value
Node 1	*NAT481C>T(W)/NAT857G>A(M)*	85	50	37.00	2.12(1.30-3.45)	**0.002**
Node 2	*NAT803A>G(M)/ NAT481C>T(W)/NAT857G>A(M)*	134	60	30.90	1.53(0.98-2.37)	0.06
Node 3	*NAT803A>G(W)/ NAT481C>T(W)/NAT857G>A(M)*	67	17	20.20	0.98(0.53-1.85)	0.96

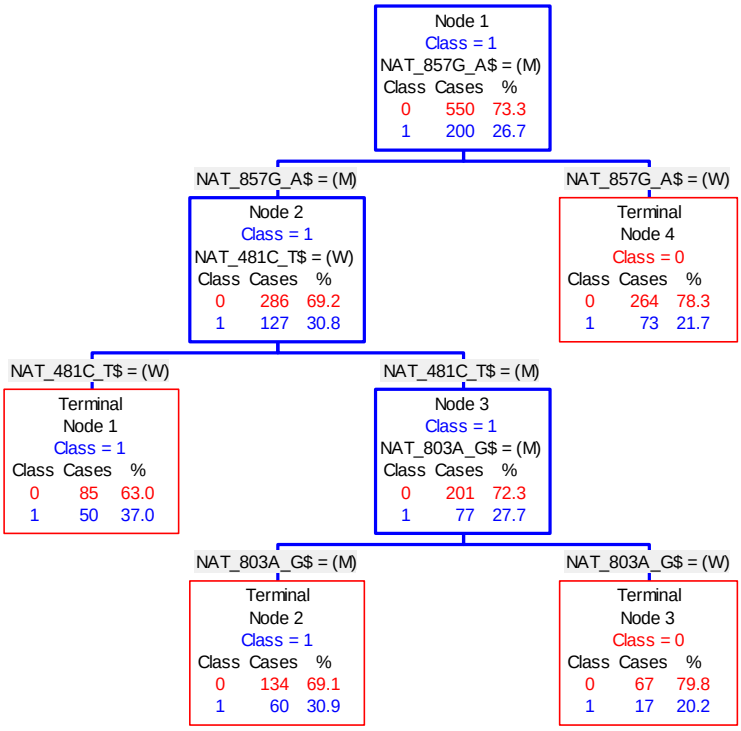


Figure 5.39: Overall risk estimates based on CART analysis of Phase II detoxification genes for ADCC subgroup (W: Homozygous wild type genotype, M: Heterozygous + Homozygous variant genotype)

Further, to explore gene-environment interactions CART analysis incorporating smoking status was also performed. Further, inspection of the CART structure suggested distinct risk genotypic combinations for smokers (Table 5.94) and non-smokers (Table 5.95, Figure 5.40). Terminal node 1 with genotypic combination *NAT857G>A(M)/NQO1609C>T(M)/ EPHX1415A>G(M)/EPHX1337T>C(M)/ NAT481C>T(W)* showed the highest risk towards lung cancer with an odds of 28.7 (OR=28.77, 95% C.I..=3.37-245.64, p=**0.002**). Node27 with the genotypic combination of *NAT590G>A(M)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT803 A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)* possessed a significant high risk (OR=20.92, 95% C.I..=4.21-103.93, p=**0.0002**). The genotypic combination of node2 *NAT803A>G(W)/NAT857G>A(W)/NQO1609C>T(M)/EPHX1415A>G(M)/EPHX1337T>C(M)/ NAT481C>T(W)* also possessed a higher risk towards lung cancer with an odds of 15.69 (95% C.I..=1.72-143.24, p=**0.015**). Further, the nodes 7, 12,13, 15, 17, 18, 21, 23, 26, 30, 34, 38, 48, 55, 57 were high risk subgroups as shown in Table 5.94.

In non-smokers Node1 with the genotypic combination of *EPHX1337T>C(W)/NAT857G>A(M)* exhibit the highest risk towards lung cancer with an odds of 3.34 (95% C.I..=1.60-6.96, p=**0.001**). Further, terminal node 2 (*SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT857G>A(M)*) also possessed a risk of 3 fold towards lung cancer predisposition (Table 5.94).

Table 5.94: Overall risk estimates based on CART analysis of Phase II detoxification genes for smokers

Terminal Node	Genotype for each terminal node	Control (n)	Case (n)	Case rate	OR (95% C.I.)	p-value
Node 1	NAT857G>A(M)/NQO1609C>T(M)/EPHX1415A>G(M)/EPHX1337T>C(M)/NAT481C>T(W)	1	13	92.90	28.77 (3.37-245.64)	**0.002**
Node 2	NAT803A>G(W)/NAT857G>A(W)/NQO1609C>T(M)/EPHX1415A>G(M)/EPHX1337T>C(M)/NAT481C>T(W)	2	7	77.80	15.69(1.72-143.24)	**0.015**
Node 3	NAT803A>G(M)/NAT857G>A(W)/NQO1609C>T(M)/EPHX1415A>G(M)/EPHX1337T>C(M)/NAT481C>T(W)	1	0	0.00		-
Node 4	NAT590G>A(W)/NQO1609C>T(W)/EPHX1415A>G(M)/EPHX1337T>C(M)/NAT481C>T(W)	2	7	77.80	3.92(0.59-26.23)	0.16
Node 5	SULT1A1638G>A(M)/NAT590G>A(M)/NQO1609C>T(W)/EPHX1415A>G(M)/EPHX1337T>C(M)/NAT481C>T(W)	1	4	80.00	5.23(0.44-62.72)	0.19
Node 6	SULT1A1638G>A(W)/NAT590G>A(M)/NQO1609C>T(W)/EPHX1415A>G(M)/EPHX1337T>C(M)/NAT481C>T(W)	6	5	45.50	2.09(0.48-9.03)	0.32
Node 7	NAT803A>G(W)/SULT1A1638G>A(W)/EPHX1415A>G(W)/NAT481C>T(W)	21	52	71.20	6.10(2.53-14.71)	**0.00006**
Node 8	NQO1609C>T(W)/NAT857G>A(W)/NAT803A>G(M)/SULT1A1638G>A(W)/EPHX1415A>G(W)/NAT481C>T(W)	0	8	100.00	-	0.98
Node 9	NQO1609C>T(M)/NAT857G>A(W)/NAT803A>G(M)/SULT1A1638G>A(W)/EPHX1415A>G(W)/NAT481C>T(W)	3	1	25.00	0.87(0.08-9.16)	0.91
Node 10	NAT590G>A(M)/NAT857G>A(M)/NAT803A>G(M)/SULT1A1638G>A(W)/EPHX1415A>G(W)/NAT481C>T(W)	4	5	55.60	2.62(0.57-12.03)	0.22
Node 11	NAT590G>A(W)/NAT857G>A(M)/NAT803A>G(M)/SULT1A1638G>A(W)/EPHX1415A>G(W)/NAT481C>T(W)	5	1	16.70	-	0.98
Node 12	NAT803A>G(M)/SULT1A1638G>A(M)/EPHX1415A>G(W)/NAT481C>T(W)	3	8	72.70	5.23(1.14-24.06)	**0.03**
Node 13	NQO1609C>T(W)/NAT857G>A(M)/NAT803A>G(W)/SULT1A1638G>A(M)/EPHX1415A>G(W)/NAT481C>T(W)	2	5	71.40	10.46(1.07-102.55)	**0.04**
Node 14	NQO1609C>T(M)/NAT857G>A(M)/NAT803A>G(W)/SULT1A1638G>A(M)/EPHX1415A>G(W)/NAT481C>T(W)	5	4	44.40	2.62(0.57-12.03)	0.22
Node 15	NQO1609C>T(M)/NAT857G>A(W)/NAT803A>G(W)/SULT1A1638G>A(M)/EPHX1415A>G(W)/NAT481C>T(W)	6	8	57.10	6.97(1.60-30.42)	**0.01**
Node 16	NQO1609C>T(W)/NAT857G>A(W)/NAT803A>G(W)/SULT1A1638G>A(M)/EPHX1415A>G(W)/NAT481C>T(W)	8	4	33.30	1.74(0.42-7.19)	0.44
Node 17	NQO1609C>T(M)/NAT857G>A(M)/EPHX1415A>G(W)/NAT481C>T(W)	8	18	69.20	7.41(2.40-22.92)	**0.0005**
Node 18	NAT590G>A(M)/NQO1609C>T(W)/NAT857G>A(M)/EPHX1415A>G(W)/NAT481C>T(W)	7	17	70.80	7.32(2.19-24.43)	**0.001**
Node 19	NAT803A>G(W)/NAT590G>A(W)/NQO1609C>T(W)/NAT857G>A(M)/EPHX1415A>G(W)/NAT481C>T(W)	4	7	63.60	3.92(0.95-16.19)	0.06
Node 20	NAT803A>G(M)/NAT590G>A(W)/NQO1609C>T(W)/NAT857G>A(M)/EPHX	5	0	0.00	-	0.99

	1415A>G(W)/NAT481C>T(W)					
Node 21	SULT1A1638G>A(W)/EPHX1415A>G(M)/NAT803A>G(W)/NAT857G>A(W)/EPHX1337T>C(W)/NAT481C>T(W)	2	6	75.00	6.54(1.12-38.00)	**0.04**
Node 22	SULT1A1638G>A(M)/EPHX1415A>G(M)/NAT803A>G(W)/NAT857G>A(W)/EPHX1337T>C(W)/NAT481C>T(W)	4	2	33.30	1.31(0.21-8.02)	0.77
Node 23	SULT1A1638G>A(M)/EPHX1415A>G(W)/NAT803A>G(W)/NAT857G>A(W)/EPHX1337T>C(W)/NAT481C>T(W)	1	4	80.00	10.46(1.07-102.55)	**0.04**
Node 24	SULT1A1638G>A(W)/EPHX1415A>G(W)/NAT803A>G(W)/NAT857G>A(W)/EPHX1337T>C(W)/NAT481C>T(W)	11	8	42.10	1.66(0.53-5.22)	0.38
Node 25	NAT803A>G(M)/NAT857G>A(W)/EPHX1337T>C(W)/NAT481C>T(W)	14	0	0.00	-	0.98
Node 26	NQO1609C>T(W)/EPHX1415A>G(W)/NAT803A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	22	37	62.70	5.75(2.38-13.92)	**0.0001**
Node 27	NAT590G>A(M)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT803A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	8	17	68.00	20.92(4.21-103.93)	**0.0002**
Node 28	NAT857G>A(W)/NAT590G>A(W)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT803A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	3	6	66.70	2.62(0.47-14.65)	0.27
Node 29	NAT857G>A(M)/NAT590G>A(W)/NQO1609C>T(M)/EPHX1415A>G(W)/NAT803A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	11	6	35.30	1.31(0.34-5.09)	0.70
Node 30	NQO1609C>T(M)/EPHX1415A>G(M)/NAT803A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	7	11	61.10	4.71(1.33-16.70)	**0.02**
Node 31	NAT857G>A(M)/NAT590G>A(W)/NQO1609C>T(W)/EPHX1415A>G(M)/NAT803A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	2	4	66.70	5.23(0.85-32.08)	0.07
Node 32	NAT857G>A(M)/NAT590G>A(W)/NQO1609C>T(W)/EPHX1415A>G(M)/NAT803A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	4	2	33.30	0.65(0.07-6.41)	0.72
Node 33	NAT590G>A(M)/NQO1609C>T(W)/EPHX1415A>G(M)/NAT803A>G(M)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	5	2	28.60	0.65(0.07-6.41)	0.72
Node 34	NAT590G>A(W)/NAT803A>G(W)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	4	8	66.70	5.23(1.34-20.38)	**0.02**
Node 35	NQO1609C>T(W)/NAT857G>A(W)/NAT590G>A(M)/NAT803A>G(W)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	4	8	66.70	3.92(0.95-16.19)	0.06
Node 36	NQO1609C>T(M)/NAT857G>A(W)/NAT590G>A(M)/NAT803A>G(W)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	11	7	38.90	1.96(0.57-6.75)	0.28
Node 37	NAT857G>A(M)/NAT590G>A(M)/NAT803A>G(W)/EPHX1337T>C(M)/SULT1A1638G>A(W)/NAT481C>T(M)	14	6	30.00	1.31(0.41-4.21)	0.65
Node 38	NAT803A>G(M)/EPHX1415A>G(W)/NAT857G>A(M)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	8	18	69.20	11.12(3.14-39.30)	**0.0002**
Node 39	NQO1609C>T(W)/NAT803A>G(W)/EPHX1415A>G(W)/NAT857G>A(M)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	2	4	66.70	5.23(0.85-32.08)	0.07
Node 40	NQO1609C>T(M)/NAT803A>G(W)/EPHX1415A>G(W)/NAT857G>A(M)/EPH	7	2	22.20	1.05(0.18-6.08)	0.96

	X1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)					
Node 41	EPHX1415A>G(M)/NAT857G>A(M)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	20	13	39.40	1.45(0.53-3.96)	0.46
Node 42	NAT803A>G(W)/NAT590G>A(W)/NAT857G>A(W)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	1	2	66.70	5.23(0.44-62.72)	0.19
Node 43	EPHX1415A>G(W)/NQO1609C>T(W)/NAT803A>G(M)/NAT590G>A(W)/NAT857G>A(W)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	3	4	57.10	3.92(0.59-26.23)	0.16
Node 44	EPHX1415A>G(M)/NQO1609C>T(W)/NAT803A>G(M)/NAT590G>A(W)/NAT857G>A(W)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	3	1	25.00	1.31(0.11-15.68)	0.83
Node 45	EPHX1415A>G(M)/NQO1609C>T(M)/NAT803A>G(M)/NAT590G>A(W)/NAT857G>A(W)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	1	2	66.70	-	0.99
Node 46	EPHX1415A>G(W)/NQO1609C>T(M)/NAT803A>G(M)/NAT590G>A(W)/NAT857G>A(W)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	5	2	28.60	0.65(0.07-6.41)	0.72
Node 47	NAT590G>A(M)/NAT857G>A(W)/EPHX1337T>C(W)/SULT1A1638G>A(W)/NAT481C>T(M)	28	7	23.30	1.14(0.38-3.42)	0.81
Node 48	NAT857G>A(M)/NAT590G>A(W)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)	4	11	73.30	7.84(1.83-33.60)	**0.006**
Node 49	EPHX1337T>C(M)/NAT857G>A(W)/NAT590G>A(W)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)	2	5	71.40	-	0.98
Node 50	EPHX1337T>C(W)/NAT857G>A(W)/NAT590G>A(W)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)	4	1	20.00	0.65(0.07-6.41)	0.72
Node 51	NQO1609C>T(M)/NAT857G>A(M)/NAT590G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)	3	5	62.50	4.36(0.91-20.90)	0.07
Node 52	NAT803A>G(W)/NQO1609C>T(W)/NAT857G>A(M)/NAT590G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)	2	3	60.00	-	0.99
Node 53	NAT803A>G(M)/NQO1609C>T(W)/NAT857G>A(M)/NAT590G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)	6	3	33.30	1.05(0.18-6.08)	0.96
Node 54	NAT857G>A(W)/NAT590G>A(M)/EPHX1415A>G(M)/SULT1A1638G>A(M)/NAT481C>T(M)	11	4	26.70	1.16(0.30-4.44)	0.83
Node 55	NAT857G>A(M)/NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)	7	10	58.80	3.66(0.98-13.62)	**0.05**
Node 56	NAT857G>A(W)/NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)	3	1	25.00	0.87(0.08-9.16)	0.91
Node 57	NAT857G>A(W)/NAT803A>G(W)/NQO1609C>T(M)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)	4	6	60.0	6.54(1.12-38.00)	**0.04**
Node 58	NAT857G>A(M)/NAT803A>G(W)/NQO1609C>T(M)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)	4	2	33.30	1.31(0.11-15.68)	0.83
Node 59	NAT857G>A(W)/NAT803A>G(W)/NQO1609C>T(W)/NAT590G>A(M)/EPHX1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)	4	5	55.60	4.36(0.91-20.90)	0.06

Node 60	NAT857G>A(M)/NAT803A>G(W)/NQO1609C>T(W)/NAT590G>A(M)/EPHX 1415A>G(W)/SULT1A1638G>A(M)/NAT481C>T(M)	8	3	27.30	1.12(0.25-5.00)	0.88
Node 61	NAT803A>G(M)/NQO1609C>T(W)/NAT590G>A(M)/EPHX1415A>G(W)/SUL T1A1638G>A(M)/NAT481C>T(M)	18	5	21.70	0.33(0.06-1.62)	0.17

Table 5.95: Overall risk estimates based on CART analysis of Phase II detoxification genes for non-smokers subgroup

Terminal Node	**Genotype for each terminal node**	**Control (n)**	**Case (n)**	**Case rate**	**OR (95% C.I.)**	**p-value**
Node 1	EPHX1337T>C(W)/NAT857G>A(M)	15	29	65.90	3.34(1.60-6.96)	**0.001**
Node 2	SULT1A1638G>A(M)/EPHX1337T>C(M)/NAT857G>A(M)	10	17	63.00	2.93(1.23-7.02)	**0.02**
Node 3	NQO1609C>T(W)/NAT803A>G(M)/ SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT857G>A(M)	4	8	66.70	3.45(0.98-12.19)	0.05
Node 4	NQO1609C>T(M)/NAT803A>G(M)/ SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT857G>A(M)	8	5	38.50	1.08(0.33-3.52)	0.90
Node 5	NAT803A>G(W)/ SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT857G>A(M)	19	7	26.90	0.64(0.25-1.64)	0.35

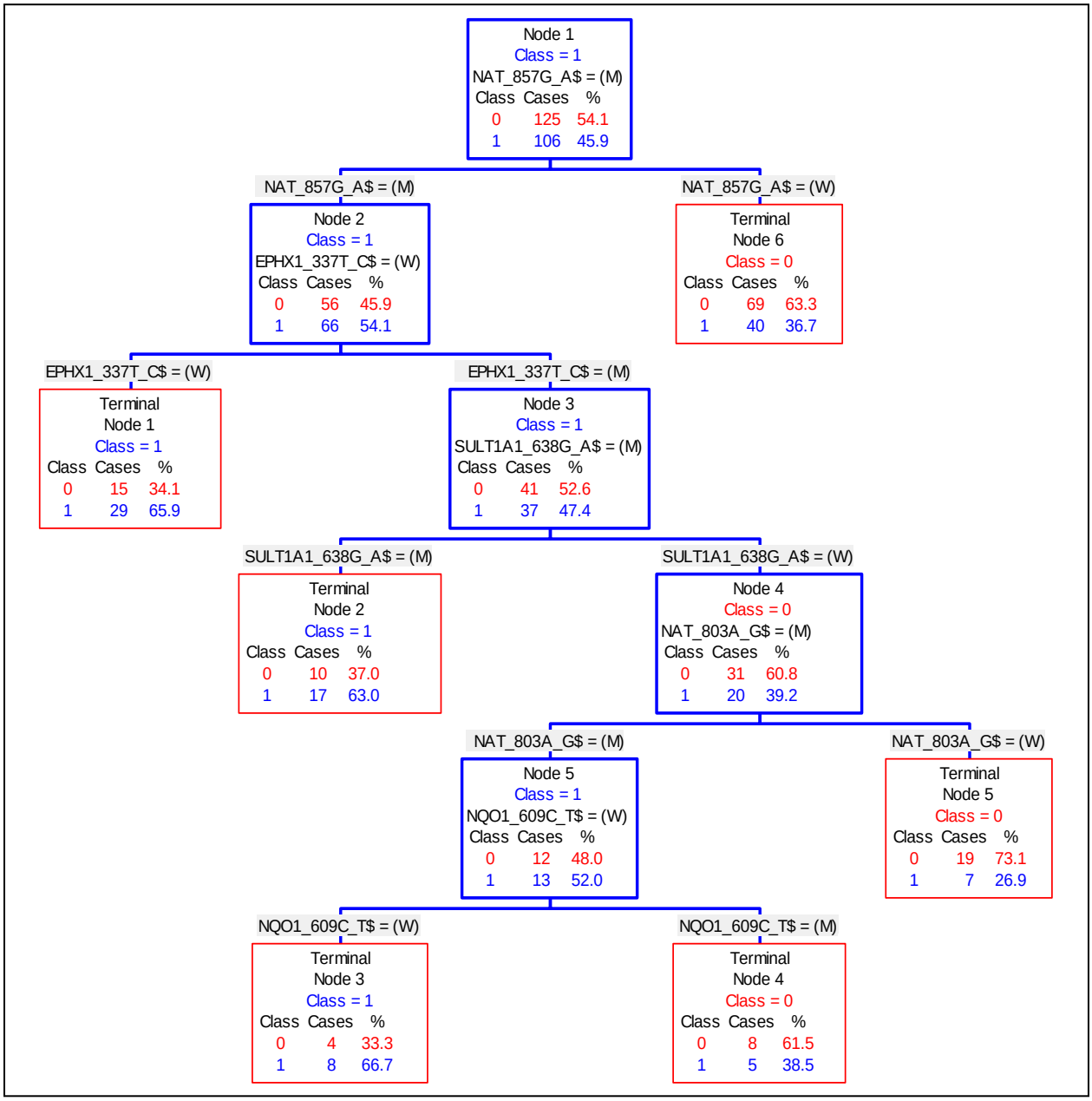


Figure 5.40: Overall risk estimates based on CART analysis of Phase II detoxification genes for non-smokers subgroup (W: Homozygous wild type genotype, M: Heterozygous + Homozygous variant genotype)

5.11.9. Association of unfavorable genotypic combinations of Phase II detoxification genes with overall survival of lung cancer patients

Different genotypic combinations of unfavorable genotypes of Phase II genes were investigated to predict lung cancer patients' overall survival, as shown in Table 5.96. The frequency of single allelic carriers for the combination of *NQO1 Pro^{187}Ser* with *SULT1A1 Arg^{213}His*, *EPHX1 His^{139}Arg*, *NAT2*5C Lys^{268}Arg* was less compared to wild type subjects, respectively, as demonstrated in Table 5.96. This genotype for the combination of *NQO1 Pro^{187}Ser* with *NAT2*5B Leu^{161}Leu*, and *NAT2*6 Arg^{197}Gln* was almost identical to the wild-type subjects, whereas, with *EPHX1 Tyr^{113}His*, and *NAT2*7 Gly^{286}Glu*, the heterozygous was over-represented than the wild type genotype subjects (Table 5.96). In the case of the polymorphic combination of *NQO1 Pro^{187}Ser* with *EPHX1 Tyr^{113}His*, the MST was observed to be better for mutant subjects (*TT+CC*) in comparison to wild subjects (*CC+TT*) (17.57 months *vs*. 10.40 months, log-rank p=0.21). For the following combination of *NQO1 Pro^{187}Ser* with *EPHX1, His^{139}Arg* MST of single allelic carriers (*CT+AG*) was observed to be higher compared to wild subjects (*CC+AA*) (12.17 months *vs*. 7.70 months, log-rank p=**0.03**) (Figure 5.41 a)). The hazard rate was sequentially lower, indicating a trend towards decreased death rate (HR'=0.65, 95% C.I.=0.44-0.95, p=**0.03**). Similarly, the subjects with combined genotype (CT+AG+TT+GG) for the same genotypic combination had higher survival than wild subjects (12.13 months *vs*. 7.70 months, log-rank p=**0.04**) and a lower death rate as observed by Cox-proportional analysis (HR'=0.66, 95% C.I.=0.45-0.96, p=**0.03**). The polymorphic *NQO1 Pro^{187}Ser* did not show any association with lung cancer risk combined with any other genotypic variant in the study.

Further, polymorphic *SULT1A1 Arg^{213}His* was analyzed for the combination of unfavorable genotypes. Survival analysis did not reveal any significant difference in the median survival time and hazard rate for the combination of *SULT1A1 Arg^{213}His* with either of the three polymorphic variants *EPHX1 Tyr^{113}His*, *NAT2*6 Arg^{197}Gln*, and *NAT2*5C Lys^{268}Arg* as shown in Table 5.96. For the combination of *SULT1A1 Arg^{213}His* with *EPHX1 His^{139}Arg*, a better survival was observed for heterozygous type subjects as subjects with wild genotype (9.80 months *vs*. 7.60 months, log-rank p=0.06) (Figure 5. 41 b)). Further, regression analysis showcased a lower hazard rate for heterozygous subjects (HR'=0.74, 95% C.I.=0.50-1.08, p=0.12). For the combination of *SULT1A1 Arg^{213}His* with *EPHX1 Tyr^{113}His*, the subjects with mutant genotype (*AA+CC*) had better survival than the wild type subjects (*GG+TT*) (19.93

months *vs*. 10.27 months log-rank p=0.35), and a lower death rate was also observed (HR'=0.53, 95% C.I.=0.16-1.79, p=0.31), but it lacked statistical significance.

Further, the frequency of single allelic carriers for the combination of *EPHX1 Tyr^{113}His* with *NAT2*5B Leu^{161}Leu*, *NAT2*6 Arg^{197}Gln*, and *NAT2*7 Gly^{286}Glu* was higher compared to wild type subjects, respectively, as shown in Table 5.96. However, for the combination of *EPHX1 Tyr^{113}His* with *NAT2*5C Lys^{268}Arg*, subjects carrying mutant genotype (*CC+GG*) had a smaller MST compared to wild type subjects (*TT+AA*) (2.80 months *vs*. 9.93 months). The adjusted hazards ratio analyzed using Cox-proportional analysis exhibited a high death rate (HR'=3.32, 95% C.I.=1.51-7.32, p=**0.003**).

A trend of higher MST of 16.17 months was observed for subjects harboring double unfavorable genotype compared to subjects with respective wild genotype in case of the combination of *EPHX1 His^{139}Arg* with *NAT2*5B Leu^{161}Leu*, *NAT2*5C Lys^{268}Arg*, and *NAT2*7 Gly^{286}Glu*. The genotypic combination of *EPHX1 His^{139}Arg* with *NAT2*6 Arg^{197}Gln* subjects comprising mutant genotype exhibit a poor survival of 2.27 months compared to 6.87 months (log-rank p=0.80) for subjects with wild genotype. Further, the genotypic combination of *EPHX1 His^{139}Arg* with *NAT2*5B Leu^{161}Leu* subjects comprising unfavorable genotype for single allele exhibited a better survival of 9.80 months than 6.87 months (log-rank p-value=0.11) for subjects lacking any unfavorable genotype. Hazards ratio obtained from univariate analysis (HR=0.75) decreased further when adjusted for various covariates the death rate significantly decreased (HR'=0.67, 95% C.I.= 0.46-0.98, p=**0.04**). Several unfavorable genotypes were also examined for other SNPs, but none showed any significant association with prognosis.

Table 5.96: Genotypic combinations and overall survival

NQO1 P187S & SULT1A1 R213H	**Cases N=224 ,n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (CC + GG)	157(70.09)	7.57				
1 (CT + GA)	63(28.13)	5.27	0.88	0.97 (0.70-1.35)	0.99 (0.70-1.40)	0.95
2 (TT + AA)	4(1.78)	2.23	0.51	1.46 (0.37-5.79)	2.66 (0.79-9.01)	0.12
1-2	67(29.91	5.27	0.96	0.99 (0.72-1.37)	1.03 (0.73-1.45)	0.86
NQO1 P187S & EPHX1 Y113H	**Cases N= 179, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (CC + TT)	79(44.13)	10.40				
1 (CT + TC)	93(51.96)	8.43	0.59	1.10 (0.77-1.56)	1.03 (0.71-1.50)	0.86
2 (TT + CC)	7(3.91)	17.57	0.21	0.54 (0.24-1.18)	0.98 (0.28-3.44)	0.97
1-2	100(55.87)	9.10	0.83	1.04 (0.73-1.46)	0.99 (0.69-1.44)	0.97
NQO1 P187S & EPHX1 H139R	**Cases N= 216, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (CC +AA)	156(72.22)	7.70				
1 (CT + AG)	59(27.31)	12.17	**0.03**	0.67 (0.48-0.94)	0.65 (0.44-0.95)	**0.03**
2 (TT + GG)	1(0.46)	4.73	0.41	2.24 (0.12-42.03)	2.25 (0.30-16.94)	0.43
1-2	60(27.78)	12.13	**0.04**	0.68 (0.49-0.95)	0.66 (0.45-0.96)	**0.03**
NQO1 P187S & NAT2*5B L161L	**Cases N= 189, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (CC +CC)	91(48.15)	7.93				
1 (CT + CT)	94(49.73)	9.20	0.24	0.82 (0.59-1.15)	0.82(0.58-1.17)	0.27
2 (TT + TT)	4(2.12)	-	0.12	0.24(0.09-0.67)	0.12(0.01-0.98)	0.05
1-2	98(51.85)	9.53	0.16	0.79(0.57-1.10)	0.79(0.56-1.12)	0.19
NQO1 P187S & NAT2*6 R197Q	**Cases N= 210, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (CC +GG)	99(47.14)	7.57				
1 (CT + GA)	102(48.57)	9.10	0.92	0.98 (0.72-1.35)	1.06 (0.76-1.49)	0.72
2 (TT + AA)	9(4.29)	5.20	0.75	1.14 (0.47-2.75)	1.10 (0.46-2.62)	0.83
1-2	111(52.86)	8.63	0.97	0.99 (0.73-1.36)	1.06 (0.76-1.47)	0.73
NQO1 P187S & NAT2*5C K268R	**Cases N= 179, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (CC +AA)	105(58.66)	9.13				
1 (CT + AG)	68(37.99)	9.13	0.65	0.92(0.65-1.31)	0.93(0.64-1.36)	0.71
2 (TT + GG)	6(3.35)	5.93	0.65	0.80(0.32-1.97)	1.13(0.39-3.28)	0.83
1-2	74(41.34)	9.13	0.60	0.91(0.65-1.28)	0.93(0.64-1.34)	0.68
NQO1 P187S & NAT2*7 G286E	**Cases N= 195, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (CC +GG)	85(43.59)	8.20				
1 (CT + GA)	103(52.82)	6.07	0.72	1.06(0.77-1.45)	1.08(0.77-1.52)	0.66
2 (TT + AA)	7(3.59)	9.83	0.45	0.68(0.29-1.60)	1.10(0.38-3.19)	0.86
1-2	110(56.41)	6.07	0.84	1.03 (0.76-1.41)	1.08(0.77-1.52)	0.64
SULT1A1 R213H & EPHX1 Y113H	**Cases N= 182, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (GG+ TT)	107(58.79)	10.27				

1 (GA + TC)	69(37.91)	7.10	0.56	1.11(0.78-1.58)	1.20(0.83-1.74)	0.33
2 (AA + CC)	6(3.30)	19.93	0.35	0.58(0.24-1.44)	0.53(0.16-1.79)	0.31
1-2	75(41.21)	7.70	0.74	1.06(0.75-1.49)	1.14(0.79-1.63)	0.49
SULT1A1 R213H & EPHX1 H139R	**Cases N= 272, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (GG+ AA)	220(80.88)	7.60				
1 (GA + AG)	52(19.12)	9.80	0.06	0.71(0.51-0.99)	0.74(0.50-1.08)	0.12
SULT1A1 R213H & NAT2*5B L161L	**Cases N= 199, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (GG+ CC)	130(65.33)	7.80				
1 (GA + CT)	69(34.67)	5.83	0.81	0.96(0.69-1.33)	0.94(0.66-1.34)	0.73
SULT1A1 R213H & NAT2*6 R197Q	**Cases N= 208, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (GG+ CC)	129(62.02)	7.23				
1 (GA + CT)	73(35.10)	9.47	0.34	0.86(0.62-1.17)	1.02(0.71-1.47)	0.90
2 (AA + TT)	6(2.88)	7.23	0.32	0.57(0.23-1.37)	0.59(0.14-2.50)	0.48
1-2	79	9.47	0.26	0.83(0.61-1.14)	1.01(0.71-1.45)	0.95
SULT1A1 R213H & NAT2*5C K268R	**Cases N= 205, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (GG+ AA)	148(72.20)	8.37				
1 (GA + AG)	55(26.83)	5.20	0.22	1.24(0.86-1.79)	1.39(0.95-2.03)	0.09
2 (AA + GG)	2(0.97)	5.93	0.74	0.72(0.13-3.80)	1.15(0.15-8.87)	0.89
1-2	57(27.80)	5.27	0.25	1.22(0.85-1.75)	1.37(0.94-1.99)	0.11
SULT1A1 R213H & NAT2*7 G286E	**Cases N= 197, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (GG+ GG)	114(57.87)	7.93				
1 (GA + GA)	79(40.10)	5.53	0.62	1.08(0.79-1.48)	1.09(0.78-1.52)	0.62
2 (AA + AA)	4(2.03)	-	0.17	0.28(0.09-0.81)	0.34(0.05-2.50)	0.29
1-2	83(42.13)	5.83	0.82	1.04(0.76-1.41)	1.06(0.76-1.47)	0.75
EPHX1 Y113H & NAT2*5B L161L	**Cases N= 189, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (TT + CC)	68(35.98)	9.13				
1 (TC + CT)	112(59.26)	7.57	0.89	0.98(0.69-1.38)	0.96(0.67-1.39)	0.85
2 (CC + TT)	9(4.76)	9.43	0.49	1.30(0.57-2.95)	1.49(0.64-3.44)	0.36
1-2	121(64.02)	7.60	0.98	0.99(0.71-1.40)	0.98(0.68-1.40)	0.89
EPHX1 Y113H & NAT2*6 R197Q	**Cases N= 191, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (TT + GG)	69(36.12)	7.53				
1 (TC + GA)	111(58.12)	8.40	0.99	1.00(0.72-1.40)	1.02(0.71-1.46)	0.91
2 (CC + AA)	11(5.76)	7.10	0.62	0.82(0.39-1.70)	0.69(0.28-1.73)	0.43
1-2	122(63.87)	8.27	0.89	0.98(0.70-1.36)	1.02(0.71-1.45)	0.93
EPHX1 Y113H & NAT2*5C K268R	**Cases N= 188, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (TT + AA)	85(45.21)	9.93				
1 (TC + AG)	88(46.81)	7.60	0.52	1.12(0.79-1.58)	1.11(0.77-1.61)	0.58
2 (CC + GG)	15(7.98)	2.80	0.31	1.37(0.69-2.72)	3.32(1.51-7.32)	**0.003**
1-2	103(54.79)	7.60	0.38	1.16(0.83-1.61)	1.17(0.82-1.67)	0.38

EPHX1 Y113H & NAT2*7 G286E	**Cases N= 149, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (TT + GG)	42(28.19)	10.30				
1 (TC + GA)	99(66.44)	9.33	0.96	1.01(0.68-1.51)	0.98(0.65-1.50)	0.94
2 (CC + AA)	8(5.37)	8.23	0.81	1.12(0.42-2.96)	1.79(0.65-4.97)	0.27
1-2	107(71.81)	9.33	0.95	1.01(0.68-1.50)	0.98(0.65-1.49)	0.93
EPHX1 H139R & NAT2*5B L161L	**Cases N= 195, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (AA + CC)	133(68.21)	7.90				
1 (AG + CT)	61(31.28)	9.80	0.11	0.75(0.54-1.05)	0.67(0.46-0.98)	**0.04**
2 (GG + TT)	1(0.51)	16.17	0.66	0.65(0.13-3.18)	1.82(0.24-13.71)	0.56
1-2	62(31.79)	9.80	0.10	0.75(0.53-1.04)	0.68(0.47-0.99)	**0.04**
EPHX1 H139R & NAT2*6 R197Q	**Cases N= 191, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (AA + GG)	131(68.59)	6.87				
1 (AG + GA)	58(30.37)	9.80	0.18	0.79(0.56-1.10)	0.79(0.54-1.15)	0.22
2 (GG + AA)	2(1.04)	2.27	0.80	1.19(0.26-5.49)	1.80(0.43-7.49)	0.42
1-2	60(31.41)	9.80	0.20	0.80(0.57-1.12)	0.81(0.56-1.16)	0.25
EPHX1 H139R & NAT2*5C K268R	**Cases N= 209, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (AA + AA)	160(76.56)	8.60				
1 (AG + AG)	48(22.96)	7.80	0.73	0.94(0.65-1.35)	0.87(0.59-1.28)	0.48
2 (GG + GG)	1(0.48)	16.17	0.70	0.69(0.13-3.51)	1.38(0.19-10.25)	0.75
1-2	49(23.44)	8.63	0.69	0.93(0.65-1.33)	0.88(0.60-1.28)	0.50
EPHX1 H139R & NAT2*7 G286E	**Cases N= 204, n(%)**	**Univariate analysis**			**Multivariate analysis**	
		MST (months)	**Log rank *p***	**Unadjusted HR[a]**	**Adjusted HR[b] (95% C.I.)**	***P***
0 (AA + GG)	127(62.25)	7.93				
1 (AG + GA)	76(37.25)	7.33	0.62	0.92(0.68-1.26)	0.89(0.64-1.24)	0.50
2 (GG + AA)	1(0.50)	16.17	0.59	0.58(0.13-2.66)	0.96(0.13-7.17)	0.97
1-2	77(37.75)	7.47	0.58	0.92(0.67-1.25)	0.89(0.64-1.24)	0.49

[a]Unadjusted Hazards ratio for Kaplan Meier analysis, [b]hazards ratio for age, sex, smoking, histology, stage, KPS, ECOG

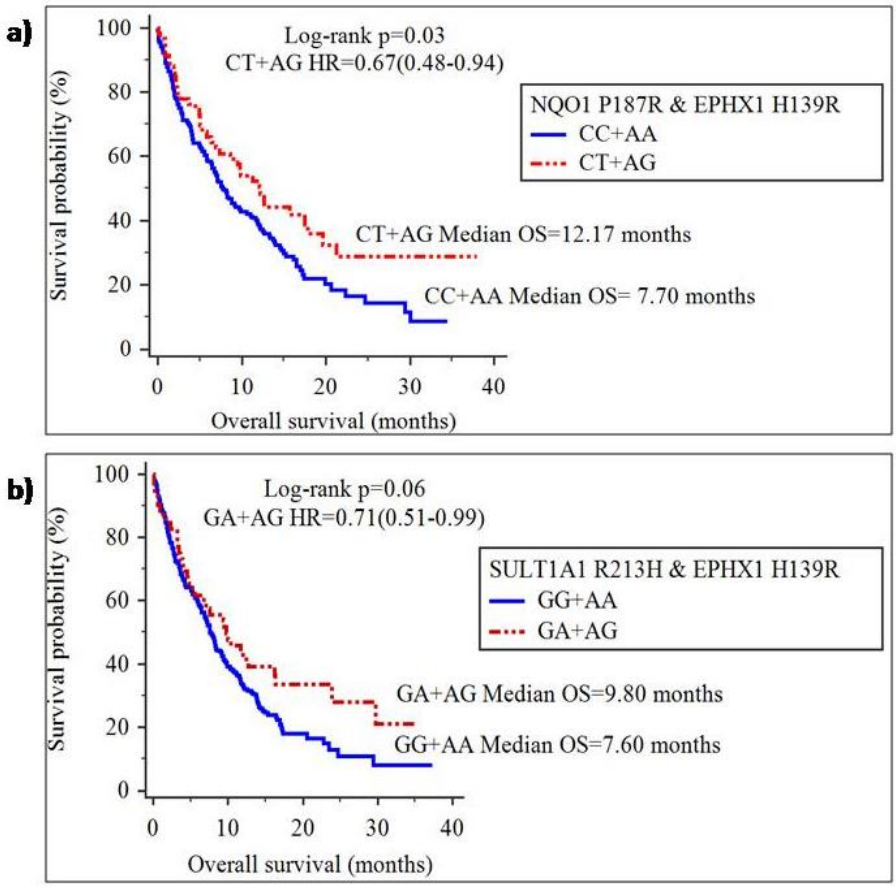


Figure 5.41: a) Kaplan Meier curves illustrating the association between overall survival in heterozygous and wild genotypes for the genotypic combination of *NQO1 Pro*^{187}*Ser* with *EPHX1 His*^{139}*Arg*. (Note: *AA* and *CC* represent wild-type genotype, *CT* and *AG* represent heterozygous genotype).

b) Kaplan Meier curves illustrating the association between overall survival in heterozygous and wild genotype for the combination of *SULT1A1 Arg*^{213}*His* with *EPHX1 His*^{139}*Arg*. (Note: *GG* and *AA* represent wild-type genotype, *GA* and *AG* represent heterozygous genotype).

CHAPTER 6

DISCUSSION

Genome-wide association studies (GWAS) have been possible because of the completion of the International HapMap Project (International HapMap Consortium, 2007; Xu and Taylor, 2009) and the development of sophisticated genotyping technology. Typically, more than 1000 patients and controls are genotyped for 300,000 to one million SNPs in these investigations. Any GWAS will yield a huge number of candidate SNPs. Because custom arrays are more expensive than regular GWAS arrays, this big SNP list poses a barrier for validation studies involving a large number of participants. There is a large, diversified, and growing literature for many diseases that can be used to find genes and chromosomal regions of interest (Xu and Taylor, 2009; Pavan *et al*., 2020). This literature contains research on illness causation as well as current genetic investigations of linkage and candidate genes. For example, information on role of different PhaseII detoxification genes, from earlier literature on cancer basically lung cancer, have led to the discovery of the genes chosen for our study as the promising lung cancer candidate genes. As a result, researchers may have a vast but limited list of genes and genomic regions that they believe merit special attention, or they may have a specific interest in certain genes or chromosomal regions.

6.1. *NQO1* polymorphism and its relationship with lung cancer susceptibility

Our study did not find any association between the mutant (*TT*) genotype and lung cancer risk in the North Indian population. Our data here are consistent with several other studies conducted (Saldivar *et al*., 2005; Timofeeva *et al*., 2010; Guo *et al*., 2012). On the contrary, other studies have reported an association of the *NQO1* polymorphism towards an increased risk for lung cancer (Eom *et al*., 2009; Tian *et al*., 2014; Masroor *et al*., 2015). In a study conducted by Saldivar *et al*. (2005), the *NQO1* genotype was associated with a marginally increased risk of lung cancer. In a previous study by Lewis *et al*. (2001) almost 4-fold increase in risk towards SCLC was observed in patients with at least one variant allele. In our study, the effect of *NQO1* polymorphism on lung cancer was contradictory from some of the previous studies conducted in which the *NQO1* variant genotype exhibited a potential protective effect (Chao *et al*., 2006;

Yang *et al*., 2007; Wang *et al*., 2008). This deviation may be because various factors, such as age, gender, exposure levels of carcinogens, may affect the NQO1 enzyme activity. It has also been reported that the association of the *NQO1C*^{609}*T* polymorphism in other types of malignancies had conflicting results wherein some studies observed an increased risk and other studies showed a lack of association for several other malignancies, e.g., colorectal cancer, breast cancer, bladder cancer, prostate cancer, cervical cancer (Chao *et al*., 2006; Yang *et al*., 2007; Wang *et al*., 2008; Sameer *et al*., 2010; Yuan *et al*., 2011; Lajin *et al*., 2013). In a meta-analysis conducted by Huang *et* al. (2017) overall, the *NQO1 C609T* polymorphism did not connect with the risk of lung cancer. Furthermore, study stratified by ethnicity, control source, and smoking status revealed no statistical significance, but the patients diagnosed with SCLC, there was a significant association found for the risk of SCLC. In a meta-analysis the pooled data for the *NQO1* polymorphism was found to be significantly associated with an increased risk of bladder cancer and prostate cancer. But further stratification on basis of ethnicity the results depict an increased risk of bladder cancer among Caucasians and the prostate cancer among Asians (Abedinzadeh *et al*., 2020). Also in a meta-analysis conducted on colorectal cancer by Gorący *et al*. (2021) an elevated risk was seen in patients carrying the mutant genotype.

NQO1 has also been found to convert benzene-derived quinones into less active hydroquinones, thereby protecting against benzene-induced hematotoxicity (Ross and Siegel, 2004). It has been reported that skin cancer generated by Benzo[*a*]pyrene and 7, 12-dimethylbenz[a]-anthracene (7,12-DMBA) are enhanced in *NQO1* knock-out mice, implying that NQO1 is engaged in the detoxification of PAHs (most likely PAH o-quinones) to inactive PAH hydroquinones (Long *et al*., 2000; Long *et al*., 2001). In some cases, NQO1 converts anticancer drugs such as mitomycin C and streptomycin to reactive metabolites that are hazardous in both normal well as tumor cells (Shimada, 2006).

We further evaluated the role of this polymorphism based on tumor histology. Our data exhibit that the *NQO1* mutant (*TT*) genotype was associated with an increased lung cancer risk in ADCC subjects and SCLC subjects, although not reaching statistical significance. The findings of our study are consistent with earlier studies wherein *TT* genotype has been reported with an increased risk of NSCLC (Tian *et al*., 2014; Masroor *et al*., 2015). In a study carried out by Lewis *et al*. (2001), patients with at least one variant allele of the *NQO1* genotype had a four-

fold increased risk of SCLC. Eom *et al*. (2009) discovered no association between different lung cancer histological cell types. Furthermore, our study found no association between *NQO1* polymorphism and smoking status. However, other studies have found an increased risk of lung cancer among heavy smokers carrying heterozygous (Eom *et al*., 2009) and mutant genotype (Masroor *et al*., 2015). In another study conducted by Banerjee (2020) in a case-control study involving patients from Eastern India the *NQO1 C609T* polymorphism was found to be related with a greater risk of lung cancer in the genotypic comparison between the two smoker groups.

6.1.1. Relation of *NQO1* polymorphism with clinic-pathological features and prognosis

We also evaluated the significance of *NQO1* polymorphism in clinic-pathological features such as stage, tumor size, lymph node invasion, and metastasis. We found that lung cancer subjects with the mutant (*TT*) genotype had a 3.5-fold higher incidence of stage III lung cancer than stage IV patients. Our findings do not agree with those of an earlier study (Masroor*et al*., 2015). Similarly, we found that patients with variant genotype had a five-fold greater risk of lymph node invasion. Furthermore, it was also found that lung cancer cases with heterozygous (*CT*) and mutant (*TT*) genotypes displayed a protective effect in lung cancer progression to other organs. Masroor *et al*. (2015) discovered similar findings in their investigation. Several studies have previously been conducted to investigate the role of NQO1 protein expression in various cancers,including NSCLC (Li *et al*., 2015), gastric ADCC (Nunobiki *et al*., 2015), ovarian carcinoma (Lin *et al*., 2014), and pancreatic ADCC (Ji *et al*., 2017), indicating that NQO1 expression was associated with tumor occurrence and that it may be a significant predictive or a prognostic marker. In a meta-analysis on colorectal cancer (CRC), the analysis of clinical parameters concerning location and tumor stage characteristics revealed a significant increase in risk for CRC for mutant genotype (Gorący *et al*., 2021).

This study evaluated the relationship between the *NQO1* polymorphism and treatment response for various chemotherapeutic regimens and their correlation with an objective response. Our finding suggests that patients with the mutant (*TT*) genotype of *NQO1* have a better CR+PR response to chemotherapy than those with wild-type genotype (*CC*). As a result, *NQO1 TT* genotype carriers were more likely to have higher sensitivity and a higher response rate to platinum therapy. Tian *et al*. (2014) found that the *TT* genotype carriers had a lower chance of responding to chemotherapy than *CC* genotype carriers.

Using survival analysis, we investigated the correlation of *NQO1* polymorphism with OS in lung cancer subjects. Even after classification based on histology, the *NQO1* polymorphism did not exhibit any significant association with overall survival in lung cancer patients. A previous study by Kolesar *et al.* (2011) found that the *NQO1* mutant genotype was related to poor survival in NSCLC patients. Masroor *et al.* (2015) discovered that NSCLC subjects with the *CT* and *TT* genotypes had a worse chance of survival. Several studies have been conducted in different cancers *viz.* lung cancer (Kolesar *et al.*, 2011; Cui *et al.*, 2014; Li *et al.*, 2015), breast cancer (Yang *et al.*, 2014), uterine cervix (Hou and Lin, 2016), pancreatic adenocarcinoma (Ji *et al.*, 2017) have found that *NQO1* polymorphism or over-expression is related with cancer progression and maybe a possible biomarker for poor prognosis. Considering NQO1's critical role in preventing the generation of ROS and formation of reactive semiquinone radicals *via* redox cycling, one can hypothesize that the poor survival of heterozygous and mutant genotype carriers is the result of elevated levels of ROS, which subsequently results from an increase in genetic instability, enhanced ROS-mediated DNA damage, and further progression of cancer (Fagerholm *et al.*, 2008; Khrunin *et al.*, 2018).

This study also evaluated the relationship between *NQO1* polymorphism and OS based on chemotherapeutic regimens. The majority of cancer patients with a similar form of cancer and stage received the same treatment. Only a few people have been observed to benefit from treatment. Knowledge of genetic differences in patients and tumors could explain why patients' respond to drugs vary. Among all the chemotherapy regimens administered, we found that patients with lung cancer who had the mutant genotype and were treated with paclitaxel cis/carboplatin had a poor OS of 3.57 months, compared to 12.20 months in those with the wild (CC) genotype (**p=0.0098**). In such patients, Cox regression analysis depicted a higher hazard rate (HR'=7.95). Previous studies on NSCLC patients indicated that the *NQO1* polymorphism played an essential role in clinical outcome in patients who underwent radiotherapy after surgery and platinum-based chemotherapy (Tian *et al.*, 2014; Song *et al.*, 2010). The loss in NQO1 activity may increase cancer riskby reducing antioxidant activity in the cell and increasing the probability of one electron reduction of quinones to semiquinones which may attack proteins, DNA, and lipids (Cui *et al.*, 2015).

6.2 *SULT1A1* polymorphism and its relationship with lung cancer susceptibility

The phase II detoxification enzyme SULT1A1 is critical in detoxifying carcinogens in cigarettes and hazardous chemicals (Wang *et al.,* 2008). It has been proposed that polymorphism in the genes that code for carcinogen metabolizing enzymes may influence an individual's susceptibility to cancer initiation. The subjects with the mutant (*His/His*), heterozygous (*Arg/His*), or combined (*Arg/His+His/His*) genotypes did not have an increased risk of lung cancer in the current study. Our data support the findings of Ihsan *et al*. (2011), who found that the mutant (*His/His*; OR=0.87) and heterozygous (*Arg/His*; OR=0.51) genotypes have a lower chance of developing lung cancer. In a study conducted in the Japanese population by Tamaki *et al*. (2011), subjects with the mutant genotype had a reduced risk or a protective effect towards lung cancer (AOR=0.76). On the contrary, our findings do not corroborate Pachouri *et al.* (2006), who observed an elevated risk of lung cancer in subjects carrying the *His/His* genotype (OR=1.40) in North India. Similarly, in a meta-analysis conducted by Zhu *et al.,* (2019) the *SULT1A1 Arg^{213}His* polymorphism was linked to digestive cancer, and it was also found to be linked to UADT (upper aerodigestive tract) cancer in subgroup analysis, however no link was identified between the *SULT1A1 Arg^{213}His* polymorphism and colorectal cancer. In a meta-analysis conducted by Liao *et al.* (2012) no association was found between the *SULT* polymorphism and lung cancer susceptibility. Various case-control studies were conducted to determine whether there is an association between *SULT1A1 Arg^{213}His* polymorphism and different cancer types. The *SULT1A1 His^{213}* allele is associated with an increased risk of stomach cancer (Boccia *et al.,* 2005), prostate cancer (Nowell *et al.*, 2004), breast cancer, and urothelial cancer (Rouprêt *et al.,* 2007). On the other hand, reduced risks for colorectal cancer (Nowell *et al.,* 2002), bladder cancer (Hung *et al.*, 2004, Zheng *et al.,* 2003), and lymphocytic leukemia (Sak *et al.,* 2019) have been reported with the *His^{213}* genotype. Chung *et al.* (2009) investigated the association between *SUTT1A1* polymorphisms and the risk of developing oral squamous cell carcinoma (OSCC) and discovered that the presence of *Arg^{213}His SULT1A1* polymorphism was not associated with the risk of developing oral cancer. It has been postulated that in the Caucasian population, the *SULT1A1 Arg^{213}His*polymorphism is relatively common (25.4%-36.5%) (Glatt and Meinl, 2004). This mutation alters the activity and stability of the enzyme SULT1A1. Several publications have suggested that *SULT1A1* polymorphism may play a role in

the pathogenesis of lung cancer (Arslan *et al*., 2009), urothelial carcinoma (Huang, 2009), and meningioma brain tumors (Bardakci, 2008; Huang, 2009).

Our study also looked at the relevance of this polymorphism in different histological subtypes of lung cancer. Our findings revealed no genetic predisposition towards SQCC. Our findings contradict the prior research, which found thatthe histological types were associated with an increased risk of lung cancer (Liang *et al.*, 2004). Furthermore, we observed that SQCC male patients with either the *Arg/His* or *His/His* genotype did not have a higher risk of modulating lung cancer susceptibility. We also evaluated if gender and *SULT1A1* polymorphism have any effect on lung cancer susceptibility in our study. Compared to female lung cancer patients, male subjects harboring a single copy of the mutant allele showed no connection with genetic risk for lung cancer. The earlier research by Wang *et al.* (2002) and Liang *et al.* (2004) found that gender stratification increased the risk of lung cancer.

As SULT1A1 is also involved in the bioactivation of tobacco smoke procarcinogens, and smoking is the major risk factor for lung cancer, the $Arg^{213}His$ *SULT1A1* polymorphism may be linked to lung carcinogenesis in smokers. Our findings suggest that smokers have a lower risk for lung cancer than non-smokers, but when the analysis was stratified by smoking status and pack-years, it was observed that smokers with pack-years greater than 40 who had the heterozygous (*Arg/His*) or *His/His* genotype did not exhibit a significantly reduced risk towards lung cancer. These findings suggest that the low enzyme activity variant *His/His* may helpreduce the bioactivation of procarcinogens (such as PAHs), lowering the risk of smoking-related cancers. Our findings are consistent with those of Ihsan *et al.* (2011), who discovered that the *Arg/His* genotype was associated with a lower risk of lung cancer among smokers. In the Japanese population, the variant genotype was associated with a lower incidence of lung cancer (Tamaki *et al.,* 2011). In a study conducted by Zheng *et al.* (2003), the authors found that light smokers (30 pack-years) with the His^{213} allele had a lower risk of bladder cancer than heavy smokers, but this was not statistically significant. Previous research found that current smokers with the *Arg/Arg* and *Arg/His* genotypes had a higher risk of developing lung cancer than smokers (Liang *et al.,* 2004, Wang *et al.,* 2002, Tasnim *et al.,* 2017). A study on oral cancer by da Silva *et al.*(2012) discovered that the association between cigarette smoking and oral cancer was higher in individuals with a *SULT1A1*1* isoform (wild-type genotype (*Arg/Arg*) and individuals with at

least one *SULT1A1*2* allele (*Arg/His+His/His*). Smoking and *SULT1A1*polymorphism were linked to an elevated risk of various other cancers,*viz.* urothelial cancer (Tamaki *et al.*, 2011), breast cancer (Lilla *et al.*, 2005), bladder cancer (Wang *et al.*, 2017).

The SULT1A1*2 allele frequency varies amongst ethnic groups, ranging from 5% to 32% (Glatt *et al*., 2000; Lee *et al*., 2013). The frequency of this allele in the control group in the current study was 6.55%. Whereas, in a study conducted by Bustami *et al*. (2018) on breast cancer patients, the frequency was 40.2%, greater than those reported for Turkish, Taiwanese, Chinese, and Koreans but close to those reported for Caucasians.

6.2.1. Relation of *SULT1A1* polymorphism in association with clinic-pathological parameters and prognosis

We also evaluated the role of *SULT1A1* polymorphism concerning clinical-pathological parameters such as stage, tumor size, lymph node invasion, metastasis, and treatment response for different chemotherapeutic regimens and their correlation with an objective response. However, no association was found with lung cancer susceptibility. Furthermore, using survival analysis, we also studied the association of *SULT1A1* polymorphism with OS in lung cancer subjects and found no association with lung cancer prognosis. Henceforth, even after histological stratification, the *SULT1A1* polymorphism did not significantly affect overall survival in lung cancer subjects.

This study also assessed the relationship between *SULT1A1* polymorphism and OS based on the chemotherapeutic regimen in this study. The majority of cancer patients with the same type of cancer and stage received the same treatment. The treatment has been observed to work well only in a few patients. Knowledge of genetic differences in patients and their tumor could explain this varied response to the drug in patients. There was no significant association between the OS and *SULT1A1 Arg^{213}His* polymorphism in lung cancer subjects treated with any chemotherapeutic regimens. A study conducted in breast cancer patients found that carriers of the *SULT1A1*2 /*2* genotype had a higher risk of death in tamoxifen-treated women's (Nowell *et al*., 2002)

6.3.*EPHX1* polymorphism and its relationship with lung cancer susceptibility

Microsomal epoxide hydrolase (EPHX1) is a Phase-II enzyme that participates in the general oxidative protection against a myriad of environmental chemicals and is primarily responsible for the xenobiotic activation of tobacco smoke carcinogens (Gresner *et al*., 2007). Our findings indicate no significant association between *EPHX1 Tyr^{113}His* (exon3) polymorphism and lung cancer susceptibility. Our findings are consistent with earlier studies that demonstrated no association between exon3 polymorphism and LC risk (Zhao *et al*., 2002; Cajas-Salazar *et al*., 2003; Graziano *et al*., 2009). Similarly, the study conducted by Timofeeva *et al*. (2010) in the German population verifies our observations since they found no significant association between the *EPHX1 Tyr^{113}His* polymorphism and risk for lung cancer. In another study, Wu *et al*. (2001) found no association between *EPHX1Tyr^{113}His* (exon3) polymorphism and lung cancer risk in African-Americans. However, some other studies have found an elevated risk of lung cancer for the variant genotype of *EPHX1 Tyr^{113}His* polymorphism, particularly in Asian subjects (Tan *et al*., 2014). An earlier investigation conducted in Chinese, the exon3 *Tyr^{113}His* polymorphism was associated with susceptibility to lung cancer (Persson *et al*., 1999). Similarly, in a study of 175 lung cancer patients from north India, the analysis exhibited a link between the *Tyr^{113}His* polymorphism and susceptibility to lung cancer (Tilak *et al*., 2011).

On the contrary, numerous additional studies that have looked at the exon3 *Tyr^{113}His* polymorphism in different populations, including South Indians (Ihsan *et al*., 2011), Germans (Rosenberger *et al*., 2008), Finns (Voho *et al*., 2006), Austrians (Gsur *et al*., 2003) and Spaniards (To-Figueras *et al*., 2001), found that the variant genotype was associated with a decreased risk of lung cancer. The pooled findings of a meta-analysis study that included 24 studies found no association between *Tyr^{113}His* polymorphism and lung cancer risk (Wang *et al*., 2013). In contrast, subgroup analysis of Caucasians and Asians suggested that *Tyr^{113}His* polymorphism was associated with a decreased and increased risk for lung cancer, respectively. Tan *et al*. (2014) did a meta-analysis and found no association between the *EPHX1Tyr^{113}His* polymorphism and risk of lung cancer susceptibility, and our findings corroborate with the findings of this meta-analysis. On the other hand, Yu *et al*. (2015) in 2015 conducted another meta-analysis study to evaluate the impact of *EPHX1 Tyr^{113}His* polymorphism and lung cancer association, and they discovered that the *Tyr^{113}His* polymorphic variant was a risk factor for lung

cancer susceptibility. The *Tyr*^{113}*His* polymorphic variant is a risk factor for various conditions, such as liver cirrhosis, Crohn's disease, diabetes mellitus, chronic obstructive pulmonary disease, gastric cancer, esophageal carcinoma, breast cancer, and ovarian lymphoblastic leukemia, and prostate cancer. In a recent meta-analysis conducted on ovarian cancer the results depict no statistical association between *Tyr*^{113}*His* polymorphic variant and ovarian cancer risk in both overall analysis and subgroup analysis by ethnicity and source of control (Jin, 2019).

On the other hand, the *Tyr*^{113}*His* polymorphic variant is also a protective factor against some previously described diseases, like ovarian cancer and breast cancer; therefore, its significance was not evident (El-Sherbeni and El-Kadi, 2014; Gautheron and Jéru, 2021).Thus, we could say that the results from different lung cancer subjects amongst different populations have been relatively inconclusive.

We have investigated the role of *EPHX1 His*^{139}*Arg* (exon4) polymorphism as a marker for lung cancer susceptibility and found a lack of any association of this polymorphism towards modulating the susceptibility for lung cancer. Our study is in line with earlier studies conducted in different populations *viz.* Chinese (Yin *et al*., 2001), Caucasian's (Park *et al*., 2005; Erkisi *et al*., 2010), Austrians (Gsur *et al*., 2003), Americans (Cajas-Salazar *et al*., 2003), and North Indians (Tilak *et al*., 2011), which identified no positive association between exon4 and LC risk. In an African-American study, Wu *et al*. (2001) found no association between *EPHX1 His*^{139}*Arg* (exon4) polymorphism and LC risk.

On the other hand, a study of North-East Indian lung cancer subjects indicated that the *EPHX1 His*^{139}*Arg* variant genotype was found to be strongly associated with an increased risk of lung cancer (Ihsan *et al*., 2011). Prior studies conducted in Italian and Mexican-American lung cancer patients (Graziano *et al*., 2009; Cajas-Salazar *et al*., 2003) found that people with the variant *EPHX1 His*^{139}*Arg* genotype had a 3-fold increased risk of developing lung cancer. In contrast, exon4 was associated with decreased lung cancer risk in a study conducted on Spanish lung cancer patients (To-Figueras *et al*., 2001). Both the heterozygous and combined genotypes of *EPHX1 His*^{139}*Arg* polymorphism were associated with an increased risk of lung cancer in a meta-analysis conducted by Liu *et al*. (2013). Another meta-analysis discovered that the *EPHX1 His*^{139}*Arg* polymorphism is associated with an increased risk of lung cancer susceptibility (Yu *et al*., 2015). The *EPHX1 His*^{139}*Arg* polymorphism has also been studied and was thought to be a

risk factor for illnesses such as preeclampsia, stomach cancer, isolated orofacial clefts, esophageal carcinoma, while its role in some of these diseases was debatable (El-Sherbeni and El-Kadi, 2014; Gautheron and Jéru, 2021).These variations in the correlation of *EPHX1* genotype with lung cancer are due to differences in ethnic composition among populations, differences in selection criteria of patients and controls, and sample size.

Furthermore, we investigated the association between *EPHX1* genetic polymorphisms and different histological subtypes of lung cancer, and our results indicate that the $Tyr^{113}His$ polymorphism conferred a positive association between *Tyr/His+His/His* genotype and SQCC. Our study correlates with the study carried out by Yin *et al*.(2001), who discovered a correlation between$Tyr^{113}His$ variant genotype and SQCC. Whereas in another study, the $His^{139}Arg$ was found to be associated with NSCLC (Zienolddiny *et al*., 2008), but data from the current analysis did not demonstrate any correlation between the $Arg^{139}His$ polymorphism and lung cancer risk based on histology. Park *et al*. (2005) discovered a substantial increase in adenocarcinoma risk for the high activity genotype of *EPHX1*. Similarly, Cajas-Salazar *et al*. (2003) found that subjects with high enzyme activity genotypes were more likely to develop ADCC. The frequency of SQCC in our study was 39.27%, followed by ADCC (36.36%) and SCLC (23.1%), which was comparable with previous studies conducted by Tilak *et al*. (2011), and Behera and Balamugesh (2004). This distribution of LC clinical subtypes differs from that of the Western population, where the predominant type of LC is ADCC, and this discrepancy may be attributed to variations in the properties of cigarettes, as well as changes in carcinogen inhaled doses and their interaction with genes (Hoffmann *et al*., 1997; Alberg *et al*., 2005; Subramanian *et al*., 2007). In a meta-analysis conducted by Bhaskar *et al*., (2020) for colorectal cancer the result depict that *EPHX1* $Tyr^{113}His$ and $His^{139}Arg$ polymorphisms are not associated with the susceptibility to colorectal cancer. Further in a study performed to investigate the relationship between polymorphisms in *EPHX1* ($Tyr^{113}His$ and $His^{139}Arg$ substitution) and its correlation with clinico-histopathological features in hepatocellular carcinoma (HCC) the $His^{139}Arg$ and wild-type $Tyr^{113}His$ were found to be associated with HCC (Ferreira *et al*., 2021).

Furthermore, when the results were retrieved based on the phenotype and the identical genotype for both the polymorphisms, no association towards lung cancer risk was attributed. According to Park *et al*. (2005), when participants were stratified based upon predicted activity

phenotype, an increase in lung cancer risk was observed among subjects with high activity genotypes compared to those with low genotype activity. A substantial increase in LC risk was identified for the high enzyme activity genotype based on enzymatic activity phenotype (Cajas-Salazar *et al*., 2003). Zhao *et al*. (2002) discovered that a high enzymatic activity phenotype was associated with an elevated LC risk. On the other hand, London *et al*. (2000) suggested that the predicted activity had no association among Caucasians, consistent with our findings. On similar lines, the findings of Zhou *et al*. (2001) corroborate our findings, as they found no association between *EPHX1* activity phenotypes and lung cancer risk.

Tobacco smoking is another well-known risk factor for lung cancer; thus, to examine the synergistic effect of smoking and *EPXH1* polymorphisms towards lung cancer susceptibility, we stratified our data based on smoking status to study this gene-environment interaction. Various toxic compounds such as PAHs and free radicals are found in cigarette smoke. Thus, either direct reactivity or metabolic activation, these compounds can produce reactive epoxides, and EPHX1 catalyzes the detoxification of numerous reactive epoxides. Nevertheless, EPHX1 is involved in the metabolic activation of PAHs (Fretland *et al*., 2000; Tan *et al*., 2014). Thus, in the activation and detoxification of procarcinogens, EPHX1 plays a dual role (Yu *et al*., 2014). Thus, increased mEH activity may give cancer protection due to improved detoxification or a higher risk of increased carcinogen activation (Lee *et al*., 2011). Our data indicate that the *EPHX1 Tyr^{113}His* polymorphic variant was found to be a risk factor specifically for the subjects who were smokers and carrying the single susceptible variant allele (*Tyr/His*, **p=0.01***)*, thus indicating that the slow phenotype is associated with increased risk of developing cancer linked to cigarette/tobacco smoking.

Furthermore, in a stratified analysis by pack-years, subjects with the homozygous variant (*His/His*) genotype and which were light smokers, as well as heavy smokers possessing the heterozygous (*Tyr/His*) variant genotype, were also found to be associated with the increased risk of lung cancer (**p=0.04**). Our findings correlate with an earlier study conducted in a North Indian population, which found that the gene-environment interaction increased the risk of LC by more than 4-fold and 2.3-fold for subjects who smoked and carried a single variant allele for both the *EPHX1 His^{113}His* and *Tyr^{113}His* genotypes, respectively (Tilak *et al*., 2011). In German lung cancer patients, the *EPHX1 Tyr*^{113}*His* and *His^{139}Arg* polymorphisms were similarly associated

with an elevated risk of lung cancer in the group of heavy smokers (Timofeeva *et al*., 2010). Another study found a clear relationship between *EPHX1 Tyr^{113}His*genotypes and the incidence of lung cancer among smokers (Yin *et al*., 2001). On the contrary, earlier case-control studies have found that the subjects who smoked and had both *His^{113}* variant alleles tend to have a decreased risk of lung cancer than subjects who smoked but had the homozygous wild-type *Tyr^{113}/Tyr^{113}* genotype (Voho *et al*., 2006). Heterozygous genotypes in *EPHX1 Tyr^{113}His* were associated with a 66% reduced risk in smokers compared to subjects with the homozygous mutant genotype. However, it was also reported in the Indian population that subjects who were heterozygous for the *EPHX1 His^{139}Arg* were more susceptible to lung cancer when compared to those patients who had both alleles of the mutant genotype (Ihsan *et al*., 2011), whereas our study did not reveal any such relationship. It should also be noted that no research has been conducted to determine the mechanism by which the homozygous mutant genotype, *i.e., His^{113}His,* enhances the risk of lung cancer.

Some sporadic findings indicate a lack of association between the *EPHX1 Tyr^{113}His* polymorphism and smoking towards modulating risk for lung cancer (Lee *et al*., 2002; Zhao *et al*., 2002). Furthermore, no association was found between the *EPHX1 His^{139}Arg* polymorphism and smoking (Lee *et al*., 2002; Zhao *et al*., 2002). A large American study found a substantial relationship between the predicted activity of the EPHX1 enzyme and cigarette smoking, indicating that cigarette smoking played a vital role in the association between *EPHX1*polymorphisms and the risk of lung cancer. In heavy smokers, smoking changed the risk direction than in non-smokers (Zhou *et al*., 2001). To-Figueras *et al*. (2001) did a study in Spain and found no association between predicted *EPHX1* activity and cigarette smoking. In several other studies, no difference in the association between the predicted activity of *EPHX1* and LC risk according to smoking status was observed (Benhamou *et al*., 1998; Lin *et al*., 2000; London *et al*., 2000; Lee *et al*., 2002; Zhao *et al*., 2002). Based on the findings of previous studies, including ours, it is plausible to speculate that individuals carrying either single or both mutant alleles for the *Tyr^{113}His* polymorphism might have an increased propensity to the action of carcinogens that are inhaled in the tobacco smoke. Bioactivation of carcinogens may increase, but detoxification efficiency may be limited because of low EPHX1 activity, as *His^{113}His* is associated with atleast a 50% reduction in enzyme activity. As a result, the imbalance in the

metabolism of carcinogens might explain the increased susceptibility to lung cancer reported in smokers with *EPHX1* variant genotypes.

Another factor under consideration was gender and its role in modulating the risk for lung cancer when stratified based on*EPHX1* polymorphism. In our study, male subjects harboring heterozygous genotype (*Tyr/His*) of *EPHX1 Tyr*^{113}*His* polymorphism had a higher risk of lung cancer susceptibility than female lung cancer subjects in both the dominant and co-dominant genetic models (**p=0.03**). Whereas, on the other hand, females did not exhibit any association towards risk for lung cancer. Similarly, after stratification based on gender,*EPHX1 His*^{139}*Arg* polymorphism did not reveal any association with lung cancer susceptibility. Timofeeva *et al*. (2010) discovered no association between gender and LC risk in their study. Another study found no association between gender and lung cancer risk (2001).

The haplotype approach was used to investigate the combined effect of *EPHX1* variant genotypes *Tyr*^{113}*His* and *His*^{139}*Arg* on lung cancer susceptibility. There is no significant difference in the overall haplotype distribution between cases and controls in our investigation. All the exploratory haplotypes did not show any significant effect on lung cancer susceptibility. Our findings support an earlier study conducted by Ihsan *et al*. (2011) in the Indian population, wherein no haplotype was significantly associated with lung cancer risk. In another study conducted in India, subjects with haplotype *His*^{113}*-His*^{139} had a slightly higher risk of lung cancer, but the difference was not statistically significant (Tilak *et al*., 2011). The same results were obtained in our study, wherein *His*^{113}*-His*^{139} exhibited a slight risk of LC but lacked statistical significance.

The intricacy of *EPHX1* gene regulation, its expression pattern, and the presence of functional genetic diversity suggest that individuals may differ significantly in their ability to metabolize various substrates, which could have implications for disease pathogenesis. The results show heterogeneity, the linkage disequilibrium, with allelic variations that influence overall enzyme activity, could be one cause of heterogeneity, whereas the other reasons could be differences in age, smoking, and ethnicity. Dietary nutrients such as fish oil may promote mEH and boost enzyme activity (Yang *et al*., 1993). The mutations in *EPHX1* have been linked to hyperchloremia (Zhu *et al*., 2003), preeclampsia (Zusterzeel *et al*., 2001; Laasanen *et al*., 2002) and fetal hydantion syndrome (Buehler *et al*., 1990). Despite one study that presents evidence of

the functional role of EPHX1 in the etiology of adverse anticonvulsant reactions such as diphenylhydantoin toxicity (Gaedigk *et al*., 1994) and both the SNPs were also found to be associated with a risk of craniofacial abnormalities in children of women who took phenytoin during the first trimester (Azzato*et al*., 2010). The significance of *EPHX1*in neurological diseases and cancer is one of the most recent discoveries in the field of EPHX1-linked pathophysiology. Like lung cancer, the EPHX1 low activity phenotype was a risk factor for chronic obstructive pulmonary disease (COPD) in Caucasian populations but not in Asian populations (Li *et al*., 2013). COPD is more common in smokers, and the previous discussion of EPHX1 and lung carcinogenesis suggests that EPHX1 plays a vital role in protecting the lung from environmental exposures (Václavíková *et al*., 2015).

6.3.1. Relation of *EPHX1* polymorphism in association with clinic-pathological features and prognosis

We also investigated the role of *EPHX1* polymorphism based on clinical-pathological features such as stage, tumor size, lymph node invasion, and metastasis. Our study found that patients harboring *His/His* genotype had a nearly two-fold (**p=0.04**) more significant risk of lymph node invasion. Our study corroborates earlier studies where lung cancer patients had a 2-fold increased odds of lymph node invasion in Turkish lung cancer subjects (Erkisi *et al*., 2010). Thus, we suggest that the mutant genotype increases the likelihood of lymph node metastasis.

This study evaluated the prognostic effect of two functional polymorphisms in the *EPHX1* pathway in lung cancer patients treated with platinum-based doublet chemotherapy. We discovered that the *EPHX1 $His^{139}Arg$* played a role in the prognosis of lung cancer. To the best of our knowledge, this is the first study conducted to investigate the effect of *EPHX1* polymorphisms on the prognosis of lung cancer patients receiving platinum-based chemotherapy. It has been postulated that mEH might have a role in suppressing tamoxifen response and causing poor prognosis in patients with breast cancer (Lin *et al*., 2007). Our findings suggest that patients with a single copy of the *$His^{139}Arg$* allele had a better prognosis than subjects carrying the wild-type genotype. Furthermore, based on the chemotherapeutic regimen, the patients who received cisplatin/carboplatin along with irinotecan and had the heterozygous genotype (*Tyr/His*) of *EPHX1 $Tyr^{113}His$* gene had a longer survival time (p=**0.01**) and hence a better prognosis. Lin

et al. (2007) discovered a significant difference in patient survival when mEH expression and Adriamycin chemotherapy were administered.

6.4. *NAT2* polymorphism and its relationship with lung cancer susceptibility

Genetic diversity in metabolic detoxification activities may modify the risk of lung cancer associated with exposure to endogenous or exogenous xenobiotics compounds (Liu *et al*., 2015). As a result, the relevance of *NAT2* polymorphisms to the risk of lung cancer is therefore of particular importance. The present study focused on the influence of genetic variants in the *NAT2* gene towards modulating the risk for lung cancer and the impact of the *NAT2* gene polymorphisms on lung cancer patients receiving platinum-based chemotherapy. Several previous studies on the *NAT2* polymorphic variants of this gene have yielded inconsistent results (Cascorbi *et al*., 1996; Oyama *et al*., 1997; Chen *et al*., 2006; Lee *et al*., 2009; Sobti *et al*., 2009). These inconsistencies may be attributed to differences in selection criteria of patients and controls, differences in ethnic composition among populations, and sample size. Smoking and exposure to other environmental pollutants are other major confounding factors.

Data from our study indicate that there was no significant association between the *590G>A (NAT2*6)* and *803A>G (NAT2*5C)* polymorphisms and lung cancer susceptibility. Our findings are consistent with the earlier studies conducted, wherein no association was found between these two *NAT2* polymorphisms and LC risk (Martinez *et al*., 1995; Bouchardy *et al*., 1998; Hansen *et al*., 2001; Zhou *et al*., 2002; Borlak and Reamon-Buettner, 2006). Perera *et al*. (2006) also found no association between the *590G>A (NAT2*6)* and *803A>G (NAT2*5C)* polymorphisms and lung cancer risk in Caucasians. Other studies in Caucasians with lung cancer found that these two polymorphic variants of the *NAT* gene did not appear to modify the risk for lung cancer (Wikman *et al*., 2001). However, on the contrary, some studies suggest that *NAT2* polymorphisms might play a role in lung cancer risk modulation. Rusinova *et al*. (2014) discovered that the *NAT2*6 590G>A* variant genotype affected lung cancer risk. Slow NAT2 acetylators have a higher risk of lung cancer, according to studies conducted in both Mongolian and Chinese Han participants. Slow acetylators for *NAT2*5* and *NAT2*6* have also been associated with an increased risk of lung cancer in the North Indian population, according to Sobti *et al*. (2009). Cui *et al*. (2011) and Liu *et al*. (2015) found no association between the slow acetylator phenotypes of NAT2 polymorphism and lung cancer risk in meta-analysis data.

In comparison to the other polymorphic variants of the *NAT2* gene, our results reveal that the *857G>A (NAT2*7)* polymorphism is associated with an increased risk of lung cancer (p=**0.005**). Our data corroborates with an earlier study conducted by Gemignani *et al*. (2007), who discovered a risk of lung cancer in subjects harboring heterozygous genotype for the *NAT2 857G>A* polymorphism. Similarly, in a meta-analysis, Cui *et al*. (2011) discovered an association between the *NAT2* polymorphism and lung cancer susceptibility.

In the case of the *NAT2 481C>T (NAT2*5B)* polymorphism, our results reveal that those subjects who had both variant alleles for the *NAT2 481C>T* polymorphism (*NAT2*5B/*5B*) had a decreased risk for lung cancer risk (OR=0.50, p=**0.0003**) when compared to *NAT2*4/*4* as reference genotype.

Since it is well known that *NAT2*5* is a slow acetylator allele, and NATs can bioactivate carcinogens like arylamine to form reactive intermediates (Duval *et al*., 2016). Thus the *NAT2* slow acetylator status expresses an enzyme with impaired N-acetylation capacity, and it is plausible that this impaired acetylation may lead to reduced production of reactive intermediates, thereby preventing DNA damage and carcinogenesis. As a result, we believe that slow acetylators have a reduced risk of lung cancer in our population, and it is plausible that slow acetylators have some protection against the disease. The above findings might provide insights about the association why lung cancer was observed among the *NAT2*5* slow acetylators and not among patients carrying other *NAT2* slow acetylator alleles. It should be noted, however, that the acetylation status of the synonymous variant c.*481C>T* is not clear because it is also associated with *NAT2*12C* (c.*481C>T*+c.*803A>G*), a fast allele. Our findings support the findings of an earlier study conducted in Central Europe, in which the authors indicated that the *NAT2*4/*5* polymorphism conferred a reduced risk of lung cancer for the *481C>TNAT2* polymorphism (McKay *et al*., 2008). Cascorbi *et al*. (1996) discovered an over-representation of the fast acetylators (*NAT2*4*/4*) in German lung cancer patients *vs*. healthy subjects, suggesting the relevance of ultra-rapid acetylation capacity in lung cancer risk. This trend was also observed in the current study.

Furthermore, we investigated the association between *NAT2* genetic polymorphisms and different histological subtypes of lung cancer, and our findings suggested that the *857G>A (NAT2*7)* polymorphism or intermediate acetylators conferred a positive association between

*NAT2*4/*7* genotype and ADCC (OR=1.81, p=**0.001**). Our findings agree with an earlier study by Seow *et al*. (1999) in Chinese lung cancer subjects in Singapore, which found a 2-fold increase in the risk of ADCC for the *NAT2*7* polymorphism. As previously stated, slow acetylators have a decreased risk of lung cancer susceptibility, and when we evaluated this according to histological subtypes, the *481C>T* polymorphism had a decreased risk for SQCC (OR=0.39, **p=0.0006**) and SCLC (OR=0.50**, p=0.001**) subjects. The slow acetylators were found to have a decreased risk of SQCC in a study of Chinese lung cancer patients (Tian *et al*., 2014).

However, our findings do not support an earlier study conducted in North Indian lung cancer patients by Sobti *et al*. (2009), who found an increased risk for SQCC, SCLC, and ADCC. The inconsistency could be due to the small sample size. In particular, our study included 550 cases and 550 controls, whereas Sobti *et al.* (2009) only studied 151 cases and 151 controls. Habalova *et al*. (2005) discovered that *NAT2*5B/*6* conferred a risk for SQCC in a study of Slovakians. In the case of both the *590G>A (NAT2*6)* and *803A>G (NAT2*5C)* polymorphisms of *NAT2* genotype, no association was found between any genotype and the different histological subtypes of lung cancer. Our findings are consistent with several previous studies conducted wherein no association was observed between the *NAT2* polymorphism and different histological subtypes (Saarikoski *et al*., 2000; Zhou *et al*., 2002; Skuladottir *et al*., 2005; Zienolddiny *et al*., 2008). Seow *et al.* (2011) discovered a marginal increase in the risk of ADCC in the Chinese population. Similarly, Chiou *et al*. (2005) discovered an increased risk of ADCC and SQCC in subjects with a fast acetylator phenotype. In another study conducted in a Caucasian population, no association was found between *NAT2* genotype and histological subtypes (Perera *et al*., 2006). Cui *et al*. (2011) found no significant association between *NAT2* genotypes and histological subtypes in a meta-analysis. All the meta-analysis studies failed to show an overall significant association with *NAT2* acetylation genotypes in lung cancer. Most studies that have evaluated the *NAT* polymorphism and lung cancer risk have shown a lack of consistency in either the design, analysis and exposure variables that have been studied, which might have resulted in discrepancies in the results. It might be possible that the influence of the *NAT2* genotypes may pose a risk for lung cancer when combined with other metabolic genes. Wikaman *et al*. (2001) discovered that combining slow acetylators of *NAT2* with rapid acetylators of *NAT1* increased the risk of ADCC of the lung. It has been shown that when the slow genotype of *NAT2* was combined with the null genotype of *GSTM1,* it may act as a catalyst,

conferring increased susceptibility to DNA adduct formation, gene mutations, and lung cancer susceptibility. Similarly, the combination of both *NAT2* slow acetylators and *CYP1A1* rapid acetylator were at the highest risk for lung adenocarcinoma.

Tobacco smoking is another well-known risk factor for lung cancer; thus, to evaluate the synergistic effect of smoking and *NAT2* polymorphisms in lung cancer susceptibility, we stratified our data based on smoking status to study the gene-environment interaction. Cigarette smoke contains a variety of toxic compounds such as PAHs and free radicals. It appears that the gene-environment interaction between *NAT2* and smoking might play a significant role in the increased risk, especially the cumulative exposure to cigarette smoke over time. Our findings show that subjects who smoked and had the mutant genotype for the *481C>T (NAT2*5/*5B)* polymorphism or were slow acetylators had a decreased risk of lung cancer (p**<0.0001**). Furthermore, when stratified by pack-years, our results revealed a decreased risk of lung cancer in both light (OR=0.44, p=**0.002**) and heavy smokers (OR=0.28, **p=0.001**). However, this association was more pronounced in the case of heavy smokers. After stratification analysis by smoking status and pack-years, the *NAT2 590G>A (NAT2*6)* and *803A>G (NAT2*5C)* polymorphisms did not demonstrate any association with lung cancer susceptibility. However, there was no relationship between smoking and lung cancer for the *857G>A (NAT2*7)* polymorphism, whereas non-smokers had a 2-fold risk of lung cancer for the intermediate acetylators. An earlier study in the Indian population found that slow acetylators for both *NAT2*5* and *NAT2*6* polymorphisms had a 2.8-fold and 3-fold risk of lung cancer in smokers and non-smokers (Sobti *et al*., 2009). Seow *et al*. (1999) found no association between the *NAT2* polymorphism and smokers among Chinese lung cancer patients. According to Sørensen *et al*. (2005), the *NAT2* rapid acetylator phenotype appears to protect against lung cancer in light smokers. A Japanese study found that the risk of lung cancer was significantly increased among light smokers who were intermediate-slow *NAT2* acetylators (Osawa *et al*., 2007). Slow acetylators increase the risk of non-smokers, while rapid acetylators increase the risk of smokers, according to Nyberg *et al*. (1998). Smokers with *NAT2* slow acetylator had a 1.84-fold risk of lung cancer in another study conducted in the Mongolian and Han populations (Zhang *et al*., 2014). Furthermore, the *NAT2* polymorphism did not appear to modify lung cancer risk based on smoking status in Caucasian or Taiwanese populations (Wikman *et al*., 2001; Lee *et al*., 2009).

Our study is the first to report a statistically significant interaction between *NAT2* genotype and smoking in terms of lung cancer modulation; *NAT2* slow acetylator genotypes (compared to fast acetylator genotypes) were found to play a modulatory role in reduced risk for lung cancer in smokers. In a previous study, cigarette smoking was not a risk factor for breast cancer among rapid acetylators (Hunter *et al.*, 1997). This *NAT2* genotype-smoking interaction may partly explain the varied and inconsistent results reported in previous studies, based on variation in baseline smoking characteristics across different study populations. Our findings reveal that the slow acetylation genotype was strongly associated with decreased lung cancer risk overall. Further, the slow acetylation genotype conferred enhanced protectivity to smoking. Secondly, all the subjects in our investigation were bidi smokers (type of Indian cigarette); thus, the differences in the metabolites produced that modify the activity of the enzymes involved in biotransformation reactions might also have contributed to the variation observed when compared to other studies. Tobacco smoking causes the production of heterocyclic amines; however, it is known that the metabolism of inhaled heterocyclic amines is different in the lung tissue as compared to those metabolized by the liver. These heterocyclic amines are activated not only by the N-hydroxylation of the CYP1A2 enzymes but also by the O-acetylation of the NAT2 enzyme and can bind to form DNA adducts, which might act as potential carcinogens. Thus, high CYP1A2 activity has increased lung cancer risk among non-smokers, but this effect was only seen when high CYP1A2 activity was combined with slow NAT2 activity. Data from a previous study has validated that the risk of lung cancer is associated with the synergistic role of both high CYP1A2 activity and NAT2 intermediate-slow acetylator activity (Dik *et al.*, 2013).

Gender was considered in terms of modulating the risk for lung cancer when stratified based on *NAT2* polymorphism. In the case of *590G>A (NAT2*6)*, polymorphism variant genotype in the case of males and females had a decreased risk of lung cancer susceptibility. The male subjects carrying the single mutant allele *(NAT2*4/*7)* for the *857G>A (NAT2*7)* polymorphism exhibited an association towards lung cancer risk. Whereas, in the case of females, 2.59 folds and 3.66 folds increased risk of lung cancer development was observed in subjects with intermediate *(NAT2*4/*7)* and slow acetylator *(NAT2*7/*7)* for the *857G>A (NAT2*7)* polymorphism. In contrast, in males, this polymorphism indicates a reduced risk for lung cancer. On the other hand, *803A>G (NAT2*5C)* polymorphism was associated with a 2.82 folds increased risk of lung cancer in the case of female subjects who were slow acetylators or

had the mutant genotype. Our data corroborates with an earlier study conducted on Chinese women in Singapore, where slow acetylator was associated with an increased risk of lung cancer (Seow*et al*., 1999). In contrast, in another study conducted in females of Taiwan, the results depict that fast acetylator was associated with an increased risk of lung cancer (Chiou *et al*., 2005). Another study conducted by Mota *et al*. (2015) found that the *NAT2* slow acetylator phenotype in males is associated with an increased risk of lung cancer susceptibility. In a meta-analysis by Cui *et al*. (2011), no significant association was found based on gender.

In a meta-analysis including 54 case-control studies the slow genotype in *NAT2* polymorphism was obviously associated with bladder cancer risk. On basis of subgroup analysis Caucasians, Asians, smokers, non-exposed individuals, high grade bladder cancer (HGBC) patients, and muscle-invasive bladder cancer (MIBC) patients all had a substantial risk with the slow genotype. Furthermore, the intermediate *NAT2* genotype was found to enhance the incidence of bladder cancer in Asians and transitional cell carcinoma (TCC) patients. In Africans, however, there was no association with the *NAT2* polymorphism (Song *et al*., 2020). In a study conducted by Zhu *et al*., (2020) to assess the effect of *NAT2* polymorphisms on risk of lung cancer and to investigate its association with smoking. The results depict no significant correlation between *NAT2* polymorphism and lung cancer. Further in stratified analysis of smoking, ethnicity, gender, no significant results were obtained. However, the subgroup analysis revealed that NAT2 individuals with intermediate-slow acetylation may have a higher risk of lung cancer (Zhu *et al*., 2020). In a systematic review and meta-analysis for *NAT2* polymorphism the results depict that patients with the non-rapid (slow+intermediate) phenotype had a considerably higher risk of lung cancer. Furthermore, when compared to slow or intermediate *NAT2* phenotypes, the rapid phenotype was associated with a lower risk of lung cancer (Zhu *et al*., 2021).

We performed haplotype analysis to enhance the effect of the combinations between the four SNPs of the *NAT2* gene. To the best of our knowledge, no reports have scrutinized the haplotype analysis of these four SNPs together. This is the first report giving us an insight into the four SNPs together in Indian data. The overall haplotype distribution differs significantly between cases and controls in our study. The Hap block 2 confers a protective effect towards lung cancer with either mutant allele for *NAT2 590G>A, 857G>A, 481C>T* and wild type for

NAT2 803A>G (p=**0.0002**). Similarly, Hap block 10 also exhibits decreased lung cancer risk and has either wild genotype for *NAT2 590G>A* or *857G>A*. The allele from other polymorphic sites was the mutant type. On the contrary, Hap5 showed an increased risk of lung cancer.

Furthermore, MDR and CART tools were applied, and some important conclusions regarding the contribution of all germline variants understudy became clear. These nonparametric techniques have lately gained importance because they provide a clear picture of the various factors that may prevent cancer predisposition. To identify potential gene-environment interactions, we have used both CART and MDR methods. Thus, using these methods can reduce the chances of making type I errors and improve the statistical power. MDR is used to determine the best interaction model within SNPs. In our study overall, (*857G>A(NAT2*7), 481C>T(NAT2*5), 803A>G(NAT2*5)*) was sown to be the best model, but it was not significant. Based on the CART results, *481C>T* polymorphism was the most significant factor determining lung cancer risk. Furthermore, terminal node 1 with *857G>A*(M) had a 2.20-fold increased risk of lung cancer (p=**0.00002**). When CART analysis was performed on histological subtypes, the ADCC population revealed that *NAT2 481C>T* was the individual risk factor for the disease (**p=0.0007**). In SQCC, the *803A>G*(W) or rapid acetylators had a 2-fold increased risk of lung cancer. Thus, results exhibit that the *NAT2* variants alone did not depict any association towards lung cancer susceptibility, but an increased risk towards lung cancer susceptibility was observed by the SNP interaction using CART.

6.4.1. Relation of *NAT2* polymorphism with clinic-pathological features and prognosis

We also looked into the significance of *NAT2* polymorphism in clinical-pathological features such as stage, tumor size, lymph node invasion, and metastasis. Our investigation found that lung cancer patients with the single variant alleles (*NAT2*4/*6*) for the *NAT2 590G>A (NAT2*6)* polymorphism had a decreased risk for lymph node invasion in both the co-dominant model as well as the dominant model. The *803A>G (NAT2*5C)* polymorphism was found to have a 3-fold odds of lymph node invasion in subjects with the mutant genotype (*NAT2 *5C/*5C*) using the co-dominant model, a 2.8-fold odds of lymph node invasion in the recessive model, and a 1.78-fold odds of lymph node invasion in the dominant model. In one study conducted by Zhou *et al*. (2002) in the Caucasian population, no significant association was

found when disease stages were compared. However, no other study has been conducted to assess the role of clinic-pathological features and their association with *NAT2* polymorphism.

This study also evaluated the prognostic effect of the functional polymorphisms in the *NAT2* pathway in lung cancer patients treated with platinum-based doublet chemotherapy. Patients who were administered with carboplatin/cisplatin and paclitaxel and had a single as well as both the variant alleles for the *857G>A (NAT2*7)* polymorphism exhibited a poor prognosis (MST=2.23, p=**0.002**) as compared to the patients who were fast acetylators. Furthermore, the lung cancer patients who were slow acetylators (*NAT2 *6/*6*) for *590G>A (NAT2*6)* polymorphism and who were given cisplatin/carboplatin and irinotecan had a longer survival time and a significantly reduced death rate when compared to patients who were fast acetylators (MST=11.20, p=**0.01**). All taken together, the *NAT2 590G>A (NAT2*6)* polymorphism appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy, as the survival time for such patients was better. To the best of our knowledge, this is the first study to investigate the effect of *NAT2* polymorphisms on the prognosis of lung cancer patients with platinum-based chemotherapy.

6.5. *MTHFR* polymorphism and its relationship with overall survival of lung cancer patients and toxicity

Both genetic and environmental factors have been identified to have a significant role in the development and progression of LC; however, the precise mechanism of both these factors is not well elucidated (Wheeler *et al*., 2013). Our findings imply that patients with the wild-type alleles for the *MTHFR A1298C* gene had less time than patients carrying either a single or both mutant alleles for the *MTHFR* gene. The *MTHFR A1298C* polymorphism was an independent prognostic factor associated with a poor prognosis in a German gastric cancer patient study (Blank *et al*., 2014). The *MTHFR C677T* polymorphism did not affect survival in patients carrying mutant genotypes. At present, the association between *MTHFR* SNPs and 5-FU efficacy is inconclusive, and few studies have been conducted to investigate the relationship between *MTHFR* SNPs and pemetrexed efficacy. Jakobsen *et al*. (2005) discovered that patients with the *MTHFR TT* genotype for *C677T* polymorphism had a considerably greater response rate than two other genotype groups in metastatic colorectal cancer patients treated with 5-FU. On the

other hand, several large-scale clinical trials have failed to find a direct correlation (Boige *et al*., 2010; Ruzzo *et al*., 2006; Goekkurt *et al*., 2009).

Furthermore, a significant meta-analysis discovered that *MTHFR C677T* is not a strong predictor of FU-based therapy efficacy (Zintzaras *et al*., 2009). Patients with an *MTHFR TT* homozygous mutation exhibited a more prolonged progression-free survival (PFS) than patients with wild-type or heterozygous NSCLC who were treated with pemetrexed, according to Smit *et al*. (2009). In contrast, Argiris *et al*. (2011) published contradictory findings. Kucharczyk *et al*. (2016) discovered minor (but not statistically significant) variations in OS among carriers of the *MTHFR C677T* polymorphism. A breast cancer study found that the variant allele (*CC*) of *A1298C* had reduced survival compared to the wild genotype (*AA*), whereas variant allele of *C677T* polymorphism had improved survival (not significant) (2006). The correlation of *C677T* polymorphism with OS in glioblastoma multiforme (GBM) did not show any therapy-specific effect (Linnebank *et al*., 2008). The *TT* genotype was found to have lower OS than the wild genotype in patients younger than 60 years and have poor prognosis in patients with a post-operative karnofsky score of ≥70. According to Faria *et al*. (2020), patients with recurrent glioblastoma who had the mutant (*TT*) genotype for *C677T* polymorphism and were undergoing perillyl alcohol (POH) based therapy had a longer survival time as compared to those patients harboring the wild genotype (*CC*).

Pemetrexed is becoming more widely utilized as first and second line and maintenance therapy for advanced NSCLC patients with "non-squamous" (*i.e*., ADCC) histology. However, unlike other chemotherapeutic agents, no known predictive criteria other than histology can separate a subset of patients who may benefit from pemetrexed more than others. In our study, the subjects were given cisplatin/carboplatin alongside pemetrexed, and the *MTHFR C677T* polymorphism did not affect OS, whereas the heterozygous and mutant genotype of *MTHFR A1298C* polymorphism had an increase in OS, but it was not significant. Our findings do not support the findings of Tiseo *et al*. (2013), who discovered a significant association between the *TT* genotype (*C677T*) and increased OS in NSCLC patients treated in second-line with pemetrexed or pemetrexed/carboplatin. Scagliotti *et al*. (2008) found that patients who received cisplatin plus pemetrexed had a longer OS duration than those who received cisplatin and gemcitabine in phase III clinical trial. Li *et al*. (2013) discovered no association between

the *MTHFR C677T* polymorphism and pemetrexed efficacy in advanced ADCC LC subjects. The study of two clinical trials for polymorphisms associated with pemetrexed efficacy, GOIRC 02–2006 (Ardizzoni *et al*., 2012) and NVALT-7 (Smit *et al*., 2009), revealed that the *TT* genotype of *MTHFR C677T* polymorphism is associated with longer PFS and OS as compared to other genotypes (*CC* and *CT*). According to Krawczyk *et al*. (2014), the *MTHFR C677T* polymorphism did not affect OS or PFS. In a study of patients with advanced chronic kidney disease (CKD) and end stage renal disease (ESRD), Jamison *et al*. (2009) discovered that patients with the *TT* genotype of *C677T* polymorphism were associated with worse survival, whereas the *CC* genotype of *A1298C* polymorphism was associated with improved survival in patients with ESRD, but not patients with advanced CKD. A Dutch study indicated that the *MTHFR C677T* polymorphism was related to outcomes in which patients carrying the *TT* genotype had inferior progression-free survival in HER2-negative gastric cancer patients treated with bevacizumab along with docetaxel, oxaliplatin, and capecitabine (Meulendijks *et al*., 2016).

We also investigated the role of clinic-pathological features concerning *MTHFR* polymorphisms and their impact on the OS of lung cancer subjects. However, no statistically significant correlations were observed. Our study supports an earlier study by Takehara *et al*. (2005), who investigated the impact of clinic-pathological features of *MTHFR C677T* polymorphisms and their role in the OS of LC subjects but found no statistically significant correlation. He did, however, identify a novel relationship between TS available type, *MTHFR* genotype, and prognosis in stage I NSCLC. In this particular disease level, patients in the TS H-group (2R/3G, 3G/3C, or 3G/3G) had more prolonged survival than those in the L-group (2R/2R, 2R/3C, or 3C/3C) when confined to the *MTHFR* T-group (*T/T* or *C/T*), but patients in the TS L-group outlived those in the H-group when limited to the *MTHFR* C-group. These findings indicated that the *MTHFR* genotype could influence the effect of the TS functional form on NSCLC tumor biology. Since *TS* and *MTHFR* compete for CH_2FH_4 in folate metabolism, this gene-gene interaction can be connected to the folate status of NSCLC tissue. Tang and colleagues (2018) investigated the impact of the *C677T* polymorphism in patients with gastric cancer receiving 5-fluorouracil (5-FU)-based chemotherapy. There was no correlation found between the *C677T* polymorphism and response rate.

This study also examined whether polymorphisms in the *MTHFR* gene were associated with increased toxicity in individuals receiving pemetrexed-platinum-based chemotherapy for lung cancer. When comparing intermediate/severe toxicity (grade 1-3) to the lack of any toxicity (grade 0), heterozygous genotype (*CT*) of the *MTHFR C677T* polymorphism had a 5.34-fold increased risk of developing neutropenia toxicity. When patients with low-grade toxicity (WHO grades 1) were compared to those with intermediate-high grades (WHO grades 2-3), patients with the *AC* (heterozygous) genotype and combined genotype (*AC+CC*) for *MTHFR A1298C* polymorphism had an increased risk of diarrhea after adjusting for confounding factors like age, gender, and performance status. When the subjects with low-intermediate grade toxicity (WHO grades 1-2) were compared to those with high grades (WHO grades 2-3), those harboring heterozygous genotype for *MTHFR A1298C* polymorphism had an increased risk for diarrhea. Our study corroborates an earlier study conducted where neutropenia was detected in mutant and heterozygous metastatic patients for *MTHFR C677T* polymorphism in patients with colorectal cancer (Polat *et al*., 2018). On 292 colorectal cancer patients, Lee *et al*. (2013) discovered that *677C>T* mutant genotype is related to a high risk of neutropenia. Yeh *et al*. (2017) discovered a strong link between the *MTHFR 677C>T*SNP and neutropenia. *MTHFR 677C>T* SNP was related to an increased incidence of diarrhea in another investigation by Chua *et al*. (2009) in colorectal cancer patients.According to the study, patients with the *AC/CC* genotype exhibited a higher frequency of leucopenia, neutropenia, fatigue, and nausea than patients with *CC* genotype in non-squamous NSCLC patients treated with pemetrexed, according to a study by Lan *et al*. (2017). Six patients in the mono-institutional community of pre and postmenopausal women with early breast cancer treated with classical CMF (cyclophosphamide, methotrexate, and fluorouracil) developed severe toxic effects (grade 4 according to NCI-CTC criteria) during the first cycle of chemotherapy (Lan *et al*., 2017). Toffoli *et al*. (2003) discovered a significant association between toxicity and *TT* genotype of *MTHFR C677T* polymorphism in another research of 43 ovarian cancer patients(12 with G3/4 WHO toxicity (6 were G3/4 thrombocytopenia, 1 with G3 neutropenia, 1 with G3 anemia, 9 were G3 `mucositis and 1 G4 mucositis cases)). Using CTCAE G3/G4, the patients with *MTHFR C677T* genotype *CC* had significantly higher liver toxicity. Nearly half of the patients hadG3/G4 hematologic toxicity (leukopenia, neutropenia), and mutant genotype had an increased probability of having G3/G4 hematologic toxicity. In a study conducted by Smit *et al*. (2009), toxicity was observed between

two arms, *i.e.*, patients treated with pemetrexed plus carboplatin, and the second were the patients treated with pemetrexed. Neutropenia, thrombocytopenia, leukocytopenia, and anemia were common in the combination arm, and the former two were significant (p=**0.002** and **<0.001**, respectively).

Tang *et al*. (2018) discovered that gastric cancer patients with the mutant genotype (*TT*) for *C677T* polymorphism experienced more hematologic toxicity than those with the *CC* or *CT* genotype. It has been discovered that the *MTHFR* gene is associated with the metabolism of 5-FU (Zhao *et al*., 2015). As a result, subjects carrying the mutant genotype for *C677T* polymorphism reduce MTHFR activity, resulting in inhibition of thymidylate synthase (TS) and increased 5-fluorouracil efficiency. Because 5-fluorouracil is cytotoxic, this could increase patient toxicity (Tang *et al*., 2018).

Overall toxicity was more common in patients with the *TT* genotype when treated with methotrexate for treatment of rheumatoid arthritis, but the proportion of patients who had toxicity did not differ across genotypes for the *A1298C* polymorphism (Urano *et al*., 2002). According to the studies mentioned above, the variation in the *MTHFR* genotype is associated with toxicity among the patients treated using chemotherapy. Patients who were negative for Helicobacter pylori infection and had the mutant genotype for *C677T* had a greater risk of developing a severe form of atrophic gastritis, a precursor lesion for gastric cancer (Kong *et al*., 2020).

Inter-individual differences in platinum-based toxicity might be due to *MTHFR* polymorphism, as *C677T* polymorphism results in an amino acid change of valine for alanine, reducing MTHFR's activity and thermal stability and *MTHFR A1298C* (*E429A*) polymorphism, may increase serum folate levels, potentially influencing the risk of cancer (Krajinovic *et al*., 2004; Parle-McDermott *et al*., 2016; Yang *et al*., 2016). Extracellular folate levels have been found to inhibit pemetrexed activity in human lung and colon cancer cell lines. Pemetrexed is amultitarget antifolate drug that inhibits purine and pyrimidine synthesis enzymes, while MTHFR is indirectly involved in purine and pyrimidine production (Chattopadhyay *et al*., 2007).Our study, which was undertaken in the North Indian lung cancer patients undergoing platinum chemotherapy, is conducted in the Indian scenario. This study adds to our understanding of MTHFR's potential role as a possible predictive marker for patients treated with

pemetrexed and platinum-based therapy for ADCC. The overall gastrointestinal toxicity was shown to be more common in individuals with heterozygous genotype for the *A1298C* polymorphism, and hematological toxicity such as neutropenia was found to be more represented in patients with a single copy of the variant allele for the *C677T* polymorphism. However, our data did not find any significant association for predicting the survival outcome and chemotherapeutic response for such patients.

6.6.*GSTP1* polymorphism and its relationship with overall survival of lung cancer patients and toxicity

The standard first-line regimen for lung cancer patients is platinum-based doublet chemotherapy; however, the efficacy of platinum-based chemotherapy has been observed to vary among individual patients (Tiseo *et al*., 2013). According to published data, the *GSTP1 IIe^{105}Val* variant might be associated with the efficacy of platinum-based chemotherapy in a lung cancer patient. Our data suggest that patients carrying the wild-type alleles for the *GSTP1 IIe^{105}Val* had a poor survival time than patients carrying either a single or both mutant alleles for the *GSTP1* gene. When lung cancer patients were stratified based on histological cell types, it was visible that ADCC patients had a better prognosis. According to Booten *et al.* (2006), the *GSTP1 IIe^{105}Val* polymorphism is correlated with NSCLC patients' survival. A Chinese study found that *GSTP1 IIe^{105}Val* influences the clinical outcome of NSCLC patients (Liu *et al*., 2015). Similarly, Deng *et al.* (2015) concluded that NSCLC patients harboring the mutant genotype of *GSTP1 IIe^{105}Val* polymorphism had a favorable recurrence-free survival of advanced NSCLC. Jia *et al.*(2016) discovered that the advanced NSCLC patients with mutant genotype (*Val/Val*) had a better OS and clinical response. However, some studies which have investigated the association between *GSTP1 IIe^{105}Val* polymorphism and the prognosis of advanced NSCLC have reported inconclusive results (Ada *et al*., 2010; Zhou *et al*., 2011; Ke *et al*., 2012; Lv *et al*., 2014; Yang *et al*., 2014; Deng *et al*., 2015; Han *et al*., 2015; Liu *et al*., 2015). It has been observed that Asian lung cancer patients who carried the *Val* allele of *GSTP1 IIe^{105}Val* polymorphism had a better survival time and clinical response than patients with the wild type genotype (Ada *et al*., 2010; Zhou *et al*., 2011; Ke *et al*., 2012; Lv *et al*., 2014; Han *et al*., 2015). Others have demonstrated that the *Val* allele of *GSTP1 IIe^{105}Val* is associated with poor survival and clinical response (Deng *et al*., 2015; Liu *et al*., 2015). On the contrary, some studies have reported contradictory

results. Lu *et al.* (2006) conducted a study in a Chinese population, and Ada *et al*. (2010) conducted a study in a Turkish population, and they reported that the *GSTP1 Ala^{114}Val* polymorphism, but not *GSTP1 IIe^{105}Val*, was associated with improved OS among patients with advanced NSCLC. Furthermore, individuals with the mutant form of the *GSTP1* gene have a reduced ability to detoxify drug metabolites, thus promoting the overall survival of NSCLC (Lu *et al*., 2006; Ke *et al*., 2012; Lv *et al*., 2014; Deng *et al*., 2015; Han *et al*., 2015).

In addition, we have also investigated the impact of *GSTP1* polymorphism and response concerning the treatment outcome of lung cancer patients who received platinum-based chemotherapy. According to our data, none of the *GSTP1 IIe^{105}Val* genotypes could act as a significant predictor of chemotherapeutic response rate or clinical benefit. So far, data have suggested that the *GSTP1* variant allele might be associated with the efficacy and clinical response to platinum-based chemotherapy in lung cancer patients (Lv *et al*., 2014; Deng *et al*., 2015; Han *et al*., 2015; Liu *et al*., 2015; Zhao *et al*., 2015; Bu *et al*., 2016; Chen *et al*., 2016; Jia *et al*., 2016). The *GSTP1 IIe^{105}Val* polymorphism was significantly associated with clinical response in NSCLC patients by Booten *et al*. (2006). In a Chinese study, Liu *et al*. (2015) discovered that the *GSTP1 IIe^{105}Val* polymorphism affects the clinical outcome of lung cancer who have received platinum-based chemotherapy. Similarly, other studies conducted in Chinese lung cancer patients have found that the *GSTP1* polymorphism correlates with chemotherapy response(Lv *et al*., 2014; Han *et al*., 2015). A meta-analysis study has reported that the polymorphism in the *GSTP1* predicts the response to platinum-based chemotherapy in East-Asian lung cancer subjects (Yang *et al*., 2014). Vlachogeorgos *et al.* (2008) found that the *GSTP1* polymorphism may predict the response to treatment of patients with advanced NSCLC, in Caucasian subjects, especially in Greek lung cancer patients. On the contrary, some studies have conferred a lack of association between the *GSTP1 IIe^{105}Val* polymorphism and response. Kalikaki *et al*. (2009) and Ada *et al*. (2010) did not find a significant association between *GSTP1* polymorphism and treatment outcome of advanced NSCLC. According to a recent meta-analysis study, the *GSTP1 IIe^{105}Val* variant allele is an excellent clinical predictor and has a better response rate in Asian lung cancer patients who have undergone platinum-based chemotherapy than Caucasian lung cancer subjects. The inconsistent results in different populations may be due to differences in ethnicity, source of patients, and sample size.

Lung cancer is a complex disease, and different ethnicities have different genetic backgrounds, which may impact the anticancer therapeutic outcome of platinum-based chemotherapy. It has been reported that *GSTP1* mediates the detoxification of platinum-based chemotherapeutic agents, and it plays a significant role in both intrinsic and acquired resistance to platinum components. Some studies have reported that platinum resistance is not associated with *GSTP1* (Galluzzi *et al*., 2012). One plausible reason for not being able to exemplify the relationship between the expression of *GSTP1* and platinum drugs at any time might be due to the polymorphism of the *GSTP1* gene. A single nucleotide substitution (*A*>*G*) at position 313 replaces isoleucine with valine, resulting in the reduction of the enzyme activity.

Our study found that the responses to cisplatin-based chemotherapy of lung cancer patients who were either homozygous/heterozygous for the *GSTP1*variant allele were similar to the responses observed in patients harboring the wild type genotype of the *GSTP1* gene. Data has suggested that *GSTP1* develops resistance to different drugs like etoposide, Adriamycin, cisplatin, and carboplatin (Nakagawa *et al*., 1988). As a result, the*GSTP1 IIe*^{105}*Val* variant allele's beneficial effect could not be directly linked to a specific drug concerned with treatment response. It might be possible that the response for a tumor to cisplatin can also be attributed to different drug combinations. However, the benefit of different chemotherapeutic drugs shall remain undetermined until it is proved that *GSTP1* detoxifies the mentioned drugs. Another reason why the results obtained here are different from other studies conducted in different tumors, like breast, colon, and leukemia, is because the sample size of patients was different, different types of tumor, different resistance mechanisms that might be dominant, and finally, type of chemotherapy regimens applied.

In this study, we investigated whether polymorphisms in the *GSTP1 IIe*^{105}*Val* were associated with increased toxicity in patients receiving platinum-based chemotherapy-treated lung cancer patients. We discovered that patients carrying the single allele of the *GSTP1 IIe*^{105}*Val* were significantly associated with an increased risk of grade 3 or 4 hematologic toxicity and severe anemia toxicity (OR=2.14, p=**0.03**). Our data also revealed that lung cancer patients with the mutant genotype for *GSTP1 IIe*^{105}*Val* had a 2.5-fold increased risk of leukopenia (p=**0.001**) compared to patients carrying wild genotype. Based onthe histological classification,

SCLC patients harboring the mutant alleles for *GSTP1 IIe^{105}Val* had more severe hematological adverse effects, especially thrombocytopenia (OR=9.75, p=**0.01**).

The existence of alterations in the GSTP1 proteins, which reduces the transport capacity for cisplatin in patients with mutant genotype (*Val/Val)*, may explain inter-individual differences in platinum-based toxicity (Stoehlmacher *et al*., 2004). GSTP1 is a phase II detoxification enzyme that aids in the elimination of platinum-based compounds. The polymorphism in the *GSTP1* gene changes the thermal stability of the enzyme and changes the conjugation capacity. The *GSTP1*polymorphism results in the altered ability of GSTP1 to detoxify chemotherapeutic agents and hence also modulate drug responses (Allan *et al*., 2001). So, patients with mutant genotype have less enzyme activity, and hence the ability to detoxify the chemotherapeutic agent used shall be less. Hence such patients might retain more of the drug and show higher toxicity than subjects with the wild-type genotype. It has been documented that individuals carrying both the mutant alleles for the *GSTP1 IIe^{105}Val* polymorphism are more susceptible to inflammatory symptoms (Pincinato *et al*., 2019). The decrease in the activity of GSTP1 in lung cancer subjects harboring the heterozygous or mutant genotype may result in high exposure to platinum-based compounds and thus proportional increase towards drug toxicity because of reduced *GSTP1* mediated catalyzed reactions, thus leading to reduced metabolism of platinum-based drugs.

The association of *GSTP1 IIe^{105}Val* genotypes with the clinical outcomes of platinum-based chemotherapy is still controversial. Deng *et al*. (2015) reported that individuals with mutant or heterozygous genotypes had a lower risk of anemia. Zhang *et al*. (2011) reported that breast cancer patients with *GSTP1 Val/Val* genotype were more likely to develop severe toxicity. Further, the study conducted by Yoshihama *et al*. (2018) patients with the *GSTP1* wild-type genotype had higher grade 3 or 4 hematological toxicity rates than patients with the mutant *GSTP1* genotype. In a study conducted by Khrunin *et al*. (2012) on ovarian cancer patients, the clinical findings confirm that platinum-based regimens are more toxic in Asian patients than Caucasians.

Our results showed that patients with *GSTP1 Val/Val* genotype were more likely to have severe anemia, leukopenia, and thrombocytopenia toxicity. In summary, our study provides information about the role of *GSTP1* polymorphisms in outcome after platinum-based doublet therapy of lung cancer. Findings from the current study are consistent with the general

assumption that subjects expressing the variant form of the GSTP1 enzyme have a reduced capacity to detoxify platinum-based chemotherapeutic drugs, which results in lower clearance and poor efficacy. Despite the poor response, the decreased enzyme activity of GSTP1 also shows a propensity towards severe toxicity. Our study has several strengths, such as many patients who were receiving platinum-based chemotherapy were enrolled, and all the patients were treated in the same hospital. Secondly, the recruitment of patients and the clinical parameters data collection was carried out independently without prior knowledge of the *GSTP1* polymorphism.

6.7. High order gene-gene interaction and toxicity profile of *NQO1, SULT1A1, NAT2,* and *EPHX1*

The high order gene-gene interaction was carried out between the different genotypes of PhaseII detoxification genes (*NQO1, SULT1A1, NAT2,* and *EPHX1*). In our study, the different combinations of some SNPs were associated with lung cancer susceptibility. The studies have been conducted on different xenobiotics metabolizing enzymes, but none of these studies evaluate the role of these SNPs in combination. In CART analysis, overall, the primary split was by the *481C>T NAT2* polymorphism. In MDR, these genes were also found to pose an effect on lung cancer risk, showing significance. Overall the five-factor model *NQO1* $P^{187}S$, *EPHX1* $Y^{113}H$, *NAT2*6* $R^{197}Q$, *NAT2*5B* $L^{161}L$, *NAT2*5C* $K^{268}R$was the best model with a CVC of 10/10, prediction error of 0.4345, and p **<0.0001**.A study conducted by Skjelbred *et al*. (2011) found that polymorphisms of *GSTT1, EPHX1, MTHFR, MTR,* and *NAT2* affect the frequency of chromosome-type aberrations (CSAs), chromatid-type aberrations (CTAs), and chromatid gaps (CTGs), exhibit interaction with smoking and age differentially. But, these polymorphisms were not evaluated in combination. Furthermore, another study by Timofeeva *et al*. (2010) associated *MPO, GSTT1, GSTM1, GSTP1, EPHX1,* and *NQO1* polymorphisms towards lung cancer risk was evaluated in individual SNPs. Similarly, the polymorphisms of phaseI and phaseII metabolic enzymes in a study conducted by Mota *et al*. (2014) the *GSTP1* and *NAT2* polymorphisms were found significant based on histology, metastasis in males who were smokers. In a study carried in NSCLC smokers, an analysis of phaseI and phaseII genetic polymorphisms was carried out on individual SNPs, but the analysis was not carried in combination (Zienolddiny *et al*., 2008)

Further, the hematological toxicity, gastrointestinal, and nephrotoxicity were evaluated on these phaseII detoxification genes. These SNPs were evaluated for adverse events caused due to the administration of platinum-based chemotherapy to the lung cancer patients involved in this study. The *NQO1 609C>T* polymorphism was associated with decreased risk of gastrointestinal toxicity (mainly diarrhea) and anorexia in lung cancer patients. The *SULT1A1* $Arg^{213}His$ was found to be associated with decreased risk for gastrointestinal toxicity. The polymorphisms (viz. $Tyr^{113}His$ and $His^{139}Arg$) of the *EPHX1* gene were not statistically associated with hematological toxicity, nephrotoxicity, or gastrointestinal toxicity. But, the role of EPHXs in cisplatin-induced toxicity appears to be linked to oxidative stress effects. The EPHXs role is inhibitory, resulting from epoxide hydrolases' ability to metabolize epoxyeicosatrienoic acids (EETs), which have a variety of activities, including anti-inflammatory actions. The majority of the information on the correlations between EET hydrolysis and cisplatin toxicity has come from investigations on cisplatin nephrotoxicity (Parrish *et al*., 2009; Liu *et al*., 2012). It has been demonstrated that the anti-inflammatory impact of EETs is reliable on EPHX2, a cytosolic partner of EPHX1. The EPHX1 also takes up EETs, though considerably smaller than EPHX2 (Decker *et al*., 2009). As a result, a function for *EPHX1* in cisplatin toxicity should not be ruled out, especially given its high expression in the kidneys. This idea is supported by the connection between SNP rs1051740 in *EPHX1* and the nephrotoxicity of the regimen utilized (Khrunin *et al*., 2018).

The subjects carrying heterozygous genotype and combined genotype for *NAT2*6* were associated with increased risk of hematological toxicity, especially leukopenia, thrombocytopenia, and anemia. Whereas in the case of NAT2*5C polymorphism, it was associated with decreased risk for hematological toxicity, especially leukopenia and anemia, but the variant genotype was associated with increased risk for nephrotoxicity. Furthermore, *NAT2*7* polymorphism for patients carrying heterozygous genotype exhibit an increased risk for thrombocytopenia but decreased risk for anorexia and alopecia. There are very few studies conducted in different cancers to evaluate toxicity due to platinum-based chemotherapy, but no study involved all the SNPs studied here in one study.SNPs rs4986993, rs15561, and rs8190870 (*NAT1*), rs1353035 (*NAT1/NAT2*), and rs17038909 (CNTN6/CNTN4) were revealed to be linked with nephrotoxicity in the GWAS and improved the performance of the machine learning model (Garcia *et al*., 2020). Cisplatin has been shown to impair *NAT1* activity in human breast cancer cells by blocking its transferase activity and thus leading it to impair murine NAT2 activity in

cultured mouse tissues (liver and kidney) (Ragunathan *et al*., 2008), which on the one hand, contributes to the therapeutic effects of cisplatin but on the other hand, may lead to cisplatin accumulation in the kidneys. In a study where induction chemotherapy in locally advanced head and neck squamous cell carcinoma (LAHNSCC) was given, the *MTHFR* was associated with an increased risk of gastrointestinal toxicity. In a review conducted by Brensen *et al*. (2020), the toxicity of different drugs and platinum-based compounds in pediatric oncology has been discussed. The *EPHX1*hasbeen associated with a reduced risk of renal toxicity in adult patients treated with platinum compounds for pediatric oncology (Khrunin *et al*., 2014). In pediatric patients treated with cisplatin or carboplatin, a mutant in *GSTP1* was related to an increased risk of hearing loss (Liberman *et al*., 2019).

In a study on the toxicity of irinotecan-based regimes (FOLFIRI (irinotecan, leucovorin, 5-fluorouracil (5-FU)), IFL (bevacizumab, irinotecan, 5-FU, leucovorin), XELIRI (capecitabine, 3-weekly irinotecan), and irinotecan monotherapy) in patients with metastatic colorectal cancer revealed that toxicity (grades 3 and 4 diarrhea) due to FOLFIRI and irinotecan monotherapy in the non-trial patients were not statistically different from the clinical trials (Tam *et al*., 2009). In another study conducted by Panasci *et al*. (1990) found that mitomycin-C and mitoxantrone chemotherapy combination used for the treatment of breast cancer causes significant gastrointestinal toxicity or alopecia; but, the period of response to this regimen appears to be less than that obtained with CMF (cyclophosphamide - methotrexate - 5-fluorouracil) or CAF (cyclophosphamide - Adriamycin - 5-fluorouracil). Furthermore, the metastatic HCC (hepatocellular carcinoma) patients given combination treatment with doxorubicin and cisplatin demonstrated modest antitumor efficacy with some tolerable side effects. Neutropenia (14.3%), thrombocytopenia (11.9%), and diarrhea (9.5%) were the most prevalent grade 3/4 toxicity in these patients (Lee *et al*., 2004). To the best of our knowledge, toxicities due to chemotherapy in other cancers have been reported in a few studies, but the toxicity due to platinum-based compounds in lung cancer patients to these SNPs has not been studied earlier.

6.8.Strengths of the study

The strengths of the current study are a large sample size compared to the previous study done in the Indian population, which could enhance the reliability of the results. Furthermore, a particular strength of this study lies in the analysis of three endpoints, *viz*. OS, chemotherapy

response, and clinic-pathological features to the Phase II polymorphic variants. As far as our knowledge is concerned, no other study has evaluated the role of these polymorphisms to OS and platinum-based doublet chemotherapy and the combinatorial analysis of these SNPs. Our study has several strengths, including that many patients receiving platinum-based chemotherapy were enrolled, and all the patients were treated in the same hospital. Secondly, the recruitment of patients and the clinical parameters data collection was carried out independently without knowing the polymorphism's studied. The stratified analysis was also performed for smoking and histology to find out high-risk posing subgroups. Our study is the first attempt in PhaseII detoxification genes for lung cancer in the Indian population to use a multifactorial approach (CART and MDR) to explore the complex interaction between these genes and their relationship towards lung cancer risk. This is the first study that involves evaluating the association of these SNPs with chemotherapeutic-related toxicities. The present findings with a larger sample size for the overall and stratified group would help develop biomarkers both for diagnosis of lung cancer and prognosis of lung cancer. Some of the findings involve:

As the PhaseII detoxification enzymes catalyze the detoxification of carcinogens and anticancer drugs, SNPs in PhaseII detoxification genes are likely to play a role in cancer susceptibility. These polymorphisms in the PhaseII genes are worth paying more attention, for they may serve as a platform for detecting and predicting individual susceptibility to lung cancer. Second, it may aid in determining the prognosis of the disease.

6.9. Limitations of the study

The limitations of this study include: more population-based study in different population of India needs to be carried out to reduce the extent of selection bias. Our research has so far been limited to North Indian population. As genetic polymorphism frequently differs between ethnic groups, this should be expanded to diverse ethnic population. Secondly, to fulfill the personalized therapeutic system, developing comprehensive and reliable SNP and haplotype analyses of diverse ethnic populations, like Indians, shall be highly beneficial, allowing for predicted genotype-drug efficacy in each individual. But further study needs to understand the role of these genes in immunotherapy and development of oncopanels for detection. Apart from other genes these genes should also be taken into consideration.

CHAPTER 7

CONCLUSION

Salient findings from the present study

- The *NQO1Pro^{187}Ser* polymorphism exhibited a significant association between clinical-pathological features, including stage and TNM.
- *NQO1 Ser/Ser (T/T)* genotype is associated with poor OS in lung cancer patients and thus poor prognosis when further stratified for chemotherapy.
- The *SULT1A1* polymorphism revealed a decreased risk for SQCC (male subjects).
- The study on *SULT1A1 Arg^{213}His* polymorphism in the case of smokers and with pack-years of above 40 who were harboring the heterozygous (*Arg/His*) or *His^{213}* allele genotype exhibited a significantly reduced risk towards lung cancer.
- The *SULT1A1 Arg^{213}His* polymorphism when stratified for chemotherapy the above genotype exhibits a poor survival and thus poor prognosis.
- The *EPHX1Tyr^{113}His*polymorphism exhibited a two-fold increased odds of lymph node invasion. The *Tyr^{113}His* polymorphism, *Tyr/His+His/His* genotype, exhibits a significant increase in the risk of SQCC.
- In *EPHX1 Tyr^{113}His* polymorphism, the *Tyr/His* and *Tyr/His+His/His* genotype was a risk factor for smokers. Further, the *Tyr/His* and *Tyr/His+His/His* genotype was associated with the risk of LC in male subjects.
- The subjects carrying the combined genotype for *His^{139}Arg* (*His/Arg+Arg/Arg*) exhibited a better OS overall, and the heterozygous genotype exhibited a better OS in the case of SCLC. The *EPHX1 Tyr^{113}His* appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy.
- The *NAT2857G>A* polymorphism exhibited a risk oflung cancer (p=**0.005**). Whereas variant alleles for the *481C>T* polymorphism had a decreased risk for LC (**p=0.0003**).

- *NAT2*4/*7* polymorphism conferred a positive association between genotype and ADCC (p=**0.001),** and *481C>T* polymorphism had a decreased risk for SQCC (**p=0.0006**) and SCLC (**p=0.001**) subjects.
- The smokers carrying mutant genotype for the *481C>T* polymorphism had a decreased risk towards LC (**p<0.0001**) even in light (**p=0.002**) as well as heavy smokers **(p=0.001**).
- In the case of females, 2.59 folds and 3.66 folds increased risk of LC development was observed in subjects with intermediate and slow acetylator for the *857G>A* polymorphism. Whereas, in the case of males, this polymorphism depicts a reduced risk for lung cancer.
- On the other hand, *803A>G* depicted a 2.82 folds risk of lung cancer in the case of female subjects who were slow acetylators.
- The*NAT2803A>G* polymorphism a 3-fold odds of lymph node invasion were observed for mutant genotype, the recessive model exhibited an odd of 2.8.
- *590G>A* appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy as the survival time for such patients was better.
- Based on toxicity profiling, lung cancer patients with the mutant genotype of *GSTP1(Val/Val)* had an increased risk of leukopenia with an odds of 2.41(**p=0.001**) as compared to subjects carrying both copies of the wild alleles (*Ile/Ile*).
- The patients with the heterozygous genotype (*Ile/Val*) had a 2.14-fold increased risk of developing severe anemia (**p=0.03**).
- Further, in SCLC patients, polymorphism of *GSTP1* was associated with thrombocytopenia (p=**0.02**). Thus, suggesting that *GSTP1 Ile^{105}Val* polymorphism could be a predictive biomarker for specifichematological toxicity, like leukopenia and anemia, but not thrombocytopenia or neutropenia.
- The patients who were carrying mutant alleles (*C/C*) for the *MTHFR* (*1298A>C*) polymorphism showed a higher trend of median survival time (MST) as compared to the patients bearing the wild type genotype (*A/A*) (MST=13.93 vs. 7.97, p=0.12). No

significant associations were noted between *MTHFR* polymorphisms and treatment response, gender, clinic-pathological state, and survival.

- Based on toxicity profiling, we observed that lung cancer patients with the *AC* (heterozygous) genotype for *MTHFR A1298C* polymorphism had a 12-fold risk of diarrhea (**p=0.018**). The patients with the heterozygous genotype (*CT*) of the *MTHFRC677T* polymorphism had a 5.34-fold increased risk of developing neutropenia toxicity (**p=0.009**) when compared to patients with the wild type genotype (*CC*)—suggesting that *MTHFR* genetic variants are associated with severe hematological toxicity like leukopenia and also with gastrointestinal toxicity.
- Thus, evidence suggests that *MTHFR* polymorphism might impact the development of pemetrexed and platinum-related toxicities but not as a clinical predictor of efficiency or as a prognostic marker.
- The five-factor model *(NQO1 $P^{187}S$, EPHX1 $Y^{113}H$, NAT2*6 $R^{197}Q$, NAT2*5B $L^{161}L$, NAT2*5C $K^{268}R$)* was the overall best model for the interaction of selected Phase II polymorphisms.
- The CART results demonstrates that *NAT2 481C>T* caused the initial split
- Individuals carrying the genotypic combination of *NAT803A>G(M)/NAT857G>A(M)/EPHX1337T>C(W)/NQO1609C>T(M)/EPHX1415A >G(W)/NAT590G>A(W)/SULT1A1638G>A(W)/NAT481C>T(M)* conferred the highest risk towards lung cancer (OR=24.35, 95%C.I.=2.94-201.26, p= **0.003**).
- Subgroups under SCLC category with the highest risk was of node 18, 39, 45 and the genotypic combination of these nodes were *NAT803A>G(M)/NAT857G>A(W)/ EPHX1337T>C(M)/NQO1609C>T(W)/SULT1A1638G>A(M)/NAT481C>T(W)*; *EPHX1415A>G(M)/NAT590G>A(W)/NAT857G>A(M)/NAT803A>G(M)/NQO1609C>T (M)/SULT1A1638G>A(W)/EPHX1337T>C(M)/NAT481C>T(M)* and *EPHX1415 A>G(M)/NAT803A>G(M)/NQO1609C>T(M)/NAT590G>A(M)/SULT1A1638G>A(M)/ EPHX1337T>C(M)/NAT481C>T(M)* with an odds of 78.
- Subjects with genotypic combination of *NAT481C>T(W)/NAT857G>A(M)* exhibited an odds of two-folds (OR=2.12, 95%C.I.= 1.30-3.45, p=**0.002**) for ADCC subtype.
- The combination of *NQO1 $Pro^{187}Ser$* with *EPHX1 $His^{139}Arg$* MST of single allelic carriers (CT+AG) was observed to be higher compared to wild subjects (CC+AA) (12.17

months *vs.* 7.70 months, log-rank p=**0.03**) with decreased death rate (HR'=0.65, 95%C.I=0.44-0.95, p=**0.03**). Similarly, the subjects with combined genotype (CT+AG+TT+GG) for the same genotypic combination had higher survival than wild subjects (12.13 months *vs.* 7.70 months, log-rank p=**0.04**) and a lower death rate as observed by Cox-proportional analysis (HR'=0.66, 95%C.I=0.45-0.96, p=**0.03**).

- In terms of analyzing toxicity profile,the mutant genotype for *NQO1* exhibits reduced risk for gastrointestinal toxicity, including constipation and anorexia. Similarly, the heterozygous and combined genotype for *SULT1A1* polymorphism had decreased risk for developing neutropenia.
- Further, the heterozygous and combined genotype for *NAT2*6* polymorphism was found to be associated with an increased risk for leukopenia and anemia,*i.e.*, hematological toxicity as well as neutropenia. At the same time,*NAT2*5C* was associated with a reduced risk of developing hematological toxicity, especially anemia.
- The heterozygous genotype for *NAT2*7* was found to be associated with 2 folds increased risk of thrombocytopenia but decreased risk of anemia (hematological toxicities) as well as decreased risk of alopecia and anorexia.

In conclusion, this is the first Indian study to highlight many PhaseII detoxification genes altogether with such a large sample size and their correlation with clinicopathological factors. The stratified analysis was also performed for smoking and histology to find out high-risk posing subgroups. Our study is thefirst attempt in PhaseII detoxification genes for lung cancer in the Indian population to usea multifactorial approach (CART and MDR) to explore the complex interaction between these genes and theirrelationship towards lung cancer risk. This is the first study in the North Indian region to analyze the SNPs for overall survival and clinical outcomes. For the first time, we have analyzed the stratified clinical outcomes with genotype for the genetic variants in the study. This is the first study thatinvolves evaluating the association of these SNPs with chemotherapeutic-related toxicities. The present findings with a larger sample size for the overall and stratified group would help develop biomarkers both for diagnosis of lung cancer and prognosis of lung cancer.

Prospects of the study

As the PhaseII detoxification enzymes catalyze the detoxification of carcinogens and anticancer drugs, SNPs in PhaseII detoxification genes are likely to play a role in cancer susceptibility. These polymorphisms in the PhaseII genes are worth paying more attention to, for theymay serve as a platform for detecting and predicting individual susceptibility tolung cancer. Second, it may aid in determining the prognosis of the disease. To fulfill the personalized therapeutic system, developing comprehensive and reliable SNP and haplotype analyses of diverse ethnic populations, like Indians, shall be highly beneficial, allowing for predicted genotype-drug efficacy in each individual.

CHAPTER 8

BIBLIOGRAPHY

- Abedinzadeh, M., Moghimi, M., Dastgheib, S. A., Maleki, H., Salehi, E., Mohammad, Z., ... & Neamatzadeh, H. (2020). Association of NAD (P) H Quinine Oxidoreductase 1 rs1800566 Polymorphism with Bladder and Prostate Cancers–a Systematic Review and Meta-Analysis. *Klinicka onkologie: casopis Ceske a Slovenske onkologicke spolecnosti*, *33*(2), 92-100.
- Ada, A. O., Kunak, S. C., Hancer, F., Bilgen, S., Suzen, S. H., Alpar, S., ... & Iscan, M. (2010). CYP and GST polymorphisms and survival in advanced non-small cell lung cancer patients. *Neoplasma*, *57*(6), 512.
- Aihara, T., Kim, Y., & Takatsuka, Y. (2002). Phase II study of weekly docetaxel in patients with metastatic breast cancer. *Annals of oncology*, *13*(2), 286-292.
- Al-Abd, A. M., Alamoudi, A. J., Abdel-Naim, A. B., Neamatallah, T. A., & Ashour, O. M. (2017). Anti-angiogenic agents for the treatment of solid tumors: potential pathways, therapy and current strategies–a review. *Journal of advanced research*, *8*(6), 591-605.
- Alberg, A. J., Brock, M. V., & Samet, J. M. (2005). Epidemiology of lung cancer: looking to the future. *Journal of clinical oncology*, *23*(14), 3175-3185.
- Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M., & Spivack, S. D. (2013). Epidemiology of lung cancer: Diagnosis and management of lung cancer: American College of Chest Physicians evidence-based clinical practice guidelines. *CHEST Journal*, *143*(5_suppl), e1S-e29S
- Allan, J. M., Wild, C. P., Rollinson, S., Willett, E. V., Moorman, A. V., Dovey, G. J., ... & Morgan, G. J. (2001). Polymorphism in glutathione S-transferase P1 is associated with susceptibility to chemotherapy-induced leukemia. *Proceedings of the National Academy of Sciences*, *98*(20), 11592-11597.
- Allocati, N., Masulli, M., Di Ilio, C., & Federici, L. (2018). Glutathione transferases: substrates, inihibitors and pro-drugs in cancer and neurodegenerative diseases. *Oncogenesis*, *7*(1), 1-15.
- Andrew, A. S., Gui, J., Sanderson, A. C., Mason, R. A., Morlock, E. V., Schned, A. R., ... & Karagas, M. R. (2009). Bladder cancer SNP panel predicts susceptibility and survival. *Human genetics*, *125*(5), 527-539.
- Ankathil, R. (2010). Tobacco, genetic susceptibility and lung cancer. *Tobacco Use Insights*, *3*, 1.
- Ansari, J., Palmer, D. H., Rea, D. W., & Hussain, S. A. (2009). Role of tyrosine kinase inhibitors in lung cancer. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, *9*(5), 569-575.

- Arand, M., Grant, D. F., Beetham, J. K., Friedberg, T., Oesch, F., & Hammock, B. D. (1994). Sequence similarity of mammalian epoxide hydrolases to the bacterial haloalkane dehalogenase and other related proteins: implication for the potential catalytic mechanism of enzymatic epoxide hydrolysis. *FEBS letters*, *338*(3), 251-256.
- Ardizzoni, A., Tiseo, M., Boni, L., Vincent, A. D., Passalacqua, R., Buti, S., ... & Smit, E. F. (2012). Pemetrexed versus pemetrexed and carboplatin as second-line chemotherapy in advanced non–small-cell lung cancer: results of the GOIRC 02-2006 randomized phase II study and pooled analysis with the NVALT7 trial. *Journal of clinical oncology*, *30*(36), 4501-4507.
- Argiris, A., Karamouzis, M. V., Gooding, W. E., Branstetter, B. F., Zhong, S., Raez, L. E., ... & Romkes, M. (2011). Phase II trial of pemetrexed and bevacizumab in patients with recurrent or metastatic head and neck cancer. *Journal of Clinical Oncology*, *29*(9), 1140.
- Armstrong, R. N. (1987). Enzyme-Catalyzed Detoxication Reactions: Mechanisms and Stereochemistr. *Critical Reviews in Biochemistry*, *22*(1), 39-88.
- Armstrong, R. N. (1997). Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chemical research in toxicology*, *10*(1), 2-18.
- Armstrong, R. N., Kedzierski, B., Levin, W., & Jerina, D. (1981). Enantioselectivity of microsomal epoxide hydrolase toward arene oxide substrates. *Journal of Biological Chemistry*, *256*(10), 4726-4733.
- Arslan, S. (2010). Genetic polymorphisms of sulfotransferases (SULT1A1 and SULT1A2) in a Turkish population. *Biochemical genetics*, *48*(11-12), 987-994.
- Arslan, S., Karadayi, S., Yildirim, M. E., Ozdemir, O., & Akkurt, I. (2011). The association between methylene-tetrahydrofolate reductase gene polymorphism and lung cancer risk. *Molecular biology reports*, *38*(2), 991-996.
- Arslan, S., Silig, Y., & Pinarbasi, H. (2009). An investigation of the relationship between SULT1A1 Arg213His polymorphism and lung cancer susceptibility in a Turkish population. *Cell Biochemistry and Function: Cellular biochemistry and its modulation by active agents or disease*, *27*(4), 211-215.
- Artac, M., Bozcuk, H., Pehlivan, S., Akcan, S., Pehlivan, M., Sever, T., ... & Savas, B. (2010). The value of XPD and XRCC1 genotype polymorphisms to predict clinical outcome in metastatic colorectal carcinoma patients with irinotecan-based regimens. *Journal of cancer research and clinical oncology*, *136*(6), 803-809.
- Asami, S., Manabe, H., Miyake, J., Tsurudome, Y., Hirano, T., Yamaguchi, R., ... & Kasai, H. (1997). Cigarette smoking induces an increase in oxidative DNA damage, 8-hydroxydeoxyguanosine, in a central site of the human lung. *Carcinogenesis*, *18*(9), 1763-1766.
- Azzato, E. M., Chen, R. A., Wacholder, S., Chanock, S. J., Klebanoff, M. A., & Caporaso, N. E. (2010). Maternal EPHX1 polymorphisms and risk of phenytoin-induced congenital malformations. *Pharmacogenetics and genomics*, *20*(1), 58-63.

- Banerjee, S. (2020). Evaluation of the risk of lung cancer associated with NAD (P) H: quinone oxidoreductase 1 (NQO1) C609T polymorphism in male current cigarette smokers from the Eastern India. *Molecular Biology Research Communications*, *9*(3), 111.
- Banoglu, E. (2000). Current status of the cytosolic sulfotransferases in the metabolic activation of promutagens and procarcinogens. *Current drug metabolism*, *1*(1), 1-30.
- Bardakci, F., Arslan, S. E. R. D. A. L., Bardakci, S., Binatli, A. O., & Budak, M. A. H. İ. R. (2008). Sulfotransferase 1A1 (SULT1A1) polymorphism and susceptibility to primary brain tumors. *Journal of cancer research and clinical oncology*, *134*(1), 109-114.
- Barker, D. F., Husain, A., Neale, J. R., Martini, B. D., Zhang, X., Doll, M. A., ... & Hein, D. W. (2006). Functional properties of an alternative, tissue-specific promoter for human arylamine N-acetyltransferase 1. *Pharmacogenetics and genomics*, *16*(7), 515.
- Bartlett, J. M., & White, A. (2003). Extraction of DNA from whole blood. In *PCR protocols* (pp. 29-31). Humana Press.
- Baumann, E. (1876). Ueber Sulfosäuren im Harn. *Berichte der deutschen chemischen Gesellschaft*, *9*(1), 54-58.
- Behera, D. (2007). New approach to the treatment of lung cancer: the molecular targeted therapy. *Indian journal of chest diseases & allied sciences*, *49*(3), 149-158.
- Behera, D., & Balamugesh, T. (2004). Lung cancer in India. *Indian Journal of Chest Diseases and Allied Sciences*, *46*, 269-282.
- Behera, D., & Balamugesh, T. (2012). Lung cancer in India. *Indian J Chest Dis Allied Science 2004; 46: 269*, *81*.
- Belogubova, E. V., Kuligina, E. S., Togo, A. V., Karpova, M. B., Ulibina, J. M., Shutkin, V. A., ... & Imyanitov, E. N. (2005). 'Comparison of extremes' approach provides evidence against the modifying role of NAT2 polymorphism in lung cancer susceptibility. *Cancer letters*, *221*(2), 177-183.
- Benhamou, S., Reinikainen, M., Bouchardy, C., Dayer, P., & Hirvonen, A. (1998). Association between lung cancer and microsomal epoxide hydrolase genotypes. *Cancer research*, *58*(23), 5291-5293.
- Benjamini, Y. and Hochberg, Y. (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the royal statistical society. Series B (Methodological)* pp. 289–300.
- Bennett, W. P., Alavanja, M. C., Blomeke, B., Vähäkangas, K. H., Castrén, K., Welsh, J. A., ... & Harris, C. C. (1999). Environmental tobacco smoke, genetic susceptibility, and risk of lung cancer in never-smoking women. *Journal of the National Cancer Institute*, *91*(23), 2009-2014.
- Berman, D. W., & Crump, K. S. (2008). Update of potency factors for asbestos-related lung cancer and mesothelioma. *Critical reviews in toxicology*, *38*(sup1), 1-47.
- Bernsen, E. C., Hagleitner, M. M., Kouwenberg, T. W., & Hanff, L. M. (2020). Pharmacogenomics as a tool to limit acute and long-term adverse effects of

chemotherapeutics: An update in pediatric oncology. *Frontiers in pharmacology*, *11*, 1184.

- Bianchet, M. A., Erdemli, S. B., & Amzel, L. M. (2008). Structure, function, and mechanism of cytosolic quinone reductases. *Vitamins & Hormones*, *78*, 63-84.
- Bianchet, M. A., Faig, M., & Amzel, L. M. (2004). Structure and mechanism of NAD [P] H: quinone acceptor oxidoreductases (NQO). *Methods in enzymology*, *382*, 144-174.
- Blanchard, R. L., Freimuth, R. R., Buck, J., Weinshilboum, R. M., & Coughtrie, M. W. (2004). A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. *Pharmacogenetics and Genomics*, *14*(3), 199-211.
- Blank, S., Rachakonda, S., Keller, G., Weichert, W., Lordick, F., Langer, R., ... & Ott, K. (2014). A retrospective comparative exploratory study on two methylentetrahydrofolate reductase (MTHFR) polymorphisms in esophagogastric cancer: the A1298C MTHFR polymorphism is an independent prognostic factor only in neoadjuvantly treated gastric cancer patients. *BMC cancer*, *14*(1), 1-15.
- Blum, M., Grant, D. M., McBride, W., Heim, M., & Meyer, U. A. (1990). Human arylamine N-acetyltransferase genes: isolation, chromosomal localization, and functional expression. *DNA and cell biology*, *9*(3), 193-203.
- Board, P. G., & Menon, D. (2013). Glutathione transferases, regulators of cellular metabolism and physiology. *Biochimica et biophysica acta (bba)-general subjects*, *1830*(5), 3267-3288.
- Boccia, S., Persiani, R., La Torre, G., Rausei, S., Arzani, D., Gianfagna, F., ... & Ricciardi, G. (2005). Sulfotransferase 1A1 polymorphism and gastric cancer risk: a pilot case-control study. *Cancer letters*, *229*(2), 235-243.
- Bocedi, A., Noce, A., Marrone, G., Noce, G., Cattani, G., Gambardella, G., ... & Ricci, G. (2019). Glutathione transferase P1-1 an enzyme useful in biomedicine and as biomarker in clinical practice and in environmental pollution. *Nutrients*, *11*(8), 1741.
- Bock, C. H., Wenzlaff, A. S., Cote, M. L., Land, S. J., & Schwartz, A. G. (2005). NQO1 T allele associated with decreased risk of later age at diagnosis lung cancer among never smokers: results from a population-based study. *Carcinogenesis*, *26*(2), 381-386.
- Bode, A. M., & Dong, Z. (2005). Signal transduction pathways in cancer development and as targets for cancer prevention. *Progress in nucleic acid research and molecular biology*, *79*, 237-297.
- Boige, V., Mendiboure, J., Pignon, J. P., Loriot, M. A., Castaing, M., Barrois, M., ... & Laurent-Puig, P. (2010). Pharmacogenetic assessment of toxicity and outcome in patients with metastatic colorectal cancer treated with LV5FU2, FOLFOX, and FOLFIRI: FFCD 2000-05. *Journal of Clinical Oncology*, *28*(15), 2556-2564.
- Boland, M. P., Knox, R. J., & Roberts, J. J. (1991). The differences in kinetics of rat and human DT diaphorase result in a differential sensitivity of derived cell lines to CB 1954 (5-(aziridin-1-yl)-2, 4-dinitrobenzamide). *Biochemical pharmacology*, *41*(6-7), 867-875.

- Booten, R., Ward, T., Heighway, J., Ashcroft, L., Morris, J., & Thatcher, N. (2006). Glutathione-S-transferase P1 isoenzyme polymorphisms, platinum-based chemotherapy, and non-small cell lung cancer. *Journal of thoracic oncology*, *1*(7), 679-683.
- Booth, J., Boyland, E., & Sims, A. P. (1961). An enzyme from rat liver catalysing conjugations with glutathione. *Biochemical Journal*, *79*(3), 516-524.
- Borlak, J., & Reamon-Buettner, S. M. (2006). N-acetyltransferase 2 (NAT2) gene polymorphisms in colon and lung cancer patients. *BMC medical genetics*, *7*(1), 1-9.
- Bouchardy, C., Mitrunen, K., Wikman, H., Husgafvel-Pursiainen, K., Dayer, P., Benhamou, S., & Hirvonen, A. (1998). N-acetyltransferase NAT1 and NAT2 genotypes and lung cancer risk. *Pharmacogenetics*, *8*(4), 291-298.
- Boukouvala, S., & Sim, E. (2005). Structural analysis of the genes for human arylamine N-acetyltransferases and characterisation of alternative transcripts. *Basic & clinical pharmacology & toxicology*, *96*(5), 343-351.
- Bu, L., Zhang, L. B., Mao, X., & Wang, P. (2016). GSTP1 Ile105Val and XRCC1 Arg399Gln gene polymorphisms contribute to the clinical outcome of patients with advanced non-small cell lung cancer. *Genet Mol Res*, *15*(2).
- Buchbinder, E. I., & Desai, A. (2016). CTLA-4 and PD-1 pathways: similarities, differences, and implications of their inhibition. *American journal of clinical oncology*, *39*(1), 98.
- Buehler, B. A., Delimont, D., Van Waes, M., & Finnell, R. H. (1990). Prenatal prediction of risk of the fetal hydantoin syndrome. *New England Journal of Medicine*, *322*(22), 1567-1572.
- Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS computational biology*, *8*(12), e1002822.
- Bustami, M., Al-Shudifat, A. E., Hussein, N., Yacoub, M., Atwa, E., Sabri, I., ... & Abu-Qatouseh, L. (2018). Impact of Genetic Polymorphism of Sulpha Transferase Genes (SULT1A) Genes on the Risk of Females with Breast Cancer in Jordan.
- Butler, M. A., Iwasaki, M., Guengerich, F. P., & Kadlubar, F. F. (1989). Human cytochrome P-450PA (P-450IA2), the phenacetin O-deethylase, is primarily responsible for the hepatic 3-demethylation of caffeine and N-oxidation of carcinogenic arylamines. *Proceedings of the National Academy of Sciences*, *86*(20), 7696-7700.
- Cabanero, M., Sangha, R., Sheffield, B. S., Sukhai, M., Pakkal, M., Kamel-Reid, S., ... & Tsao, M. S. (2017). Management of EGFR-mutated non-small-cell lung cancer: practical implications from a clinical and pathology perspective. *Current Oncology*, *24*(2), 111-119.
- Cadenas, E. (1995). Antioxidant and prooxidant functions of DT-diaphorase in quinone metabolism. *Biochemical pharmacology*, *49*(2), 127-140.
- Cajas-Salazar, N., Au, W. W., Zwischenberger, J. B., Sierra-Torres, C. H., Salama, S. A., Alpard, S. K., & Tyring, S. K. (2003). Effect of epoxide hydrolase polymorphisms on

chromosome aberrations and risk for lung cancer. *Cancer genetics and cytogenetics*, *145*(2), 97-102.

- Carter, B. W., Glisson, B. S., Truong, M. T., & Erasmus, J. J. (2014). Small cell lung carcinoma: staging, imaging, and treatment considerations. *Radiographics*, *34*(6), 1707-1721.
- Cascorbi, I., Brockmöller, J., Mrozikiewiz, P. M., Bauer, S., Loddenkemper, R., & Roots, I. (1996). Homozygous rapid arylamine N-acetyltransferase (NAT2) genotype as a susceptibility factor for lung cancer. *Cancer research*, *56*(17), 3961-3966.
  cases with surgical specimens. *The American Journal of Surgical Pathology*, *26*(9), 1184-1197.
- Castell, J. V., Donato, M. T., & Gómez-Lechón, M. J. (2005). Metabolism and bioactivation of toxicants in the lung. The in vitro cellular approach. *Experimental and Toxicologic Pathology*, *57*, 189-204.
- Cenas, N., Nivinskas, H., Anusevicius, Z., Sarlauskas, J., Lederer, F., & Arnér, E. S. (2004). Interactions of quinones with thioredoxin reductase: a challenge to the antioxidant role of the mammalian selenoprotein. *Journal of Biological Chemistry*, *279*(4), 2583-2592.
- Center NC. Cancer Gene Panel Test OncoGuideTM NCC Oncopanel System added to Health Insurance Coverage list. Available online: https://www.ncc.go.jp/en/information/press_release/20190717/20190717152024.html
- Chao, C., Zhang, Z. F., Berthiller, J., Boffetta, P., & Hashibe, M. (2006). NAD (P) H: quinone oxidoreductase 1 (NQO1) Pro187Ser polymorphism and the risk of lung, bladder, and colorectal cancers: a meta-analysis. *Cancer Epidemiology and Prevention Biomarkers*, *15*(5), 979-987.
- Chattopadhyay, S., Moran, R. G., & Goldman, I. D. (2007). Pemetrexed: biochemical and cellular pharmacology, mechanisms, and clinical applications. *Molecular cancer therapeutics*, *6*(2), 404-417.
- Chen, H. C., Cao, Y. F., Hu, W. X., Liu, X. F., Liu, Q. X., Zhang, J., & Liu, J. (2006). Genetic polymorphisms of phase II metabolic enzymes and lung cancer susceptibility in a population of Central South China. *Disease markers*, *22*(3), 141-152.
- Chen, J. B., Wang, F., Wu, J. J., & Cai, M. (2016). Glutathione S-transferase pi polymorphism contributes to the treatment outcomes of advanced non-small cell lung cancer patients in a Chinese population. *Genet Mol Res*, *15*(3).
- Chen, L. Z., Harris, P. C., Apostolou, S., Baker, E., Holman, K., Lane, S. A., ... & Callen, D. F. (1991). A refined physical map of the long arm of human chromosome 16. *Genomics*, *10*(2), 308-312.
- Cheng, D. T., Mitchell, T. N., Zehir, A., Shah, R. H., Benayed, R., Syed, A., ... & Berger, M. F. (2015). Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation

sequencing clinical assay for solid tumor molecular oncology. *The Journal of molecular diagnostics*, *17*(3), 251-264.

- Chiou, H. L., Wu, M. F., Chien, W. P., Cheng, Y. W., Wong, R. H., Chen, C. Y., ... & Lee, H. (2005). NAT2 fast acetylator genotype is associated with an increased risk of lung cancer among never-smoking women in Taiwan. *Cancer letters*, *223*(1), 93-101.
- Cho, W. C. S. (2007). Potentially useful biomarkers for the diagnosis, treatment and prognosis of lung cancer. *Biomedicine & pharmacotherapy*, *61*(9), 515-519.
- Choi, S. W., & Mason, J. B. (2002). Folate status: effects on pathways of colorectal carcinogenesis. *The Journal of nutrition*, *132*(8), 2413S-2418S.
- Christmann, M., Tomicic, M. T., Roos, W. P., & Kaina, B. (2003). Mechanisms of human DNA repair: an update. *Toxicology*, *193*(1-2), 3-34.
- Chu, P. G., & Weiss, L. M. (2002). Expression of cytokeratin 5/6 in epithelial neoplasms: an immunohistochemical study of 509 cases. *Modern pathology*, *15*(1), 6-10.
- Chu, Q. C., Markman, B., Leighl, N., Krug, L., Rudin, C., Lathers, D., … & Kolaitis, G. (2016). A phase 1/2 trial of a monoclonal antibody targeting fucosyl GM1 in relapsed/refractory small cell lung cancer (SCLC): Safety and preliminary efficacy.
- Chua, W., Goldstein, D., Lee, C. K., Dhillon, H., Michael, M., Mitchell, P., ... & Iacopetta, B. (2009). Molecular markers of response and toxicity to FOLFOX chemotherapy in metastatic colorectal cancer. *British journal of cancer*, *101*(6), 998-1004.
- Chung, Y. T., Hsieh, L. L., Chen, I. H., Liao, C. T., Liou, S. H., Chi, C. W., ... & Liu, T. Y. (2009). Sulfotransferase 1A1 haplotypes associated with oral squamous cell carcinoma susceptibility in male Taiwanese. *Carcinogenesis*, *30*(2), 286-294.
- Cohen, S. M., & Arnold, L. L. (2011). Chemical carcinogenesis. *Toxicological Sciences*, *120*(suppl_1), S76-S92.
- Colby, T. V. (1995). Tumors of the lower respiratory tract. *Atlas of Tumor Pathology*, 309-317.
- Corrigan, A., Walker, J. L., Wickramasinghe, S., Hernandez, M. A., Newhouse, S. J., Folarin, A. A., ... & Marinaki, A. M. (2014). Pharmacogenetics of pemetrexed combination therapy in lung cancer: pathway analysis reveals novel toxicity associations. *The pharmacogenomics journal*, *14*(5), 411-417.
- Cottrell, C. E., Al-Kateb, H., Bredemeyer, A. J., Duncavage, E. J., Spencer, D. H., Abel, H. J., ... & Pfeifer, J. D. (2014). Validation of a next-generation sequencing assay for clinical molecular oncology. *The Journal of molecular diagnostics*, *16*(1), 89-105.
- Coughtrie, M. W. (2016). Function and organization of the human cytosolic sulfotransferase (SULT) family. *Chemico-biological interactions*, *259*, 2-7.
- Coughtrie, M. W., Sharp, S., Maxwell, K., & Innes, N. P. (1998). Biology and function of the reversible sulfation pathway catalysed by human sulfotransferases and sulfatases. *Chemico-biological interactions*, *109*(1-3), 3-27.

- Craft, J. A., Baird, S., Lament, M., & Burchell, B. (1990). Membrane topology of epoxide hydrolase. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, *1046*(1), 32-39.
- Cruz, C. S. D., Tanoue, L. T., & Matthay, R. A. (2011). Lung cancer: epidemiology, etiology, and prevention. *Clinics in chest medicine*, *32*(4), 605-644.
- Cui, D., Wang, Z., Zhao, E., Ma, J., & Lu, W. (2011). NAT2 polymorphism and lung cancer risk: a meta-analysis. *Lung Cancer*, *73*(2), 153-157.
- Cui, D., Wang, Z., Zhao, E., Ma, J., & Lu, W. (2011). NAT2 polymorphism and lung cancer risk: a meta-analysis. *Lung Cancer*, *73*(2), 153-157.
- Cui, K., Lu, A. Y., & Yang, C. S. (1995). Subunit functional studies of NAD (P) H: quinone oxidoreductase with a heterodimer approach. *Proceedings of the National Academy of Sciences*, *92*(4), 1043-1047.
- Cui, X., Jin, T., Wang, X., Jin, G., Li, Z., & Lin, L. (2014). NAD (P) H: quinone oxidoreductase-1 overexpression predicts poor prognosis in small cell lung cancer. *Oncology reports*, *32*(6), 2589-2595.
- Cui, X., Li, L., Yan, G., Meng, K., Lin, Z., Nan, Y., ... & Li, C. (2015). High expression of NQO1 is associated with poor prognosis in serous ovarian carcinoma. *BMC cancer*, *15*(1), 1-8.
- da Silva, T. D., Felipe, A. V., de Lima, J. M., Oshima, C. T. F., & Forones, N. M. (2011). N-Acetyltransferase 2 genetic polymorphisms and risk of colorectal cancer. *World journal of gastroenterology: WJG*, *17*(6), 760.
- Darvin, P., Toor, S. M., Nair, V. S., & Elkord, E. (2018). Immune checkpoint inhibitors: recent progress and potential biomarkers. *Experimental & molecular medicine*, *50*(12), 1-11.
- Dash, R., Ali, M., Dash, N., Azad, M., Kalam, A., Hosen, S. M., ... & Moon, I. S. (2019). Structural and dynamic characterizations highlight the deleterious role of SULT1A1 R213H polymorphism in substrate binding. *International journal of molecular sciences*, *20*(24), 6256.
- Daubner, S. C., & Matthews, R. G. (1982). Purification and properties of methylenetetrahydrofolate reductase from pig liver. *Journal of Biological Chemistry*, *257*(1), 140-145.
- de Figueiredo Teixeira, R. L., Lopes, M. Q. P., Suffys, P. N., & Santos, A. R. (2013). Tuberculosis Pharmacogenetics: State of The Art. *Tuberculosis: Current Issues in Diagnosis and Management*, 105.
- de Groot, P. M., Wu, C. C., Carter, B. W., & Munden, R. F. (2018). The epidemiology of lung cancer. *Translational lung cancer research*, *7*(3), 220.
- de Groot, P., & Munden, R. F. (2012). Lung cancer epidemiology, risk factors, and prevention. *Radiologic Clinics*, *50*(5), 863-876.

- De Luca, A., Pellizzari Tregno, F., Sau, A., Pastore, A., Palumbo, C., Alama, A., ... & Caccuri, A. M. (2013). Glutathione S-transferase P 1-1 as a target for mesothelioma treatment. *Cancer science*, *104*(2), 223-230.
- Decker, M., Arand, M., & Cronin, A. (2009). Mammalian epoxide hydrolases in xenobiotic metabolism and signalling. *Archives of toxicology*, *83*(4), 297-318.
- Decker, M., Arand, M., & Cronin, A. (2009). Mammalian epoxide hydrolases in xenobiotic metabolism and signalling. *Archives of toxicology*, *83*(4), 297-318.
- den Braver-Sewradj, S. P., den Braver, M. W., Toorneman, R. M., Van Leeuwen, S., Zhang, Y., Dekker, S. J., ... & Vos, J. C. (2018). Reduction and scavenging of chemically reactive drug metabolites by NAD (P) H: quinone oxidoreductase 1 and NRH: quinone oxidoreductase 2 and variability in hepatic concentrations. *Chemical research in toxicology*, *31*(2), 116-126.
- Deng, J. H., Deng, J., Shi, D. H., Ouyang, X. N., & Niu, P. G. (2015). Clinical outcome of cisplatin-based chemotherapy is associated with the polymorphisms of GSTP1 and XRCC1 in advanced non-small cell lung cancer patients. *Clinical and Translational Oncology*, *17*(9), 720-726.
- Desjardins, P., & Conklin, D. (2010). NanoDrop microvolume quantitation of nucleic acids. *JoVE (Journal of Visualized Experiments)*, (45), e2565.
- Desnoyer, A., Broutin, S., Delahousse, J., Maritaz, C., Blondel, L., Mir, O., ... & Paci, A. (2020). Pharmacokinetic/pharmacodynamic relationship of therapeutic monoclonal antibodies used in oncology: part 2, immune checkpoint inhibitor antibodies. *European Journal of Cancer*, *128*, 119-128.
- Detterbeck, F. C., Boffa, D. J., Kim, A. W., & Tanoue, L. T. (2017). The eighth edition lung cancer stage classification. *Chest*, *151*(1), 193-203.
- Di Pietro, G., Magno, L. A. V., & Rios-Santos, F. (2010). Glutathione S-transferases: an overview in cancer research. *Expert opinion on drug metabolism & toxicology*, *6*(2), 153-170.
- Dik, V. K., van Oijen, M. G., Uiterwaal, C., van Gils, C., van Duijnhoven, F. J., Cauchi, S., ... & Bueno-de-Mesquita, H. B. (2013). Mo1139 Coffee Consumption Genetic Polymorphisms in CYP1A2 and NAT2, and Colorectal Cancer Risk. *Gastroenterology*, *5*(144), S-589.
- Dikshit, R. P., & Kanhere, S. (2000). Tobacco habits and risk of lung, oropharyngeal and oral cavity cancer: a population-based case-control study in Bhopal, India. *International Journal of epidemiology*, *29*(4), 609-614.
- Dinkova-Kostova, A. T., & Talalay, P. (2010). NAD (P) H: quinone acceptor oxidoreductase 1 (NQO1), a multifunctional antioxidant enzyme and exceptionally versatile cytoprotector. *Archives of biochemistry and biophysics*, *501*(1), 116-123.
- Donnadieu, N., Paesmans, M., & Sculier, J. P. (1991). Chimiothérapie des cancers bronchiques non à petites cellules: meta-analyse de la littérature en fonction de l'extension de la maladie. *Revue des maladies respiratoires*, *8*(2), 197-204.

- Dourado, D. F., Fernandes, P. A., & Ramos, M. J. (2008). Mammalian cytosolic glutathione transferases. *Current protein & peptide science*, *9*(4), 325-337.
- Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. *Nature reviews cancer*, *3*(1), 11-22.
- Duthie, S. J. (1999). Folic acid deficiency and cancer: mechanisms of DNA instability. *British medical bulletin*, *55*(3), 578-592.
- Duval, R., Xu, X., Bui, L. C., Mathieu, C., Petit, E., Cariou, K., ... & Rodrigues-Lima, F. (2016). Identification of cancer chemopreventive isothiocyanates as direct inhibitors of the arylamine N-acetyltransferase-dependent acetylation and bioactivation of aromatic amine carcinogens. *Oncotarget*, *7*(8), 8688.
- Duvoix, A., Schnekenburger, M., Delhalle, S., Blasius, R., Borde-Chiché, P., Morceau, F., ... & Diederich, M. (2004). Expression of glutathione S-transferase P1-1 in leukemic cells is regulated by inducible AP-1 binding. *Cancer letters*, *216*(2), 207-219.
- Dyda, F., Klein, D. C., & Hickman, A. B. (2000). GCN5-related N-acetyltransferases: a structural overview. *Annual review of biophysics and biomolecular structure*, *29*(1), 81-103.
- Eisenhauer, E., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., ...& Rubinstein, L. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer*, *45*(2), 228-247.
- Ellis, G. K., Bone, H. G., & Chlebowski, R. (2008). ADJUVANT THERAPY. *J Clin Oncol*, *26*, 4875-4882.
- El-Sherbeni, A. A., & El-Kadi, A. O. (2014). The role of epoxide hydrolases in health and disease. *Archives of toxicology*, *88*(11), 2013-2032.
- Eom, S. Y., Zhang, Y. W., Kim, S. H., Choe, K. H., Lee, K. Y., Park, J. D., ... & Kim, H. (2009). Influence of NQO1, ALDH2, and CYP2E1 genetic polymorphisms, smoking, and alcohol drinking on the risk of lung cancer in Koreans. *Cancer Causes & Control*, *20*(2), 137-145.
- Eranster, L. (1958). Soluble diaphorase in animal tissues. *Acta chem. scand.*, *12*, 595.
- Erkisi, Z., Yaylim-Eraltan, I., Turna, A., Görmüs, U., Camlica, H., & Isbir, T. (2010). Polymorphisms in the microsomal epoxide hydrolase gene: role in lung cancer susceptibility and prognosis. *Tumori Journal*, *96*(5), 756-763.
- Fagerholm, R., Hofstetter, B., Tommiska, J., Aaltonen, K., Vrtel, R., Syrjäkoski, K., ... & Nevanlinna, H. (2008). NAD (P) H: quinone oxidoreductase 1 NQO1* 2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. *Nature genetics*, *40*(7), 844-853.
- Faig, M., Bianchet, M. A., Talalay, P., Chen, S., Winski, S., Ross, D., & Amzel, L. M. (2000). Structures of recombinant human and mouse NAD (P) H: quinone oxidoreductases: species comparison and structural changes with substrate binding and release. *Proceedings of the National Academy of Sciences*, *97*(7), 3177-3182.

- Faig, M., Bianchet, M. A., Winski, S., Hargreaves, R., Moody, C. J., Hudnott, A. R., ... & Amzel, L. M. (2001). Structure-based development of anticancer drugs: complexes of NAD (P) H: quinone oxidoreductase 1 with chemotherapeutic quinones. *Structure*, *9*(8), 659-667.
- Faria, G. M., Soares, I. D., Salazar, M. D. A., Amorim, M. R., Pessoa, B. L., da Fonseca, C. O., & Quirico-Santos, T. (2020). Intranasal perillyl alcohol therapy improves survival of patients with recurrent glioblastoma harboring mutant variant for MTHFR rs1801133 polymorphism. *BMC cancer*, *20*(1), 1-10.
- Fathy, M., Hamed, M., Youssif, O., Fawzy, N., & Ashour, W. (2014). Association between environmental tobacco smoke exposure and lung cancer susceptibility: modification by antioxidant enzyme genetic polymorphisms. *Molecular diagnosis & therapy*, *18*(1), 55-62.
- Ferreira, G. D., Fernandes, G. M. D. M., Penteado, C., Cória, V. R., Galbiatti-Dias, A. L. D. S., Russo, A., ... & Goloni-Bertollo, E. M. (2021). Polymorphisms in xenobiotic metabolism-related genes in patients with hepatocellular carcinoma: a case–control study. *Xenobiotica*, *51*(6), 737-744.
- Flammang, T. J., & Kadlubar, F. F. (1986). Acetyl coenzyme A-dependent metabolic activation of N-hydroxy-3, 2'-dimethyl-4-aminobiphenyl and several carcinogenic N-hydroxy arylamines in relation to tissue and species differences, other acyl donors, and arylhydroxamic acid-dependent acyltransferases. *Carcinogenesis*, *7*(6), 919-926.
- Fleiss, J. L., & Gross, A. J. (1991). Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *Journal of clinical epidemiology*, *44*(2), 127-139.
- Fong, K. M., Sekido, Y., Gazdar, A. F., & Minna, J. D. (2003). Lung cancer• 9: Molecular biology of lung cancer: clinical implications. *Thorax*, *58*(10), 892-900.
- Forgacs, E., Zöchbauer-Müller, S., Oláh, E., & Minna, J. D. (2001). Molecular genetic abnormalities in the pathogenesis of human lung cancer. *Pathology Oncology Research*, *7*(1), 6-13.
- Frampton, G. M., Fichtenholtz, A., Otto, G. A., Wang, K., Downing, S. R., He, J., ... & Yelensky, R. (2013). Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature biotechnology*, *31*(11), 1023-1031.
- Freimuth, R. R., Wiepert, M., Chute, C. G., Wieben, E. D., & Weinshilboum, R. M. (2004). Human cytosolic sulfotransferase database mining: identification of seven novel genes and pseudogenes. *The pharmacogenomics journal*, *4*(1), 54-65.
- Fretland, A. J., & Omiecinski, C. J. (2000). Epoxide hydrolases: biochemistry and molecular biology. *Chemico-biological interactions*, *129*(1-2), 41-59.
- Fretland, A. J., Doll, M. A., Gray, K., Feng, Y., & Hein, D. W. (1997). Cloning, Sequencing, and Recombinant Expression ofNAT1, NAT2, andNAT3Derived from the C3H/HeJ (Rapid) and A/HeJ (Slow) Acetylator Inbred Mouse: Functional

Characterization of the Activation and Deactivation of Aromatic Amine Carcinogens. *Toxicology and applied pharmacology*, *142*(2), 360-366.

- Froese, D. S., Kopec, J., Rembeza, E., Bezerra, G. A., Oberholzer, A. E., Suormala, T., ... & Yue, W. W. (2018). Structural basis for the regulation of human 5, 10-methylenetetrahydrofolate reductase by phosphorylation and S-adenosylmethionine inhibition. *Nature communications*, *9*(1), 1-13.
- Frosst, P., Blom, H. J., Milos, R., Goyette, P., Sheppard, C. A., Matthews, R. G., ... & Rozen, R. (1995). A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nature genetics*, *10*(1), 111-113.
- Fukuda, M., Oka, M., Soda, H., Kinoshita, A., Fukuda, M., Nagashima, S., ... & Kohno, S. (2004). Phase II study of irinotecan combined with carboplatin in previously untreated non-small-cell lung cancer. *Cancer chemotherapy and pharmacology*, *54*(6), 573-577.
- Fukuse, T., Hirata, T., Naiki, H., Hitomi, S., & Wada, H. (2000). Prognostic significance of cyclin E overexpression in resected non-small cell lung cancer. *Cancer research*, *60*(2), 242-244.
- Funai, K., Yokose, T., Ishii, G. I., Araki, K., Yoshida, J., Nishimura, M., ... & Ochiai, A. (2003). Clinicopathologic characteristics of peripheral squamous cell carcinoma of the lung. *The American journal of surgical pathology*, *27*(7), 978-984.
- GAEDIGK, A., LEEDER, J. S., & GRANT, D. M. (1997). Tissue-specific expression and alternative splicing of human microsomal epoxide hydrolase. *DNA and cell biology*, *16*(11), 1257-1266.
- Gaedigk, A., Spielberg, S. P., & Grant, D. M. (1994). Characterization of the microsomal epoxide hydrolase gene in patients with anticonvulsant adverse drug reactions. *Pharmacogenetics*, *4*(3), 142-153.
- Gajalakshmi, V., Hung, R. J., Mathew, A., Varghese, C., Brennan, P., & Boffetta, P. (2003). Tobacco smoking and chewing, alcohol drinking and lung cancer risk among men in southern India. *International journal of cancer*, *107*(3), 441-447.
- Galluzzi, L., Senovilla, L., Vitale, I., Michels, J., Martins, I., Kepp, O., ... & Kroemer, G. (2012). Molecular mechanisms of cisplatin resistance. *Oncogene*, *31*(15), 1869-1883.
- Gamage, N. U., Duggleby, R. G., Barnett, A. C., Tresillian, M., Latham, C. F., Liyou, N. E., ... & Martin, J. L. (2003). Structure of a human carcinogen-converting enzyme, SULT1A1: structural and kinetic implications of substrate inhibition. *Journal of Biological Chemistry*, *278*(9), 7655-7662.
- Gamage, N., Barnett, A., Hempel, N., Duggleby, R. G., Windmill, K. F., Martin, J. L., & McManus, M. E. (2006). Human sulfotransferases and their role in chemical metabolism. *Toxicological sciences*, *90*(1), 5-22.
- Gandini, S., Botteri, E., Iodice, S., Boniol, M., Lowenfels, A. B., Maisonneuve, P., & Boyle, P. (2008). Tobacco smoking and cancer: A meta-analysis. *International journal of cancer*, *122*(1), 155-164.

- Garcia, E. P., Minkovsky, A., Jia, Y., Ducar, M. D., Shivdasani, P., Gong, X., ... & Dong, F. (2017). Validation of OncoPanel: a targeted next-generation sequencing assay for the detection of somatic variants in cancer. *Archives of Pathology and Laboratory Medicine*, *141*(6), 751-758.
- Garcia, S. L., Lauritsen, J., Zhang, Z., Bandak, M., Dalgaard, M. D., Nielsen, R. L., ... & Gupta, R. (2020). Prediction of nephrotoxicity associated with cisplatin-based chemotherapy in testicular cancer patients. *JNCI cancer spectrum*, *4*(3), pkaa032.
- Garfall, A. L., Fraietta, J. A., & Maus, M. V. (2014). Immunotherapy with chimeric antigen receptors for multiple myeloma. *Discovery medicine*, *17*(91), 37-46.
- Gastman, B., Agarwal, P. K., Berger, A., Boland, G., Broderick, S., Butterfield, L. H., ... & Kaufman, H. L. (2020). Defining best practices for tissue procurement in immuno-oncology clinical trials: Consensus statement from the Society for Immunotherapy of Cancer Surgery Committee. *Journal for immunotherapy of cancer*, *8*(2).
- Gautheron, J., & Jéru, I. (2021). The Multifaceted Role of Epoxide Hydrolases in Human Health and Disease. *International Journal of Molecular Sciences*, *22*(1), 13.
- Gemignani, F., Landi, S., Szeszenia-Dabrowska, N., Zaridze, D., Lissowska, J., Rudnai, P., ... & Brennan, P. (2007). Development of lung cancer before the age of 50: the role of xenobiotic metabolizing genes. *Carcinogenesis*, *28*(6), 1287-1293.
- George, J., Lim, J. S., Jang, S. J., Cun, Y., Ozretić, L., Kong, G., ... & Thomas, R. K. (2015). Comprehensive genomic profiles of small cell lung cancer. *Nature*, *524*(7563), 47-53.
- Ghattas, M. H., & Amer, M. A. (2012). Possible role of microsomal epoxide hydrolase gene polymorphism as a risk factor for developing insulin resistance and type 2 diabetes mellitus. *Endocrine*, *42*(3), 577-583.
- Glatt, H. (2000). Sulfotransferases in the bioactivation of xenobiotics. *Chemico-biological interactions*, *129*(1-2), 141-170.
- Glatt, H., & Meinl, W. (2004). Pharmacogenetics of soluble sulfotransferases (SULTs). *Naunyn-Schmiedeberg's archives of pharmacology*, *369*(1), 55-68.
- Glatt, H., Boeing, H., Engelke, C. E., Ma, L., Kuhlow, A., Pabel, U., ... & Meinl, W. (2001). Human cytosolic sulphotransferases: genetics, characteristics, toxicological aspects. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *482*(1-2), 27-40.
- Glatt, H., Engelke, C. E., Pabel, U., Teubner, W., Jones, A. L., Coughtrie, M. W., ... & Meinl, W. (2000). Sulfotransferases: genetics and role in toxicology. *Toxicology letters*, *112*, 341-348.
- Goekkurt, E., Al-Batran, S. E., Hartmann, J. T., Mogck, U., Schuch, G., Kramer, M., ... & Stoehlmacher, J. (2009). Pharmacogenetic analyses of a phase III trial in metastatic gastroesophageal adenocarcinoma with fluorouracil and leucovorin plus either oxaliplatin or cisplatin: a study of the arbeitsgemeinschaft internistische onkologie. *Journal of Clinical Oncology*, *27*(17), 2863-2873.

- Goldstein, J. A., & Faletto, M. B. (1993). Advances in mechanisms of activation and deactivation of environmental chemicals. *Environmental health perspectives*, *100*, 169-176.
- Goodsell, D. S. (1999). The molecular perspective: the ras oncogene. *Stem cells*, *17*(4), 235-236.
- Gorący, J., Bogacz, A., Uzar, I., Wolek, M., Łochyńska, M., Ziętek, P., ... & Kamiński, A. (2021). The Analysis of NADPH Quinone Reductase 1 (NQO1) Polymorphism in Polish Patients with Colorectal Cancer. *Biomolecules*, *11*(7), 1024.
- Goričar, K., Kovač, V., Jazbec, J., Zakotnik, B., Lamovec, J., & Dolžan, V. (2015). Genetic variability of DNA repair mechanisms and glutathione-S-transferase genes influences treatment outcome in osteosarcoma. *Cancer epidemiology*, *39*(2), 182-188.
- Goyette, P., Pai, A., Milos, R., Frosst, P., Tran, P., Chen, Z., ... & Rozen, R. (1998). Gene structure of human and mouse methylenetetrahydrofolate reductase (MTHFR). *Mammalian Genome*, *9*(8), 652-656.
- Goyette, P., Sumner, J. S., Milos, R., Duncan, A. M., Rosenblatt, D. S., Matthews, R. G., & Rozen, R. (1994). Human methylenetetrahydrofolate reductase: isolation of cDNA, mapping and mutation identification. *Nature genetics*, *7*(2), 195-200.
- Graziano, C., Comin, C. E., Crisci, C., Novelli, L., Politi, L., Messerini, L., ... & Porfirio, B. (2009). Functional polymorphisms of the microsomal epoxide hydrolase gene: a reappraisal on a early-onset lung cancer patients series. *Lung Cancer*, *63*(2), 187-193.
- Graziano, S. L., Gamble, G. P., Newman, N. B., Abbott, L. Z., Rooney, M., Mookherjee, S., ... & Poiesz, B. J. (1999). Prognostic significance of K-ras codon 12 mutations in patients with resected stage I and II non–small-cell lung cancer. *Journal of Clinical Oncology*, *17*(2), 668-668.
- Gresner, P., Gromadzinska, J., & Wasowicz, W. (2007). Polymorphism of selected enzymes involved in detoxification and biotransformation in relation to lung cancer. *Lung cancer*, *57*(1), 1-25.
- Gsur, A., Zidek, T., Schnattinger, K., Feik, E., Haidinger, G., Hollaus, P., ... & Micksche, M. (2003). Association of microsomal epoxide hydrolase polymorphisms and lung cancer risk. *British journal of cancer*, *89*(4), 702-706.
- Guenther, B. D., Sheppard, C. A., Tran, P., Rozen, R., Matthews, R. G., & Ludwig, M. L. (1999). The structure and properties of methylenetetrahydrofolate reductase from Escherichia coli suggest how folate ameliorates human hyperhomocysteinemia. *Nature structural biology*, *6*(4), 359-365.
- Guidelines for the Early Detection of Cancer. American Cancer Society website. https://www.cancer.org/healthy/find-cancer-early/cancer-screeningguidelines/american-cancer-society-guidelines-for-the early detection-of-cancer.html. Published 2018

- Guo, S., Gao, M., Li, X., Li, Y., Chu, S., Zhu, D., & Niu, W. (2012). Lack of association between NADPH quinone oxidoreductase 1 (NQO1) gene C609T polymorphism and lung cancer: a case-control study and a meta-analysis. *PloS one*, *7*(10), e47939.
- Gupta, D., Boffetta, P., Gaborieau, V., & Jindal, S. K. (2001). Risk factors of lung cancer in Chandigarh, India. *Indian Journal of Medical Research*, *113*, 142.
- Gupta, R. (2006). Smoking, educational status & health inequity in India. *Indian Journal of Medical Research*, *124*(1), 15.
- Gupta, U., Mir, S. S., Garg, N., Agarwal, S. K., Pande, S. and Mittal, B. (2016), 'Association study of inflammatory genes with rheumatic heart disease in north Indian population: A multi-analytical approach', *Immunology letters* 174, 53–62.
- Habalova, V., Salagovic, J., Kalina, I., & Stubna, J. (2005). A pilot study testing the genetic polymorphism of N-acetyltransferase 2 as a risk factor in lung cancer. *Neoplasma*, *52*(5), 364-368.
- Habuchi, O. (2000). Diversity and functions of glycosaminoglycan sulfotransferases. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1474*(2), 115-127.
- Hajos, A. K., & Winston, G. W. (1991). Dinitropyrene nitroreductase activity of purified NAD (P) H-quinone oxidoreductase: role in rat liver cytosol and induction by Aroclor-1254 pretreatment. *Carcinogenesis*, *12*(4), 697-702.
- Hammock, B. D., Grant, D., & Storms, D. (1997). Epoxide hydrolases. *Comprehensive toxicology*, *3*, 283-305.
- Han, B., Guo, Z., Ma, Y., Kang, S., Wang, Y., Wei, Q., & Wu, X. (2015). Association of GSTP1 and XRCC1 gene polymorphisms with clinical outcome of advanced non-small cell lung cancer patients with cisplatin-based chemotherapy. *International journal of clinical and experimental pathology*, *8*(4), 4113.
- Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *cell*, *100*(1), 57-70.
- Hansen, J. L., Schmeing, T. M., Klein, D. J., Ippolito, J. A., Ban, N., Nissen, P., ... & Steitz, T. A. (2001, January). Progress toward an understanding of the structure and enzymatic mechanism of the large ribosomal subunit. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 66, pp. 33-42). Cold Spring Harbor Laboratory Press.
- Hartsfield Jr, J. K., Sutcliffe, M. J., Everett, E. T., Hassett, C., Omiecinski, C. J., & Saari, J. A. (1998). Assignment1 of microsomal epoxide hydrolase (EPHX1) to human chromosome 1q42. 1 by in situ hybridization. *Cytogenetic and Genome Research*, *83*(1-2), 44-45.
- Hassett, C., Alcher, L., Sidhu, J. S., & Omieclnskl, C. J. (1994). Human microsomal epoxide hydrolase: genetic poloymorphism and functional expression in vitro of amino acid variants. *Human molecular genetics*, *3*(3), 421-428.
- Hayes, J. D., & Strange, R. C. (2000). Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology*, *61*(3), 154-166.

- Hayes, J. D., Flanagan, J. U., & Jowsey, I. R. (2005). Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.*, *45*, 51-88.
- Hayes, R. B. (1997). The carcinogenicity of metals in humans. *Cancer Causes & Control*, *8*(3), 371-385.
- Hecht, S. S. (1999). Tobacco smoke carcinogens and lung cancer. *JNCI: Journal of the National Cancer Institute*, *91*(14), 1194-1210.
- Hecht, S. S. (2011). Tobacco smoke carcinogens and lung cancer. In *Chemical carcinogenesis* (pp. 53-74). Humana Press.
- Hein, D. W. (1988). Acetylator genotype and arylamine-induced carcinogenesis. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *948*(1), 37-66.
- Hein, D. W., Doll, M. A., Fretland, A. J., Leff, M. A., Webb, S. J., Xiao, G. H., ... & Feng, Y. (2000). Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiology and Prevention Biomarkers*, *9*(1), 29-42.
- Hein, D. W., Doll, M. A., Rustan, T. D., Gray, K., Feng, Y., Ferguson, R. J., & Grant, D. M. (1993). Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases. *Carcinogenesis*, *14*(8), 1633-1638.
- Hempel, N., Gamage, N., Martin, J. L., & McManus, M. E. (2007). Human cytosolic sulfotransferase SULT1A1. *The international journal of biochemistry & cell biology*, *39*(4), 685-689.
- Henderson, A. P., Bleasdale, C., Delaney, K., Lindstrom, A. B., Rappaport, S. M., Waidyanatha, S., ... & Golding, B. T. (2005). Evidence for the formation of Michael adducts from reactions of (E, E)-muconaldehyde with glutathione and other thiols. *Bioorganic chemistry*, *33*(5), 363-373.
- Hickman, D., Risch, A., Buckle, V., Spurr, N. K., Jeremiah, S. J., McCarthy, A., & Sim, E. (1994). Chromosomal localization of human genes for arylamine N-acetyltransferase. *Biochemical Journal*, *297*(3), 441-445.
- Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature reviews genetics*, *6*(2), 95-108.
- Hoffmann, D. H. I. (1997). The changing cigarette, 1950-1995. *Journal of Toxicology and Environmental Health Part A*, *50*(4), 307-364.
- Hoffmann, D., Hoffmann, I., & El-Bayoumy, K. (2001). The less harmful cigarette: a controversial issue. A tribute to Ernst L. Wynder. *Chemical Research in Toxicology*, *14*(7), 767 -790.
- Hou, L., & Lin, L. (2016). NQO1 overexpression is associated with poor prognosis in the uterine cervix carcinoma. *Practical Oncology Journal*, *30*(4), 351-355.

- Hou, S. M., Ryberg, D., Fält, S., Deverill, A., Tefre, T., Børresen, A. L., ... & Lambert, B. (2000). GSTM1 and NAT2 polymorphisms in operable and non-operable lung cancer patients. *Carcinogenesis*, *21*(1), 49-54.
- https://ghr.nlm.nih.gov/condition/lung-cancer
- https://www.genecards.org/cgi-bin/carddisp.pl?gene=MTHFR&keywords=MTHFR
- https://www.genecards.org/cgi-bin/carddisp.pl?gene=SULT1A1
- https://www.medicinenet.com/lung_cancer/article.htm
- https://www.who.int/news-room/fact-sheets/detail/cancer
- Huang, J., Lin, H., Wu, X., Jin, W., & Zhang, Z. (2017). NQO1 C609T polymorphism and lung cancer susceptibility: Evidence from a comprehensive meta-analysis. *Oncotarget*, *8*(60), 102301.
- Huang, M. T., Miwa, G. T., Cronheim, N., & Lu, A. Y. H. (1979). Rat liver cytosolic azoreductase. *J. biol. Chem*, *254*, 11223-11227.
- Huang, S. K., Chiu, A. W. H., Pu, Y. S., Huang, Y. K., Chung, C. J., Tsai, H. J., ... & Hsueh, Y. M. (2009). Arsenic methylation capability, myeloperoxidase and sulfotransferase genetic polymorphisms, and the stage and grade of urothelial carcinoma. *Urologia internationalis*, *82*(2), 227-234.
- Huang, J., Lin, H., Wu, X., Jin, W., & Zhang, Z. (2017). NQO1 C609T polymorphism and lung cancer susceptibility: evidence from a comprehensive meta-analysis. *Oncotarget*, *8*(60), 102301.
- Huang, X., Cao, M., Wang, L., Wu, S., Liu, X., Li, H., ... & Fang, B. (2015). Expression of sulfotransferase SULT1A1 in cancer cells predicts susceptibility to the novel anticancer agent NSC-743380. *Oncotarget*, *6*(1), 345.
- Huang, Y., & Chang, Y. (2011). Epidermal growth factor receptor (EGFR) phosphorylation, signaling and trafficking in prostate cancer. *Prostate cancer-from bench to bedside*, *143*, 72.
- Huggins, C. (1979). *Experimental leukemia and mammary cancer: induction, prevention, cure*. University of Chicago Press.
- Hung, R. J., Boffetta, P., Brennan, P., Malaveille, C., Hautefeuille, A., Donato, F., ... & Porru, S. (2004). GST, NAT, SULT1A1, CYP1B1 genetic polymorphisms, interactions with environmental exposures and bladder cancer risk in a high-risk population. *International Journal of Cancer*, *110*(4), 598-604.
- Hunter, D. J., Hankinson, S. E., Hough, H., Gertig, D. M., Garcia-Closas, M., Spiegelman, D., ... & Kelsey, K. (1997). A prospective study of NAT2 acetylation genotype, cigarette smoking, and risk of breast cancer. *Carcinogenesis*, *18*(11), 2127-2132.
- IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, World Health Organization, & International Agency for Research on Cancer. (2004). *Tobacco smoke and involuntary smoking* (No. 83). Iarc.

- Ibarrola-Villava, M., Martin-Gonzalez, M., Lazaro, P., Pizarro, A., Lluch, A., & Ribas, G. (2012). Role of glutathione S-transferases in melanoma susceptibility: association with GSTP1 rs1695 polymorphism. *British Journal of Dermatology*, *166*(6), 1176-1183.
- Igari, S., Ohtaki, A., Yamanaka, Y., Sato, Y., Yohda, M., Odaka, M., ... & Yamada, K. (2011). Properties and crystal structure of methylenetetrahydrofolate reductase from Thermus thermophilus HB8. *PLoS One*, *6*(8), e23716.
- Ihsan, R., Chauhan, P. S., Mishra, A. K., Yadav, D. S., Kaushal, M., Sharma, J. D., ... & Saxena, S. (2011). Multiple analytical approaches reveal distinct gene-environment interactions in smokers and non smokers in lung cancer. *PloS one*, *6*(12), e29431.
- International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*(7164), 851.
- Irigaray, P., & Belpomme, D. (2010). Basic properties and molecular mechanisms of exogenous chemical carcinogens. *Carcinogenesis*, *31*(2), 135-148.
- Jackson, M. R., Craft, J. A., & Burchell, B. (1987). Nucleotide and deduced amino acid sequence of human liver microsomal epoxide hydrolase. *Nucleic acids research*, *15*(17), 7188.
- Jaiswal, A. K. (1991). Human NAD (P) H: quinone oxidoreductase (NQO1) gene structure and induction by dioxin. *Biochemistry*, *30*(44), 10647-10653.
- Jaiswal, A. K., McBride, O. W., Adesnik, M., & Nebert, D. W. (1988). Human dioxin-inducible cytosolic NAD (P) H: menadione oxidoreductase. cDNA sequence and localization of gene to chromosome 16. *Journal of Biological Chemistry*, *263*(27), 13572-13578.
- Jakobsen, A., Nielsen, J. N., Gyldenkerne, N., & Lindeberg, J. (2005). Thymidylate synthase and methylenetetrahydrofolate reductase gene polymorphism in normal tissue as predictors of fluorouracil sensitivity. *Journal of clinical oncology*, *23*(7), 1365-1369.
- Jamison, R. L., Shih, M. C., Humphries, D. E., Guarino, P. D., Kaufman, J. S., Goldfarb, D. S., ... & Veterans Affairs Site Investigators. (2009). Effect of the MTHFR C677T and A1298C polymorphisms on survival in patients with advanced CKD and ESRD: a prospective study. *American journal of kidney diseases*, *53*(5), 779-789.
- Jancova, P., Anzenbacher, P., & Anzenbacherova, E. (2010). Phase II drug metabolizing enzymes. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*, *154*(2), 103-16.
- Järup, L., & Pershagen, G. (1991). Arsenic exposure, smoking, and lung cancer in smelter workers- a case-control study. *American Journal of Epidemiology*, *134*(6), 545-551.
- Jayalekshmy, P. A., Akiba, S., Nair, M. K., Gangadharan, P., Rajan, B., Nair, R. K., & Sugahara, T. (2008). Bidi smoking and lung cancer incidence among males in Karunagappally cohort in Kerala, India. *International journal of cancer*, *123*(6), 1390-1397.
- Jayant, K., & Pakhale, S. S. (1989). Toxic constituents in bidi smoke. *Tobacco and Health: The Indian Scene. Bombay: Tata Memorial Centre*, 101-10.

- Jayanthi, M. K., Prathima, C., & Rao, M. S. (2015). NQO1, role in lung Cancer: a review. *Journal of Pharmaceutical Sciences and Research*, *7*(6), 319.
- Jennne, J. W. (1965). Partial purification and properties of the isoniazid transacetylase in human liver. Its relationship to the acetylation of p-aminosalicylic acid. *The Journal of clinical investigation*, *44*(12), 1992-2002.
- Jennne, J. W. (1965). Partial purification and properties of the isoniazid transacetylase in human liver. Its relationship to the acetylation of p-aminosalicylic acid. *The Journal of clinical investigation*, *44*(12), 1992-2002.
- Ji, M., Jin, A., Sun, J., Cui, X., Yang, Y., Chen, L., & Lin, Z. (2017). Clinicopathological implications of NQO1 overexpression in the prognosis of pancreatic adenocarcinoma. *Oncology letters*, *13*(5), 2996-3002.
- Ji, Y. N., Wang, Q., & Suo, L. J. (2012). CYP1A1 Ile462Val polymorphism contributes to lung cancer susceptibility among lung squamous carcinoma and smokers: a meta-analysis.
- Jia, W., Sun, J. Y., Jia, K. Y., & Liu, X. C. (2016). Role of GSTM1, GSTT1, and GSTP1 IIe105Val gene polymorphisms in the response to chemotherapy and overall survival of advanced non-small cell lung cancer. *Genet Mol Res*, *15*(3).
- Jin, Y. (2019). Association between EPHX1 polymorphism rs1051740 and the risk of ovarian cancer: a meta-analysis. *Artificial Cells, Nanomedicine, and Biotechnology*, *47*(1), 2338-2342.
- Jindal, S. K., & Behera, D. (1990). Clinical spectrum of primary lung cancer-review of Chandigarh experience of 10 years. *Lung India*, *8*(2), 94.
- Jindal, S. K., Aggarwal, A. N., Chaudhry, K., Chhabra, S. K., D Souza, G. A., Gupta, D., ... & Asthma Epidemiology Study Group. (2006). Tobacco smoking in India: prevalence, quit-rates and respiratory morbidity. *Indian Journal of Chest Diseases and Allied Sciences*, *48*(1), 37.
- Johansson, B. G. (1972). Agarose gel electrophoresis. *Scandinavian Journal of Clinical and Laboratory Investigation*, *29*(sup124), 7-19.
- Jones, A. L., Hume, R., Bamforth, K. J., & Coughtrie, M. W. (1992). Estrogen and phenol sulfotransferase activities in human fetal lung. *Early human development*, *28*(1), 65-77.
- Jussawalla, D. J., & Jain, D. K. (1979). Lung cancer in Greater Bombay: correlations with religion and smoking habits. *British journal of cancer*, *40*(3), 437-448.
- Kaiser, U., Schilli, M., Haag, U., Neumann, K., Kreipe, H., Kogan, E., & Havemann, K. (1996). Expression of bcl-2—protein in small cell lung cancer. *Lung cancer*, *15*(1), 31-40.
- Kakuta, Y., Petrotchenko, E. V., Pedersen, L. C., & Negishi, M. (1998). The sulfuryl transfer mechanism; crystal structure of a vanadate complex of estrogen sulfotransferase and mutational analysis. Journal of Biological Chemistry, 273, 27325–27330.

- Kalikaki, A., Kanaki, M., Vassalou, H., Souglakos, J., Voutsina, A., Georgoulias, V., & Mavroudis, D. (2009). DNA repair gene polymorphisms predict favorable clinical outcome in advanced non–small-cell lung cancer. *Clinical lung cancer*, *10*(2), 118-123.
- Kanwal, R., Pandey, M., Bhaskaran, N., MacLennan, G. T., Fu, P., Ponsky, L. E., & Gupta, S. (2014). Protection against oxidative DNA damage and stress in human prostate by glutathione S-transferase P1. *Molecular carcinogenesis*, *53*(1), 8-18.
- Ke, H. G., Li, J., Shen, Y., You, Q. S., Yan, Y., Dong, H. X., ... & Shen, Z. Y. (2012). Prognostic significance of GSTP1, XRCC1 and XRCC3 polymorphisms in non-small cell lung cancer patients. *Asian Pacific Journal of Cancer Prevention*, *13*(9), 4413-4416.
- Kepa, J. K., Traver, R. D., Siegel, D., Winski, S. L., & Ross, D. (1997). Regulation and function of NAD (P) H: quinone oxidoreductase(NQO 1). *Reviews in Toxicology*, *1*(7), 53-73.
- Keum, J. S., Kong, G., Yang, S. C., Shin, D. H., Park, S. S., Lee, J. H., & Lee, J. D. (1999). Cyclin D1 overexpression is an indicator of poor prognosis in resectable non-small cell lung cancer. *British journal of cancer*, *81*(1), 127-132.
- Khrunin, A. V., Khokhrin, D. V., Moisseev, A. A., Gorbunova, V. A., & Limborska, S. A. (2014). Pharmacogenomic assessment of cisplatin-based chemotherapy outcomes in ovarian cancer. *Pharmacogenomics*, *15*(3), 329-337.
- Khrunin, A., Ivanova, F., Moisseev, A., Khokhrin, D., Sleptsova, Y., Gorbunova, V., & Limborska, S. (2012). Pharmacogenomics of cisplatin-based chemotherapy in ovarian cancer patients of different ethnic origins. *Pharmacogenomics*, *13*(2), 171-178.
- Khrunin, A., Moisseev, A., Gorbunova, V., & Limborska, S. (2018). Ethnic Differences in Susceptibility to the Effects of Platinum-Based Chemotherapy. *Ovarian Cancer: From Pathogenesis to Treatment*, 309.
- Khuder, S. A. (2001). Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung cancer*, *31*(2-3), 139-148.
- Kim, J. H., & Hong, Y. C. (2015). Interactive effect of smoking and NQO1 haplotypes on lung cancer risk. *Journal of Korean medical science*, *30*(3), 221-226.
- Kim, Y. I. (1999). Folate and carcinogenesis: evidence, mechanisms, and implications. *The Journal of nutritional biochemistry*, *10*(2), 66-88.
- Kirschbaum, M. H., & Yarden, Y. (2000). The ErbB/HER family of receptor tyrosine kinases: a potential target for chemoprevention of epithelial neoplasms. *Journal of Cellular Biochemistry*, *77*(S34), 52-60.
- Kitamura, H., Yazawa, T., Okudela, K., Shimoyamada, H., & Sato, H. (2008). Molecular and genetic pathogenesis of lung cancer: differences between small-cell and non-small-cell carcinomas. *The Open Pathology Journal*, *2*(1).
- Kitteringham, N. R., Davis, C., Howard, N., Pirmohamed, M., & Park, B. K. (1996). Interindividual and interspecies variation in hepatic microsomal epoxide hydrolase activity: studies with cis-stilbene oxide, carbamazepine 10, 11-epoxide and

naphthalene. *Journal of pharmacology and experimental therapeutics*, *278*(3), 1018-1027.

- Kiyohara, C., Yoshimasu, K., Takayama, K., & Nakanishi, Y. (2006). EPHX1 polymorphisms and the risk of lung cancer: a HuGE review. *Epidemiology*, 89-99.
- Knox, R. J., Boland, M. P., Friedlos, F., Coles, B., Southan, C., & Roberts, J. J. (1988). The nitroreductase enzyme in Walker cells that activates 5-(aziridin-1-yl)-2, 4-dinitrobenzamide (CB 1954) to 5-(aziridin-1-yl)-4-hydroxylamino-2-nitrobenzamide is a form of NAD (P) H dehydrogenase (quinone)(EC 1.6. 99.2). *Biochemical pharmacology*, *37*(24), 4671-4677.
- Koifman, S. (2012). SULT1A1 genetic polymorphisms and the association between smoking and oral cancer in a case-control study in Brazil. *Frontiers in oncology*, *2*, 183.
- Koike, H., Nakazato, H., Ohtake, N., Matsui, H., Okugi, H., Shibata, Y., ... & Suzuki, K. (2008). Further evidence for null association of phenol sulfotransferase SULT1A1 polymorphism with prostate cancer risk: a case–control study of familial prostate cancer in a Japanese population. *International urology and nephrology*, *40*(4), 947-951.
- Kolesar, J. M., Dahlberg, S. E., Marsh, S., McLeod, H. L., Johnson, D. H., Keller, S. M., & Schiller, J. H. (2011). The NQO1* 2/* 2 polymorphism is associated with poor overall survival in patients following resection of stages II and IIIa non-small cell lung cancer. *Oncology reports*, *25*(6), 1765-1772.
- Kondo, T., Yamada, K., Noda, K., Nakayama, H., & Kameda, Y. (2002). Radiologic–prognostic correlation in patients with small pulmonary adenocarcinomas. *Lung Cancer*, *36*(1), 49-57.
- Kong, S., Ye, F., Dang, Y., Hua, Y., & Zhang, G. (2020). Association of MTHFR C677T polymorphism with severity and localization of chronic atrophic gastritis patients without Helicobacter pylori infection: a case control study. *BMC cancer*, *20*(1), 1-9.
- König-MeredizSven, A. (2015). Design and structural requirements of the potent and safe TLR-9 agonistic immunomodulator MGN1703. *Nucleic acid therapeutics*.
- Kosaka, T., Yatabe, Y., Endoh, H., Kuwano, H., Takahashi, T., & Mitsudomi, T. (2004). Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer research*, *64*(24), 8919-8923.
- Krajinovic, M., Lamothe, S., Labuda, D., Lemieux-Blanchard, E., Théorêt, Y., Moghrabi, A., & Sinnett, D. (2004). Role of MTHFR genetic polymorphisms in the susceptibility to childhood acute lymphoblastic leukemia. *Blood*, *103*(1), 252-257.
- Krawczyk, P., Kucharczyk, T., Kowalski, D. M., Powrózek, T., Ramlau, R., Kalinka-Warzocha, E., ... & Milanowski, J. (2014). Polymorphisms in TS, MTHFR and ERCC1 genes as predictive markers in first-line platinum and pemetrexed therapy in NSCLC patients. *Journal of cancer research and clinical oncology*, *140*(12), 2047-2057.
- Kucharczyk, T., Krawczyk, P., Powrózek, T., Kowalski, D. M., Ramlau, R., Kalinka-Warzocha, E., ... & Milanowski, J. (2016). The effectiveness of pemetrexed monotherapy

depending on polymorphisms in TS and MTHFR genes as well as clinical factors in advanced NSCLC patients. *Pathology & Oncology Research*, *22*(1), 49-56.

- Kumar, K., Kandula, M., Bhavani, V., Laxmi, A., & Jamil, K. (2013). Sulfotransferase 1A1 (SULT1A1) Polymorphism and Breast Cancer Risk: A Case-control Study in South India. *American Journal of Cancer Science*, *2*(2), 108-115.
- Kumar, R., Prakash, S., Kushwah, A. S., & Vijayan, V. K. (2010). Breath carbon monoxide concentration in cigarette and bidi smokers in India. *Religion*, *56*, 34-7.
- Kutzbach, C., & Stokstad, E. L. R. (1971). Mammalian methylenetetrahydrofolate reductase Partial purification, properties, and inhibition by S-adenosylmethionine. *Biochimica et Biophysica Acta (BBA)-Enzymology*, *250*(3), 459-477.
- Laasanen, J., Romppanen, E. L., Hiltunen, M., Helisalmi, S., Mannermaa, A., Punnonen, K., & Heinonen, S. (2002). Two exonic single nucleotide polymorphisms in the microsomal epoxide hydrolase gene are jointly associated with preeclampsia. *European journal of human genetics*, *10*(9), 569-573.
- Lacourciere, G. M., Vakharia, V. N., Tan, C. P., Morris, D. I., Edwards, G. H., Moos, M., & Armstrong, R. N. (1993). Interaction of hepatic microsomal epoxide hydrolase derived from a recombinant baculovirus expression system with an azarene oxide and an aziridine substrate analog. *Biochemistry*, *32*(10), 2610-2616.
- Lajin, B., & Alachkar, A. (2013). The NQO1 polymorphism C609T (Pro187Ser) and cancer susceptibility: a comprehensive meta-analysis. *British journal of cancer*, *109*(5), 1325-1337.
- Lam, T. K., Gallicchio, L., Lindsley, K., Shiels, M., Hammond, E., Tao, X. G., ... & Alberg, A. J. (2009). Cruciferous vegetable consumption and lung cancer risk: a systematic review. *Cancer Epidemiology and Prevention Biomarkers*, *18*(1), 184-195.
- Lam, T. K., Gallicchio, L., Lindsley, K., Shiels, M., Hammond, E., Tao, X. G., ... & Alberg, A. J. (2009). Cruciferous vegetable consumption and lung cancer risk: a systematic review. *Cancer Epidemiology and Prevention Biomarkers*, *18*(1), 184-195.
- Lambert, A. J., Friedlos, F., Boland, M. P., & Knox, R. J. (1992). A CB 1954 analogue that is highly cytotoxic in human tumour cell lines. *Br J Cancer*, *65*(Suppl. XVI), 59.
- Lan, G., Lin, L., Chen, X., Chen, L., & Chen, X. (2017). Correlation between methylenetetrahydrofolate reductase (MTHFR) C677T polymorphisms and pemetrexed chemotherapy efficacy/toxicity in non-squamous non-small cell lung cancer. *Medical science monitor: international medical journal of experimental and clinical research*, *23*, 5683.
- Lantz, D. A., Dougherty, T. H., & Lucca, M. J. (1989). Primary angiosarcoma of the heart causing cardiac rupture. *American Heart Journal*, *118*(1), 186-188.

- Larsson, S. C., Håkansson, N., Giovannucci, E., & Wolk, A. (2006). Folate intake and pancreatic cancer incidence: a prospective study of Swedish women and men. *Journal of the National Cancer Institute*, *98*(6), 407-413.
- Laurieri, N., Kawamura, A., Westwood, I. M., Varney, A., Morris, E., Russell, A. J., ... & Sim, E. (2014). Differences between murine arylamine N-acetyltransferase type 1 and human arylamine N-acetyltransferase type 2 defined by substrate specificity and inhibitor binding. *BMC Pharmacology and Toxicology*, *15*(1), 1-13.
- Le Marchand, L., Donlon, T., Hankin, J. H., Kolonel, L. N., Wilkens, L. R., & Seifried, A. (2002). B-vitamin intake, metabolic genes, and colorectal cancer risk (United States). *Cancer Causes & Control*, *13*(3), 239-248.
- Lee, C. N., Yu, M. C., Bai, K. J., Chang, J. H., Fang, C. L., Hsu, H. L., ... & Liu, H. E. (2009). NAT2 fast acetylator genotypes are associated with an increased risk for lung cancer with wildtype epidermal growth factor receptors in Taiwan. *Lung Cancer*, *64*(1), 9-12.
- Lee, H., Wang, Q., Yang, F., Tao, P., Li, H., Huang, Y., & Li, J. Y. (2012). SULT1A1 Arg213His polymorphism, smoked meat, and breast cancer risk: a case–control study and meta-analysis. *DNA and cell biology*, *31*(5), 688-699.
- Lee, J., Park, J. O., Kim, W. S., Park, S. H., Park, K. W., Choi, M. S., ... & Park, K. (2004). Phase II study of doxorubicin and cisplatin in patients with metastatic hepatocellular carcinoma. *Cancer chemotherapy and pharmacology*, *54*(5), 385-390.
- Lee, K. H., Chang, H. J., Han, S. W., Oh, D. Y., Im, S. A., Bang, Y. J., ... & Kim, T. Y. (2013). Pharmacogenetic analysis of adjuvant FOLFOX for Korean patients with colon cancer. *Cancer chemotherapy and pharmacology*, *71*(4), 843-851.
- Lee, M. N., Takawira, D., Nikolova, A. P., Ballou, D. P., Furtado, V. C., Phung, N. L., ... & Trimmer, E. E. (2009). Functional role for the conformationally mobile phenylalanine 223 in the reaction of methylenetetrahydrofolate reductase from Escherichia coli. *Biochemistry*, *48*(32), 7673-7685.
- Lee, S. J., Kim, W. Y., Jarrar, Y. B., Kim, Y. W., Lee, S. S., & Shin, J. G. (2013). Single Nucleotide Polymorphisms in SULT1A1 and SULT1A2 in a Korean population. *Drug metabolism and pharmacokinetics*, DMPK-12.
- Lee, W. J., Brennan, P., Boffetta, P., London, S. J., Benhamou, S., Rannug, A., ... & Taioli, E. (2002). Microsomal epoxide hydrolase polymorphisms and lung cancer risk: a quantitative review. *Biomarkers*, *7*(3), 230-241.
- Leventakos, K., & Mansfield, A. S. (2016). Advances in the treatment of non–small cell lung cancer: focus on nivolumab, pembrolizumab, and atezolizumab. *BioDrugs*, *30*(5), 397-405.
- Lewis, D. F., Lake, B. G., & Bird, M. G. (2005). Molecular modelling of human microsomal epoxide hydrolase (EH) by homology with a fungal (Aspergillus niger) EH

crystal structure of 1.8 Å resolution: Structure–activity relationships in epoxides inhibiting EH activity. *Toxicology in vitro*, *19*(4), 517-522.

- Lewis, S. J., Cherry, N. M., Niven, R. M., Barber, P. V., & Povey, A. C. (2001). Polymorphisms in the NAD (P) H: quinone oxidoreductase gene and small cell lung cancer risk in a UK population. *Lung Cancer*, *34*(2), 177-183.
- Li, H., Fu, W. P., & Hong, Z. H. (2013). Microsomal epoxide hydrolase gene polymorphisms and risk of chronic obstructive pulmonary disease: A comprehensive meta-analysis. *Oncology letters*, *5*(3), 1022-1030.
- Li, R., Bianchet, M. A., Talalay, P., & Amzel, L. M. (1995). The three-dimensional structure of NAD (P) H: quinone reductase, a flavoprotein involved in cancer chemoprotection and chemotherapy: mechanism of the two-electron reduction. *Proceedings of the National Academy of Sciences*, *92*(19), 8846-8850.
- Li, W. J., Jiang, H., Fang, X. J., Ye, H. L., Liu, M. H., Liu, Y. W., ... & Zhang, Q. Y. (2013). Polymorphisms in thymidylate synthase and reduced folate carrier (SLC19A1) genes predict survival outcome in advanced non-small cell lung cancer patients treated with pemetrexed-based chemotherapy. *Oncology letters*, *5*(4), 1165-1170.
- Li, Y., Lindsay, J., Wang, L. L., & Zhou, S. F. (2008). Structure, function and polymorphism of human cytosolic sulfotransferases. *Current drug metabolism*, *9*(2), 99-105.
- Li, Z., Zhang, Y., Jin, T., Men, J., Lin, Z., Qi, P., ... & Yan, G. (2015). NQO1 protein expression predicts poor prognosis of non-small cell lung cancers. *BMC cancer*, *15*(1), 1-9.
- Liang, G., Miao, X., Zhou, Y., Tan, W., & Lin, D. (2004). A functional polymorphism in the SULT1A1 gene (G638A) is associated with risk of lung cancer in relation to tobacco smoking. *Carcinogenesis*, *25*(5), 773-778.
- Liang, S. H., Hassett, C., & Omiecinski, C. J. (2005). Alternative promoters determine tissue-specific expression profiles of the human microsomal epoxide hydrolase gene (EPHX1). *Molecular pharmacology*, *67*(1), 220-230.
- Liao, S. G., Liu, L., Zhang, Y. Y., Wang, Y., & Wang, Y. J. (2012). SULT1A1 Arg213His polymorphism and lung cancer risk: a meta-analysis. *Asian Pacific Journal of Cancer Prevention*, *13*(2), 579-583.
- Liberman, P. H. P., Goffi-Gomez, M. V. S., Schultz, C., Jacob, P. L., de Paula, C. A. A., Sartorato, E. L., ... & Carraro, D. M. (2019). Contribution of the GSTP1 c. 313A> G variant to hearing loss risk in patients exposed to platin chemotherapy during childhood. *Clinical and translational oncology*, *21*(5), 630-635.
- Lilla, C., Risch, A., Kropp, S., & Chang-Claude, J. (2005). SULT1A1 genotype, active and passive smoking, and breast cancer risk by age 50 years in a German case–control study. *Breast Cancer Research*, *7*(2), 1-9.

- Lim, W., Ridge, C. A., Nicholson, A. G., & Mirsadraee, S. (2018). The 8th lung cancer TNM classification and clinical staging system: review of the changes and clinical implications. *Quantitative imaging in medicine and surgery*, *8*(7), 709.
- Lin, L., Qin, Y., Jin, T., Liu, S., Zhang, S., Shen, X., & Lin, Z. (2014). Significance of NQO1 overexpression for prognostic evaluation of gastric adenocarcinoma. *Experimental and molecular pathology*, *96*(2), 200-205.
- Lin, P., Hsueh, Y. M., Ko, J. L., Liang, Y. F., Tsai, K. J., & Chen, C. Y. (2003). Analysis of NQO1, GSTP1, and MnSOD genetic polymorphisms on lung cancer risk in Taiwan. *Lung cancer*, *40*(2), 123-129.
- Lin, P., Wang, S. L., Wang, H. J., Chen, K. W., Lee, H. S., Tsai, K. J., ... & Lee, H. (2000). Association of CYP1A1 and microsomal epoxide hydrolase polymorphisms with lung squamous cell carcinoma. *British journal of cancer*, *82*(4), 852-857.
- Lin, T. S., Huang, H. H., Fan, Y. H., Chiou, S. H., & Chow, K. C. (2007). Genetic polymorphism and gene expression of microsomal epoxide hydrolase in non-small cell lung cancer. *Oncology reports*, *17*(3), 565-572.
- Lin, Y. S., Hung, S. C., Wei, Y. H., & Tarng, D. C. (2009). GST M1 polymorphism associates with DNA oxidative damage and mortality among hemodialysis patients. *Journal of the American Society of Nephrology*, *20*(2), 405-415.
- Linnebank, M., Semmler, A., Moskau, S., Smulders, Y., Blom, H., & Simon, M. (2008). The methylenetetrahydrofolate reductase (MTHFR) variant c. 677C> T (A222V) influences overall survival of patients with glioblastoma multiforme. *Neuro-oncology*, *10*(4), 548-552.
- Liska, D. J. (1998). The detoxification enzyme systems. *Altern Med Rev*, *3*(3), 187-198.
- Liu, C., Cui, W., Cong, L., Wang, L., Ruan, X., Jia, J., ... & Zhang, X. (2015). Association between NAT2 polymorphisms and lung cancer susceptibility. *Medicine*, *94*(49).
- Liu, G., Zhou, W., & Christiani, D. C. (2005, June). Molecular epidemiology of non-small cell lung cancer. In *Seminars in respiratory and critical care medicine* (Vol. 26, No. 03, pp. 265-272). Copyright© 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
- Liu, H., Li, H. Y., Chen, H. J., Huang, Y. J., Zhang, S., & Wang, J. (2013). EPHX1 A139G polymorphism and lung cancer risk: a meta-analysis. *Tumor Biology*, *34*(1), 155-163.
- Liu, K., Lin, Q., Ding, H., Jin, Y., & Chen, G. (2015). Predictive potential role of GSTs gene polymorphisms in the treatment outcome of advanced non-small cell lung cancer patients. *International journal of clinical and experimental medicine*, *8*(11), 20918.
- Liu, Y., Webb, H. K., Fukushima, H., Micheli, J., Markova, S., Olson, J. L., & Kroetz, D. L. (2012). Attenuation of cisplatin-induced renal injury by inhibition of soluble epoxide

hydrolase involves nuclear factor κB signaling. *Journal of Pharmacology and Experimental Therapeutics*, *341*(3), 725-734.

- Loeb, K. R., & Loeb, L. A. (2000). Significance of multiple mutations in cancer. *Carcinogenesis*, *21*(3), 379-385.
- London, S. J., Smart, J., & Daly, A. K. (2000). Lung cancer risk in relation to genetic polymorphisms of microsomal epoxide hydrolase among African-Americans and Caucasians in Los Angeles County. *Lung Cancer*, *28*(2), 147-155.
- Long, D. J., Waikel, R. L., Wang, X. J., Perlaky, L., Roop, D. R., & Jaiswal, A. K. (2000). NAD (P) H: quinone oxidoreductase 1 deficiency increases susceptibility to benzo (a) pyrene-induced mouse skin carcinogenesis. *Cancer research*, *60*(21), 5913-5915.
- Long, D. J., Waikel, R. L., Wang, X. J., Roop, D. R., & Jaiswal, A. K. (2001). NAD (P) H: quinone oxidoreductase 1 deficiency and increased susceptibility to 7, 12-dimethylbenz [a]-anthracene-induced carcinogenesis in mouse skin. *Journal of the National Cancer Institute*, *93*(15), 1166-1170.
- Loprieno, N. (1975). International Agency for Research on Cancer (IARC) monographs on the evaluation of carcinogenic risk of chemicals to man*: "Relevance of data on mutagenicity". Mutation Research/Environmental Mutagenesis and Related Subjects, 31(3), 201*.
- Lotfi, F., Bahrehmand, F., Vaisi-Raygani, A., Khodarahmi, R., Tanhapour, M., Kiani, A., ... & Pourmotabbed, T. (2018). Cytochrome P450 (CYP450, 2D6* A), N-Acetyltransferase-2 (NAT2* 7, A) and Multidrug Resistance 1 (MDR1 3435 T) Alleles Collectively Increase Risk of Ulcerative Colitis. *Archives of Iranian medicine*, *21*(11), 530-535.
- Lou, Y., Li, R., Xiong, L., Gu, A., Shi, C., Chu, T., ... & Han, B. (2013). NAD (P) H: quinone oxidoreductase 1 (NQO1) C609T polymorphism and lung cancer risk: a meta-analysis. *Tumor Biology*, *34*(6), 3967-3979.
- Lu, C., Spitz, M. R., Zhao, H., Dong, Q., Truong, M., Chang, J. Y., ... & Wu, X. (2006). Association between glutathione S-transferase π polymorphisms and survival in patients with advanced nonsmall cell lung carcinoma. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *106*(2), 441-447.
- Lung cancer - Genetics Home Reference. (2017). Retrieved October 31, 2017, from
- *Lung Cancer* (2019).
- Lung Cancer Screening. Radiology Info.com website. https://www.radiologyinfo.org/en/info.cfm?pg=screening-lung. Revised August 10, 2018.
- Luyapan, J., Ji, X., Li, S., Xiao, X., Zhu, D., Duell, E. J., ... & Gui, J. (2020). A new efficient method to detect genetic interactions for lung cancer GWAS. *BMC Medical Genomics*, *13*(1), 1-15.

- Lv, H., Han, T., Shi, X., Yao, Y., Yao, Y., Qiu, W., ... & Liang, J. (2014). Genetic polymorphism of GSTP1 and ERCC1 correlated with response to platinum-based chemotherapy in non-small cell lung cancer. *Medical Oncology*, *31*(8), 86.
- Mahasneh, A., Jubaili, A., El Bateiha, A., Al-Ghazo, M., Matalka, I., & Malkawi, M. (2012). Polymorphisms of arylamine N-acetyltransferase2 and risk of lung and colorectal cancer. *Genetics and molecular biology*, *35*, 725-733.
- Maher, B. (2008). The case of the missing heritability: when scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. Brendan Maher shines a light on six places where the missing loot could be stashed away. *Nature*, *456*(7218), 18-22.
- Malec, V., Gottschald, O. R., Li, S., Rose, F., Seeger, W., & Hänze, J. (2010). HIF-1α signaling is augmented during intermittent hypoxia by induction of the Nrf2 pathway in NOX1-expressing adenocarcinoma A549 cells. *Free Radical Biology and Medicine*, *48*(12), 1626-1635.
- Malec, V., Gottschald, O. R., Li, S., Rose, F., Seeger, W., & Hänze, J. (2010). HIF-1α signaling is augmented during intermittent hypoxia by induction of the Nrf2 pathway in NOX1-expressing adenocarcinoma A549 cells. *Free Radical Biology and Medicine*, *48*(12), 1626-1635.
- Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C., & Boffetta, P. (2016). Risk factors for lung cancer worldwide. *European Respiratory Journal*, *48*(3), 889-902.
- Malson, J. L., Sims, K., Murty, R., & Pickworth, W. B. (2001). Comparison of the nicotine content of tobacco used in bidis and conventional cigarettes. *Tobacco control*, *10*(2), 181-183.
- Mandal, R. K., Nissar, K., & Mittal, R. D. (2012). Genetic variants in metabolizing genes NQO1, NQO2, MTHFR and risk of prostate cancer: a study from North India. *Molecular biology reports*, *39*(12), 11145-11152.
- Mandal, R. K., Nissar, K., & Mittal, R. D. (2012). Genetic variants in metabolizing genes NQO1, NQO2, MTHFR and risk of prostate cancer: a study from North India. *Molecular biology reports*, *39*(12), 11145-11152.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747-753.
- Martin, D. N., Boersma, B. J., Howe, T. M., Goodman, J. E., Mechanic, L. E., Chanock, S. J., & Ambs, S. (2006). Association of MTHFR gene polymorphisms with breast cancer survival. *BMC cancer*, *6*(1), 1-10.
- Martínez, C., Agundez, J. A., Olivera, M., Martín, R., Ladero, J. M., & Benítez, J. (1995). Lung cancer and mutations at the polymorphic NAT2 gene locus. *Pharmacogenetics*, *5*(4), 207-214.

- Masroor, M., Jain, A., Javid, J., Mir, R., Prashant, Y., Imtiyaz, A., ... & Saxena, A. (2015). Clinical significance of the NQO1 C609T polymorphism in non small cell lung adenocarcinoma patients. *Asian Pacific Journal of Cancer Prevention*, *16*(17), 7653-7658.
- Massion, P. P., & Carbone, D. P. (2003). The molecular basis of lung cancer: molecular abnormalities and therapeutic implications. *Respiratory research*, *4*(1), 1-15.
- Matas, N., Thygesen, P., Stacey, M., Risch, A., & Sim, E. (1997). Mapping AAC1, AAC2 and AACP, the genes for arylamine N-acetyltransferases, carcinogen metabolising enzymes on human chromosome 8p22, a region frequently deleted in tumours. *Cytogenetic and Genome Research*, *77*(3-4), 290-295.
- Mathur, P., Sathishkumar, K., Chaturvedi, M., Das, P., Sudarshan, K. L., Santhappan, S., ... & ICMR-NCDIR-NCRP Investigator Group. (2020). Cancer statistics, 2020: report from national cancer registry programme, India. *JCO Global Oncology*, *6*, 1063-1075.
- Matthews, R. G. (1986). [58] Methylenetetrahydrofolate reductase from pig liver. In *Methods in enzymology* (Vol. 122, pp. 372-381). Academic Press.
- Matthews, R. G., Sheppard, C., & Goulding, C. (1998). Methylenetetrahydrofolate reductase and methionine synthase: biochemistry and molecular biology. *European journal of pediatrics*, *157*(2), S54-S59.
- Matthews, R. G., Vanoni, M. A., Hainfeld, J. F., & Wall, J. (1984). Methylenetetrahydrofolate reductase. Evidence for spatially distinct subunit domains obtained by scanning transmission electron microscopy and limited proteolysis. *Journal of Biological Chemistry*, *259*(19), 11647-11650.
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews genetics*, *9*(5), 356-369.
- McCarthy, W. J., Meza, R., Jeon, J., & Moolgavkar, S. H. (2012). Chapter 6: lung cancer in never smokers: epidemiology and risk prediction models. *Risk Analysis: An International Journal*, *32*, S69-S84.
- McKay, J. D., Hashibe, M., Hung, R. J., Wakefield, J., Gaborieau, V., Szeszenia-Dabrowska, N., ... & Brennan, P. (2008). Sequence variants of NAT1 and NAT2 and other xenometabolic genes and risk of lung and aerodigestive tract cancers in Central Europe. *Cancer Epidemiology and Prevention Biomarkers*, *17*(1), 141-147.
- Medline Plus, MTHFR gene, 2021. https://medlineplus.gov/genetics/gene/mthfr/#function
- Melville, A., & Eastwood, A. (1998). Management of lung cancer. *Quality in health care: QHC*, *7*(3), 170.
- Metro, G., Chiari, R., Mare, M., Giannarelli, D., Tofanetti, F. R., Minotti, V., ... & Crinò, L. (2011). Carboplatin plus pemetrexed for platinum-pretreated, advanced non-small cell lung cancer: a retrospective study with pharmacogenetic evaluation. *Cancer chemotherapy and pharmacology*, *68*(6), 1405-1412.

- Meulendijks, D., De Groot, J. W. B., Los, M., Boers, J. E., Beerepoot, L. V., Polee, M. B., ... & Cats, A. (2016). Bevacizumab combined with docetaxel, oxaliplatin, and capecitabine, followed by maintenance with capecitabine and bevacizumab, as first-line treatment of patients with advanced HER2-negative gastric cancer: A multicenter phase 2 study. *Cancer*, *122*(9), 1434-1443.
- Millar, D. S., Paul, C. L., Molloy, P. L., & Clark, S. J. (2000). A Distinct Sequence (ATAAA) n Separates Methylated and Unmethylated Domains at the 5′-End of theGSTP1 CpG Island* 210. *Journal of Biological Chemistry*, *275*(32), 24893-24899.
- Miller, J. A. (1994). Recent studies on the metabolic activation of chemical carcinogens. *Cancer research*, *54*(7 Supplement), 1879s-1881s.
- Minchin, R. F., Hanna, P. E., Dupret, J. M., Wagner, C. R., Rodrigues-Lima, F., & Butcher, N. J. (2007). Arylamine N-acetyltransferase I. *The international journal of biochemistry & cell biology*, *39*(11), 1999-2005.
- Mishra, G. A., Pimple, S. A., & Shastri, S. S. (2012). An overview of the tobacco problem in India. *Indian journal of medical and paediatric oncology: official journal of Indian Society of Medical & Paediatric Oncology*, *33*(3), 139.
- Mishra, S., Joseph, R. A., Gupta, P. C., Pezzack, B., Ram, F., Sinha, D. N.,& Jha, P. (2016). Trends in bidi and cigarette smoking in India from 1998 to 2015, by age, gender and education. *BMJ Global Health*, *1*(1), e000005.
- Mishra, S., Joseph, R. A., Gupta, P. C., Pezzack, B., Ram, F., Sinha, D. N., ... & Jha, P. (2016). Trends in bidi and cigarette smoking in India from 1998 to 2015, by age, gender and education. *BMJ global health*, *1*(1), e000005.
- Mittal, R. D., & Srivastava, D. L. (2007). Cytochrome P4501A1 and microsomal epoxide hydrolase gene polymorphisms: gene–environment interaction and risk of prostate cancer. *DNA and cell biology*, *26*(11), 791-798.
- Moffat, G. J., McLaren, A. W., & Wolf, C. R. (1996). Sp1-mediated Transcriptional Activation of the Human Pi Class Glutathione S-Transferase Promoter (∗). *Journal of Biological Chemistry*, *271*(2), 1054-1060.
- Moore, J. H., & Williams, S. M. (2009). Epistasis and its implications for personal genetics. *The American Journal of Human Genetics*, *85*(3), 309-320.
- Morceau, F., Duvoix, A., Delhalle, S., Schnekenburger, M., Dicato, M., & Diederich, M. (2004). Regulation of glutathione S-transferase P1-1 gene expression by NF-kappaB in tumor necrosis factor alpha-treated K562 leukemia cells. *Biochemical pharmacology*, *67*(7), 1227-1238.
- Mota, P., Silva, H. C., Soares, M. J., Pego, A., Loureiro, M., Cordeiro, C. R., & Regateiro, F. J. (2015). Genetic polymorphisms of phase I and phase II metabolic enzymes as modulators of lung cancer susceptibility. *Journal of cancer research and clinical oncology*, *141*(5), 851-860.

- Mueller, B. M., Romerdahl, C. A., Gillies, S. D., & Reisfeld, R. A. (1990). Enhancement of antibody-dependent cytotoxicity with a chimeric anti-GD2 antibody. *The Journal of Immunology*, *144*(4), 1382-1386.
- Nagaishi, C., & Nagasawa, N. (1972). Functional Anatomy and Histology of the Lung. University Park Press.
- Nagata, K., & Yamazoe, Y. (2000). Pharmacogenetics of sulfotransferase. *Annual review of pharmacology and toxicology*, *40*(1), 159-176.
- Nakagawa, K., Yokota, J., Wada, M., Sasaki, Y., Fujiwara, Y., Sakai, M., ... & Saijo, N. (1988). Levels of glutathioneS transferase π mRNA in human lung cancer cell lines correlate with the resistance to cisplatin and carboplatin. *Japanese journal of cancer research*, *79*(3), 301-304.
- National Research Council. (1999). The mechanistic basis of radon-induced lung cancer. In *Health Effects of Exposure to Radon: BEIR VI*. National Academies Press (US).
- Navid, F., M Santana, V., & C Barfield, R. (2010). Anti-GD2 antibody therapy for GD2-expressing tumors. *Current cancer drug targets*, *10*(2), 200-209.
- Nebert, D. W., Roe, A. L., Vandale, S. E., Bingham, E., & Oakley, G. G. (2002). NAD (P) H: quinone oxidoreductase (NQO1) polymorphism, exposure to benzene, and predisposition to disease: a HuGE review. *Genetics in Medicine*, *4*(2), 62-70.
- Negishi, M., Pedersen, L. G., Petrotchenko, E., Shevtsov, S., Gorokhov, A., Kakuta, Y., & Pedersen, L. C. (2001). Structure and function of sulfotransferases. *Archives of biochemistry and biophysics*, *390*(2), 149-157.
- Nguyen, H. L., Yang, X., & Omiecinski, C. J. (2013). Expression of a novel mRNA transcript for human microsomal epoxide hydrolase (EPHX1) is regulated by short open reading frames within its 5′-untranslated region. *Rna*, *19*(6), 752-766.
- Nicholson, S. A., Beasley, M. B., Brambilla, E., Hdtasleton, P. S., Colby, T. V., Sheppard, M. N.,& Travis, W. D. (2002). Small cell lung carcinoma (SCLC): a clinicopathologic study of 100
- Nimmagadda, D., Cherala, G., & Ghatta, S. (2006). Cytosolic sulfotransferases. *Indian Journal of Experimental Biology.*
- Nishikawa, A., Mori, Y., Lee, I. S., Tanaka, T., & Hirose, M. (2004). Cigarette smoking, metabolic activation and carcinogenesis. *Current drug metabolism*, *5*(5), 363-373.
- Nooreldeen, R., & Bach, H. (2021). Current and Future Development in Lung Cancer Diagnosis. *International Journal of Molecular Sciences*, *22*(16), 8661.
- Noronha, V., Dikshit, R., Raut, N., Joshi, A., Pramesh, C. S., George, K., ... & Prabhash, K. (2012). Epidemiology of lung cancer in India: Focus on the differences between non-smokers and smokers: A single-centre experience. *Indian journal of cancer*, *49*(1), 74.
- Noronha, V., Pinninti, R., Patil, V. M., Joshi, A., & Prabhash, K. (2016). Lung cancer in the Indian subcontinent. *South Asian journal of cancer*, *5*(03), 095-103.

- Notani, P. N., Rao, D. N., Sirsat, M. V., & Sanghvi, L. D. (1977). A study of lung cancer in relation to bidi smoking in different religious communities in Bombay. *Indian journal of cancer*, *14*(2), 115-121.
- Notani, P., & Sanghvi, L. D. (1974). A retrospective study of lung cancer in Bombay. *British journal of cancer*, *29*(6), 477-482.
- Nowell, S., Coles, B., Sinha, R., MacLeod, S., Ratnasinghe, D. L., Stotts, C., ... & Lang, N. P. (2002). Analysis of total meat intake and exposure to individual heterocyclic amines in a case-control study of colorectal cancer: contribution of metabolic variation to risk. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *506*, 175-185.
- Nowell, S., Ratnasinghe, D. L., Ambrosone, C. B., Williams, S., Teague-Ross, T., Trimble, L., ... & Lang, N. P. (2004). Association of SULT1A1 phenotype and genotype with prostate cancer risk in African-Americans and Caucasians. *Cancer Epidemiology and Prevention Biomarkers*, *13*(2), 270-276.
- Nowell, S., Sweeney, C., Winters, M., Stone, A., Lang, N. P., Hutchins, L. F., ... & Ambrosone, C. B. (2002). Association between sulfotransferase 1A1 genotype and survival of breast cancer patients receiving tamoxifen therapy. *Journal of the National Cancer Institute*, *94*(21), 1635-1640.
- Nuciforo, P., & Fraggetta, F. (2004). Cancer stem cell theory: pathologists' considerations and ruminations about wasting time and wrong evaluations. *Journal of clinical pathology*, *57*(7), 782-782.
- Nunobiki, O., Ueda, M., Akise, H., Izuma, S., Torii, K., Okamoto, Y., ... & Higashida, T. (2015). GSTM1, GSTT1, and NQO1 polymorphisms in cervical carcinogenesis. *Human cell*, *28*(3), 109-113.
- Nyberg, F., Hou, S. M., Hemminki, K., Lambert, B., & Pershagen, G. (1998). Glutathione S-transferase mu1 and N-acetyltransferase 2 genetic polymorphisms and exposure to tobacco smoke in nonsmoking and smoking lung cancer patients and population controls. *Cancer Epidemiology and Prevention Biomarkers*, *7*(10), 875-883.
- Oakley, A. (2011). Glutathione transferases: a structural perspective. *Drug metabolism reviews*, *43*(2), 138-151.
- Oesch, F. (1974). Purification and specificity of a human microsomal epoxide hydratase. *Biochemical journal*, *139*(1), 77-88.
- Oesch, F., Herrero, M. E., Hengstler, J. G., Lohmann, M., & Arand, M. (2000). Metabolic detoxification: implications for thresholds. *Toxicologic pathology*, *28*(3), 382-387.
- Oesch, F., Raphael, D., Schwind, H., & Glatt, H. R. (1977). Species differences in activating and inactivating enzymes related to the control of mutagenic metabolites. *Archives of toxicology*, *39*(1), 97-108.

- Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., ... & Goldman, A. (1992). The α/β hydrolase fold. *Protein Engineering, Design and Selection*, *5*(3), 197-211.
- Omiecinski, C. J., Vanden Heuvel, J. P., Perdew, G. H., & Peters, J. M. (2011). Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. *Toxicological sciences*, *120*(suppl_1), S49-S75.
- Oren, M. (2003). Decision making by p53: life, death and cancer. *Cell Death & Differentiation*, *10*(4), 431-442.
- Osawa, Y., Osawa, K. K., Miyaishi, A., Higuchi, M., Tsutou, A., Matsumura, S., ... & Takahashi, J. (2007). NAT2 and CYP1A2 polymorphisms and lung cancer risk in relation to smoking status. *Asian Pacific journal of cancer prevention*, *8*(1), 103-108.
- Oyama, T., Kawamoto, T., Mizoue, T., Yasumoto, K., Kodama, Y., & Mitsudomi, T. (1997). N-acetylation polymorphism in patients with lung cancer and its association with p53 gene mutation. *Anticancer research*, *17*(1B), 577-581.
- Pachouri, S. S., Sobti, R. C., Kaur, P., Singh, J., & Gupta, S. K. (2006). Impact of polymorphism in sulfotransferase gene on the risk of lung cancer. *Cancer genetics and cytogenetics*, *171*(1), 39-43.
- Panasci, L., Shenouda, G., Begin, L., Pollak, M., Reinke, A., & Margolese, R. (1990). Mitomycin C and mitoxantrone chemotherapy for advanced breast cancer: efficacy with minimal gastrointestinal toxicity and alopecia. *Cancer chemotherapy and pharmacology*, *26*(6), 457-460.
- Pandey, A., Desai, A., Ostwal, V., Patil, V., Kulkarni, A., Kulkarni, R., ... & Banavali, S. D. (2016). Outcome of operable oral cavity cancer and impact of maintenance metronomic chemotherapy: A retrospective study from rural India. *South Asian journal of cancer*, *5*(02), 052-055.
- Pandey, A., Singh, A., Singh, S., Shahi, H., Kumar, A., Kumari, A., & Das, A. (2020). Smoking pattern among rural Indian cancer patients: A prospective survey. *Indian Journal of Medical and Paediatric Oncology*, *41*(02), 209-212.
- Pang, K. S., Schwab, A. J., Goresky, C. A., & Chiba, M. (1994). Transport, binding, and metabolism of sulfate conjugates in the liver. *Chemico-biological interactions*, *92*(1-3), 179-207.
- Panov, S. Z. (2005). Molecular biology of the lung cancer. *Radiology and Oncology*, *39*(3).
- Pao, W., Miller, V. A., Politi, K. A., Riely, G. J., Somwar, R., Zakowski, M. F., ... & Varmus, H. (2005). Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS medicine*, *2*(3), e73.
- Parang, S., & Bhavin, J. (2021). LDCT Screening in Smokers in India—A Pilot, Proof-of-Concept Study. *Indian Journal of Radiology and Imaging*, *31*(02), 318-322.

- Pardoll, D. M. (2012). The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer*, *12*(4), 252-264.
- Park, J. Y., Chen, L., Elahi, A., Lazarus, P., & Tockman, M. S. (2005). Genetic analysis of microsomal epoxide hydrolase gene and its association with lung cancer risk. *European journal of cancer prevention: the official journal of the European Cancer Prevention Organisation (ECP)*, *14*(3), 223.
- Park, M. Y., & Hastie, T. (2008). Penalized logistic regression for detecting gene interactions. *Biostatistics*, *9*(1), 30-50.
- Parle-McDermott, A., Mills, J. L., Molloy, A. M., Carroll, N., Kirke, P. N., Cox, C., ... & Scott, J. M. (2006). The MTHFR 1298CC and 677TT genotypes have opposite associations with red cell folate levels. *Molecular genetics and metabolism*, *88*(3), 290-294.
- Parrish, A. R., Chen, G., Burghardt, R. C., Watanabe, T., Morisseau, C., & Hammock, B. D. (2009). Attenuation of cisplatin nephrotoxicity by inhibition of soluble epoxide hydrolase. *Cell biology and toxicology*, *25*(3), 217-225.
- Pavan, S., Delvento, C., Ricciardi, L., Lotti, C., Ciani, E., & D'Agostino, N. (2020). Recommendations for choosing the genotyping method and best practices for quality control in crop genome-wide association studies. *Frontiers in Genetics*, *11*, 447.
- Pedersen, L. C., Petrotchenko, E., Shevtsov, S., & Negishi, M. (2002). Crystal structure of the human estrogen sulfotransferase-PAPS complex: evidence for catalytic role of Ser137 in the sulfuryl transfer reaction. *Journal of Biological Chemistry*, *277*(20), 17928-17932.
- Pejchal, R., Campbell, E., Guenther, B. D., Lennon, B. W., Matthews, R. G., & Ludwig, M. L. (2006). Structural perturbations in the Ala→ Val polymorphism of methylenetetrahydrofolate reductase: how binding of folates may protect against inactivation. *Biochemistry*, *45*(15), 4808-4818.
- Pejchal, R., Sargeant, R., & Ludwig, M. L. (2005). Structures of NADH and CH3-H4folate complexes of Escherichia coli methylenetetrahydrofolate reductase reveal a spartan strategy for a ping-pong reaction. *Biochemistry*, *44*(34), 11447-11457.
- Peng, C. T., Chen, J. C., Yeh, K. T., Wang, Y. F., Hou, M. F., Lee, T. P., ... & Chang, J. G. (2003). The relationship among the polymorphisms of SULT1A1, 1A2 and different types of cancers in Taiwanese. *International journal of molecular medicine*, *11*(1), 85-89.https://doi.org/10.3892/ijmm.11.1.85
- Peng, H., Zhu, Q. S., Zhong, S., & Levy, D. (2013). Transcription of the human microsomal epoxide hydrolase gene (EPHX1) is regulated by an HNF-4α/CAR/RXR/PSF complex. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1829*(10), 1000-1009.
- Perera, F. P., Tang, D., Brandt-Rauf, P., Santella, R. M., Mooney, L. A., Tu, Y. H., ... & Bell, D. A. (2006). Lack of Associations Between Cancer and Albumin Adducts, ras p21 Oncoprotein Levels, and CYP1A1, CYP2D6, NAT1, NAT2 in a Nested Case-Control

Study of Lung Cancer Within the Physicians Health Study. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *15*(7), 1417.

- Pérez-Ramírez, C., Cañadas-Garre, M., Alnatsha, A., Villar, E., Delgado, J. R., Faus-Dáder, M. J., & ÿngel Calleja-Hernández, M. (2016). Pharmacogenetic predictors of toxicity to platinum based chemotherapy in non-small cell lung cancer patients. *Pharmacological research*, *111*, 877-884.
- Persson, I., Johansson, I., Yue, Q. Y., Duan, L. S., Bertilsson, L., & Ingelman-Sundberg, M. (1999). Genetic polymorphism of xenobiotic metabolizing enzymes among Chinese lung cancer patients. *International journal of cancer*, *81*(3), 325-329.
- Petersen, D. D., Gonzalez, F. J., Rapic, V., Kozak, C. A., Lee, J. Y., Jones, J. E., & Nebert, D. W. (1989). Marked increases in hepatic NAD (P) H: oxidoreductase gene transcription and mRNA levels correlated with a mouse chromosome 7 deletion. *Proceedings of the National Academy of Sciences*, *86*(17), 6699-6703.
- Pey, A. L., Megarity, C. F., & Timson, D. J. (2019). NAD (P) H quinone oxidoreductase (NQO1): an enzyme which needs just enough mobility, in just the right places. *Bioscience reports*, *39*(1).
- Pfeifer, G. P., Denissenko, M. F., Olivier, M., Tretyakova, N., Hecht, S. S., & Hainaut, P. (2002). Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene*, *21*(48), 7435-7451.
- Phan, G. Q., & Rosenberg, S. A. (2013). Adoptive cell transfer for patients with metastatic melanoma: the potential and promise of cancer immunotherapy. *Cancer Control*, *20*(4), 289-297.
- Pillot, G. A., Read, W. L., Hennenfent, K. L., Marsh, S., Gao, F., Viswanathan, A., ... & Govindan, R. (2006). A phase II study of irinotecan and carboplatin in advanced non-small cell lung cancer with pharmacogenomic analysis. *Journal of Thoracic Oncology*, *1*(9), 972-978.
- Pincinato, E. C., Costa, E. F., Lopes-Aguiar, L., Nogueira, G. A., Lima, T. R., Visacri, M. B., ... & Lima, C. S. (2019). GSTM1, GSTT1 and GSTP1 Ile105Val polymorphisms in outcomes of head and neck squamous cell carcinoma patients treated with cisplatin chemoradiation. *Scientific reports*, *9*(1), 1-11.
- Polat, N. D., Degirmencioglu, S., Yaren, A., Demiray, A., Akca, H., & Dogu, G. G. (2018). GSTP1, TSER, MTHFR C677T and MTHFR A1298C gene single nucleotide polymorphisms associated with toxicity and survival in patients with colorectal cancer treated with 5-fluorouracil-based chemotherapy. *Cancer Rep. Rev*, *2*(2), 1-7.
- Prade, L., Huber, R., Manoharan, T. H., Fahl, W. E., & Reuter, W. (1997). Structures of class pi glutathione S-transferase from human placenta in complex with substrate, transition-state analogue and inhibitor. *Structure*, *5*(10), 1287-1295.
- Prasad, R., Singhal, S., & Garg, R. (2009). Bidi smoking and lung cancer. *Bioscience trends*, *3*(2).

- Prasad, R., Tandon, S., Kumar, S., Pant, M. C., Sinha, K. N., & Mukerji, P. K. (1998). A case control study on tobacco smoking and lung cancer. *Lung India*, *16*(2), 60-64.
- Quint, L. E., Tummala, S., Brisson, L. J., Francis, I. R., Krupnick, A. S., Kazerooni, E. A.,& Orringer, M. B. (1996). Distribution of distant metastases from newly diagnosed non-small cell lung cancer. *The Annals of Thoracic Surgery*, *62*(1), 246-250.
- Ragunathan, N., Dairou, J., Pluvinage, B., Martins, M., Petit, E., Janel, N., ... & Rodrigues-Lima, F. (2008). Identification of the xenobiotic-metabolizing enzyme arylamine N-acetyltransferase 1 as a new target of cisplatin in breast cancer cells: molecular and cellular mechanisms of inhibition. *Molecular pharmacology*, *73*(6), 1761-1768.
- Rahman, M., & Fukui, T. (2000). Bidi smoking and health. *Public health*, *114*(2), 123-127.
- Rahman, M., Sakamoto, J., & Fukui, T. (2003). Bidi smoking and oral cancer: a meta-analysis. *International journal of cancer*, *106*(4), 600-604.
- Rami-Porta, R., Crowley, J. J., & Goldstraw, P. (2009). Review the revised TNM staging system for lung cancer. *Ann Thorac Cardiovasc Surg*, *15*(1), 5.
- Reck, M., Bondarenko, I., Luft, A., Serwatowski, P., Barlesi, F., Chacko, R., ... & Lynch, T. J. (2013). Ipilimumab in combination with paclitaxel and carboplatin as first-line therapy in extensive-disease-small-cell lung cancer: results from a randomized, double-blind, multicenter phase 2 trial. *Annals of Oncology*, *24*(1), 75-83.
- Reddy, K. S., & Gupta, P. C. (2004). Tobacco control in India. *New delhi: ministry of health and family welfare, Government of India*, 43-47.
- Reed, J. C. (1997). Double identity for proteins of the Bcl-2 family. *Nature*, *387*(6635), 773-776.
- Riely, G. J., Kris, M. G., Rosenbaum, D., Marks, J., Li, A., Chitale, D. A., ... & Ladanyi, M. (2008). Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma. *Clinical cancer research*, *14*(18), 5731-5734.
- Ritchie, M. D. and Motsinger, A. A. (2005), 'Multifactor dimensionality reduction for detecting gene–gene and gene–environment interactions in pharmacogenomics studies.
- Rivlin, N., Brosh, R., Oren, M., & Rotter, V. (2011). Mutations in the p53 tumor suppressor gene: important milestones at the various steps of tumorigenesis. *Genes & cancer*, *2*(4), 466-474.
- Robertson, J. A., Chen, H. C., & Nebert, D. W. (1986). NAD (P) H: menadione oxidoreductase. Novel purification of enzyme cDNA and complete amino acid sequence, and gene regulation. *Journal of Biological Chemistry*, *261*(33), 15794-15799.
- Rom, W. N., Hay, J. G., Lee, T. C., Jiang, Y., & Tchou-Wong, K. M. (2000). Molecular and genetic aspects of lung cancer. *American journal of respiratory and critical care medicine*, *161*(4), 1355-1367.

- Rosenberger, A., Illig, T., Korb, K., Klopp, N., Zietemann, V., Wölke, G., ... & Wichmann, H. E. (2008). Do genetic factors protect for early onset lung cancer? A case control study before the age of 50 years. *BMC cancer*, *8*(1), 1-12.
- Ross, D., & Siegel, D. (2004). NAD (P) H: quinone oxidoreductase 1 (NQO1, DT-diaphorase), functions and pharmacogenetics. *Methods in enzymology*, *382*, 115-144.
- Ross, D., Kepa, J. K., Winski, S. L., Beall, H. D., Anwar, A., & Siegel, D. (2000). NAD (P) H: quinone oxidoreductase 1 (NQO1): chemoprotection, bioactivation, gene regulation and genetic polymorphisms. *Chemico-biological interactions*, *129*(1-2), 77-97.
- Rota, M., Bosetti, C., Boccia, S., Boffetta, P., & La Vecchia, C. (2014). Occupational exposures to polycyclic aromatic hydrocarbons and respiratory and urinary tract cancers: an updated systematic review and a meta-analysis to 2014. *Archives of toxicology*, *88*(8), 1479-1490.
- Rouprêt, M., Cancel-Tassin, G., Comperat, E., Fromont, G., Sibony, M., Molinié, V., ... & Cussenot, O. (2007). Phenol sulfotransferase SULT1A1* 2 allele and enhanced risk of upper urinary tract urothelial cell carcinoma. *Cancer Epidemiology and Prevention Biomarkers*, *16*(11), 2500-2503.
- Rubin, B. P., Skarin, A. T., Pisick, E., Rizk, M., & Salgia, R. (2001). Use of cytokeratins 7 and 20 in determining the origin of metastatic carcinoma of unknown primary, with special emphasis on lung cancer. *European Journal of Cancer Prevention*, *10*(1), 77-82.
- Ruscoe, J. E., Rosario, L. A., Wang, T., Gaté, L., Arifoglu, P., Wolf, C. R., ... & Tew, K. D. (2001). Pharmacologic or genetic manipulation of glutathione S-transferase P1–1 (GSTπ) influences cell proliferation pathways. *Journal of Pharmacology and Experimental Therapeutics*, *298*(1), 339-345.
- Rusinova, G. G., Azizova, T. V., Viazovskaia, N. S., Glazkova, I. V., & Osovets, S. V. (2014). Role of gene polymorphisms of phase II of xenobiotic biotransformation from glutathione-S-transferase and N-acetyltransferase families in susceptibility to lung cancer among Mayak workers. *Radiatsionnaia biologiia, radioecologiia*, *54*(4), 350-359.
- Ruzza, P., Rosato, A., Rossi, C. R., Floreani, M., & Quintieri, L. (2009). Glutathione transferases as targets for cancer therapy. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, *9*(7), 763-777.
- Ruzzo, A., Graziano, F., Kawakami, K., Watanabe, G., Santini, D., Catalano, V., ... & Magnani, M. (2006). Pharmacogenetic profiling and clinical outcome of patients with advanced gastric cancer treated with palliative chemotherapy. *J Clin Oncol*, *24*(12), 1883-1891.
- Saarikoski, S. T., Reinikainen, M., Anttila, S., Karjalainen, A., Vainio, H., Husgafvel-Pursiainen, K., & Hirvonen, A. (2000). Role of NAT2 deficiency in susceptibility to lung cancer among asbestos-exposed individuals. *Pharmacogenetics and Genomics*, *10*(2), 183-185.

- Sabari, J. K., Lok, B. H., Laird, J. H., Poirier, J. T., & Rudin, C. M. (2017). Unravelling the biology of SCLC: implications for therapy. *Nature reviews Clinical oncology*, *14*(9), 549-561.
- Sak, K., Kasemaa, K., & Everaus, H. (2019). Polymorphisms in COMT and SULT1A1 Genes and Chronic Lymphocytic Leukemia Risk in Estonia.
- Saldivar, S. J., Wang, Y., Zhao, H., Shao, L., Lin, J., Spitz, M. R., & Wu, X. (2005). An association between a NQO1 genetic polymorphism and risk of lung cancer. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, *582*(1-2), 71-78.
- Sameer, A. S., Shah, Z. A., Syeed, N., Rasool, R., Afroze, D., & Siddiqi, M. A. (2010). NAD (P) H: quinone oxidoreductase 1 (NQO1) Pro187Ser polymorphism and colorectal cancer predisposition in the ethnic Kashmiri population. *Asian Pac J Cancer Prev*, *11*(1), 209-213.
- Samet, J. M. (1989). Radon and lung cancer. *JNCI: Journal of the National Cancer Institute*, *81*(10), 745-758.
- Samet, J. M., Avila-Tang, E., Boffetta, P., Hannan, L. M., Olivo-Marston, S., Thun, M. J., & Rudin, C. M. (2009). Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clinical Cancer Research*, *15*(18), 5626-5645.
- Samet, J. M., Avila-Tang, E., Boffetta, P., Hannan, L. M., Olivo-Marston, S., Thun, M. J., & Rudin, C. M. (2009). Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clinical Cancer Research*, *15*(18), 5626-5645.
- Santos, P. A. R. D., Rocha, R. S. D., Pipkin, M., Silveira, M. L. D., Cypel, M., Rios, J. O., & Pinto, J. A. L. D. F. (2007). Concordance between clinical and pathological staging in patients with stages I or II non-small cell lung cancer subjected to surgical treatment. *Jornal Brasileiro de Pneumologia*, *33*, 647-654.
- Sasco, A. J., Secretan, M. B., & Straif, K. (2004). Tobacco smoking and cancer: a brief review of recent epidemiological evidence. *Lung cancer*, *45*, S3-S9.
- Saunders, L. R., Bankovich, A. J., Anderson, W. C., Aujay, M. A., Bheddah, S., Black, K., ... & Dylla, S. J. (2015). A DLL3-targeted antibody-drug conjugate eradicates high-grade pulmonary neuroendocrine tumor-initiating cells in vivo. *Science translational medicine*, *7*(302), 302ra136-302ra136.
- Scagliotti, G. V., Parikh, P., Von Pawel, J., Biesma, B., Vansteenkiste, J., Manegold, C., ... & Gandara, D. (2008). Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *Journal of clinical oncology*, *26*(21), 3543-3551.
- Scagliotti, G., Hanna, N., Fossella, F., Sugarman, K., Blatter, J., Peterson, P., ... & Shepherd, F. A. (2009). The differential efficacy of pemetrexed according to NSCLC histology: a review of two Phase III studies. *The oncologist*, *14*(3), 253-263.
- Schabath, M. B., Spitz, M. R., Lerner, S. P., Pillow, P. C., Hernandez, L. M., Delclos, G. L., ... & Wu, X. (2005). Case-control analysis of dietary folate and risk of bladder cancer. *Nutrition and cancer*, *53*(2), 144-151.

- Schnekenburger, M., Karius, T., & Diederich, M. (2014). Regulation of epigenetic traits of the glutathione S-transferase P1 gene: from detoxification toward cancer prevention and diagnosis. *Frontiers in pharmacology*, *5*, 170.
- Schnekenburger, M., Morceau, F., Duvoix, A., Delhalle, S., Trentesaux, C., Dicato, M., & Diederich, M. (2003). Expression of glutathione S-transferase P1-1 in differentiating K562: role of GATA-1. *Biochemical and biophysical research communications*, *311*(4), 815-821.
- SEER Stat Fact Sheets: Lung and Bronchus Cancer. National Cancer Institute website. http://seer.cancer.gov/statfacts/html/lungb.html. Updated April 2021.
- Segura-Aguilar, J., & Lind, C. (1989). On the mechanism of the Mn3+-induced neurotoxicity of dopamine: Prevention of quinone-derived oxygen toxicity by DT diaphorase and superoxide dismutase. *Chemico-biological interactions*, *72*(3), 309-324.
- Senggunprai, L., Yoshinari, K., & Yamazoe, Y. (2009). Selective role of sulfotransferase 2A1 (SULT2A1) in the N-sulfoconjugation of quinolone drugs in humans. *Drug metabolism and disposition*, *37*(8), 1711-1717.
- Seow, A., Zhao, B., Lee, E. J., Poh, W. T., Teh, M., Eng, P., ... & Lee, H. P. (2001). Cytochrome P4501A2 (CYP1A2) activity and lung cancer risk: a preliminary study among Chinese women in Singapore. *Carcinogenesis*, *22*(4), 673-677.
- Seow, A., Zhao, B., Poh, W. T., Teh, M., Eng, P., Wang, Y. T., ... & Lee, H. P. (1999). NAT2 slow acetylator genotype is associated with increased risk of lung cancer among non-smoking Chinese women in Singapore. *Carcinogenesis*, *20*(9), 1877-1881.
- Seow, A., Zhao, B., Poh, W. T., Teh, M., Eng, P., Wang, Y. T., ... & Lee, H. P. (1999). NAT2 slow acetylator genotype is associated with increased risk of lung cancer among non-smoking Chinese women in Singapore. *Carcinogenesis*, *20*(9), 1877-1881.
- Shan, X., Wang, L., Hoffmaster, R., & Kruger, W. D. (1999). Functional characterization of human methylenetetrahydrofolate reductase in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *274*(46), 32613-32618.
- Shankar, A., Saini, D., Dubey, A., Roy, S., Bharati, S. J., Singh, N., ... & Rath, G. K. (2019). Feasibility of lung cancer screening in developing countries: challenges, opportunities and way forward. *Translational lung cancer research*, *8*(Suppl 1), S106.
- Shankar, A., Saini, D., Dubey, A., Roy, S., Bharati, S. J., Singh, N., ... & Rath, G. K. (2019). Feasibility of lung cancer screening in developing countries: challenges, opportunities and way forward. *Translational lung cancer research*, *8*(Suppl 1), S106.
- Sheehan, D., Meade, G., FOLEY, V. M., & DOWD, C. A. (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical journal*, *360*(1), 1-16.
- Shen, H., Wei, Q., Pillow, P. C., Amos, C. I., Hong, W. K., & Spitz, M. R. (2003). Dietary folate intake and lung cancer risk in former smokers: a case-control analysis. *Cancer Epidemiology and Prevention Biomarkers*, *12*(10), 980-986.

- Sheweita, S. A., & Tilmisany, A. K. (2003). Cancer and phase II drug-metabolizing enzymes. *Current drug metabolism*, *4*(1), 45-58.
- Shi, Q., Zhang, Z., Li, G., Pillow, P. C., Hernandez, L. M., Spitz, M. R., & Wei, Q. (2005). Sex differences in risk of lung cancer associated with methylene-tetrahydrofolate reductase polymorphisms. *Cancer Epidemiology and Prevention Biomarkers*, *14*(6), 1477-1484.
- Shimada, T. (2006). Xenobiotic-metabolizing enzymes involved in activation and detoxification of carcinogenic polycyclic aromatic hydrocarbons. *Drug metabolism and pharmacokinetics*, *21*(4), 257-276.
- *Sholl, L. M. (2016). The molecular pathology of lung cancer. Surgical pathology clinics, 9(3), 353-378.*
- Sholl, L. M., Do, K., Shivdasani, P., Cerami, E., Dubuc, A. M., Kuo, F. C., ... & MacConaill, L. E. (2016). Institutional implementation of clinical tumor profiling on an unselected cancer population. *JCI insight*, *1*(19).
- Shrubsole, M. J., Gao, Y. T., Cai, Q., Shu, X. O., Dai, Q., Hébert, J. R., ... & Zheng, W. (2004). MTHFR polymorphisms, dietary folate intake, and breast cancer risk: results from the Shanghai Breast Cancer Study. *Cancer Epidemiology and Prevention Biomarkers*, *13*(2), 190-196.
- Siegel, D., Anwar, A., Winski, S. L., Kepa, J. K., Zolman, K. L., & Ross, D. (2001). Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD (P) H: quinone oxidoreductase 1. *Molecular pharmacology*, *59*(2), 263-268.
- Siemianowicz, K., Gminski, J., Garczorz, W., Slabiak, N., Goss, M., Machalski, M., & Magiera-Molendowska, H. (2003). Methylenetetrahydrofolate reductase gene C677T and A1298C polymorphisms in patients with small cell and non-small cell lung cancer. *Oncology reports*, *10*(5), 1341-1344.
- Silverman, J. F., & Geisinger, K. R. (1996). *Fine Needle Aspiration Cytology of the Thorax And Abdomen*. WB Saunders Company.
- Sim, E., Abuhammad, A., & Ryan, A. (2014). Arylamine N-acetyltransferases: from drug metabolism and pharmacogenetics to drug discovery. *British journal of pharmacology*, *171*(11), 2705-2725.
- Sim, E., Fullam, E., & Wakefield, L. (2010). Arylamine N-acetyltransferases. *Compr Toxicology*, *4*, 385–412.
- Singh, S. (2015). Cytoprotective and regulatory functions of glutathione S-transferases in cancer cell proliferation and cell death. *Cancer chemotherapy and pharmacology*, *75*(1), 1-15.
- Sinha, D. N., Gupta, P. C., & Pednekar, M. S. (2003). Tobacco use in a rural area of Bihar, India. *Indian Journal of Community Medicine*, *28*(4), 167-170.

- Skjelbred, C. F., Svendsen, M., Haugan, V., Eek, A. K., Clausen, K. O., Kure, E. H., ... & Hansteen, I. L. (2011). Influence of GSTM1, GSTT1, GSTP1, NAT1, NAT2, EPHX1, MTR and MTHFR polymorphism on chromosomal aberration frequencies in human lymphocytes. *Carcinogenesis*, *32*(3), 399-405.
- Skuladottir, H., Autrup, H., Autrup, J., Tjoenneland, A., Overvad, K., Ryberg, D., ... & Olsen, J. H. (2005). Polymorphisms in genes involved in xenobiotic metabolism and lung cancer risk under the age of 60 years: A pooled study of lung cancer patients in Denmark and Norway. *Lung Cancer*, *48*(2), 187-199.
- Smit, E. F., Burgers, S. A., Biesma, B., Smit, H. J., Eppinga, P., Dingemans, A. M. C., ... & Groen, H. J. (2009). Randomized phase II and pharmacogenetic study of pemetrexed compared with pemetrexed plus carboplatin in pretreated patients with advanced non–small-cell lung cancer. *Journal of Clinical Oncology*, *27*(12), 2038-2045.
- Smith, C. A., Wadelius, M., Gough, A. C., Harrison, D. J., Wolf, C. R., & Rane, A. (1997). A simplified assay for the arylamine N-acetyltransferase 2 polymorphism validated by phenotyping with isoniazid. *Journal of medical genetics*, *34*(9), 758-760.
- Sobti, R. C., Kaur, P., Kaur, S., Janmeja, A. K., Jindal, S. K., Kishan, J., & Raimondi, S. (2009). Impact of interaction of polymorphic forms of p53 codon 72 and N-acetylation gene (NAT2) on the risk of lung cancer in the North Indian population. *DNA and cell biology*, *28*(9), 443-449.
- Song, S. Y., Jeong, S. Y., Park, H. J., Park, S. I., Kim, D. K., Kim, Y. H., ... & Choi, E. K. (2010). Clinical significance of NQO1 C609T polymorphisms after postoperative radiation therapy in completely resected non-small cell lung cancer. *Lung Cancer*, *68*(2), 278-282.
- Song, Y., Qi, X., & Liu, X. (2020). N-acetyltransferase 2 polymorphism is associated with bladder cancer risk: An updated meta-analysis based on 54 case-control studies. *Gene*, *757*, 144924.
- Sørensen, M., Autrup, H., Tjønneland, A., Overvad, K., & Raaschou-Nielsen, O. (2005). Genetic polymorphisms in CYP1B1, GSTA1, NQO1 and NAT2 and the risk of lung cancer. *Cancer letters*, *221*(2), 185-190.
- Soria, J. C., Jang, S. J., Khuri, F. R., Hassan, K., Liu, D., Hong, W. K., & Mao, L. (2000). Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. *Cancer Research*, *60*(15), 4000-4004.
- Spiro, S. G., & Porter, J. C. (2002). Lung cancer—where are we today? Current advances in staging and nonsurgical treatment. *American journal of respiratory and critical care medicine*, *166*(9), 1166-1196.
- Steenland, K., Mannetje, A., Boffetta, P., Stayner, L., Attfield, M., Chen, J., ... & Checkoway, H. (2001). Pooled exposure–response analyses and risk assessment for lung cancer in 10 cohorts of silica-exposed workers: an IARC multicentre study. *Cancer Causes & Control*, *12*(9), 773-784.

- Stoehlmacher, J., Park, D. J., Zhang, W., Yang, D., Groshen, S., Zahedy, S., & Lenz, H. J. (2004). A multivariate analysis of genomic polymorphisms: prediction of clinical outcome to 5-FU/oxaliplatin combination chemotherapy in refractory colorectal cancer. *British journal of cancer*, *91*(2), 344-354.
- Strassburg, C. P., Strassburg, A., Kneip, S., Barut, A., Tukey, R. H., Rodeck, B., & Manns, M. P. (2002). Developmental aspects of human hepatic drug glucuronidation in young children and adults. *Gut*, *50*(2), 259-265.
- Strott, C. A. (2002). Sulfonation and molecular action. *Endocrine reviews*, *23*(5), 703-732.
- Su, Z., Dias-Santagata, D., Duke, M., Hutchinson, K., Lin, Y. L., Borger, D. R., ... & Pao, W. (2011). A platform for rapid detection of multiple oncogenic mutations with relevance to targeted therapy in non–small-cell lung cancer. *The Journal of Molecular Diagnostics*, *13*(1), 74-84.
- Subramanian, J., & Govindan, R. (2007). Lung cancer in never smokers: a review. *Journal of clinical oncology*, *25*(5), 561-570.
- Sugamori, K. S., Brenneman, D., Wong, S., Gaedigk, A., Yu, V., Abramovici, H., ... & Grant, D. M. (2007). Effect of arylamine acetyltransferase Nat3 gene knockout on N-acetylation in the mouse. *Drug metabolism and disposition*, *35*(7), 1064-1070.
- Sumner, J. S., & Matthews, R. G. (1992). Stereochemistry and mechanism of hydrogen transfer between NADPH and methylenetetrahydrofolate in the reaction catalyzed by methylenetetrahydrofolate reductase from pig liver. *Journal of the American Chemical Society*, *114*(18), 6949-6956.
- Sumner, J., Jencks, D. A., Khani, S., & Matthews, R. G. (1986). Photoaffinity labeling of methylenetetrahydrofolate reductase with 8-azido-S-adenosylmethionine. *Journal of Biological Chemistry*, *261*(17), 7697-7700.
- Sunaga, N., Kohno, T., Yanagitani, N., Sugimura, H., Kunitoh, H., Tamura, T., ... & Yokota, J. (2002). Contribution of the NQO1 and GSTT1 polymorphisms to lung adenocarcinoma susceptibility. *Cancer Epidemiology and Prevention Biomarkers*, *11*(8), 730-738.
- Sunami, K., Ichikawa, H., Kubo, T., Kato, M., Fujiwara, Y., Shimomura, A., ... & Kohno, T. (2019). Feasibility and utility of a panel testing for 114 cancer-associated genes in a clinical setting: a hospital-based study. *Cancer science*, *110*(4), 1480-1490.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *71*(3), 209-249.
- Takehara, A., Kawakami, K., Ohta, N., Oyama, K., Ota, Y., Oda, M., & Watanabe, G. (2005). Prognostic significance of the polymorphisms in thymidylate synthase and methylenetetrahydrofolate reductase gene in lung cancer. *Anticancer research*, *25*(6C), 4455-4461.

- Takei, H., Asamura, H., Maeshima, A., Suzuki, K., Kondo, H., Niki, T., ... & Matsuno, Y. (2002). Large cell neuroendocrine carcinoma of the lung: a clinicopathologic study of eighty-seven cases. *The Journal of thoracic and cardiovascular surgery*, *124*(2), 285-292.
- Talalay, P., & Dinkova-Kostova, A. T. (2004). Role of nicotinamide quinone oxidoreductase 1 (NQO1) in protection against toxicity of electrophiles and reactive oxygen intermediates. *Methods in enzymology*, *382*, 355-364.
- Tam, V. C., Rask, S., Koru–Sengul, T., & Dhesy–Thind, S. (2009). Generalizability of toxicity data from oncology clinical trials to clinical practice: toxicity of irinotecan-based regimens in patients with metastatic colorectal cancer. *Current Oncology*, *16*(6), 13-20.
- Tamaki, Y., Arai, T., Sugimura, H., Sasaki, T., Honda, M., Muroi, Y., ... & Hiratsuka, M. (2011). Association between cancer risk and drug metabolizing enzyme gene (CYP2A6, CYP2A13, CYP4B1, SULT1A1, GSTM1, and GSTT1) polymorphisms in Japanese cases of lung cancer. *Drug metabolism and pharmacokinetics*, 1107190203-1107190203.
- Tamiya, M., Tamiya, A., Suzuki, H., Moriizumi, K., Nakahama, K., Taniguchi, Y., ... & Kumagai, T. (2019). Which is better EGFR-TKI followed by osimertinib: afatinib or gefitinib/erlotinib?. *Anticancer research*, *39*(7), 3923-3929.
- Tamura, T., Kurishima, K., Nakazawa, K., Kagohashi, K., Ishikawa, H., Satoh, H., & Hizawa, N. (2015). Specific organ metastases and survival in metastatic non-small-cell lung cancer. *Molecular and clinical oncology*, *3*(1), 217-221.
- Tan, X., Wang, Y. Y., Chen, X. Y., Xian, L., Guo, J. J., Liang, G. B., & Chen, M. W. (2014). Quantitative assessment of the effects of the EPHX1 Tyr113His polymorphism on lung and breast cancer. *Genet Mol Res*, *13*, 7437-7446.
- Tang, C., Yu, S., Jiang, H., Li, W., Xu, X., Cheng, X., ... & Liu, T. (2018). A meta-analysis: methylenetetrahydrofolate reductase C677T polymorphism in gastric cancer patients treated with 5-Fu based chemotherapy predicts serious hematologic toxicity but not prognosis. *Journal of Cancer*, *9*(6), 1057-1066.
- Tasnim, T., Al-Mamun, M. M. A., Nahid, N. A., Islam, M. R., Apu, M. N. H., Bushra, M. U., ... & Hasnat, A. (2017). Genetic variants of SULT1A1 and XRCC1 genes and risk of lung cancer in Bangladeshi population. *Tumor Biology*, *39*(11), 1010428317729270.
- Taxak, N., & Bharatam, P. V. (2014). Drug metabolism. *Resonance*, *19*(3), 259-282.
- Tian, F., Zhang, Y., Ren, Y., Shen, L., Wu, W., & Zhou, B. (2014). N-Acetyltransferase 2 (NAT2) gene polymorphism and exposure to smoking in lung cancer of Chinese males. *Medical Oncology*, *31*(8), 90.
- Tian, F., Zhang, Y., Ren, Y., Shen, L., Wu, W., & Zhou, B. (2014). N-Acetyltransferase 2 (NAT2) gene polymorphism and exposure to smoking in lung cancer of Chinese males. *Medical Oncology*, *31*(8), 90.

- Tian, G., Wang, M., & Xu, X. (2014). The role of NQO1 polymorphisms in the susceptibility and chemotherapy response of Chinese NSCLC patients. *Cell biochemistry and biophysics*, *69*(3), 475-479.
- Tilak, A. R., Kumar, S., Jain, M., Pant, M. C., Das, B. C., Guleria, R., ... & Kumar, A. (2011). Association of functionally important polymorphism of microsomal epoxide hydrolase gene (EPHX1) with lung cancer susceptibility. *Cancer investigation*, *29*(6), 411-418.
- Timofeeva, M., Kropp, S., Sauter, W., Beckmann, L., Rosenberger, A., Illig, T., ... & Wichmann, H. E. (2010). Genetic polymorphisms of MPO, GSTT1, GSTM1, GSTP1, EPHX1 and NQO1 as risk factors of early-onset lung cancer. *International journal of cancer*, *127*(7), 1547-1561.
- Tiseo, M., Bordi, P., Bortesi, B., Boni, L., Boni, C., Baldini, E., ... & Ardizzoni, A. (2013). ERCC1/BRCA1 expression and gene polymorphisms as prognostic and predictive factors in advanced NSCLC treated with or without cisplatin. *British journal of cancer*, *108*(8), 1695-1703.
- Toffoli, G., Russo, A., Innocenti, F., Corona, G., Tumolo, S., Sartor, F., ... & Boiocchi, M. (2003). Effect of methylenetetrahydrofolate reductase 677C→ T polymorphism on toxicity and homocysteine plasma level after chronic methotrexate treatment of ovarian cancer patients. *International journal of cancer*, *103*(3), 294-299.
- To-Figueras, J., Gené, M., Gómez-Catalán, J., Piqué, E., Borrego, N., & Corbella, J. (2001). Lung cancer susceptibility in relation to combined polymorphisms of microsomal epoxide hydrolase and glutathione S-transferase P1. *Cancer letters*, *173*(2), 155-162.
- Tomasini, P., Barlesi, F., Mascaux, C., & Greillier, L. (2016). Pemetrexed for advanced stage nonsquamous non-small cell lung cancer: latest evidence about its extended use and outcomes. *Therapeutic advances in medical oncology*, *8*(3), 198-208.
- Tong, Y. H., Zhang, B., Yan, Y. Y., Fan, Y., Yu, J. W., Kong, S. S., ... & Lin, N. M. (2017). Dual-negative expression of Nrf2 and NQO1 predicts superior outcomes in patients with non-small cell lung cancer. *Oncotarget*, *8*(28), 45750.
- Travis, W. D., & Harris, C. (2004). Pathology and genetics of tumours of the lung, pleura, thymus and heart, Feance: IARC Press, 2004
- Trimmer, E. E., Ballou, D. P., Ludwig, M. L., & Matthews, R. G. (2001). Folate activation and catalysis in methylenetetrahydrofolate reductase from Escherichia coli: roles for aspartate 120 and glutamate 28. *Biochemistry*, *40*(21), 6216-6226.
- Tuberculosis country profile. WHO. Available online: https://extranet.who.int/sree/ Reports?op=Replet&name=%2FWHO_HQ_Reports%2F G2%2FPROD%2FEXT%2FTBCountryProfile&ISO2=I N&LAN=EN&outtype=html
- Type of Tumors (2016).

- Underwood C, Glass C. Small cell carcinoma. PathologyOutlines.com website. https://www.pathologyoutlines.com/topic/lungtumorsmallcell.html. Accessed August 10th, 2021.
- Urano, W., Taniguchi, A., Yamanaka, H., Tanaka, E., Nakajima, H., Matsuda, Y., ... & Kamatani, N. (2002). Polymorphisms in the methylenetetrahydrofolate reductase gene were associated with both the efficacy and the toxicity of methotrexate used for the treatment of rheumatoid arthritis, as evidenced by single locus and haplotype analyses. *Pharmacogenetics and Genomics*, *12*(3), 183-190.
- URL: http://pathology.jhu.edu/pc/BasicTypes1.php?area=ba
- URL: https://www.cancer.org/cancer/lungcancer.html
- Václavíková, R., Hughes, D. J., & Souček, P. (2015). Microsomal epoxide hydrolase 1 (EPHX1): Gene, structure, function, and role in human disease. *Gene*, *571*(1), 1-8.
- Van Bladeren, P. J. (2000). Glutathione conjugation as a bioactivation reaction. *Chemico-biological interactions*, *129*(1-2), 61-76.
- Van Loo, B., Kingma, J., Arand, M., Wubbolts, M. G., & Janssen, D. B. (2006). Diversity and biocatalytic potential of epoxide hydrolases identified by genome analysis. *Applied and environmental microbiology*, *72*(4), 2905-2917.
- Vanoni, M. A., & Matthews, R. G. (1984). Kinetic isotope effects on the oxidation of reduced nicotinamide adenine dinucleotide phosphate by the flavoprotein methylenetetrahydrofolate reductase. *Biochemistry*, *23*(22), 5272-5279.
- Vasiliou, V., Theurer, M. J., Puga, A., Reuter, S. F., & Nebert, D. W. (1994). Mouse dioxin-inducible NAD (P) H: menadione oxidoreductase: NMO1 cDNA sequence and genetic differences in mRNA levels. *Pharmacogenetics*, *4*(6), 341-348.
- Veronese, M. E., Burgess, W., Zhu, X. I. A. O. Y. I., & McManus, M. E. (1994). Functional characterization of two human sulphotransferase cDNAs that encode monoamine-and phenol-sulphating forms of phenol sulphotransferase: substrate kinetics, thermal-stability and inhibitor-sensitivity studies. *Biochemical Journal*, *302*(2), 497-502.
- Vlachogeorgos, G. S., Manali, E. D., Blana, E., Legaki, S., Karagiannidis, N., Polychronopoulos, V. S., & Roussos, C. (2008). Placental isoform glutathione S-transferase and P-glycoprotein expression in advanced nonsmall cell lung cancer: Association with response to treatment and survival. *Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society*, *114*(6), 519-526.
- Voho, A., Metsola, K., Anttila, S., Impivaara, O., Järvisalo, J., Vainio, H., ... & Hirvonen, A. (2006). EPHX1 gene polymorphisms and individual susceptibility to lung cancer. *Cancer letters*, *237*(1), 102-108.
- Walia, H. K., Sharma, P., Singh, N., & Sharma, S. (2022). Immunotherapy in Small Cell Lung Cancer Treatment: a Promising Headway for Future Perspective. *Current Treatment Options in Oncology*, 1-27.

- Walia, H. K., Singh, N., & Sharma, S. (2021). Association of NQO1Pro187Ser polymorphism with clinical outcomes and survival of lung cancer patients treated with platinum chemotherapy. *Personalized Medicine*, (0).
- Walia, H. K., Verma, U., Singh, N., & Sharma, S. (2019). Polymorphisms in GSTM1 and GSTT1 influence the response and treatment outcome in lung cancer patients treated with platinum-based chemotherapy. *British journal of biomedical science*, *76*(4), 198-200.
- Walraven, J. M., Trent, J. O., & Hein, D. W. (2007). Computational and experimental analyses of mammalian arylamine N-acetyltransferase structure and function. *Drug metabolism and disposition*, *35*(6), 1001-1007.
- Walter, J. B., & Pryce, D. M. (1955). The histology of lung cancer. *Thorax*, *10*(2), 107.
- Wang, S., Zhu, J., Zhang, R., Wang, S., & Gu, Z. (2013). Association between microsomal epoxide hydrolase 1 T113C polymorphism and susceptibility to lung cancer. *Tumor Biology*, *34*(2), 1045-1052.
- Wang, X. C., Wang, J., Tao, H. H., Zhang, C., & Xu, L. F. (2017). Combined effects of NQO1 Pro187Ser or SULT1A1 Arg213His polymorphism and smoking on bladder cancer risk: Two meta-analyses. *International journal of occupational medicine and environmental health*, *30*(5), 791.
- Wang, Y. H., Juang, G. D., Hwang, T. I., Shen, C. H., Shao, K. Y., & Chiou, H. Y. (2008). Genetic polymorphism of sulfotransferase 1A1, cigarette smoking, hazardous chemical exposure and urothelial cancer risk in a Taiwanese population. *International journal of urology*, *15*(12), 1029-1034.
- Wang, Y. H., Lee, Y. H., Tseng, P. T., Shen, C. H., & Chiou, H. Y. (2008). Human NAD (P) H: quinone oxidoreductase 1 (NQO1) and sulfotransferase 1A1 (SULT1A1) polymorphisms and urothelial cancer risk in Taiwan. *Journal of cancer research and clinical oncology*, *134*(2), 203-209.
- Wang, Y., Spitz, M. R., Tsou, A. M. H., Zhang, K., Makan, N., & Wu, X. (2002). Sulfotransferase (SULT) 1A1 polymorphism as a predisposition factor for lung cancer: a case-control analysis. *Lung Cancer*, *35*(2), 137-142.
- WATABE, T., & KANEHIRA, S. (1970). Solubilization of epoxide hydrolase from liver microsomes. *Chemical and Pharmaceutical Bulletin*, *18*(6), 1295-1296.
- Weber, W. W. (1987). *The acetylator genes and drug response*. Oxford University Press, USA.
- Weber, W. W. Hein. DW (1985) N-acetylation pharmacogenetics. *Pharmacol. Rev*, *37*, 25-79.
- Weiskopf, K. (2017). Cancer immunotherapy targeting the CD47/SIRPα axis. *European journal of cancer*, *76*, 100-109.
- Weiss. (2010, April 5). An Introduction to Lung Cancer. Retrieved October 17, 2021, from *http://cancergrace.org/lung/2010/04/05/an-introduction-to-lung-cancer/*
- Wheeler, H. E., Gamazon, E. R., Stark, A. L., O'Donnell, P. H., Gorsic, L. K., Huang, R. S., ... & Dolan, M. E. (2013). Genome-wide meta-analysis identifies variants associated

with platinating agent susceptibility across populations. *The pharmacogenomics journal*, *13*(1), 35-43.

- Wikman, H., Thiel, S., Jäger, B., Schmezer, P., Spiegelhalder, B., Edler, L., ... & Risch, A. (2001). Relevance of N-acetyltransferase 1 and 2 (NAT1, NAT2) genetic polymorphisms in non-small cell lung cancer susceptibility. *Pharmacogenetics and Genomics*, *11*(2), 157-168.
- Wilborn, T. W., Comer, K. A., Dooley, T. P., Reardon, I. M., Heinrikson, R. L., & Falany, C. N. (1993). Sequence analysis and expression of the cDNA for the phenol-sulfating form of human liver phenol sulfotransferase. *Molecular pharmacology*, *43*(1), 70-77.
- Willey, J. C., Coy, E., Brolly, C., Utell, M. J., Frampton, M. W., Hammersley, J., ... & Cairns, K. (1996). Xenobiotic metabolism enzyme gene expression in human bronchial epithelial and alveolar macrophage cells. *American journal of respiratory cell and molecular biology*, *14*(3), 262-271.
- Williamson, S. K., Crowley, J. J., Lara Jr, P. N., McCoy, J., Lau, D. H., Tucker, R. W., ... & Gandara, D. R. (2005). Phase III trial of paclitaxel plus carboplatin with or without tirapazamine in advanced non–small-cell lung cancer: Southwest Oncology Group Trial S0003. *Journal of Clinical Oncology*, *23*(36), 9097-9104.
- Windmill, K. F., Gaedigk, A., de la M. Hall, P., Samaratunga, H., Grant, D. M., & McManus, M. E. (2000). Localization of N-acetyltransferases NAT1 and NAT2 in human tissues. *Toxicological Sciences*, *54*(1), 19-29.
- Wistuba, I. I., Behrens, C., Virmani, A. K., Mele, G., Milchgrub, S., Girard, L., ... & Minna, J. D. (2000). High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Research*, *60*(7), 1949-1960.
- Wistuba, I. I., Mao, L., & Gazdar, A. F. (2002). Smoking molecular damage in bronchial epithelium. *Oncogene*, *21*(48), 7298-7306.
- Wu, H., Dombrovsky, L., Tempel, W., Martin, F., Loppnau, P., Goodfellow, G. H., ... & Plotnikov, A. N. (2007). Structural basis of substrate-binding specificity of human arylamine N-acetyltransferases. *Journal of Biological Chemistry*, *282*(41), 30189-30197.
- Wu, K., House, L., Liu, W., & Cho, W. (2012). Personalized targeted therapy for lung cancer. *International journal of molecular sciences*, *13*(9), 11471-11496.
- Wu, X., Gwyn, K., Amos, C. I., Makan, N., Hong, W. K., & Spitz, M. R. (2001). The association of microsomal epoxide hydrolase polymorphisms and lung cancer risk in African–Americans and Mexican–Americans. *Carcinogenesis*, *22*(6), 923-928.
- Wu, Y., Fan, Y., Xue, B., Luo, L., Shen, J., Zhang, S., ... & Yin, Z. (2006). Human glutathione S-transferase P1-1 interacts with TRAF2 and regulates TRAF2–ASK1 signals. *Oncogene*, *25*(42), 5787-5800.

- Xiao, J., Zheng, Y., Zhou, Y., Zhang, P., Wang, J., Shen, F., ... & Wang, M. (2014). Sulfotransferase SULT1A1 Arg213His polymorphism with cancer risk: a meta-analysis of 53 case-control studies. *PloS one*, *9*(9), e106774.
- Xu, Y., Wan, B., Chen, X., Zhan, P., Zhao, Y., Zhang, T., ... & Song, Y. (2019). The association of PD-L1 expression with the efficacy of anti-PD-1/PD-L1 immunotherapy and survival of non-small cell lung cancer patients: a meta-analysis of randomized controlled trials. *Translational lung cancer research*, *8*(4), 413.
- Xu, Z., & Taylor, J. A. (2009). SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic acids research*, *37*(suppl_2), W600-W605.
- Xue, J., Yang, S., & Seng, S. (2014). Mechanisms of cancer induction by tobacco-specific NNK and NNN. *Cancers*, *6*(2), 1138-1156.
- Yadav, U., Kumar, P., & Rai, V. (2018). NQO1 gene C609T polymorphism (dbSNP: rs1800566) and digestive tract cancer risk: a meta-analysis. *Nutrition and cancer*, *70*(4), 557-568.
- Yamada, K., Chen, Z., Rozen, R., & Matthews, R. G. (2001). Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. *Proceedings of the National Academy of Sciences*, *98*(26), 14853-14858.
- Yamada, T., Morisseau, C., Maxwell, J. E., Argiriadi, M. A., Christianson, D. W., & Hammock, B. D. (2000). Biochemical evidence for the involvement of tyrosine in epoxide activation during the catalytic cycle of epoxide hydrolase. *Journal of Biological Chemistry*, *275*(30), 23082-23088.
- Yang, E. K., Radominska, A., Winder, B. S., & Dannenberg, A. J. (1993). Dietary lipids coinduce xenobiotic metabolizing enzymes in rat liver. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, *1168*(1), 52-58.
- Yang, M., Choi, Y., Hwangbo, B., & Lee, J. S. (2007). Combined effects of genetic polymorphisms in six selected genes on lung cancer susceptibility. *Lung cancer*, *57*(2), 135-142.
- Yang, Y., & Xian, L. (2014). The association between the GSTP1 A313G and GSTM1 null/present polymorphisms and the treatment response of the platinum-based chemotherapy in non-small cell lung cancer (NSCLC) patients: a meta-analysis. *Tumor Biology*, *35*(7), 6791-6799.
- Yang, Y., Yang, L. J., Deng, M. Z., Luo, Y. Y., Wu, S., Xiong, L., ... & Liu, H. (2016). MTHFR C677T and A1298C polymorphisms and risk of lung cancer: a comprehensive evaluation. *Genetics and Molecular Research*, *15*(2).
- Yang, Y., Zhang, Y., Wu, Q., Cui, X., Lin, Z., Liu, S., & Chen, L. (2014). Clinical implications of high NQO1 expression in breast cancers. *Journal of Experimental & Clinical Cancer Research*, *33*(1), 1-9.

- Yatabe, Y., Sunami, K., Goto, K., Nishio, K., Aragane, N., Ikeda, S., ... & Dosaka-Akita, H. (2020). Multiplex gene-panel testing for lung cancer patients. *Pathology international*, *70*(12), 921-931.
- Yeh, C. C., Lai, C. Y., Chang, S. N., Hsieh, L. L., Tang, R., Sung, F. C., & Lin, Y. K. (2017). Polymorphisms of MTHFR C677T and A1298C associated with survival in patients with colorectal cancer treated with 5-fluorouracil-based chemotherapy. *International journal of clinical oncology*, *22*(3), 484-493.
- Yin, L., Pu, Y., Liu, T. Y., Tung, Y. H., Chen, K. W., & Lin, P. (2001). Genetic polymorphisms of NAD (P) H quinone oxidoreductase, CYP1A1 and microsomal epoxide hydrolase and lung cancer risk in Nanjing, China. *Lung Cancer*, *33*(2-3), 133-141.
- Yong, Y. and Lin, H. (2005), 'Shesis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci',*Cell research* 15(2), 97.
- Yoshihama, T., Fukunaga, K., Hirasawa, A., Nomura, H., Akahane, T., Kataoka, F., ... & Mushiroda, T. (2018). GSTP1 rs1695 is associated with both hematological toxicity and prognosis of ovarian cancer treated with paclitaxel plus carboplatin combination chemotherapy: a comprehensive analysis using targeted resequencing of 100 pharmacogenes. *Oncotarget*, *9*(51), 29789.
- Yu, W., Lin, Z., & Qu, B. (2015). Impact of epoxide hydrolase 1 polymorphisms on lung cancer susceptibility in Asian populations. *Cell biochemistry and biophysics*, *71*(2), 813-820.
- Yuan, W., Xu, L., Chen, W., Wang, L., Fu, Z., Pang, D., & Li, D. (2011). Evidence on the association between NQO1 Pro187Ser polymorphism and breast cancer risk in the current studies: a meta-analysis. *Breast cancer research and treatment*, *125*(2), 467-472.
- Yuan, Y., Qian, Z. R., Sano, T., Asa, S. L., Yamada, S., Kagawa, N., & Kudo, E. (2008). Reduction of GSTP1 expression by DNA methylation correlates with clinicopathological features in pituitary adenomas. *Modern Pathology*, *21*(7), 856-865.
- Yuan, Z. J., Zhou, W. W., Liu, W., Wu, B. P., Zhao, J., Wu, W., ... & Luo, Y. (2015). Association of GSTP1 and RRM1 polymorphisms with the response and toxicity of gemcitabine-cisplatin combination chemotherapy in chinese patients with non-small cell lung cancer. *Asian Pacific Journal of Cancer Prevention*, *16*(10), 4347-4351.
- Yvon, A. M. C., Wadsworth, P., & Jordan, M. A. (1999). Taxol suppresses dynamics of individual microtubules in living human tumor cells. *Molecular biology of the cell*, *10*(4), 947-959.
- Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., ... & Berger, M. F. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nature medicine*, *23*(6), 703-713.

- Zhang, B. L., Tong, S. U. N., Zhang, B. N., Zheng, S., Ning, L. Ü., Xu, B. H., ... & Lin, D. X. (2011). Polymorphisms ofGSTP1is associated with differences of chemotherapy response and toxicity in breast cancer. *Chinese medical journal*, *124*(2), 199-204.
- Zhang, J. W., Yu, W. J., Sheng, X. M., Chang, F. H., Bai, T. Y., Lv, X. L., ... & Liu, S. Z. (2014). Association of CYP2E1 and NAT2 polymorphisms with lung cancer susceptibility among Mongolian and Han populations in the Inner Mongolian region. *Asian Pacific Journal of Cancer Prevention*, *15*(21), 9203-9210.
- Zhang, J. Y., Fen Wang, Y., & Prakash, C. (2006). Xenobiotic-metabolizing enzymes in human lung. *Current drug metabolism*, *7*(8), 939-948.
- Zhang, K., Chen, D., Ma, K., Wu, X., Hao, H., & Jiang, S. (2018). NAD (P) H: quinone oxidoreductase 1 (NQO1) as a therapeutic and diagnostic target in cancer. *Journal of medicinal chemistry*, *61*(16), 6983-7003.
- Zhang, P., Zhang, Y., Yang, H., Li, W., Chen, X., & Long, F. (2015). Association between EPHX1 rs1051740 and lung cancer susceptibility: a meta-analysis. *International journal of clinical and experimental medicine*, *8*(10), 17941.
- Zhao, H., Spitz, M. R., Gwyn, K. M., & Wu, X. (2002). Microsomal epoxide hydrolase polymorphisms and lung cancer risk in non-Hispanic whites. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, *33*(2), 99-104.
- Zhao, R., & Chen, G. (2015). Role of GSTP1 Ile105Val and XRCC1 Arg194Trp, Arg280His and Arg399Gln gene polymorphisms in the clinical outcome of advanced non-small cell lung cancer. *International journal of clinical and experimental pathology*, *8*(11), 14909.
- Zhao, Y., Li, X., & Kong, X. (2015). MTHFR C677T polymorphism is associated with tumor response to preoperative chemoradiotherapy: a result based on previous reports. *Medical science monitor: international medical journal of experimental and clinical research*, *21*, 3068-3076.
- Zheng, L., Wang, Y., Schabath, M. B., Grossman, H. B., & Wu, X. (2003). Sulfotransferase 1A1 (SULT1A1) polymorphism and bladder cancer risk: a case-control study. *Cancer letters*, *202*(1), 61-69.
- Zhou, F., Yu, Z., Jiang, T., Lv, H., Yao, R., & Liang, J. (2011). Genetic polymorphisms of GSTP1 and XRCC1: prediction of clinical outcome of platinum-based chemotherapy in advanced non-small cell lung cancer (NSCLC) patients. *Swiss Med Wkly*, *141*, w13275.
- Zhou, W., Liu, G., Thurston, S. W., Xu, L. L., Miller, D. P., Wain, J. C., ... & Christiani, D. C. (2002). Genetic polymorphisms in N-acetyltransferase-2 and microsomal epoxide hydrolase, cumulative cigarette smoking, and lung cancer. *Cancer Epidemiology and Prevention Biomarkers*, *11*(1), 15-21.
- Zhou, W., Thurston, S. W., Liu, G., Xu, L. L., Miller, D. P., Wain, J. C., ... & Christiani, D. C. (2001). The interaction between microsomal epoxide hydrolase polymorphisms and

cumulative cigarette smoking in different histological subtypes of lung cancer. *Cancer Epidemiology and Prevention Biomarkers*, *10*(5), 461-466.

- Zhou, W., Thurston, S. W., Liu, G., Xu, L. L., Miller, D. P., Wain, J. C., ... & Christiani, D. C. (2001). The interaction between microsomal epoxide hydrolase polymorphisms and cumulative cigarette smoking in different histological subtypes of lung cancer. *Cancer Epidemiology and Prevention Biomarkers*, *10*(5), 461-466.
- Zhu, Q. S., Qian, B., & Levy, D. (2004). CCAAT/enhancer-binding protein α (C/EBPα) activates transcription of the human microsomal epoxide hydrolase gene (EPHX1) through the interaction with DNA-bound NF-Y. *Journal of Biological Chemistry*, *279*(29), 29902-29910.
- Zhu, J., Liu, Y., Peng Du, C. H., Hu, G., Li, S., Ye, M., ... & Tian, Q. (2019). Meta-analysis of the association between the rs9282861 polymorphism in SULT1A1 and digestive tract cancer risk. *Int J Clin Exp Med*, *12*(3), 2912-2920.
- Zhu, K., Xu, A., Zhang, B., Li, P., Jiang, H., Zhou, S., & Wang, R. (2020). Effects of NAT2 Polymorphism on Lung Cancer Risk, and the Interaction Between Smoking and NAT2: A Meta-analysis.
- Zhu, K., Xu, A., Xia, W., Li, P., Zhang, B., Jiang, H., ... & Wang, R. (2021). Association between NAT2 polymorphism and lung cancer risk: a systematic review and meta-analysis. *Frontiers in oncology*, 102.
- Zhu, Q. S., Xing, W., Qian, B., von Dippe, P., Shneider, B. L., Fox, V. L., & Levy, D. (2003). Inhibition of human m-epoxide hydrolase gene expression in a case of hypercholanemia. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1638*(3), 208-216.
- Zienolddiny, S., Campa, D., Lind, H., Ryberg, D., Skaug, V., Stangeland, L. B., ... & Haugen, A. (2008). A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of non-small cell lung cancer in smokers. *Carcinogenesis*, *29*(6), 1164-1169.
- Zilfou, J. T., & Lowe, S. W. (2009). Tumor suppressive functions of p53. *Cold Spring Harbor perspectives in biology*, *1*(5), a001883.
- Zintzaras, E., Ziogas, D. C., Kitsios, G. D., Papathanasiou, A. A., Lau, J., & Raman, G. (2009). MTHFR gene polymorphisms and response to chemotherapy in colorectal cancer: a meta-analysis. *Pharmacogenomics*, *10*(8), 1285-1294.
- Zou, J., Hallberg, B. M., Bergfors, T., Oesch, F., Arand, M., Mowbray, S. L., & Jones, T. A. (2000). Structure of Aspergillus niger epoxide hydrolase at 1.8 Å resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases. *Structure*, *8*(2), 111-122.
- Zupa, A., Sgambato, A., Bianchino, G., Improta, G., Grieco, V., La Torre, G., ... & Cittadini, A. (2009). GSTM1 and NAT2 polymorphisms and colon, lung and bladder cancer risk: a case-control study. *Anticancer research*, *29*(5), 1709-1714.

- Zusterzeel, P. L., Peters, W. H., Visser, W., Hermsen, K. J., Roelofs, H. M., & Steegers, E. A. (2001). A polymorphism in the gene for microsomal epoxide hydrolase is associated with pre-eclampsia. *Journal of medical genetics*, *38*(4), 234-237.

Appendix

Appendix-I ETHICS COMMITTEE REPORT

Institutional Ethics Committee
Postgraduate Institute of Medical Education and Research, Chandigarh

Prof. K.N. Pathak
Chairman

Phone No. 0172-2755266

Dr. Samir Malhotra
Convener

No. PGI/IEC/2018/ 001310 Dated: - 08-09-2018

To: Dr. Navneet Singh, Dept. of Pulmonary Medicine, PGIMER, Chandigarh.
From: IEC Office
Date: 29.08.2018
IEC #: IEC-04/2018-884
Subject: Genetic polymorphism in phase I/II and ABC transporter gens in lung cancer pateint undergoing platinum based chemotherapy and its association with toxcity and clinical outcome.

==========

The Institutional Ethics Committee of PGIMER, Chandigarh reviewed and discussed the above mentioned Project in its meeting, which was held on 14/15.05.2018 and sent the observations for your clarification, after receiving your reply the same is APPROVED by Institutional Ethics committee. The following documents were reviewed:-

(a) Study Protocol (including amendments, if any)
(b) Patient Information Sheet and Informed Consent Form (including updates, if any)
(c) Principal Investigator's current CV.
(d) IEC form

The following members of the IEC were present at the meeting held on 29.08.2018 at 03:00 PM in the Board Room of Kairon Block, PGIMER, Chandigarh:-

Sr. No	Name of the member	Designation	Affiliation to PGI
1.	Prof. K. N. Pathak	Chairman	No
2.	Justice V. K. Jhanji	Member (Legal)	No
3.	Dr. Virender Kumar	Member (Lay Person)	No
4.	Dr. O. P. Katare	Member (Basic Scientist)	No
5.	Dr. Sukhbir Kaur,	Member (Basic Science)	No
6.	Dr. Anil Monga	Member (Social Scientist)	No
7.	Dr. Pallab Ray	Member (Basic Science)	Yes
8.	Dr. Ashish Bhalla	Member (Clinician)	Yes
9.	Dr. Neelam Aggarwal	Member (Clinician)	Yes
10.	Dr. Ritambra Nadda	Member (Basic Sciences)	Yes
11.	Dr. Kanya Mukhopadhyay	Member (Clinician)	Yes
12.	Dr. Manish Modi	Member (Clinician)	Yes
13.	Dr. Nusrat Shafiq	Member (Pharmacologist)	Yes
14.	Dr. P.V.M. Lakshmi	Member	Yes
15.	Dr. Samir Malhotra	Convener	Yes

The Institutional Ethics Committee expects to be informed about the progress of the study, please use the Annual Progress Report Form available on the PGIMER website. Please note that it is the responsibility of the PI to inform the IEC/DSMC regarding any SAE occurring in the course of the study, or any changes in the protocol and patient information sheet /informed consent form, if applicable. You are also requested to provide a copy of the final report in the Study Completion Form available on the PGIMER website. Please be advised that your study can be periodically monitored by the IEC.

Yours sincerely,

(Samir Malhotra)
Convener
Instt. Ethics Committee, PGIMER, Chandigarh.

INSTITUTIONAL ETHICS COMMITTEE (IEC)

Department of Biotechnology

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed-to-be-University u/s 3 of the UGC Act, 1956)
Thapar Technology Campus,
Bhadson Road, Patiala - 147 004 (Punjab) India
Phone : +91-175-2393021
Email : registrar@thapar.edu
URL : www.thapar.edu

Review Letter number IEC/TIET/DBT/09/03/18

Dated: 27.09.2018

The Institutional ethics committee of Thapar Institute of Engineering & Technology, Patiala, in its Sixth (06) meeting held on 27.09.2018, in the office of Head, Department of Biotechnology, reviewed and discussed the project submitted by Dr. Siddharth Sharma as PI from Department of Biotechnology.

TIET/DBT/IEC/09/03/18: Genetic polymorphism in Phase I/II and ABC transporter genes in Lung Cancer Patient undergoing Platinum based chemotherapy and its association with toxicity and clinical outcomes.

The members reviewed the following documents and raised queries which were satisfactorily answered by Dr. Siddharth Sharma.

a) Proposal & Protocol

b) Available Safety Information

c) Cell line/ Genetic material storage and discard procedure

It is understood that the study would be conducted under the PI's direction strictly as per the submitted protocol. In case of any deviation from the approved methods, the study documents will have to be resubmitted to ethics committee for re-approval.

Members	**Signature**
Dr. Manoj Baranwal (Member)	
Dr. Siddharth Sharma (Member)	
Dr. Vikas Handa (Member)	
Dr. Sanjai Saxena (Member Secretary)	

The present approval is for 01 year only; the investigators are advised to take a re-approval after one year. 06 Monthly compliance report is a must for seeking re-approval.

Signature

Prof. Moushumi Ghosh

Chairman, Institutional Ethics Committee

HEAD
Department of Biotechnology
Thapar Institute of Engineering & Technology
Patiala - 147004, Punjab

POSTGRADUATE INSTITUTE OF MEDICAL EDUCATION & RESEARCH, CHANDIGARH
MEDICAL EDUCATION AND RESEARCH CELL
PHONE: 0172-2755568

No: 79/179-Edu-18/4913 Dated: 05/12/2018

To

Dr. Navneet Singh
Additional Professor
Department of Pulmonary Medicine
PGIMER, Chandigarh

Subject: Collaborative Research project titled **"Genetic Polymorphism in Phase I/II and ABC Transporter Genes in Lung Cancer patient undergoing platinum based chemotherapy and its association with toxicity and clinical outcomes"**

Please refer to your letter no. 2765 dated 02/11/2018 on the subject cited above.

Approval of the ICRC Committee is herewith conveyed for the collaborative research project titled "Genetic Polymorphism in Phase I/II and ABC Transporter Genes in Lung Cancer patients undergoing platinum based chemotherapy and its association with toxicity and clinical outcomes"

5/12

Sub-Dean (Research)
Medical Education & Research Cell
PGIMER, Chandigarh

Appendix-II PATIENT CONSENT FORM

Patient consent form

Name of the participant: _____

Name of the Principal (Co-) Investigators: _____

Name of the Institution: _____

Name and address of the sponsoring (funding) agency(ies): _____

_____

Documentation of the informed consent

I, … … … … … … … … … … . have read the information in this form (or it has been read to me). I was free to ask any questions and they have been answered. I am over 18 years of age and, exercising my free power of choice, hereby give my consent to be included as a participant in the project "**Genetic Polymorphisms in Phase I/II and ABC Transporter Genes in Lung Cancer patient undergoing platinum based chemotherapy and its association with toxicity and clinical outcomes**"

(1) I have read and understood consent form and the information provided to me.

(2) I have had the consent document explained to me.

(3) I have been explained about the nature of the study.

(4) My rights and responsibilities have been explained to me by the investigator.

(5) I have the right to not be a subject in the above study and am not willing to give sample

(6) I have informed the investigator of all the treatments I am taking or have taken in the past … … . months including any *desi* (alternative) treatments.

(7) Patient data and confidentiality has been explained to me.

(8) I hereby give permission to the investigators to release the information obtained from me as result of participation in this study to the sponsors, regulatory authorities, Government agencies, and ethics committee. I understand that they may inspect my original records.

(9) My identity will be kept confidential if my data are publicly presented.

(10) I have had my questions answered to my satisfaction.

(11) I have decided to be in the research study.

I am aware, that if I have any questions during this study, I should contact the concerned authorities. By signing this consent from, I attest that the information given in this document. I will be given a copy of this consent document.

For adult participants

Name and signature / thumb impression of the participant (or legal representative if participant incompetent):

_____ (Name) _____ (Signature)

Date: _____ Time: _____

Name and signature of impartial witness (required for illiterate patients):

_____ (Name) _____ (Signature)

Date:_____ Time: _____

Address and contact number of the impartial witness: _____

_____

Name and signature of the Investigator or his representative obtaining consent:

_____ (Name) _____ (Signature)

_____ (Date)

Investigator Certificate

I certify that all the elements including the nature of the research study and possible risks of the above study as described in this consent document have been fully explained to the subject. In my judgment, the participant possesses the legal capacity to give informed consent to participate in this research and is voluntarily and knowingly giving informed consent to participate,

Signature of the Investigator: _____ Dated:_____

Name of the Investigator: _____

Appendix-III PATIENT INFORMATION SHEET

PATIENT INFORMATION SHEET

Department of Biotechnology, Thapar University &
Dept. of Pulmonary Medicine, PGIMER, Chandigarh

Investigator: Siddharth Sharma, Ph.D
Co-Investigator: DrNavneet Singh, M.D, D.M

Name of Patient _____

Address(Rural/Urban) _____

C.R Number _____

Date of Sampling _____

Type of Cancer _____

Stage of Cancer _____

Sex _____

Age _____

Education (Years) _____

Occupation _____

Marital Status _____

Smoking Status

1. Smoker or Non-Smoker _____

2. Type of Smoking

a) Cigarette b) Beedi c) Hookah d) Cigarette+Beedi e) Beedi+Hookah f) Any other form of smoke

3. No. of Cigarettes/ Beedis smoked

a) Per day_____

b) Per Week _____

c) Amount of tobacco added to hookah _____

4. a) Age at which started Smoking _____

b) How many years have you smoked _____

5. Time taken to finish cigarette (in min) _____

6. Brand of smoke preferred _____

7. Do you chew tobacco/betelnut

a) Yes b) No

If yes amount chewed _____

8. Any familial case of Smoking related cancer in your family: a) Yes b) No

9. Are you only one to smoke in your family

Alcohol Consumption

1. Do you consume alcohol
   a) Yes b) No
2. Intake of Alcohol
   a) Daily b) 3-5 times a week c) 1-2 times a week d) once a week e) once a month
3. How much alcohol do you consume in ml_____
4. What type of alcohol do you consume _____

Documentation of the informed consent

I, … … … … … … … … … … .have read the information in this form (or it has been read to me). I was free to ask any questions and they have been answered. I am over 18 years of age and, exercising my free power of choice, hereby give my consent to be included as a participant in the project "**Association ofSingle Nucleotide Polymorphisms and Methylation status in the Wnt and AhR signaling pathways with risk for occurrence of Lung Cancer in a North Indian Population** "

(1) I have read and understood consent form and the information provided to me.

(2) I have had the consent document explained to me.

(3) I have been explained about the nature of the study.

(4) My rights and responsibilities have been explained to me by the investigator.

(5) I have the right to not be a subject in the above study and am not willing to give sample

(6) I have informed the investigator of all the treatments I am taking or have taken in the past … … . months including any *desi* (alternative) treatments.

(7) Patient data and confidentiality has been explained to me.

(8) I hereby give permission to the investigators to release the information obtained from me asresult of participation in this study to the sponsors, regulatory authorities, Government agencies, and ethics committee. I understand that they may inspect my original records.

(9) My identity will be kept confidential if my data are publicly presented.

(10) I have had my questions answered to my satisfaction.

(11) I have decided to be in the research study.

I am aware, that if I have any questions during this study, I should contact the concerned authorities. By signing this consent from, I attest that the information given in this document. I will be given a copy of this consent document.

Appendix-IV

- 0.5M EDTA: Dissolved 9.306g of disodium salt of EDTA in 20ml of deionised water, and then adjusted the pH to 8.0 by 1 M sodium hydroxide. Sterilized the solution by autoclaving.
- 10% SDS: Dissolved 1g of SDS in 10ml of deionised water.
- 3. 100mM Tris-Cl (pH 8.0): Dissolved 0.32g of Tris-Cl in 10 ml of deionised water, then adjusted the pH to 8.0 by 1M sodium hydroxide. Sterilized the solution by autoclaving.
- 10mg/ml Proteinase K: Dissolved 10mg Proteinase K in 1ml of double distilled water. Sterilized the solution by autoclaving.
- 1mg/ml BSA: Dissolved 100mg of BSA in 100ml of deionised sterile water and kept at 4 ˚C overnight.
- 5M Sodium chloride (NaCl): Dissolved 5.85g of sodium chloride in 20ml of deionised water. Sterilized the solution by autoclaving.
- 5X TBE buffer: Dissolved 54g of Tris base and 27.5g of boric acid in 980ml of double distilled water and then added 20ml of 0.5 EDTA. Sterilized the solution by autoclaving.
- Ethidium Bromide (10mg/ml): Dissolved 1g of ethidium bromide in 100ml of water. Mixed the solution properly.
- Magnesium chloride ($MgCl_2$) (100mM): Dissolved 0.41gms of MgCl2 in 20ml of deionised water and sterilized by autoclaving.
- Sucrose (1M): Dissolved 3.41 g of sucrose in 10 ml of deionised water and sterilized by autoclaving.
- TE buffer (pH 8.0): Added 1ml of 100mM Tris-Cl (pH 8.0) and 200 µl of 0.5M EDTA solution to 8.8 ml of deioinsed water. Sterilized the solution.
- Triton X- 100 (10%): Took 100 µl of TritonX-100 and mixed with 900 µl of deionised water and mixed properly**.**

Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Personalized Medicine

Association of *NQO1Pro187Ser* polymorphism with clinical outcomes and survival of lung cancer patients treated with platinum chemotherapy

Harleen Kaur Walia[1], Navneet Singh[2] & Siddharth Sharma*[,1]

[1]Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India

[2]Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research (PGIMER), Chandigarh, India

*Author for correspondence: Tel.: +91 950 168 8366; siddharthsharma.phd@thapar.edu

Background: The study was carried out to evaluate the association of *NQO1 P187S* polymorphism in North Indian lung cancer (LC) patients. We determined the effect of this polymorphic variant on the survival of LC patients. **Patients & methods/results:** This case–control study comprised a total of 1100 subjects. The genotyping was carried out using PCR-RFLP and statistical analysis was carried out. The variant TT genotype exhibited 3.5-fold higher odds in subjects with stage III ($p = 0.0006$), fivefold higher odds of lymph-node invasion ($p = 0.007$) and an odd of <1 in case of metastasis ($p = 0.0028$). Patients possessing TT genotype and administered with paclitaxel, exhibited a poor survival (3.57 vs 12.20 months; hazard ratio = 7.95; $p = 0.0098$). **Conclusion:** These results suggest that *NQO1* variant genotype was not found to modulate risk toward LC. However, the variant genotype was found to be strongly correlated with stage III LC, lymph node invasion and was found to be positively correlating with metastasis.

First draft submitted: 23 July 2020; Accepted for publication: 23 March 2021; Published online: 11 May 2021

Keywords: association • chemotherapy • cisplatin • lung cancer • *NQO1* • overall survival • pack years • polymorphism • smoking • variant genotype

Cancer is the leading cause of mortality worldwide, accounting for an estimated 1.8 million deaths and 2.1 million new cases in 2018 [1]. Among the different types of cancer, lung cancer (LC) has the highest cases of mortality and morbidity in both developed and developing countries. Both tobacco smoking and exposure to different carcinogenic compounds are one of the major risk factors for predisposition toward LC [2]. Some of the known carcinogens found in tobacco smoke are 4-(methyl-nitrosamino)-1-(3-pyridyl)-1-butanone, polycyclic aromatic hydrocarbons, aromatic amines, nitroso-compounds, etc. [3].

The biotransformation of xenobiotics which includes carcinogens is carried out by an inbuilt metabolic system, which involves both phase I and II enzymes. The phase I reactions include conversions to polar metabolite(s) by certain reactions *viz.* aliphatic and aromatic hydroxylation, *N*- and *S*-dealkylation or oxidation and deamination. The main enzyme in phase I reactions is mediated by CYP450. The phase II biotransformation reactions convert it to more polar form by the conjugation reactions. These include sulfation, acetylation, glucuronidation, methylation, glutathione and amino acid conjugation [3,4].

NADPH: quinone oxidoreductase 1 (NQO1) is an important flavoenzyme which is a phase II cytosolic enzyme and is involved in catalyzing the metabolism of quinones [5,6]. It is involved in reduction of two electrons of the carcinogenic quinones to its reduced form, hydroquinones [7]. The occurrence of LC is implicated to the accumulation of the toxicants, thus leading to an increased reactive oxygen species production in human tissues, which furthermore result in upregulation of antioxidant selective genes. The growing tumor and its microenvironment are additional sites for the generation of reactive oxygen species which stem from accelerated mitochondrial activity that is needed for the rapid growth of cells and division in the proliferative stage and thus may be linked to oncogenic stimulus [8]. Thus, the role of phase II antioxidant enzymes such as NQO1 and glutathione-S-transferase which

XENOBIOTICA
https://doi.org/10.1080/00498254.2021.1963008

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

Genetic polymorphism of *Arg*^{213}*His* variant in the *SULT1A1* gene is associated with reduced susceptibility to lung cancer in North Indian population

Harleen Kaur Walia[a], Navneet Singh[b] and Siddharth Sharma[a]

[a]Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India; [b]Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research (PGIMER), Chandigarh, India

ABSTRACT

1. Sulfotransferases (SULTs) are phase II detoxification enzymes that is involved in the biotransformation of many compounds including tobacco carcinogens. A polymorphism in the *SULT1A1* (*Arg*^{213}*His*) gene results in reduced enzyme activity.
2. We investigated the association between the *SULT1A1* (Arg^{213}/His) genotype and lung cancer (LC). This case-control study comprised of 550 cases and controls, matched on age, gender and smoking status.
3. The variant genotype exhibited no association with LC risk, even after stratification on basis of histological subtypes. Male LC patients carrying the variant *His*^{213} allele (**$p = 0.02$**) did not exhibit an increased risk towards LC. Smokers harbouring the *Arg/His* genotype did demonstrate a reduced risk towards LC (AOR = 0.70; **$p = 0.019$**). Furthermore, the LC subjects who were heavy smokers and harbouring the *Arg/His* genotype (AOR = 0.28; **$p = 0.019$**) did not show a genetic predisposition towards LC susceptibility. The subjects who smoked pack years of above 40 and carrying the *His/His* (AOR = 0.28; **$p = 0.036$**) genotype were found to have a reduced risk for LC. Furthermore, 473 subjects were analysed in regards to overall survival, wherein the *His/His* genotype exhibited better OS than *Arg/Arg* genotype (11.30 vs. 8.07 months).
4. This study provides evidence of no genetic predisposition towards LC risk associated with *SULT1A1 Arg*^{213}*His* polymorphism in relation to tobacco smoking.

ARTICLE HISTORY
Received 15 June 2021
Revised 27 July 2021
Accepted 28 July 2021

KEYWORDS
SULT1A1; polymorphism; lung cancer; association; overall survival

Introduction

Lung cancer is one of the leading causes of death due to cancer worldwide and is the most commonly diagnosed cancer. In India, lung cancer constitutes about 5.9% of all cancers with approximately 67,795 new incident cases reported in 2018 and is the second most common cause of cancer in males (Bray et al. 2018). Both the environmental factors (*i.e.* tobacco use, smoking and dietary consumption of carcinogens) and genetic factors play a role in the high incidence rate of lung cancer (Ihsan et al. 2011; de Groot et al. 2018). The variety of procarcinogens present in tobacco smoke is activated by metabolic enzymes *viz.* N-acetyltransferases, cytochrome P450, sulfotransferases and myeloperoxidases to genotoxic molecules. The metabolic enzymes' functionality may be affected due to mutation or polymorphism in these enzymes resulting in the variable rate of detoxification or activation of procarcinogens by these metabolic enzymes (Ihsan et al. 2011). The risk for lung cancer may be modulated due to genetic polymorphism in the metabolic enzymes.

The human sulfotransferases (SULTs) is an important phase II detoxification enzyme that catalyses the sulphonate conjugation of various endogenous compounds such as neurotransmitters, many drugs, steroid and thyroid hormones, xenobiotics and pro-carcinogenic agents (Xiao et al. 2014; Dash et al. 2019). The sulfotransferases are found to be differentially expressed in different types of human tissues such as the lung, brain, colon, liver, platelet as well as other sites (Jones et al. 1992; Li et al. 2008). Sulfonation is commonly known to be a detoxifying mechanism that produces more water-soluble and therefore fewer toxic metabolites; however, this conjugation reaction of some compounds creates electrophiles that may bind to DNA to form carcinogen-DNA adducts (Coughtrie et al. 1998; Banoglu, 2000). So far, 13 human cytosolic SULT isoforms have been recognized and further grouped as four key families: SULT1, SULT2, SULT4, SULT6 (Blanchard et al. 2004; Kumar et al. 2013). The most expressed member of the phenol SULT1 family is SULT1A1. The SULT1A1 enzyme is expressed by the *SULT1A1* gene, which maps to chromosome 16p11.2-p12.1. Apart from the inactivation of toxic compounds SULT1A1 isoform is also involved in the bioactivation of procarcinogens such as polycyclic aromatic hydrocarbons (PAHs) and heterocyclic amines (HCAs), both of which are present in tobacco smoke (Bardakci et al. 2008; Koike et al. 2008; Kumar et al. 2013).

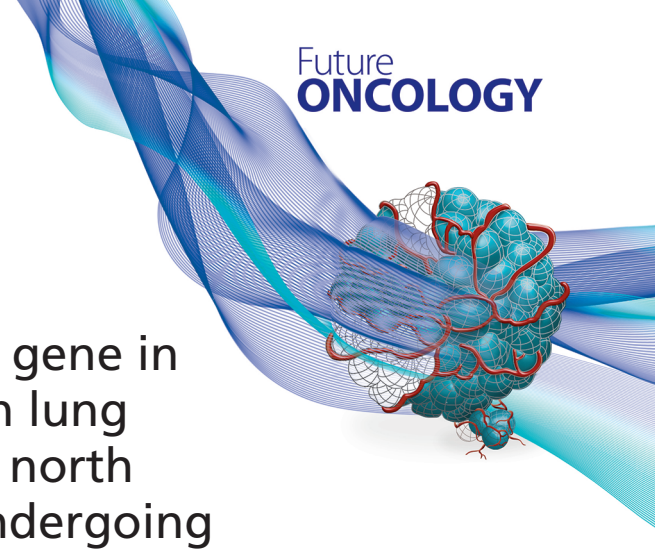
CONTACT Siddharth Sharma siddharthsharma.phd@thapar.edu Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab, 147004, India

Supplemental data for this article can be accessed here.

© 2021 Informa UK Limited, trading as Taylor & Francis Group

Research Article

For reprint orders, please contact: reprints@futuremedicine.com



Genetic polymorphisms in the *mEH* gene in relation to tobacco smoking: role in lung cancer susceptibility and survival in north Indian patients with lung cancer undergoing platinum-based chemotherapy

Harleen Kaur Walia[1], Navneet Singh[2] & Siddharth Sharma*[,1]

[1]Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala 147004, India

[2]Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research (PGIMER), Chandigarh 160012, India

*Author for correspondence: Tel.: +91 95 0168 8366; siddharthsharma.phd@thapar.edu

Aim: Epoxide hydrolase is involved in oxidative defenses and is responsible for the activation of carcinogens. The relationship between EPHX1 polymorphisms (Tyr^{113}His and His^{139}Arg) and overall survival (OS) and lung cancer (LC) risk was investigated. **Methods:** The study comprised 550 cases and 550 controls. Genotyping and statistical analysis were applied. **Results:** The variant genotypes of EPHX1 polymorphisms exhibited no association with LC risk. The Tyr^{113}His polymorphism exhibited twofold increased odds of lymph node invasion (p = 0.04). The Tyr/His genotype is a risk factor for smokers. Subjects carrying the combined genotype for His^{139}Arg showed better median survival time (MST) and the heterozygous genotype revealed better MST in the case of small-cell lung cancer (SCLC; 11.30 vs 6.73 months; log-rank test: p = 0.02). The heterozygous genotype (His139Arg) had longer MST in patients receiving cisplatin/carboplatin and irinotecan (11.30 vs 7.23; log-rank test: p = 0.007) **Conclusion:** The Tyr^{113}His polymorphism is associated with LC risk in smokers and is a potential prognostic factor for OS in patients with SCLC after irinotecan.

Lay abstract: Microsomal epoxide hydrolase (mEH) is an enzyme that plays a defensive role against chemicals. In this study, the relationship between the variation of the epoxide hydrolase and the risk of lung cancer was investigated and its role in the survival of patients with lung cancer was evaluated. The study comprised 550 cases and 550 controls. Genotyping was carried out using molecular biology tools and was followed by statistical analysis. The variant genotype of the *EPHX1* gene was not associated with the risk of lung cancer, even based on histology. The variant form of the *EPHX1* gene was found to be a risk factor for smokers. The Tyr/His genotype was associated with the risk of lung cancer in male subjects. Patients carrying the variant form of the *EPHX1* gene (His^{139}Arg) experienced better survival. The heterozygous genotype of the *EPHX1*Tyr^{113}His gene was related to longer survival time in patients who received cisplatin/carboplatin along with irinotecan. The *EPHX1* Tyr^{113}His polymorphism is associated with LC risk in smokers and is a potential prognostic factor for OS of patients with small-cell lung cancer (SCLC) after irinotecan therapy and might increase the likelihood of lymph node metastasis; *EPHX1*His^{139}Arg exhibited better survival, especially in patients with SCLC.

First draft submitted: 5 April 2021; Accepted for publication: 25 August 2021; Published online: 21 October 2021

Keywords: association • lung cancer • microsomal epoxide hydrolase • overall survival • platinum-based chemotherapy • polymorphism

Lung cancer (LC) has become a global health burden due to its high mortality and incidence rates, and about 50% of all cases are diagnosed in developing countries [1]. Evidence of LC in genome-wide nonsmoking individuals

MTHFR polymorphism as a predictive biomarker for gastrointestinal and hematological toxicity in North Indian adenocarcinoma patients

Left running head: H. K. WALIA ET AL.

AQ0

Harleen Kaur Walia[a] AQ6 Navneet Singh[b] AQ6 Siddharth Sharma[a]

[a]Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India; AQ1

[b]Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research (PGIMER), Chandigarh, India

Supplemental data for this article is available online at https://doi.org/10.1080/1120009X.2021.1997008.

CONTACT Siddharth Sharma siddharthsharma.phd@thapar.edu Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab 147004, India. AQ2

Abstract

Background

In the present study, we aimed to investigate the relationship between the methylenetetrahydrofolate reductase (*MTHFR*) *C677T* and *A1298C*polymorphisms and overall survival and treatment response for North Indian adenocarcinoma patients. Furthermore, we have also analyzed the correlation between hematological and gastrointestinal toxicity of PEM chemotherapy and *MTHFR* polymorphisms in patients with adenocarcinoma.

Methods

The polymorphisms of the *MTHFR* gene in north Indian adenocarcinoma patients were assessed by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). Univariate and multivariate analyses correlated genotyping data with the response, clinical benefit, toxicity, and overall survival (OS) using log-rank test and Cox proportional hazards model was calculated.

Results

FULL TEXT LINKS

Future Medicine
futuremedicine.com

Pharmacogenomics. 2022 Jan;23(2):97-118. doi: 10.2217/pgs-2021-0080. Epub 2021 Dec 16.

Association of *NAT-2* gene polymorphisms toward lung cancer susceptibility and prognosis in North Indian patients treated with platinum-based chemotherapy

Harleen Kaur Walia [1], Navneet Singh [2], Siddharth Sharma [1]

Affiliations

PMID: 34911343 DOI: 10.2217/pgs-2021-0080

Abstract

Aim: The present study has been carried out to evaluate the association of the N-acetyl transferase 2 (*NAT2*) variants in North Indian lung cancer patients and healthy controls. Furthermore, we have also determined the effect of the polymorphic variants of the *NAT2* gene on the clinical outcomes and overall survival among lung cancer (LC) subjects treated with platinum-based doublet chemotherapy. **Methods:** This case-control study comprised a total of 550 cases and 550 healthy controls. The genotyping was carried out using polymerase chain reaction restriction fragment length polymorphism (PCR-RFLP) and the statistical analysis was carried out using MedCalc. **Results:** There was a lack of any significant association for both 590G>A and 803A>G polymorphisms toward risk for LC, but 857G>A polymorphism exhibited a risk toward LC ($p = 0.005$). Whereas, variant alleles for the 481C>T polymorphism had a decreased risk for LC ($p = 0.0003$). Further, 857G>A polymorphism conferred a positive association between genotype and ADCC ($p = 0.001$) and 481C>T polymorphism had a decreased risk for SQCC (OR = 0.39, $p = 0.0006$) and SCLC ($p = 0.001$) subjects. The smokers carrying mutant genotype for the 481C>T polymorphism had a decreased risk toward LC ($p < 0.0001$) even in light ($p = 0.002$) as well as heavy smokers ($p = 0.001$). In case of females, 2.59-fold and 3.66-fold increased risk of LC development was observed in subjects with intermediate and slow acetylator for the 857G>A polymorphism. Whereas, in case of males this polymorphism depicts a reduced risk for LC. On the other hand, 803A>G depicted a 2.82-fold risk of LC in case of female subjects who were slow acetylators. Our study exhibits a significant difference in the overall haplotype distribution between cases and controls. In our study overall, (857G>A, 481C>T, 803A>G) was found to be best model, but was not significant using MDR. Considering the CART results 481C>T polymorphism came out to be the most significant factor in determining the LC risk. For the 803A>G polymorphism, a threefold odds of lymph node invasion were observed for mutant genotype, the recessive model exhibited an odds ratio of 2.8. 590G>A appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy as the survival time for such patients was better. **Conclusion:** These results suggest that *NAT2* variant genotype for 590G>A and 803A>G was not found to modulate risk toward LC, but 857G>A polymorphism exhibited a risk toward LC and 481C>T polymorphism had a decreased risk for LC. *NAT2* 590G>A appears to be a potential prognostic factor for OS of SCLC patients after irinotecan therapy and 481C>T came out to be significant factor using CART.

Keywords: NAT2; association; chemotherapy; cisplatin; lung cancer; overall survival; pack-years; polymorphism; smoking.

Related information

MedGen

Curr. Treat. Options in Oncol.
DOI 10.1007/s11864-022-00949-1

Lung Cancer (TA Leal and N Sethakorn, Section Editors)

Immunotherapy in Small Cell Lung Cancer Treatment: a Promising Headway for Future Perspective

Harleen Kaur Walia, M.Tech[1]
Parul Sharma, M.E.[1]
Navneet Singh[2]
Siddharth Sharma[1,*]

Address
[*,1]Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab, 147004, India
Email: siddharthsharma.phd@thapar.edu
[2]Department of Pulmonary Medicine, Post-Graduate Institute of Medical Education & Research, Chandigarh, 160012, India

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Harleen Kaur Walia and Parul Sharma contributed equally to this work.
This article is part of the Topical Collection on *Lung Cancer*

Keywords Lung cancer · SCLC · Immunotherapy · Check-point Inhibitors · PD-1 · CTLA-4

Opinion statement

Despite advancements in clinical research, both prognosis and treatment for SCLC patients are still in the nascent stage. SCLC is a fatal disease with high tumor mutational burden and is strongly associated with exposure to tobacco. This leads to the development of potential neo-antigens, inhibition of immune responses, and development of paraneoplastic disorders. Surgery, radiation, and chemotherapy are widely accepted treatments for cancer globally, and most recently, immunotherapy has now become the "fourth pillar" of SCLC treatment. Various immune checkpoint pathways regulate the activation of T cells at multiple stages during an immune response. T cell checkpoint inhibitors such as anti-PD1 (pembrolizumab, nivolumab), anti-PDL1, and anti-CTLA-4 (tremelimumab, ipilimumab) have potential to show anti-cancer activity along with the promise to prolong survival in patients with SCLC. Treatment with the CTLA-4-specific antibodies can restore the immune response by increasing the accumulation and survival of T-cells whereas monoclonal antibodies block either PD-1 or its ligands that prevent downregulation of effector T-cell, which enables the T-cells to mediate the death of tumor cells. Furthermore, monoclonal antibody in combination with chemotherapy has attained

Published online: 28 February 2022

XENOBIOTICA
https://doi.org/10.1080/00498254.2022.2069527

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

Predictive role of polymorphic variants of phase II drug metabolising enzyme in modulating toxicity in North Indian lung cancer patients undergoing chemotherapy

Harleen Walia[a], Parul Sharma[a], Navneet Singh[b] and Siddharth Sharma[a]

[a]Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India; [b]Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research, Chandigarh, India

ABSTRACT

1. Genetic polymorphism of drug-metabolising enzymes such as *NQO1, SULT1A1, EPHX1*, and *NAT2* alters its activity which hampers the detoxification and disposal of chemotherapeutic compounds. Thus, in the present study, we have comprehensively investigated the associations between SNPs of the Phase II detoxifying genes and its relationship towards platinum-induced toxicity of lung cancer patients.
2. *NQO1* (609 C > T), *SULT1A1* (Arg^{213} His), *EPHX1* (Tyr^{113}His, His^{139}Arg), and *NAT2* (481 C > T, 803 A > G, 590 G > A, 857 G > A) were evaluated in our study for their associated adverse events caused due to the administration of platinum-based chemotherapy to the lung cancer patients.
3. For *NQO1* 609 C > T polymorphism, the *TT* genotype showed reduced risk of constipation (OR = 0.10, $p = 0.04$) and anorexia (OR = 0.15, $p = \mathbf{0.03}$). For *SULT1A1 Arg^{213}His,* heterozygous genotype (*Arg/His*) (AOR = 0.38, $p = \mathbf{0.006}$) and combined genotype (*Arg/His + His/His*) were not associated with increased risk of nephrotoxicity (AOR = 0.38, $p = \mathbf{0.004}$). For *NAT2,* heterozygous (*NAT2*4/*6*) and combined genotypes (*NAT2*4/*4+*4/*6*) for *NAT2*6* polymorphism exhibit 2.4 folds ($p = \mathbf{0.005}$), and two-folds ($p = \mathbf{0.01}$) increased risk of hematological toxicity. The heterozygous (AOR = 0.45, $p = \mathbf{0.004}$) and variant genotype (AOR = 0.39, $p = \mathbf{0.02}$) for *NAT2*5C* had decreased risk for hematological toxicity. The heterozygous genotype for *NAT2*7* polymorphism showed two-fold increased risk for developing thrombocytopenia.
4. This study provides association of *NAT2* polymorphic variants in predicting haematological toxicity.

ARTICLE HISTORY
Received 23 March 2022
Revised 20 April 2022
Accepted 20 April 2022

KEYWORDS
Phase II enzymes; lung cancer; polymorphism; toxicity; risk

Introduction

Lung carcinoma is the most prevalent cause of mortality from cancer, contributing to 18.0% of all cancer deaths (Siegel et al. 2012). According to the International Agency for Research on Cancer (IARC), less developed nations account for more than half of lung cancer cases (Molina et al. 2008; Bray 2014). The 5-year survival rate of lung cancer is estimated to be just 15%, with a median survival time of fewer than ten months. The standard treatment for lung cancer is the combination of platinum-based therapies such as cisplatin and carboplatin, along with third-generation agents (docetaxel, paclitaxel, gemcitabine, and irinotecan) (Walia et al. 2021). However, treatments with platinum-based regimens, on the other hand, are linked with adverse reactions that might cause patients to experience common ailments and prohibit them from receiving further treatment (Qi et al. 2016).

Genetic factors are thought to play a significant role in modulating the level of carcinogen in an individual that might lead to risk differences among people who have had the same exposure to carcinogen. (Shields 1999). This may be due to differences in the rate of metabolism or detoxification of xenobiotic chemicals and any alteration in the balance between these two processes effects the rate of metabolism of carcinogens. Inhaled pro-carcinogens get bioactivated into DNA-reactive intermediates by phase I drug-metabolising enzymes, but they are also detoxicated by 'phase II' enzymes such sulfotransferases (*SULT*), N-acetyltransferases (*NAT*), and NADPH quinone oxidoreductase 1 (*NQO1*) (Nishikawa et al. 2004). Role of phase II reactions are typically detoxifying, the conjugates generated might also have adverse effects (Lv et al. 2014; Qi et al. 2016). Individual predisposition to develop chemically induced carcinoma is greatly influenced by a cellular imbalance between phase I and phase II detoxifying enzymes (Sheweita and Tilmisany, 2003).

Genetic polymorphism of drug-metabolising enzymes such as *NQO1, SULT1A1, EPHX1*, and *NAT2* alters its activity or

CONTACT Siddharth Sharma siddharthsharma.phd@thapar.edu Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, Punjab-147004, India

Supplementary data for this article can be accessed here

© 2022 Informa UK Limited, trading as Taylor & Francis Group