

Studies on genetic polymorphism in nucleotide excision repair genes in relation to lung cancer patients

A Thesis

Submitted in partial fulfillment of the requirements for the award of the degree of

**Doctor of philosophy
in
Department of Biotechnology**

by

**Shweta Lawania
(Reg No: 901400010)**

**Under supervision of
Dr Siddharth Sharma
Associate Professor**



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

**Thapar Institute of Engineering & Technology
Patiala, India
October, 2018**

Candidate Declaration

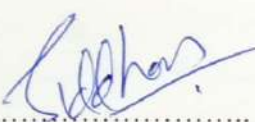
I hereby certify that the work, which is being presented in the thesis, entitled "**Studies on genetic polymorphism in nucleotide excision repair genes in relation to lung cancer patients**", in partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy in Biotechnology** from Thapar University, Patiala is an authentic record of my own work carried out during the period Jan 2015 to Sept 2018 under the supervision of **Dr. Siddharth Sharma**. I have also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

The matter presented in this thesis has not been submitted elsewhere for the award of any other degree or diploma from any institution.

Date: 28/5/2019


.....
Shweta Lawania
Candidate

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


.....
Dr. Siddharth Sharma
Supervisor

Abstract

Background

DNA repair genes safeguard the genomic integrity from the exogenous as well as endogenous assaults. Molecular alterations in the repair pathway genes may lead to improper repair and ultimately carcinogenesis.

Objectives

To evaluate the role of single nucleotide polymorphic variants of Nucleotide Excision Repair (NER) genes *i.e.* *XPC* (*Lys*⁹³⁹*Gln*, *Ala*⁴⁹⁹*Val*), *XPA* (*A23G*, *G709A*), *XPD* (*Lys*⁹³⁹*Gln*, *Asp*³¹²*Asn*, *Arg*¹⁵⁶*Arg*), *XPG* (*Phe*⁶⁷⁰*Leu*, *Asp*1104*His*) and *XPF* (*673C>T*, *11985A>G* and *Arg*⁴¹⁵*Gln*) with lung cancer susceptibility, overall survival and clinical response. Gene-gene and gene-environment interaction were also assessed to understand their association with lung cancer.

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 for each polymorphic variant was 740 (370 cases, 370 controls). Following this association analysis was carried out using logistic regression to obtain adjusted odds ratio and significance. Data mining analysis was performed including both Multi-dimensionality reduction (MDR) and Classification and Regression tree (CART) analysis to find the possible interaction between interacting SNP-SNP and gene-gene. Overall survival analysis for 311 cases was performed using Kaplan Meier survival analysis and Cox-regression analysis which gave adjusted hazards ratio. Multivariate logistic regression analysis for 202 samples was also conducted to evaluate association of different polymorphic variants with clinical response.

Results

*XPC Lys*⁹³⁹*Gln* possessed risk for lung cancer (OR: 2.30; $p=0.0007$). *Lys*⁹³⁹*Gln* polymorphism also showed higher risk for all sub-groups of lung cancer. For subjects with mutant genotype (CC) had a 4 and 3-fold increased risk for ADCC ($p<0.0001$) and SQCC ($p=0.009$). The mutant

genotype (CC) for *Lys*⁹³⁹*Gln* also presented a higher risk of developing lung cancer in heavy smokers (OR: 3.71; $p=0.005$). Another *XPC* polymorphic variant *i.e.* *Ala*⁴⁹⁹*Val* showed a protective association towards lung cancer (OR: 0.25; $p=0.003$). The survival analysis suggested no significant association between the *XPC* variants and survival of lung cancer patients. Further, the study showed association of A23G polymorphism with mutant genotype (GG) suggesting a 2.5-fold elevated risk for the disease ($p=0.0007$). For lung cancer histological subtypes: SQCC ($p=0.01$), ADCC ($p=0.05$) and SCLC patients ($p=0.0007$) an increased lung cancer risk was observed. Smokers with mutant genotype (GG) exhibited a high risk for the lung cancer when compared to wild genotype (AA) for A23G polymorphism ($p=0.0002$). In G709A polymorphism, the heterozygous genotype (GA) displayed a protective effect towards the disease for subjects who were smokers ($p=0.001$). No significant association was observed between *XPA* polymorphism and overall survival in both Kaplan-Meier analysis and Cox regression model. The study further revealed association of mutant genotype (CC) with an increased lung cancer risk ($p=0.01$) and also with its all histological subtypes *i.e.* ADCC ($p=0.03$), SQCC ($p=0.008$) and SCLC ($p=0.02$) for *XPD Lys*⁷⁵¹*Gln* polymorphism. Smokers carrying the *Lys*⁷⁵¹*Gln* variant allele were associated with 2-fold increased risk of lung cancer ($p=0.01$). *XPD Lys*⁷⁵¹*Gln* was analyzed as the best one factor model using Multidimensionality reduction (MDR) analysis and was associated with increased risk for lung cancer ($p<0.0001$). Classification and regression tree (CART) analysis depicted *XPD Arg*¹⁵⁶*Arg* as the highest risk group associated towards lung cancer susceptibility. A subgroup analysis based on chemotherapeutic regimens suggested that patients administered Irinotecan–platinum, for *XPD Lys*⁷⁵¹*Gln* mutant genotype (CC) was significantly related to better OS as compared to wild type genotype (AA) (25.43 vs 5.63, Log rank $p=0.01$). The Cox model revealed a significant effect (HR=0.21, 95%CI=0.08-0.57, $p=0.002$), whereas a poor survival was reported for subjects who were carrying the mutant genotype (AA) for *XPD Asp*³¹²*Asn* polymorphism and treated with same regimen as above in comparison with wild type genotype (GG) (1.86 vs 7.28, Log rank $p=0.01$). Patients treated with Pemetrexed–platinum doublets and carrying the heterozygous genotype of *Asp*³¹²*Asn* polymorphism had a poor survival (HR=2.01, 95%CI=1.14-3.56, $p=0.01$). *XPG Phe*^{670L}*eu* was associated with protective effect in lung cancer and its histological subtypes. It was observed that 673 C>T polymorphism might be associated with increased lung cancer risk. 11985 A>G and *Arg*⁴¹⁵*Gln* polymorphism does not show any association with lung cancer risk. The genetic

variants of 673 C>T polymorphism were associated with a better survival in ADCC patients while 11985 A>G polymorphism was associated with poor survival in SCLC patients. Lung cancer patients treated with irinotecan cisplatin/irinotecan carboplatin regimen indicated their association with shorter survival time and high risk of death in 11985 A>G polymorphism. The inter-gene analysis showed that individuals carrying multiple copies of risk variants have decreased risk for lung cancer. MDR analysis suggested prevalence of gene-gene interaction as well as gene-environment interaction. *XPG Phe⁶⁷⁰Leu* was predicted as the highest risk factor. CART analysis did not confirm gene-gene interaction but gene-smoking interaction was observed for NER SNPs. The combinatorial studies suggested that there was no association observed between polymorphic combinations of NER genes with overall survival in lung cancer patients. When applied on histological subtypes, the polymorphic were associated with poor survival showing a hazardous effect in SCLC patients.

Conclusion

This preliminary study provides an insight into the potential of genetic differences in NER genes to modulate lung cancer susceptibility. *XPC Lys⁹³⁹Gln* and *XPB Lys⁷⁵¹Gln* were certainly modifying risk of lung cancer in North Indian population. *XPG Phe⁶⁷⁰Leu* also acted as an important risk modifier of lung cancer. *ERCC2* functional SNPs (*Lys⁷⁵¹Gln* and *Asp³¹²Asn*) independently affect the OS of patients treated with irinotecan cisplatin/ irinotecan carboplatin regimen and may act as an important predictive parameter in survival of lung cancer patients. This study in future can facilitate assessment of genetic variations of *XPB/ERCC2* and help patients to make therapeutic decisions for individualized therapy in advanced lung cancer patients.

List of Publications

International Publications

- Shweta Lawania, Navneet Singh, Digamber Behera and Siddharth Sharma. (2017). *XPC* Polymorphism and risk for lung cancer in North Indian patients treated with platinum based chemotherapy and its association with clinical outcomes. *Pathology and Oncology Research* 24(2); 353-366 (Impact factor-1.7).
- Shweta Lawania, Navneet Singh, Digamber Behera and Siddharth Sharma. (2017). *XPB* polymorphism towards lung cancer susceptibility, overall survival and clinical response. *Future Oncology* 13(22); 1945:1965 (Impact factor-2.1).
- Shweta Lawania, Siddharth Sharma, Navneet Singh and Digamber Behera. (2018). *XPF* polymorphism towards lung cancer susceptibility and survival in patients treated with platinum based chemotherapy. *Future oncology* 14(11); 1071-1089 (Impact factor-2.1).
- Shweta Lawania, Navneet Singh, Digamber Behera and Siddharth Sharma. (2018). Association of *XPA* polymorphisms towards lung cancer susceptibility and its predictive role for overall survival and clinical response to platinum based chemotherapy in North Indians. *Biochemical Genetics* 56(4); 375-396 (Impact factor-1.52).
- Shweta Lawania, Navneet Singh, Digamber Behera and Siddharth Sharma. (2018). *XPG* polymorphism and its association with lung cancer susceptibility, survival and clinical response in North Indian patients treated with platinum based doublet chemotherapy. *Future oncology* 15(2); <https://doi.org/10.2217/fon-2018-0408> (Impact factor-2.1).

Poster presentation in conferences

Shweta Lawania, Siddharth Sharma, Navneet Singh and Digamber Behera “*XPB* polymorphism and its association with lung cancer risk, Overall survival and response in North Indian Patients treated with platinum based chemotherapy” *2nd Indian Cancer Congress*, 8-12 November 2017, Bengaluru.

Acknowledgements

Writing thesis concludes to the final steps of my Ph.D journey. A decision which happened to me all of a sudden, accounts the search of my best abilities and the duration of 3.5 years contributed in my professional as well as personal growth as a person. This work cannot be completed without the support and encouragement of numerous well wishers which includes colleagues, family and friends. My area of work lies in the premises of the Department of Biotechnology, Thapar University which has become an eternal part of mine during my research period.

Firstly, I would like to express my sincere gratitude to my supervisor **Dr. Siddharth Sharma**, Associate Professor, for his guidance, motivation and immense support. His knowledge helped me create understanding about the area of cancer research which was quite new to me when I enrolled as a Ph.D student. He has been supportive throughout my Ph.D journey and advised me in my professional and personal front which helped me to grow as a researcher. I would also like to thank him for his valuable suggestions, constructive criticism and concrete suggestions during the execution of research work which helped me complete my work within a time frame. I would like to thank him for his constant encouragement and enthusiasm which kept me going throughout this journey.

I would also like to extend my special thanks to **Dr. Navneet Singh**, Additional Professor, and **Dr. Digambar Behera**, Head, Department of Pulmonary Medicine, PGIMER for their invaluable support and guidance. They provided us the opportunity to have access to the hospital and also encouraged research. They also helped to resolve issues and problems which were faced by us while carrying out work and also gave suggestions for improving my work whenever required.

I am also thankful to **Dr. Moushmi Ghosh**, Head, Department of Biotechnology, Thapar University, as she supported research based environment in the laboratories and also emphasized on resolving infrastructure based issues for smooth working of student. I thank my Doctoral committee members, **Dr. Vikas Handa**, **Dr. Manoj Barnwal** and **Dr. B. N. Chudasma** for their helpful suggestions and comments during my progress report presentations.

I would like to thank my lab mates Charu Bahl and Amrita Singh for their help during my Ph.D period and making it a learning experience for me. I also thank them for maintaining a positive approach towards the work. I would like to thank **Gurpreet Kaur, Noorpreet Dhanjal, Devendra Sillu, Chhavi Pahwa, Monika Bharti, Kritika Sharma** and **Tamanna Mahajan** for supporting me throughout my Ph.D and kept on encouraging me to keep going in every situation.

I am also indebted to my friends **Jyoti Badhouria, Anshul Goel** and **Apoorva Singh**, who always been a source of strength during my Ph.D phase. They have always given their crucial time in listen to my thoughts and provided me with better suggestions.

A family is always a person's backbone and I am blessed to have such an encouraging and loving family. They always stood by my side whenever I required in every aspect whether it is emotionally or financially. I therefore, dedicate my thesis to my parents **Dr. V.K Lawania** and **Mrs Sadhana Lawania**, without their blessings this would not have been possible. They always dreamed of me getting a doctorate and supported me in my thick and thin throughout my Ph.D period. They have always been a perennial source of encouragement in my life. I would also like to pay gratitude to my mother-in-law **Mrs Aruna Sharma**, who firmly supported my step of pursuing Ph.D and has been encouraging throughout the period. I would also like to thank my brother, **Gaurav Lawania** for providing his wise words whenever required.

Lastly I would like to thank the man in my life, my husband **Maj Somesh Sharma**, who agreed with the desires and aspirations of his wife and supported me to follow my dreams. I would also thank him for all the sacrifices he made and has always been a source of encouragement. He has been a pillar of strength for me. His immense support and love has kept me focused throughout the time. I feel privileged to have such loving people around me.

Thank you all for your encouragement!!!!

Shweta Lawania

Table of Contents

Title	Page No.
Abstract.....	iii
List of publications.....	vi
Table of contents.....	ix
List of figures.....	xx
List of Tables.....	xxiv
List of Notations.....	xxx
List of abbreviations.....	xxxii
Chapter 1 Introduction.....	1
1.1 Introduction.....	1
Chapter 2 Review of literature.....	8
2.1 Lung anatomy.....	8

2.2 Lung cancer.....	10
2.3 Classification of lung cancer.....	10
2.3.1 Adenocarcinoma.....	11
2.3.2 Squamous cell carcinoma.....	11
2.3.3 Large cell carcinoma.....	12
2.3.4 Small cell carcinoma.....	12
2.4 TNM staging.....	13
2.5 Epidemiology of lung cancer worldwide.....	17
2.6 Lung cancer in India.....	18
2.7 Risk Factor.....	19
2.7.1 Smoking.....	19
2.7.2 Passive smoking.....	20
2.7.3 Diet and alcohol.....	20
2.7.4 Occupational exposure.....	20
2.7.4.1 Abestos.....	20
2.7.4.2 Exposure to radon gas.....	21
2.7.4.3 Metal and mixed occupation exposure.....	21
2.7.4.4 PAH and Diesel exhaust and air pollution.....	21
2.8 Lung carcinogenesis.....	22
2.9 Metabolic activation and excretion of Smoke carcinogens.....	22
2.9.1 Molecular alterations in Lung carcinogenesis.....	23
2.9.2 Activation of Proto-oncogenes.....	23
2.9.3 Tumor suppressor genes.....	24
2.9.4 Escaping the programmed cell death.....	25
2.9.5 Diagnosis of Lung cancer.....	26

2.10 Lung cancer treatments.....	27
2.11 Platinum based Chemotherapy regimens.....	28
2.11.1 Docetecel cisplatin/ Doectacel carboplatin.....	28
2.11.1.1 Mechanism of action.....	29
2.11.2 Irinotecan cisplatin/ Irinotecan carboplatin.....	30
2.11.2.1 Mechanism of action.....	30
2.11.3 Pemetrexed cisplatin/Pemetrexed carboplatin.....	30
2.11.3.1 Mechanism of action.....	31
2.12 DNA damage.....	31
2.13 Genome stability.....	32
2.14 DNA repair.....	32
2.14.1 Nucleotide excision repair.....	33
2.15 Nucleotide excision repair in Humans.....	34
2.15.1 Transcription-coupled Nucleotide excision repair (TC-NER).....	35
2.15.2 Global genome Nucleotide excision repair (GG-NER).....	38
2.16 NER protein.....	40
2.16.1 XPC.....	40
2.16.2 TFIIH.....	41
2.16.3 XPA.....	42
2.16.4 RPA.....	43
2.16.5 XPG.....	43
2.16.6 ERCC1-XPF.....	44
2.17 Polymorphism in NER genes and cancer risk.....	44
2.18 <i>XPC</i> polymorphism.....	45
2.18.1 <i>XPC</i> polymorphism and their association with lung Cancer risk in Caucasian and Asian Population.....	46
2.18.2 <i>XPC</i> polymorphism and its association with other cancers.....	47
2.19 <i>XPA</i> polymorphism.....	48

2.19.1 <i>XPA</i> polymorphism and lung cancer risk in Caucasian and Asian population.....	48
2.19.2 <i>XPA</i> polymorphism and its association with other cancers.....	50
2.20 <i>XPB</i> polymorphism.....	50
2.20.1 <i>XPB</i> polymorphism and lung cancer risk in Caucasian and Asian population	51
2.20.2 <i>XPB</i> polymorphism and other cancer risk.....	52
2.21 <i>XPD</i> polymorphism.....	53
2.21.1 <i>XPD</i> polymorphism and lung cancer risk in Asian and Caucasian population	53
2.21.2 <i>XPD</i> polymorphism and other cancer risk.....	53
2.22 <i>XPG</i> polymorphism.....	55
2.22.1 <i>XPG</i> polymorphism and lung cancer risk in Asian and Caucasian population....	55
2.22.2 <i>XPG</i> polymorphism and other cancer risk	55
2.23 <i>NER</i> polymorphism and clinical data	56
2.23.1 Association of <i>XPD</i> polymorphism with overall survival and clinical outcomes.....	57
2.23.2 Association of <i>XPA</i> polymorphism with overall survival and clinical outcomes.....	58
2.23.3 Association of <i>XPB</i> polymorphism with overall survival and clinical outcomes...	59
2.21.4 Association of <i>XPD</i> polymorphism with overall survival and clinical outcomes...	60
2.21.5 Association of <i>XPG</i> polymorphism with overall survival and clinical outcomes...	61
Chapter 3 Gaps in the study.....	62
Chapter 4 Material and methods.....	64
4.1 Study Design.....	64

4.1.1 Chemotherapy regimen.....	66
4.1.2 Follow up and response assessment.....	66
4.1.3 Clinical response.....	66
4.2 Genomic DNA isolation.....	68
4.3 Qualitative and quantitative estimation of genomic DNA.....	69
4.4 Genotyping of polymorphic sites <i>XPC</i> , <i>XPA</i> , <i>XPB</i> , <i>XPG</i> and <i>XPF</i> gene.....	69
4.4.1 Genotyping of <i>XPC</i> polymorphism by PCR-RFLP.....	70
4.4.2 Genotyping for <i>XPA</i> variants.....	71
4.4.3 Genotyping of <i>XPB</i> variants.....	72
4.4.4 Genotyping of <i>XPG</i> and <i>XPF</i> variants.....	72
4.5 Statistical analysis.....	75
Chapter 5 Results.....	78
5.1 Genomic DNA isolation.....	80
5.2 PCR amplification.....	81
5.3 Genotyping of NER genes.....	81
5.3.1 Genotyping of <i>XPC</i> variants.....	81
5.3.2 Genotyping of <i>XPA</i> variants.....	82
5.3.3 Genotyping of <i>XPB</i> variants.....	83
5.3.4 Genotyping of <i>XPG</i> variants.....	84
5.3.5 Genotyping of <i>XPF</i> variants.....	85
5.4 Epidemiological and demographical statistics of North Indian lung cancer	
Patients.....	87
5.4.1 Regional distribution of lung cancer patients.....	87
5.4.2 Distribution of study subjects on the basis of smoking status.....	88

5.4.3 Distribution of subjects on the basis age.....	90
5.4.4 Distribution of lung cancer patients on the basis of TNM staging.....	91
5.4.5 Stratification of lung cancer patients on the basis of histology.....	92
5.4.6 Distribution of lung cancer patients on the basis of performance status.....	92
5.4.7 Distribution of lung cancer patients as per different chemotherapy regimen.....	93
5.5 Statistical analysis for NER genes.....	94
5.5.1 Demographic characteristics	94
5.5.2 Clinical characteristics of lung cancer patients evaluated for detect the prognostic relevance of polymorphism of different NER genes.....	95
5.5.3 <i>XPC</i> gene.....	98
5.5.3.1 Genetic and allelic distribution of <i>XPC</i> polymorphism.....	98
5.5.3.2 Genetic and allelic distribution on the basis of histological subtypes.....	100
5.5.3.3 Effect of smoking status on association of <i>XPC</i> polymorphism with lung cancer risk.....	102
5.5.3.4 <i>XPC</i> haplotypes and lung cancer risk.....	105
5.5.3.5 Association between lung cancer risk and different combinations of <i>XPC</i> genotypes.....	106
5.5.3.6 Relationship between <i>XPC</i> Genotypic variants and Clinico-pathological Characteristics.....	106
5.5.3.7 Association of <i>XPC</i> SNPs with overall survival of lung cancer patients.....	110
5.5.3.8 Association of <i>XPC</i> SNPs with overall survival in lung cancer histological Subtypes.....	111
5.5.3.9 <i>XPC</i> polymorphism and their association with overall survival in lung cancer patients treated with different chemotherapy regimens.....	114
5.5.3.10 Association of <i>XPC</i> polymorphism with clinical response.....	118

5.5.4 <i>XPA</i>	119
5.5.4.1 Association of <i>XPA</i> polymorphism with lung cancer susceptibility.....	119
5.5.4.2 Interactive effect of smoking on the <i>XPA</i> polymorphism and their association with lung cancer risk.....	122
5.5.4.3 Haplotype analysis of <i>XPA</i> variants.....	125
5.5.4.4 Association between lung cancer risk and different combinations of <i>XPA</i> genotypes.....	126
5.5.4.5 Relationship of Clinico-pathological features with <i>XPA</i> polymorphic variants.....	127
5.5.4.6 Association of <i>XPA</i> SNPs with overall survival of lung cancer patients.....	130
5.5.4.7 Association of <i>XPA</i> SNPs with overall survival in lung cancer histological subtypes.....	130
5.5.4.8 Association of <i>XPA</i> SNPs and Overall survival of lung cancer patients treated with different chemotherapeutic regimen.....	135
5.5.4.9 Association of <i>XPA</i> polymorphism with treatment response for different chemotherapeutic regimens and also their correlation with objective response.....	138
5.5.5 <i>XPB</i> gene.....	139
5.5.5.1 <i>XPB</i> polymorphic variants and their association with lung cancer susceptibility.....	139
5.5.5.2 Effect of smoking on the <i>XPB</i> polymorphism and their association with lung cancer risk.....	143
5.5.5.3 Haplotype analysis of <i>XPB</i> variants.....	146
5.5.5.4 Combinatorial effect of polymorphic variants of <i>XPB</i> on lung cancer risk.....	147
5.5.5.5 Relationship of Clinico-pathological features with <i>XPB</i> polymorphic variants.....	148
5.5.5.6 MDR analysis.....	152

5.5.5.7 CART analysis.....	153
5.5.5.8 Association of <i>XPD</i> SNPs with overall survival of lung cancer patients.....	155
5.5.5.9 Association of <i>XPD</i> SNPs with overall survival in lung cancer histological Subtypes.....	155
5.5.5.10 Association of <i>XPD</i> SNPs and Overall survival of lung cancer patients treated with different chemotherapeutic regimen	162
5.5.5.11 Association of <i>XPD</i> polymorphism and clinical outcomes.....	167
5.5.6 <i>XPG</i>	168
5.5.6.1 <i>XPG</i> variants and its association with lung cancer susceptibility.....	168
5.5.6.2 Interactive effect of smoking on the <i>XPG</i> polymorphism.....	172
5.5.6.3 Haplotype analysis.....	175
5.5.6.4 Combined effect of <i>XPG</i> SNPs on lung cancer risk.....	175
5.5.6.5 Relationship of Clinico-pathological features with <i>XPG</i> polymorphic variants.....	176
5.5.6.6 Association of <i>XPG</i> polymorphic variants and Overall survival (OS) of lung cancer patients.....	179
5.5.6.7 Association of <i>XPG</i> polymorphic variants and Overall survival (OS) in lung cancer subtypes.....	179
5.5.6.8 <i>XPG</i> polymorphic variants and their association with Overall survival in lung cancer patients treated with different chemotherapy regimens.....	183
5.5.6.9 Association of <i>XPG</i> polymorphism with clinical response in lung cancer patients...	188
5.5.7 <i>XPF</i>	189
5.5.7.1 <i>XPF</i> polymorphism and their association with lung cancer risk.....	189
5.5.7.2 Effect of smoking on <i>XPF</i> polymorphism.....	192
5.5.7.3 Haplotype analysis of <i>XPF</i> variants.....	195

5.5.7.4 Combined effect of <i>XPF</i> SNPs on lung cancer risk.....	195
5.5.7.5 Relationship of Clinico-pathological features with <i>XPF</i> polymorphic variants.....	197
5.5.7.6 MDR analysis applied on <i>XPF</i> variants.....	201
5.5.7.7 CART analysis applied on <i>XPF</i> variants.....	201
5.5.7.8 <i>XPF</i> polymorphism and their association with overall survival in lung cancer patients.....	204
5.5.7.9 Association of <i>XPF</i> SNPs with overall survival in lung cancer histological subtypes.....	205
5.5.7.10 <i>XPF</i> polymorphic variants and their association with Overall survival in lung cancer patients treated with different chemotherapy regimens.....	209
5.5.7.11 Association of <i>XPF</i> polymorphism and also their correlation with objective Response.....	215
5.6 High order Gene-gene interaction.....	216
5.6.1 Distribution of combination of genotypes on different polymorphism and their association with lung cancer risk.....	216
5.6.2 Distribution of combination of genotypes on different polymorphism and their association with histological subtypes.....	223
5.6.2.1 Association of combination genotype on different polymorphism and their association with SQCC.....	223
5.6.2.2 Association of combination genotype on different polymorphism and their association with ADCC.....	228
5.6.2.3 Association of combination genotype on different polymorphism and their association with SCLC.....	230
5.6.3 MDR analysis and lung cancer risk.....	231
5.6.4 MDR and association with histological subtypes of lung cancer.....	232

5.6.5 MDR analysis and role of gene-environment interaction in lung cancer risk....	234
5.6.6 CART analysis.....	235
5.6.7 CART analysis and its association with Smoking parameter.....	239
5.6.8 Association of NER genes polymorphic combinations and Overall Survival of Lung cancer patients.....	240
5.6.9 Association of NER gene polymorphic combinations with overall survival according to histological subtype of Lung cancer patients.....	242
5.6.10 Association of different unfavorable genotypic combinations of NER gene with overall survival of Lung cancer patients.....	249
Chapter 6 Discussion.....	256
6.1 <i>XPC</i> polymorphism and its relationship with lung cancer risk and overall survival	256
6.2 <i>XPC</i> polymorphism and its relationship with lung cancer risk and overall survival	259
6.3 <i>XPB</i> polymorphisms and its role in lung cancer risk and overall survival	262
6.4 <i>XPD</i> polymorphisms and their association with lung cancer pre-disposition and Overall survival	266
6.5 <i>XPF</i> polymorphisms and their association lung cancer susceptibility and overall survival	269
6.6 Gene-gene interaction and gene-environment interaction.....	272
6.7 MDR and CART analysis in genetic variants of NER.....	273
6.8 Gene-gene interaction and Overall survival.....	275
Chapter 7 Conclusion.....	277

7.1 Future prospects of the study.....	279
References.....	280
Appendix.....	320
A.1 Ethics Committee Report.....	320
A.2 Patient Information Sheet.....	321
A.3 III.....	325

List of Figures

Figure No.	Title	Page No.
2.1	Schematic representation of Lung tissue; Anatomical features of lung showing trachea, bronchioles, alveoli and diaphragm	9
2.2	Microscopic images of different histological subtypes of lung cancer; a: Adenocarcinoma; b: squamous cell carcinoma; c: Large Cell Carcinoma; d: Small Cell Lung Carcinoma (Source: Weiss, 2005)	14
2.3	Schematic representation of major molecular abnormalities taking place during lung cancer pathogenesis. (Source: Panov, 2005)	27
2.4	Simplified description of main DNA lesions, cause behind these lesions and DNA repair pathways implemented on them (Source: Fleck, 2004; Houtgraaf <i>et al</i> , 2006)	33
2.5	Nucleotide excision repair (NER) depicting mode of action of its two sub-pathways: (1) Transcription coupled repair (2) Global genome repair (Fuso <i>et al</i> , 2006)	35
2.6	Transcription-coupled repair (Source: Fousteri <i>et al.</i> , 2006; Sarasin and Sary, 2007)	37
2.7	Global genome repair in Mammals (Source: Sarasin and Sary, 2007)	34
2.8	Structure of XPC protein derived from the yeast (<i>Saccharomyces cerevisiae</i>) Rad4 homolog. (Source: Clement <i>et al.</i> , 2010)	41
2.9	Structure of XPA protein (Source: Coin <i>et al.</i> , 2008)	42
2.10	Diagrammatic representation of XPG protein and its domains (Source: Zhu <i>et al.</i> , 2012)	44
2.11	NER pathway works in repair of DNA damage. Unrepaired lesion or malfunctioning of proteins associated with NER could lead to unregulated growth and cancer (Zhang <i>et al</i> , 2009)	45
4.1	STRAD diagram explains the selection criteria of the patients for	67

	overall survival analysis	
5.1	Agarose gel electrophoresis representing isolated genomic DNA from blood sample in (A) Cases (B) Controls	80
5.2	Representative example of agarose gel electrophoresis showing PCR amplification for <i>XPG Asp¹¹⁰⁴His</i>	81
5.3	Restriction digestion pattern of <i>XPC</i> variants	82
5.4	Restriction digestion pattern of <i>XPA</i> variants	83
5.5	Restriction digestion pattern of <i>XPB</i> variants	84
5.6	Restriction digestion pattern of <i>XPG variants</i>	85
5.7	Restriction digestion pattern of <i>XPF</i> variants	86
5.8	Regional distribution of the North Indian lung cancer patients	87
5.9	Regional distribution of Lung cancer on the basis of gender (a) Males (b) Females	88
5.10	Representative distribution of smoking status in the study population (a) cases (b) controls	89
5.11	Graph showing distribution of male population according to smoking status	90
5.12	Age wise gender distribution of lung cancer patients	90
5.13	Distribution of Lung cancer patients according to clinical stage in North Indians	91
5.14	Distribution of lung cancer cases as per their histological subtype	92
5.15	Pie chart illustrating number of cancer patients reported with different Karnofsky performance status	93
5.16	Pie chart illustrating number of cancer patients reported with ECOG performance status	94
5.17	Distribution of lung cancer patients according to their treatment regimen	95
5.18	Kaplan Meier graphs representing relationship of <i>XPC Ala⁴⁹⁹Val</i> polymorphism with OS in (A) ADCC patients with heterozygous (CT) genotype (B) ADCC patients with CT+TT genotype (C) patients treated with docetaxel cis/carboplatin regimen for heterozygous (CT)	117

	genotype	
5.19	Kaplan Meier curve depicting association of A23G polymorphism with overall survival in (A) ADCC patients with mutant genotype (GG) (B) SCLC patients with mutant genotype (GG)	134
5.20	(A) Interaction dendrogram gained from the multifactor dimensionality reduction (MDR) showing gene-gene interaction for <i>XPD Lys⁷⁵¹Gln</i> , <i>XPD Asp³¹²Asn</i> and <i>XPD Arg¹⁵⁶Arg</i> (B) Classification and regression tree (CART) analysis depicting high order interaction among three polymorphisms (<i>Lys⁷⁵¹Gln</i> , <i>Asp³¹²Asn</i> and <i>Arg¹⁵⁶Arg</i>) of <i>XPD</i>	154
5.21	Kaplan Meier curves illustrating association between overall survival in patients with (A) Mutant genotype (CC) for SCLC showing <i>XPD Lys⁷⁵¹Gln</i> polymorphism (B) Mutant genotype (AA) for SCLC and (C) Heterozygous genotype (GA) depicting <i>XPD Asp³¹²Asn</i> polymorphism	161
5.22	Kaplan Meier curves illustrating association between overall survival in lung cancer patients with (A) mutant genotype (CC) treated with irinotecan cisplatin/carboplatin regimen showing <i>XPD Lys⁷⁵¹Gln</i> polymorphism (B) mutant genotype (AA) treated with irinotecan cisplatin/ carboplatin regimen depicting <i>XPD Asp³¹²Gln</i> polymorphism (C) Heterozygous genotype (GA) treated with pemetrexed cisplatin/carboplatin regimen showing <i>XPD Asp³¹²Asn</i> polymorphism	166
5.23	Kaplan Meier curve depicting association between (A) <i>Phe⁶⁷⁰Leu</i> polymorphism and patients having mutant genotype (AA) in ADCC patients (B) <i>Asp¹¹⁰⁴His</i> polymorphism and patients having mutant genotype (CC) in SCLC patients (C) <i>Phe⁶⁷⁰Leu</i> polymorphism and patients having mutant genotype (AA) treated with irinotecan-cisplatin/carboplatin regimen (D) <i>Asp¹¹⁰⁴His</i> polymorphism and patients having combined genotype (GC+CC) treated with pemetrexed-cisplatin/ carboplatin regimen	186-187
5.24	(A) Interaction dendrogram gained from the multifactor dimensionality	203

	reduction (MDR) showing gene-gene interaction for <i>XPF</i> 673 C>T, 11985 A>G and G415A (B) Classification and regression tree (CART) analysis depicting high order interaction among three polymorphisms of <i>XPF</i>	
5.25	Kaplan Meier curves illustrating association between overall survival in patients with (a) Heterozygous genotype (CT) for ADCC showing <i>XPF</i> 673C>T polymorphism (b) Heterozygous genotype (AG) for SCLC showing 11985 A>G polymorphism and (c) Heterozygous genotype (GA) for ADCC in G415A polymorphism (d) Kaplan Meier curves illustrating association between overall survival in lung cancer patients with heterozygous genotype (AG) treated with irinotecan/cisplatin/carboplatin regimen showing 11985 A>G polymorphism.	213-214
5.26	MDR depicting the association between the all the NER genes involved in the present study with lung cancer risk (A) a 4-factor model illustrating association with lung cancer risk (B) a 3-factor model depicting ADCC risk (C) a 4-factor model representing SQCC risk (D) a 6-factor model showing SCLC risk	234
5.27	MDR graph representing association of gene-environmental factor in NER SNPs	235
5.28	Classification and Regression Tree analysis showing high order NER gene-gene interactions in SCLC patients	238
5.29	CART tree explaining overall NER SNP interaction on smoking	239
5.30	Recursive partitioning method (STREE) explaining association of NER SNPs with overall survival in lung cancer patients	253
5.31	Kaplan-Meier graph explaining overall survival in unfavorable genotypic combinations of NER SNPs (a) Terminal node 1-4 when compared with reference node (node 9) (b) Terminal node 5-8 when compared with reference node (9).	254

List of tables

Table No.	Tables	Page No.
2.1	TNM stage groupings (Source: Weissferdt and Moran, 2012)	15
2.2	Tumor, node, and metastasis (TNM) definitions (Source: Weissferdt and Moran, 2012)	16
5.1	Demographic characteristics among cases and control patients in North Indian population	96
5.2	Patients characteristics and Clinical parameters	97
5.3	Association of XPC polymorphism and its association with lung cancer susceptibility	99
5.4	Distribution of genotypes of XPC polymorphism depicting their association with Lung cancer risk and its histological subtypes	101
5.5	Distribution of genotypes of XPC polymorphism according to their smoking status and its association with Lung Cancer risk	104
5.6	XPC Haplotypes and their association with lung cancer risk	105
5.7	Combination of genotypes of XPC polymorphic variants and their relationship with the risk of lung cancer	106
5.8	Relationship of Clinico-pathological features with XPC Lys ⁹³⁹ Gln and XPC Ala ⁴⁹⁹ Val genotypes in lung cancer patients	109
5.9	Association of XPC polymorphism and overall survival analysis in lung cancer patients	110
5.10 (a)	Association of XPC polymorphism and overall survival analysis in ADCC patients	112
5.10 (b)	Association of XPC polymorphism and overall survival analysis in SQCC patients	113
5.10 (c)	Association of XPC polymorphism and overall survival analysis in SCLC patients	113
5.11	Association of XPC polymorphism and overall survival analysis according to regimen	116
5.12	Genotypic Response of XPC polymorphism to platinum based	118

	chemotherapy and their association with lung cancer risk	
5.13	Distribution of genotypes of <i>XPA</i> polymorphism depicting their association with Lung cancer risk and its histological subtypes	121
5.14	Distribution of genotypes of <i>XPA</i> polymorphism according to their smoking status and its association with Lung Cancer risk	124
5.15	<i>XPA</i> Haplotypes and their association with lung cancer risk	125
5.16	Combination of genotypes of <i>XPA</i> polymorphic variants and their relationship with the risk of lung cancer	126
5.17	Relationship of Clinic-pathological features with <i>XPA</i> A23G and <i>XPA</i> G709A genotypes in lung cancer patients	129
5.18	Association of <i>XPA</i> polymorphism and overall survival analysis	131
5.19 (a)	Association of Overall survival and lung cancer risk in ADCC patients	131
5.19 (b)	Association of Overall survival and lung cancer risk in SQCC patients	132
5.19 (c)	Association of Overall survival and lung cancer risk in SCLC patients	132
5.20	Association of <i>XPA</i> polymorphism and overall survival analysis according to regimen	137
5.21	Genotypic Response of <i>XPA</i> polymorphism to platinum based chemotherapy and their association with lung cancer risk	139
5.22(a)	Distribution of genotypes of <i>XPB</i> polymorphism depicting their association with lung cancer risk	141
5.22(b)	Distribution of genotypes of <i>XPB</i> polymorphism depicting their association with lung cancer subtypes	142
5.23	Distribution of genotypes of <i>XPB</i> polymorphism according to their smoking status and its association with Lung Cancer risk	145
5.24	<i>XPB</i> Haplotypes and their association with lung cancer risk	146
5.25	Combination of genotypes of <i>XPB</i> polymorphic variants and their relationship with the risk of lung cancer	148
5.26	Distribution of Clinico-pathological features and their association	151

	with lung cancer risk in patients with <i>XPD</i> polymorphism	
5.27	Multifactor Dimensionality reduction (MDR) analysis showing interactions of <i>XPD</i> variants with lung cancer risk	152
5.28	Risk estimates based on CART analysis	153
5.29	Association of <i>XPD</i> polymorphism and overall survival analysis in lung cancer patients	157
5.30 (a)	Association of <i>XPD</i> polymorphism and overall survival analysis in ADCC patients	158
5.30 (b)	Association of <i>XPD</i> polymorphism and overall survival analysis in SQCC patients	159
5.30 (c)	Association of <i>XPD</i> polymorphism and overall survival analysis in SCLC patients	160
5.31	Association of <i>XPD</i> polymorphism and overall survival according to regimen	164
5.32	Genotypic Response of <i>XPD</i> polymorphism to platinum based chemotherapy and their association with lung cancer risk	168
5.33(a)	Distribution of genotypes of <i>XPG</i> polymorphism depicting their association with lung cancer risk	170
5.33	Distribution of genotypes of <i>XPG</i> polymorphism depicting their association with lung cancer subtypes	171
5.34	Distribution of genotypes of <i>XPG</i> polymorphism according to their smoking status and its association with lung Cancer risk	174
5.35	<i>XPG</i> Haplotypes and their association with lung cancer risk	175
5.36	Combination of genotypes of <i>XPG</i> polymorphic variants and their relationship with the risk of lung cancer	176
5.37	Relationship of Clinico-pathological features with <i>XPG</i> <i>Phe</i> ⁶⁷⁰ <i>Leu</i> and <i>XPG</i> <i>Asp</i> ¹¹⁰⁴ <i>His</i> genotypes in lung cancer patients	178
5.38	Association of <i>XPG</i> polymorphism and overall survival in Lung cancer risk	181
5.39 (a)	Association of <i>XPG</i> polymorphism and overall survival analysis in ADCC patients	181

5.39 (b)	Association of <i>XPG</i> polymorphism and overall survival analysis in SQCC patients	182
5.39 (c)	Association of <i>XPG</i> polymorphism and overall survival analysis in SCLC patients	182
5.40	Association of <i>XPG</i> polymorphism and overall survival according to regimen	185
5.41	Genotypic Response of <i>XPG</i> polymorphism to platinum based chemotherapy and their association with lung cancer risk	189
5.42	Distribution of genotypes of <i>XPF</i> polymorphism depicting their association with Lung cancer risk and its histological subtypes	191
5.43	Distribution of genotypes of <i>XPF</i> polymorphism according to their smoking status and its association with Lung Cancer risk	194
5.44	<i>XPD</i> Haplotypes and their association with lung cancer risk	195
5.45	Combination of genotypes of <i>XPF</i> polymorphic variants and their relationship with the risk of lung cancer	197
5.45	Relationship of Clinico-pathological features with <i>XPF</i> genotypes in lung cancer patients	200
5.46	Multifactor Dimensionality reduction (MDR) analysis showing interactions of <i>XPF</i> variants with lung cancer risk	201
5.47	Risk estimates for <i>XPF</i> variants based on CART analysis	202
5.48	Association of <i>XPF</i> polymorphism and overall survival analysis in lung cancer patients	203
5.49 (a)	Association of <i>XPF</i> polymorphism and overall survival analysis in ADCC patients	207
5.49 (b)	Association of <i>XPF</i> polymorphism and overall survival analysis in SQCC patients	207
5.49 (c)	Association of <i>XPF</i> polymorphism and overall survival analysis in SCLC patients	208
5.50	Association of <i>XPF</i> polymorphism and overall survival according to regimen	211

5.51	Genotypic Response of <i>XPF</i> polymorphism to platinum based chemotherapy and their association with lung cancer risk	216
5.52	Distribution of genotypes depicting dual combination and their association with lung cancer risk	220
5.53 (a)	Distribution of genotypes depicting dual combination and their association with SQCC patients	225
5.53 (b)	Distribution of genotypes depicting dual combination and their association with ADCC patients	227
5.53 (c)	Distribution of genotypes depicting dual combination and their association with SCLC patients	228
5.54	Distribution of genotypes depicting triple combination and their association with lung cancer risk	231
5.55	Multifactor Dimensionality reduction (MDR) analysis showing interactions of NER variants with lung cancer risk	233
5.56	Multifactor Dimensionality reduction (MDR) analysis showing gene interaction in NER variants with ADCC	233
5.57	Multifactor Dimensionality reduction (MDR) analysis showing gene interaction in NER variants with SQCC	233
5.58	Multifactor Dimensionality reduction (MDR) analysis showing gene interaction in NER variants with SCLC	233
5.59	Multifactor Dimensionality reduction (MDR) analysis showing gene-environment interaction in NER variants with lung cancer risk	235
5.60	Risk estimates based on CART analysis for all NER variants	237
5.61	Risk estimates based on CART analysis for all NER variants in ADCC patients	237
5.62	Risk estimates based on CART analysis for all NER variants in SQCC patients	237
5.63	Risk estimates based on CART analysis for all NER variants in SCLC patients	237
5.64	Risk estimates based on CART analysis among smokers for all NER variants	239

5.65	Association of polymorphic combinations with overall survival in lung cancer patients	241
5.66	Association of polymorphic combinations NER genes with overall survival in ADCC patients	245
5.67	Association of polymorphic combinations NER genes with overall survival in SCLC patients	247
5.68	Association of polymorphic combinations NER genes with overall survival in SQCC patients	248
5.69	Overall survival in NER SNPs by applying STREE	251

List of Notations

A	Adenine
α	Alpha
R	Arginine
N	Asparagine
D	Aspartic acid
B	Aspartic acid or Asparagine
bp	Base pair
β	Beta
cm	Centimetre
χ^2	Chi-square
D'	Coefficient of linkage disequilibrium
C	Cytosine
°	Degree
°C	Degree Celsius
df	Degree of Freedom
$\leq$	Equals to or greater than
r^2	Extent of linkage disequilibrium
γ	Gamma
E	Glutamic Acid
Q	Glutamine
Z	Glutamine or glutamic acid
G	Gaunine
g	Gram
H	Histidine
h	Hour
HCl	Hydrochloride
I	Isoleucine
<	Less than
L	Leucine

l	Litre
K	Lysine
M	Methionine
µg	Microgram
µl	Microliter
µm	Micrometer
µM	Micromolar
mg	Milligram
mg/m ²	Milligram per square meter
mm	Millimetre
min	Minute
M	Molar
ng/µl	Nanograms per microlitre
%	Percentage
F	Phenylalanine
±	Plus Minus
P	Proline
s	Second
S	Serine
AgNO ₃	Silver nitrate
NaCl	Sodium Chloride
T	Thymidine
Y	Tyrosine
W	Tryptophan
V	Valine
v	Volume

List of Abbreviations

ADCC	Adenocarcinoma
AMP	Adenosine monophosphate
Arg	Arginine
Asp	Asparagine
β-TrCp	Beta transducin containing protein
BSA	Bovine Serum Albumin
CAK	Cdk activating kinase
CART	Classification And Regression Tree Analysis
CPD	Cyclobutane pyrimide dimer
CR	Complete Response
CS	Cockayne Syndrome
CSA	Cockayne syndrome complementation group A
CSB	Cockayne syndrome complementation group B
CSC's	Cancer stem cells
CVC	Cross validation consistency
Cys	Cysteine
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribose Nucleoside Triphosphates
DSBR	Double Strand Break Repair
EDTA	Ethylene diamine tetra acetic acid
EMA	Epithelial Membrane Antigen
EGFR	Epidermal growth factor receptor
ERCC	Excision Repair Cross Complementing
FDR	False discovery rate
GG-NER	Global genome nucleotide excision repair
H	Heterozygote
hHR23B	(human)Homologue B of yeast RAD23 protein
His	Histidine

HR ¹	Adjusted hazards ratio
HR	Hazard ratio
HWE	Hardy-Weinberg equilibrium
IASLC	International Association for the Study of Lung Cancer
ICMR	Indian Council Of Medical Research
Ile	Isoleucine
LD	Linkage disequilibrium
M	Mutant
MAF	Minor Allele Frequency
MDR	Multifactor Dimensionality Reduction
m-RNA	Messenger Ribonucleic Acid
MST	Median survival time
NER	Nucleotide Excision Repair
NSCLC	Non-Small Cell Lung Cancer
OS	Overall Survival
OR	Odds ratio
PAGE	Polyacrylamide Gel Electrophoresis
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PD	Progressive Disease
PGIMER	Postgraduate Institute of Medical Education and Research
PR	Partial Response
Pro	Proline
RBC	Red Blood Cells
RECIST	Response Evaluation Criteria In Solid Tumors
RFLP	Restriction Fragment Length Polymorphism
RPA	Replication Protein A
RPC	Replication Protein C
RNA Pol II	RNA polymerase II
SCF	Stem cell factor
SCLC	Small Cell Lung Cancer

SD	Stable Disease
SER	Serine
SNP	Single Nucleotide Polymorphism
SQCC	Squamous Cell Carcinoma
TC-NER	Transcription coupled nucleotide excision repair
TEMED	N, N, N', N' Tetramethylethylene-1, 2-Diamine
TFIIH	Transcription factor IIH
Thr	Threonine
Trp	Tryptophan
TNM	Tumour Node And Metastasis
AOR	Adjusted Odds ratio
VEGF	Vascular Endothelial Growth Factor
W	Wild
WBC	White Blood Cells
WHO	World Health Organization
XP	Xeroderma Pigmentosum
XPA	Xeroderma Pigmentosum Complementation Group A
XPC	Xeroderma Pigmentosum Complementation Group C
XPB	Xeroderma Pigmentosum Complementation Group D
XPG	Xeroderma Pigmentosum Complementation Group G
XPF	Xeroderma Pigmentosum Complementation Group F

CHAPTER-1

Introduction

Lung cancer is a malignant tumor growth which has been characterized by the uncontrolled growth of cells in the tissue of the lung. Primarily in the beginning of the century, lung cancer was considered as the rare disease but now it has reached to an epidemic level (Noronha *et al.*, 2016). According to GLOBOCAN survey, there has been 1.8 million lung cancer cases (contributing 12.7% of world population) reported every year worldwide making it the most commonly diagnosed form of malignancy. This survey further accounts that 1.38 million deaths occur due to lung cancer all over the world. In India, lung cancer constitutes **6.9% of all new cancer cases** and 9.3% of all cancer related deaths in both the genders; it is the commonest cause of cancer related mortality in men [GLOBOCAN, 2012]. The overall 5-year survival rate of lung cancer is dismal with approximately 15 per cent in developed countries and 5 per cent in developing countries. Lung cancer has been majorly classified in terms of non small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) **which accounts 85% and 15% of the total world population** respectively. In Western countries and most of the Asian countries adenocarcinoma has surpassed squamous cell carcinoma as the most common form of lung cancer (Malik *et al.*, 2015). Most of the previous Indian studies have described squamous cell carcinoma as the most commonly found histology, however, a recent study represented adenocarcinoma as the most common subtype reported in Indian population (Noronha *et al.*, 2016). There is also increasing incidence of lung cancer in females and non smokers (Malik *et al.*, 2015). This pattern of lung cancer incidence differs as per geographic region and ethnicity which largely reflect the prevalence and pattern of smoking. It is **documented** that tobacco smoke is the major contributor and risk factor for lung cancer and 90% of all cases is due to smoking (Behera, 2012). However, not all smokers develop lung cancer and it has been estimated that fewer than 20% of smokers develop the disease (Carlsten *et al.*, 2008).

Lung cancer is the consequence of the **environmental factors as well as genetic pre-disposition** attributed due to carcinogenic exposure. To overcome the damaging effects of the pernicious agents, alteration in the activity of the tumor-suppressor genes and proto-oncogenes occur. Cell machinery on activation checks for the cell check points in cell cycle which halts the cell progression or rectifies the damage. If the damage is detected and correctly repaired by the DNA repair pathways, the cell cycle further progresses, otherwise resulting into oncogenesis. In this whole process, DNA repair system regarded as the safeguard for human genome preventing constant impertinence implicated by the carcinogenic and mutagenic agents such as cigarette/tobacco smoke, environmental exposure, oxidizing agents, ROS agents, alkylating agents, ionizing radiations, occupational hazards *etc* (Sellers *et al.*, 1992).

Exogenous factors like smoking, ultraviolet rays, chemical mutagens, air pollution are important risk factors which contribute in generation of reactive oxygen species (ROS). On the other hand, endogenous factor like oxidation, methylation and deamination produce reactive oxygen species (ROS) inside the cell. It acts as a risk factor when an imbalance occurs between biochemical antioxidants and ROS (Fong *et al.*, 2003). Oxidative stress occurs in cancerous cells when compared with normal cells therefore it might be related to oncogenic stimulus (Przybyłowska *et al.*, 2013). ROS developed in this process is responsible for oxidative damage and development of cytotoxic DNA lesions such as 8-hydroxoguanine (8-oxo-G) and 7, 8-dihydro-8oxo-2-deoxyguanosine (Przybyłowska *et al.*, 2013). These lesions are cytotoxic, mutagenic in nature contributing towards neuro-degeneration, ageing and cancer. The exogenous factor such as tobacco smoke has high content of aromatic amines *e.g.* 2- Naphthylamine, 4-aminobiphenyl, N-nitrosamines, similarly heavy fuel combustion produces elevated levels of Polycyclic aromatic hydrocarbon (PAH) *e.g.* B[α]P in the environment (Gelboin, 1980). These carcinogens when inhaled are bio-activated by xenobiotic metabolism of cells. During this process, electrophiles and free radicals are generated (MacLeod and Tang, 1985). PAH form DNA adducts which are present in respiratory epithelium and act as an important factor for induction of lung carcinogenesis (Wiencke *et al.*, 1995). Evolutionary conserved DNA repair pathways have evolved over the time in cells which actively remove the DNA lesions created by these mutagens (Houtgraaf *et al.*, 2006). Therefore, lung cancer susceptibility is modulated by the damage removal rate which is genetically determined by DNA repair capacity (DRC).

Nucleotide excision repair (NER) is one such multi-step process which involves thirty different proteins that orchestrates together for the removal of damaged DNA by dual incision. It is a “cut” and “patch” mechanism essential for the elimination of bulky DNA adducts (deLaat *et al.*, 1999). This repair pathway has also been essential for the repair of the UV-induced lesion such as CPD (cyclobutane primidine dimer) and 6-4 PP (pyrimidine (6-4) pyrimidone photoproduct) (Morita *et al.*, 2010). The NER pathway has been further divided into two part sub pathways: the pathway which takes part in removal of those lesions which at one point blocks on going transcription known as transcription-coupled nucleotide excision repair (TC-NER), or the mechanism for removal of lesion present at any position in genome known as global genome nucleotide excision repair (GG-NER) (Morita *et al.*, 2010). NER is present in every organism and its general mechanism consists of five steps: (1) Recognition of DNA damage (2) Unwinding DNA helix and verification of lesion at both ends (3) double incision at the both ends of the lesion (4) excision of the oligonucleotide and resynthesis of DNA fragment (5) Ligation of residual nick. *XPC*, *XPD*, *XPF*, *XPG* and *XPA* are some important genes of NER pathway (Hoeijmakers, 2001).

Xeroderma pigmentosum complementation group C gene (*XPC*) is an important gene which encodes a 940 amino-acid protein involved in initial DNA damage recognition in global genome nucleotide excision repair (GG-NER) (Matutani *et al.*, 1997). *XPC*-RAD23B recognizes a wide variety of lesions that thermodynamically destabilizes DNA duplexes including exogenous agents such as benzyl-[α]-pyrene. ***XPC* polymorphisms cause amino acid substitution in the functional domain which alters the DNA repair capacity and its susceptibility towards the lung cancer.** Two common polymorphic sites of the *XPC* gene have been investigated in different forms of cancer: A-to-C transition in exon 15 resulting in a lysine-to-glutamine transition at position 939 (*Lys*⁹³⁹*Gln*; rs2228001), located in the interaction domain with TFIIH and the second one involves a C-to-T transition in exon 8 which results in the substitution of alanine for valine at codon 499 (*Ala*⁴⁹⁹*Val*; rs2228000) in the interaction domain of *XPC* interacts with hHRAD23 protein. These sites are suggested to be in linkage disequilibrium with each other (Khan *et al.*, 2000).

Xeroderma pigmentosum group A (*XPA*) is another important gene which maps at chromosome 9 at 9q22.3 having 12 introns and 6 exons is a 273 amino acid zinc finger DNA-binding protein associated with excision of wide range of lesions such as benzo[α]pyrene induced DNA bulky

adducts (Wakasugi and Sancar 1999). XPA protein recognizes the DNA damage *via* interaction with replication protein A (Wu *et al.*, 2003), transcription factor IIH and excision repair cross complementing group 1- xeroderma pigmentosum group F (Batty and Wood 2000). *XPA* gene is part of both the sub pathways of nucleotide excision repair *i.e.* global genome and transcription coupled pathways (Volker *et al.* 2001), therefore, it interacts with many repair factors which are part of DNA repair process (Laat *et al.* 1999). ***XPA* gene with A23G variant has potential to influence the protein function which might subsequently affect the DNA repair capacity (DRC).** As *XPA* gene interacts with many repair factors, therefore, it shows a strong affinity towards the damaged DNA, it is likely to play central role in correct positioning of repair machinery around the tumor lesion (Jelonek *et al.* 2010). A→G polymorphism is present at position 23 in the 5' non-coding region with four nucleotides present in the upstream of the start codon of the gene (Butkiewicz *et al.*, 2000). Studies on A23G polymorphism showed association of the protective effect for the G-allele against the lung cancer risk and homozygous carriers of G-allele shows 13% higher DNA repair capacity than A-allele carriers in Asian population (Wu *et al.*, 2003). **Although other studies on Caucasians (Butkiewicz *et al.*, 2004) East Asian (Wu *et al.*, 2003) and American African (Wu *et al.*, 2003) population differ from these studies suggesting association of variant genotype with increased risk for lung cancer.** The studies conducted on the association of *XPA* A23G polymorphism with cancer are reported with contradictorily different results (Butkiewicz *et al.*, 2004, Hirschhorn *et al.*, 2002). The tumor sites or different ethnicities might be the possible reason for such discrepancy (Hirschhorn *et al.*, 2002). The other known polymorphism of *XPA* is codon 228 present in exon 6 associated with *Arg228Gln* (G709A, rs1805160) substitution (Butkiewicz *et al.*, 2000). It is a missense mutation of *XPA* which resulted out of DNA repair defects (Porter *et al.*, 2005). *XPA* G709A polymorphism reported to be associated with improved NER kinetics (Porter *et al.*, 2005) although any epidemiological study has not yet suggested its role in cancer. **The functional significance of these polymorphism are still unknown.**

The third gene which is closely studied is Xeroderma pigmentosum complementation group D (*XPD*) also known as Excision repair cross complementation group 2 (*ERCC2*). It is an ATP-dependent DNA helicase essential for transcription and NER activity (Hoejimaerker *et al.*, 1996). *XPD* protein encodes 760 amino acids with the polypeptide size of 87KDa. *XPD Arg¹⁵⁶Arg* (rs238406) polymorphic variant is located in exon 6 and represents a silent polymorphism

leading to C→A transition, resulting in an Arg→Arg substitution at codon 156 (Benhamou & Sarasain, 2002). Another SNP of the *XPB* gene is present in codon 312 (rs1799793) of exon 10, resulting into G→A substitution, which changes aspartate (Asp) to asparagine (Asn) (Shen *et al.*, 1998). The third SNP present on codon 751 (rs13181) and located at exon 23 of *XPB* gene is characterized by A→C substitution resulting in a change of lysine (Lys) to glutamine (Gln) (Shen *et al.*, 1998). Various epidemiological studies have suggested association of *XPB* polymorphism with different types of cancer such as head and neck (Strugis *et al.*, 1998), melanoma (Tomescu *et al.*, 2001), basal cell carcinoma (Vogel *et al.*, 2001), lung (Butkiewicz *et al.*, 2001) and bladder cancer (Matullo *et al.*, 2001) although the results reported were inconsistent.

ERCC5 or *XPG* gene (Xeroderma pigmentosum complementation group G) maps at chromosome 13q32-33 and encodes 1186 amino acid is the structure-specific endonuclease which plays an essential role in two incision steps in NER pathway (Wakasugi *et al.*, 1997). The protein shows extensive homology with two other nucleases *i.e.* RAD2 (a yeast DNA repair protein) and bacteriophage T4 RNase H. These enzymes cleave single to double strand DNA at junctions forming a flap or bubble-like structure (Shen *et al.*, 1996). *XPG* catalyzes incision approximately at five nucleotides from 3' site of the damage and is also involved non-enzymatically in subsequent 5' incision by *XPF/ERCC1* heterodimer during NER process (Cheng *et al.*, 1998). **Two polymorphic variants of *XPG* have been identified *i.e.* *Phe*⁶⁷⁰*Leu* and *Asp*¹¹⁰⁴*His* but the functional significance of these polymorphisms are still unknown.** *Phe*⁶⁷⁰*Leu* polymorphism is located on exon 9 leading to C→A transition (*C2207A*; rs2228959), thus resulting in phenylalanine (Phe) to Leucine (Leu) amino acid change (Kumar *et al.*, 2003). *Asp*¹¹⁰⁴*His* is present at exon 15 subsequently leading to G→C substitution (*G3507C*; rs17655) which changes aspartate (Asp) to histidine (His) amino acid (Jeon *et al.*, 2003). **As *XPG* is a rare polymorphism, very few epidemiological studies have been reported till now** (Au *et al.*, 2007; Hussain *et al.*, 2009). Emmert *et al* (2001) stated association of *XPG Asp*¹¹⁰⁴*His* with increased breast cancer risk whereas Hussain *et al* (2009) reported the association of *XPG* exon15 polymorphism with reduced gastric cancer risk.

Another gene contributing to endonuclease activity during NER pathway is Xeroderma pigmentosum group F (*XPF*) which is also known as *ERCC4* (Excision repair cross

complementation group 4). It is located on chromosome 16p13.2 and consists of 11 exons which span approximately 28.2kb (Liu *et al.*, 1993). The main function of *XPF* gene is to make 5'-incision during NER process. *XPF* protein is composed of 916 amino acids and an ERCC4 domain which belong to nuclease family (Tsodikov *et al.*, 2005). ERCC1 binds to form complex with *XPF* which acts like a structure-specific endonuclease responsible for 5' primer incision during DNA excision repair (Ahmad *et al.*, 2008). This complex also plays an important role in DNA inter-strand crosslink as well as double strand breaks removal (Ahmed *et al.*, 2008). **There are 580 SNPs in the *XPF* gene reported till date but only few have shown susceptibility to cancers like breast, lung, endometrium and colorectal. *Arg*⁴¹⁵*Gln* (rs1800067) results in an arginine-to-glutamine transition at codon 415 and this transition affects protein interaction, hence hampers *XPF/ERCC1* complex activity during NER pathway and alters genetic susceptibility towards cancer** (Shen *et al.*, 1998). A cytosine-to-thymine change has been observed for *XPF* 673 *C>T* (rs316028) polymorphism, thus affecting the transcriptional activity of the gene. One of the least studied polymorphism is *XPF*11985A>G (rs254942) stating alanine-to-glycine transition. Few epidemiological studies have been reported on *XPF* variants till now. A study by Smith *et al* (1998) suggested association of *XPF Arg*⁴¹⁵*Gln* polymorphism towards increased risk for breast cancer as well as a study on laryngeal cancer documented its relationship of *XPF Arg*⁴¹⁵*Gln*, rs1800067 polymorphism (Krupa *et al.*, 2001).

The repair pathway genes discussed above for susceptibility were also looked for their influence on survival. *XPB* protein of NER pathway is critical in repairing damages implicated by damaging agents such as platinum based drugs such as cisplatin and carboplatin (Fukuda *et al.*, 2005). Studies have shown that *XPF* and *XPB* could bind to platinum containing DNA duplexes (Zhu *et al.*, 2009). Thus it has been proposed that the *XPB* protein is critical for repairing other types of DNA damage induced by the platinum-based anticancer drugs (such as cisplatin and carboplatin). Therefore it might be plausible that genetic variants in the *XPB* gene might modulate the DRC (Fukuda *et al.*, 2005) and hence influence the anticancer effect of platinum based drugs. Therefore, it is clear that these genes play role in the prognosis and survival of lung cancer patients.

Chemotherapy has been the major process for lung cancer treatment (Fleming *et al.*, 2012). From the available chemotherapy regimens, platinum based doublets (**When cisplatin/carboplatin is**

administered along with another chemotherapy drug which can be either paclitaxel, docetaxel, irinotecan, gemcitabine or vinorelbine) have been associated with improved overall survival (OS). Platinum based compounds such as cisplatin forms inter as well as intra DNA strands leading to bulky adduct formation resulting into double helix destabilization and hence inhibition of DNA replication (Lavanya *et al.*, 2015). DNA repair is a composite process and is carried out by series of DNA repair pathways such as NER. NER pathway is responsible for the repair of majority of platinum-DNA adducts followed by restoration of DNA segment. SNPs of NER genes modulate repair capacity which may contribute to individual's variations towards chemotherapeutic regimens (Shi *et al.*, 2015).

For lung cancer patients, platinum based doublet chemotherapy has been most commonly used treatment. However, the response rate amongst patients is only 20% (Jemal *et al.*, 2010). The patients having similar tumor stage and histological subtype show different chemotherapy response along with a different prognosis suggesting role of molecular marker in predicting the treatment sensitivity and clinical outcomes (Molina *et al.*, 2008). Many mechanisms play an important role in repair of DNA damage caused by these platinum compounds including NER pathway which may influence anticancer effect of these agents (Molina *et al.*, 2008). Few studies have emphasized on the association of various polymorphisms in NER pathway gene which may lead to alteration in the repair efficiency of DNA which may influence treatment response and cancer survival (Yuli *et al.*, 2013).

Therefore, the primary reason for evaluating these sequence variants of Nucleotide excision repair pathway genes (*XPC*, *XPA*, *XPB*, *XPD*, *XPG* and *XPF*) is that there are very few reports on Indian population which have evaluated such large number of repair genes together in lung cancer. The high risk posing subgroups within the sample population were identified by stratifying the subjects on the basis of their histology, clinic-pathological features and smoking habits. Haplotypic analysis was also performed for SNPs present on the same gene. Further, the study also assessed the high order gene-gene and gene-environment interactions using Multifactor dimensionality reduction (MDR) and Classification and regression tree analysis (CART). The SNPs of all the above mentioned NER genes and their association with overall survival and clinical response among lung cancer subjects has also been elaborated in this study.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Lung Anatomy

Respiratory system can be divide into two parts *i.e.* airways anatomy and lung anatomy. Airways anatomy consists of extra-thoracic and intra-thoracic airways. Lung anatomy includes lung parenchyma which is responsible for gas exchange at alveolar level. Lungs are the paired pyramid shaped air sacs situated in the upper thoracic cavity. Lungs are surrounded by the chest wall made up of ribs and muscles present between ribs. Lungs are separated *via* mediastinum where organs like heart and other organs are present (Nagaishi and Nagasawn, 1972). Diaphragm separates abdomen from respiratory cavity *via* thin muscle. Lungs are further extended laterally from heart to the ribs on the both sides of the chest and continue posterior towards the spine. Lung is cone shaped from the inferior end while concave from the basal region. Left lung is smaller than the right one as it accommodates two-third of the heart on the left region of the body as shown in Fig 2.1. Left lung comprises of two lobes separated by left oblique fissure. Right lung has three lobes which are separated through a horizontal fissure while right oblique fissure creates the partition between middle and lower lobe (OpenStax 2014).

Lungs are covered with the pleural cavity which is formed constituting visceral and parietal pleura. Visceral pleura is made up of simple squamous epithelium layer present on the inner side of the lungs. The upper region of visceral pleura is known as parietal pleura. Between the two layers lies a cavity known as pleural cavity which promotes expanding of lungs during inhalation (ElaineN. Marieb, 2012). A serous fluid is present between these two layers of pleural cavity which lubricates the region in order to avoid irritation while breathing (ElaineN. Marieb, 2012). Hilum is the point through which blood vessels, primary bronchus and nerves penetrate inside the lungs.

Nose is the main body part through which air enters and then further passes through pharynx, larynx and trachea sequentially. Pharynx has three parts “Nasopharynx”, “Oropharynx” and “Laryngopharynx”. Nasopharynx is lined with ciliated pseudostratified columnar epithelial tissue which is situated next to nasal cavity. Oropharynx is located next to Nasopharynx, it provides passage for food, air, drink (TerryDesJardins, 2008). It is lined with stratified squamous epithelial tissue. Laryngopharynx is part of pharynx connected to larynx via glottis. The main function of this part is that it helps in swallowing of food, drink and air. Larynx also known as voice box is made up of nine cartilages, of which eight are hyaline cartilage connective tissue and the ninth cartilage *i.e.* epiglottis is made up of elastic cartilage connective tissue. Epiglottis helps in swallowing of food by covering larynx through a flap so that entry of food in larynx can be prevented. Air then enters in Trachea (windpipe). The trachea is a membranous structure, made up of 18-20 C-shaped hyaline cartilage connective tissue (Tu.J:Inthavong, 2013). Posterior wall of the trachea is made up of smooth muscle which contributes in contraction and relaxation of trachea. Trachea further splits into right and left bronchi in the mediastinum. Bronchi are hollow tubes which are made up of hyaline cartilage like trachea and are lined with ciliated pseudo-stratified epithelium (Dalavanga, 2017). C-shaped rings are present inside the hyaline cartilage which is found in the interior of bronchi, so that they are prevented from collapsing and blocking airflow inside the lungs. Primary bronchi further splits into secondary and tertiary bronchi until it divides individually for each lobe (Fig 2.1). This division supports an even distribution of air flow in each lobe of the lung. Tertiary bronchi further splits into bronchioles and these bronchioles into terminal bronchioles which then gets divided into respiratory bronchioles and at the end terminating forming an alveolar duct (Empey, 1978). Alveolar ducts leads to cluster of alveoli called alveolar sacs where gaseous exchange occur (Dalavanga, 2017).

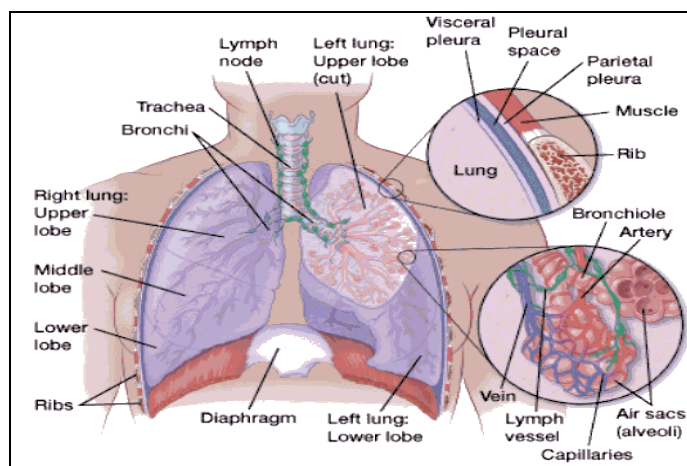


Fig 2.1 Schematic representation of Lung tissue; Anatomical features of lung showing trachea, bronchioles, alveoli and diaphragm

2.2 Lung Cancer

Lung cancer is the condition of aberrant cell behavior resulting in progressive pathological changes such as formation of malignant neoplastic lesions. The cause for cancer can be any error or mutation in cell's DNA (Loeb and Loeb, 2000). These mutations affect cells in a way that there occurs a disruption in the control over cell death and cell growth which results in growing of tumor as a mass. A tumor mass can either be “malignant” or “benign” (*Type of Tumors*, 2016). Benign tumors can be easily removed and do not spread to other body parts. Tumors which are able to invade other tissues and spread in other body parts when enters through blood or lymphatic system are malignant in nature (Alberg *et al.*, 2016). This process of spread of tumor to other sites is called metastasis. The most prominent sites for metastasis of lung tumor are adrenal glands, bones, liver and brain (Tamura *et al.*, 2015). The abnormal growth in different regions of the lung tissue leads to the formation of a solid tumor which derives all the blood supply and inherently impairs the proper functioning of the lung tissue by invading the adjacent cells and consequently other organs through lymph node infiltration. The lung tissue subsequently suffers from mucous formation, fibrosis and necrosis due to malignant properties of the tumor (Chang *et al.*, 2015). The major clinical implications which are evident in the condition of lung carcinoma include coughing, shortness of breath, change in voice, harsh sounds while breathing, bronchitis, loss of appetite, weakness, hemoptysis, dysphonia and asthenia. Lung cancers may occur in any part of the lung, but 90%-95% of cancers of the lung arise from the epithelial cells. Cell lines appear to cover the larger and smaller airways (bronchi and bronchioles) and because of this reason, lung cancers are sometimes called bronchogenic cancers or bronchogenic carcinomas. Cancers also can arise from the pleura called mesotheliomas or rarely from supporting tissues within the lungs, for example, the blood vessels (Tamura *et al.*, 2015).

2.3 Classification of lung cancer

Lung cancers are broadly classified into two types: small cell lung cancers (SCLC) and non-small cell lung cancers (NSCLC). This classification is based upon the microscopic appearance of the tumor cells, specifically the size of the cells. These two types of cancers grow and spread in different ways and may have different treatment options. SCLC comprises about 20% of lung cancers and is the most aggressive and rapidly growing of all lung cancers. SCLC is strongly

related to cigarette smoking and only 1% of these tumors occurring in nonsmokers. SCLC metastasizes rapidly to many sites within the body and are most often discovered after they have spread extensively. NSCLC are the most common lung cancers, accounting for about 80% of all lung cancers (Noronha *et al.*, 2016). NSCLC can be divided into several main types that are named based upon the type of cells found in the tumor: Adenocarcinoma (ADCC), Squamous cell carcinoma (SQCC) and large cell carcinoma (LCC).

2.3.1 Adenocarcinoma

The term adenocarcinoma is derived from words “adeno” which stands for pertaining to a gland and “carcinoma” which describes cancer that is developed in epithelial cells. They produce mucin which result into different growth patterns such as acinar, papillary and bronchiovascular. It is usually located around peripheral nodules than central or perihilar nodules. Morphology of cells and architecture of their cell clusters depends on the cytological identification of the adenocarcinoma. Cells are arranged in sheets or a three dimensional morulae, acini and papillae along with fibrovascular cores. Another typical feature is that the cellular borders of these cells are delineated. Single, eccentric and oval shaped nuclei are visible in these cells. Prominent, single macronucleoli which are round or irregular are evident as illustrated in fig 2.2(A). The invasiveness of this tumor is visualized by fibrosis, along with necrosis, cavitation and hemorrhagic area (Walter and Pryce, 1955). These cells express the epithelial markers such as AE1/AE3, CAM 5.2, epithelial membrane antigen (EMA) and carcinoembryonic antigen (CEA) (Colby, 1995). Differentiation between primary and metastatic ADCC can be performed with the help of immuno-phenotype for cytokeratin (CK) 7 and CK20. Adenocarcinoma can be differentiated into different grades on the basis of extent of variation of the architectural frame of tumor cells from the normal lung tissue (Travis *et al.*, 2015). The three grades include grade 1 as well differentiated, grade 2 as moderately and grade 3 as poorly differentiated (Travis and Harris, 2004). Nearly 40% of lung cancers are adenocarcinomas, which usually originates in peripheral lung tissue (Liu *et al.*, 2010).

2.3.2 Squamous cell carcinoma

It is a type of NSCLC which is prominently found among smokers. 90% of the smoking population reported to have SQCC. It is characterized by keratinization and intercellular connections

formed between bronchial epithelial cells (Funai *et al.*, 2003). Squamous cell carcinoma occurs in central part of lung (lobar), mainstem or segmental bronchi. It begins with a precancerous lesion known as squamous metaplasia or dysplasia located in bronchial epithelium. This tumor often shows irregular friable, grey white cut surface with a large area having central necrosis with or without cavity formation as indicated in fig 2.2(B). This tumor is locally aggressive whereas metastasis to distant organs is less frequent (Funai *et al.*, 2003). Histopathologically squamous cell carcinoma varies from well-differentiated squamous cell neoplasm with keratin formation to poorly differentiated intracellular bridges with minimal residual squamous cell features. SQCCs spread locally within chest directly making use of mediastinal structure. However, carcinomas which are poorly differentiated metastasize early to distant sites (Funai *et al.*, 2003). Histochemically they appear thyroid transcription factor (TTF1) negative and CK 5, 6 P63, P40 positive. Cytogenetics shows there is characteristic amplification of telomeric 3q regions along with a gain of 3q24 in majority of squamous cell carcinoma along with deletion on short arm of chromosome 3. 84% of SQCC are observed to be EGFR positive whereas HER2 and K-ras expression is rare in SQCC subtype.

2.3.3 Large cell carcinoma

These epithelial tumors are poorly differentiated and are malignant in nature. Only 9% of lung cancer patients are diagnosed with this subtype (Travis and Harris, 2004). It was previously known as large cell neuroendocrine carcinoma. This tumor detected at the periphery of lungs. These tumors have soft pink tan with frequent necrosis leading to occasional hemorrhage and leads to rare cavitation as indicated in figure 2.2(C). LCC consist of sheets or nest of large polygonal cells with vesicular nuclei and prominent nucleoli and moderate cytoplasm. LCC is an aneuploid neoplasm with highest mean number of chromosome and DNA content of all lung cancer types. LCC show K-ras mutations, p53 mutations and Rb alteration like in other NSCLCs [Takei *et al.*, 2002].

2.3.4 Small cell lung cancer

Small-cell lung carcinoma is an aggressive form of lung cancer that is strongly associated with cigarette smoking and has a tendency for early dissemination. Small-cell carcinoma of the lung

usually presents in the central airways and infiltrates the submucosa leading to narrowing of bronchial airways (Carter *et al.*, 2014). It is sometimes called "oat cell carcinoma" due to the flat cell shape and scanty cytoplasm. It is thought to originate from neuroendocrine cells [Amine Precursor Uptake and Decarboxylation (APUD) cells] in the bronchus called Feyrter cells (Carter *et al.*, 2014). They express a variety of neuroendocrine markers, and may lead to ectopic production of hormones like atypical ductal hyperplasia (ADH) and adrenocorticotrophic hormone (ACTH) that may result in paraneoplastic syndromes. Small-cell carcinoma is most often more rapidly and widely metastatic than non-small cell lung carcinoma (and hence staged differently). SCLC cells are white-tan soft friable peri-hilar masses which have a huge nodal involvement. They have a scant cytoplasm with ill-defined cell borders therefore they appear flat in shape and also sometimes also known as "oat cell carcinoma". They are finely granular chromatin and without or inconspicuous nucleoli. SCLC cell appears to be round, oval or spindle-shaped. Necrosis is extensively observed in them as shown in fig 2.2(D). SCLC are CD56 positive, have all neuroendocrine markers, transcriptional intermediary factor (TIF) positive (Nicholson *et al.*, 2002). SCLC cells are aneuploid. There is a chromosomal imbalance observed in maximum number of SCLC population giving large 3p deletions which are associated with 3q isochromosome formation. Chromosomal sub-regions are amplified characteristically in SCLC. SCLC cells usually result in brain metastasis as the diseases progresses. They show a high rate of p53 mutations. Myc amplification and methylation of caspase8, antiapoptotic gene are characteristics of SCLC. Over-expression of E2F1 and inactivation of retinoblastoma (RB) gene are unanimously found in SCLC (Nicholson *et al.*, 2002).

2.4 TNM staging

Tumor-node-metastasis (TNM) classification defines lung cancer staging in 7th edition of TNM staging system by International Association for the study of lung cancer (IASLC) (Mirsadraee *et al.*, 2012). This includes the size and the degree of loco-regional invasions by primary tumor (T), extent of lymph node involvement (N) and the absence or presence of distant metastases (M).

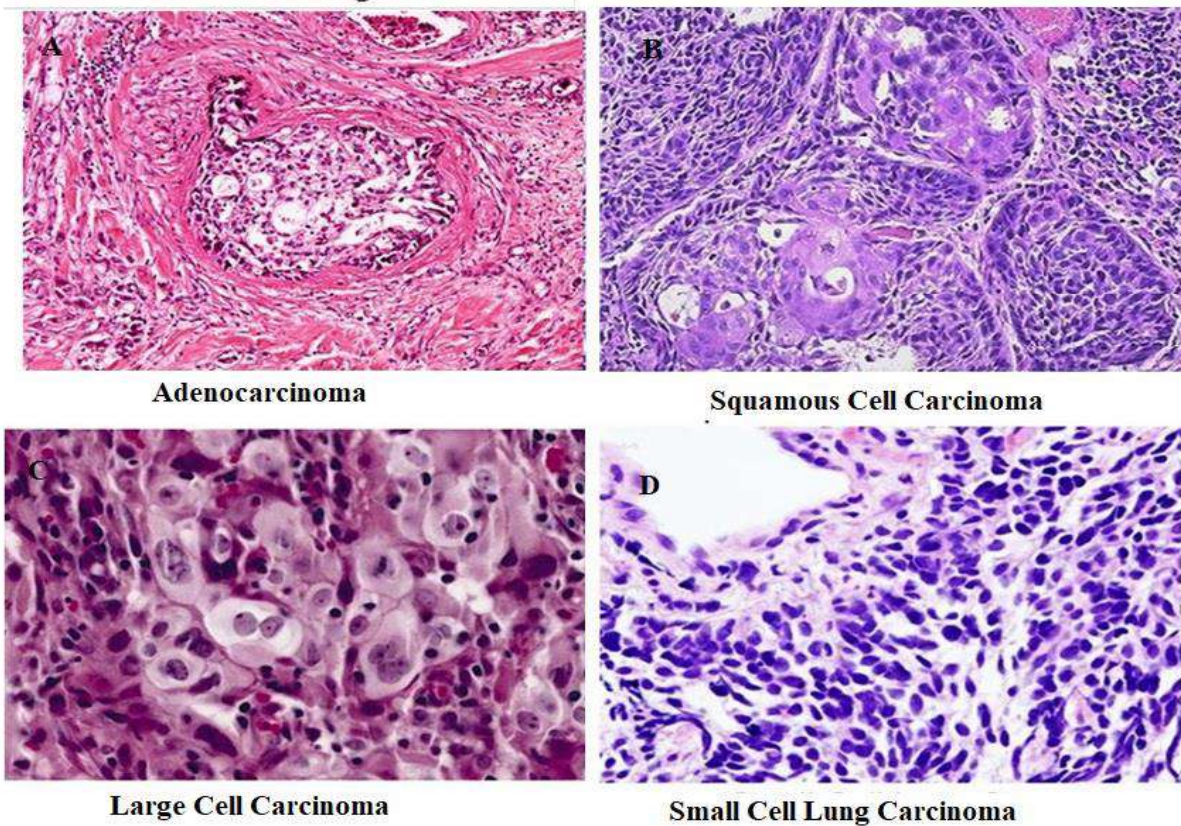


Fig 2.2: Microscopic images of different histological subtypes of lung cancer; a: Adenocarcinoma; **b:** Squamous cell carcinoma; **c:** Large Cell Carcinoma; **d:** Small Cell Lung Carcinoma (**Source:** Weiss, 2005)

This classification system helps clinicians in defining prognosis in lung cancer patients and evaluates treatment for them. Staging of lung cancer is a significant factor in determining the prognosis of the disease (Rami-Porta *et al.*, 2009). Stage of lung tumor can be determined by two ways pathological and clinical. Clinical stage is determined by result of imaging test, biopsies and physical test and pathological stag is determined after surgery in addition to the determination of clinical stage (Santos *et al.*, 2007). Classification of stage summarized in Table 2.1. Determination of the metastatic status remains the primary target of clinicians, since any distant metastasis indicates stage IV. The tumor descriptors are subdivided on the basis of size and location which help to better define tumor. Table 2.2 indicates the details of TNM classification and parameters used that help in staging.

Occult carcinoma	TX	N0	M0
Stage 0	Tis	N0	M0
Stage IA	T1a	N0	M0
	T1b	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0	M0
	T1a	N1	M0
	T1b	N1	M0
	T2a	N1	M0
Stage IIB Stage IIIA	T2b	N1	M0
	T3	N0	M0
	T1a	N2	M0
	T1b	N2	M0
	T2a	N2	M0
	T2b	N2	M0
	T3	N1	M0
	T3	N2	M0
	T4	N0	M0
	T4	N1	M0
Stage IIIB Stage IV	T1a	N3	M0
	T1b	N3	M0
	T2a	N3	M0
	T2b	N3	M0
	T3	N3	M0
	T4	N2	M0
	T4	N3	M0
	Any T	Any N	M1a
	Any T	Any N	M1b

Table 2.1: TNM stage groupings (Source: Weissferdt and Moran, 2012)

Table 2.2: Tumor, node, and metastasis (TNM) definitions (Source: Weissferdt and Moran, 2012)

<i>Primary tumor (T)</i>	
TX	Primary tumor cannot be assessed, or tumor proven by the presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy
T0	No evidence of primary tumor
Tis	Carcinoma in situ
T1	Tumor 3 cm or less in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the lobar bronchus (i.e., not in the main bronchus) ^a
T1a	Tumor 2 cm or less in greatest dimension
T1b	Tumor more than 2 cm but 3 cm or less in greatest dimension
T2	Tumor more than 3 cm but 7 cm or less or tumor with any of the following features (T2 tumors with these features are classified T2a, if 5 cm or less): involves main bronchus, 2 cm or more distal to the carina; invades visceral pleura (PL1 or PL2); associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung
T2a	Tumor more than 3 cm but 5 cm or less in greatest dimension
T2b	Tumor more than 5 cm but 7 cm or less in greatest dimension
T3	Tumor more than 7 cm or one that directly invades any of the following: parietal pleural (PL3), chest wall (including superior sulcus tumors), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium; or tumor in the main bronchus (less than 2 cm distal to the carina ^a but without involvement of the carina); or associated atelectasis or obstructive pneumonitis of the entire lung or separate tumor nodule(s) in the same lobe
T4	Tumor of any size that invades any of the following: mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina, separate tumor nodule(s) in a different ipsilateral lobe
<i>Regional lymph nodes (N)</i>	
NX	Regional lymph nodes cannot be assessed
N0	No regional node metastases
N1	Metastases to ipsilateral peri-bronchial and/or ipsilateral hilar node(s) and intrapulmonary nodes, including involvement by direct extension
N2	Metastases to ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastases in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular node(s)
<i>Distant metastasis (M)</i>	
M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumor nodule(s) in a contralateral lobe tumor with pleural nodules or malignant pleural (or pericardial) effusion
M1b	Distant metastasis (in extrathoracic organs)

2.5 Epidemiology of lung cancer world wide

Lung cancer regarded as the most common cancer in the global scenario. Annually, 1.61 million new cases of lung cancer were reported worldwide representing 1.38 million deaths in 2008 (Noronha, 2016). In 2015, there were 8.8 million deaths recorded due to lung cancer suggesting that out of 6 reported cancer-related deaths worldwide one death is recorded due to lung cancer (WHO, 2018). It has been estimated that there were 222,500 new cancer cases reported in the year 2017 (WHO, 2018). A report by the American cancer society has published certain important facts about the incidence and mortality occurring due to lung cancer. It illustrated that lung cancer is the second most common solid tumor in both the gender after prostate cancer in men and breast cancer in women. American cancer society in its survey for the year 2016 elaborated that 14% of patients diagnosed with lung cancer are men (117,920) and 13% are women (106,470). The number of deaths due to lung cancer estimated by American cancer society in 2017 in United States may reach upto 155, 870 cases. Also, the estimated incidence in 2017 in US will rise to 222,500 cases [Key Statistics for Lung Cancer, 2017].

As per Globacon survey, North America ranks first with the highest incidence rate of lung cancer while Africa stands at the last position. As per the regional history, Hungary marks the highest lung cancer risk with 51.6% rate per 100,000 (Ferlay et al., 2015). Australia ranks fifth as per the region where maximum patients diagnosed with lung cancer for the year 2013. It is estimated that the incidence rate would remain the same with increase in number of patients for 2017 with 12,434 new cases compared to 11,174 cases in 2013 (Lung cancer statistics, 2015). Cancer Research association reported lung cancer as the third most reported cancer in UK for the year 2014. 46,403 new lung cancer cases suggesting 35,895 deaths were reported. A survival rate of 5% for lung cancer in 10 years was observed. There were 410,000 new cancer cases diagnosed in the year 2012 and 35900 deaths by the end of the year 2014 (Lung cancer statistics UK, 2015).

A higher mortality rate in males (19,600 deaths) was observed as compared to female (16,300 deaths) for the year 2014. Geographically, the maximum proportions of males having lung cancer are reported in Europe, North America and South America (Müller-Hermelink *et al.*, 2004). Other countries showing a moderate rate of lung cancer incidence include China, Japan and South East Asia. Southern Asia (Pakistan and India) and sub-Saharan Africa have shown the

lowest rate of lung cancer incidence among men (Parkin *et al.*, 2002). This pattern among females mainly showcases the historical scheme of tobacco smoking around the globe. The rate of lung cancer occurrence can be easily correlated with the changing pattern of smoking habits in females among the developed and developing nations. North America and North West Europe exhibit the highest rate of lung cancer incidence followed by Australia, New Zealand and China (Parkin *et al.*, 2002).

2.6 Lung cancer in India

In the early 19th century, the incidence of lung cancer was rare in India. The history of the early data published related to lung cancer epidemiology on Indian population reflects a strong impact of industrialization and smoking trends. During the start of 1950s, 27.4 per million cases were reported and by 1959 it increased to 78.6 per million all around the world (Behera and Balamugesh, 2004). In 2012, Overall 1.8 million lung cancer cases were reported world-wide. Of these cases, lung cancer cases in India constitute 6.9 per cent of all new cancer cases and 9.3 per cent of all cancer related deaths in both sexes (Behera, 2012). It is the most common cancer and cause of cancer related mortality in men. The highest reported incidences were from Mizoram in both males and females. (Age adjusted rate 28.3 and 28.7 per 100,000 population in males and females, respectively) (Malik and Raina, 2015). The time trends of lung cancer show a significant rise in Delhi, Chennai and Bengaluru in both sexes. The contribution of lung cancer found maximum in 11 registered states which varied from 21.5% in Sikkim policy centre of biomedical research (PBCR) to 48.2% in Manipur state. Similarly, in case of females, the contribution of lung cancer was found to be maximum in 8 registries and the prevalence varied from 24.1% in Chennai PBCR to 72.2% in Manipur PBCR. However, it has been concluded that the incidence rate of lung cancer in females is much lower approximately 3% [*Three-Year Report of Population Based Cancer Registries 2012-2014*, 2016].

A report by Indian Council of medical research (ICMR) also accounts that nearly 17.3 lakh new cancer cases and 8.8 lakh deaths due to cancer will be registered by 2020 ranking India second on the list. Indian epidemiological surveys report of ICMR for the year 2016 indicated a rise in 1.14 lakh new lung cancer cases, in which 83,000 were males and 31,000 were females [*New*

cancer cases in India by 2020, 2015]. Globocon 2012, revealed there were 63,759 deaths reported in India due to lung cancer making it the third largest reason for cancer death. Furthermore, the mortality was higher in males with 48,697 deaths compared to 15,062 dead females (Noronha *et al.*, 2016). Over a time of 24 years (1982-2005), ICMR has observed lung cancer cases per one lakh has increased by around 100% in Bengaluru, 40% in Delhi and 160% in Chennai whereas a reduction of 60% was observed in Mumbai (LUNG CANCER Focus: INDIA, 2007).

2.7 Risk Factor

2.7.1 Smoking

Tobacco smoke is considered as the major risk factor leading to lung cancer. Smokers are more prone to lung cancer risk than non-smokers. The severity of the risk increases with the duration of smoking which are counted in terms of pack years. It has been suggested worldwide that 90% of male lung cancer deaths and 75-80% female lung cancer deaths are due to smoking (Ankathil, 2010). Many epidemiological studies have supported the idea that tobacco smoke has carcinogenic effects (Fleiss and Gross, 1991; Sasco *et al.*, 2004; Mishra *et al.*, 2012). Squamous cell carcinoma subtype is majorly found in smokers. People who were ex-smokers have 19% higher probability of having lung cancer as compared to non-smokers. But when the comparison is set between ex-smokers and smoker, the chances of lung cancer risk in ex-smokers are less as compared to smokers (Mishra *et al.*, 2012). Smoke from burning of tobacco contain nearly 1010 particles where 95% of it consists of gases chiefly nitrogen, oxygen, carbon dioxide, carbon monoxide, water, volatile sulphur compounds, acetaldehyde *etc.* (Hecht, 1999). There are nearly 3500 carcinogenic compounds in particulate phase of smoke. IARC (International Agency for Research on Cancer) has evaluated 55 carcinogens till date. Major particulate substances present are nicotine (0.2-0.6% of the weight of tobacco smoke), catechols (1%), polycyclic aromatic hydrocarbons (PAHs) (0.00002-0.00007%), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) (Hoffmann *et al.*, 2001). In Indian subcontinent, Bidis are used which consist of tobacco flakes rolled in a leaf of tendu plant. They are the major form of tobacco consumption along with hookah and cigarettes. 73% of smoking population in India consumes bidi in compared to cigarettes (27%). Bidis have a higher content of nicotine than cigarettes (21.2 vs 13.5mg/g)

(Mishra *et al.*, 2012). Smoking index of an individual can be calculates the cumulative risk due to smoking. Smoking status is quantified in terms of pack years. Pack years is defined as the product of average number of cigarettes smoked per day and the total duration of smoking in years (Jindal *et al.*,1988)

2.7.2 Passive Smoking

Passive smoking is also considered equally harmful as there are reports where non-smokers show high risk of lung cancer (McCarthy *et al.*, 2012). Passive smoking or the inhalation of tobacco smoke by non smokers who share living or working quarters with smokers, also is an established risk factor for the development of lung cancer (McCarthy *et al.*, 2012). Research has shown that nonsmokers who reside with a smoker have a 24% increase in risk for developing lung cancer when compared with nonsmokers who do not reside with a smoker. Never smokers have reported 10-15% of all lung cancer cases (Noronha *et al.*, 2016). Side-stream smoke is reported to have carcinogens such as aromatic amines, PAH, Nitrosamines. Side stream has four times higher B[α]P than in main stream of smoke. Non-smokers exposed to second hand smoke are observed to have DNA adducts from tobacco carcinogens in urine samples (de Groot and Munden, 2012).

2.7.3 Diet and alcohol

The changes in food habits with time have caused damage to our health. The dietary habits have changed over the time from home made food to packaged food with heavy spices and oils. A diet rich in vegetables, fruits, cruciferous vegetables have iso-thiocyanates present in them, which are responsible for reducing lung cancer risk (Lam *et al.*, 2009). Smokers which consume higher alcohol consumption are considered at lung cancer risk (Malhotra *et al.*, 2016).

2.7.4 Occupational exposure

Different occupation hazards are also responsible for lung cancer risk:

2.7.4.1 Asbestos

Asbestos fibers are silicate fibers that can persist for a lifetime in lung tissue. They are present in the workplace areas such as factory work, insulation, carpentry, demolition. They are widely used in the past as both thermal and acoustic insulation. Asbestos fibre become lodged in the

lining of the lung and is observed to be responsible for mesothelioma type of lung cancer (Berman and Crump, 2008). Cigarette smoking drastically increases the chance of developing an asbestos-related lung cancer in workers exposed to asbestos. Asbestos workers who do not smoke have a fivefold greater risk of developing lung cancer than nonsmokers (Reddy *et al.*, 2004).

2.7.4.2 Exposure to radon gas

Radon gas is a natural radioactive gas which is a natural decay product of uranium that emits a type of ionizing radiation. Radon gas is a known cause of lung cancer resulting in 12% of lung cancer deaths world-wide [Global Adult Tobacco Survey (GATS) India, 2009-2010]. About 21,000 lung-cancer-related deaths are annually reported in U.S. due to radon gas. It is the second leading cause of lung cancer after smoking. Radon gas can travel up through soil and enter homes through gaps in the foundation, pipes, drains, or other openings. Radon gas attacks the lung lining and produce radiations which can cause ultimately lung cancer (Gawe lek *et al.*, 2017).

2.7.4.3 Metals and mixed occupation exposures

People working on Chromium (IV) compounds, nickel miners have reported an increased risk of lung cancer (Hayes, 1997). High risk of lung cancer is also observed in people exposed to arsenic which may come along in drinking water. People who are in occupation related to silica such as poultry, brick making, stone cutting, earth mining, have also reported an increased risk of lung cancer, with evidence of exposure-response relationship in the high-exposure range (Steenland *et al.*, 2001).

2.7.4.4 PAH and Diesel exhaust and air pollution

Combustion of organic material produces polycyclic aromatic hydrocarbons (PAH). Aluminium, coal gasification, coke production, roofing, tar distillation produces PAH and an increased risk of lung cancer has been reported in people exposed to high PAH (Rota *et al.*, 2014). Motor vehicle engine exhaust represents a high risk causative factor for lung cancer (Rota *et al.*, 2014). Coal burning in non-ventilated houses, working on chullahs, fumes from high temperature cooking

using unrefined vegetables oil is also one of the reasons for increasing lung cancer rate among women (Melville and Eastwood, 1998).

2.8 Lung carcinogenesis

Lung cancer initially occurs when the epithelial cells lining the respiratory tract are exposed to carcinogens in smoke over a period of time (Cruz *et al.*, 2011). Tobacco smoke carcinogens are first metabolically activated and then sequentially they are excreted from the body (Hecht, 1999). During this process DNA-adducts formation occur (Phillips, 2002). Repair system of cells removes the DNA adducts and renders the normal DNA. If DNA damages persist for longer duration, this leads to the mutations (Cruz *et al.*, 2011). Oncogenes and Tumor suppressor genes if become the target of these mutations, they tend to lead towards unregulated growth and ultimately carcinogenesis (Cruz *et al.*, 2011).

2.9 Metabolic activation and excretion of Smoke carcinogens

The carcinogens present in smoke are substrates for drug metabolic system. The enzymes that act upon on smoke carcinogens are cytochromes P450 (Phase I), Glutathione-S transferases and UDP-glucuronosyl transferases (Phase II) (Hecht, 1999). They sequentially act on the carcinogenic substrates and ultimately excrete them out of the body. The first step is activation of carcinogenic compounds which is done by cytochrome P450 (Hecht, 2012). The CYPs performs either of N- and O-dealkylation, aliphatic and aromatic hydroxylation, N-and S-oxidation or deamination (Taxak and Bharatam, 2014). The activated carcinogens formed during the process can actively bind with DNA and form DNA adducts. The by-products of Phase I when polar they are directly excreted out of the body, however otherwise Phase I is followed by conjugation of a polar compound to the Phase I byproduct. The Phase II conjugation is performed by "transferase" enzymes" (Xue *et al.*, 2014). The process is done by glutathione S-transferases (GSTs), UDP-glucuronosyl transferases (UGTs), sulfatases (STs), arylamine N-acetyl transferases (NATs) and methyltransferases which convert activated carcinogens in harmless more hydrophilic compounds that are easily excreted out of the body (Nishikawa *et al.*, 2004). The GSTs perform cytosolic dimerization, UGTs do glucuronidation, STs are responsible for sulfation of aliphatic alcohols and phenols. Phase II detoxification is followed by Phase III where membrane transporters come into activity. ABC transporters are important genes that form Phase

III (Omiecinski *et al.*, 2010). A balance between the activation and detoxification of carcinogens must exist. There occurs an inter-individual difference in the balance between activation and detoxification of carcinogens that ultimately affects the cancer susceptibility. Certain carcinogens directly forms adducts without any activation (Cohen and Arnold, 2010).

2.9.1 Molecular alterations in Lung carcinogenesis

Molecular alterations as DNA adducts were observed in respiratory epithelium (Kadara *et al.*, 2016). These DNA-adducts are either repaired or may lead to apoptosis of cell (Roos and Kaina, 2006). In case cell is neither repaired nor apoptosized genetic instability occurs through acquired genetic and epigenetic mechanisms in cells leading to mutational events and ultimately cancer. Whole chromosome gain or loss, DNA segment amplification or deletion, gene silencing through methylation all these methods leads to genetic alteration. These alterations are evident in karyotyping and molecular profiles of a tumor. Mutations giving rise to lung carcinogenesis follow a sequence. Molecular changes with allelic loss in 3p regions such as 3p21, 3p22-24 *etc.* and 9p21 and to some extent in 8p21-23, 13q14 and 17p13 (Hecht, 2012). A certain sequence is observed for allelic loss 3p-8p-9p (Wistuba *et al.*, 2002). Pathogenesis of lung cancer results after continuous accumulation of genetic alteration over the time which involves activation of oncogenes, inactivation of certain tumor suppressor genes, allelic losses, loss of heterozygosity (LOH), chromosomal imbalance.

2.9.2 Activation of Proto-oncogenes

Tumor cells show production of growth factors for autocrine and paracrine cell growth. Cells that produce growth factors and express its receptors show autocrine growth. Paracrine growth occurs when growth factors stimulate growth in adjacent cells. ERBB family is a group of receptor tyrosine kinases which is responsible for growth in case of NSCLCs (Kirschbaum and Yarden, 2000).

Epidermal growth factor receptor (EGFR) is responsible for events such as cell division, motility, neo-vascularization by ligand activated cascade of auto-phosphorylation of mediators and finally effector molecules. Signal transduction stimulated by EGFR occurs through

PIK3/AKT/mTOR, RAS/RAF/MAPK and JAK/STAT pathways (Cohen and Arnold, 2010). Tyrosine kinase is constitutively expressed due to mutations occurring in EGFR gene. Mutated EGFR is considered responsible for oncogenic differentiation, proliferation, neovascularisation, invasion, metastasis *etc.* Studies have reported that NSCLC subjects harbor most common mutations in four exons namely 18-21. Mutation in exon 19 is the most common one which is frame deletion mutation. Second most common mutation is in exon 21, this is a missense substitution (Kosaka *et al.*, 2004). EGFR mutations are mostly reported in adenocarcinoma subjects (Cooper *et al.*, 2013). Tyrosine kinases inhibitors are extensively explored for their role in lung cancer since tyrosine kinases (growth related protein kinases) act as a therapeutic target for cancer. Similarly, ATP-binding domain of tyrosine kinases behaves as a target for drug development (Thomas *et al.*, 2012). Gefitinib and erlotinib are two TKIs which are observed to have improved survival in lung cancer subjects (Ansari *et al.*, 2009). HER2/neu (Human epidermal growth factor 2) and c-myc. HER2/neu is another member of ERBB family. It is a transmembrane glycoprotein of 185 kDa. It is expressed in NSCLCs and majorly in adenocarcinomas (Fong *et al.*, 2003). Its exon 20 shows gene activation by mutation. It was first observed that c-myc was over-expressed in 80-90% SCLCs (Rom *et al.*, 2000).

The position of these phospho-proteins is nucleus and it belongs to transcription factors family. It further interacts with the histone deacetylase which directly controls the chromatin conformation responsible for affecting gene transcription (Forgacs *et al.*, 2001). KRAS is an important member of RAS family of proto-oncogene and encodes an important G-protein. It plays an important role in signaling pathways leading to differentiation, survival, and cell proliferation. In normal state Ras is inactive and GDP (guanosine diphosphate) bound (Forgacs *et al.*, 2001). It gets activated when GDP switches to GTP and activates upstream growth factor receptors. The activated Ras-GTP consequently binds and activates MAPK, RAS/RAF/MEK/MAPK (Goodsell, 1999). The above signal cascade plays a role in lung cancer with a mutation in K-RAS (Downward, 2003). Following K-ras mutation lung epithelial cells can transform by activating ERK-MAP pathways (Massion and Carbone, 2003). Mutated K-ras was observed in 15-20% NSCLC in adenocarcinomas. Most common mutation in K-ras occurs in codon 12, 13 and 61 (Riely *et al.*, 2008). K-ras mutation is reported in smokers with G-T transversion.

2.9.3 Inactivation of Tumor suppressor genes

(TSGs) or recessive oncogenes inactivated in the process include Tp53 (17p13), RB (13q14), and p16INK4a (9p21). There are certain new tumor suppressor genes at short arm of chromosome 3 3p21(DTT1), 3p14.2(FHIT), 3p21(RASFF1), 3p22-24 (BAP-1) (Wistuba *et al.*, 2002). Inactivation of recessive oncogenes is a two step process, primarily point mutation occurs in first allele followed by inactivation of second allele may be by translocation or may be promoter methylation or chromosomal deletion (Wistuba *et al.*, 2002). *Tp53* is located on 17p13 chromosome and encodes for a protein of 53kDa and behaves as a transcription factor (Rivlin *et al.*, 2011). Carcinogenic load or damaged DNA induces p53 to dys-regulate cell-cycle and induces apoptosis and ultimately tumor formation (Zilfou and Lowe, 2009). Mutations in TSG p53 are most commonly observed in SQCC and SCLC subtype of lung cancer (Wistuba *et al.*, 2002). p53 (17p13) is likely to be used as a biomarker as over two-third of lung cancer report mutated p53. NSCLCs have reported *Tp53* mutations to be associated with a smoking history of patients (Sholl, 2016). There are reports where mutations in *Tp53* is correlated with resistance towards lung cancer treatment. Next important tumor suppressor gene involved in lung carcinogenesis is Rb(13q14). Rb cooperates with p53 at transcriptional level, cell cycle progression and to maintain equilibrium between cell differentiation and proliferation. Non-sense mutations, chromosomal deletions are normal gene behaviour observed in lung cancer (Forgacs *et al.*, 2001). It is observed to be over-expressed in 90% SCLC and 20% NSCLC cases (Sholl, 2016). Rb acts as a control for cell to go from G1 to S phase of cell cycle. p16 (9p21) is another TSG. It is also termed as CDKN2A. Over 40% NSCLC report inactivated p16. Inactivation may occur by loss of heterogeneity, point mutation.

2.9.4 Escaping the programmed cell death: Apoptosis or programmed cell death has been the most important metabolic tool for controlling the repercussions that arise from the exposure towards carcinogens, DNA damage, and necrotic cells. The inherent mechanism of the cells to undergo the apoptotic pathway maintains the homeostasis of the human body. These cellular pathways are dysregulated whenever a tumor formation is initiated [Panov, 2005]. *Bcl-2 (B-cell lymphoma-2)* gene was the first oncogene found to function through the production of an inhibitor of apoptosis. The *bcl-2* gene family consists of more than 15 members, which either promote or inhibit the apoptosis [Reed, 1997]. Bcl-2 overexpression, detected by

immunohistochemistry was found in 75%-95% of SCLC tumors, 25%-30% of the squamous cell carcinomas and in 10% of adenocarcinomas [Kaiser *et al.*,1996].

The pathogenesis of lung tissue can be demonstrated through fig 2.3 which is assembled on the basis of earlier reports. Removal of 3p21.3 allele is triggered in many epithelial cells of respiratory system at distinct sites due to the exposure of tobacco and other pollutants. Further, chromosomal loss of tumor suppressor gene located at 3p21.3 takes place which makes it haplo-insufficient. The next set of event is characterized by inactivation of growth regulating genes such as *Rb*, *p53* or *p16* due to mutation or epigenetic silencing. This subsequently allows the clonal proliferation of cells which were initially transformed. It is proposed that during the pathogenesis of the SCLC neoplastic cells arise directly either from normal or hyperplastic epithelial cells without passing through characteristic preneoplastic intermediate pathological stages (parallel theory of lung cancer pathogenesis). On the contrary, the NSCLC pathogenesis is accompanied with sequential morphological changes (sequential theory) [Panov, 2005].

Some of the oncogenic driver mutations have gained clinical importance and are recently been employed in therapeutic regimen of the ADCC patients along with the ongoing chemotherapy which largely comprises of the platinum based drugs. These potential targets have been of great importance in increasing the average survival time of lung cancer patients, particularly ADCC patients [Sholl, 2016].

2.9.5 Diagnosis of Lung cancer

The diagnosis of lung cancer is finalized after a series of tests. The sequence of the tests starts with physical examination, medical history *etc.* followed by complete blood count, hepatic transaminase, calcium levels and measurement of alkaline phosphatase and chest radiography. Confirmation of tumor is done by imaging tests by Computed tomography scan (CT-scan), Positron emission tomography (PET-scan), Magnetic resonance imaging 27(MRI), bone scan depending on the severity of the disease respectively. The step followed by imaging test is confirmation of histology using techniques immunohistochemical characterization. In order to carry out this molecular testing, a section of the lung tumor tissue is required. The approach used

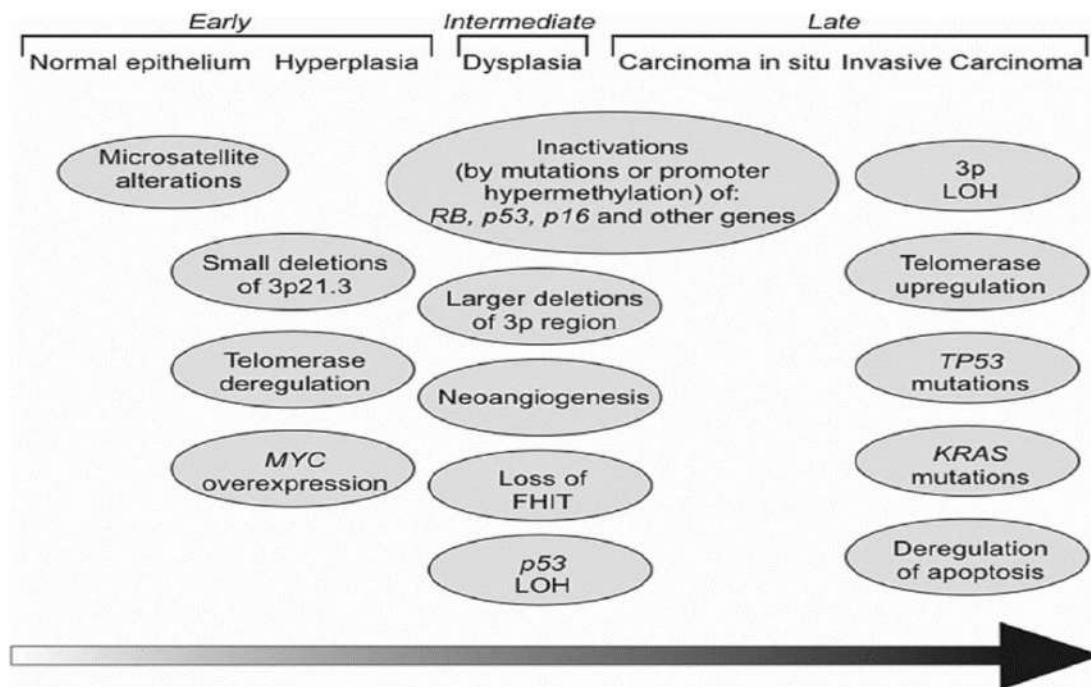


Fig 2.3: Schematic representation of major molecular abnormalities taking place during lung cancer pathogenesis. (Source: Panov, 2005)

to obtain the tissue largely depends on the location and size of the tumor and its metastatic accessibilities. For the centrally located lesions, fibre optic bronchoscopy is used, whereas for lesions located near the periphery CT-guided transthoracic needle aspiration is a typical method. Endobronchial ultrasound and electromagnetic navigation are the techniques of the new diagnostics (Latimer and Mott, 2015).

2.10 Lung cancer Treatment

Treatment for lung cancer depends on stage of lung cancer and type of lung cancer. It's also required that extent of spread of lung cancer must be known before administering the treatment for cancer. Treatment for lung cancer includes surgery of tumor, chemotherapy or radiotherapy (RadiologyInfo.org, 2017). When tumor has not spread to distant areas and is limited to small area, it is treated with local treatment such as radiotherapy and surgery. Chemotherapy and targeted therapy are preferred when tumor has spread to distant areas, as in stage III B and IV. At times chemotherapy is given before or after local treatment to remove the remaining tumor cells (Chang, 2011). Chemotherapy can be given either before surgery as neo-adjuvant to shrink

cancerous cells or along with radiotherapy as concurrent therapy where cancer has metastasizes to nearby important organs due to which there is no possibility of surgery or after surgery as adjuvant chemotherapy to kill any cancer cells that are left behind or as main treatment. When chemotherapy is given as main therapy cisplatin/carboplatin along with either paclitaxel, docetaxel, gemcitabine or vinorelbine is given in case of NSCLC. SCLC subjects are given cisplatin/carboplatin along with irinotecan (Behera, 2007).

Targeted therapy unlike chemotherapy targets the molecules expressed on the surface of cancerous cells. The therapy attacks these targets either by blocking or destroying the targets. Anti-angiogenesis therapy is used to limit the supply of blood to tumor to avoid its growth and ultimately starve it to death *i.e.* anti-VEGF therapy (Al-Abd *et al.*, 2017). Vascular endothelial growth factor (VEGF) is responsible for angiogenesis in tumor. VEGF promotes survival of cells and further inhibit their apoptosis. It is a growth factor that even regulates the permeability of vascular endothelial cells. Cancerous cells produce vascular endothelial growth factor. Anti-angiogenic compounds are hypothesized to act either by making the blood flow normal within the tumor and thereby improving the efficacy of the treatment. Bevacizumab is FDA-approved VEGF inhibitor used in colon cancer. EGFR inhibitors eg. Erlotinib, Gefitinib, Afatinib block the growth of cancerous cells that are EGFR positive. EGFR provide for the development of cancerous cells growth and development (Cabanero *et al.*, 2017). Matrix Metallo protease (MMPs) inhibitors: MMPs cause the breakdown of structure present between cells to allow for the growth of healthy new tissues. These MMPs are therefore responsible for the blood vessel development tissue growth and wound healing in normal cells. Cancerous cells invade MMPs for their growth and metastasize to other tissues and organs of the body.

2.11 Platinum based Chemotherapy regimens

2.11.1 Docetecel cisplatin/ Doectacel carboplatin

Docetaxel is the anti-mitotic chemotherapy drug majorly used for the breast, ovarian and NSCLC treatment. It reversibly binds with micrtotubule with high affinity. Docetaxel is a semi-synthetic taxoid derived from 10-deacetyl bacatin III which result into formation of microtubule assembly and inhibition of tubulin depolymerization which leads to stabilization of microtubule. This drug

gives 23-33% treatment success and 21% objective response in NSCLC patients as well as new cancer cases respectively. Cisplatin has a mechanism which is different from docetaxel is also used for treatment in NSCLC patients majorly. Studies have shown cisplatin-based chemotherapy in NSCLC shows a superior response rate and survival than those regimens which does not contain cisplatin (Donadieu *et al.*, 1991), therefore use of cisplatin regarded as the independent predictor of improved survival (Aihara *et al.*, 2002).

2.11.1.1 Mechanism of action

Cisplatin blocks DNA replication which stops the cell proliferation in carcinogenic cells. On administration, one of the two chloride ligands is displaced by water giving the aquo-complex $cis-[PtCl(NH_3)_2(H_2O)]^+$, this process is known as aquation. The aqua ligand in $cis-[PtCl(NH_3)_2(H_2O)]^+$ is displaced by the *N*-heterocyclic bases on DNA. Guanine preferentially binds subsequently for the formation of $[PtCl(guanine-DNA)(NH_3)_2]^+$, leading to DNA cross-linking interfering mitosis. The damaged DNA activates the DNA repair mechanism which activates apoptosis when it appears impossible to repair the damage. Although cisplatin is frequently designated as an alkylating agent but in reality it has no alkyl group and it therefore cannot carry out alkylating reactions. Cisplatin used in combination with another chemotherapy drug has been a cornerstone in treatment of many cancers. Initially platinum responsiveness is high but majority of cancer patients eventually relapse with the cisplatin-resistance. There are many mechanisms which show this resistance which includes cellular uptake and efflux of drug, increased drug detoxification, inhibition of apoptosis and increased DNA repair.

As it is known that docetaxel exerts by promoting and stabilizing the microtubule assembly by preventing microtubule depolymerization and stabilizing microtubule assembly in the absence of GTP (Williamson *et al.*, 2005). This results a decrease in number of free tubulin which are required for microtubule formation which inhibits mitotic cell division between metaphase and anaphase preventing cancer cell progeny (Yvon *et al.*, 1999). As microtubules do not disassemble in the presence of docetaxel they accumulate inside cell which leads to initiation of apoptosis. Apoptosis has been encouraged by the blocking of apoptosis-blocking bcl-2 oncoprotein. Both *in vitro* and *in vivo* analysis show the anti-neoplastic activity of docetaxel to be effective against a wide range of known cancer cells, cooperate with other anti-neoplastic agents activity, and have greater cytotoxicity. The main mode of therapeutic action of docetaxel

is the suppression of microtubule dynamic assembly and disassembly, rather than microtubule bundling leading to apoptosis, or the blocking of bcl-2.

2.11.2 Irinotecan cisplatin/ Irinotecan carboplatin

Irinotecan (Camposter) is a camptothecin-derived compound whose active compound SN-38 interacts with topoisomerase I forms a stabilizing complex with DNA (Fukuda *et al.*, 2004). Along with its novel cytotoxic mechanism, irinotecan differs from other chemotherapeutic agents in the multidrug-resistant cell lines which retain irinotecan sensitivity (Tusuro *et al.*, 1988). It is used as an approved drug for the metastatic colon cancer in USA. For colon cancer treatment it is either infused alone or with fluorouracil. And for small cell lung cancer treatment it is used in combination with cisplatin. When used as a single agent, the response rate in NSCLC may vary from 15-34% (Pillot *et al.*, 2006).

2.11.2.1 Mechanism of action

Irinotecan is a prodrug (CPT-11) that is metabolized into its active form, SN-38, when inside the cell. Irinotecan and SN-38 may be transported out of the cell via multidrug transporters such as ABCB1 (Fukuda *et al.*, 2004). Irinotecan is converted into SN-38 via human carboxylesterases but can also be converted to inactive metabolites via CYP3A4 and CYP3A5, and SN-38 can be inactivated through glucuronidation to SN-38G via UGT1A1 (Artrc *et al.*, 2010). This inactivation leads to inhibition of both DNA replication and transcription. The UGT1A1*28 allele provides the strongest support for a role of pharmacogenetics in predicting toxicity from irinotecan therapy (Fukuda *et al.*, 2004).

2.11.3 Pemetrexed cisplatin/Pemetrexed carboplatin

Pemetrexed is an anti-metabolite which inhibits thymidylate synthase (TS) and other folate-dependent enzymes. It is mainly used for the treatment of NSCLC and pleural mesothelioma (Metro *et al.*, 2011). Corticosteroids and supplementation with vitamin B12 and folic acid reduces its toxicity. The most common adverse effects are fatigue, nausea, vomiting, mucositis, and rash. Pemetrexed is approved for second-line treatment in NSCLC (Tomasini *et al.*, 2016). Pemetrexed is usually infused in combination with a platinum based drug carboplatin to treat NSCLC.

2.11.3.1 Mechanism of action

Pemetrexed is chemically similar to folic acid and belong to class of chemotherapy drugs known as folate antimetabolites. It inhibits three enzymes i.e. thymidylate synthase (TS), dihydrofolate reductase (DHFR) and Glycinamide ribonucleotide formyltransferase (GRAFT) used in purine and pyrimidine synthesis. By inhibiting formation of purine and pyrimidine nucleotides, pemetrexed arrests DNA and RNA formation which are usually required for survival of both normal and cancer cells (Metro *et al.*, 2011).

2.12 DNA damage

Cell nucleus is considered as the control centre of eukaryotic cell, as the cell store the information about heredity in form of deoxyribonucleic acid (DNA). Genetic information is an important aspect as it constitutes the basis for cell function and its identity (Clancy, 2008). The correct functioning of the cell is associated with safeguarding the genome. DNA is composed of two complementary strands which are joined together by hydrogen bond that maintains its helical structure providing biochemical stability to the molecule. Although the DNA is present inside the cell nucleus considered as the first line defense protection for genetic information as well as making protection shield around DNA from harmful processes which might be active elsewhere in cell. But despite all these things, DNA can be still compromised in many ways. In day to day routine, genome is exposed to wide variety of DNA damaging events which can be chemical or physical in nature, originating inside or outside the cell. A broad range of DNA injuries are considered as its consequence (Weismuller *et al.*, 2002). The genotoxic components which might lead to DNA damage compromises of cigarette smoke, air pollutants and few components like vinyl chloride present in vegetables etc. These substances composed of bulky molecules which link themselves directly to DNA distorting its structure. Several form of radiations also cause DNA damage. Ionizing radiation results into damage bases and may also lead to DNA breaks either it is single or double stranded. Ultraviolet radiations (UV rays) absorbed directly by DNA which induces cross-linking of adjacent bases. DNA synthesis machinery occasionally causes errors during replication which lead to insertions/deletions or mismatches in bases or can also encourage the formation of strand brakes. The spontaneous

hydrolysis of nucleotides might lead to disappearance of genome giving abasic sites (Bernstein *et al.*, 2008).

2.13 Genome Stability

The cell activity is hampered by the DNA damage and consequently it also affects well being of an organism. DNA damage leads to instantaneous cellular consequences which causes cell cycle arrest and even cell death. The genomic lesions might ultimately lead to termination of nuclear event like transcription and translation (Hoeijmakers, 2009). During these processes which involves the translocation of a polymerase along the DNA fiber, the polymerization machinery can get obstruct due to DNA stimulating it to making mistakes. If the lesion is present in the transcribed strand of a gene, expression of that gene can come to a halt as the RNA polymerase stalls. Continuous blocking of DNA polymerase can be regarded as the strong p53-dependent apoptotic signal (Bernstein C *et al.*, 2013). In addition, the importance of efficient handling of such challenges is clearly underlined during the replication process by the discoveries of a large number of damage-specific polymerases that have different specificities depending on the type of lesion in the DNA template (Kannouche and Strydom, 2003). In addition to these trans-lesion polymerases, cells may opt to replicate past an injury by switching templates to the newly synthesized complementary strand, a process that involves recombination (recombination-dependent daughter-strand gap repair (McGlynn and Lloyd, 2002). Nevertheless, if DNA damage is not properly dealt with, it may give rise to stable mutations or chromosomal abnormalities (Bernstein C *et al.*, 2013). Long-term consequences of mutations are the onset of carcinogenesis, aging, and, when present in germ cells, may cause hereditary diseases. It is therefore not surprising that, in the course of evolution, all life forms have invested in accumulating an impressive repertoire of biochemical systems to get rid of DNA damage (Bernstein C *et al.*, 2013).

2.14 DNA repair

In order to deal various errors which might lead to genomic instability, a complex network of mechanisms becomes active in organism with the sole aim to protect genome against various lesions. These mechanisms are known as DNA repair pathways which play exhaustive role in safeguarding genome (Gavande *et al.*, 2016). In case of cancer, DNA repair pathways play two

roles: Inhibition or alteration of one or more repair proteins leads to cancer development or cancer like syndromes. The high repair efficiency of these proteins shows resistance against highly DNA damaging cancer drugs.

As DNA damage can occur in many forms the cells employ different DNA repair mechanisms. Each pathway has a different specificity and strategy to resolve problems caused by these injuries. 130 gene products are identified for genome maintenance when damage occurs in human DNA (Morita *et al.*, 2010). Once damage is healed up, restoring of chromatin resulting in completion of cell cycle occurs. There are mainly four DNA repair pathways in mammals: (1) Nucleotide excision repair (NER) (2) Base excision repair (BER) (3) Double strand break repair (DSBR) (4) Mismatch repair (MR) (Fig 2.4). These pathways intricate a multi-protein processes which includes Several repair factors are often recruited for more than one repair pathway that sometimes lead to functional overlap in DNA repair machinery.

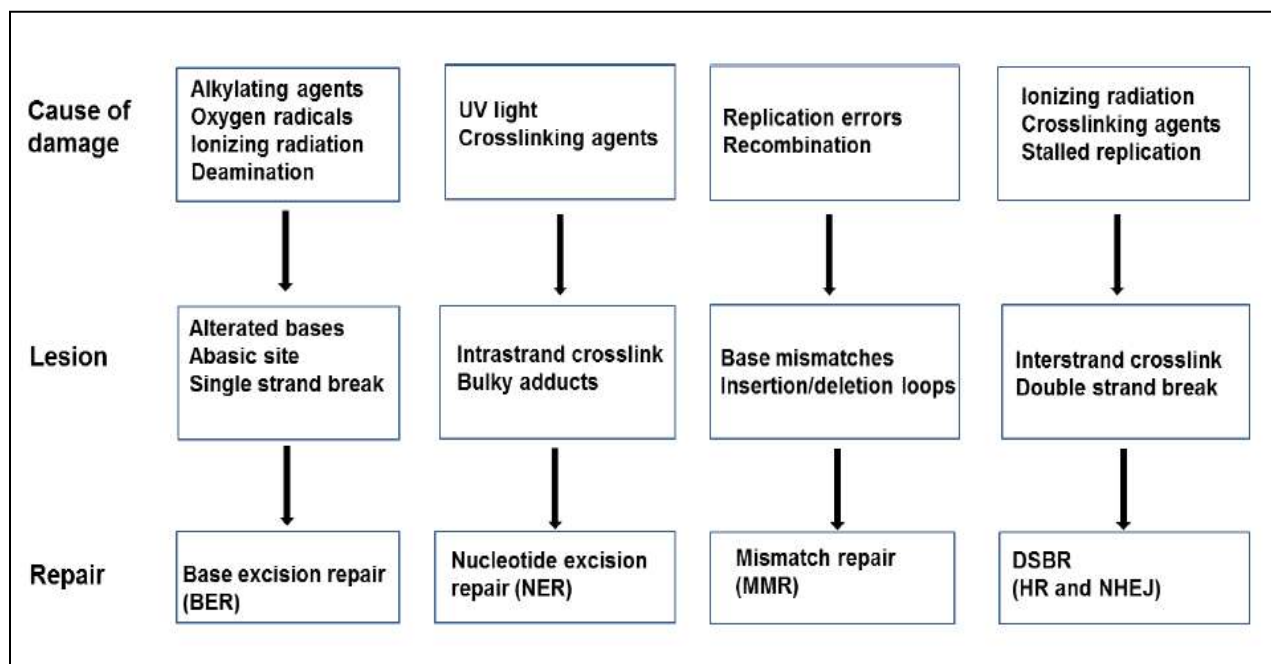


Fig 2.4: Simplified description of main DNA lesions, cause behind these lesions and DNA repair pathways implemented on them (**Source:** Fleck, 2004; Houtgraaf *et al*, 2006)

2.14.1 Nucleotide excision repair

Nucleotide excision repair (NER) is known as most versatile pathway among DNA repair systems. It is simple and employs cut and patch mechanism to remove large lesions in an error-

free manner (Morita *et al.*, 2010). This pathway is activated when it comes in contact with wide range of helix-distorting lesions basically induced by UV photoproducts (cyclobutane pyrimidine (CPD) and pyrimidine (6,4) photoproducts), bulky chemical adducts and oxidative lesions. NER is present in every organism and its general mechanism consists of five steps: (1) Recognition of DNA damage (2) Unwinding DNA helix and verification of lesion at both ends (3) double incision at the both ends of the lesion (4) excision of the oligonucleotide and resynthesis of DNA fragment (5) Ligation of residual nick.

In *E.Coli*, NER removes lesion by using small set of proteins i.e. Uvr A, Uvr B, Uvr C and Uvr D. However, in Eukaryotes the process is more complex as it involves 30 proteins implicated directly in the repair process (Morita *et al.*, 2010). NER process does not use large number of factors to recognize damage. Instead the process of detection is much more general; the insults *via* chemical modifications in DNA are generally recognized by the presence of injury or distortion in DNA (Truglio *et al.*, 2006). NER therefore has a preference for bulky adducts showing a positive correlation between extent a DNA helix deformation and presence of injury. This strategy allows this pathway to repair broadway spectrum of chemically distinct lesions such as cisplatin adducts, benzopyrene *etc.* UV-induced photo-adducts also induces formation of intra-strand cross-links between two flanking pyrimidine bases such as cyclobutane pyrimidine (CPD) and pyrimidine (6,4) photoproducts (Kim *et al.*, 1995).

2.15 Nucleotide Excision Repair pathway in Humans

NER is more than a multi-step pathway as there are many aspects which are applied mechanistically using two different sub-pathways *i.e.* global genome repair (GGR) and Transcription-coupled repair (TCR). As their names implicate, GGR detects lesions present throughout the genome whereas TCR functions only in transcriptionally active regions. Although the steps involved in these pathways are mechanistically overlapping but primary steps of these pathways differ a lot from each other (Fig 2.5).

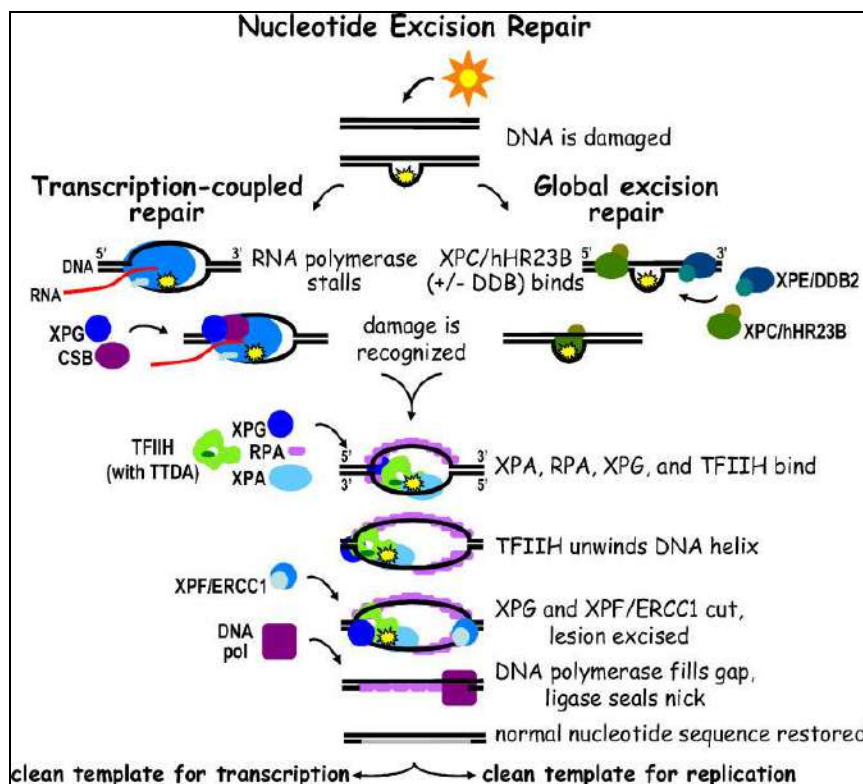


Fig 2.5: Nucleotide excision repair (NER) depicting mode of action of its two sub-pathways: (1) Transcription coupled nucleotide excision repair (2) Global genome nucleotide excision repair (Fuso *et al*, 2006)

2.15.1 Transcription-coupled Nucleotide excision repair (TC-NER)

The presence of an alternative pathway of NER came from the observation that active genes have a faster repair as compared to inactive genes. The faster rate for this pathway is observed because it leads to the removal of enhanced lesions of a gene (Pani and Nudle, 2017). TC-NER differs from GG-NER only in initial steps. The process of TC-NER pathway is explained as follows:

Transcription-coupled repair has a specific meaning with specific consequences in the way the system is considered. At the site of lesion an arrested RNA polymerase provides a strong signal for cell apoptosis. To resume transcription process and stop/delay apoptosis, cell utilizes the pathway to remove transcription lesion is transcription coupled nucleotide excision repair (Ljungman, 2005). When a bulky lesion is present on the transcribed strand of an active gene, it blocks RNA polymerase II (RNA P IIo). This process initiated TC-NER. Cockayne syndrome complementary group B (CSB) interacts loosely with the elongating polymerase with which the process becomes more tightly bound leading to transcriptional arrest (Brueckner and Cramer,

2007). CSB gene encodes 168 KDa protein related to SWI/SNF family of ATP dependent chromatin remodeling factors. CSB is a DNA dependent ATPase and has a DNA binding activity. If RNA polymerase II gets arrested at CPD (cyclobutane pyrimidine dimer), then CSB stimulates transcription elongation by the addition of a nucleotide to the new transcript present *in vitro*. The *in vitro* transcription assays have defined the assembly of NER factors in arrested RNA polymerase II complex only when RNA is utilized in its pure state. CSB/RNA polymerase II complex interacts with transcription factor TFII H subunits and therefore play a significant role in its recruitment (Citterio *et al.*, 2000). In support of this model, it has been shown in studies that CSB and XPG binds to each other in a cooperative manner for repair of cisplatin lesion which has previously lead to the RNA polymerase arrest. When the UV radiation act upon the complex, the interaction between CSB and chromatin bound RNA polymerase II complex gets enhanced (Sarkar *et al.*, 2005). In this whole process the fate of polymerase still remains unclear. *In vitro* studies have suggested the release of CSB after XPF carries out incision activity (Fousteri *et al.*, 2006). However, recent findings suggests that polymerase may stay stalled at the site of damage while repair is performed by the XPG which interacts with RNA polymerase II leading to removal of donating polymerase without causing incision (Fousteri *et al.*, 2006). However *in vivo* evidence presented that NER specific endonuclease works as a single repair complex in combination with stalled RNA polymerase II. *In vivo* studies also explained that CSB dependent CSA recruitment to the RNA polymerase II complex was consistent with the previous findings (Tantin *et al.*, 1997).

CSA, is a multi-protein complex with E3 ubiquitin-ligase activity and RNA Polymerase II/CSB assembly with CSA-DDB1 E3-ubiquitin ligase/CSN complex which further allow the recruitment of HMGN1, HATp300, XAB2 and TF IIS proteins (Fousteri *et al.*, 2006). For recruitment of p300 the CSB is utilized alone while for recruitment of HMGN1, XAB2 and TF IIS requires both CSB and CSA for letting the bind with stalled polymerase (Fousteri *et al.*, 2006). These conformational changes in RNA Polymerase II are required for the accessibility to repair proteins (Fousteri *et al.*, 2006). Backtracking RNA Polymerase II is considered as the main mechanism which plays important role in restarting transcription (Fousteri *et al.*, 2006). During the process of backtracking the small DNA repair enzymes like photolyase could bind directly and reverse CPD (Toranletti *et al.*, 2001). Further TF IIS enables the cleavage of

protruding 3'mRNA by RNA polymerase II so that the transcription could be restarted on lesion removal (Fousteri *et al.*, 2006). This backtracking process is also utilized in chromatin remodeling events which involves high mobility protein group like HMGN1. P300 also have a role in chromatin remodeling while performing both transcription and DNA repair. XAB2 interacts with *XPA* which is suggested to be a key protein at the juncture of both GGR and TCR as these both mechanism shares some common steps (Bowman *et al.*, 1997). Once TF IIH and other factors are established rest all steps in TCR pathway and GGR pathway are identical. The identical steps include binding of *XPA*, *RPA*, *XPG* and TF IIH with the repair complex, of which TF IIH results into unwinding of DNA. Further *XPG* and *XPF-ERCC1* complex excise the lesion and ligation is performed by DNA polymerase leading to seal the gap (Fuso *et al.*, 2006).

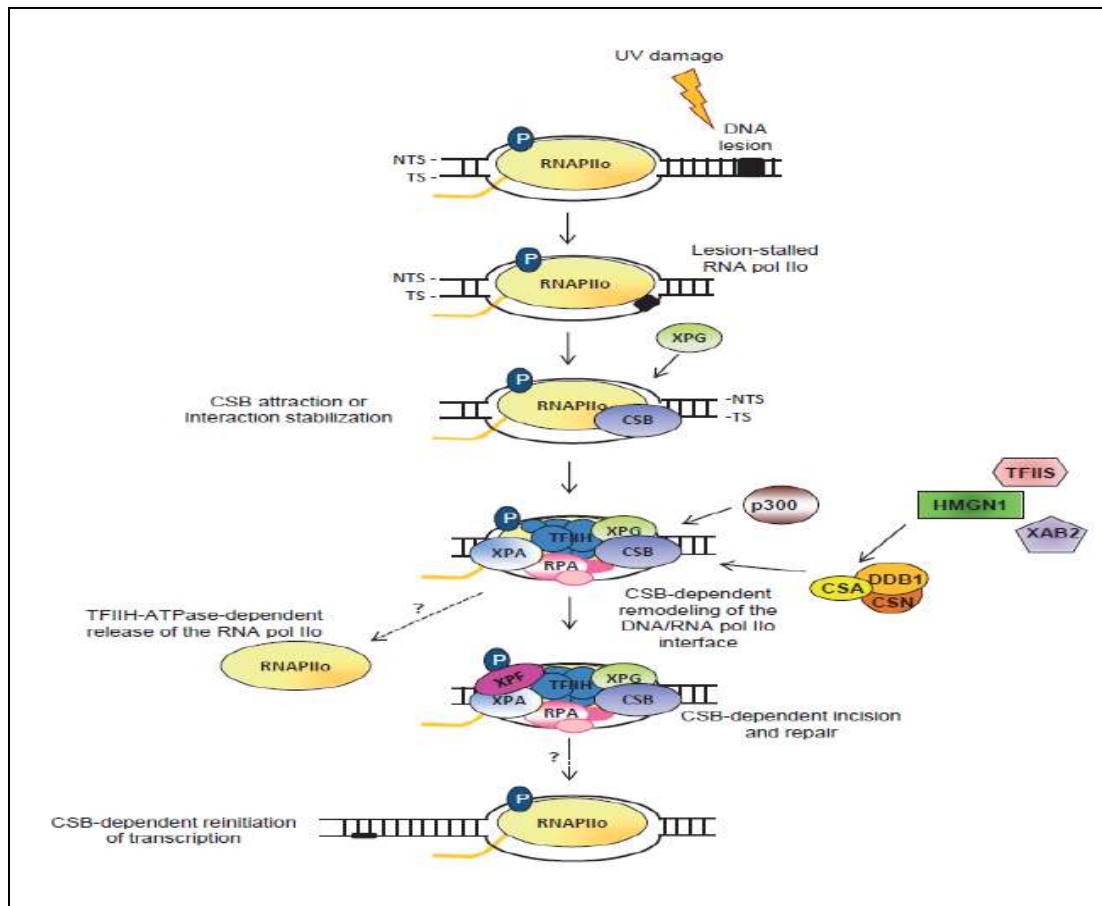


Fig 2.6: Transcription-coupled repair. Transcription is arrested when the RNAPII encounters a lesion. CSB becomes tightly bound and facilitates the recruitment of TCR factors. *XPG*, *TFIIH*, *XPA* and *ERCC 1-XPF* localize to the arrested elongation complex with *RPA* arriving once single stranded DNA is formed. The chromatin remodeling factors p300 and HMG N1 loosen the nucleosome structure behind the polymerase. The RNAPII reverses direction, backtracking from the lesion allowing cleavage of the 3' mRNA which is activated by *TFIIS*. After backtracking, *TFIIH* extends the denatured region around the lesion in order to allow subsequent incision by *XPG* and *ERCC 1-XPF*. (Source: Fousteri *et al.*, 2006; Sarasin and Stary, 2007)

2.15.2 Global Genome Nucleotide Excision Repair (GG-NER)

This repair pathway can resolve, DNA damage from any part of the genome and hence plays an important role in removing majority of lesions. When the DNA lesion is more complex, the faster repair kinetics of this sub pathway works. Therefore, the process of DNA damage sensing defines the efficiency of this sub-pathway.

The precise mechanism of damage recognition of GG-NER has been a matter of discussion past years, although considerable progress has been elucidated but this question still has been unanswered. Different proteins are responsible for the recognition of initial damage-detection step: XPC-HR23B (Sugasawa *et al.*, 2001), XPA (Asahina *et al.*, 1994), RPA (Reardon and Sancar, 2002) and several combinations of other factors (Costa *et al.*, 2003). These proteins show a high affinity for damaged DNA than for undamaged DNA templates. XPC-HR23B and damaged DNA binding complex has a specific action in this sub-pathway. Recognition and presence of lesion is confirmed by this XPC-HR23B complex (Sugasawa *et al.*, 2001). This lead to formation of a unwounded intermediate state of DNA around the damage which leads to formation of an open complex of 25-30bp. This demarcation around these adducts confers the requirement of polarity to the complex so that endonucleases, XPG and ERCC1-XPF can incise the damaged strands through the junctions present between double and single stranded DNA (Wood *et al.*, 1999). Although there are many accessory or stimulatory factors present in NER such as damage recognition, demarcation and double incision process which can be performed *in vitro* by using these 6 factors: XPC-HR23B, XPA, RPA, TFIIH, XPG and ERCC1-XPF. The gap after dual incision is filled by the DNA polymerase which also involves the presence of replication factor C (RFC), proliferating cell nuclear antigen (PCNA) and DNA polymerase δ and ξ , Ligase I action which subsequently seal the residual nick (Costa *et al.*, 2003). The biochemical properties and putative role of these NER proteins has been expressed in section below.

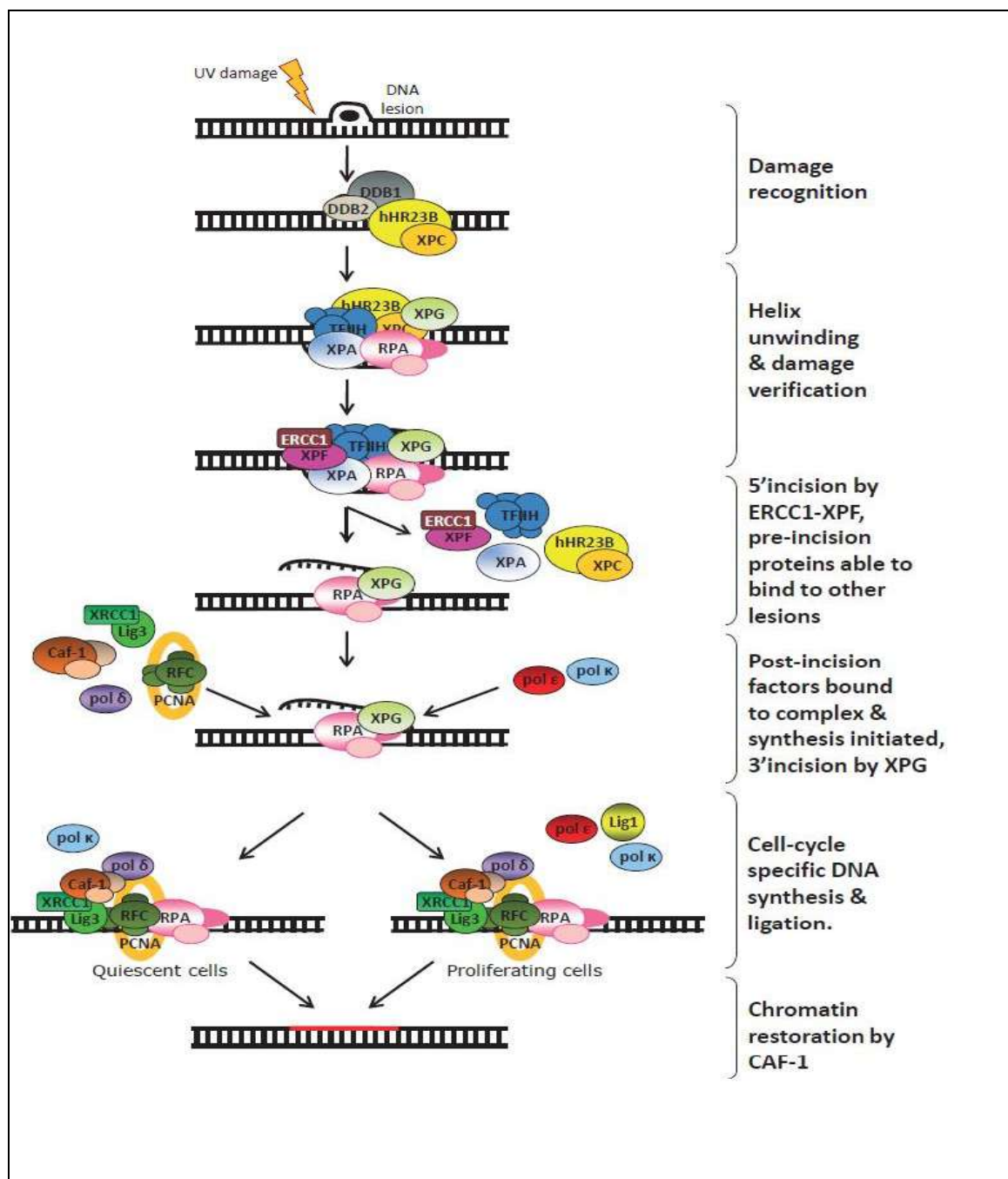


Fig 2.7: Global genome repair in Mammals (Sarasin and Stary, 2007)

2.16 NER Proteins

2.16.1 XPC

Xeroderma pigmentosum group C is a 125KDa protein stabilized by hetero-dimerization with 58KDa HR23B protein. This complex plays an important role in damage recognition step during GG-NER. In GG-NER, the recognition proteins distinguish between damaged and undamaged DNA which has gone under helical distortions (Sugasawa *et al.*, 2002). The initial damage recognition process for this pathway has been a matter of debate in the past studies (Zhu *et al.*, 2006). Few studies have emphasized on the fact that XPA or RPA-XPA complex first binds to the site of damage, thereafter recruiting XPC-hR23B and other core NER factors (Tapias *et al.*, 2004). While *in vitro* studies site clearly that XPC-hR23B complex binds to the damage first. Finally, *in vivo* analysis by Volker *et al.*, 2001 suggest XPC-hHR23B as the principal damage recognition complex as well as initiator of NER pathway.

XPC protein comprises of 940 amino acid located at chromosome 3p25.1 and consisted of 18 exons. Several domains are mapped to its sequence (Clement *et al.*, 2010) (Fig 2.7). As per the crystal structure of RAD 4, the yeast homologue of human XPC, there are four structural domains which are present in this protein (Clement *et al.*, 2010). XPC has a transglutaminase-homology domain (TGD) and three structurally related β -hairpin domains (BHD 1-3) which are characterized by two opposing β -sheets linked by a short hairpin. B-turn sequence connects BHD2 and BHD3 (Li *et al.*, 1997). XPC binds to damaged DNA in two ways, firstly with TGD and BHD1 domains attaches to the double stranded part present next to lesion. Then BHD2 and BHD3 form a clamp like structure which encloses the native strand opposite the damaged site. After that, BHD3 inserts its β -hairpin inside the lesion leading to flipping out two unpaired bases. XPC is present in cells in a bound state with hHR23B and Centrin 2 (Clement *et al.*, 2010). hHR23B is a protein consisting of two ubiquitin-associated domain that stabilizes and enhances XPC efficiency to bind with damaged DNA (Uchinda *et al.*, 2002). A loss of hHR23B protein gives a decrease in XPC levels. Centrin 2, a 18-KDa chromosomal factor also attached to the XPC-hHR23B complex. XPC has a DNA binding capacity whereas main function of

hHR23Band Centrin 2, a calcium binding factor is to protect the initiation complex from degradation and stimulate its activity in DNA repair (Bunick *et al.*, 2006).

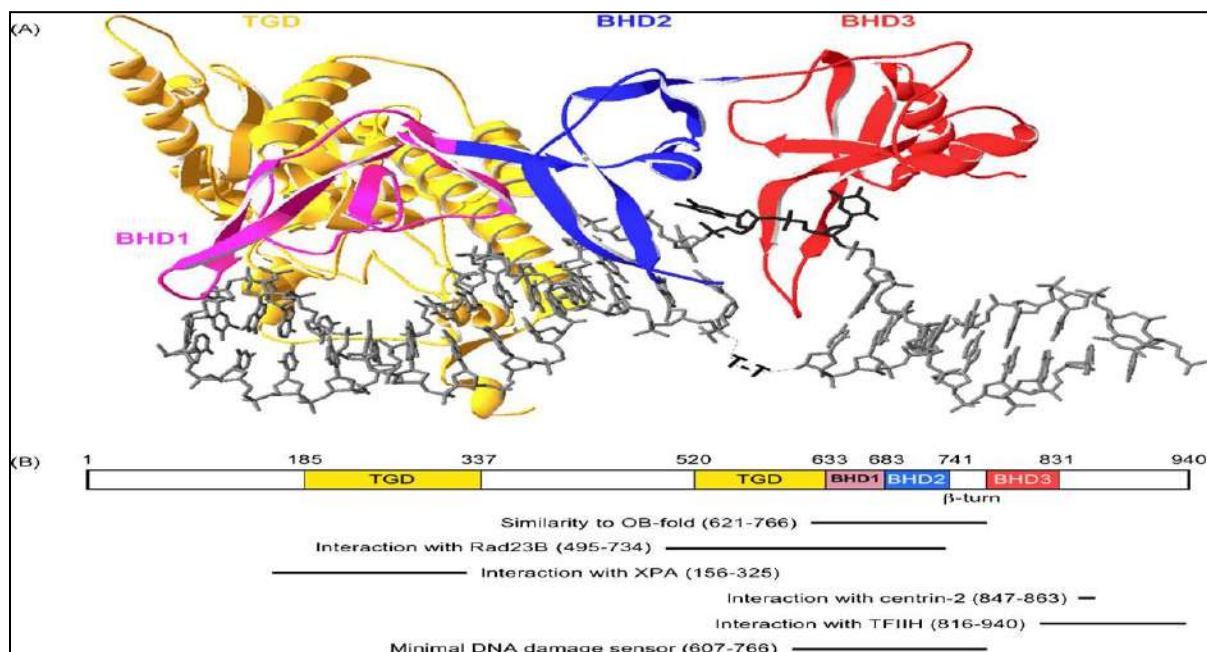


Fig 2.8: Structure of XPC protein derived from the yeast (*Saccharomyces cerevisiae*) Rad4 homolog. (A) Ribbon diagram of the crystal structure of a Rad4 protein fragment (residues 123–632) in complex with heteroduplex DNA carrying a CPD lesion. Multiple Rad4 domains interact with the DNA substrate: TGD (gold), BHD1 (magenta), BHD2 (blue) and BHD3 (red). T-T denotes the CPD, which is totally expelled from the duplex. (B) Scheme of the homologous domains in the human XPC sequence (Clement *et al.*, 2010)

2.16.2 TFIIH

The basal transcription factor IIH (TF IIH) is one of the most important factors of NER pathway. Apart from its significant role in the process of RNA Pol II transcription and NER DNA repair, it has been implicated to play interesting role in RNA Pol I transcription, trans-activation and cell cycle regulation (Coin *et al.*, 2008). It is a multimeric protein complex and consists of ten different gene products: *XPB* (p89), p62, p52, p34, *XPB* (p80), CDK7, Cyclin H, MAT1 and TFB5, which forms different sub-complexes. The core complex comprises of p62, p52, p44, p34, *XPB* which *XPB* links with CAK sub complex (Cdk-activating kinase), which consists of CDK7, Cyclin H, MAT 1. The CAK complex phosphorylates several substrates, among others the carboxy-terminal domain (CTD) of RNA Pol II but its activity remains expendable for in vitro NER (Coin *et al.*, 2008).. There are two large subunits of TF IIH *i.e.* *XPB* and *XPB* which posses ATP-dependent helicase activity in 3'→5' and 5'→3' direction respectively. They play a plausible role in unwinding DNA helix and forming an opening complex around the lesion which is a necessity for dual incision to occur (Coin *et al.*, 2007). *XPB* which is also known as

ERCC2 is a protein which majorly takes part in TC-NER. *XPD* gene encodes 2.3 KB mRNA having 22 exons and 21 introns (Cleaver *et al.*, 2009). *XPD* protein is a 760 amino acid polypeptide of 87 KDa size. It is an integral part of human transcription initiation factor TF IIH and has a ATP-dependent helicase activity. It belongs to *RAD3/XPD* subfamily of helicases.

2.16.3 XPA

After unwinding of DNA duplex, a pre-incision complex is formed which fully opens DNA. *XPA* is a 36KDa scaffold protein with no enzymatic activity of its own (Cleaver *et al.*, 2009). *XPA* has a zinc finger domain which has been regarded as the part of minimal DNA binding domain and has zinc binding to four cysteine residues forming a complex as shown in fig 2.8. Substitution in any of the four cysteine residues leads to NER activity reduction. Structural studies have provided evidence that zinc finger motif is necessary for protein-protein interaction with RPA (Park *et al.*, 2002). *XPA*-RPA binds with damaged DNA with some specificity which results into an enhanced interaction between the two proteins. Priority, it has been a notion that *XPA*-RPA complex was originally responsible for damage recognition but many *in vivo* and *in vitro* studies have provided a substantial proof that recruitment of *XPA* to DNA lesions occur over TF IIH is recruited (Coin *et al.*, 2006). The recruitment of *XPA* is dependent on the *XPC* recruitment. *XPA* shows binding affinity with certain DNA structures which suggest that those DNA conformations would emerge only during the action of TF IIH activity (Cleaver *et al.*, 2009). During that period, the recruitment and assembly of *XPA* and RPA into the pre-incision complex may occur. As mentioned above, evidence suggested that the detachment of CAK from core TF IIH is catalyzed by *XPA* which therefore facilitates enlargement and stabilization of DNA open structure (Coin *et al.*, 2008).

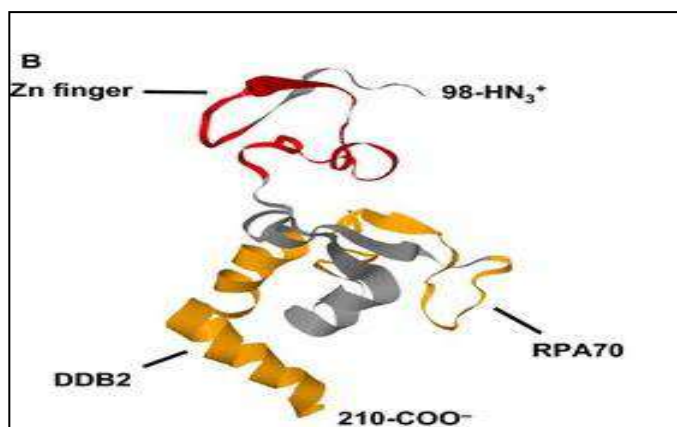


Fig 2.9: Structure of XPA protein
(Source: Coin *et al.*, 2008)

2.16.4 RPA

RPA is a protein which is composed of three subunits 70KDa, 32 KDa and 14KDa, plays important role in replication, recombination and various DNA repair pathways (Fanning *et al.*, 2006). This protein shows a high affinity for single stranded DNA (ss-DNA), RNA binds to ss-DNA via two mechanisms, firstly RPA binds to the region of 8-10 nucleotides which is created by the helicase activity of TF IIH and then it elongates to the stretch of 30 nucleotides creating a wedge of separate DNA strands around the lesion (Hermanson-Miller and Taruchi, 2002). RPA also showed an important role in facilitating the correct position of endonuclease which stimulates its function and as well as leads to the formation of open complex (Fanning *et al.*, 2006). RPA interacts with other NER factors such as *XPA*, *XPG*, *ERCC1-XPF* and is also required for dual incision. RPA also shows association of undamaged DNA strand and required for complex formation. *In vitro* studies have suggested that RPA initially binds with 8-10 nucleotides at 5'-end of the lesion and then further stretches to the 3'-end allowing correct positioning of the *XPG* and *ERCC1-XPF*, 5'-end of the RPA provides a stable binding location to *XPG* endonuclease (deLaat *et al.*, 1998).

2.16.5 XPG

After the pre-incision complex assembly single stranded breaks are introduced by *XPG* and *XPF-ERCC1* (Zhu *et al.*, 2012). Xeroderma pigmentosum complementation group G also known as *ERCC 5* is a 133-KDa structure specific endonuclease of FEN 1 family which encodes 1186 amino acid (Friedberg *et al.*, 2006). FEN 1 family member consists of structure specific endonuclease. The primary structure of human *ERCC 5* proteins has N- and I- domains which are highly conserved and together form a nuclear core. Any mutation in conserved residues present in the active site which includes Glu77, Glu 791 and Glu 812, might lead to the reduction or abolishment of catalytic activity of the protein (Fig 2.10) (Zhu *et al.*, 2012). N- and I- nuclease domain are separated from each other *via* 600 amino acid space region which is highly acidic for protein-protein interactions including with TF IIH and RPA (Zhu *et al.*, 2012). Therefore, it recruits *ERCC5* to the sites of NER. *ERCC 5* has ubiquitin-binding motif (UBM), a PIP domain which mediates interaction with PCNA. The interaction between *ERCC 5* and PCNA triggers the 3'-incision in NER. *ERCC 5* cleaves a 5'-flap, splayed arm and a variety of substrates present at ss or ds-DNA junction with 5'-overhang and makes a 3'-incision (Zhu *et al.*, 2012).

Being structure specific endonuclease and also a 5'→3' exonuclease, *XPG* has its role in both the sub-pathways of NER (Zhu *et al.*, 2012). In TC-NER, preferentially removes DNA damage in the transcribed DNA while GG-NER lesions are removed from complete genome (Tapias *et al.*, 2004). Other than its endonuclease activity, *XPG* has secondary functions also. It plays a supportive role of TF IIH in receptor-mediated transcription. While studies on *S. cerevisiae* demonstrated role of RAD 2 which is a homologue of *XPG* in RNA polymerase II mediated transcription (Mocquet *et al.*, 2008). *ERCC5* is thought to have possible role in removing oxidative damage in BER pathway (Mocquet *et al.*, 2008). Numerous studies using tumor lines and tissues have indicated *ERCC5* a key role in carcinogenesis and its deficiency might lead to DNA repair defects, genomic instability and failure of gene modulation during transcription (Staresincic *et al.*, 2009). *ERCC5* gene is located on chromosomes 13q22.q23 and has 15 exons that range from 61-1074 bp and introns of 250-5763bp range which spans 32Kb.

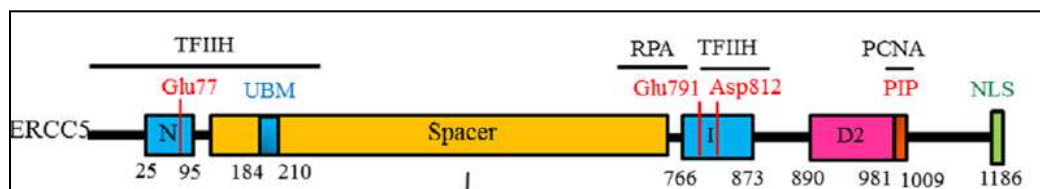


Fig 2.10: (Diagrammatic representation of *XPG* protein and its domains (**Source:** Zhu *et al.*, 2012))

2.16.6 *ERCC1-XPF*

ERCC1 (33 KDa) forms a heterodimer with 103 KDa *XPF* which cleaves DNA at the 5'-end of open complex between the ds and ss-DNA (Matsunaga *et al.*, 1996). The interaction between *ERCC1-XPA* is a crucial step in their recruitment to the damaged site. *ERCC1-XPF* makes incision 15-24 nucleotides away in DNA at the 5'-end of the lesion. *XPF* protein has 916 amino acids containing an *ERCC4* domain which belongs to the nuclease family (Zhu *et al.*, 2012). This domain consists of essential meiotic endonuclease 1 (EME 1) which acts as an essential component of holliday junction and resolves it to interact with MUS 81 (Enzlin and Schrer, 2002). *ERCC4* domain forms a tight complex with *ERCC1*. It acts as a structure-specific DNA repair endonuclease which is responsible for 5'-primer incision during process. Other than NER, this complex also plays an important role in removal of DNA intra-strand cross-links (ICC) and DNA double strand breaks as well (Staresincic *et al.*, 2009).

2.17 Polymorphism in NER genes and cancer risk

Numerous studies have emphasized on the inconsistent results, genetic variations of mutations to nucleotide excision repair (NER) genes which can cause cancer by affecting their repair efficiency. DNA damage has been the primary underlying cancer cause and deficiencies in DNA repair genes likely underlie many forms of cancer (Manelyte *et al.*, 2010). If DNA repair is deficient, DNA damage tends to accumulate. Such excess DNA damage may increase mutations due to error-prone translesion synthesis as shown in fig 2.11. Excess DNA damage may also increase epigenetic alterations due to errors during DNA repair. Such mutations and epigenetic alterations may give rise to cancer. Single nucleotide polymorphism (SNPs) and non synonymous coding single nucleotide polymorphisms (ns SNPs) are usually present at a very low level (1%) in human population (Karahalil *et al.*, 2012). If these genes are present in the regulatory sequences than it might result into mutations which can cause a negative impact on the DNA repair capacity increase the likelihood of cancer development. As functional impact of these polymorphism has not been fully explored, these polymorphism in NER genes or their regulatory sequence induces phenotypical changes as well as they contribute to cancer risk. A study by Sakoda *et al* (2012) suggested the association of NER genes with lung cancer risk. This result indicates that some inherited polymorphic variations in NER genes may result in pre-disposition to lung cancer and potentially to other cancers also.

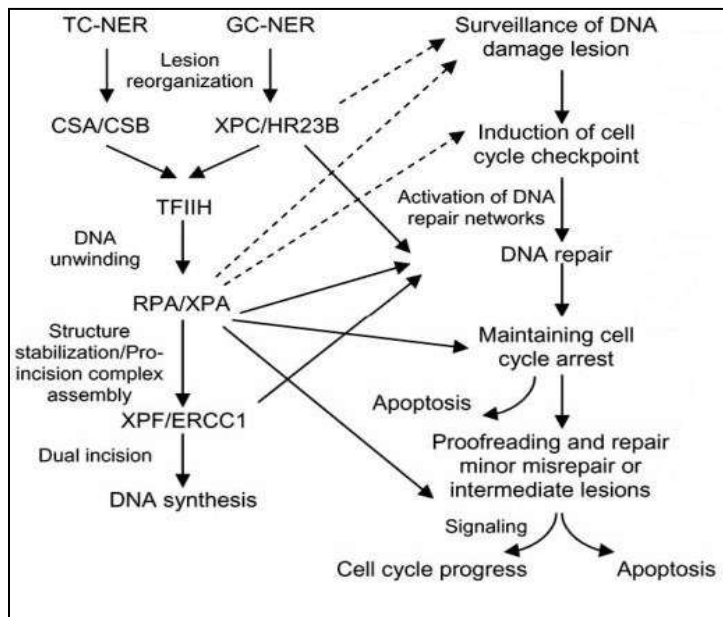


Fig 2.11: NER pathway works in repair of DNA damage. Unrepaired lesion or malfunctioning of proteins associated with NER could lead to unregulated growth and cancer (Zhang *et al.*, 2009)

2.18 XPC polymorphism

It is a key member in NER repair system and forms *XPC-HR23B* complex which recognize target lesions. This polymorphism has three SNP's out of which two are non-synonymous: *Lys*⁹³⁹*Gln* (A33512C, rs2228001) is present at exon 15 shows an A-to-C transition resulting in a lysine-to-glutamine transition at position 939. *Ala*⁴⁹⁹*Val* (C21151T, rs2228000)] present at exon 8 involves a C-to-T transition which results in the substitution of alanine for valine at codon 499 in the interaction domain with hHRAD23 protein which might alter NER capacity and modulate cancer risk. The third *XPC* SNP is located at intron-9 and designated as PAT+/-.

2.19.1 XPC polymorphism and their association with lung cancer risk in Caucasian and Asian Population

There has been a dearth of reports explaining association of *XPC* polymorphism with lung cancer susceptibility. It has been stated in many studies that both the *Lys*⁹³⁹*Gln* and *Ala*⁴⁹⁹*Val* SNP are associated with a relatively high risk for cancer progression (Hu *et al.*, 2005). Few studies have emphasized on association of *XPC Lys*⁹³⁹*Gln* polymorphism with lung cancer in Caucasian and Asian population as explained further in this section (Jin *et al.*, 2014, Vodika *et al.*, 2004). A study by Vogel *et al.*, 2005 suggested association of *Lys*⁹³⁹*Gln* polymorphism with increase lung cancer risk in Finnish population. On the same lines a study by Nielsen *et al.*, 2008 and Chang *et al.*, 2011 have emphasized on the fact that the homozygous variant genotype (CC) of *Lys*⁹³⁹*Gln* polymorphism contributed to increased lung cancer risk in Latin American population. While study by Letkova *et al.*, 2013 and Sakoda *et al.*, 2012 conferred association of *Lys*⁹³⁹*Gln* polymorphism with decreased cancer risk in Caucasian Slovaks and African Americans. Chang *et al.*, 2011 further elucidated association of *Lys*⁹³⁹*Gln* polymorphism with decreased lung cancer risk in African American patients.

While when the same study was applied on Asian population a prominence of increased risk for lung cancer risk was expressed by each and every study. Study by Shen *et al.*, 2005 showed almost 3-fold risk for lung cancer in Chinese patients showing *Lys*⁹³⁹*Gln* polymorphism. This study was further supported by Bai *et al.*, 2007, Hu *et al.*, 2005 and Lee *et al.*, 2005 who also depicted relationship of *A*⁹³⁹*C* polymorphism with increased cancer risk in Asian population. When Jin *et al.*, 2014, performed a subgroup analysis according to ethnicity for these studies a

significant increased risk was observed for homozygote Gln/Gln genotype with lung cancer in Asian population, while no association was found among Caucasian population. The studies by Lopez-Cima *et al.*, 2007 and Letkova *et al.*, 2013 further elaborated prevalence of lung cancer in those patients who were smokers and having *XPC Lys⁹³⁹Gln* polymorphism indicating a persistent role of environmental factor in causing the disease.

Association of *XPC Ala⁴⁹⁹Val* with cancer risk has been rarely studied. Till now very few studies have been conducted on this polymorphism. Three studies on Asian population suggested no association of this polymorphism with lung cancer risk (Lee *et al.*, 2005, Shen *et al.*, 2005 and Bai *et al.*, 2007) while study by Hu *et al.*, 2005 depicted association of *C⁴⁹⁹T* polymorphism with increased lung cancer risk in Chinese population. Only a single study on Caucasian population has applied with this polymorphism which stated role of this polymorphism with increased risk in African American population (Skoda *et al.*, 2012).

Another important polymorphism of *XPC* is PAT +/- polymorphism. For PAT polymorphism, Liu *et al.*, 2011 found a significant risk in patients with mutant genotype while patients with heterozygous genotype was significantly associated with reduced risk of lung cancer in Asian population. However, In Caucasian population that the carriers of the heterozygous genotype had significantly reduced susceptibility to lung cancer compared with carriers of the wild genotype (Marin *et al.*, 2004, Lopez-Cima *et al.*, 2007). Zhu *et al.*, 2013 reported another meta-analysis about *XPC* polymorphism and lung cancer risk. In that study they were not able to find any significant relationship between *XPC* polymorphism and lung cancer. The role of *XPC* polymorphism with lung cancer risk has not been explored in Indian population till now.

2.19.2 XPC polymorphism and its association with other cancers

A large number of studies have elaborated role of *XPC* polymorphism with prostate cancer, bladder and breast cancer. Wu *et al* (2003) studied association of *XPC Lys⁹³⁹Gln* polymorphism with prostate cancer risk in Asian, Caucasian and African population. In the subgroup analysis by ethnicity there was no significant association found in the three races (Asian, Caucasian and African). While a study by Zhang *et al.*, 2014 and Mittal *et al.*, 2012 have emphasized on the association of *A⁹³⁹C* polymorphism with increased prostate cancer risk in Chinese and Indian

population. Zhou *et al.*, 2013 suggested that PAT+/-polymorphism was associated with prostate cancer risk in Caucasian, Africans and Asian population but rs2228001 did not contributed in any such risk. A study by Zhang *et al.*, 2013 suggested association of *XPC Lys⁹³⁹Gln* polymorphism with bladder cancer risk in Asian, Caucasian and mixed population. A study by Zhu *et al.*, 2008 indicated that subjects carrying rs2228000 (*Ala⁴⁹⁹Val* polymorphism) CT genotypes had a better DRC when compared with those carrying the rs2228000 CC genotype showing a protective effect for gastric cancer risk. Song *et al.*, 2013 presented a significant association between the *Ala⁴⁹⁹Val* polymorphism with high risk of recurrence in oropharynx squamous cell carcinoma. A study by Strom *et al.*, 2010 also predicted a significant association of mutant genotype (CC) for *C⁴⁹⁹T* polymorphism with acute myeloid leukemia. Stern *et al.*, 2009 illustrated association of *XPC* polymorphism with bladder cancer risk. Datkhile *et al.*, 2016 also suggested association of *A⁹³⁹C* polymorphism with risk of head and neck cancer in Indian population. Interestingly, a study by Zhi *et al.*, 2012 suggested that *XPC* plays an additional role in the removal of oxidative damage and is involved in cell cycle regulation in the DNA damage response. The study depicted that *XPC Lys⁹³⁹Gln* was not associated with increased bladder cancer risk. The finding by Sak *et al.*, 2005 also expressed no role of *XPC Lys⁹³⁹Gln* polymorphisms in bladder cancer. Although some studies showed the *XPC Lys⁹³⁹Gln* polymorphism significantly elevated the human bladder cancer risk (Chen *et al.*, 2007).

2.19 XPA polymorphism

Xeroderma pigmentosum complementation group A (*XPA*) gene maps at chromosome 9 at 9q22.3 having 12 introns and 6 exons is a 273 amino acid zinc finger DNA-binding protein associated with excision of wide range of lesions such as benzo[α]pyrene induced DNA bulky adducts (Wakasugi and Sancar 1999). *XPA A23G* polymorphism (rs1800975) is (A→G) present at position 23 in the 5' non-coding region with four nucleotides present in the upstream of the start codon of the gene (Butkiewicz *et al.* 2000). The other known polymorphism of *XPA* is codon 228 present in exon 6 associated with *Arg228Gln* (G709A, rs1805160) substitution (Butkiewicz *et al.* 2000).

2.19.1 XPA polymorphism and lung cancer risk in Caucasian and Asian population

The studies illustrating role of XPA A23G polymorphism with lung cancer risk in patients with different ethnicity has been stated. Park *et al.*, 2002 and Wu *et al.*, 2003 found that G variant allele was significantly associated with decreased risk of lung in Korean, Caucasian and Mexican–American populations, respectively. When Liu *et al.*, 2012 stratified their results according to ethnicity for lung cancer patients, the results showed that the association was much stronger in Asians than that in Caucasians. They also observed that there was a significantly increased risk for other cancers subgroup in the homozygous and the recessive model. Lou *et al.*, 2014 suggested that A allele of rs1800975 (A23G polymorphism) might be a danger factor for lung cancer in the East Asians but not in the Caucasians. Butkiewicz *et al.*, 2004 conferred that combined (heterozygous+mutant) genotype might be associated with increased lung cancer risk in Caucasian population. While study by Raaschou-Nielson *et al.*, 2008 and De Ruyk *et al.*, 2007 suggested no association between XPA polymorphism and lung cancer in Caucasian population. Voegel *et al.*, 2005, Ziennolddiny *et al.*, 2009 and Sakoda *et al.*, 2012 elucidated a protective role of rs1800975 for the disease among Caucasians. Many studies in Asian population have emphasized on association of A23G polymorphism with increased lung cancer risk (Park *et al.*, 2002; Zhu *et al.*, 2001 and Quian *et al.*, 2011). Zheng *et al.*, 2013 elaborated that no risk association was observed between XPA polymorphism and lung cancer susceptibility. Wu *et al.*, 2003 also performed study on African American and Mexican American population demonstrating their association of A23G polymorphism with almost 2-fold risk of lung cancer. Lou *et al.*, 2014 suggested that in the subgroup analysis it can be concluded that rs1800975 (A23G) polymorphism was associated with squamous cell carcinoma and the AA genotype increased risk, but no association was found for other subtype like adenocarcinoma in both Asian and Caucasian population. . A study by Wang *et al.*, 2007 have reported that the XPA (A23G) G allele was associated with a reduced lung cancer risk for Caucasians, Mexican-Americans, and African-Americans. Wu *et al.*, 2003 further reported association of A23G polymorphism with higher DNA repair capacity when applied to controls. XPA A23G polymorphism alters the transcription as well as translation of the gene as this polymorphism is located in the vicinity of translation initiation codon which may alter translation efficiency. The nearby proximal nucleotides to the AUG initiation codon are important for the initiation of translation because the 40S ribosomal subunit binds initially at the 5' -end of the mRNA. The consensus sequence

around the start codon is known as the Kozak consensus sequence. The R at position -3 and the G just downstream of the start codon are especially important, and the lack of these bases leads to read-through of the start codon. However, there has been no precise explanation of the mechanism by which the recognition of the start codon is aided by a purine at position -3, which is the core nucleotide of the Kozak consensus (Zheng *et al.*, 2013). The polymorphism *XPA* A23G is a A/G transversion occurring four nucleotides upstream of the start codon of *XPA* and possibly improving the Kozak sequence. The sequences (CCAGAGAUGG) around the predicted initiator methionine codon of the *XPA* gene agree with the Kozak's consensus sequence at positions -3 and +4. Although both the A and polymorphic variant G nucleotides at the -4 position of the *XPA* gene do not correspond to the original consensus Kozak sequence containing the nucleotide C at position -4, it is possible that a nucleotide substitution of A to G at position-4 preceding the AUG codon may affect ribosomal binding and thus alter the efficiency of *XPA* protein synthesis (Zheng *et al.*, 2013).

2.19.2 *XPA* polymorphism and its association with other cancers

Pan *et al.*, 2009 found that the G variant allele in A23G polymorphism was associated with increased the risk of esophageal cancer in Caucasian population while Feng *et al.*, 2008 and Guo *et al.*, 2008 stated that G variant allele might play a protective role in the development of esophageal cancer in Asian population. Three studies on colorectal cancer were all from Caucasians and the results suggested that the G variant allele increased the risk of colorectal cancer significantly (Hansen *et al.*, 2007; Jelenok *et al.*, 2010 and Gil *et al.*, 2011). Jelenok *et al.*, 2010 and Crew *et al.*, 2007 suggested that individual's with Caucasian and mixed ethnicity showed a non-significant borderline association with breast cancer. A study by Hall *et al.*, 2006 interpreted that A23G polymorphism was not the related with the head and neck cancer. Sugimura *et al.*, 2006 and Bau *et al.*, 2007 illustrated an insignificant association of *XPA* A23G polymorphism with oral cancer. A study on Hispanic population suggested that the homozygous G allele was associated with a significantly increased risk for Oral lesions. Dong *et al.*, 2008 and Palli *et al.*, 2010 found that A23G polymorphism showed a 2-fold risk for gastric cancer. Lin *et al.*, 2008 depicted association of A23G polymorphism in Caucasian population with renal cancer. A study on skin cancer in population of mixed ethnicity found that homozygous variant genotype was linked with risk for melanomas (Miller *et al.*, 2006). Weiss *et al.*, 2005 found that A23G polymorphism contributes to endometrial cancer in population of mixed ethnicities.

2.20 XPD Polymorphism

Xeroderma pigmentosum complementation group D (*XPD*)/Excision repair cross complementation group 2 (*ERCC2*) is an ATP-dependent DNA helicase. Many SNPs are located on *XPD* locus but we emphasized only on SNPs present on codon 156, 312 and 751. *XPD* *Arg*¹⁵⁶*Arg* (rs238406) polymorphic variant is located in exon 6 represents a silent polymorphism leading to C→A transition, resulting in an Arg→Arg substitution at codon 156 (Benhamou and Sarasain). Another SNP of the *XPD* gene is present in codon 312 (rs1799793) of exon 10, resulting into G→A substitution, which changes aspartate (Asp) to asparagine (Asn) (Shen *et al.*, 1998). The third SNP present on codon 751 (rs13181) and located at exon 23 of *XPD* gene is characterized by A→C substitution resulting in change of lysine (Lys) to glutamine (Gln) (Shen *et al.*, 1998).

2.20.1 XPD polymorphism and lung cancer risk in Caucasian and Asian population

Epidemiological studies have indicated that above mentioned XPD polymorphic variations i.e. *Lys*⁷⁵¹*Gln*, *Asp*³¹²*Asn* and *Arg*¹⁵⁶*Arg* may influence cancer risk (Shen *et al.*, 2001, Spitz *et al.*, 1998, Butkiewicz *et al.*, 2001). The most frequently studied *XPD* polymorphism are *Lys*⁷⁵¹*Gln* and *Asp*³¹²*Asn*. These polymorphisms are reported to be associated with increased lung cancer risk. Lunn *et al.*, 2000 reported that the *XPD* 751 Lys/Lys genotype has been associated with suboptimal DRC in lung cancer patients. A study by Butkiewicz *et al.*, 2001 and Ziennoldinny *et al.*, 2006 further emphasized on the association of *Lys* A⁷⁵¹C polymorphism with increased lung cancer risk in Caucasians. However, study by Hou *et al.*, 2002 and Misra *et al.*, 2003 suggested the role of *Lys*⁷⁵¹*Gln* polymorphism with decreased cancer risk. Matullo *et al.*, 2006, De Ruyk *et al.*, 2007 and Lopez-Cima *et al.*, 2007 were not able to find any association between *Lys*⁷⁵¹*Gln* and lung cancer risk in Caucasians. Among Asian population, Xing *et al.*, 2003 and Liang *et al.*, 2003 suggested more than 10-fold risk for the disease in individuals carrying Gln/Gln genotype. Studies by Kong *et al.*, 2008 illustrated a 4-fold risk in patients showing *Lys*⁷⁵¹*Gln* polymorphism. Only a study by Liang *et al.*, 2003 suggested presence of protective effect in patients with *Lys*⁷⁵¹*Gln* polymorphism. An overall subgroup analysis by Zhan *et al.*, 2010 elaborated that *Lys*⁷⁵¹*Gln* polymorphism was associated with increased lung cancer susceptibility in both Asian and Caucasian population. Another most studied polymorphism of *XPD* is

*Asp*³¹²*Asn*. This polymorphism is reported to be associated with the increased lung cancer risk in numerous studies. Study by David-Beabes *et al.*, 2001, Popanda *et al.*, 2004, Vogal *et al.*, 2005, Matullo *et al.*, 2006, De Ruyk *et al.*, 2007, Harms *et al.*, 2004 and Lopez-Cima *et al.*, 2007 reported relationship of homozygous rare polymorphism with increased cancer risk for Caucasian population. In Asian population, study by Park *et al.*, 2002 emphasized on a 3-fold risk in lung cancer for *Asp*³¹²*Asn* polymorphism. Liang *et al.*, 2003, Yin *et al.*, 2005 and Zhou *et al.*, 2002 further explained association of heterozygous as well as homozygous variant allele with lung cancer risk while studies by Chen *et al.*, 2002 and Shen *et al.*, 2005 found a protective effect for this polymorphism in Asians. Findings by Seker *et al.*, also elaborated a protective role of 312*Asn/Asn* (mutant) genotype in lung cancer patients. Very few studies have been reported on *XPB Arg*¹⁵⁶*Arg* polymorphism. This polymorphism does not show any amino acid change, although instead of any polymorphic substitution which might alter change the translation rate leading to mRNA instability and hence cause cancer. Catana *et al.*, 2012 is the single study on lung cancer patients for this polymorphism. Catana *et al* further reported a significant risk for ADCC subtype of lung cancer in individuals with the *XPB Arg*¹⁵⁶*Arg* variant genotype.

2.20.2 *XPB* polymorphism and other cancer risk

Numerous studies have been reported on *ERCC2* polymorphism but the most studied polymorphism has been on Codon 751 and 312. Dybdahl *et al.*, 1999 reported basal cell carcinoma risk for 751*Lys/Lys* genotype while a protective effect was observed for the variant allele. A study on head and neck cancer suggested 751*Gln/Gln* (mutant) genotype as the most common reported genotype which is found to be related with a modest risk (Strugis *et al.*, 2000). Moller *et al.*, 2000 reported no relationship between *Lys*⁷⁵¹*Gln* polymorphism and DRC. While Spitz *et al.*, 2008 stated association of *XPB Lys*⁷⁵¹*Gln* and *Asp*³¹²*Asn* genotypes with less optimal DRC. A study on glioma patients indicated that the cases were less likely than the controls to carry variants at codon 312 and 751 (Butkiewicz *et al.*, 2001). Bernard-Gallon *et al.*, 2008 showed a 6-fold risk for breast cancer in patients having heterozygous as well as mutant genotype. Yan *et al.*, 2014, Shi *et al.*, 2004 and Jelenok *et al.*, 2010 suggested association of *Asp*³¹²*Asn* polymorphism with increased breast cancer risk among Caucasians. While two breast cancer studies on population of mix ethnicity emphasized on the protective effect for the disease with *Asn/Asn* genotype (Shen *et al.*, 2006, Jogersen *et al.*, 2007). Study by Mechanic *et al.*, 2006

correlated the mutant genotype (CC) of A⁷⁵¹C polymorphism with increased DRC in Caucasians. Han *et al.*, 2005, Lovatt *et al.*, 2005 demonstrated a link between the mutant genotype for both Lys⁷⁵¹Gln and Asp³¹²Asn polymorphism with protective effect in the basal cell carcinoma patients. Debniak *et al.*, 2006 was significantly associated with increased skin cancer risk for XPD Lys⁷⁵¹Gln polymorphism. Kertat *et al.*, 2008 showed an increased risk in cutaneous melanoma in patients having Lys⁷⁵¹Gln polymorphism. Jiaze *et al.*, 2007 and Ji *et al.*, 2010 conferred that heterozygous genotype (Asp/Asn) showed a protective effect for head and neck cancer patients. Prablan *et al.*, 2010 was associated increased bladder cancer risk for Asp³¹²Asn polymorphism. Majumder *et al.*, 2007 emphasized that no significant association between ERCC2 polymorphism and oral leucopenia. Harth *et al.*, 2008 also supported the fact that heterozygous genotype (Asp/Asn) associated with head and neck cancer risk. Few studies have been reported on this Arg¹⁵⁶Arg polymorphism. The studies conducted by Yin *et al.*, 2013 and Strugis *et al.*, 2007 on Danish and Caucasian subjects carrying heterozygous genotype were associated with significant basal cell carcinoma and glioma risk for Arg¹⁵⁶Arg polymorphism.

2.21 XPG polymorphism

Xeroderma Pigmentosum Complementation Group G/ Excision repair cross complementation group 5 which lies at locus13q22-q33 have 15 exons and 14 introns. Two polymorphic variants of XPG has been included in the present study *i.e.* Phe⁶⁷⁰Leu (rs2228959) and Asp¹¹⁰⁴His (rs17655) but functional significance of these polymorphisms are still unknown. The Phe⁶⁷⁰Leu polymorphism is located on exon 9 leads to C→A transition (C2207A), thus resulting in phenylalanine (Phe) to Leucine (Leu) amino acid change. The Asp¹¹⁰⁴His is present at exon 15, results in G→A substitution (G3507C) which changes aspartate (Asp) to histidine (His) amino acid.

2.21.1 XPG polymorphism and lung cancer risk in Asian and Caucasian population

XPG polymorphism and their association with lung cancer risk has not been much explored and few studies has been reported till date. Jeon *et al.*, 2003, Sakiyama *et al.*, 2005 and Hung *et al.*, 2008 depicted an increased risk for the disease in Asian subjects while a study by Shen *et al.*, 2005 did not show any association for lung cancer risk with Asp¹¹⁰⁴His genotypes. Three more studies illustrating association with lung cancer were concluded. A study by Chang *et al.*, 2008

further suggested a protective effect for *Asp¹¹⁰⁴His* polymorphism with lung cancer risk in Hispanics. While two studies in African (Chang *et al.*, 2006) and mixed population (Cui *et al.*, 2006) population suggested an increased risk for patients carrying His/His genotype.

2.21.2 XPG polymorphism and other cancer risk

A study by Kumar *et al.*, 2003 reported lack of association in genotypic variants with *Phe⁶⁷⁰Leu* polymorphism towards breast cancer. For *Asp¹¹⁰⁴His* polymorphism, Kumar *et al.*, 2003 and Rajaraman *et al.*, 2008 stated that mutant genotype (AA) was found to be associated with breast cancer risk in Caucasian population. Emmert *et al.*, 2000 stated association of *XPG Asp¹¹⁰⁴His* with increased risk for breast cancer risk whereas exon 9 polymorphism was not detected in the study population. Crew *et al.*, 2007 and Mechanic *et al.*, 2006 reported no association of *Asp¹¹⁰⁴His* polymorphism in African Americans and whites towards breast cancer. Similarly, Lavanya *et al.*, 2012 have also reported no association of *Asp¹¹⁰⁴His* polymorphism towards breast cancer susceptibility in Indian population. A study on Asian (Ming-Shiean *et al.*, 2010) and Caucasian (Jorgensen *et al.*, 2007 and Smith *et al.*, 2008) predicted heterozygous as well as mutant genotype have a protective effect in breast cancer patients. Four studies on skin cancer indicated GA+AA genotype was found to have a protective effect (Blankenburg *et al.*, 2005, Povey *et al.*, 2007, Figl *et al.*, 2010 and Wang *et al.*, 2007). An increased risk for skin cancer was observed in patients for *Asp¹¹⁰⁴His* polymorphism (Ibarrola-Villava *et al.*, 2011). Hussain *et al.* 2009 further reported association of *XPG* exon15 polymorphism with reduced gastric cancer risk. Three studies on bladder cancer (Sanyal *et al.*, 2004, Garcia-Closas *et al.*, 2006, Narter *et al.*, 2009) described relationship of *Asp¹¹⁰⁴His* polymorphism with protective effect in patients carrying mutant genotype. While a study by Wu *et al.*, 2006 on Chinese population and Rouissi *et al.*, 2011 in Tunisian population a higher risk for bladder cancer in individual with combined genotype (GA+AA). A study on non-Hispanic white population documented weak association of *XPG* SNPs with small cell carcinoma of head and neck cancer. Studies on *Asp¹¹⁰⁴His* polymorphism depicted its association with head and neck cancer suggesting a protective relationship in those patients comprising of mutant genotype (Abbasi *et al.*, 2009, Ma *et al.*, 2012). Study by Sugimara *et al.*, 2006 illustrated association of *Asp¹¹⁰⁴His* polymorphism with increased head and neck cancer risk. Pardini *et al.*, 2008 and Canbay *et al.*, 2011 deciphered a

protective association between *Asp*¹¹⁰⁴*His* polymorphism and colorectal cancer. A study by Gil *et al.*, 2011 showed 2-fold risk of colorectal cancer for *Asp*¹¹⁰⁴*His* polymorphism.

2.22 XPF polymorphism

Xeroderma pigmentosum complementation group F/ Excision repair cross complementation group 4 spans 28.2kb has 11 exons and located on chromosome 16p13.2-p13.13. The most studied polymorphism is *Arg*⁴¹⁵*Gln* (rs1800067) which is a transition at codon 415 resulting in arginine-to-glutamine amino acid change. This transition results into protein interaction which hampers *XPF/ERCC1* complex activity during NER pathway and alters genetic susceptibility towards cancer. The other two polymorphism which are not much studied but are important for the present study are: 673C>T (rs316028) which is a cytosine-to-threonine transition and therefore affects transcriptional activity of the gene. The other one is 11985A>G (rs254942) polymorphism resulting in an alanine-to-glycine transition.

2.22.1 XPF polymorphism and lung cancer risk in Asian and Caucasian population

XPF polymorphism and their association towards lung cancer susceptibility has not been much explored till date. A study by Hung *et al.*, 2008 suggested association of *XPF Arg*⁴¹⁵*Gln* polymorphism with decreased lung cancer risk in both Asian and Caucasian population. Another study explained the association of lung cancer with *Arg*⁴¹⁵*Gln* polymorphism suggested the relationship of homozygous variant genotype with increased risk in Latin population (Chang *et al.*, 2008). On the contrary a study by Yu *et al.*, 2010 and Shao *et al.*, 2008 has emphasized on the decreased risk towards lung cancer for 673 C>T polymorphism. To our knowledge, no prior data on 11985 A>G polymorphism and its role in cancer risk has been reported. Only a single study by Liu *et al.*, 2015 suggested no role of 11985 A>G polymorphism towards lung cancer risk.

2.23.2 XPF polymorphism and other cancer risk

Many studies reported a significant association between breast cancer risk and *XPF* polymorphism. For *Arg*⁴¹⁵*Gln* polymorphism, Mechanic *et al.*, 2006, Crew *et al.*, 2007, Jorgensen *et al.*, 2010 and Kohlhase *et al.*, 2014 have depicted a protective effect towards breast cancer in Caucasian population. A study by Han *et al.*, 2009 did not predict any association between *XPF* polymorphism and breast cancer. Whereas two studies in African American population

predicted association of heterozygous+mutant (GA+AA) genotype for *XPF Arg⁴¹⁵Gln* polymorphism with breast cancer risk (Smith *et al.*, 2008, Mechanic *et al.*, 2006). Rajaraman *et al.*, 2008 assessed a relationship between *XPF* polymorphism and breast cancer risk among Caucasians. Steck *et al.*, also elaborated association of rs1800067 (*Arg⁴¹⁵Gln*) polymorphism with protective effect against colon cancer in Caucasian as well as African American population. Wyss *et al.*, depicted an increased risk for head and neck cancer risk in Caucasian as well as African American patients showing *Arg⁴¹⁵Gln* polymorphism. Whereas Yu *et al.*, 2011 suggested a protective effect in Caucasians carrying *G⁴¹⁵A* genotypes. Santos *et al.*, 2013 expressed an increased thyroid cancer risk in Japanese patients carrying GA+AA genotype for *Arg⁴¹⁵Gln* polymorphism. Cheng *et al.*, 2013 and Wang *et al.*, 2013 both studies emphasized on presence of protective effect towards glioma cancer in Asian patients. Rajaraman *et al.*, 2010 showed an increased glioma risk in patients with combined genotype (GA+AA) and are of Caucasian ethnicity. Two studies on colorectal cancer deciphered association of *XPF Arg⁴¹⁵Gln* polymorphism with protective effect (Gil *et al.*, 2012, Huang *et al.*, 2006) whereas Joshi *et al.*, 2009 and Moreno *et al.*, 2006 did not show any association between *XPF* polymorphism and Glioma cancer.

A study by McWilliams *et al.*, 2008, Agalliu *et al.*, 2010, Doherty *et al.*, 2011 and Krupa *et al.*, 2011 showed association of *Arg⁴¹⁵Gln* polymorphism with increased pancreatic, endometrial, prostate and laryngeal cancer risk in Caucasian population. Few studies have been reported on the role of the 673C>T polymorphism and their association with different cancers. A study on Caucasian population found a strong association of 673C>T polymorphism towards increased risk for bladder cancer (Shi *et al.*, 2012). Another study by Liu *et al.*, 2015 negated any association of 673C>T variant towards esophageal cancer risk. A study on Romanian population predicted no association of 11985A>G polymorphism with acute myeloid leukemia risk (Baneseu *et al.*, 2016).

2.24 NER polymorphism and clinical data

Lung cancer is most commonly diagnosed cancer world-wide. It has been found that in Indian population over the last 10 years the survival rate in lung cancer patients have increased from median survival time of 11 months to an overall 5-year survival rate of 17.8% (Parikh *et al.*, 2016). This benefit is due to availability of target chemotherapy drugs and appropriate selection

of patients for treatment. Further the crude lung cancer cases reported in China were 46.08 percent of 100,000 people out of which the 5-year survival rate is less than 20% (Zhou *et al.*, 2014). While 5-year mortality rate of Chinese population was 45.6% per 100,000 people (Zhou *et al.*, 2014). An average 5-year survival rate for NSCLC is quite low illustrating <10% survival among Polish population (Butkiewicz *et al.*, 2012). However, the 5-year survival rate of NSCLC remains less than 15% among developing countries (Malik *et al.*, 2015). The efficacy of platinum-based chemotherapy varies among Caucasians, with a response rate of 26–60% (Bahl *et al.*, 2006). Ardizzoni *et al.*, 2007 suggested there is considerable heterogeneity in therapeutic efficacy of platinum-based chemotherapy between different patients of different ethnicity, and the response rate of platinum doublets is only 25–30% in NSCLC.

As reported in previous studies platinum based chemotherapeutics are the most common regimens used for the treatment of lung cancer patients and genetic factors are recognized as the important parameters for the efficacy of individuals towards the drugs (Vincent *et al.*, 2009, Bala *et al.*, 2016). Many population based studies supported a relationship between polymorphism involved in DNA repair and decreased chemotherapy response (Shi *et al.*, 2018). A poor survival has been observed in NSCLC patients which is partly due to resistance shown by the drugs (Sullivan *et al.*, 2014). It has been suggested that the genetic variants as well as protein expression might affect the response to platinum based chemotherapy as well as its prognosis (Shi *et al.*, 2018, Wu *et al.*, 2013) Platinum based chemotherapy shows its anti-cancer activity by formation of bulky platinum-DNA adducts which result in the destabilization of the DNA helix leading to the blockage of replication which further results in inhibition of transcription process. These adducts are repaired by the NER-BER pathways and plays an important role in evaluating the sensitivity of the tumor cells to the platinum based treatment (Sullivan *et al.*, 2014).

2.23.1 Association of *XPC* polymorphism with overall survival and clinical outcomes

Few studies on *XPC* polymorphism have emphasized on their relationship with the clinical outcomes. The investigations by Matakideu *et al.*, 2006 suggested no significant association of *Lys*⁹³⁹*Gln* variant between histology and stage of diagnosis in lung cancer patients. They further emphasized on an increased mean survival for mutant genotype in *XPC Lys*⁹³⁹*Gln* variant. Both the *Lys*⁹³⁹*Gln* and *Ala*⁴⁹⁹*Val* SNP are associated with a relatively high risk for cancer progression.

Xiao *et al.*, 2010 reported that the heterozygous *XPC Lys⁹³⁹Gln* genotype was found to be associated in an effective response for stage III NSCLC patients undergoing platinum-based chemotherapy in. Furthermore, the polymorphic status of *XPC Lys⁹³⁹Gln A/C* might be a predictive factor for treatment response in advanced NSCLC patients (Xiao *et al.*, 2010). Yang *et al.*, suggested a longer survival in patients having combined genotype for *Lys⁹³⁹Gln* polymorphism. Sanyal *et al.*, 2007 emphasized on the modest influence of *XPC* variants on overall survival in patients after lung cancer diagnosis depending on their tumor histology and chemotherapy regimen with which they have been treated. Sakoda *et al.*, 2012 estimated that for NSCLC and SCLC patients modestly influence overall survival after lung cancer diagnosis, depending on tumor histology and receipt of chemotherapy. Kang *et al.*, 2012 suggested association of DNA repair gene *XPC* with clinical outcome for epithelial ovarian cancer treated with platinum based chemotherapy. Gln/Gln genotype was detected with the increased risk of death when compared with Lys/Lys genotype in Asian population. Fleming *et al.*, 2012 found association of heterozygous+mutant (combined) genotype with prolonged survival in ovarian cancer patients of Caucasian ethnicity.

2.23.2 Association of *XPA* polymorphism with overall survival and clinical outcomes

The only studied polymorphism till date *XPA* is *A23G* and its association with overall survival and clinical outcomes has been explored in many studies. A study on Chinese population showed association of *XPA* polymorphism with increased death risk only in female patients (Su *et al.*, 2015). Alternatively, the association of *XPA* variants with lung cancer suggested a poorer outcome might be explained by the fact that inefficient DNA repair leads to additional instability and more aggressive tumors (Wei *et al.*, 2000, Cheng *et al.*, 2000). A study defining association of acute myeloblastic leukemia with *XPA* polymorphism showed an allele dosage effect of *XPA* AG with respect to outcome (Su *et al.*, 2015). Thus, the proportion of individuals with resistant disease was 6.5% for *XPA* AA, 11% for those AG, and 30% for patients with the GG genotype. Consistent with this, actuarial probability of overall survival at 5 years was 50% for *XPA* AA, 29% for AG, and 0% for GG. In the multivariate analysis, adjusted for the most important clinical and biologic characteristics, *XPA* GG genotype was the most important factor associated with chemoresistance and with overall survival (Monzo *et al.*, 2006). Sun *et al.*, 2015 did not found any association between *XPA A23G* polymorphism and overall survival in osteosarcoma

patients. The critical role of *XPA* in NER pathway influencing cisplatin resistance have been evaluated by Reidl *et al.*, 2003 and Martin *et al.*, 2008. The relationship of *XPA* expression to cisplatin resistance and clinical prognosis of patients treated with platinum-based chemotherapy is still controversial according to previous reports (Bowden *et al.*, 2013, Weaver *et al.*, 2005, Liu *et al.*, 2012). Ding *et al.*, 2012 suggested that *XPA* gene's single nucleotide polymorphism (SNP) when located at -4G/A position *i.e.* the fourth nucleotide before start codon, it has been related significantly with increased risk of progression and death in non-small cell lung cancer patients when treated with radiotherapy and platinum-based chemotherapy among Asians. Yang *et al.*, 2012 demonstrated the significant prognostic effect of the *XPA* A23G for esophageal squamous cell carcinoma (ESCC). The A/A and A/G genotypes at *XPA* 5' -UTR have been observed to be associated with higher expression levels of *XPA* mRNA compared with G/G genotypes in healthy volunteer blood donors suggesting its association with better survival in patients with acute myeloid leukemia (Monzo *et al.*, 2006). Sugiyama *et al.*, 2007 also supported that patients having A/G genotype at *XPA* 5'UTR were more susceptible of developing oral cancer. Carles *et al.* 2006 found association of A/G genotype with poor response to radiotherapy in patients with stage I to II head-and-neck cancer. Clinical impact on prognosis for *XPA* polymorphism was also demonstrated in patients with cancers of the bladder and colorectum (Gu *et al.*, 2005).

2.23.3 Association of *XPD* polymorphism with overall survival and clinical outcomes

The association studies depicting role of *XPD* polymorphism with OS in lung cancer patients using survival analysis has been explained in this section. The studies conducted on Asian population by Provencio *et al.*, 2012 and Spanish patients by Liu *et al.*, 2011 suggest that *Lys*⁷⁵¹*Gln* polymorphism might not be a genetic determinant for survival in lung cancer patients treated with platinum based chemotherapy. When segregated on the basis of histology a protective effect was observed for individuals with SCLC carrying mutant (CC) genotype (HR: 0.49; p=0.04) for *XPD Lys*⁷⁵¹*Gln* polymorphism. There are few studies which did not observe any association between OS and *XPD Lys*⁷⁵¹*Gln* polymorphism in NSCLC patients of Caucasian ethnicity (Li *et al.*, 2013, Ludovini *et al.*, 2011, Mathiaux *et al.*, 2011, , Booton *et al.*, 2006, Kalikaki *et al.*, 2009 and Gandara *et al.*, 2009). Yao *et al.*, 2009, Zhang *et al.*, 2012 and Ren *et al.*, 2012 validated that there was no association between *Lys*⁷⁵¹*Gln* polymorphism and OS in Asian subjects diagnosed with NSCLC. Camps *et al.*, 2003 suggest a poor survival in Spanish population illustrated *Lys*⁷⁵¹*Gln* polymorphism in NSCLC patients. Cheng *et al.*, 2013 suggested

a longer survival for SQCC patients having *Gln/Gln* genotype in Chinese population. Further, Li *et al.*, 2010 also proposed a better survival for NSCLC patients carrying *Gln/Gln* genotype for *Lys⁷⁵¹Gln* polymorphism in Chinese population. Studies by Li *et al.*, 2010 and Wu *et al.*, 2012 conferred no significant relationship between *Asp³¹²Asn* polymorphism and OS in lung cancer patients of Asian and Caucasian ethnicity. Joerger *et al.*, 2012 concluded a lower risk of death in Caucasian population for patients showing *Asp³¹²Asn* genotypes. Furthermore, Zhang *et al.*, 2012, Gurubhagavatula *et al.*, 2004, Isla *et al.*, 2004 and Tibaldi *et al.*, 2004 did not observed any association between OS and *Asp³¹²Asn* polymorphism in NSCLC patients of Caucasian ethnic background. Meta-analytical study conducted Yin *et al.*, 2011 found no association of clinical outcomes with *XPD* polymorphism in Asian subjects. Studies by Li *et al.*, 2013, Vinolas *et al.*, 2004 and Ryu *et al.*, 2011 demonstrated no association of *Lys⁷⁵¹Gln* and *Asp³¹²Asn* polymorphism with clinical response in NSCLC patients. However a study conducted in Caucasian population (Mathiaux *et al.*, 2011) marked a significant association of *XPD Lys⁷⁵¹Gln* polymorphism with favorable objective response. Platinum based doublet chemotherapy is the standard care for first line treatment for lung cancer patients showing variable resistance with different chemotherapeutic efficacy and prognosis. The heterogeneity may be due to different parameters used in clinical setting like TNM staging, performance status and histology. SNPs are considered as biomarkers for cancer prognosis due to their modification in gene function related with phenotypic drug sensitivity.

2.24.4 Association of *XPG* polymorphism with overall survival and clinical outcomes

Studies have reported correlation of *XPG* polymorphism with chemotherapy response and prognosis in NSCLC patients. Xiang *et al.*, 2017 suggested no association between *His¹¹⁰⁴Asp* polymorphism and NSCLC in the overall analysis. The study emphasized on the fact that stratified analysis by ethnicity yields a significant difference in the estimates of effect among Asians and the results are only observed in the homozygous genotype. Matakdidou *et al.* 2006 reported Caucasian patients with the wild-type genotype had a median survival of 24.5 months, whereas patients with heterozygous and homozygous variant genotype for *His¹¹⁰⁴Asp* polymorphism had a median survival of 17.8 and 14.7 months respectively. Few other studies on analysis of *His¹¹⁰⁴Asp* demonstrated no significant effect on the mutant genotype and OS of patients diagnosed with advanced NSCLC (Sullivan *et al.*, 2014, Zhang *et al.*, 2013, Yuli *et al.*,

2013). Studies by Bartolucci *et al.*, 2009, Sun *et al.*, 2009, Butkiewicz *et al.*, 2011 also supported the notion of no significant association exist between *XPG* polymorphism (*Asp*¹¹⁰⁴*His*) and clinical outcomes in NSCLC patients. A study on Chinese population predicted association of *XPG* polymorphism towards NSCLC prognosis (Zou *et al.*, 2015). Wang *et al.*, 2015 also suggested association of *XPG* genes with colorectal cancer prognosis.

2.24.5 Association of *XPF* polymorphism with overall survival and clinical outcomes

The clinical studies on *XPF* polymorphism has been rarely explored and not many studies have been reported previously on this gene. Huang *et al.*, 2014, Kornguth *et al.*, 2005 and Fareed *et al.*, 2010 tried to understand the role of *XPF* polymorphism with clinical response in NSCLC and gastro-esophageal cancer but all the three studies were not able to find any significant association. Vaezi *et al.*, 2011 found that expression level of *XPF* polymorphism in Squamous cell carcinoma patients of head and neck cancer showed correlation with clinical response to DNA damaging agents. DeLoia *et al.*, 2012 and Alexander *et al.*, 2012 suggested association of *XPF* polymorphism with clinical outcomes in patients diagnosed with ovarian and gastric cancer.

Chapter-3

Gaps in the study and objectives

3.1 Study population

There have been very few reports on known NER genes of DNA repair pathway in North Indian population. NER genes and their association with lung cancer susceptibility in North Indian population have not been much explored. The combined effect of these repair genes polymorphism towards lung cancer risk is also less explored. There have been no studies reported in Indian population elaborating role of the different SNPs in the NER genes with clinical outcomes of lung cancer patients undergoing platinum based doublet chemotherapy. As many DNA repair genes carry polymorphisms with the potential to modulate gene function altering DNA repair capacity and impairing the DNA repair mechanisms which are involved in removing DNA adducts, therefore identification of polymorphism of DNA repair genes and understanding their correlation with survival in lung cancer patients and susceptibility to lung cancer would yield insights how these gene polymorphism adversely or favorably influence the chemotherapy outcomes among North Indians. Some remarkable have been observed when lung cancer patients treated with different chemotherapy regimens and their association with NER gene polymorphism such as *XPD*. It is known that selection of chemotherapeutic regimen could improve response and hence overall survival of cancer patients. Personalized chemotherapy as per the molecular marker could help to improve the response rate as well as survival of lung cancer patients. This study in future can facilitate assessment of genetic variations of genes like *XPD/ERCC2* and help patients to make therapeutic decisions for individualized therapy in advanced lung cancer patients.

3.2 Objectives of the study

1. To study the single nucleotide polymorphic variants in Nucleotide Excision Repair (NER) pathway.

2. To evaluate the role of NER gene polymorphism in lung cancer susceptibility.
3. To test association between gene polymorphism and clinical outcomes.

CHAPTER-4

Material and methods

4.1 Study design

This case-control study was designed for assessing role of various NER genes with lung cancer predisposition. The study comprises of 370 lung cancer cases and 370 healthy controls. Eligible cases were recruited from Department of Pulmonary Medicine, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India. The study has been reviewed and approved by the Ethics Committee of Post Graduate Institute of Medical Education and Research (PGIMER) Chandigarh having approval number PGI/IEC/2014/305. The study was carried out under ethical standards established by Declaration of Helsinki (DoH) 1964 and its amendments.

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

- (1) Diagnosis of lung cancer was confirmed by cytology or histology.
- (2) Stage III or IV disease.
- (3) No age, gender, smoking, histology and staging restrictions was applied.
- (4) Untreated and intent to treat with definitive chemotherapy (Treated with platinum agent's cisplatin or caboplatin, either first or second line.)
- (5) An Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 0–2.
- (6) At least one bi-dimensionally measurable lesion, according to the RECIST criteria
- (7) Adequate organ function, defined as absolute neutrophil count >1500/ μ L, platelet count >100,000/ μ L, and levels of creatinine, liver enzymes and alanine aminotransferase (ALT) less than two times the upper limits of normal (ULN).
- (8) Written informed consent obtained.

The exclusion criteria included the following considerations:

- Patients having prior history of any other carcinoma.

- Active infection or immunosuppression (HIV).
- Patients receiving any kind of systemic steroids.
- Patients suffering from chronic form of diarrhoea due to any cause.
- Patients who are not undergoing chemotherapy.

370 unrelated individuals with no history of lung or any other cancer visited PGIMER for regular health check-ups were recruited as control group. Each control was matched with the cases on basis of age, sex and smoking parameters. A detailed questionnaire was filled by a trained interviewer for each case and control. The questionnaire consisted of major demographic details such as age, gender, place, smoking and drinking habits. Smokers reported tobacco habits such as smoking of cigarette and/or beedi (a native cigarette like stick of coarse tobacco hand-rolled in a dry tembuhurni leaf). Cumulative smoking exposure in pack years (PY) was calculated using formula: [(cigarettes or beedis per day/20) X years smoked], $PY \leq 25$ were light smokers and $PY > 25$ were heavy smokers. Details regarding histology, TNM classification, clinical staging, regimen of treatment, number of chemotherapeutic cycles and performance status (KPS and ECOG) as per the Response Evaluation Criteria in Solid tumors (RECIST) were retrieved from the medical records of each patient from the hospital. The follow up data of patients were telephonically obtained by utilizing the information present in the medical records.

4.1.1 Chemotherapy regimen

Platinum- based doublet chemotherapeutic drug combination was administered to all the lung cancer patients enrolled in the study. This combination comprised of different drugs such as docetaxel, irinotecan or pemetrexed along with a platinum based drug such as cisplatin or carboplatin. Docetaxel is a taxane, however irinotecan is a topoisomerase I poison and pemetrexed is an anti-folate. These drugs were administered intravenously at a fixed dosage according to the surface area of the body. Where in 75mg/m^2 docetaxel, 500mg/m^2 pemetrexed or 75mg/m^2 irinotecan was given as 1 hr infusion which was then followed by cisplatin 70mg/m^2 infused for over 3 hrs. 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 premedication routine. Tumor response assessment was carried out after

four cycles of chemotherapy. In certain cases, due to high toxicity and massive disease progression reported by radiological assessment chemotherapy procedure was stopped even before the completion of four cycles. The patients showing a good objective response were administered with additional two cycles of chemotherapy.

4.1.2 Follow up and response assessment

A total of 311 lung cancer subjects were followed up for survival through telephonic conversations. Every two months until death or the end of the study period patients were called and inquired about health. Primarily we assessed overall survival and for that we considered survival time from the date of diagnosis till the date of last follow up or death due to other reasons. A five year survival data was considered for study which includes patients enrolled from September 2012-2017. STRAD diagram explains the selection criteria of the patients for overall survival analysis as shown in Fig 4.1

4.1.3 Clinical Response

Objective tumor response is one of primary endpoints for efficacy in cancer clinical trials. It has been widely utilized by clinicians in evaluation and guidance of cancer treatments as well as in prediction of clinical outcomes. The most commonly used standard for determining the tumor response is a set of published rules called Response Evaluation Criteria in Solid Tumors (RECIST) (Wolchok *et al.*, 2009).

- RECIST criteria, which defines when cancer patients improve ("respond"), stay the same ("stable"), or get worsen ("progression") during antitumor treatments.
- There are four evaluable categories of tumor response for target lesions:
- **Complete Response (CR)** –all target lesions have disappeared during the treatment. Lymph nodes selected must return to normal size (<10mm).
- **Partial Response (PR)** –at least 30% decrease from baseline sum of the longest diameter.
- **Stable Disease (SD)** –no significant decrease or increase in the size of target lesions.
- **Progressive Disease (PD)** – $\geq 20\%$ increase over the smallest sum of the longest diameter and with at least 5mm increase, or appearance of any new lesion

- **Overall response :** The best response recorded from the start of the treatment until disease progression/recurrence (taking as reference for PD the smallest measurements recorded since the treatment started)(Yu *et al.*,2013)

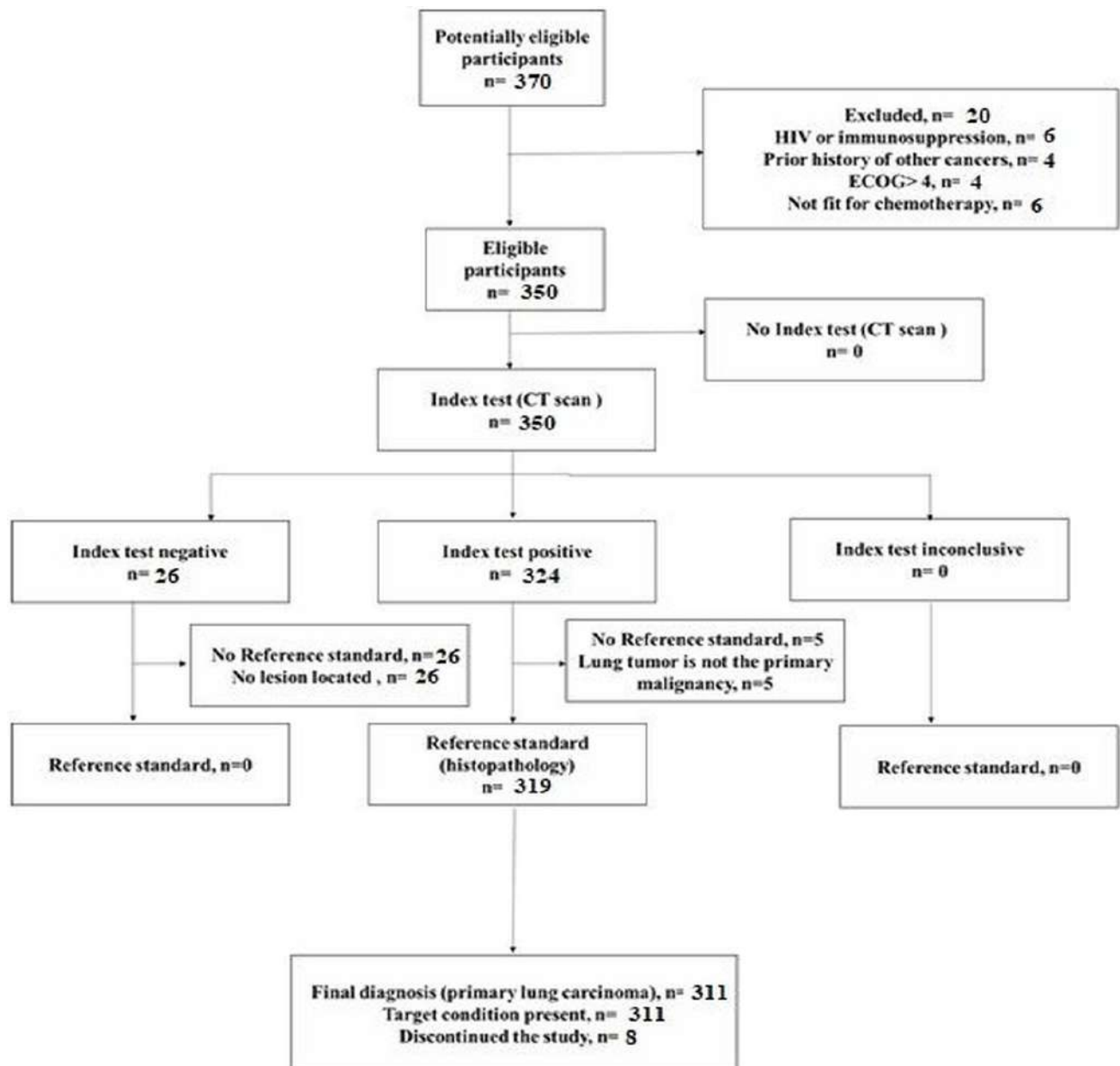


Fig 4.1: STRAD diagram explains the selection criteria of the patients for overall survival analysis

4.2 Genomic DNA Extraction

Genomic DNA was isolated from peripheral blood lymphocytes of cases and controls according to the protocol of Bartlett and Stirling (2003).

- 3ml of blood sample was treated with equal volume of washing buffer having 320mM sucrose, 1.5% Triton X-100, 20mM Tris-HCl (pH 8.0) and 5mM MgCl₂ for separating red blood cells from the blood lymphocyte. The sample was centrifuged at 3,500 rpm for 10 min each time. This step was repeated thrice.
- RBC pellet was lysed at 45⁰C overnight using an equal volume of WBC lysis buffer having 400mM Tris-HCl (pH 8.0), 150mM NaCl, 60mM EDTA, 1% SDS, and 100µg/ml proteinase K.
- An equal volume of 150µl phenol/144µl chloroform/60µl isoamyl alcohol (25:24:1) solution was added and contents were mixed gently by inverting microfuge tube twice or thrice. The mixture was centrifuged at 8000 rpm for 10 minutes at 4°C, two clear layers were observed. Aqueous layer was separated and the above step was repeated again.
- To the aqueous layer, an equal volume of chloroform/isoamyl alcohol (24:1) solution was added. Centrifuged at 6500 rpm for 5 min, aqueous layer has been collected in another tube.
- DNA was precipitated using the equal volume of chilled isopropanol or 2.5 times volume of ethanol which was obtained after centrifugation at 10,000rpm for 5min. Freeze it at -20°C for 1-2 hours.
- Decant ethanol and pellet was washed with 1ml of ice-cold 70% ethanol by centrifuging at 10,000 for 5 min. The pellet was then air-dried.
- DNA pellet was then suspended in 200µl of sterile Tris-EDTA buffer.

4.3 Qualitative and quantitative estimation of genomic DNA

DNA quality was assessed using agarose gel electrophoresis. 0.8% agarose gel with 0.5µg/ml ethidium bromide immersed in 0.5X TBE was used to visualize the isolated DNA sample. The sample was mixed with 6X loading dye and diluted with sterile water before loading in the wells. The gel was visualized under UV transilluminator showing the presence or absence of DNA (Johansson, 1972).

Quantity of DNA was determined by spectrophotometric analysis using Nanodrop (NanoDrop ND-1000 spectrophotometer). A single unit of absorbance (1OD) of double stranded DNA at 260 nm is equivalent to 50ng/µl of DNA. Henceforth, the amount of DNA in each sample is calculated by estimating its absorbance at 260 nm. Also, the purity of the sample was estimated by obtaining the ratio of absorbance at 260 nm to absorbance at 280 nm. For a pure DNA sample, the ratio should come out to be 1.8. A ratio less than 1.8 indicates protein contamination in the DNA sample and a ratio of more than 1.8 indicates the presence of RNA or phenol contaminants [Desjardins and Conklin, 2010]. After complete examination of concentration and purity, DNA was diluted to 100ng/µl and was stored at -20°C for further use as template.

4.4 Genotyping of polymorphic sites *XPC*, *XPA*, *XPB*, *XPD*, *XPG* and *XPF* gene

4.4.1 Genotyping of *XPC* polymorphism by PCR-RFLP

Genotyping analysis was performed using PCR- RFLP technique using primers, annealing temperature and restriction enzymes as previously reported by Hirata *et al.*, 2009 which are tabulated in table 4.1. A standard PCR was used for the amplification of the polymorphic site. The amplification for both sites was carried out in 25ul volume with approximately 50ng of DNA, 0.5µM of each primer, 200µM of dNTP, 1X PCR buffer, 1.5mM of MgCl₂, 100µg BSA and 1U of Taq Polymerase (DNAzyme II DNA Polymerase, Thermo Scientific). The 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 annealing temperature of the primer for 30s, primer extension at 72°C for 30s and final extension step at 72°C for 5 mins for the *XPC* polymorphism.

Restriction fragment length polymorphism

After the successful amplification, the PCR product was treated with the suitable restriction enzymes which are obtained from New England Biolabs, Ipswich, MA, USA. 10µl of each PCR product was treated with 5 units of enzyme in the presence of 1X respective NEB buffer making reaction volume of 20µl. This reaction was incubated at the desired temperature which was 37°C for all enzymes used for the present study except Taq I polymerase which has an incubation temperature of 65°C for 12-16 hours.

The digested products were resolved on 2.5% agarose gel 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 2µl of 6X loading buffer containing sucrose, bromophenol blue and xylene cyanol. The resolved gel is visualized under the UV trans-illuminator and restriction pattern is observed [Rasmussen, 2012]. Further the results were documented using Gel Doc™ XR+ Gel Documentation System by Bio rad. To avoid biasness in results scoring of gel was done by two independent observers. The scoring of gel was performed according to 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 the cases or control. 15% samples were randomly genotyped twice to confirm the reproducibility of the results which was found to be 100%.

4.4.2 Genotyping for XPA variants

The genotyping of the two polymorphic variants of XPA gene i.e. A23G and G709A was carried out by PCR-RFLP. The primers and restriction enzymes used in this protocol have been previously detailed by Gunes *et al* (2009). The PCR protocol used for the amplification of these genotypic variants is as follows. The 25µl reaction mixture comprised of 1X PCR buffer, 1.5mM MgCl₂, with 0.5µM of both forward and reverse primer, 200µM of each dNTP's, 100µg/ml bovine serum albumin (BSA) and 1U Taq polymerase (DNAzyme II DNA Polymerase, Thermo Scientific) and approximately 200ng DNA. The annealing temperature was optimized for each

variant and their respective restriction enzymes have been mentioned in table 4.1. The cycling conditions used in this case are similar as mentioned in case of previous gene. The amplified products were digested with the respective restriction enzymes (2U) and were incubated at desired temperature overnight. RFLP products were resolved on the agarose gel electrophoresis/polyacrylamide gel electrophoresis depending on the difference in the size of fragments to be separated.

4.4.3 Genotyping of *XPB* variants

Genotyping was performed using PCR-RFLP methods for three *XPB* variants i.e. *Lys*⁷⁵¹*Gln*, *Asp*³¹²*Asn*, *Arg*¹⁵⁶*Arg*. The primers were taken as described in previous studies (Yin *et al.*, 2005, Yu *et al.*, 2004, Li *et al.*, 2007) and have been represented in Table 4.1. The protocol for the PCR-RFLP was similar as used in previous gene. The amplified product was digested with the respective enzymes defined for each SNP. The amplified products were resolved using agarose gel/polyacrylamide gel electrophoresis on the basis of size. For all polymorphism, 15% samples were repeated twice using same procedure and the reproducibility results were 100%.

Polyacrylamide gel electrophoresis

- Polyacrylamide gel has been used to separate nucleic acids and proteins in a solid support under the presence of electric field in a buffer. PAGE gel is a cross linked polymer of acrylamide and bisacrylamide. Pore size is determined by the total concentration of monomeric units (acrylamide and bis-acrylamide) and ratio of acrylamide: bis acrylamide. Polymerization is initiated by ammonium per sulphate (APS) and catalyzed by TEMED (N, N, N', N'-Tetramethylethylenediamine). Polyacrylamide gels have a rather 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 visualised using silver staining. This protocol used was given by Sambrook *et al.*, [2001] with slight modifications.
- Vinyl addition polymerization initiated by a free radical-generating system.
- 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 bisacrylamide, 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.

Silver staining

It is a highly sensitive technique, relies on the reduction of silver cations into insoluble silver metal by nucleic acids. Reduced silver grains deposit in the gel around DNA bands, creating a latent image. Image is developed to visibility by soaking the gel in solution of silver cations and reducing agent. Silver granules in latent image catalyze the further reduction and deposition of silver from the solution. Bands manifest as dark brown or black regions. Methanol and Glacial Acetic Acid are used as fixative for the DNA in gel. Formaldehyde acts as reducing agent (Reduces Ag^{2+} to Ag) and Sodium Carbonate is used as developing enhancer. The solutions which were used included:

- Fixative Solution- Water: Methanol: Glacial acetic acid (50:40:10).
- Staining solution-0.1gm of AgNO_3 in 100ml of de-ionized water followed by addition of 150 μl of 37% formaldehyde.
- Developing solution-3gm of Sodium Carbonate in 100ml de-ionized water and 150 μl of 37%formaldehyde and 20 μl of 10mg/ml sodium thiosulfate.

After putting the gel in fixative for 30 minutes, the gel was washed with sterile deionized water at least thrice and then, it was kept in staining solution for 30 min in dark. After staining, the gel was carefully washed with deionized water three times and then the developing solution was added and gel was kept on shaking. The bands were observed.

4.4.4 Genotyping of *XPG* and *XPF* variants

Two *XPG* polymorphisms *i.e.* $\text{Phe}^{670}\text{Leu}$ and $\text{Asp}^{1104}\text{His}$ were detected using Polymerase chain reaction- Restriction fragment length polymorphism (PCR-RFLP) assay as described earlier in the material and methods section (Jeon *et al.*, 2003 and Kumar *et al.*, 2003). Three *XPF* polymorphisms *i.e.* $673\text{ C}>\text{T}$, $11985\text{ A}>\text{G}$ and $\text{Arg}^{415}\text{Gln}$ were also detected using PCR-RFLP assay as mentioned previous studies (Liu *et al.*, 2015 and Lavanya *et al.*, 2015). The primers

were taken from the priory reported studies as mentioned in Table 4.1. The amplified product was digested with the respective restriction enzymes defined for each SNP. The amplified products were resolved using agarose gel/polyacrylamide gel electrophoresis on the basis of size. For all polymorphism, 15% samples were repeated twice using same procedure and the reproducibility results were 100%.

Table 4.1: List of primers and restriction enzymes used for NER genes and their genetic variants

Gene, SNP ID	Substitution	Primer sequence: forward and reverse	Annealing temperatur e	Product (bp)	Restriction enzyme	Restriction pattern
<i>XPC Lys⁹³⁹Gln</i> rs 2228001	Exon15	F: ACC AGC TCT CAA GCA GAA GC R: CTG CCT CAG TTT GCC TTC TC	56°C	281	<i>Pvu II</i>	A (281) C (242)
<i>XPC Ala⁴⁹⁹Val</i> rs 2228000	Exon8	F: TAA GGA CCC AAG CTT GCC CG R: CCC ACT TTT CCT CCT GCT CAC AG	63°C	253	<i>SacI</i>	C (152) T (131+21)
<i>XPA A23G</i> rs 1800975	5'-UTR	F:TTAACTGCGCAGGCGCTCTC ACTC R:AAAGCCCCGTCGGCCGCCG CCAT	60°C	158	<i>MspI</i>	A (158) G (132+26)
<i>XPA G709A</i> rs 1805160	Exon6	F: TTTTCAGAATTGCGTC R: TTCATATGTCAGTTCATG	46°C	143	<i>TaqI</i>	G(127+16) A(143)
<i>XPD Lys⁷⁵¹Gln</i> rs 13181	Exon 23	F: GCC CGC TCT GGA TTA TACG R: CTA TCA TCT CCT GGC CCC C	57°C	294	<i>PstI</i>	A (290+146) C(227+146+63)
<i>XPD Asp³¹²Asn</i> rs1799793	Exon 10	F: CTG TTG GTG GGT GCC CGT ATC TGT TGG TCT R: TAA TAT CGG GGC TCA CCC TGC AGC ACT TCC	63°C	249	<i>StyI</i>	G (507+244) A(474+244+33)
<i>XPDArg¹⁵⁶Arg</i> rs238406	Exon 6	F: TGG AGT GCT ATG GCA CGA TCT CT R: CCA TGG GCA TCA AATTCC TGG GA	60°C	209	<i>HinfI</i>	C (288) A (206+82)
<i>XPG</i> <i>Phe⁶⁷⁰Leu</i> rs2228959	Exon 9	F: GACAGTCCAGTTTACCT R: CACGTCGTCCCTCTCA	60°C	400	<i>EcoRI</i>	A (125+275) C (400)

<i>XPG Asp¹¹⁰⁴His</i> rs17655	Exon 15	F: GACCTGCCTCTCAGAATCATC R: CCTCGCACGTCTTAGTTTCC	56°C	271	<i>NlaIII</i>	G (271) C (227+33)
<i>XPF 673C>T</i> rs316028	Exon 6	F: GGG AGG CAA ACA GAG GTC TGA ATT R: TGC GAT TAC TCC CCA TCC TTC TT	58°C	137	<i>EcoRI</i>	C (114+23) T (137)
<i>XPF 11985 A>G</i> rs254942	Exon 6	F: GGA GTC AAG AAA CAG CCA ACC TAG TA R: AGG AAG ACA GGA TGA CAG CCA G	58°C	129	<i>RsaI</i>	A (129) G (104+25)
<i>XPF Arg⁴¹⁵Gln</i> rs1800067	Exon 8+31	F:CTTCGGGTGAAGGAATAA G R:TTCTCAAAGGTTTTCTGT AG	60°C	380	<i>XmnI</i>	G (380) A (284+96)

4.5 Statistical Analysis

Statistical analysis for Demographic parameters:

To compare demographic features like sex and smoking status Chi square test was applied. For continuous variables such as age and pack years smoked students t-test has been used. To differentiate between observed and expected genotypic frequencies of controls and cases Hardy Weinberg Equilibrium (HWE) was applied.

Statistical analysis for Genotypic and Allelic frequencies: The Hardy–Weinberg equilibrium (HWE) theory ($p^2+2pq+q^2=1$; where p is the frequency of the wild-type allele and q is the frequency of the variant allele) was used for controls to calculate the genotype frequencies of all the polymorphisms using χ^2 test. The minor allele frequency was also calculated for each polymorphic site. Pearson's χ^2 test was used to compare the allelic and genotypic frequencies between cases and controls to find whether significant difference occurs between the two groups [Thakkestian *et al.*, 2005].

Association analysis: Unconditional Multivariate logistic regression was used to find out the adjusted odds ratio along with 95% Confidence Intervals (CI) associated with each genotype and various genotypic combinations. The odds ratio was adjusted for age, gender and smoking as these might act as confounding factors. All the variables used in the study were dichotomous except the continuous variables such as age. In the logistic regression analysis, the occurrence of lung cancer was taken as the dependent variable and genotype was selected as the independent variable which can predict the susceptibility for lung cancer along with various confounding factors such as age, sex and smoking status. The wild genotype was used as a reference group in case of each polymorphic site [Begg and Zhang, 1994; Karimi *et al.*, 2013].

Stratified analysis: The same approach was used for evaluating the risk associated with these polymorphic sites in the various sub groups formed on the basis of histology (SQCC, ADCC and SCLC), smoking status (smokers and non-smokers) and clinic- pathological parameters such as stage, primary tumor extension, lymph node invasion and metastasis. Also, the different genotypic combinations

including the double and triple inter SNP combinations were evaluated using the same approach. All p-values were two sided, and a p-value of <0.05 was considered statistically significant [Begg and Zhang, 1994]. All the statistical analyses were performed with Medcalc version 16.8.1 (Medcalc Software, Ostend, Belgium) and SPSS Version 20.0.

Linkage Disequilibrium and haplotype analysis: Haplotype and linkage disequilibrium (LD) analysis was carried out using the SHEsis software (<http://202.120.31.177/myAnalysis.php>) [Shi and He, 2009]. The polymorphic sites which share a common chromosomal loci and belong to the same gene were evaluated for LD in order to obtain D' and r^2 values. Only those haplotype blocks were evaluated where the case and control frequency was more than 0.03, rest of the haplotypes were not considered for finding their correlation with lung cancer predisposition. The difference in the frequencies was calculated using global-chi square test. The association was estimated by finding out the odds ratio for each haplotype.

False discovery rate analysis: Multiple comparison corrections were carried out for genotypic combinations (SNP- SNP) using FDR (False discovery rate) method by online software based on Benjamini and Hochberg approach (<http://sdmproject.com/utilities/?show=FDR>). The adjusted p values are mentioned and FDR adjusted $p < 0.1$ was considered as significant [Benjamini and Hochberg, 1995].

MDR: The major goal of the present study was to uncover the hidden interactions taking place in the presence of multiple factors which cannot be determined by traditional statistical methods. Nonparametric approach was applied by using Multifactor dimensionality reduction method in order to analyze various gene-gene and gene to environment interactions that contribute maximally to the scenario of lung cancer predisposition. MDR software package version 0.5.1 which is available online has been applied to the sample data (<http://www.epistasis.org>). The aim of this method was to find main factor and make combinations with multiple factors such as genotypes which were frequently associated with cases than with controls. For search of the best n-factor model ($N=1,2,3,4,5$), the data were divided into training set which is 9/10 of data and a testing set which is 1/10 of the whole data. Training set is utilized for best model search and a high risk pattern associated with it depending upon the ratio of cases to control and the testing set

controls goodness of fit of the model. MDR method gives two scores for each n-factor model, a mean prediction error percentage and cross-validation consistency (CVC) frequency. Mean prediction error percentage consist of the section of individuals for whom an incorrect class prediction was made, therefore best model will have the lowest mean prediction error. CVC represents the number of times a particular combination of factor is identified in each possible testing set, therefore the best model will have maximum CVC. The complete test gets repeated 10 times using different random seeds to reduce the biasness in the results due to chance division of data into training and testing sets. To evaluated magnitude of prediction error and CVC frequency, we permuted the cases and control status in the data set and the analysis was repeated 500 times and then prediction error and CVC of each n-factor models was obtained under null hypothesis of no association. By comparing those data, p-values associated with each prediction error and CVC was obtained.

CART analysis: Another high order interaction between all the polymorphic variants was analyzed using CART (classification and regression tree analysis). CART was applied using the CART software (6.0, Salford Systems, San Diego, California). CART is a binary recursive-partitioning tool which results in a decision tree to identify the subgroup of subjects at higher risk. The recursive-partitioning algorithm starts at the first node and then uses statistical hypothesis-testing approach i.e. formal interference based recursive modeling to determine first locally optimal split and the each subsequent split of the data set, with each multiplicity adjusted p-values for tree growth control ($p < 0.05$). The process continues until the terminal node which keeps on splitting reach to subsequently significant nodes or reach the minimum size of subjects (< 10) in analysis. The tree obtained has various nodes and sub nodes depicting their role in modulating the disease susceptibility. The node with a lower case rate is used as a reference to find out adjusted odds ratio and 95% C.I for all other genotypes depicted in other nodes [Gupta *et al.*, 2016].

Overall survival and Hazards ratio analysis: For finding out the role of these polymorphisms in predicting the survival of the lung cancer patients, firstly the survival time (in months) of each patient was calculated. Survival time can be also be censored when there is a follow-up time but the event which is death of the patient in this case has not yet occurred or is not known to have

occurred. The concept of Kaplan Meier was used to find out the Median survival time (MST) and survival curve for each category in all the polymorphic sites. Univariate analysis called log rank test was used as a statistical method to compare the MST of patients based on the genotype of that particular variant. In order to whether these polymorphic variants can act as independent predictor variable, multivariate Cox proportional Hazard was used where Hazard ratio along with 95% C.I adjusted with various covariates such as age, sex, smoking, stage, KPS, ECOG and histology was obtained. Cox's proportional hazards model is analogous to a multiple regression model and enables the difference between survival times of particular groups of patients to be tested while allowing for other factors [Mahimkar *et al.*, 2012]. In this model, the response (dependent) variable is the 'hazard'. The hazard is the probability of dying that patients have survived up to a given point in time or the risk for death at that moment. Kaplan-Meier and Cox Regression method was also used in various subgroups stratified on the basis of sex, smoking, histology and performance status [Efron, 1988]. All the statistical analysis was performed with Medcalc version 16.8.1 (Medcalc Software, Ostend, Belgium).

STREE analysis: Further, high order gene interactions were taken into account by using the recursive partitioning approach. This approach produces the initial spilt taking into consideration the primary factor affecting the OS of patients. The software used for making the decision tree is STREE program (<http://c2s2.yale.edu/software/stree>). Log rank method was employed as the splitting criteria. The recursive method keeps on producing the offspring nodes till statistically insignificant spilt occurs. The resulting tree was binary, and each terminal node represented a subgroup of patients carrying a particular genotypic combination and hence shows a varied survival time and prognosis. HRs and their 95% CIs for each terminal node were calculated using the multivariate Cox proportional hazard model adjusted for sex, age, histology, KPS, ECOG, smoking and stage of tumor.

CHAPTER-5

RESULTS

5.1 Genomic DNA isolation

The isolated genomic DNA was qualitatively estimated by running 0.8% agarose gel electrophoresis. Fig 5.1(A) and 5.1(B) illustrates a good quality of intact genomic DNA isolated from the blood lymphocyte collected from lung cancer cases and healthy controls. Further, total estimation of nucleic acid was performed by measuring absorbance at 260 and 280nm.

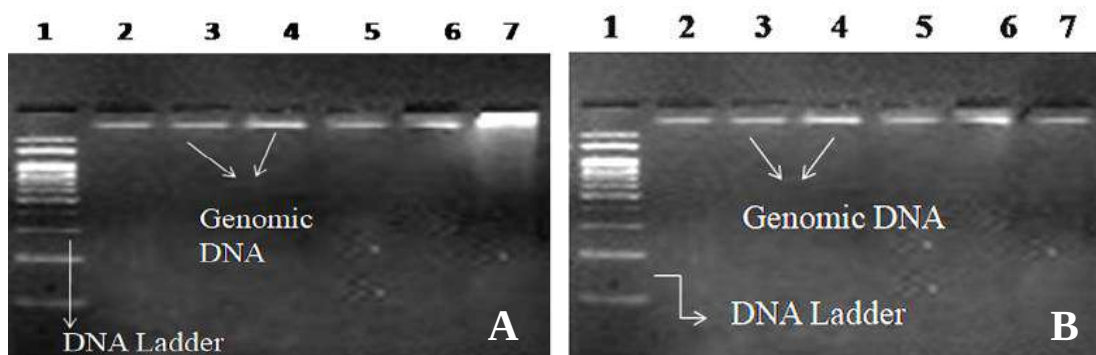


Fig 5.1: Agarose gel electrophoresis representing isolated genomic DNA from blood sample in (A) Cases (B) Controls

5.2 PCR amplification

PCR amplified samples were confirmed by 1.7% agarose gel electrophoresis for each SNP. Successful amplification was verified using a 100bp molecular weight marker. Fig 5.2 illustrates an example representing amplification of genomic DNA for *XPG Asp¹¹⁰⁴His* polymorphism. The size of amplified PCR product was 271bp.

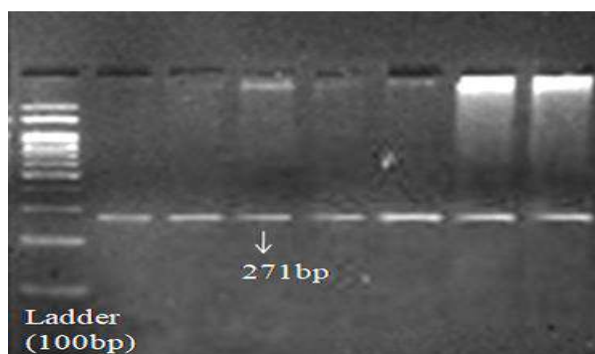


Fig 5.2: Representative example of agarose gel electrophoresis showing PCR amplification for *XPG Asp¹¹⁰⁴His*

5.3 Genotyping of Nucleotide Excision Repair genes

5.3.1 Genotyping of *XPC* variants

The restriction patterns of two polymorphic variants of *XPC* (*Lys⁹³⁹Gln* and *Arg⁴⁹⁹Gln*) were confirmed by using following methods. In case of *XPC Lys⁹³⁹Gln* the resulting fragments were separated on 2.5 % agarose gel electrophoresis. The size of the amplified PCR product was 281bp for *Lys⁹³⁹Gln* polymorphism. The variant C allele had a *PvuII* restriction site that resulted in bands of 150 and 131bp and the wild-type A allele lacked the restriction site and therefore produced a single 281bp band (Fig 5.3A). For the *XPC Ala⁴⁹⁹Val* gene, the size of the PCR product was 152bp. The digested products were separated on a 6% polyacrylamide gel and were silver stained. The wild-type (152bp) C allele produces fragments of 131 and 21bp and the mutant (T) allele produce a single 152bp fragment (Fig 5.3B).

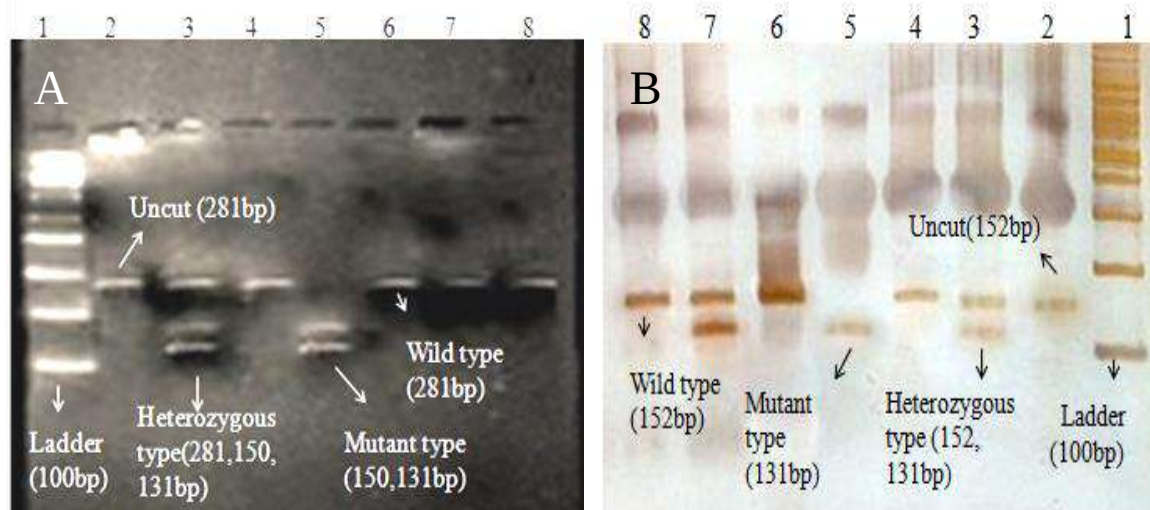


Fig 5.3: Restriction digestion pattern of XPC variants (A) *Lys939Gln* polymorphism (Wild type (AA) : 281bp; Mutant type (CC): 150bp and 131bp; Heterozygous type(AC) : 281bp,150bp and 131bp. Lane 1: marker; Lane 2: Uncut; Lane 4,6,7,8: AA genotype; Lane 5: CC genotype; Lane 3: AC genotype) (B) *Ala499Val* polymorphism (Wild type (GG): 152bp; Mutant type (AA): 131bp; Heterozygous type (GA): 152bp and 131bp. Lane 1:marker; Lane2; Uncut; Lane 3,7: CT genotype; Lane 5: TT genotype; Lane 4,6,8: CC genotype)

5.3.2 Genotyping of XPA variants

Restriction pattern of XPA variants (A23G and G709A) were confirmed on 8% polyacrylamide gel electrophoresis. The PCR amplification was confirmed by 1.7% agarose gel electrophoresis. The PCR product for A23G polymorphism was 158bp in size (Fig 5.4A). For A23G polymorphism, wild type (A) allele (*i.e.* 23A) produced single band at 158bp when digested with *Msp1* enzyme and variant (G) allele (*i.e.* 23G) resulted in two bands at 132 and 16bp. For G709A polymorphism, a wild genotype (G) allele (*i.e.* 709G) produced two bands at 127 and 16bp and polymorphic (A) allele (*i.e.* 709A) has one band representing the entire 143bp fragment when restriction digested with *Taq1* enzyme (Fig 5.4B).

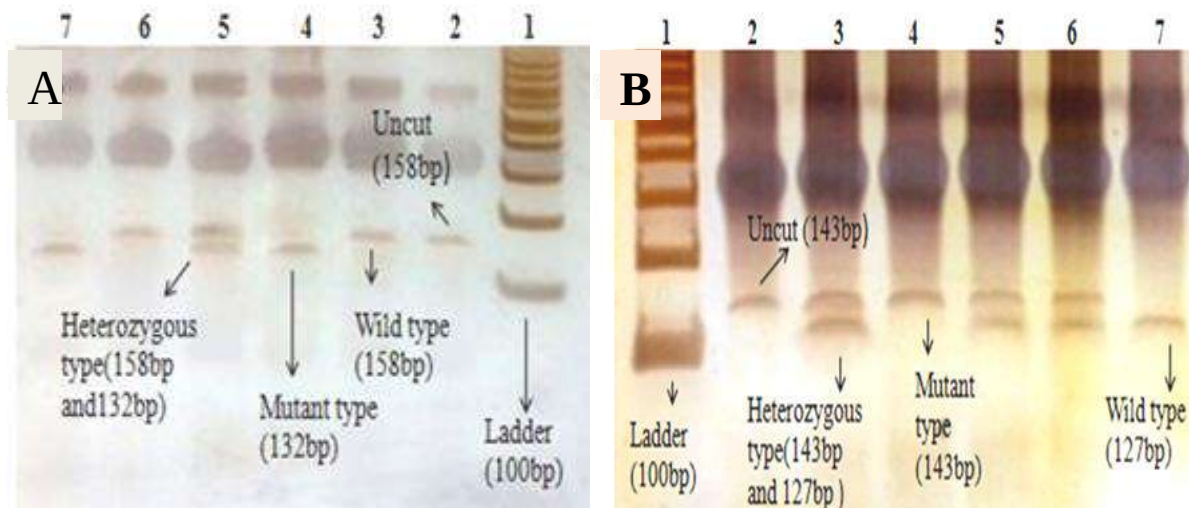


Fig 5.4: Restriction digestion pattern of XPA variants (A) A23G polymorphism (Wild type (AA) : 158bp; Mutant type (GG): 132bp and 26bp (the 26bp product was too small to resolve accurately); Heterozygous type (AG) : 158bp, 132bp and 26bp. Lane 1: marker; Lane 2: Uncut; Lane 3,6: AA genotype; Lane 4,7: GG genotype; Lane 5: AG genotype) (B) G709A polymorphism (Wild type (GG): 127bp and 16bp (16bp product was too small to resolve accurately); Mutant type (AA): 143bp; Heterozygous type (GA): 143bp, 127bp and 16bp. Lane 1: marker; Lane 2: Uncut; Lane 3,5,6: GA genotype; Lane 4: AA genotype; Lane 7: GG genotype)

5.3.3 Genotyping of XPD variants

Three polymorphic variants of XPD gene (*Lys*⁷⁵¹*Gln*, *Asp*³¹²*Asn* and *Arg*¹⁵⁶*Arg*) were genotyped. In case of XPD *Lys*⁷⁵¹*Gln* and *Asp*³¹²*Asn* the resulting fragments were separated on 2.5 % agarose gel electrophoresis. The size of the amplified PCR product was 436bp for *Lys*⁷⁵¹*Gln* polymorphism. The variant C allele had a restriction site at 227, 146 and 63bp when treated with *Pst*I restriction enzyme. Restriction pattern of 290 and 146bp represented wild-type A allele (Fig 5.5A). For the *Asp*³¹²*Asn* gene, the size of the PCR product was 751bp. The wild-type G allele produces fragments of 507, 244 and 33bp and the mutant (A) allele produce bands at 474, 244 and 33bp fragment when digested with *Sty*I restriction enzyme (Fig 5.5B). For *Arg*¹⁵⁶*Arg* polymorphism, the PCR product was obtained of size 288bp and then the products were digested using *Hinf*I enzyme. The digested products were separated on an 8% polyacrylamide gel and were silver stained. The wild type genotype (CC) was detected at 288bp. The variant genotype (CC) produced fragments at 288, 206 and 82bp (Fig 5.5C).

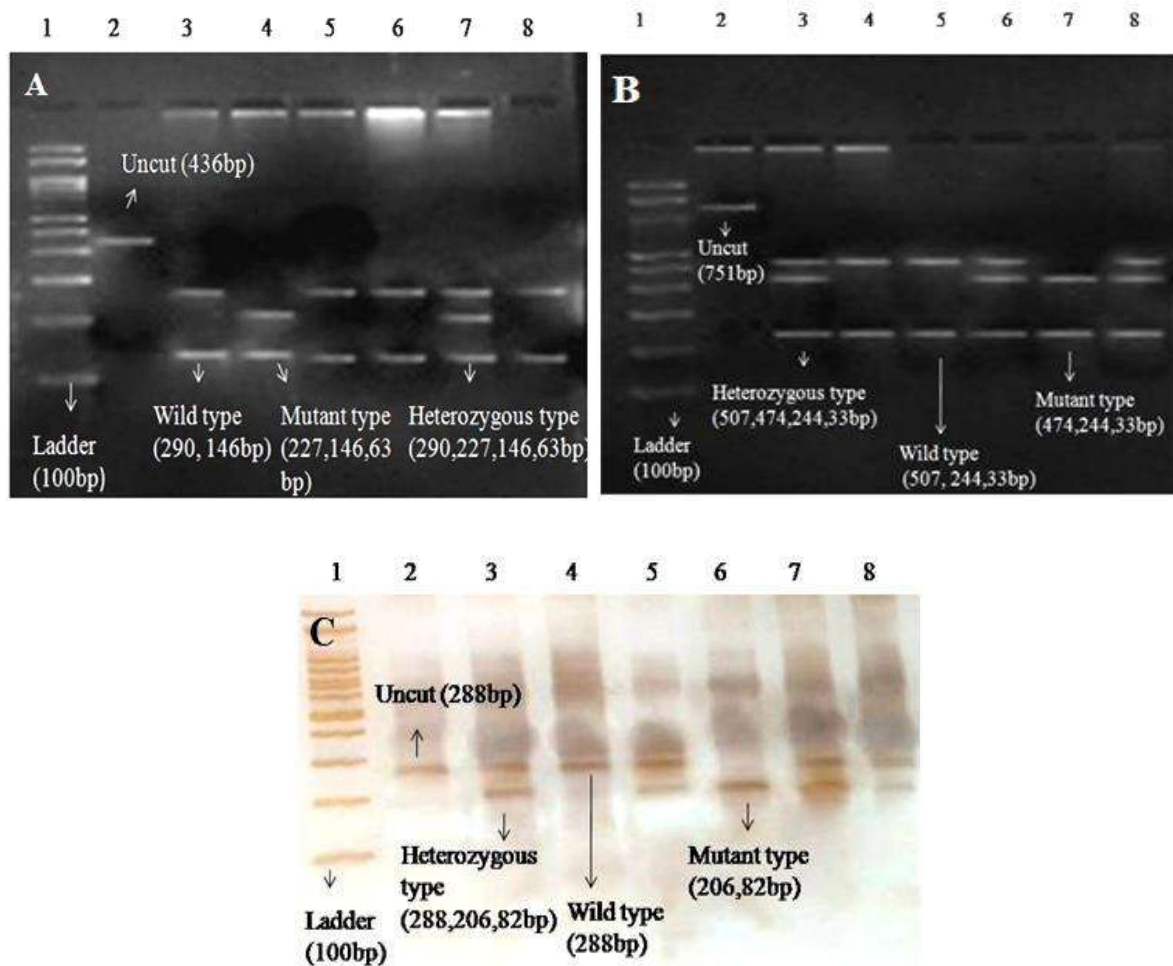


Fig 5.5: Restriction digestion pattern of XPD variants (A) *Lys*⁷⁵¹*Gln* polymorphism (Wild type (AA) : 290 and 147bp; Mutant type (CC): 227bp, 146bp and 63bp (the 63bp product was too small to resolve accurately); Heterozygous type (AC) : 290bp, 227bp, 146bp and 63bp. Lane 1: marker; Lane 2: Uncut; Lane 3,5,6,8: AA genotype; Lane 4: CC genotype; Lane 7: AC genotype) (B) *Asp*³¹²*Asn* polymorphism (Wild type (GG): 507bp, 474bp and 33bp (33bp product was too small to resolve accurately); Mutant type (AA): 474bp, 244bp, 33bp; Heterozygous type (GA): 507bp, 474bp, 244bp and 33bp. Lane 1: marker; Lane 2: Uncut; Lane 3,6,8: GA genotype; Lane 7: AA genotype; Lane 4,5: GG genotype) (C) *Arg*¹⁵⁶*Arg* polymorphism (Wild type (CC) : 288bp; Mutant type (AA): 206bp and 82bp (the 82bp product was too small to resolve accurately); Heterozygous type (AC) : 288bp, 206bp and 82bp. Lane 1: marker; Lane 2: Uncut; Lane 6: AA genotype; Lane 4: CC genotype; Lane 3,5,7,8: CA genotype).

5.3.4 Genotyping of XPG variants

Restriction pattern for two XPG polymorphic variants (*Phe*⁶⁷⁰*Leu* and *Asp*¹¹⁰⁴*His*) were confirmed on 2.5% agarose gel electrophoresis. For *Phe*⁶⁷⁰*Leu* polymorphism, mutant type (A) allele produced a single fragment at 400bp and wild (CC) type genotype produced two bands at 125 and 275bp when treated with *ECor1* enzyme (Fig 5.6A). For *Asp*¹¹⁰⁴*His* polymorphism,

when digested using *Nla III* restriction enzyme wild genotype (G) allele produced a single fragment at 271bp while mutant genotype (A) allele resulting bands at 227 and 33bp (Fig 5.6B).

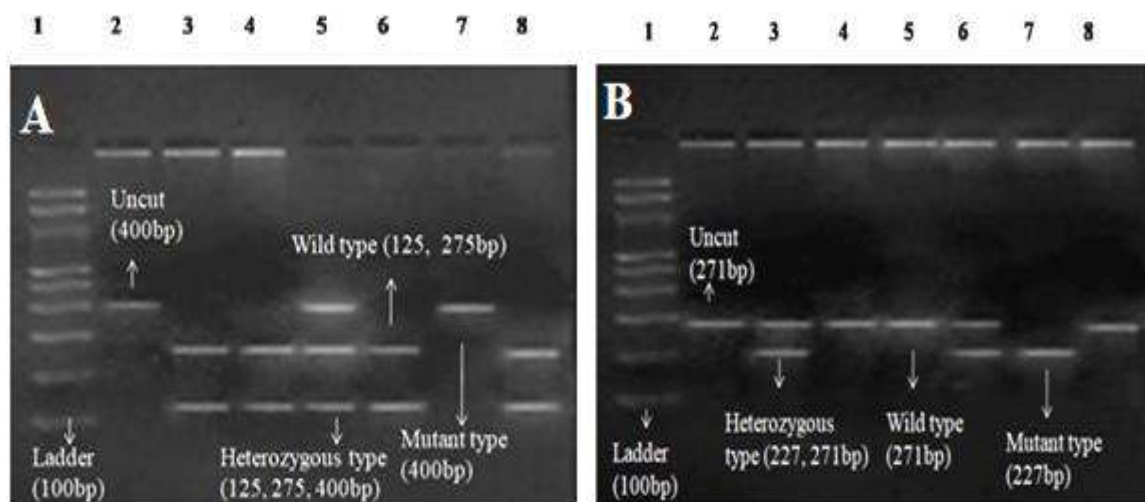


Fig 5.6: Restriction digestion pattern of XPG (A) *Phe*⁶⁷⁰*Leu* polymorphism: Wild type (CC) : 125, 275bp; Mutant type (AA): 400bp ;Heterozygous type(CA) : 125bp, 275bp and 400bp. Lane 1: marker; Lane 2: Uncut; Lane 3,4,6,8: CC genotype; Lane7: AA genotype; Lane 5: CA genotype) (B) Restriction digestion pattern of *Asp*¹¹⁰⁴*His* polymorphism: Wild type (GG): 271bp; Mutant type (CC): 227bp (33bp product was too small to resolve accurately); Heterozygous type (GC): 227bp and 271bp. Lane 1:marker; Lane2; Uncut; Lane3,6: GC genotype; Lane 7: CC genotype; Lane 4,5,8: GG genotype)

5.3.5 Genotyping of XPF variants

Three polymorphic variants of *XPF* gene (673 C>T, 11985A>G and *Arg*⁴¹⁵*Gln*) were genotyped. Restriction pattern of *XPF* 673C>T and 11985A>G were confirmed using 8% polyacrylamide agarose gel electrophoresis and were silver stained whereas resulting fragments for *Arg*⁴¹⁵*Gln* polymorphism were separated on 2.5% agarose gel electrophoresis. The size of the amplified PCR product was 137bp for 673C>T polymorphism. The variant T allele had a restriction site at 137bp while fragments at 114 and 23bp represented wild-type C allele when digested with *EcoRI* restriction enzyme (Fig 5.7A). For the 11985A>G polymorphism, PCR amplification confirmed fragment at 129bp. The variant genotype GG produced fragments at 104 and 25bp after restriction digestion with *RsaI*. The wild-type (A) allele produce a band at 129bp (Fig 5.7B). For *Arg*⁴¹⁵*Gln* polymorphism, the PCR product was obtained of size 380bp. After

restriction digesting with *Xmn1*, the wild type genotype (GG) was detected at 380bp. The variant genotype (AA) produced fragments at 284 and 96bp (Fig 5.7C).

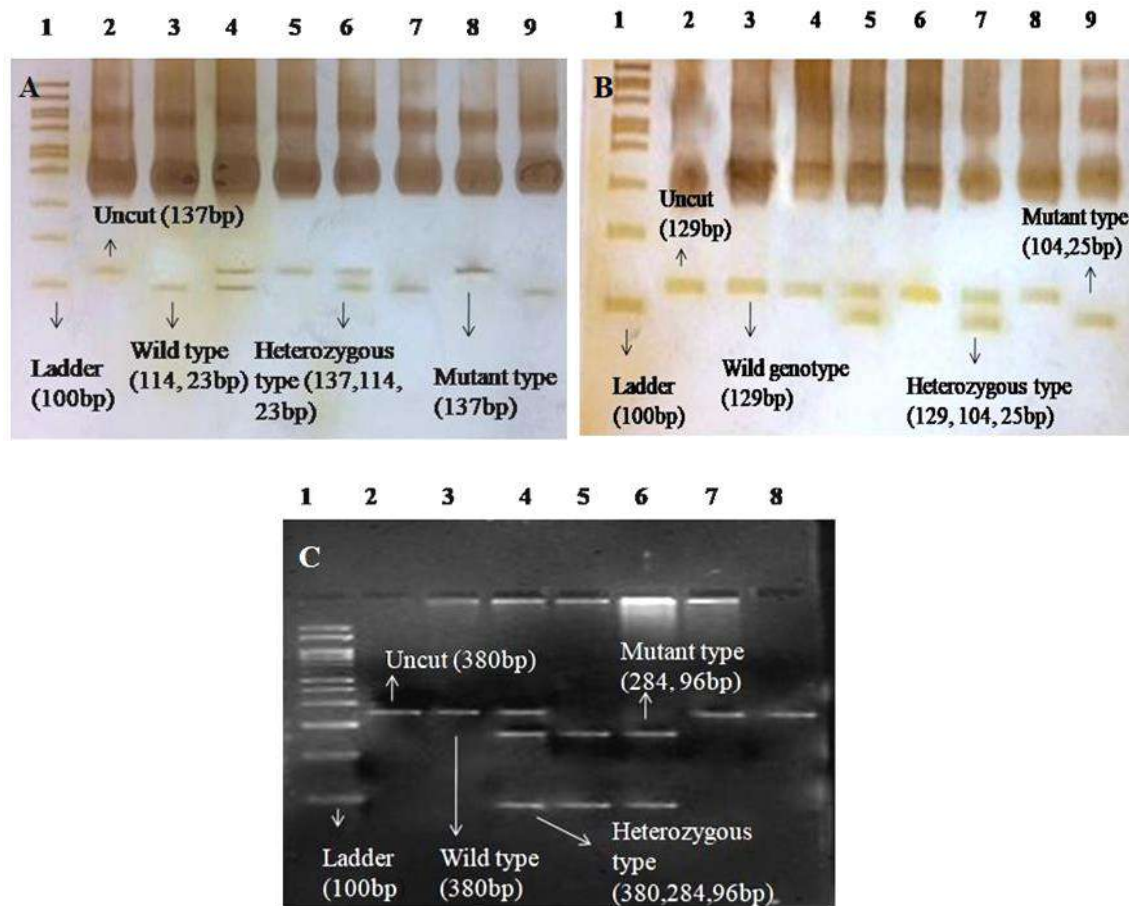


Fig 5.7: Restriction digestion pattern of *XPF* variants (A) 673C>T polymorphism (Wild type (CC) : 114 and 23bp (the 23bp product was too small to resolve accurately); Mutant type (CC): 227bp, 146bp and 63bp; Heterozygous type (AC) : 290bp, 227bp, 146bp and 63bp. Lane 1: marker; Lane 2: Uncut; Lane 3,5,6,8: AA genotype; Lane 4: CC genotype; Lane 7: AC genotype) (B) 11985 A>G polymorphism (Wild type (AA): 129bp; Mutant type (GG): 104 and 25bp; Heterozygous type (AG): 129, 104 and 25bp. Lane 1: marker; Lane 2: Uncut; Lane 5,7: AG genotype; Lane 2,3,4,6,8: AA genotype; Lane 9: GG genotype) (C) Arg⁴¹⁵Gln polymorphism (Wild type (GG) : 380bp; Mutant type (AA): 284bp and 96bp; Heterozygous type (GA) : 380bp, 284bp and 96bp. Lane 1: marker; Lane 2: Uncut; Lane 3,7,8: GG genotype; Lane 5,6: AA genotype; Lane 4: GA genotype).

5.4 Epidemiological and demographical statistics of North Indian lung cancer patients

5.4.1 Regional distribution of lung cancer patients

The blood samples were collected from Department of Pulmonary Medicine, PGIMER, Chandigarh. It is a specialized multi-facility government hospital where people from every part of Northern India come for treatment when referred from the locally situated hospitals. The study included 370 lung cancer cases reported from Punjab, Haryana, Himachal Pradesh, Chandigarh, Uttarakhand, Jammu and Kashmir. From total number of patients considered for study maximum number of patients were reported 103 (27.83%) from Punjab, 93 (25.13%) from Himachal Pradesh, 75 (20.27%) from Haryana and 44 (11.89%) from Uttar Pradesh. 7.29%, 4.32% patients who came for treatment were from Chandigarh and Uttarakhand respectively. Patients from Jammu and Kashmir and Chattisgarh were 5(1.35%), 1(0.27%) respectively in number as shown in fig 5.8.

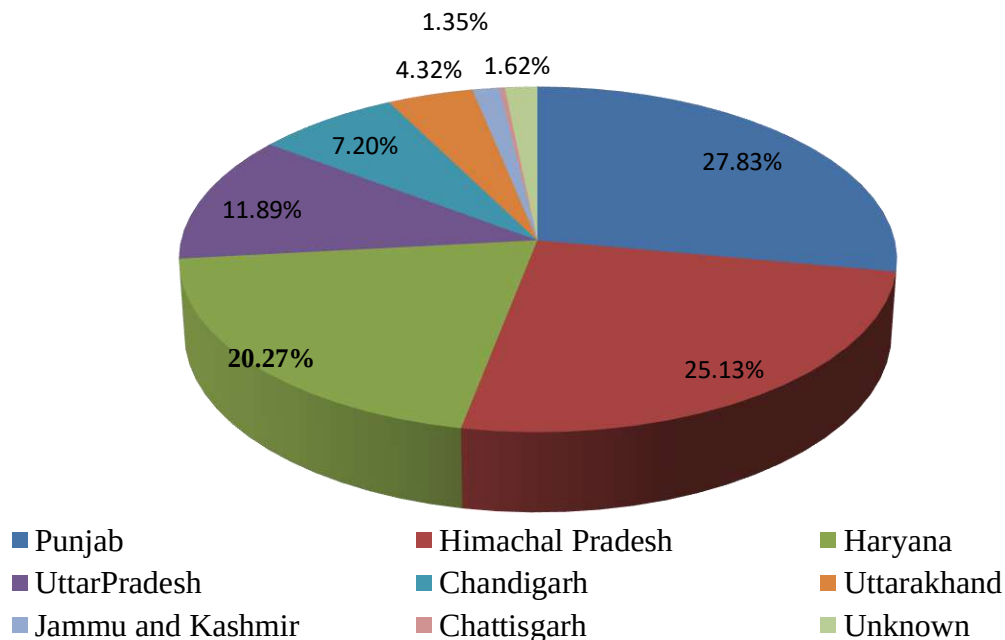


Fig 5.8: Regional distribution of the North Indian lung cancer patients

The subjects analyzed for the regional distribution were further segregated on the basis of gender. Of the total population, there were 319 (86.21%) males and 51 (13.78%) were females (Fig 5.9). In controls, the same distribution of males (86.21%) and females (13.78%) were present. The maximum numbers of males were present in the region of Punjab (28.80%), Himachal Pradesh (23.82%) and Haryana (21.31%) whereas for females the maximum number of cases came from Himachal Pradesh (33.33%), Punjab (21.56%), Chandigarh (15.68%) and Haryana (13.72%). Total number males reported for treatment to hospital from Uttar Pradesh, Chandigarh, Uttarakhand, Jammu and Kashmir were 12.85%, 5.95%, 4.38%, 1.56% respectively. Female patients came from the parts of Uttar Pradesh, Uttarakhand and Chattisgarh were 5.88%, 3.92%, 1.96% respectively.

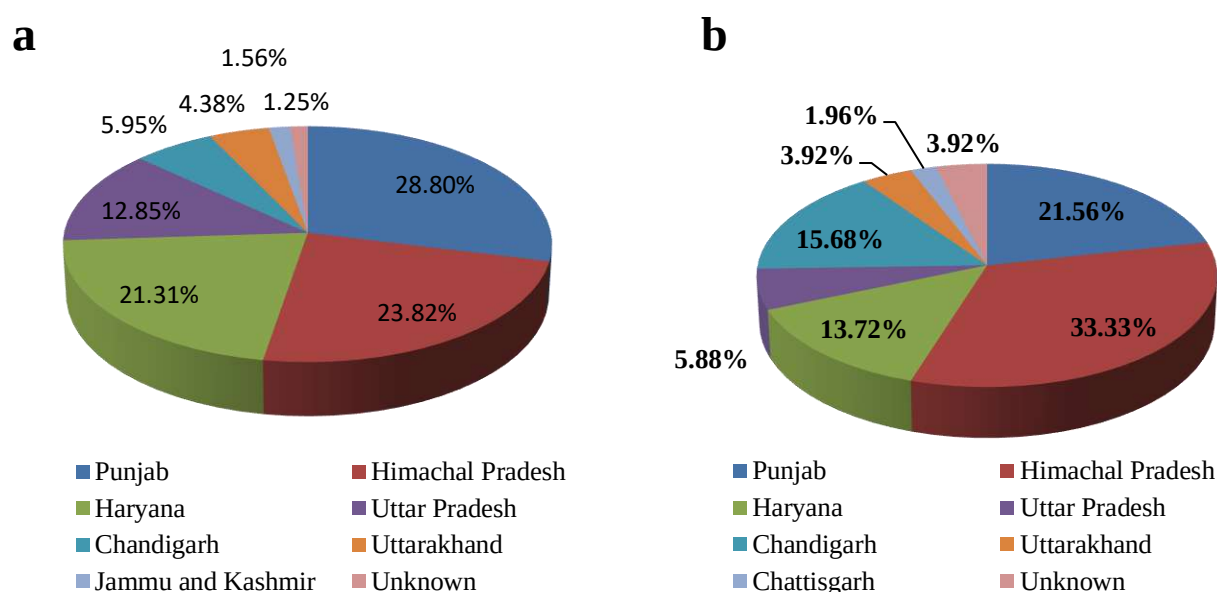


Fig 5.9: Regional distribution of Lung cancer on the basis of gender (a) Males (b) Females

5.4.2 Distribution of study subjects on the basis of smoking status

The present epidemiological data was segregated on the basis of the smoking status for all cases and controls. The study population was dichotomized on the basis of their smoking habit as smokers and non-smokers. As the study was majorly targeting smoking population for lung cancer patients, therefore, the study included a higher number of smokers in comparison with non-smokers. Similarly, in order to nullify the sample bias attempts were made to match controls with cases as per the smoking status. 81.89% (303) and 73.29% (271) were smokers among cases and controls while 18.18% (67) and 26.75% (99) reported as non-smokers as shown in fig 5.10.

The quantitative estimation of smoking was calculated in terms of pack years for each subject. The mean pack years were higher in the case group when compared to the control group (24.97 ± 35.04 vs 17.88 ± 19.45 ; $p < 0.0001$). This shows a higher prevalence of pack years among smokers as compared to non-smokers.

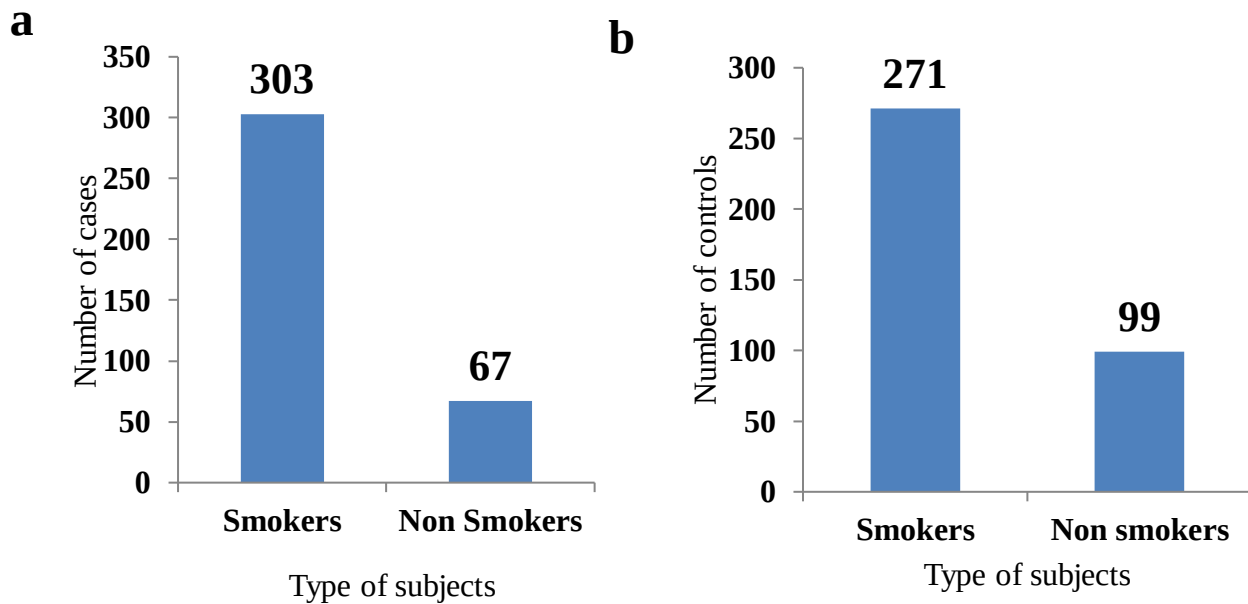


Fig 5.10: Representative distribution of smoking status in the study population (a) cases (b) controls

The data further elaborates different type of smoking habits prevailing in the country which includes beedi, cigarette and a combined intake of both. The majority of the population were found as a beedi smokers was 85.14% cases and 91.85% controls where only 5.28% cases and 9.59% controls were cigarette smokers. Smoking pattern of the study subjects was also analyzed on the basis of their gender. Smokers percentage among males in cases was 89.65% as compared to those in control group (84.63%) as shown in fig 5.11.

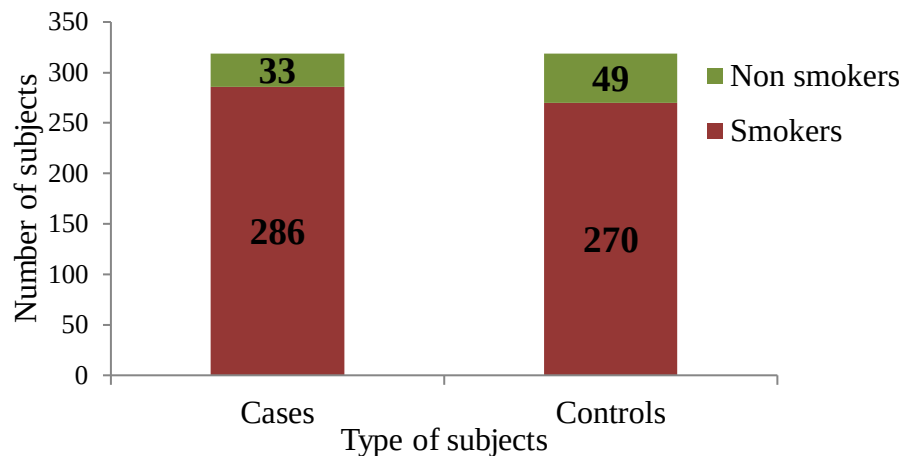


Fig 5.11: Graph showing distribution of male population according to smoking status

5.5.3 Distribution of subjects on the basis age

Age is one the important demographic parameter associated with the case-control studies. The mean age of cases was 58.11 ± 10.44 years. The range was estimated from 29 to 86 years. The median age was calculated as 59 years. The cases were matched with controls with a difference of ± 10 years. Mean age for controls was 53.83 ± 10.18 years with a range of 23-83 years. Median age among controls was 53 years. The distribution of age as a parameter on the basis of gender has been represented in Fig 5.12. A slightly higher percentage of males were reported to have age ≤ 59 years as compared to those having age less than 59 years (82.4% vs 89.8%) among lung cancer patients. Similarly, 10.1 % (19) females were found to have age ≤ 59 years and 17.5% (32) were reported with age less than 59 years.

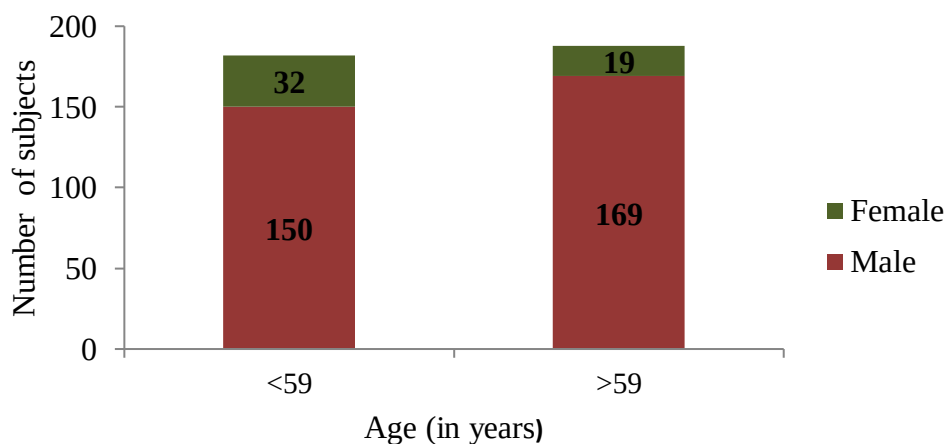


Fig 5.12: Age wise gender distribution of lung cancer patients

5.4.4 Distribution of lung cancer patients on the basis of TNM staging

Clinico-pathological parameters of the cases were collected from the records present at the hospital. The patients were classified as per the clinical stage of tumor as stage I, II, III and IV. Majority of the patients recorded for the study were in stage III and IV while a low number of patients were reporting in their stage I and II of the tumor (Fig 5.13). 0.81% (3) and 2.70% (10) patients were in stage I and II. 45.94% (175) and 44.05% (163) patients were in stage III and IV of cancer. Lung cancer patients were further stratified on the basis of various TNM staging parameters such as primary tumor extension (T), lymph node involvement (N) and metastasis (M). Lung cancer subjects with T1 and T2 tumor size constituted 4.32% (16) and 11.35% (42) while T3 and T4 were 22.97% (85) and 51.62% (191) of the total subjects. The frequencies for lymph node involvement were classified as: N0 with 16.3% of the patients whereas N1, N2, N3 with 12%, 37%, and 24% respectively. Lastly for metastatic involvement, the available data includes 50.54% (187) cases which did not have any metastasis (M0) while 42.70% (158) cases had involvement of distant metastasis (M1).

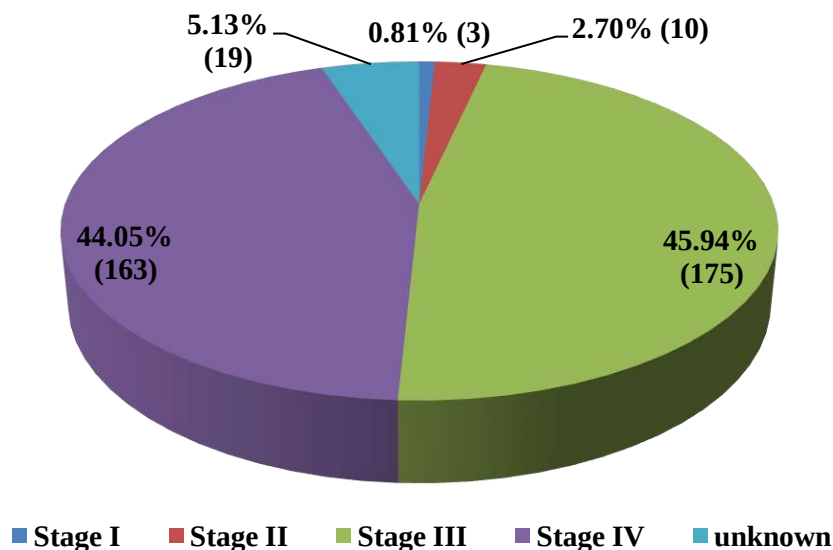


Fig 5.13: Distribution of Lung cancer patients according to clinical stage in North Indians

5.4.5 Stratification of lung cancer patients on the basis of histology

Lung cancer had been extensively classified into three histological subtypes: Adenocarcinoma (ADCC), Squamous cell carcinoma (SQCC) and Small cell lung cancer (SCLC). The data regarding the histological subtypes of the patient was collected from the medical records of the hospital. In the present study, 135 (36.48%) were diagnosed with SQCC, 120 (32.43%) with ADCC and 111(30%) with SCLC (Fig 5.14). The number of patients detected with SQCC was highest as compared to patients detected with SCLC and ADCC depicting it to be the highest type of occurring cancer among North Indians.

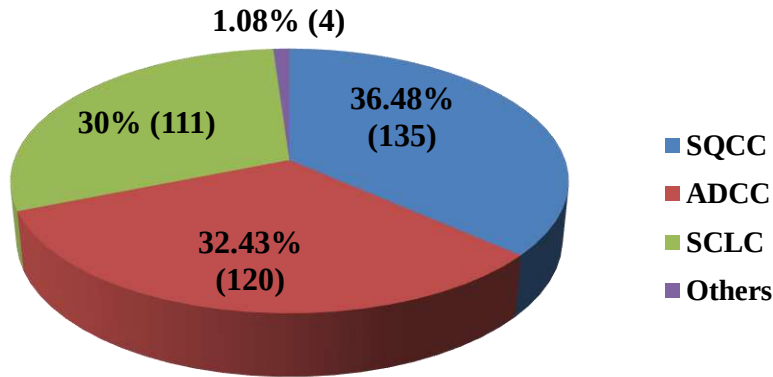


Fig 5.14: Distribution of lung cancer cases as per their histological subtype

5.4.6 Distribution of lung cancer patients on the basis of performance status

To evaluate performance of lung cancer patients while seeking for treatment in order to estimate prognosis of the disease, the patients were classified on the basis of two methods that is Karnofsky performance status (KPS) and Eastern Cooperative Oncology group performance status (ECOG-PS). These methods define the functional state of the patients in which they reported to hospital. There were 37.02% (136) patients with KPS score of 100-90, 41.89% (155) patients with score of 80-70 and 12.70% (47) patients having performance status ≤ 60 as indicated in fig 5.15. Similarly, when patients were stratified as per their ECOG performance

status, it was observed that 14.86% (55) of patients had ECOG status group (0), 32.16% (119) had ECOG status (1), 33.78% (125) patients had ECOG status of (2) and 12.97% (48) had ECOG status of (3-4) as shown in fig 5.16.

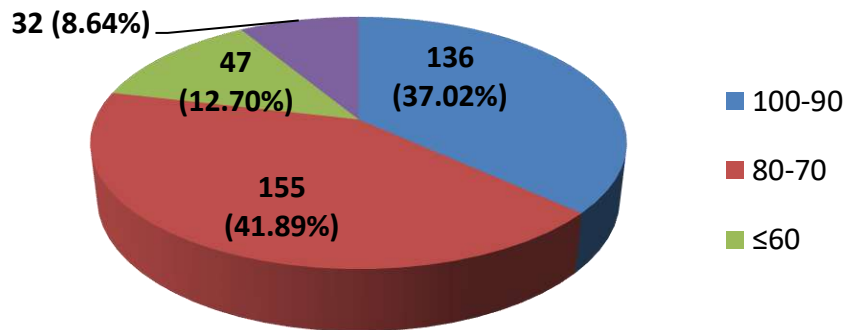


Fig 5.15: Pie chart illustrating number of cancer patients reported with different Karnofsky performance status

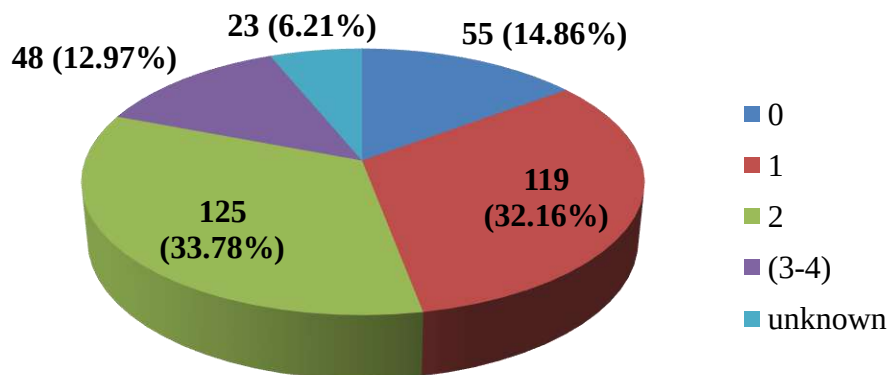


Fig 5.16: Pie chart illustrating number of cancer patients reported with ECOG performance status

5.4.7 Distribution of lung cancer patients as per different chemotherapy regimen

As per the histological subtype (ADCC, SQCC, SCLC) of the patient different treatment regimens are given to the patients. In the present study, a platinum based doublet chemotherapy

was administered to the patients which included platinum based drug along with another drug for the treatment. Out of 370 lung cancer cases, maximum number of patients, that is 87(23.51%) were treated with docetaxel-cis/carboplatin regimen. 86 (23.25%) and 78 (21.08%) were treated with irinotecan-cis/carboplatin and pemetrexed-cis/carboplatin regimen respectively. Only very few patients were treated with placitacel-cis/carboplatin (4.59%) and gemcitabine-cis/carboplatin (0.81%) regimen. Treatment data for 92 (24.86%) patients was not available.

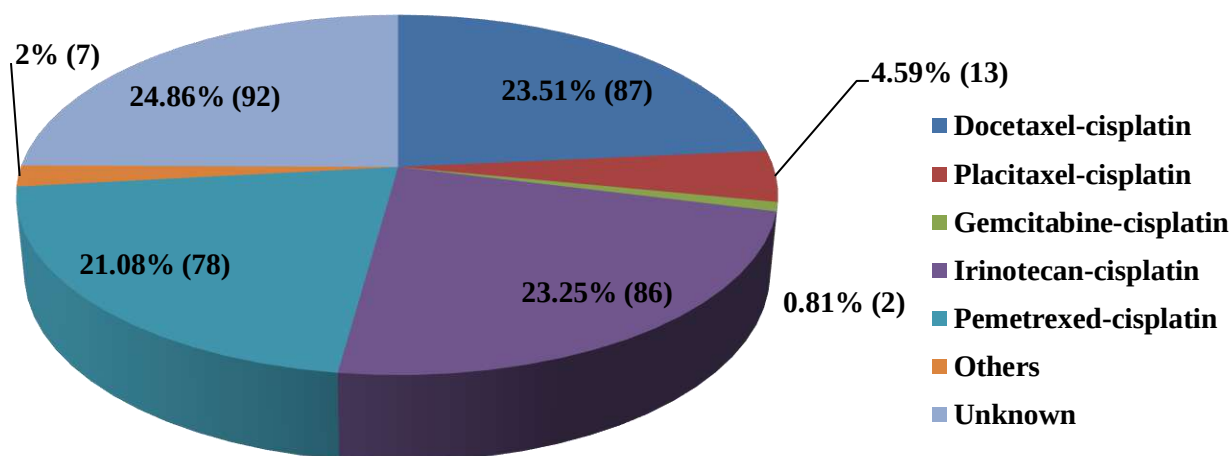


Fig 5.17: Distribution of lung cancer patients according to their treatment regimen

5.5 Statistical analysis for NER genes

5.5.1 Demographic characteristics

The sample population was characterized on the basis of various demographic characteristics such as age, gender, smoking parameter, pack years, histological subtypes, TNM staging and other clinical parameters as summarized in table 5.1. A total of 740 subjects were recruited which included 370 cases and 370 healthy controls. Mean age of lung cancer cases was 58.11 ± 10.44 (range 29-86) whereas mean age of controls was 53.83 ± 10.18 (range 22-83). There were 319 (86.21%) males and 51(13.78%) females in both the control and case groups respectively. Hence, no significant difference had been observed between the distribution of males and females, suggesting adequate matching ($p=0.91$). There were 81.89% cases identified

as smokers and 18.10% were non-smokers when compared with controls having 74.59% of smokers and 25.4% non-smokers. Smoking status was further elaborated in terms of pack years. Pack years were significantly higher amongst cases (24.97 ± 35.04) in comparison with controls (17.88 ± 19.45 ; $p=0.0001$). From the total of 370 cases, 135(36.48%) patients were classified with Squamous cell carcinoma (SQCC), 120 (32.43%) with Adenocarcinoma (ADCC) and 111(30%) with SCLC. Out of 370 cases, TNM staging data were available for 351 (94.86 %) patients. The TNM breakup of the 351 cases were as follows: Stage I: 3(0.81%), II: 10(2.7%), III: 175(47.29%), IV: 163 (44.05%). Lung cancer subjects with T1 and T2 tumor size constituted 4.32% (16) and 11.35% (42) while T3 and T4 were 22.97% (85) and 51.62% (191) of the total subjects. The frequencies for lymph node involvement were classified as: N0 with 16.3% of the patients whereas N1, N2, N3 with 12%, 37%, and 24% respectively. Lastly, the available data includes 49.18% cases which did not have any metastasis (M0) while 44.05% cases predict involvement of distant metastasis (M1).

5.5.2 Clinical characteristics of lung cancer patients evaluated for detecting the prognostic relevance of polymorphism of different NER genes

Various demographic and clinical parameters in the present study are summarized in table 5.2. A total of 311 lung cancer patients were considered for the survival analysis in the present study. The study population was characterized for various epidemiological variables such as age, gender and smoking habits. The mean age of the study subjects came out to be 57.73 ± 10.39 years. Out of 311 subjects, 87.45% were males and the remaining 12.54% were females. As smoking is an established cause for occurrence of lung cancer, so majority of sample population comprised of smokers (82.95%), while 17.04% were characterized as non-smokers. Different clinical parameters such as histology, performance status and chemotherapeutic regimen have been also listed in table1. The patients are classified on the basis of their histology into adenocarcinoma (32.79%), squamous cell carcinoma (37.62%) and small cell lung carcinoma (28.03%). 48.33% were diagnosed with stage III while 44.37% were in stage IV of cancer. In order to evaluate the performance status of the patients, two criteria's namely ECOG and KPS score were used. Out of 311 patients, 39.54% showed KPS score of 100-90, 44.38% had KPS score of 80-70 while 14.46% patients harbored KPS score (≤ 60). From 311 patients, 47.8% and 50.80% had an ECOG of 0-1 and 2-4. Further, we stratified the sample population on the basis of their chemotherapeutic regimen, highest percentage of patients were treated with docetaxel and

Table 5.1: Demographic characteristics among cases and control patients in North Indian population			
Variable	Cases, n (%) N=370	Controls, n (%) N=370	<i>p</i>^a
Age (years) Mean ± SD Range	58.11±10.44	53.83±10.18	<0.0001
Gender Male Female	319 (86.21%) 51 (13.78%)	319 (86.21%) 51 (13.78%)	0.91
Smoking Status Smokers Non – Smokers	303 (81.89%) 67 (18.10%)	271 (73.24%) 99 (26.75%)	0.0063
Pack Years Mean ± SD	24.97 ± 35.04	17.88 ± 19.45	<0.0001
Histological Types SQCC ADCC SCLC Others Unknown	135 (36.48%) 120 (32.43%) 111 (30%) 4 (1.08%) 0 (0.0)		
TNM Staging I II III IV Unclassified	3 (0.81%) 10 (2.70%) 175 (45.94%) 163 (44.05%) 19 (5.13%)		
Tumor Size Tx T1 T2 T3 T4 Unknown	11 (2.97%) 16 (4.32%) 42 (11.35%) 85 (22.97%) 191 (51.62%) 25 (6.75%)		
Lymph Node Involvement N0 N1 N2 N3 UNKNOWN	46 (12.43%) 36 (9.72%) 157 (42.43%) 106 (28.64%) 25 (6.75%)		
Metastasis M0 M1 Unknown	187 (50.54%) 158 (42.70%) 25 (6.75%)		

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

Table 5.2: Patients characteristics and clinical parameters		
Patient characteristics	Total (N)	Cases n (%)
Age (in years) Mean ± SD Range	311	57.73 ± 10.59
		28-86
	Median age	131
	180	>60
Male	311	272 (87.45%)
Female		39 (12.54%)
Smokers	311	258 (82.95%)
Non-smokers		53 (17.04%)
Histology		
ADCC	311	102 (32.79%)
SCLC		117 (37.62%)
SQCC		90 (28.03%)
Others		2 (0.6%)
Overall Survival		
Dead	311	256 (82.31%)
Alive		55 (17.68%)
TNM Staging		
III	311	151 (48.55%)
IV		138 (44.37%)
Others		22 (7.07%)
Performance Status		
KPS(100-90)	311	123 (39.54%)
KPS(80-70)		138 (44.37%)
KPS(≤60)		45 (14.46%)
unknown		5 (1.60%)
ECOG	311	
ECOG(0-1)		148 (47.8%)
ECOG(2-4)		158 (50.80%)
Unknown		5 (1.60%)
Chemotherapy Regimen		
Docetaxel/ Cisplatin	237	76 (32.06%)
Irinotecan/Cisplatin		70 (29.53%)
Paclitaxel/ Cisplatin		13 (5.48%)
Pemetrexed/Cisplatin		69 (29.11%)
Gemacitabine/Cisplatin		2 (0.84%)
Others		7 (2.95%)

Abbreviations: ADCC= Adenocarcinoma; SCLC=Small cell lung cancer; SQCC=Squamous cell carcinoma; KPS=Karnofsky performance status; ECOG=Eastern cooperative and oncology group; MST=Median survival time; (a)= Hazard ratio derived using Kaplan Meier survival curve; (b) p-value derived after applying Cox regression model; All *p*-values are two – sided. *p*< 0.05 was considered statistically significant

cisplatin (32.06%) doublet combination. 29.53% patients were given irinotecan and cisplatin doublet combination while 29.11% treated with pemetrexed and cisplatin doublet combination as shown in table 5.2.

5.5.3 XPC gene

5.5.3.1 Genetic and allelic distribution of XPC polymorphism

Table 5.3 demonstrates the allelic and genotypic frequency distribution of the two XPC polymorphic variants, *Lys⁹³⁹Gln* and *Ala⁴⁹⁹Val*. Genotypic frequencies of XPC *Lys⁹³⁹Gln* among controls (n=370; $\chi^2=0.002$, df=1, p=0.95) and cases (n=370, $\chi^2=0.222$; df=1; p=0.63) were in Hardy-Weinberg equilibrium, suggesting no sample biasness for control and cases. However, the genotypic frequencies for XPC *Ala⁴⁹⁹Val* polymorphism for both control (n=370; $\chi^2=23.2$; df=1; p=0.00) and cases (n=370; $\chi^2=32.71$; df=1; p=0.00) were not in Hardy-Weinberg Equilibrium, suggesting population stratification, migration and random mating within the sample set. The wild genotype (AA) for XPC *Lys⁹³⁹Gln* gene in lung cancer subjects was found to be less prevalent as compared to the controls (35.94 vs 44.86%). However as shown in table 5.3 the frequency of the mutant genotype (CC) for XPC *Lys⁹³⁹Gln* was 10.08% for cases as compared to controls (17.02%). Therefore, a significant difference among the genotypic distribution between cases and controls was observed ($\chi^2=9.07$, df=2, p=**0.01**). Minor allelic frequency (MAF) calculated for XPC *Lys⁹³⁹Gln* in controls and cases was 0.33, 0.40 respectively, suggesting a significant difference among the sample population. In case of XPC *Ala⁴⁹⁹Val*, the frequency of the mutant genotype (TT) was more represented in the control group in comparison with cases (6.21% vs 1.89%) whereas wild genotype (CC) was more presented in cases (45.67%) than controls (37.29%). Overall there was a significant difference in the distribution of genotypes between the controls and cases ($\chi^2=12.22$, df=2, p=**0.002**). The MAF calculated for controls and cases was 0.34 and 0.28 respectively, suggesting that both the variant allele was found to be more pronounced in the controls than cases.

Table 5.3 further showed relationship of XPC polymorphic variants towards lung cancer susceptibility. Adjusted ORs and 95% CIs were calculated using the logistic regression analysis using the more common genotype as the reference group. Homozygous wild genotype (AA) (*Lys⁹³⁹Gln*) and CC (*Ala⁴⁹⁹Val*) were considered as reference genotypes with which the variant

genotypes were compared. In case of the *XPC Lys⁹³⁹Gly* polymorphism, it was observed that individuals carrying both the mutant alleles (CC) had a 2-fold elevated risk for lung cancer which was found to be significant (OR=2.30; 95%CI:1.41-3.73; $p=0.0007$). All the bold values in the tables represent significant p-values. The subjects with heterozygote genotypes (AC) showed a less pronounced risk as compared to the mutant genotype (OR=1.44; 95%CI: 1.04-1.99; $p=0.02$). When both the genotypes were combined as a single genotype (AC+CC) and compared with the reference genotype a significant association towards lung cancer risk was also observed (OR=1.60; 95%CI: 1.18-2.17; $p=0.002$). For *Ala⁴⁹⁹Val* polymorphism, mutant genotype depicted a significant protective effect when compared with the reference genotype (OR=0.25; 95%CI: 0.10-0.63; $p=0.003$). The combined genotype (CT+TT) also illustrated a significant protective effect among lung cancer subjects as compared to wild type genotype (OR=0.72; 95%CI: 0.53-0.97; $p=0.03$).

Table 5.3: Association of XPC polymorphism and its association with lung cancer susceptibility				
OVERALL LUNG CANCER				
<i>Lys⁹³⁹Gln</i>	Controls n (%) N=370	Cases N (%) N=370	AOR (95% CI) ^a	<i>p</i>^b
AA	166(44.86)	133(35.94)	1.00 (Reference)	
AC	164(44.32)	174(47.02)	1.44 (1.04-1.99)	0.02
CC	40(10.8)	63(17.02)	2.30 (1.41-3.73)	0.0007
AC+CC	204(55.13)	237(64.05)	1.60 (1.18-2.17)	0.002
A Allele	496(67.03)	440(59.46)		
C Allele	244(32.97)	300(40.54)		
MAF	0.33	0.40		
<i>Ala⁴⁹⁹Val</i>	Controls n (%) N=370	Cases n(%) N=370	AOR (95% CI) ^a	<i>p</i>^b
CC	138(37.29)	169(45.67)	1.00 (Reference)	
CT	209(56)	194(52.43)	0.76(0.56-1.04)	0.088
TT	23(6.21)	7 (1.89)	0.25(0.10–0.63)	0.003
TC+TT	232(62.02)	201(54.32)	0.72(0.53-0.97)	0.03
C Allele	485 (65.54)	532 (71.89)		
T Allele	255 (34.45)	208(28.10)		
MAF	0.34	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 χ^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.

5.5.3.2 Genetic and allelic distribution on the basis of lung cancer subtypes

Lung cancer cases were further segregated into histological subgroups in order to evaluate the risk of susceptibility amongst these subgroups as shown in Table 5.4. SQCC patients with heterozygous genotype (CC) were more represented as compared to wild genotype (AA) (50.37% vs 31.11%). 18.51% cases with SQCC had mutant genotype (CC). Subjects diagnosed with ADCC, 49.1% had heterozygous genotype, 32.5% had wild type genotype and 18.3% were mutants for *Lys⁹³⁹Gln* polymorphism. In case of SCLC subgroup, 45.9%, 41.4%, 12.6% patients were having wild, heterozygous, mutant genotype respectively. As indicated in Table 5.4 the subjects carrying the mutant genotype (CC) for *XPC Lys⁹³⁹Gly* polymorphism were at higher risk for all sub-groups of lung cancer as compared to subjects with heterozygous genotype. It was observed that subjects with mutant genotype (CC) had a 4 and 3-fold increased risk for ADCC (OR=4.72; 95%CI: 2.16-10.32; $p<0.0001$) and SQCC (OR=3.06, 95%CI: 1.58-5.92; $p=0.009$) which was found to be highly significant. A 2-fold risk was also observed in SCLC individuals with mutant genotype (CC) (OR=2.30, 95%CI: 1.41- 3.73; $p=0.0007$). SQCC and ADCC patients depicted almost a 2-fold risk for subjects with heterozygous genotype for *Lys⁹³⁹Gln* polymorphism (OR=1.81, 95%CI: 1.14-2.88; $p=0.01$ and OR=2.15, 95%CI: 2.28-3.63; $p=0.003$). When both the heterozygous and mutant genotype were combined as a single genotype (AC+CC) showed a significant risk for SQCC (OR=2.03, 95%CI: 1.31-3.15; $p=0.001$).

For *Ala⁴⁹⁹Val* polymorphism, it was observed that in SQCC patients: 45.92% had wild type genotype, 51.85% were heterozygous and 2.2% had mutant genotype. On the contrary, 48.3%, 50.8% and 0.83% subjects represented distribution of wild type, heterozygous or mutant genotype in patients diagnosed with ADCC. An overrepresentation of heterozygous genotype (CT) was observed in SCLC patients when compared with the reference genotype (CC) (54.9% vs 42.34%). A non significant SQCC risk was observed for all the three genotypes (CT= OR=0.69, 95%CI: 0.45-1.05; $p=0.82$; TT= OR=0.28, 95%CI: 0.07-0.98; $p=0.39$; CT+TT OR=0.64, 95%CI: 0.42-0.98; $p=0.84$). ADCC subjects with mutant genotype showed the lowest risk for lung cancer (OR=0.12, 95%CI: 0.15-0.96; $p=0.04$). For SCLC patients, both heterozygous and mutant genotype depicted a reduced risk for lung cancer (CA=OR=0.76, 95%CI: 0.48-1.20; $p=0.25$ and TT= OR=0.44, 95%CI: 0.12-1.50; $p=0.21$).

Table 5.4: Distribution of genotypes of XPC polymorphism depicting their association with Lung cancer risk and its histological subtypes

		SQCC			ADCC			SCLC		
<i>Lys⁹³⁹Gln</i>	Controls n (%) N =370	Cases n(%) N=135	AOR (95% CI) ^a	<i>P^b</i>	Cases n(%) N =120	AOR (95% CI) ^a	<i>p^b</i>	Cases n(%) N=111	AOR (95% CI) ^a	<i>P^b</i>
AA	166(44.86)	42(31.11)	1.00 (Reference)		39(32.5)	1.0(Reference)		51(45.9)	1.00 (Reference)	
AC	164(44.32)	68(50.37)	1.81(1.14-2.88)	0.011	59(49.1)	2.15(2.28-3.63)	0.003	46(41.4)	1.12(0.69-1.82)	0.62
CC	40(10.8)	25(18.51)	3.06(1.58-5.92)	0.0009	22(18.3)	4.72 (2.16 10.3)	<0.0001	14(12.6)	2.30(1.41-3.73)	0.0007
AC+CC	204(55.13)	93(68.88)	2.03(1.31-3.15)	0.0014	81(67.5)	0.86(0.47-1.56)	0.06	60 (54.0)	1.19(0.76-1.87)	0.43
		SQCC			ADCC			SCLC		
<i>Ala⁴⁹⁹Val</i>	Controls n (%) N =370	Cases n(%) N =135	AOR (95% CI) ^a	<i>P^b</i>	Cases n(%) N =120	AOR (95% CI) ^b	<i>P^b</i>	Cases n(%) N =111	AOR (95% CI) ^a	<i>P^b</i>
CC	138(37.29)	62(45.92)	1.00 (Reference)		58(48.3)	1.00 (Reference)		47(42.34)	1.00 (Reference)	
CT	209(56)	70(51.85)	0.69 (0.45-1.05)	0.82	61(50.8)	0.749(0.48-1.1)	0.19	61(54.9)	0.76(0.48-1.2)	0.25
TT	23(6.21)	3 (2.2)	0.28 (0.07–0.98)	0.39	1 (0.83)	0.12(0.15-0.96)	0.04	3 (2.7)	0.44(0.12-1.50)	0.21
TC+TT	232(62.02)	73 (54.07)	0.64 (0.42-0.98)	0.84	62 (51.6)	0.70 (0.45-1.08)	0.11	64 (57)	0.91(0.57-1.43)	0.68

(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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. Abbreviations: ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma. All the bold number indicates the significant values.

5.5.3.3 Effect of smoking status on association of *XPC* polymorphism with lung cancer risk

Table 5.5 shows the correlation between *XPC* polymorphic variants *Lys*⁹³⁹*Gln* and *Ala*⁴⁹⁹*Val* with smoking and its association towards lung cancer. The sample population is divided into two subgroups: smokers and non-smokers. The smokers were further divided into heavy 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 mutant genotype (CC) were less represented in controls as compared to the cases (9.42% vs 16.5%). However, smokers harboring wild genotype (AA) were more prevalent in controls in comparison with cases (47.4% vs 36.6%). Lung cancer subjects who were smokers and carrying the mutant (CC) genotype of *XPC Lys*⁹³⁹*Gln* gene were found to have 2-fold elevated risk for lung cancer as compared (OR=2.55; 95% CI: 1.46-4.45; $p=0.001$) to smokers having wild type genotype (AA), a similar trend was also seen in subjects who were having heterozygous genotype (OR=1.53; 95%CI: 1.06-2.20; $p=0.02$), however the effect was less pronounced as shown in Table 5.5. Smokers having combined genotype (AC+CC) showed an increased risk for lung cancer when compared with the reference genotype (OR: 1.71; 95%CI: 1.21-2.42; $p=0.002$).

Furthermore, it was observed that heavy smokers with the mutant (CC) genotype for *Lys*⁹³⁹*Gln* were at significantly higher risk for lung cancer (OR=3.71, 95%CI=1.46-9.45; $p=0.005$) as compared to the light smokers (OR=2.18, 95%CI=1.10-4.28; $p=0.02$). A 2-fold risk was observed in those heavy smokers who were carrying a combined genotype (AC+CC) (OR: 1.84; 95%CI: 1.07-3.16; $p=0.02$) when compared with the combined genotype in light smokers (OR: 1.43; 95%CI: 0.94-2.17; $p=0.09$). For the heterozygous genotype (AC) a non significant risk was observed in both heavy (OR: 1.54; 95%CI: 0.87-2.72; $p=0.13$) as well as light smokers (OR: 1.24; 95%CI: 0.77-1.97; $p=0.36$) when compared with the reference wild genotype (AA). In case of non smokers, a non-significant weak association was observed for lung cancer risk as shown in table 5.5.

In case of *Ala*⁴⁹⁹*Val* polymorphism, no such association was observed between smoking status and lung cancer susceptibility with any genotypic combination as shown

in table 5.5. Smokers with heterozygous genotype (CT) were more pronounced in cases as compared with controls (56.79% vs 36.23%). The TT genotype was over expressed in controls than cases (7.2 % vs 2.3%). A protective effect was observed for all the three genotypic combinations among smokers for *Ala⁴⁹⁹Val* polymorphism *i.e.* CT (OR: 0.80; 95%CI: 0.57-1.14; p=0.23), TT (OR: 0.33; 95%CI: 0.13-0.83; p=**0.02**) and CT+TT (OR: 0.75; 95%CI: 0.53-1.14; p=0.11). A significant protective effect was observed for mutant genotype (AA) when compared with the reference genotype (OR: 0.33; 95%CI: 0.13-0.83; p=**0.02**). Among heavy smokers, the mutant genotype was more prevalent in controls when compared with cases (9.61% vs 2.12%). Wild genotype (CC) was more represented in cases as compared to controls (46.09% vs 37.5%). Light smokers exhibited an over representation of wild genotype (CC) in cases as compared to controls (41.97% vs 35.63%) whereas heterozygous genotype was more expressed among controls than cases (58.62% vs 55.55%). A non significant protective effect was observed in all the genotypes in cases who were heavy smokers *i.e.* CT (OR: 0.95; 95%CI: 0.49-1.82; p=0.88), TT (OR: 0.29; 95%CI: 0.05-1.50; p=1.43) and CT+TT (OR: 0.88; 95%CI: 0.96-1.66; p=0.69). Light smokers with mutant genotype (CT) showed a lower risk for the disease but the results were not significant (OR: 0.51; 95%CI: 0.14-1.77; p=0.29). Non smokers did not show any prominent significant risk for the disease.

Table 5.5: Distribution of genotypes of <i>XPC</i> polymorphism according to their smoking status and its association with Lung Cancer risk								
	SMOKERS				NON SMOKERS			
<i>Lys⁹³⁹Gln</i>	Controls n (%) N =276	Cases n (%) N=303	AOR (95% CI) ^a	<i>p^b</i>	Controls n (%) N =94	Cases n (%) N=67	AOR (95% CI) ^a	<i>p^b</i>
AA	131(47.4)	111(36.6)	1.00 (Reference)		35 (37.23)	22(32.8)	1.00 (Reference)	
AC	119(43.1)	142(46.8)	1.53(1.06-2.2)	0.021	45 (47.87)	32(47.7)	1.25(0.59-2.61)	0.55
CC	26(9.42)	50(16.5)	2.55(1.46-4.45)	0.001	14 (14.89)	13(19.4)	2.05(0.76-5.56)	0.15
AC+CC	145(52.5)	192 (51.89)	1.71(1.21-2.42)	0.002	59 (62.76)	45(67.16)	1.40(0.70-2.79)	0.32
<i>Ala⁴⁹⁹Val</i>	Controls n (%) N =276	Cases n (%) N=303	AOR (95% CI) ^a	<i>p^b</i>	Controls n (%) N =94	Cases n (%) N=67	AOR (95% CI) ^a	<i>p^b</i>
AA	156(56.52)	133(43.89)	1.00 (Reference)		38(40.4)	36(53.7)		
AC	100(36.23)	163(53.79)	0.80(0.571-1.14)	0.23	53(56.3)	31(46.2)	0.61(0.316-1.19)	0.150
CC	20 (7.26)	7(2.3)	0.33(0.133-0.835)	0.02	3(3.19)			
AC+CC	120(43.47)	170(56.10)	0.75(0.537-1.14)	0.116	56(59.5)	31(46.2)	0.58(0.302-1.13)	0.115
	HEAVY SMOKER				LIGHT SMOKER			
<i>Lys⁹³⁹Gln</i>	Controls n (%) N =104	Cases n (%) N=141	AOR (95% CI) ^a	<i>p^b</i>	Controls n (%) N =174	Cases n (%) N=162	AOR (95% CI) ^a	<i>p^b</i>
AA	47(45.14)	45 (31.91)	1.00 (Reference)		85 (48.85)	66 (40.74)	1.00 (Reference)	
AC	49 (47.11)	69 (48.93)	1.54 (0.87-2.72)	0.13	71 (40.80)	73 (45.06)	1.24 (0.77-1.97)	0.36
CC	8 (7.69)	27(19.14)	3.71 (1.46-9.45)	0.005	18 (10.34)	23 (14.19)	2.18 (1.10-4.28)	0.02
AC+CC	57 (54.80)	96 (68.08)	1.84 (1.07-3.16)	0.02	89 (51.14)	96 (59.25)	1.43 (0.943-2.17)	0.09
<i>Ala⁴⁹⁹Val</i>	Controls n (%) N =104	Cases n (%) N=141	AOR (95% CI) ^a	<i>p^b</i>	Controls n (%) N =174	Cases n (%) N=162	AOR (95% CI) ^a	<i>p^b</i>
CC	39(37.5)	65(46.09)	1.00 (Reference)		62(35.63)	68(41.97)	1.00 (Reference)	
CT	55(52.8)	73(51.77)	0.95(0.49-1.82)	0.88	102(58.62)	90(55.55)	0.86(0.54-1.36)	0.52
TT	10(9.61)	3(2.12)	0.29(0.05-1.50)	1.43	10(5.74)	4(2.46)	0.51 (0.14-1.77)	0.29
CT+TT	65(62.5)	76(53.90)	0.88(0.96-1.66)	0.69	112(64.36)	94(58.02)	0.82 (0.52-1.30)	0.41

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

5.5.3.4 XPC haplotypes and lung cancer risk

For the current study, we analyzed haplotypes using SHEsis program platform (Table 5.6). For two SNPs linkage disequilibrium and haplotype frequencies were analyzed for two SNPs (*Lys*⁹³⁹*Gln* and *Ala*⁴⁹⁹*Val*). The two SNPs were in linkage disequilibrium in this study population ($D=0.039$, $r^2=0.00$). The haplotypes were composed of two coding SNPs. There was a statistically significant difference in the overall haplotype distribution between cases and controls (Global χ^2 test=18.811, $p=0.0003$). We considered the individuals with 939A-499C haplotype as the reference group for OR estimations. AC haplotype was more represented among controls in comparison with cases (44.9% vs 41.8%) whereas CC haplotype was more prevalent among cases than controls (30.1% vs 20.7%). The C-C haplotype was associated with increased risk of lung cancer (OR=1.65; 95%CI; 1.30-2.09; $p=3.31e-005$). CT genotype was less expressed in cases as compared to controls (10.5% vs 11.9%). Among individuals having the A-T haplotype there was a reduced risk of lung cancer (OR=0.73; 95%CI; 0.56-0.94; $p=0.01$). All other haplotypes (CT) did not show any significant effect (OR=0.86; 95%CI; 0.62-1.19; $p=0.38$).

Table 5.6: XPC Haplotypes and their association with lung cancer risk				
<i>Lys</i> ⁹³⁹ <i>Gln</i> & <i>Ala</i> ⁴⁹⁹ <i>Val</i>	Case frequency, n (%)	Control frequency, n (%)	OR (95% CI) ^a	<i>P</i> ^b
A-C	41.8% (309)	44.9% (331)	0.88 (0.719-1.08)	0.23
A-T	17.6% (130)	22.6% (167)	0.73 (0.56-0.94)	0.01
C-C	30.1% (222)	20.7% (153)	1.65(1.301-2.09)	3.31e-005
C-T	10.5% (77)	11.9% (87)	0.86 (0.626-1.19)	0.38

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either haplotypic distribution or frequencies between the cases and controls. Global χ^2 test=18.811, *p*^b-value = 0.0003.

5.5.3.5 Association between lung cancer risk and different combinations of *XPC* genotypes

Different genotypic combinations of two *XPC* variants were analyzed to find out their association with lung cancer risk in order to evaluate SNP-SNP interaction between the two *XPC* polymorphic variants (Table 5.7). A 2-fold risk of lung cancer was observed in subjects carrying either variant alleles (AC+CT) for both the polymorphic variants, however this effect was not significant (OR=1.90; 95%CI: 0.91-3.94; $p=0.45$) as its frequency in heterozygote's was roughly similar in both cases and controls (60.28% vs 58.33%). When both the SNP-SNP combinations of heterozygous and mutant genotype were combined as a single genotype a weak was observed (OR=1.29; 95%CI: 0.82-2.04; $p=0.26$). No association was observed in other combined *XPC* genotypes in the current study.

Table 5.7: Combination of genotypes of *XPC* polymorphic variants and their relationship with the risk of lung cancer

Genotypes	Cases, n (%) N = 141	Controls, n(%) N = 156	Adjusted OR (95% CI) ^a	p^b
<i>Lys⁹³⁹Gln & Ala⁴⁹⁹Val</i>				
0 (AA+CC)	56 (39.71%)	64 (41.02%)	1.00 (Reference)	
1(AC+CT)	85 (60.28%)	91 (58.33%)	1.90 (0.91-3.94)	0.45
2(CC+TT)	-	1 (0.64%)	-	-
3 (1+2)	85 (60.28%)	92 (58.97%)	1.29 (0.82-2.04)	0.26

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding p -values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status. (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 3: combined hetero and mutant genotype

5.5.3.6 Relationship between *XPC* Genotypic variants and Clinico-pathological Characteristics

To determine association between the two polymorphic variants of *XPC* (*Lys⁹³⁹Gln*, *Ala⁴⁹⁹Val*) and clinico-pathological parameters we divided patients on the basis of stage of cancer (III vs IV), primary tumor extension (T3+T4 vs T1+T2) and metastatic status (positive vs negative) (Table 5.8).

In order to investigate the relationship of different genotypes of *XPC* polymorphic variants, the genotypic frequencies were calculated for each polymorphic site for both stage III and stage IV patients as shown in table 5.8. The comparison in the genotypic frequencies was carried out to find out whether any significant difference lies in the genotypic frequencies among the two subgroups made on the basis of stage. Overall, none of the polymorphic site depicted any significant difference in the genotypic frequencies among the two subgroups, thus, pertaining to the lack of evident association between these genotypes and stage of the tumor. This can be clearly seen in case of *Lys*⁹³⁹*Gln* where the genotypic frequencies of the subjects carrying variant genotype (CC) was 15.33% in stage III and 19.4% among stage IV category. Similarly, the genotypic frequencies of heterozygotes in cases were almost identical among stage III (46.85%) and stage IV patients (46.01%). No significant OR was observed for both heterozygous (OR= 0.80, 95%CI: 0.48-1.15, *p*= 0.39) and mutant (OR= 0.60, 95%CI: 0.30-1.17, *p*= 0.13) genotypes for *Lys*⁹³⁹*Gln* polymorphism. Further, *Ala*⁴⁹⁹*Val* 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 (46.85%) than stage IV (44.78%) while heterozygous genotype was almost equally expressed in stage III (52%) and stage IV (52.76%) patients.

The next clinical parameter which has been accounted for consideration is primary tumor extension. The patients were divided into two categories on the basis of tumor size as shown in table 5.8. A higher proportion of lung cancer patients were reported in the T3+T4 category (276) whereas TX+T1+T2 subgroup consists of only 69 patients. The genotypic frequencies were estimated for each polymorphic site and it was observed that *XPC Lys*⁹³⁹*Gln* heterozygous genotype (AC) was represented in 47.46% patients with T3/T4 tumor as compared to 31.88% patients with T1/T2 tumor. A 3.5 fold tumor extension (T3/T4) for subjects carrying heterozygous genotype was observed as compared to patients with T1/T2 tumor size (OR=3.54, 95%CI: 1.71-7.51, *p*=**0.0006**) for the *XPC Lys*⁹³⁹*Gln* polymorphism. It was also observed that frequency of mutant genotype (CC) in subjects with TX+T1+T2 extension was almost similar to the lung cancer cases with T3+T4 extension (15.94% vs 17.75%) indicating no correlation with tumor extension. For combined genotype (AC+CC), T3/T4 extension showed an increased risk when compared with wild genotype for subgroup with TX+T1+T2 extension (OR= 1.93, 95% C.I.: 1.08-3.43, *p*=**0.02**). In case of *Ala*⁴⁹⁹*Val*, the percentage of heterozygotes (CT) in lesser in

TX+T1+T2 subgroup as compared to T3+T4 tumor subgroup (42.02% vs 53.26%). On the contrary, a strong protective effect was observed in case of patients carrying the heterozygous genotype (OR=0.91, 95% CI: 0.53-1.57, p= 0.55). The wild genotype was highly prevalent in T2+T1+T0 subgroup when compared with T3+T4 subgroup (55.07% vs 45.28%).

Table 5.8 also demonstrates the relationship of these different genotypes with metastatic status of the patients. These were further classified on the basis of metastatic status into M0 and M1 group. Patients showing positive distant metastasis (M1) were compared with the group lacking metastasis (M0) which was also regarded as the reference group. Among *Lys⁹³⁹Gln* genotypic variants, there were 18.51% patients who showed metastasis while there were 17.55% patients who did not show any metastasis in lung cancer patients harboring the double variant alleles (CC). The frequency of subjects harboring wild genotype (AA) was 39.36% for distant metastasis and 35.44% for subjects lacking metastasis. The combined variants (AC+CC) illustrated a lower frequency for metastatic group than non metastatic group (60.6% vs 64.55%). 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 case of *Ala⁴⁹⁹Val* polymorphism the percentage of subjects carrying homozygous wild genotype (CC) and lacking metastasis were 45.50% while 47.34% lung cancer patients had distant metastasis. The representation of double variant allele (TT) was almost similar in both M1 (1.59%) and M0 (2.11%) groups. The frequency of combined variants (CT+TT) was 46.27% for patients with distant metastasis and 54.49% for those who were lacking metastasis. The patients with combined variant genotype (CT+TT) for *Ala⁴⁹⁹Gln* polymorphism showed a progression of lung cancer to other organs when patients who did not show metastasis were considered as reference (OR: 1.13; 95%CI: 0.72-1.77; p=0.05).

Table 5.8: Relationship of Clinico-pathological features with *XPC Lys⁹³⁹Gln* and *XPC Ala⁴⁹⁹Val* genotypes in lung cancer patients

Variable	Clinical stage		AOR (95% CI) ^a	<i>p</i> ^b	TNM stage		AOR (95% CI) ^a	<i>p</i> ^b	Metastasis		AOR (95% CI) ^a	<i>p</i> ^b
	<i>III</i>	<i>IV</i>			<i>T3+T4</i>	<i>T2+T1+T0</i>			<i>Positive</i>	<i>Negative</i>		
<i>Lys⁹³⁹Gln</i>	175	163			276	69			188	189		
AA	59 (33.17)	63 (38.65)	Reference		96 (34.78)	36 (52.17)	Reference		74 (39.36)	67 (35.44)	Reference	
AC	82 (46.85)	75 (46.01)	0.80 (0.48-1.15)	0.39	131 (47.46)	22 (31.88)	3.54(1.71-7.35)	0.0006	81 (43.08)	87 (46.03)	0.79 (0.50-1.27)	0.34
CC	34 (19.4)	25 (15.33)	0.60 (0.30-1.17)	0.13	49 (17.75)	11 (15.94)	1.71(0.74-3.93)	0.20	33 (17.55)	35 (18.51)	0.59 (0.10-3.29)	0.48
AC+CC	116 (66.28)	100 (61.34)	0.75 (0.47-1.19)	0.23	180 (65.21)	33 (47.82)	1.93(1.08-3.43)	0.02	114 (60.63)	122 (64.55)	1.12 (0.71-1.75)	0.61
<i>Ala⁴⁹⁹Val</i>												
CC	82 (46.85)	73 (44.78)	Reference		124 (44.92)	38 (55.07)	Reference		89 (47.34)	86 (45.50)	Reference	
CT	91 (52)	86 (52.76)	0.91 (0.53-1.57)	0.74	147 (53.26)	29 (42.02)	0.91 (0.53-1.57)	0.74	96 (51.06)	99 (52.38)	1.13 (0.72-1.77)	0.58
TT	2 (1.14)	4 (2.45)	0.59 (0.10-3.29)	0.55	5 (1.81)	2 (2.89)	0.59 (0.10-3.29)	0.55	3 (1.59)	4 (2.11)	1.07 (0.22-5.01)	0.92
CT+TT	93 (53.14)	90 (55.21)	0.85 (0.49-1.44)	0.58	152 (55.07)	31 (44.92)	0.85 (0.49-1.44)	0.55	99 (52.65)	103 (54.49)	1.13 (0.72-1.77)	0.05

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status.(b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 3: combined hetero and mutant genotype

5.5.3.7 Association of XPC SNPs with overall survival of lung cancer patients

Overall survival analysis of 311 lung cancer patients was evaluated as shown in Table 5.9. Univariate analyses using Kaplan-Meier curves and log-rank tests was performed in order to evaluate the association of the XPC polymorphism towards overall survival in lung cancer patients as shown in Table 5.9. The distribution of observed genotypes in the patients agreed with those estimated from the Hardy–Weinberg equilibrium ($\chi^2=0.29$; $p=0.58$; for XPC *Lys⁹³⁹Gln*). The frequency of genotypes of lung cancer patients showing *Lys⁹³⁹Gln* polymorphism were as follows: AA (35.36%), AC (46.94%), CC (17.64%) and *Ala⁴⁹⁹Val*: CC (43.40%), CT (54.34%), TT (2.25%). In case of *Lys⁹³⁹Gln* polymorphism survival time of mutant genotype was slightly higher than that of patients with wild genotype (9.13 vs 7.33 months; logrank $p=0.52$). There was no significant association was observed between *Lys⁹³⁹Gln* polymorphism and OS. As shown in Table 5.9, the lung patients carrying homozygous variant allele (TT) for XPC *Ala⁴⁹⁹Val* were associated with longer survival as compared to subjects with wild type genotype (CC) (MST= 22.16 months vs 7.56 months, log-rank $p=0.06$). In the multivariate Cox proportional hazards regression analysis, no significant association was observed between the *Ala⁴⁹⁹Val* mutant genotype and overall lung cancer survival (HR¹=0.43; 95%CI: 0.15-1.18; $p=0.10$).

Table 5.9: Association of XPC polymorphism and overall survival analysis in lung cancer patients

Genotype	No. n% N=311	Dead n% N=256	Alive n% N=55	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	p-value
<i>Lys⁹³⁹Gln</i>								
AA	110 (35.36)	93	17	7.33	1.00(Reference)		1.00(Reference)	
AC	146 (46.94)	119	27	7.56	0.93 (0.70-1.22)	0.60	0.93 (0.70-1.24)	0.64
CC	55 (17.68)	44	11	9.13	0.89 (0.62-1.26)	0.52	0.90 (0.74-1.09)	0.37
AC+CC	201 (64.64)	163	38	8.30	0.91 (0.70-1.78)	0.51	0.92 (0.70-1.21)	0.57
<i>Ala⁴⁹⁹Val</i>								
CC	135 (43.40)	115	20	7.56	1.00(Reference)		1.00(Reference)	
CT	169 (54.34)	137	32	8.03	0.92 (0.71-1.18)	0.51	0.88 (0.68-1.13)	0.32
TT	7 (2.25)	4	3	22.16	0.41 (0.21-0.80)	0.06	0.43 (0.15-1.18)	0.10
CT+TT	176 (56.59)	141	35	8.03	0.89 (0.69-1.14)	0.35	0.85 (0.66-1.09)	0.22

(a)-Hazard ratios, 95% confidence intervals and their corresponding p -values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding p -values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG .

5.5.3.8 Association of *XPC* SNPs with overall survival in lung cancer histological subtypes

Overall survival in *XPC* polymorphism and their association with different histological subtypes of lung cancer has been explained in Table 5.10 (a, b, c). The genotypic distribution and overall survival of ADCC patients has been tabulated in table 5.10 (a), SQCC patients in table 5.10 (b) and SCLC patients in table 5.10 (c).

In case of *Lys*⁹³⁹*Gln* polymorphism, the percentage of heterozygous patients was 49.03%, 51.28% and 38.89% among ADCC, SQCC and SCLC subtypes. *Lys*⁹³⁹*Gln* polymorphism did not show any significant association between overall survival and ADCC subtype. The number of patients carrying mutant genotype was low in the present study for *Ala*⁴⁹⁹*Val* polymorphism. Only one ADCC patient was found to have a mutant genotype, therefore, we combined mutant and heterozygous genotype as a single genotype for further study. Out of 104 ADCC patients, 51 (49.03%) subjects showed wild type genotype which was almost similar to the number of patients showing heterozygous genotype *i.e.* 52 (50%) for *Ala*⁴⁹⁹*Val* polymorphism. ADCC patients carrying the heterozygous (CT) genotype showed a better survival as compared to reference genotype (10.66 months vs 6.76 months). A lower hazard ratio was obtained from both univariate (HR= 0.62, 95% CI: 0.40-0.95, log rank ***p*=0.02**; Fig 5.18A) as well as multivariate analysis (HR¹= 0.54, 95% CI: 0.34-0.86, ***p*=0.01**) for heterozygous genotype (CT). Also, ADCC subjects carrying combined genotype (CT+TT) showed a MST of 12.06 months pertaining to a lower hazards ratio in case of univariate analysis (HR= 0.60, 95% CI: 0.39-0.92, log rank ***p*=0.01**; Fig 5.18B) and multivariate analysis (HR¹= 0.52, 95% CI: 0.33-0.83, *p*=0.06) as shown in table 5.10 a).

In case of SQCC patients, the subjects harboring mutant genotype (CC) for *Lys*⁹³⁹*Gln* polymorphism showed a poor median survival time in comparison with those patients with wild genotype (6.86 vs 8.03 months, log rank *p*=0.80). A decreased death probability was also observed in these patients (HR¹= 0.94, 95% CI: 0.48-1.80, *p*=0.85). The MST of SQCC subjects carrying mutant (TT) genotype for *Ala*⁴⁹⁹*Val* polymorphism was higher than those subjects who were carrying wild genotype (CC) (12.13 vs 6.86 months; log rank *p*=0.68). However, the hazards ratio obtained after both univariate (HR= 0.73, 95% CI: 0.28-2.28) and multivariate

analysis ($HR^1 = 0.95$, 95% CI: 0.37-5.14, $p=0.70$) was not significantly associated with the OS of lung cancer subtype.

In case of SCLC patients and *Lys⁹³⁹Gln* polymorphism, variant genotype showed (CC) a better OS as compared to wild type genotype (AA) (MST= 12.63 months vs 6.73 months, log rank $p=0.27$), although no significant association was observed for this polymorphism with lung cancer risk when multivariate analysis was applied ($HR^1 = 0.88$, 95%CI: 0.39-1.96, $p=0.75$). For heterozygous+mutant genotype, a better survival was observed when compared with reference genotype (AA) in SCLC patients (9 vs 6.73 months), a trend of low death ratio was observed but results were not significant ($HR^1 = 0.84$, 95% CI: 0.50-1.40, $p=0.52$). In case of *Ala⁴⁹⁹Val* polymorphism, patients with SCLC showed no significant association for OS and lung cancer risk. SCLC patients with double variant allele (TT) had the overall median survival time of 16.16 months which was higher than homozygous reference type (CC) (MST=5.83 months). None of the genotypic variant showed a significant difference was among patients carrying variant genotype and lung cancer risk.

Table 5.10 (a): Association of XPC polymorphism and overall survival analysis in ADCC patients								
Genotype	No. n% N=104	Dead n% N=86	Alive n% N=18	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	p-value
<i>Lys⁹³⁹Gln</i>								
AA	31 (29.80)	28	3	7.16	1.00(Reference)		1.00(Reference)	
AC	51 (49.03)	40	11	8.03	0.86 (0.53-1.41)	0.56	0.78 (0.46-1.31)	0.35
CC	22 (21.15)	18	4	9.46	0.88 (0.49-1.58)	0.68	1.02 (0.49-2.10)	0.95
AC+CC	73 (70.15)	58	15	9	0.86 (0.54-1.37)	0.51	0.88 (0.54-1.44)	0.62
<i>Ala⁴⁹⁹Val</i>								
CC	51 (49.03)	45	6	6.76	1.00(Reference)		1.00(Reference)	
CT	52 (50)	41	11	10.66	0.62 (0.40-0.95)	0.02	0.54 (0.34-0.86)	0.01
TT	1 (0.96)		1					
CT+TT	53 (50.96)	41	12	12.06	0.60 (0.39-0.92)	0.01	0.52 (0.33-0.83)	0.06

(a)-Hazard ratios, 95% confidence intervals and their corresponding p -values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months and (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding p -values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=overall survival

Table 5.10 (b): Association of XPC polymorphism and overall survival analysis in SQCC patients								
Genotype	No. n% N=117	Dead n% N=95	Alive n% N=22	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P-value
<i>Lys⁹³⁹Gln</i>								
AA	37 (31.62)	29	8	8.03	1.00(Reference)		1.00(Reference)	
AC	60 (51.28)	50	10	6.46	1.08 (0.68-1.70)	0.73	1.00 (0.60-1.66)	0.98
CC	20 (17.09)	16	4	6.86	1.07 (0.58-2.00)	0.80	0.94 (0.48-1.80)	0.85
AC+CC	80 (68.37)	66	14	6.86	1.08 (0.70-1.66)	0.72	0.98 (0.62-1.57)	0.96
<i>Ala⁴⁹⁹Val</i>								
CC	50 (42.73)	41	9	6.86	1.00(Reference)		1.00(Reference)	
CT	64 (54.70)	53	11	5.26	1.18 (0.53-1.34)	0.47	0.99(0.62-1.60)	0.99
TT	3 (2.56)	1	2	12.13	0.73 (0.28-2.28)	0.68	0.95 (0.37-5.14)	0.70
CT+TT	67 (57.26)	54	13	8.26	1.16 (0.67-1.70)	0.76	1.08 (0.66-1.75)	0.74

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months and (b)-Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=overall survival

Table 5.10 (c): Association of XPC polymorphism and overall survival analysis in SCLC patients								
Genotype	No. n% N=90	Dead n% N=75	Alive n% N=15	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P-value
<i>Lys⁹³⁹Gln</i>								
AA	42 (46.67)	36	6	6.73	1.00(Reference)		1.00(Reference)	
AC	35 (38.89)	29	6	8.4	0.84 (0.52-1.37)	0.49	0.89 (0.51-1.56)	0.69
CC	13 (14.44)	10	3	12.63	0.68 (0.36-1.29)	0.27	0.88 (0.39-1.96)	0.75
AC+CC	48 (53.33)	39	9	9	0.80 (0.50-1.27)	0.33	0.84 (0.50-1.40)	0.52
<i>Ala⁴⁹⁹Val</i>								
CC	34 (37.78)	29	5	5.83	1.00(Reference)		1.00(Reference)	
CT	53 (58.88)	43	10	8.26	1.09 (0.68-1.75)	0.68	1.09 (0.66-1.79)	0.72
TT	3 (3.33)	3	-	16.16	0.74 (0.25-2.16)	0.62	1.38 (0.37-5.14)	0.62
CT+TT	56 (62.22)	46	10	8.26	1.07 (0.67-1.70)	0.76	1.07 (0.66-1.74)	0.77

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months and (b)-Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall Survival

5.5.3.9 XPC polymorphism and their association with overall survival in lung cancer patients treated with different chemotherapy regimens

The overall survival of patients treated with different chemotherapy regimens and their association with *XPC* polymorphic variants has been illustrated in Table 5.11. Homozygous wild genotype (AA) was considered as reference. In case of *Lys*⁹³⁹*Gln* polymorphism, it was conferred that lung cancer patients carrying mutant genotype (CC) and treated with docetaxel cisplatin/carboplatin regimen showed a better survival as compared to patients carrying the wild type genotype (AA) (MST=9.66 vs 7.56 months, Log rank $p=0.81$). On the contrary, patients carrying mutant genotype (CC) were observed for *Ala*⁴⁹⁹*Val* polymorphism suggesting a better survival of 13.13 months when compared with wild genotype (MST=9.86 months; log rank $p=0.48$). A poor survival period with lower hazard ratio was observed for heterozygous patients (CT) treated with docetaxel cis/carboplatin regimen ($HR^1=1.96$; 95%CI: 1.07-3.61; $p=0.02$; MST= 5.93 months; Fig 5.18C). The patients carrying combined genotype (CT+TT) for *Ala*⁴⁹⁹*Val* polymorphism and treated with docetaxel cis/carboplatin regimen showed a survival of 6.67 months in comparison with the wild type genotype (9.36 months).

As shown in table 5.11, the patients carrying both the variant alleles genotype (AA) for the *Lys*⁹³⁹*Gln* polymorphism and treated with irinotecan cisplatin/irinotecan carboplatin regimen showed a longer OS and a low death ratio when compared with those patients harboring the wild type genotype (AA) (MST=9.63 vs 7.23 months; $HR^1=0.41$; 95%CI: 0.10-1.66; $p=0.36$), although the results observed were insignificant. For *Ala*⁴⁹⁹*Val* polymorphism, patients having combined (CT+TT) genotype depicted survival of 9 months as compared to patients carrying wild type genotype (MST=7.41, log rank $p=0.62$), showing a decreased mortality rate ($HR^1=0.96$; 95%CI: 0.51-1.85; $p=0.89$). There was no significant association was observed for irinotecan cis/carboplatin regimen and OS in lung cancer patients for *XPC* polymorphic variants.

Furthermore, In case of *Lys*⁹³⁹*Gln* polymorphism the patients treated with pemetrexed cisplatin/carboplatin regimen, the frequency distribution of heterozygous subjects (AC) was 43.47% which was higher than the wild type genotype (33.33%). Individuals with combined genotype (AC+CC) showed a survival of 9.33 months when assessed in comparison with wild genotype (AA) (MST= 10.33 months; log rank $p=0.81$). The lung cancer subjects who were treated with pemetrexed cisplatin/carboplatin regimen and carrying a single copy of variant allele

(CT) for *Ala⁴⁹⁹Val* polymorphism showed a better OS as compared to subjects receiving same chemotherapeutic regimen but have a wild type genotype (CC) (MST=12.13 vs 8.35 months; log rank $p=0.71$). Furthermore, patients having combined genotype (CT+TT) and treated with pemetrexed cisplatin/carboplatin regimen depicted a lower hazard ratio when compared with the wild genotype (HR= 0.69; 95%CI: 0.38-1.29; $p=0.25$). Although the results observed for *Ala⁴⁹⁹Val* polymorphism and its association with OS in lung cancer patients treated with pemetrexed cis/carboplatin regimen was not significant.

Table 5.11: Association of <i>XPC</i> polymorphism and overall survival analysis according to regimen								
REGIMEN – Docetaxel cisplatin/carboplatin								
Genotype	N=76	Alive N=62	Dead N=14	MST (months)	Crude HR^a (95% CI)	Adjusted HR¹ (95% CI)^b	Log rank P	p^b
<i>Lys⁹³⁹Gln</i>								
AA	26 (34.21)	21	5	7.56	1.00 (Reference)	1.00 (Reference)		
AC	40 (52.63)	32	8	8.03	0.92 (0.53-1.60)	0.70 (0.37-1.33)	0.77	0.28
CC	10 (13.15)	9	1	9.66	1.09 (0.49-2.44)	0.61 (0.23-1.62)	0.81	0.32
AC+CC	50 (65.78)	41	9	8.63	0.94 (0.55-1.62)	0.73 (0.40-1.34)	0.83	0.32
<i>Ala⁴⁹⁹Val</i>								
CC	46 (60.52)	35	11	9.86	1.00 (Reference)	1.00 (Reference)		
CT	25 (32.89)	23	2	5.93	1.58 (0.90-2.79)	1.96 (1.07-3.61)	0.07	0.02
TT	5 (6.57)	4	1	13.13	0.69 (0.28-1.70)	0.83 (0.49-1.41)	0.48	0.50
CT+TT	30 (39.47)	27	3	6.67	1.35 (0.80-2.26)	1.51(0.86-2.65)	0.23	0.14
REGIMEN – Irinotecan cisplatin/ carboplatin								
Genotype	N=70	Alive N=53	Dead N=17	MST (months)	Crude HR^a (95% CI)	Adjusted HR¹ (95% CI)^b	Logrank P	p^b
<i>Lys⁹³⁹Gln</i>								
AA	30 (42.85)	24	6	7.23	1.00 (Reference)	1.00 (Reference)		
AC	32 (45.71)	24	8	8.63	0.83 (0.47-1.46)	1.05 (0.56-1.97)	0.51	0.85
CC	8 (11.42)	5	3	9.63	0.64 (0.27-1.51)	0.41 (0.10-1.66)	0.36	0.21
AC+CC	40 (57.14)	29	11	9	0.79 (0.46-1.38)	0.89 (0.49-1.61)	0.40	0.70
<i>Ala⁴⁹⁹Val</i>								
CC	30 (42.85)	23	7	7.41	1.00 (Reference)	1.00 (Reference)		
CT	39 (55.71)	29	10	9	1.11 (0.64-1.92)	0.94 (0.49-1.82)	0.68	0.87
TT	1 (1.42)	1	-					
CT+TT	40 (57.14)	30	10	9	1.13 (0.65-1.94)	0.96 (0.51-1.85)	0.62	0.89
REGIMEN-Pemtrexed cisplatin/carboplatin								
Genotype	N=69	Alive N=55	Dead N=14	MST (months)	Crude HR^a (95% CI)	Adjusted HR¹ (95% CI)^b	Log rank P	p^b
<i>Lys⁹³⁹Gln</i>								
AA	23 (33.33)	19	4	10.33	1.00 (Reference)	1.00 (Reference)		
AC	30 (43.47)	23	7	9.33	0.93 (0.50-1.71)	0.90 (0.41-1.96)	0.81	0.79
CC	16 (23.18)	13	3	9.51	1.01 (0.50-2.06)	0.92 (0.34-2.49)	0.95	0.87
AC+CC	46 (66.67)	36	10	9.33	0.94 (0.54-1.66)	0.95 (0.46-1.95)	0.85	0.90
<i>Ala⁴⁹⁹Val</i>								
CC	34 (49.27)	28	6	8.35	1.00 (Reference)	1.00 (Reference)		
CT	34 (49.27)	27	7	12.13	0.72 (0.42-1.24)	0.71 (0.39-1.30)	0.71	0.27
TT	1 (1.44)	-	1					
CT+TT	35 (50.72)	27	8	12.13	0.70 (0.41-1.20)	0.69 (0.38-1.29)	0.69	0.25

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months and (b)-Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall Survival; MST=Median survival time

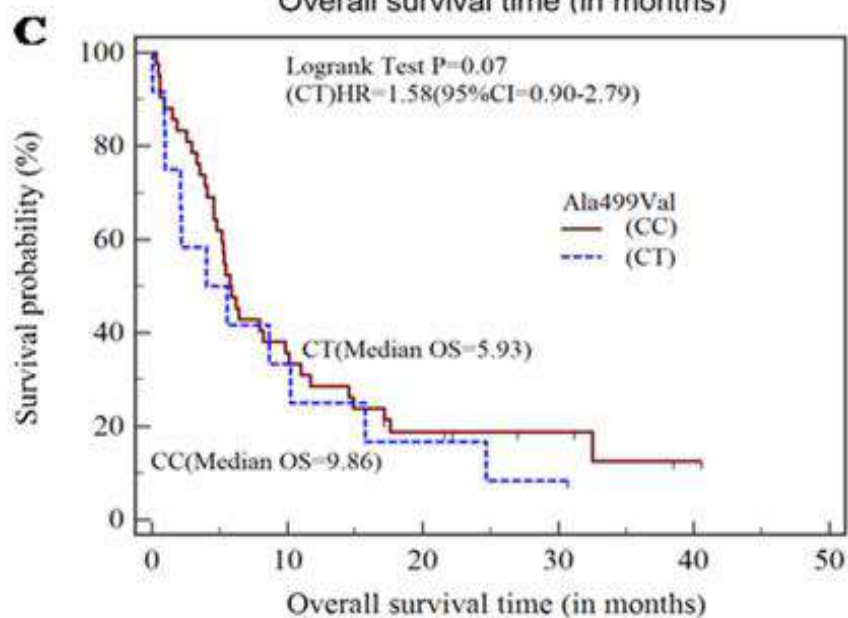
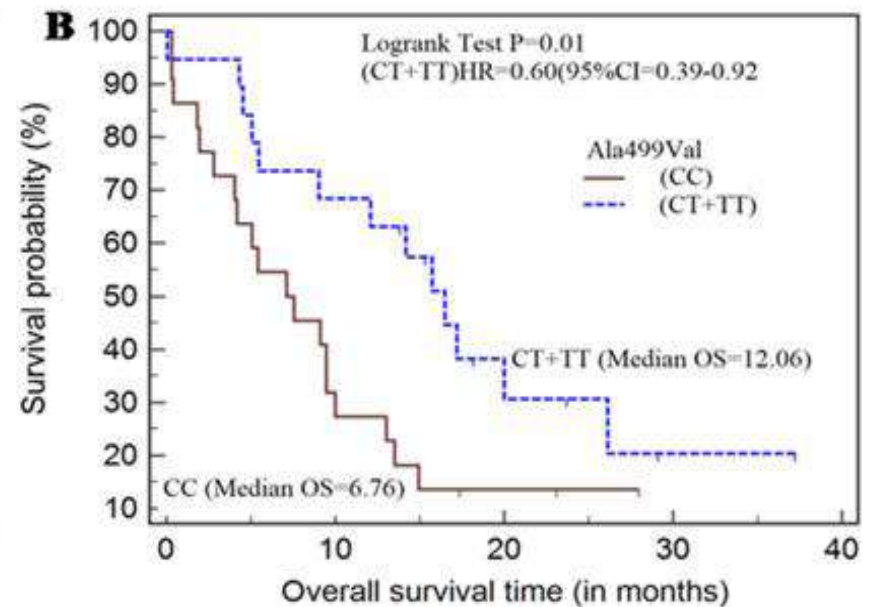
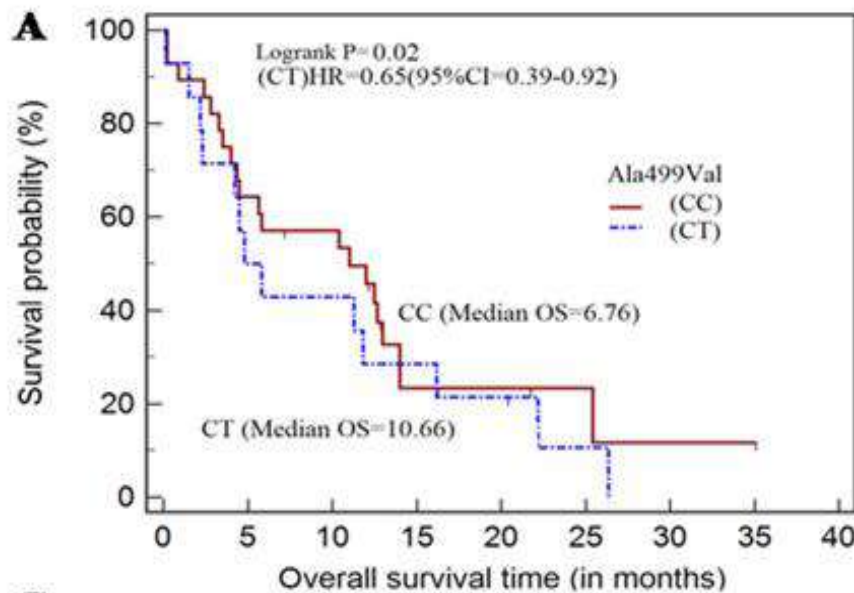


Fig 5.18: Kaplan Meier graphs representing relationship of *XPC Ala⁴⁹⁹Val* polymorphism with OS in (A) ADCC patients with heterozygous (CT) genotype (B) ADCC patients with CT+TT genotype (C) patients treated with docetaxel cis/carboplatin regimen for heterozygous (CT) genotype

5.5.3.10 Association of *XPC* polymorphism with clinical response

202 patients were studied for their response to the chemotherapeutic regimen as per their genotypes for two *XPC* polymorphic variants as shown in the table 5.12. The patients were divided into responders and non responders on the basis of their response towards chemotherapy. There were four response groups which were made as per RECIST criteria: CR (complete response), PR (partial response), SD (stable disease) and PD (progressive disease). (CR+PR) were categorized as responders and (SD+PD) were grouped under non responders. Odds ratio was calculated for each polymorphic site and it was adjusted with various factors such as age, gender, smoking, stage, histology, KPS and ECOG. None of the *XPC* polymorphism revealed any significant difference for response criteria. In case of *Lys⁹³⁹Gln* polymorphism, the representation of patients having AC+CC genotype was higher in responders (66.91%) than non responders (57.29%). Further, the subjects having *Ala⁴⁹⁹Val* polymorphism, CT+TT genotype was slightly higher (59.37%) among non responders as compared to responders (56.60%) but the difference was not found to be statistically significant. Similarly, the frequency of heterozygous genotype (CT) was slightly lower in responders as compared to non responders (54.71% vs 58.33%) but no significant association for this genotype was observed.

Table 5.12. Genotypic Response of *XPC* polymorphism to platinum based chemotherapy and their association with lung cancer risk

Genotype	Responders (CR+PR)(N=106)	Non- Responders (SD+PD)(N=96)	AOR (95% CI) ^a	<i>p</i> ^b
<i>Lys⁹³⁹Gln</i>				
AA	35 (33.01%)	41 (42.70%)	(Reference)	
AC	52 (49.05%)	45 (46.87%)	0.97 (0.48-1.97)	0.95
CC	19 (17.92%)	10 (10.41%)	1.27 (0.52-3.10)	0.58
AC+CC	71 (66.91%)	55 (57.29%)	1.09 (0.58-2.02)	0.91
<i>Ala⁴⁹⁹Val</i>				
CC	46 (43.39%)	39 (40.62%)	(Reference)	
CT	58 (54.71%)	56 (58.33%)	0.83 (0.44-1.93)	0.82
TT	2 (1.88%)	1 (1.04%)	1.86 (0.62-3.33)	0.59
CT+TT	60 (56.60)	57 (59.37%)	0.89 (0.42-1.91)	0.80

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status, histology, stage, KPS and ECOG. (b)- Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. CR-Complete response; PR-Partial response; SD-Stable disease; PD-Progressive disease

5.5.4 XPA

5.5.4.1 Association of XPA polymorphism with lung cancer susceptibility

The allelic and genotypic distribution of two XPA variants, *i.e.* A23G and G709A are shown in Table 5.13. The controls for both the SNPs *i.e.* XPA A23G (N=370, $\chi^2= 2.84$, df=1, p=0.09) and G709A polymorphism (N=370, $\chi^2= 2.55$, df=1, p=0.1) were in accordance with HWE. Furthermore, the analysis revealed that in XPA A23G polymorphism, the lung cancer patients had low frequency of wild type genotype (AA) as compared to controls (39.45% vs 48.91%) whereas homozygous variant (GG) was over represented in cases when compared to control group (14.59% vs 6.48%) as shown in Table 5.13. Therefore, overall a significant difference was observed for genotypic distribution amongst cases and controls ($\chi^2= 15.35$, df=2, p=**0.0005**) for XPA A23G polymorphism. The minor allele frequency (MAF) for cases (0.38) was higher as compared to controls (0.29). Table 5.13, further showed the association of lung cancer risk with XPA polymorphic variants. Adjusted OR's and 95% CI were calculated using logistic regression analysis. The most common genotype was considered as the reference groups. For both the XPA polymorphism the homozygous wild genotype was regarded as the reference group with which variant genotypes were compared. Regression analysis showed that subjects with mutant genotype (GG) for A23G polymorphism possessed a 2.5-fold elevated risk for the disease (OR: 2.56; 95%CI: 1.49-4.39; p= **0.0007**). Whereas individuals with heterozygous variant (AG) showed a less pronounced risk for lung cancer though it was found to be non-significant (OR: 1.21; 95%CI: 0.88-1.65; p=0.23). When both genotypes were combined and used as a single genotype (AG+GG) and compared with homozygous wild genotype (AA), a slight risk was observed which was found to be significant (OR: 1.37; 95%CI: 1.02-1.35; p=**0.03**).

The study group was further comprehended into histological subgroups to evaluate severity of risk towards the disease. Out of 135 SQCC patients, 38.51% had wild type genotype, 49.62% were heterozygous and 11.85% had a mutant genotype. Out of 120 ADCC cases, 44.16% had wild type genotype while 41.66% had heterozygous genotype. From total of 111 SCLC patients, 36.03% had wild, 47.4% were heterozygous and 16.21% had a mutant genotype. The subjects carrying both the variant allele (GG) for A23G polymorphism were at higher risk of developing

lung cancer in all the subgroups. A significant 2-fold risk was observed for individuals with SQCC (OR: 2.58; 95%CI: 1.17-5.70; $p=0.018$) and ADCC (OR: 2.01; 95%CI: 0.97-4.16; $p=0.05$) whereas this risk was found to be highly compounded (3-fold) for SCLC patients which was found to be significant (OR: 3.76; 95%CI: 1.75-8.09; $p=0.0007$). SCLC patients with combined genotype (AG+GG) depicted a 1.6-fold of risk towards the disease (OR: 1.61; 95%CI: 1.02-2.56; $p=0.04$).

For the other *XPA* variant *G709A*, frequency of wild genotype (GG) was over represented in cases in comparison with controls (57.02% vs 47.83%). The frequency of homozygous variant (GA) was higher in controls compared to cases (45.13% vs 32.70%) whereas the frequency of mutant genotype (AA) was over represented amongst cases as compared to controls (10.27% vs 7.02%; $\chi^2=1.89$; $p=0.16$). The MAF in cases and controls was 0.27 and 0.30 respectively. For *XPA G709A* polymorphism, the heterozygous genotype (GA) showed a highly significant protective effect (OR: 0.54; 95%CI: 0.38-0.78; $p=0.001$), whereas no such relation was observed for the mutant genotype (OR: 0.97, 95%CI: 0.51-1.86; $p=0.94$). When both heterozygous and mutant genotype combined as a single genotype (GA+AA) and compared with the reference genotype (GG) a significant protective effect was observed (OR: 0.64; 95%CI: 0.47-0.87; $p=0.004$) as shown in table 5.14.

When stratified according to histological subtypes, heterozygous genotype (GA) was less represented in SQCC (27.40%), ADCC (38.33%) and SCLC (34.33%) when compared with homozygous reference genotype (GG) in SQCC (60.74%), ADCC (51.66%) and SCLC (56.75%) patients respectively. The frequency of homozygous variant genotype (AA) was over represented in all the three histological subtype *i.e.* SQCC (11.85%), ADCC (10%) and SCLC (9%) as compared to controls (7.02%). After applying logistic regression analysis a significant protective effect for heterozygous genotype (GA) was found patients with SQCC (OR: 0.45; 95%CI: 0.28-0.75; $p=0.001$) and SCLC histological sub-group (OR: 0.60; 95%CI: 0.37-0.97; $p=0.03$) when compared with the reference genotype. For the subjects carrying combined genotype (GA+AA) a more pronounced protective effect was observed for SQCC in *G709* polymorphism (OR: 0.56; 95%CI: 0.36-0.76; $p=0.007$) when compared with the wild type genotype (GG). There was no significant association observed for ADCC and *G709A* polymorphism.

Table 5.13: Distribution of genotypes of XPA polymorphism depicting their association with Lung cancer risk and its histological subtypes													
	OVERALL LUNG CANCER				SQCC			ADCC			SCLC		
A23G	Controls n (%) N =370	Cases N (%) N=370	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N=135	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =120	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N=111	AOR (95% CI) ^a	<i>p</i> ^b
AA	181 (48.91)	146 (39.45)	1.00 (Reference)		52 (38.51)	1.00 (Reference)		53 (44.16)	1.0(Reference)		40 (36.03)	1.00 (Reference)	
AG	165 (44.59)	170 (45.94)	1.21 (0.88-1.65)	0.23	67 (49.62)	1.23 (0.79-1.92)	0.33	50 (41.66)	0.93 (0.59-1.46)	0.76	53 (47.74)	1.37 (0.84-2.24)	0.20
GG	24 (6.48)	54 (14.59)	2.56 (1.49-4.39)	0.0007	16 (11.85)	2.58 (1.17-5.70)	0.018	17 (14.16)	2.01 (0.97-4.16)	0.05	18 (16.21)	3.76 (1.75-8.09)	0.0007
AG+GG	189 (51.08)	224 (30.27)	1.37 (1.02-1.85)	0.03	83 (61.48)	1.36 (0.89-2.08)	0.14	67 (55.83)	1.24 (0.80-1.90)	0.32	71 (48.37)	1.61 (1.02-2.56)	0.04
A Allele	527 (71.21)	462 (62.43)											
G Allele	213 (28.78)	278 (37.56)											
MAF	0.29	0.38											
	OVERALL LUNG CANCER				SQCC			ADCC			SCLC		
G709A	Controls n (%) N =370	Cases n(%) N=370	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =135	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =120	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =111	AOR (95% CI) ^a	<i>p</i> ^b
GG	177 (47.83)	211 (57.02)	1.00 (Reference)		82 (60.74)	1.00 (Reference)		62 (51.66)	1.00 (Reference)		63 (56.75)	1.00 (Reference)	
GA	167 (45.13)	121 (32.70)	0.54 (0.38-0.78)	0.001	37 (27.40)	0.45 (0.28-0.73)	0.001	46 (38.33)	0.74 (0.47-1.16)	0.19	38 (34.23)	0.60 (0.37-0.97)	0.03
AA	26 (7.02)	38 (10.27)	0.97 (0.51–1.86)	0.94	16 (11.85)	1.21 (0.58-2.49)	0.60	12 (10)	0.99 (0.44-2.19)	0.98	10 (9.00)	0.92 (0.40-2.11)	0.84
GA+AA	293 (79.18)	159 (42.97)	0.64 (0.47-0.87)	0.004	53 (39.25)	0.56 (0.36-0.86)	0.007	58 (48.33)	0.78 (0.51-1.20)	0.26	48 (43.24)	0.65 (0.41-1.02)	0.06
G Allele	521 (70.40)	543 (73.37)											
A Allele	219 (29.59)	197 (26.62)											
MAF	0.30	0.27											

(a)-AdjustedOdds 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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.Abbreviations: ADCC=Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma. All the bold number indicates the significant values.

5.5.4.2 Interactive effect of smoking on the *XPA* polymorphism and their association with lung cancer risk

To understand the effect of gene-environment interaction, the study group was divided into smoking and non smoking group considering their smoking status (Table 5.14). The smokers were further divided into light and heavy smokers as per the number of pack years they smoked. The individual who had smoked ≤ 25 pack years were categorized as light smokers whereas individuals who has smoked >25 pack years were defined as heavy smokers. The wild type genotype was regarded as reference for both the polymorphism. For A23G polymorphism, the wild type genotype (AA) was more expressed in cases than controls for smoking population (38.28% vs 46.12%). Similarly, the number of smokers with mutant genotype (GG) was higher in cases as compared to controls (14.8% vs 5.1%). It was observed that smokers with mutant genotype (GG) exhibited a 3.6-fold elevated risk for the lung cancer when compared to reference group (AA) for A23G polymorphism (OR: 3.61; 95%CI: 1.82-7.04; $p=0.0002$). No such association was visible for the heterozygous genotype (AG) as its percentage in both cases (46.86%) and controls (48.70%) was almost same. When mutant genotype were combined with heterozygous genotype (AG+GG), the frequency of variant alleles was higher in cases (50.54%) than controls (39.40%). After applying logistic regression analysis, a slight risk for lung cancer was observed for combined genotype (OR: 1.40; 95%CI: 0.99-1.98; $p=0.05$). In case of heavy smokers, no significant association was observed between *XPA* A23G polymorphism and lung cancer susceptibility. The mutant (GG) genotype was more prominent in cases in comparison to controls (14.64% vs 8.33%) whereas the wild genotype (AA) was more represented in controls than cases (41.66% vs 36.30%). Individuals carrying mutant genotype showed 2-fold risk for lung cancer among heavy smokers which was not significant (OR: 1.87; 95%CI: 0.79-4.54; $p=0.16$).

The subjects harboring the mutant genotype (GG) and who were light smokers expressed a 7-fold increased susceptibility (OR: 6.78; 95%CI: 2.36-19.49; $p=0.0004$) towards lung cancer when compared with the reference group (AA). In case of mutant genotype (GG), it was seen that frequency distribution of mutant genotype was over expressed in cases than controls (15.06% vs 3.06%) among light smokers showing A23G polymorphism. When heterozygous and mutant genotype combined as a single genotype (AG+GG), light smokers showed 1.5-fold of lung cancer risk (OR: 1.52; 95%CI: 0.95-2.43; $p=0.07$), although this result

was not significant. No significant association was observed between non smokers and A23G polymorphism.

In *G709A* polymorphism, the frequency of the wild type genotype (GG) was more expressed in cases than controls (57.75% vs 47.97%) whereas heterozygous genotype (GA) was more prevalent in controls than cases among smokers (45.01% vs 31.68%) as shown in table 5.14. Heterozygous genotype (GA) displayed a protective effect towards the disease for subjects who were smokers as compared to non smokers and was statistically significant (OR: 0.54; 95%CI: 0.51-1.86; $p=0.001$). Our data also shows over representation of combined genotype (GA+AA) in controls (52.02%) than cases (42.24%). Smokers with combined genotype (GA+AA) showed a significant protective association when compared with wild genotype (OR: 0.68; 95%CI: 0.43-0.85, $p=0.004$). We have further classified our data into heavy and light smokers, it was observed that the subjects who were heavy smokers showed more prominence in cases than controls for wild genotype (59.87% vs 44.44%) whereas heterozygous genotype was more expressed in controls than cases (50% vs 31.84%). Heterozygous genotype (GA) showed a significant protective effect towards the disease among heavy smokers when compared with the reference genotype (GG) (OR: 0.49; 95%CI: 0.29-0.84; $p=0.01$). A significant association was also seen in heavy smokers with combined variant (GA+AA) genotype of this polymorphic site (OR: 0.53; 95%CI: 0.31-0.89; $p=0.01$).

In *G709A* polymorphism, an over presentation of controls in comparison to cases was observed for heterozygous genotype (GA) in light smokers (41.71% vs 31.50%) whereas reference genotype (GG) was more represented among cases than controls (55.47% vs 50.30%). A protective effect was observed for single variant genotype (GA) in light smokers (OR: 0.59; 95%CI: 0.35-0.98; $p=0.04$). No significant association was observed between non smoking group and lung cancer risk with any genotype of *G709A* polymorphism.

Table 5.14: Distribution of genotypes of XPA polymorphism according to their smoking status and its association with Lung Cancer risk								
	SMOKERS				NON SMOKERS			
XPA A23G	Controls n (%) N =271	Cases n (%) N=303	AOR (95% CI) ^a	p ^b	Controls n (%) N =99	Cases n (%) N=67	AOR (95% CI) ^a	p ^b
AA	125 (46.12)	116 (38.28)	1.00 (Reference)		56 (56.56)	30 (44.77)	1.00 (Reference)	
AG	132 (48.70)	142 (46.86)	1.17 (0.82-1.68)	0.37	33 (33.33)	28 (41.79)	1.49 (0.75-2.98)	0.25
GG	14 (5.1)	45 (14.8)	3.61 (1.82-7.04)	0.0002	10 (10.10)	9 (13.43)	1.39 (0.48-4.02)	0.54
AG+GG	146 (39.40)	187 (50.54)	1.40 (0.99-1.98)	0.05	43 (43.43)	37 (55.22)	1.46 (0.77-2.78)	0.23
XPA G709A								
GG	130 (47.97)	175 (57.75)	1.00 (Reference)		46 (46.46)	36 (53.73)	1.00 (Reference)	
GA	122 (45.01)	96 (31.68)	0.54 (0.38-0.78)	0.001	46 (46.46)	25 (37.31)	0.60 (0.33-1.27)	0.21
AA	19 (7.01)	32 (10.56)	0.97 (0.51-1.86)	0.94	7 (7.07)	6 (8.95)	0.84 (0.24-2.87)	0.78
GA+AA	141 (52.02)	128 (42.24)	0.68 (0.43-0.85)	0.004	53 (53.53)	31 (46.2)	0.69 (0.36-1.31)	0.25
	HEAVY SMOKER				LIGHT SMOKER			
XPA A23G	Controls n (%) N =108	Cases n (%) N=157	AOR (95% CI) ^a	p ^b	Controls n (%) N =163	Cases n (%) N=146	AOR (95% CI) ^a	p ^b
AA	45 (41.66)	57 (36.30)	1.00 (Reference)		80 (49.07)	59 (40.41)	1.00 (Reference)	
AG	54 (50)	77 (49.04)	1.12 (0.65-1.92)	0.68	78(47.85)	65 (44.52)	1.20 (0.74-1.96)	0.45
GG	9 (8.33)	23 (14.64)	1.87 (0.79-4.54)	0.16	5 (3.06)	22 (15.06)	6.78 (2.36-19.49)	0.0004
AG+GG	63 (58.33)	100 (63.69)	1.23 (0.73-2.07)	0.43	83 (50.92)	87 (59.58)	1.52 (0.95-2.43)	0.07
XPA G709A								
GG	48 (44.44)	94 (59.87)	1.00 (Reference)		82 (50.30)	81 (55.47)	1.00 (Reference)	
GA	54 (50)	50 (31.84)	0.49 (0.29-0.84)	0.01	68 (41.71)	46 (31.50)	0.59 (0.35-0.98)	0.04
AA	6 (5.55)	13 (8.28)	0.80 (0.27-2.39)	0.70	13 (7.97)	19 (13.01)	1.17 (0.52-2.64)	0.69
GA+AA	60 (55.55)	63 (40.12)	0.53 (0.31-0.89)	0.01	81 (49.69)	65 (44.52)	0.68 (0.42-1.09)	0.11

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding p-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

5.5.4.3 Haplotype analysis of *XPA* variants

For the present study, haplotype frequencies and linkage disequilibrium were analyzed for two SNPS using SHEsis program platform (Table 5.15). SNPs in the current study were in weak linkage disequilibrium for the study population ($D=0.081$, $r^2=0.001$). Comparison of overall haplotypic profile showed a statistically significant distribution between cases and control group (Global $\chi^2=14.73$, $p=0.002$). A-A haplotype was more prevalent in controls comparison with cases (22.2% vs 17%). The subjects carrying A-A haplotype showed a less pronounced protective effect (OR: 0.71; 95%CI: 0.55-0.92; $p=0.01$). The haplotypes with both mutant allele (G-G) were associated with increased risk of lung cancer (OR: 1.42; 95%CI: 1.12-1.80; $p=0.003$), as the case frequency (27.9%) was higher than that of controls (21.4%). All other haplotypes did not show any significant effect. G-A haplotype possessed risk for lung cancer (OR: 1.33; 95%CI: 0.92-1.92; $p=0.12$) while A-G haplotype was associated with decreased risk for the disease (OR: 0.86; 95%CI: 0.70-1.06; $p=0.16$).

Table 5.15: *XPA* Haplotypes and their association with lung cancer risk

<i>XPA</i> A23G & <i>XPAG709A</i>	Case frequency, n (%)	Control frequency, n (%)	OR (95% CI) ^a	p^b -value
A-A	0.170 (126)	0.222 (164)	0.71 (0.55-0.92)	0.01
A-G	0.455 (336)	0.490 (363)	0.86 (0.70-1.06)	0.16
G-A	0.097 (71)	0.074 (55)	1.33 (0.92-1.92)	0.12
G-G	0.279 (207)	0.214 (158)	1.42 (1.12-1.80)	0.003

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding p -values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either haplotypic distribution or frequencies between the cases and controls. Global χ^2 test=14.734, p^b -value = 0.002.

5.5.4.4 Association between lung cancer risk and different combinations of *XPA* genotypes

In order to evaluate SNP-SNP interaction between the two *XPA* polymorphic variants different genotypic combinations of two *XPA* variants were analyzed to estimate their association towards risk for lung cancer as shown in table 5.16. The heterozygous genotype (Group 1) for both SNPs was more prevalent in controls (47.27%) than cases (38.73%) whereas wild genotype (Group 0) was more expressed in cases (59.85%) than controls (52.12). A decreased risk of lung cancer was observed in subjects carrying either variant alleles (AG+GA) for both the polymorphic variants, however this effect was not significant (OR=0.63; 95%CI: 0.39-1.02; $p=0.06$). Since frequency of mutant genotype was less pronounced in both cases and control we combined it with the heterozygous genotype as a single genotype, a non significant protective effect for the disease was observed (OR=0.66; 95%CI: 0.47-1.05; $p=0.08$) as prevalence of this genotype was more in controls (47.87%) as compared to cases (40.14%). There was no significant association was observed in other combined *XPA* genotypes in the current study as shown in table 5.16.

Table 5.16 : Combination of genotypes of *XPA* polymorphic variants and their relationship with the risk of lung cancer

Genotypes	Cases, n (%) N = 142	Controls, n(%) N = 165	Adjusted OR (95% CI) ^b	p^a
<i>XPA</i> A23G & G709A				
0 (AA+GG)	85 (59.85)	86 (52.12)	1.00 (Reference)	
1(AG+GA)	55 (38.73)	78 (47.27)	0.63 (0.39-1.02)	0.06
2(GG+AA)	2 (1.40)	1 (0.60)	2.65 (0.22-31.98)	0.44
3 (1+2)	57 (40.14)	79 (47.87)	0.66 (0.47-1.05)	0.08

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding p -values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status. (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 3: combined hetero and mutant genotype

5.5.4.5 Relationship of Clinico-pathological features with *XPA* polymorphic variants

To evaluate association of two polymorphic variants of *XPA* (A23G, G709A) on various clinic-pathological features, we segregated patients according to clinical stage (III vs IV), primary tumor extension (T3+T4 vs T0+T1+T2) and metastatic status (positive vs negative) (Table 5.17).

On the basis of clinical stage, the patients were stratified into stage III and stage IV. The genotypic distribution for each polymorphic site has been estimated. There was no significant association found between A23G polymorphism and clinical staging. In case of A23G polymorphism, the heterozygous patients (AG) were found at slightly higher frequency among stage IV cases (46.62%) as compared to stage III patients (45.71%) as shown in Table 5.17. For homozygous variant type, the prevalence of subjects having GG genotype was higher among stage III (16.57%) subgroup as compared to stage IV (12.88%) subgroup. Logistic regression analysis showed that patients in stage IV and carrying mutant genotype (GG) were at lower risk when compared with patients in stage III of disease (OR= 0.76, 95% CI: 0.38-1.51, $p=0.44$). In case of G709A polymorphism the frequency of subjects with heterozygous genotype was 31.42% and 36.80% among stage III and stage IV subgroup respectively. Logistic regression analysis exhibited no association with lung cancer staging as indicated in table 5.17 when compared with reference genotype (OR: 1.18; 95%CI: 0.71-1.73; $p=0.63$). Similarly, the frequency of combined variant genotype (GA+AA) was lower in stage III (42.28%) in comparison to stage IV (46.62%). Hence, both *XPA* variants lacked significant relationship with clinical staging.

The next parameter evaluated for these two polymorphic sites of *XPA* was primary tumor extension, the patients were stratified into two groups namely, T3+T4 and T1+T2+T0. For A23G variant, heterozygous genotype (AG) was highly represented in patients with tumor size for T3+T4 extension (48.18%) as compared to the patients having T1+T2+T0 (33.33%) tumor stage. A 2-fold high risk was observed for lung cancer patients carrying heterozygous genotype (AG) and having T3+T4 tumor stage with T0+T1+T2 stage as a reference (OR: 1.80; 95%CI: 0.99-3.28; $p=0.05$). In case of G709A polymorphism, the distribution of heterozygous genotype (GA) in population was almost similar among T3+T4 and T1+T2 subgroup (34.42% and 33.33%). Similarly, the combined variants (GA+AA) were similar in frequency for G709A polymorphism

(43.84% vs 44.92%) among T3+T4 and T1+T2 subgroup. There was no significant association was observed between *G709A* polymorphism and primary tumor extension.

Relationship of different genotypes of *A23G* and *G709A* with metastatic status is demonstrated in table 5.17. In case of *A23G* polymorphism, the genotypic frequencies were evaluated for the two groups and comparison showed 46.52% of lung cancer subjects did not show any metastasis (M0) and 43.67% had positive metastasis for patients carrying heterozygous genotype (AG). The frequency of subjects exhibiting homozygous wild genotype (AA) was 41.77% for positive metastasis (M1) and 37.96% for subjects lacking metastasis (M0). The combined variants (AG+GG) were also in almost similar frequency for metastatic and non metastatic group (58.22% vs 62.03%). 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 5.17. In case of *G709A* polymorphism the percentage of homozygous variant (AA) subjects lacking metastasis were 10.16% while 9.49% showed distant metastasis. The representation of heterozygous genotype was almost similar in both M1 (33.68%) and M0 (34.81%) groups. However, there was no significant frequency distribution of subjects was observed. Similar to the previous variant *G709A* polymorphism did not contribute towards progression of lung cancer among other organs as shown in table 5.17.

Table 5.17: Relationship of Clinic-pathological features with XPA A23G and XPA G709A genotypes in lung cancer patients

Variable	Clinical stage		AOR (95% CI) ^a	<i>p</i> ^b	TNM stage		AOR (95% CI) ^a	<i>p</i> ^b	Metastasis		AOR (95% CI) ^a	<i>p</i> ^b
	III	IV			T3+T4	T2+T1+T0			Positive	Negative		
XPA A23G	175	163			276	69			158	187		
AA	66 (37.71)	66 (40.49)	Reference		104 (37.68)	33 (47.82)	Reference		66 (41.77)	71 (37.96)	Reference	
AG	80 (45.71)	76 (46.62)	1.06 (0.65-1.72)	0.80	133 (48.18)	23 (33.33)	1.80 (0.99-3.28)	0.05	69 (43.67)	87 (46.52)	0.92 (0.56-1.50)	0.75
GG	29 (16.57)	21 (12.88)	0.76 (0.38-1.51)	0.44	39 (14.13)	13 (18.84)	0.96 (0.45-2.04)	0.93	23 (14.55)	29 (15.50)	0.90 (0.45-1.77)	0.76
AG+GG	109 (62.28)	97 (59.50)	0.98 (0.62-1.54)	0.93	172(62.31)	36 (52.17)	1.51 (0.88-2.58)	0.13	92 (58.22)	116 (62.03)	0.91 (0.58-1.44)	0.70
XPA G709A												
GG	101 (57.71)	87 (53.37)	Reference		155 (56.15)	38 (55.07)	Reference		88 (55.69)	105 (56.14)	Reference	
GA	55 (31.42)	60 (36.80)	1.18 (0.73-1.90)	0.49	95 (34.42)	23 (33.33)	1.05 (0.58-1.88)	0.85	55 (34.81)	63 (33.68)	0.96 (0.59-1.56)	0.89
GG	19 (10.85)	16 (9.81)	0.89 (0.42-1.90)	0.77	26 (9.42)	8 (11.59)	0.80 (0.33-1.96)	0.63	15 (9.49)	19 (10.16)	0.87 (0.40-1.92)	0.74
GA+GG	74 (42.28)	76 (46.62)	1.11 (0.71-1.73)	0.63	121 (43.84)	31 (44.92)	1.00 (0.58-1.72)	0.97	80 (50.63)	82 (43.85)	0.94 (0.60-1.48)	0.81

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status.(b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

5.5.4.6 Association of XPA SNPs with overall survival of lung cancer patients

Overall survival analysis of 311 lung cancer patients was conducted as shown in Table 5.18. First, univariate analysis was performed using Kaplan-Meier curves and logrank test and then multivariate Cox hazard proportional analysis was performed to evaluate the association of XPA polymorphism with overall survival of patients. The distribution of observed genotypes in the patients agreed with those estimated from HWE ($\chi^2= 0.08$; $df=1$; $p=0.76$) for A23G polymorphic variant. The homozygous wild type genotype *i.e.* AA (A23G) and GG (G709A) was considered as reference.

For A23G polymorphism, it was observed that patients carrying single variant allele (AG) was found to be associated with shorter survival period when compared with patients harboring the wild genotype (AA) (7.56 vs 8.20 months, logrank $p=0.95$). There was no significant association observed for A23G polymorphism and overall survival in lung cancer patients on applying both univariate and multivariate analysis. For G709A polymorphism a constant survival was observed in heterozygous genotype (GA) as compared to the reference genotype (GG) (8.03 vs 8.03 months; logrank $p=0.31$). Lung cancer patients with mutant genotype (AA) showed survival of 8.3 months in comparison with the reference genotype (GG) showing overall survival of 8.03 months. No significant association was observed between G709A polymorphism and overall survival in both Kaplan-Meier analysis and Cox regression model.

5.5.4.7 Association of XPA SNPs with overall survival in lung cancer histological subtypes

Overall survival in XPA polymorphism and their association with different histological subtypes of lung cancer has been explained in Table 5.19 (a, b, c). The genotypic distribution and overall survival of ADCC patients has been tabulated in table 5.19 (a), SQCC patients in table 5.19 (b) and SCLC patients in table 5.19 (c). In case of A23G polymorphism, the percentage of patients carrying heterozygous genotype was 41.34%, 47.86% and 37.78% among ADCC, SQCC and SCLC subtypes. ADCC patients carrying the heterozygous (AG) genotype showed a poor OS as

Table 5.18: Association of XPA polymorphism and overall survival analysis

Genotype	No. n% N=311	Dead n% N=256	Alive n% N=55	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P-value
XPA A23G								
AA	126 (40.51)	103 (40.23)	23 (41.81)	8.20	1.00(Reference)		1.00(Reference)	
AG	142 (45.65)	119 (46.48)	23 (41.81)	7.56	0.99(0.76-1.29)	0.95	0.94(0.70-1.25)	0.68
GG	43 (13.82)	34 (13.28)	9 (16.36)	9.23	0.99(0.67-1.45)	0.96	0.97(0.79-1.19)	0.82
AG+GG	185 (59.48)	153 (59.76)	32 (58.18)	7.93	0.99(0.77-1.27)	0.93	0.96(0.74-1.24)	0.76
XPA G709A								
GG	175 (56.27)	138 (53.90)	37 (67.27)	8.03	1.00(Reference)		1.00(Reference)	
GA	104 (33.44)	91 (35.54)	13 (23.63)	8.03	1.14(0.87-1.49)	0.31	1.15(0.87-1.53)	0.30
AA	32 (10.28)	27 (10.54)	5 (9.09)	8.3	0.94(0.62-1.41)	0.77	0.96 (0.77-1.18)	0.71
GA+AA	136 (43.72)	118 (46.09)	18 (32.72)	8.03	1.08(0.85-1.39)	0.49	1.06(0.82-1.36)	0.64

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

Table 5.19 (a) : Association of Overall survival and lung cancer risk in ADCC patients

Genotype	No. n% N= 104	Dead n% N= 86	Alive n% N= 18	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P-value
XPA A23G								
AA	46 (44.23)	36 (41.86)	10 (55.55)	8.8	1.00(Reference)		1.00(Reference)	
AG	43 (41.34)	37 (43.02)	6 (33.33)	6.1	1.14 (0.72-1.81)	0.55	1.40 (0.83-2.39)	0.20
GG	15 (14.42)	13 (15.11)	2 (11.11)	5.13	1.41 (0.70-2.83)	0.27	2.05 (0.95-4.40)	0.06
AG+GG	58 (55.76)	50 (58.13)	8 (44.44)	5.85	1.19 (0.78-1.83)	0.40	1.53 (0.84-2.41)	0.15
XPA G709A								
GG	54 (51.92)	42 (48.83)	12 (66.67)	8.03	1.00(Reference)		1.00(Reference)	
GA	41 (39.42)	37 (43.02)	4 (22.22)	9	1.17 (0.75-1.84)	0.46	1.23 (0.77-1.98)	0.38
AA	9 (8.65)	7 (8.13)	2 (11.11)	9.46	0.83 (0.39-1.75)	0.64	0.73 (0.31-1.72)	0.48
GA+AA	50 (48.07)	44 (51.16)	6 (33.33)	9	1.09 (0.71-1.67)	0.66	1.06 (0.67-1.66)	0.79

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. ADCC=Adenocarcinoma; OS=Overall survival

Table 5.19 (b): Association of Overall survival and lung cancer risk in SQCC patients								
Genotype	No. n% N= 117	Dead n% N= 95	Alive n% N= 22	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P-value
XPA A23G								
AA	46 (39.31)	38 (40)	8 (36.36)	6.23	1.00(Reference)		1.00(Reference)	
AG	56 (47.86)	44 (46.31)	12 (54.54)	8.03	0.82 (0.53-1.27)	0.38	0.77 (0.48-1.23)	0.28
GG	15 (12.82)	13 (13.68)	2 (9.09)	8.63	1.08 (0.57-2.07)	0.78	0.95 (0.47-1.88)	0.88
AG+GG	71 (60.68)	57 (60)	14 (63.63)	8.03	0.87 (0.57-1.32)	0.51	0.82 (0.53-1.28)	0.39
XPA G709A								
GG	69 (58.97)	54 (56.84)	15 (68.18)	7.56	1.00(Reference)		1.00(Reference)	
GA	34 (29.05)	29 (30.52)	5 (22.72)	7.26	1.06 (0.67-1.68)	0.77	0.88 (0.52-1.48)	0.63
AA	14 (11.96)	12 (12.63)	2 (9.09)	12.56	0.90 (0.49-1.65)	0.74	0.84 (0.44-1.62)	0.62
GA+AA	48 (41.02)	41 (43.15)	7 (31.81)	6.86	1.01 (0.67-1.52)	0.92	0.87 (0.55-1.37)	0.52

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. SQCC=Squamous cell carcinoma; OS=Overall survival

Table 5.19 (c): Association of Overall survival and lung cancer risk in SCLC patients								
Genotype	No. n% N= 90	Dead n% N= 75	Alive n% N= 15	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P-value
XPA A23G								
AA	34 (41.11)	29 (38.67)	5 (33.33)	6.3	1.00(Reference)		1.00(Reference)	
AG	43 (37.78)	38 (50.67)	5 (33.33)	9.8	0.82 (0.45-1.48)	0.51	0.73 (0.38-1.40)	0.35
GG	13 (14.44)	8 (10.67)	5 (33.33)	15.36	0.46 (0.19-1.11)	0.15	0.35 (0.11-1.10)	0.07
AG+GG	56 (62.22)	46 (61.33)	10 (66.66)	10.66	0.73 (0.41-1.32)	0.28	0.69 (0.37-1.27)	0.24
XPA G709A								
GG	52 (57.78)	42 (56)	10 (66.66)	9	1.00(Reference)		1.00(Reference)	
GA	29 (32.22)	25 (33.33)	4 (26.67)	7.3	1.08 (0.65-1.78)	0.74	1.16 (0.67-1.99)	0.58
AA	9 (10)	8 (10.67)	1 (6.67)	5.8	1.08 (0.50-2.37)	0.82	1.23 (0.55-2.73)	0.60
GA+AA	38 (42.22)	33 (44)	5 (33.33)	5.8	1.08 (0.68-1.71)	0.72	1.07 (0.67-1.72)	0.75

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. SCLC=Small cell lung cancer; OS=Overall survival

compared to ADCC subjects carrying wild genotype (6.1 months vs 8.8 months; logrank $p=0.55$). Hazards Ratio obtained from multivariate analysis was 1.40 with insignificant p values of 0.20. Also, the ADCC subjects carrying double variant allele (GG) or mutant alleles showed MST of 5.13 months pertaining to a increased hazards ratio in both univariate analysis ($HR=1.41$, 95% CI: 0.70-2.83, log rank $p=0.27$) and multivariate analysis ($HR^1=2.05$, 95% CI: 0.95-4.40, $p=0.06$; Fig 5.19A) when compared with the subjects harboring wild genotype (AA).

In case of *G709A* polymorphism, ADCC patients having mutant (GG) genotype showed a better MST (9.46 months, log rank $p=0.64$) and a decreased death probability was observed in comparison with wild type genotype ($HR^1=0.73$, 95% CI: 0.31-1.72, $p=0.48$) as shown in table 5.19 (a). On the contrary, the MST of SCLC subjects carrying mutant (AA) genotype for *G709A* polymorphism was 5.8 months which was lower as compared to 9 months in case of lung cancer subjects having wild (GG) genotype (logrank $p=0.82$). However, the hazards ratio obtained after both univariate and multivariate analysis was not significantly associated with the mutant genotype (AA) as shown in table 5.19 (c). Similarly, In *A23G* polymorphism SCLC patients having variant genotype showed a better OS as compared to subjects with wild type genotype (MST= 15.36 months vs 6.3 months, log rank $p=0.15$; Fig 5.19B). On applying Cox regression model, a lower hazard ratio was observed for SCLC patients having mutant genotype for *A23G* polymorphism ($HR^1=0.35$, 95%CI: 0.11-1.10, **$p=0.07$**). For combined genotype (AG+GG), a better survival was observed when compared with reference genotype (AA) in SCLC patients (10.66 vs 6.3 months), a trend of low death rate was observed but results were not significant ($HR^1=0.69$, 95% CI: 0.37-1.27, $p=0.24$).

In patients diagnosed with SQCC, no significant association was observed among OS and lung cancer risk. In *G709A* polymorphism, SQCC patients showed a favorable trend as the overall median survival time was higher among individuals carrying mutant genotype as compared to the wild type (12.56 vs 7.56 months; log rank $p=0.74$). No significant association was observed among patients carrying variant genotype and lung cancer risk on applying multivariate analysis ($HR^1:0.84$; 95%CI: 0.44-1.62; $p=0.62$). In SQCC patients showing *A23G* polymorphism, heterozygous genotype (AG) (MST=8.03 months; logrank $p=0.38$) and mutant genotype (GG) (MST= 8.03 months; logrank $p=0.78$) showed a better survival than the reference genotype (AA)

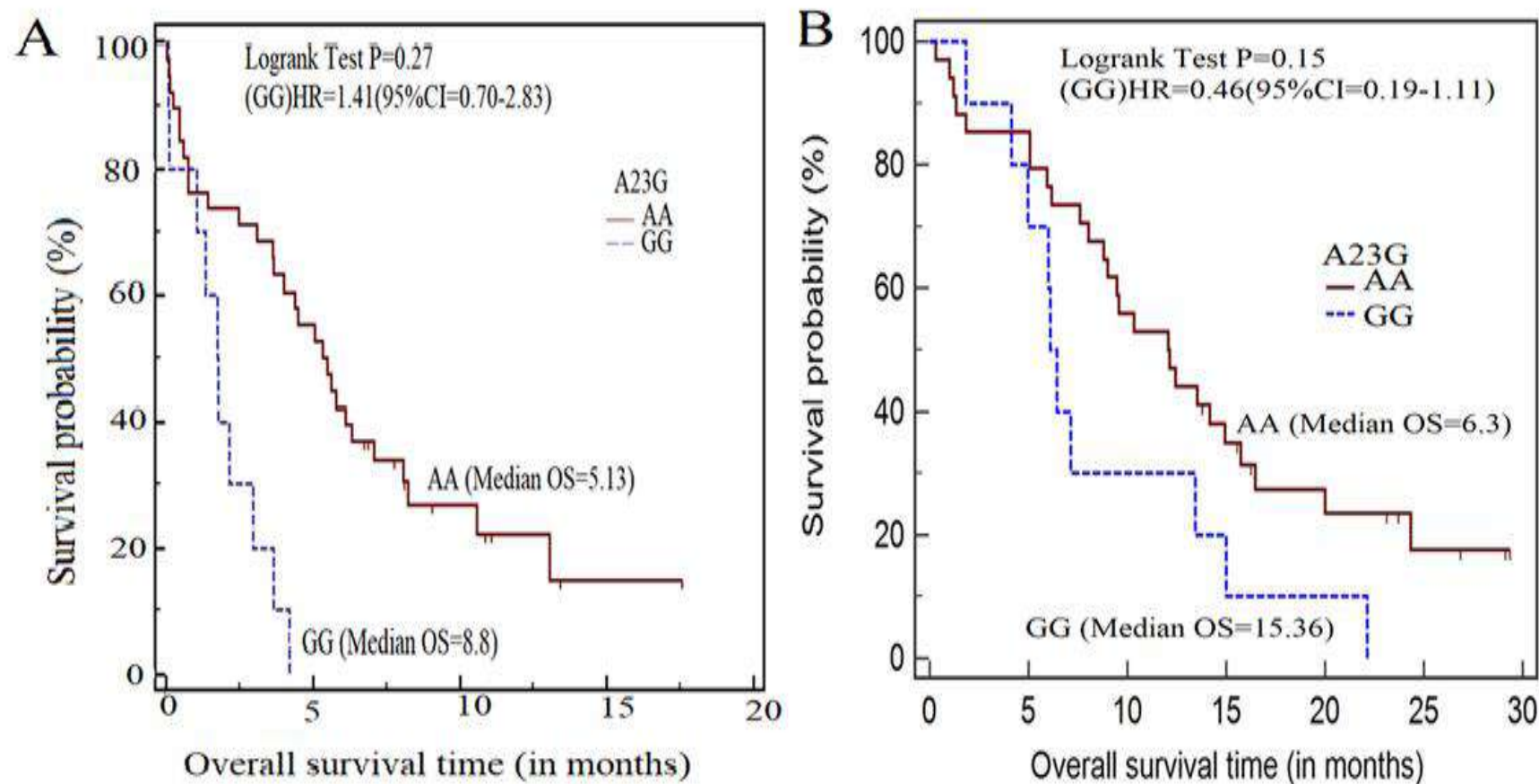


Fig 5.19: Kaplan Meier curve depicting association of A23G polymorphism with overall survival in **(A)** ADCC patients with mutant genotype (GG) **(B)** SCLC patients with mutant genotype (GG)

(MST=6.23 months). There was no significant result observed between OS and A23G polymorphism in SQCC patients.

5.5.4.8 Association of XPA SNPs and Overall survival of lung cancer patients treated with different chemotherapeutic regimen

Lung cancer patients were treated with different chemotherapeutic regimens and their association with OS as shown in Table 5.20. The present study population had been treated with different chemotherapeutic regimen but the analysis was only applied on those regimens with which maximum number of patients was treated in the study. Therefore, the OS of patients treated with three regimens *i.e.* docetaxel cis/carboplatin, irinotecan cis/carboplatin and pemetrexed cis/carboplatin and their association with XPA polymorphism has been elaborated in Table 5.20.

In A23G polymorphism, the patients treated with docetaxel cis/carboplatin regimen for all the three genotypic variants (heterozygous, mutant and combined) showed survival of 7 months when compared with the wild type genotype (AA) which have median survival of 7.56 months (table 5.20). No significant association can be stated between XPA polymorphism and lung cancer survival rates. For G709A polymorphism, patients with mutant genotype (AA) and treated with docetaxel cis/carboplatin regimen showed a better survival in patients when compared with the reference genotype (GG) (15.01 vs 7.56 months; log rank $P=0.77$). Lower hazard ratio was observed for patients carrying mutant genotype but result was not significant ($HR^1= 0.58$, 95% CI: 0.17-1.96, $p=0.38$).

Furthermore, the patients harboring mutant genotype and treated with irinotecan cisplatin/carboplatin regimen depicted a poor survival for A23G polymorphism when compared with the wild genotype (AA) (4.46 vs 10.10 months) whereas for G709A polymorphism the patients carrying mutant genotype (AA) treated with same regimen illustrated a better survival in comparison with reference genotype (9.86 vs 7.23 months) as shown in table 5.20. In A23G polymorphism patients with heterozygous genotype expressed low hazard ratio but results were not reported to be significant ($HR^1= 0.89$, 95% CI: 0.30-1.45, $p=0.31$; MST=9.80 months). While for G709A polymorphism an insignificant high risk for death was observed in lung cancer patients ($HR^1= 1.15$, 95% CI: 0.57-2.31, $p=0.68$).

The third regimen with which maximum patients in the present study have been studied is pemetrexed cis/carboplatin. For both the polymorphism no significant association could be established with OS. The patients carrying variant genotype (GG) for A23G polymorphism showed a poor survival when compared with the wild genotype (AA) (5.43 vs 9.46 months; log rank $p=0.28$). The combined genotype (AG+GG) which was present in 56.52% on the study population elaborated in high death ratio for A23G polymorphism although results were not significant as 61.9% patients reported did not survive during the study ($HR^1= 1.48$, 95% CI: 0.76-2.87, $p=0.24$; MST=6.20 months). In G709A polymorphism, individuals carrying heterozygous genotype and treated with pemetrexed cisplatin/carboplatin regimen showed a better survival than those carrying wild genotype (GG) (9.43 vs 7.56 months). A poor survival was observed for patients with mutant genotype treated with the above regimen in comparison to homozygous reference genotype (6.76 vs 7.56 months) as shown in table 5.20. There was no significant difference observed for G709A polymorphism and OS in patients treated with pemetrexed cis/carboplatin regimen.

Table 5.20: Association of XPA polymorphism and overall survival analysis according to regimen

REGIMEN – Docetaxel cisplatin/carboplatin								
Genotype	N=76	Alive (N=16)	Dead (N=60)	MST (months)	Crude HR^a (95% CI)	Adjusted HR¹ (95% CI)^b	Log P	p
XPA A23G								
AA	28 (36.84)	6	22	7.56	1.00 (Reference)	1.00 (Reference)		
AG	39 (51.31)	10	29	7.00	0.96 (0.53-1.71)	0.71 (0.37-1.36)	0.90	0.31
GG	9 (11.84)	0	9	7.00	1.80 (0.72-4.50)	1.31 (0.53-3.23)	0.12	0.55
AG+GG	48 (63.15)	10	38	7.00	1.09 (0.63-1.90)	0.84 (0.46-1.55)	0.74	0.59
XPA G709A								
GG	46 (60.52)	11	35	7.56	1.00 (Reference)	1.00 (Reference)		
GA	26 (34.21)	4	22	6.23	1.24 (0.69-2.23)	1.20 (0.62-2.34)	0.89	0.57
AA	4 (5.26)	1	3	15.01	0.63 (0.23-1.69)	0.58 (0.17-1.96)	0.77	0.38
GA+AA	30 (39.47)	5	25	6.86	1.11 (0.64-1.92)	1.03 (0.56-1.90)	0.97	0.90
REGIMEN – Irinotecan cisplatin/ carboplatin								
Genotype	N=70	Alive (N=21)	Dead (N=49)	MST (months)	Crude HR^a (95% CI)	Adjusted HR¹ (95% CI)^b	Log P	p
XPA A23G								
AA	30 (42.85)	8	22	10.10	1.00 (Reference)	1.00 (Reference)		
AG	29 (41.42)	9	20	9.80	1.08 (0.56-2.06)	0.67 (0.30-1.45)	0.80	0.31
GG	11 (15.71)	4	7	4.46	1.12 (0.46-2.75)	0.98 (0.30-1.79)	0.78	0.98
AG+GG	40 (57.14)	13	27	6.73	1.10 (0.61-2.00)	0.89 (0.44-1.79)	0.73	0.74
XPA G709A								
GG	39 (55.71)	15	24	7.23	1.00 (Reference)	1.00 (Reference)		
GA	22 (31.42)	5	17	8.26	1.29 (0.65-2.55)	1.15 (0.57-2.31)	0.44	0.68
AA	9 (12.85)	1	8	9.86	1.19 (0.51-2.80)	1.89 (0.63-5.68)	0.65	0.25
GA+AA	31 (44.28)	6	25	9.86	1.26 (0.69-2.28)	1.19 (0.64-2.23)	0.43	0.57
REGIMEN-Pemtrexed cisplatin/carboplatin								
Genotype	N=69	Alive (N=18)	Dead (N=51)	MST (months)	Crude HR ^a (95% CI)	Adjusted HR¹ (95% CI)^b	Log P	p
XPA A23G								
AA	30 (43.47)	9	21	9.46	1.00 (Reference)	1.00 (Reference)		
AG	27 (39.13)	6	21	6.43	1.47 (0.76-2.83)	1.48 (0.71-3.06)	0.22	0.29
GG	12 (17.39)	3	9	5.43	1.53 (0.64-3.70)	1.91 (0.74-4.94)	0.28	0.17
AG+GG	39 (56.52)	9	30	6.20	1.47 (0.82-2.63)	1.48 (0.76-2.87)	0.18	0.24
XPA G709A								
GG	37 (53.62)	12	25	7.56	1.00 (Reference)	1.00 (Reference)		
GA	29 (42.02)	5	24	9.43	0.96 (0.53-1.73)	1.11 (0.58-2.13)	0.89	0.74
AA	3 (4.34)	1	2	6.76	1.23 (0.25-6.02)	0.58 (0.10-3.19)	0.77	0.53
GA+AA	32 (46.37)	6	26	9.23	0.99 (0.55-1.76)	1.12 (0.59-2.12)	0.97	0.71

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. MST=Median survival time; OS=Overall survival

5.5.4.9 Association of *XPA* polymorphism with treatment response for different chemotherapeutic regimens and also their correlation with objective response

The patients were analyzed for the response to the chemotherapeutic regimen in relation to their genotypes for different *XPA* variants as shown in table 5.21. The patients were divided into responders and non responders on the basis of the RECIST criteria where they are annotated into four groups, CR (complete response), PR (partial response), SD (stable disease) and PD (progressive disease). Responders comprises of CR+PR and non responders consist of SD+PD. Adjusted odds ratio was calculated for each polymorphic site considering wild genotype as the reference. Confounding factors taken into consideration for adjustment were age, gender, smoking, stage, histology, KPS and ECOG. The objective response study was conducted on 202 patients. None of all the polymorphic site A23G and G709A gene revealed a significant difference in the frequency of patients showing good and bad response. In case of A23G polymorphism, the representation of patients having AG+GG genotype was similar among both responders (61.32%) and non responders (61.45%). The percentage of AG genotype was higher among responders (45.28%) in comparison to non responders (41.67%). Further, the variant (GG) patients showing a good response (16.03%) were not significantly represented as compared to those exhibiting a poor response (19.79%). In case of G709A polymorphism, the proportion of subjects having GA+AA genotype was lower (39.62%) among responders as compared to non responders (46.87%) but the difference was not found to be statistically significant. The distribution of G709A showed that a higher representation of heterozygotes among non responders than responders (36.45% and 25.47% respectively).

Table 5.21. Genotypic Response of XPA polymorphism to platinum based chemotherapy and their association with lung cancer risk				
Genotype	Responders (CR+PR)(N=106)	Non- Responders (SD+PD)(N=96)	AOR (95% CI)^a	p^b
A23G				
AA	41 (38.67%)	37 (38.54%)	(Reference)	
AG	48 (45.28%)	40 (41.67%)	1.10 (0.57-2.11)	0.77
GG	17 (16.03%)	19 (19.79%)	1.16 (0.50-2.67)	0.72
AG+GG	65 (61.32%)	59 (61.45%)	1.11 (0.61-2.02)	0.71
G709A				
GG	64 (60.37%)	51 (53.12%)	(Reference)	
GA	27 (25.47%)	35 (36.45%)	1.48 (0.77-2.84)	0.23
AA	15 (14.15%)	10 (10.41)	0.88 (0.34-2.28)	0.80
GA+AA	42 (39.62)	45 (46.87%)	1.27 (0.70-2.30)	0.43

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status, histology, stage, KPS and ECOG. (b)- Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. CR- Complete response; PR-Partial response; SD-Stable disease; PD-Progressive disease

5.5.5 XPD gene

5.5.5.1 XPD polymorphic variants and their association towards lung cancer susceptibility

The allelic and genotypic distribution of the three *XPD* variants *i.e.* *Lys*⁷⁵¹*Gln*, *Asp*³¹²*Asn* and *Arg*¹⁵⁶*Arg* and their association with risk of lung cancer has been summarized in Table 5.22. The genotypic frequencies of both polymorphism with *XPD Lys*⁷⁵¹*Gln* ($\chi^2=0.38$, df=1, *p*=0.53) and *XPD Asp*³¹²*Asn* ($\chi^2=2.21$, df=1, *p*=0.13) were in agreement with HWE for controls. The genotypic frequencies of *Arg*¹⁵⁶*Arg* polymorphism were not in agreement with HWE for controls ($\chi^2=5.30$, df=1, *p*=0.02). The analysis revealed over representation of wild type genotype (AA) in controls when compared with cases (54.86% vs 50.54%) in *XPD Lys*⁷⁵¹*Gln* polymorphism, whereas a higher frequency of homozygous variant alleles (CC) was observed in cases when compared with controls (14.32% vs 7.5%). Therefore, a significant overall genotypic distribution was observed amongst cases and controls ($\chi^2=8.67$, df= 2, *p*=**0.01**). Hence, the minor allele frequency (MAF) for cases (0.32) was higher as compared to controls (0.26). We further assessed the association between the *XPD* variants and lung cancer risk as shown in Table 5.22(a). Adjusted OR's and 95% CI were calculated using logistic linear regression analysis. Wild type genotype was regarded as the reference group. For *XPD Lys*⁷⁵¹*Gln* polymorphism, the analysis showed association of mutant genotype (CC) with almost 2-fold elevated risk for lung

cancer (OR: 1.96; 95%CI: 1.17-3.27; $p=0.01$) whereas heterozygous genotype did not show any significant association (OR: 1.03; 95%CI: 0.75-1.43; $p=0.81$). When segregated according to histological subtypes, for all the three subtypes of lung cancer as shown in table 5.22(b), the genotypic distribution in SQCC patients exhibited less distribution of heterozygous genotype (AC) in comparison with wild genotype (AA) (35.55% vs 51.85%). A similar result was observed for patients with ADCC (35% vs 50%) and SCLC (35.13% vs 48.64%) histological subtypes; thus indicating lower expression of heterozygous genotype when compared with the wild genotype. The mutant genotype (CC) was very less prominent in patients diagnosed with SQCC (12.59%), ADCC (15%) and SCLC (16.21%) patients respectively. An elevated risk for mutant genotype (CC) was reported for all three histological subtypes: SQCC (OR: 1.85; 95%CI: 0.91-3.76; $p=0.008$), ADCC (OR: 2.07; 95% CI: 1.05-4.08; $p=0.03$) and SCLC (OR: 2.27; 95%CI: 1.12-4.63; $p=0.02$).

For $Asp^{312}Asn$ polymorphism, the representation of all the three genotypes was almost similar in both controls and cases *i.e.* wild type genotype (GG) (47.56% vs 46.48%), heterozygous (GA) (45.13% vs 45.67%) and mutant (AA) (7.29% vs 7.83%) genotype as depicted in table 5.22(b). Therefore, a non-significant overall genotypic distribution for both controls and cases was observed ($\chi^2=0.12$, $df=2$, $p=0.93$). MAF for A allele was similar in both cases and controls (0.30 and 0.31). No significant association was observed amongst $Asp^{312}Asn$ genotypic variants and lung cancer risk after logistic regression analysis. When categorized according to histological subtypes, the SQCC subjects with wild genotype were more prevalent than heterozygous genotype (48.88% vs 40%) whereas SCLC patients showed more prominence of heterozygous genotype in comparison with the wild genotype (49.54% vs 43.24%). After applying logistic regression analysis, there was no significant association observed for $Asp^{312}Asn$ and all the histological subtypes of lung cancer.

For $Arg^{156}Arg$ polymorphism, the wild genotype (CC) was more represented in cases as compared to controls (46.75% vs 41.35%) whereas heterozygous genotype was over represented (CA) in controls as compared to cases (50% vs 45.40%) as shown in table 5.22(b). The mutant genotype (AA) was found equally represented in both controls and cases (8.64% vs 7.83%). The MAF for controls (0.34) was higher in comparison to cases (0.30). There was no significant association observed for $Arg^{156}Arg$ genotypes and lung cancer susceptibility. For histological

subtypes of lung cancer, it was observed that 34.81% of SQCC patients carrying heterozygous genotype, 55.55% had wild type genotype and 9.62% expressed mutant genotype (AA). The heterozygous genotype (CA) showed a significant protective effect in subjects with SQCC (OR: 0.63; 95%CI: 0.40-0.99; $p=0.04$) when compared with reference genotype (CC). SQCC patients with combined genotype (CA+AA) also depicted a protective effect for lung cancer risk (OR: 0.69; 95%CI: 0.45-1.05; $p=0.08$). The heterozygous genotype was slightly more represented than wild genotype in cases detected with ADCC (53.33% vs 40%) and SCLC (48.14% vs 44.14%). There was no significant association observed for $Arg^{156}Arg$ polymorphism and ADCC or SCLC.

Table 5.22 (a): Distribution of genotypes of XPD polymorphism depicting their association with lung cancer risk

OVERALL LUNG CANCER				
<i>Lys⁷⁵¹Gln</i>	Controls N(%) N =370	Cases N (%) N=370	AOR (95% CI) ^a	<i>p^b</i>
AA	203 (54.86)	187 (50.54)	1.00 (Reference)	
AC	139 (37.56)	130 (35.13)	1.03 (0.75-1.43)	0.81
CC	28 (7.5)	53 (14.32)	1.96 (1.17-3.27)	0.01
AC+CC	167 (45.13)	183 (49.45)	1.21 (0.90-1.63)	0.20
A Allele	545 (73.64)	504 (68.10)		
C Allele	195 (26.35)	236 (31.89)		
MAF	0.26	0.32		
OVERALL LUNG CANCER				
<i>Asp³¹²Asn</i>	Controls N(%) N =370	Cases N(%) N=370	AOR (95% CI) ^a	<i>p^b</i>
GG	176 (47.56)	172 (46.48)	1.00 (Reference)	
GA	167 (45.13)	169 (45.67)	1.06 (0.78-1.44)	0.69
AA	27 (7.29)	29 (7.83)	1.17 (0.65-2.10)	0.58
GA+AA	194 (52.43)	198(53.51)	1.07 (0.79-1.45)	0.62
G Allele	519 (70.13)	513 (69.32)		
A Allele	221 (29.86)	227 (30.67)		
MAF	0.30	0.31		
OVERALL LUNG CANCER				
<i>Arg¹⁵⁶Arg</i>	Control N(%) N =370	Cases N(%) N=370	AOR (95% CI) ^a	<i>p^b</i>
CC	153 (41.35)	173 (46.75)	1.00 (Reference)	
CA	185 (50)	168(45.40)	0.82 (0.60-1.13)	0.23
AA	32 (8.64)	29 (7.83)	0.83 (0.47-1.46)	0.52
CA+AA	217 (58.64)	197 (53.24)	0.83 (0.61-1.12)	0.23
C Allele	491 (66.35)	525 (70.94)		
A Allele	249 (33.64)	225 (30.40)		
MAF	0.34	0.30		

Table 5.22 (b): Distribution of genotypes of <i>XPD</i> polymorphism depicting their association with lung cancer subtypes										
		SQCC			ADCC			SCLC		
<i>Lys⁷⁵¹Gln</i>	Controls n (%) N=370	Cases n(%) N=135	AOR (95% CI) ^a	<i>p^b</i>	Cases n(%) N=120	AOR (95% CI) ^a	<i>p^b</i>	Cases n(%) N=111	AOR (95% CI) ^a	<i>p^b</i>
AA	203(54.86)	70 (51.85)	1.00 (Reference)		60 (50)	1.0(Reference)		54 (48.64)	1.00 (Reference)	
AC	139(37.56)	48 (35.55)	0.94 (0.60-1.48)	0.80	42 (35)	1.02 (0.64-1.63)	0.91	39 (35.13)	1.15 (0.70-1.87)	0.56
CC	28 (7.5)	17 (12.59)	1.85 (0.91-3.76)	0.008	18 (15)	2.07 (1.05-4.08)	0.03	18 (16.21)	2.27 (1.12-4.63)	0.02
AC+CC	167(45.13)	65 (48.14)	1.09 (0.71-1.65)	0.67	60 (50)	1.23 (0.80-1.89)	0.3	57 (51.35)	1.35 (0.86-2.12)	0.18
		SQCC			ADCC			SCLC		
<i>Asp³¹²Asn</i>	Controls n (%) N=370	Cases n(%) N=135	AOR (95% CI) ^a	<i>p^b</i>	Cases n(%) N=120	AOR (95% CI) ^a	<i>p^b</i>	Cases n(%) N=111	AOR (95% CI) ^a	<i>p^b</i>
GG	176(47.56)	66 (48.88)	1.00 (Reference)		56(46.66)	1.00 (Reference)		48 (43.24)	1.00 (Reference)	
GA	167(45.13)	54 (40)	0.86 (0.55-1.33)	0.51	58(48.33)	1.19 (0.77-1.84)	0.42	55 (49.54)	1.23 (0.77-1.96)	0.37
AA	27 (7.29)	15 (11.11)	1.62 (0.78-3.37)	0.19	6 (5)	0.89 (0.33-2.34)	0.81	8 (7.20)	1.29 (0.53-3.14)	0.56
GA+AA	194(52.43)	69 (51.11)	0.75 (0.48-1.16)	0.20	64(53.33)	1.15 (0.75-1.76)	0.51	63 (56.75)	1.24 (0.79-1.95)	0.34
		SQCC			ADCC			SCLC		
<i>Arg¹⁵⁶Arg</i>	Controls n (%) N=370	Cases n(%) N=135	AOR (95% CI) ^a	<i>p^b</i>	Cases n(%) N=120	AOR (95% CI) ^a	<i>p^b</i>	Cases n(%) N=111	AOR (95% CI) ^a	<i>p^b</i>
CC	153(41.35)	75 (55.55)	1.00 (Reference)		48 (40)	1.00(Reference)		49 (44.14)	1.00 (Reference)	
CA	185 (50)	47 (34.81)	0.63 (0.40-0.99)	0.04	64(53.33)	1.04 (0.67-1.63)	0.83	54 (48.64)	0.95 (0.59-1.51)	0.84
AA	32 (8.64)	13 (9.62)	1.07 (0.49-2.34)	0.84	8 (66.67)	0.73 (0.31-1.73)	0.48	8 (7.20)	0.87 (0.37-2.07)	0.76
CA+AA	217(58.64)	60 (44.44)	0.69 (0.45-1.05)	0.08	72 (60)	1.00 (0.65-1.54)	0.97	62 (55.85)	0.82 (0.52-1.31)	0.41

(a)-AdjustedOdds 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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. Abbreviations: ADCC=Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma. All the bold number indicates the significant values.

5.5.5.2 Effect of smoking on the *XPD* polymorphism and their association with lung cancer risk

We investigated whether the three *XPD* variants in the present study could modify the association between exposure to smoking and lung cancer risk (Table 5.23). In case of *XPD* *Lys*⁷⁵¹*Gln* polymorphic site, the frequency distribution suggested that smokers with mutant genotype (CC) were more prevalent among cases than controls (14.85% vs 7.22%) whereas the wild genotype (AA) was more expressed in controls (54.75%) than cases (50.16%). We also observed that among smokers the CC genotype (mutant) was more strongly associated with lung cancer than wild genotype (AA) (OR: 2.11; 95%CI: 1.75-3.86; p=**0.01**) for *Lys*⁷⁵¹*Gln* polymorphism. In case of heavy smokers, the lung cancer patients with mutant genotype (CC) were more represented in cases than controls (14.89% vs 7.29%) whereas the wild genotype (AA) was more prominent in controls than cases (54.16% vs 49.64%). The percentage of heterozygous genotype (AC) for heavy smokers among controls and cases was 38.54% and 35.46% respectively. For *Lys*⁷⁵¹*Gln* polymorphism, heavy smokers with mutant genotype were associated with 2-fold risk for lung cancer (OR: 2.31; 95%CI: 0.90-5.94; p= 0.08). Individuals carrying mutant (CC) and heterozygous genotype (AC) when combined as a single genotype were associated with increased risk amongst heavy smokers for *Lys*⁷⁵¹*Gln* polymorphism when compared with the homozygous wild genotype (OR: 1.77; 95%CI: 1.00-3.13; p=**0.04**). There was no significant association observed for *Lys*⁷⁵¹*Gln* polymorphism and lung cancer risk in non smokers as shown in table 5.24.

For *Asp*³¹²*Asn* polymorphism, smokers with wild type genotype (GG) were more illustrated in controls than cases (47.52% vs 44.22%) while the mutant genotype (AA) was more represented in cases than controls (9.24% vs 7.22%). The percentage of heterozygous genotype (GA) was almost similar in both cases and controls (46.53% vs 45.24%). There was no significant association observed between lung cancer susceptibility and *Asp*³¹²*Asn* polymorphism in smoking population (Table 5.23). Among light smokers, heterozygous genotype (GA) was more represented in cases than controls (50% vs 41.31%) whereas the wild genotype (GG) was more frequent among controls than cases (52.09% vs 39.50%). The mutant genotype was at higher frequency in cases than controls (10.49% vs 6.58%). An increased risk was reported for all the genotypic variants *i.e.* mutant (AA), heterozygous (GA) and combined genotype (GA+AA) of *Asp*³¹²*Asn* when compared with the reference genotype in light smokers (AA genotype: OR:

2.45; 95%CI: 1.05-5.69; $p=0.03$, GA genotype: OR: 1.65; 95%CI: 1.02-2.65; $p=0.03$, GA+AA genotype: OR: 2.03; 95%CI: 1.28-3.23, $p=0.02$). *Asp³¹²Asn* polymorphism does not possess any significant association with lung cancer risk among non-smokers as shown in table 5.23.

For the last SNP i.e. *Arg¹⁵⁶Arg* variants did not confer any contribution towards lung cancer susceptibility neither in smokers nor in non-smokers as shown in table 5.23. The frequency distribution of smokers with mutant (AA) genotype was more in cases when compared with controls (8.25% vs 7.22%) while the wild genotype (CC) was found to be present in 47.85% of cases and 44.85% in controls respectively. As a result, no visible trend for risk was observed in case of smokers carrying the mutant (AA) genotype (OR=1.12, 95% C.I: 0.58-2.16, $p=0.72$), although result was not significant. Also, in case of non-smokers a lower frequency of subjects having the mutant (AA) genotype was observed among cases (5.97%) in comparison to controls (12.14%) but the association analysis exhibited a lack of any significant correlation between lung cancer risk and this polymorphic site. When smokers were further divided into light and heavy smokers, the frequency of heterozygous genotype had almost similar distribution among cases and controls (43.26% vs 46.26%) in heavy smokers. The wild genotype (CC) showed over expression in controls in comparison with cases (51.04% vs 47.51%) The variant genotype showed an increased lung cancer risk among heavy smokers, although results were not significant (OR: 1.43; 95%CI: 0.65-1.91, $p=0.68$). Among light smokers, the variant genotype showed decreased risk for the disease (OR: 0.84; 95%CI: 0.36-1.97, $p=0.69$) but the results were not significant as controls and cases both possess almost similar frequency distribution (8.38% vs 7.40%).

Table 5.23: Distribution of genotypes of <i>XPD</i> polymorphism according to their smoking status and its association with lung cancer risk								
	SMOKERS				NON SMOKERS			
<i>Lys</i> ⁷⁵¹ <i>Gln</i>	Controls n (%) N =263	Cases n (%) N=303	AOR (95% CI) ^a	<i>p</i> ^b	Controls n (%) N =107	Cases n (%) N=67	AOR (95% CI) ^a	<i>p</i> ^b
AA	144 (54.75)	152 (50.16)	1.00 (Reference)		59 (55.14)	35 (52.28)	1.00 (Reference)	
AC	100 (38.02)	106 (34.98)	1.06 (0.73-1.53)	0.73	39 (36.44)	24 (35.82)	1.00 (0.50-1.99)	0.98
CC	19 (7.22)	45 (14.85)	2.11 (1.15-3.86)	0.01	9 (8.41)	8 (11.94)	1.51 (0.52-4.33)	0.44
AC+CC	119 (32.16)	151 (49.83)	1.23 (0.87-1.74)	0.22	48 (44.85)	32 (47.76)	1.12 (0.59-2.11)	0.71
<i>Asp</i> ³¹² <i>Asn</i>								
GG	125 (47.52)	134 (44.22)	1.00 (Reference)		51 (47.66)	38 (56.71)	1.00 (Reference)	
GA	119 (45.24)	141 (46.53)	1.12 (0.79-1.61)	0.50	48 (44.85)	28 (41.79)	0.77 (0.40-1.47)	0.43
AA	19 (7.22)	28 (9.24)	1.56 (0.82-2.98)	0.17	8 (7.76)	1 (1.49)	0.20 (0.02-1.73)	0.14
GA+AA	138 (52.47)	169 (55.77)	1.19 (0.84-1.68)	0.30	56 (52.33)	29 (43.28)	0.69 (0.36-1.31)	0.26
<i>Arg</i> ¹⁵⁶ <i>Arg</i>								
CC	117 (44.48)	145 (47.85)	1.00 (Reference)		36 (33.64)	29 (43.28)	1.00 (Reference)	
CA	127 (48.28)	133 (43.89)	0.88 (0.62-1.27)	0.51	58 (54.20)	34 (50.74)	0.70 (0.36-1.37)	0.30
AA	19 (7.22)	25 (8.25)	1.12 (0.58-2.16)	0.72	13 (12.14)	4 (5.97)	0.39 (0.11-1.38)	0.14
CA+AA	146 (55.51)	158 (52.14)	0.92 (0.65-1.29)	0.63	71 (66.35)	38 (56.71)	0.64 (0.33-1.23)	0.18
	HEAVY SMOKER				LIGHT SMOKER			
<i>Lys</i> ⁷⁵¹ <i>Gln</i>	Controls n (%) N =96	Cases n (%) N=141	AOR (95% CI) ^a	<i>p</i> ^b	Controls n (%) N =167	Cases n (%) N=162	AOR (95% CI) ^a	<i>p</i> ^b
AA	52 (54.16)	70 (49.64)	1.00 (Reference)		92 (55.08)	82 (50.61)	1.00 (Reference)	
AC	37 (38.54)	50 (35.46)	1.62 (0.87-3.00)	0.12	63 (37.72)	56 (34.56)	1.01 (0.62-1.64)	0.94
CC	7 (7.29)	21 (14.89)	2.31 (0.90-5.94)	0.08	12 (7.18)	24 (14.81)	1.93 (0.87-4.27)	0.10
AC+CC	44 (45.83)	71 (50.35)	1.77 (1.00-3.13)	0.04	75 (44.91)	80 (49.38)	1.17 (0.74-1.83)	0.48
<i>Asp</i> ³¹² <i>Asn</i>								
GG	38 (39.58)	70 (49.64)	1.00 (Reference)		87 (52.09)	64 (39.50)	1.00 (Reference)	
GA	50 (52.08)	60 (42.55)	0.64 (0.36-1.12)	0.12	69 (41.31)	81 (50)	1.65 (1.02-2.65)	0.03
AA	8 (8.33)	11 (7.80)	0.78 (0.28-2.16)	0.64	11 (6.58)	17 (10.49)	2.45 (1.05- 5.69)	0.03
GA+AA	58 (60.41)	71 (50.35)	0.76 (0.43-1.32)	0.33	80 (47.90)	98 (60.49)	2.03 (1.28-3.23)	0.02
<i>Arg</i> ¹⁵⁶ <i>Arg</i>								
CC	49 (51.04)	67 (47.51)	1.00 (Reference)		68 (40.71)	78 (48.14)	1.00 (Reference)	
CA	42 (43.75)	61 (43.26)	1.05 (0.60-1.84)	0.84	85 (50.89)	72 (44.44)	0.78 (0.49-1.25)	0.31
AA	5 (5.20)	13 (9.21)	1.43 (0.46-4.44)	0.52	14 (8.38)	12 (7.40)	0.84 (0.36-1.97)	0.69
CA+AA	47 (48.95)	74 (52.48)	1.11 (0.65-1.91)	0.68	99 (59.28)	84 (51.85)	0.79 (0.50-1.25)	0.32

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

5.5.5.3 Haplotype analysis of *XPB* variants

We analyzed haplotypes using SHEsis program platform (table 5.24). Our analysis showed that the three SNPs were in weak linkage disequilibrium with each other in this study population: *XPB Lys⁷⁵¹Gln* and *XPB Asp³¹²Asn* ($D=0.019$, $r^2=0.00$); *XPB Asp³¹²Asn* and *XPB Arg¹⁵⁶Arg* ($D=0.035$, $r^2=0.01$); *XPB Lys⁷⁵¹Gln* and *XPB Arg¹⁵⁶Arg* ($D=0.065$, $r^2=0.004$). The haplotypes were composed of three coding SNP that were located across 68.734kb on 19q13.3 region. There was no statistically significant difference in the overall haplotype distribution between cases and controls (global $\chi^2=9.73$, $p=0.20$). All the exploratory haplotypes did not show any significant effect. The AAA haplotype was more expressed among controls (8%) than cases (6%), thus illustrated a close to significant protective effect in lung cancer patients (OR: 0.69; 95%CI: 0.46-1.02; $p=0.06$). CGC haplotype was slightly less represented in controls (12%) than cases (14%). While CGA (6%) AAC (14% vs 13%), (CAC (6% vs 5%) and CAA (4% vs 3%) haplotype was almost similarly expressed in cases and controls. All these haplotypes did not represent any significant association with lung cancer risk.

Table 5.24: *XPB* Haplotypes and their association with lung cancer risk

<i>Lys⁷⁵¹Gln, Asp³¹²Asn & Arg¹⁵⁶Arg</i>	Case frequency, n (%)	Control frequency, n (%)	OR (95% CI) ^a	p^b -value
AAA	0.06 (46)	0.08 (65)	0.69 (0.46-1.02)	0.06
AAC	0.14 (104)	0.13 (96)	1.09 (0.81-1.48)	0.53
AGA	0.13(97.94)	0.15 (116)	0.81 (0.61-1.09)	0.17
AGC	0.34 (256)	0.36 (268)	0.93 (0.75-1.15)	0.51
CAA	0.04 (308)	0.03 (22)	1.41 (0.81-2.46)	0.22
CAC	0.06 (46)	0.05 (38)	1.22 (0.78-1.91)	0.36
CGA	0.06 (50)	0.06 (45)	1.10 (0.72-1.67)	0.63
CGC	0.14 (110)	0.12 (90)	1.25 (0.93-1.69)	0.13

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding p -values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either haplotypic distribution or frequencies between the cases and controls. Global χ^2 test=9.73, p^b -value = 0.20.

5.5.5.4 Combinatorial effect of polymorphic variants of *XPD* on lung cancer risk

To study SNP-SNP interaction, different genotypes of *XPD* variants (*Lys*⁷⁵¹*Gln*, *Asp*³¹²*Asn* and *Arg*¹⁵⁶*Arg*) were evaluated to predict their relationship towards lung cancer susceptibility (table 5.25). However, no significant association was observed between the combined *XPD* variants and lung cancer risk. Different SNP-SNP combinations were applied for three *XPD* SNPs. Group 0 which composed of wild type genotype was regarded as the reference. Different dual combinations between three SNPs were evaluated to check whether any combinatorial relationship between the *XPD* SNPs towards lung cancer susceptibility. The first combination between *Lys*⁷⁵¹*Gln* and *Asp*³¹²*Asn* polymorphism indicated no significant association towards lung cancer susceptibility. The wild genotype (AA+GG) (group 0) was more represented in controls than cases (62.58% vs 58.33%) whereas heterozygous genotype (AC+GA) (group 1) was more frequent in cases than controls (41.66% vs 37.41%). On applying logistic regression analysis, heterozygous genotype (Group 1) for the above dual combination showed no association towards lung cancer risk when compared with the reference genotype (OR: 1.23; 95%CI: 0.77-1.93; p=0.37). In case of *Asp*³¹²*Asn* and *Arg*¹⁵⁶*Arg* combination, wild type genotype (GG+CC) (Group 0) was found to be represented in both cases and control (50.29% vs 50%). The heterozygous genotype (47.95% vs 47.43%) as well as mutant genotype (1.75% vs 2.56%) was also similarly represented in both cases and controls. There was no significant relationship which can be concluded through this SNP-SNP interaction.

The other dual combination which has been applied was between *Arg*^{156A}*arg* and *Lys*⁷⁵¹*Gln* polymorphism, the heterozygous genotype (CA+AC) (group 1) was more prevalent in controls (44.23%) than cases (36.84%) whereas the variant genotype (AA+CC) (group 2) was slightly more expressed in cases (2.63%) than controls (3.86%) as shown in table 5.25. Logistic regression analysis showed a protective trend could be observed for this polymorphic combination in heterozygous genotype (OR: 0.76; 95%CI: 0.47-1.22; p=0.25) when compared with the reference group but results were not significant. The only possible triple combination with these SNPs was *Lys*⁷⁵¹*Gln*, *Arg*¹⁵⁶*Arg* and *Asp*³¹²*Asn*. It indicates over representation of wild genotype (group 0) in cases (56.41%) than controls (40.90%) while heterozygous genotype (group 1) was less presented in controls (43.58%) than cases (59.09%). There was no significant

association which can be concluded from this triple combination explaining its association with lung cancer susceptibility.

Table 5.25: Combination of genotypes of XPD polymorphic variants and their relationship with the risk of lung cancer				
Genotypes	Cases (n%)	Controls (n%)	Adjusted OR (95% CI)^b	p-value^a
<i>Lys⁷⁵¹Gln, Asp³¹²Asn and Arg¹⁵⁶Arg</i>	N=78	N=66		
0 (AA+GG+CC)	44 (56.41)	27 (40.90)	1.00 (Reference)	
1(AC+GA+CA)	34 (43.58)	39 (59.09)	1.24 (0.62-2.47)	0.54
2(CC+AA+AA)	-	-	-	-
3(1+2)	-	-	-	-
<i>Lys⁷⁵¹Gln and Asp³¹²Asn</i>	N=156	N=155		
0 (AA+GG)	91 (58.33)	97 (62.58)	1.00 (Reference)	
1(AC+GA)	65 (41.66)	58 (37.41)	1.23 (0.77-1.93)	0.37
2 (CC+AA)	-	-	-	-
3(1+2)	-	-	-	-
<i>Asp³¹²Asn and Arg¹⁵⁶Arg</i>	N=171	N=156		
0 (GG+CC)	86 (50.29)	78 (50)	1.00 (Reference)	
1 (GA+CA)	82 (47.95)	74 (47.43)	0.96 (0.61-1.52)	0.88
2 (AA+AA)	3 (1.75)	4 (2.56)	0.71 (0.14-3.47)	0.67
3(1+2)	85 (49.70)	78 (47.43)	0.98 (0.59-1.38)	0.65
<i>Arg¹⁵⁶Arg and Lys⁷⁵¹Gln</i>	N= 152	N=156		
0 (CC+AA)	92 (60.52)	81 (51.92)	1.00 (Reference)	
1 (CA+AC)	56 (36.84)	69 (44.23)	0.76 (0.47-1.22)	0.25
2 (AA+CC)	4 (2.63)	6 (3.84)	0.74 (0.18-2.91)	0.66
3(1+2)	60 (39.47)	75 (48.07)	0.72 (0.66-1.55)	0.94

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status. (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 3: combined hetero and mutant genotype

5.5.5.5 Relationship of Clinico-pathological features with XPD polymorphic variants

To evaluate association of the three polymorphic variants of *XPD* (*Lys⁷⁵¹Gln*, *Asp³¹²Asn*, *Arg¹⁵⁶Arg*) on various clinico-pathological features, we segregated patients according to clinical stage (III vs IV), primary tumor extension (T3+T4 vs T0+T1+T2) and metastatic status (positive vs negative) (Table 5.26).

On the basis of clinical stage, the patients were stratified into stage III and stage IV as shown in table 5.26. In case of *Lys⁷⁵¹Gln* polymorphism, the lung cancer patients with heterozygous genotype (AC) were found to have a similar frequency among stage IV (33.71%) as compared to stage III (34.33%). On the contrary, the mutant genotype (CC) was more prevalent in stage IV

subjects of the disease as compared to stage III patients (16.56% vs 12%). On applying logistic linear regression analysis, lung cancer patients having mutant genotype (CC) have higher risk for disease in stage IV patients comparison with stage III patients but results were not significant (OR= 1.69, 95% CI: 0.87-3.30, $p=0.12$). In case of *Asp³¹²Asn* polymorphism, homozygous variant genotype (AA) was more expressed among stage III (10.28%) subgroup as compared to stage IV (7.36%) subgroup. The study subjects carrying single allelic variant genotype (GA) and diagnosed with clinical stage IV was similarly represented in stage III (44.78% vs 46.28%). Since the difference was not significant logistic regression analysis suggested no association towards lung cancer (OR= 0.93, 95% CI: 0.58-1.47, $p=0.75$) as shown in table 5.27. In case of *Arg¹⁵⁶Arg* polymorphism, the single allelic variant (CA) was 39.42% and 48.46% among stage III and Stage IV respectively. The percentage of subjects carrying mutant genotype (AA) was 10.28% for stage III and 6.13% for stage IV subjects. Regression analysis indicated no involvement of stage and the polymorphism towards lung cancer predisposition as shown in table 5.26.

The other parameter evaluated for *XPD* variants was primary tumor extension as shown in table 5.26. The patients were stratified into two groups namely, T3+T4 and T1+T2+T0. The investigative study of *Lys⁷⁵¹Gln* polymorphism for its correlation with primary tumor extension, suggested that frequency of subjects harboring homozygous variant genotype (CC) was higher among T3+T4 tumor size (23.80%) as compared subjects with T1+T2+T0 tumor extension (18.84%). Regression analysis revealed no significant association between polymorphic variant and primary tumor extension (OR: 0.49; 95%CI: 0.23-1.06; $p=0.07$) for *Lys⁷⁵¹Gln* polymorphic variant. In case of *Asp³¹²Asn* polymorphism, the distribution of patients carrying heterozygous genotype (GA) was almost similar among T3+T4 and T1+T2+T0 subgroup (45.28% vs 46.37%). Similarly, the percentage of individuals carrying mutant genotype (AA) was higher in T3+T4 subgroup in comparison with T1+T2+T0 subgroup (9.42% vs 4.34%). When compared to wild genotype (GG), no correlation between T3/T4 and T1/T2/T0 was observed (OR: 2.00; 95%CI: 0.55-7.22; $p=0.28$). In case of *Arg¹⁵⁶Arg* polymorphism, the heterozygous genotype (CA) had a comparable difference in the frequency for both T0+T1+T2 (52.17%) and T3+T4 (42.02%) subgroups, however the difference was not significant (OR: 0.66; 95%CI: 0.38-1.16; $p=0.15$).

The relationship between metastatic status and *XPD* polymorphic variants were detailed in table 5.26. The genotypic frequencies of combined variant (AC+CC) for *Lys*⁷⁵¹*Gln* polymorphism exhibited a significant difference between patients who showed metastasis (53.79%) and those showing absence of metastasis (44.91%). On applying logistic regression analysis a significant association was observed between *Lys*⁷⁵¹*Gln* polymorphism with metastasis (OR= 1.58, 95% CI: 1.01-2.49, *p*=**0.04**) as shown in table 5.26. In case of *Asp*³¹²*Asn* polymorphism, lung cancer subjects with distant metastasis and carrying the mutant genotype (AA) were more expressed as compared to patients without metastasis harboring mutant genotype (6.32% vs 10.16%). This polymorphic variant of *XPD* did not show any correlation with metastasis as shown in table 5.27 (OR: 0.60; 95%CI: 0.25-1.44; *p*=0.26). In case of *Arg*¹⁵⁶*Arg* polymorphism, the percentage of subjects carrying mutant genotype (AA) showed a significant difference between in distribution of patients lacking metastasis (10.16%) and those who have metastasis (3.16%). Regression analysis showed a close to significant association for *Arg*¹⁵⁶*Arg* polymorphism and metastasis for mutant genotype (AA) suggesting a protective association among patients having lung cancer metastasis (M1) was compared to those patients who lack metastasis (M0) (OR= 0.42, 95% CI: 0.15-1.15, *p*=0.09) as shown in table 5.26.

Table 5.26: Distribution of Clinico-pathological features and their association with lung cancer risk in patients with *XPD* polymorphism

Variable	Clinical stage		AOR (95% CI) ^a	<i>p</i> ^b	TNM stage		AOR (95% CI) ^a	<i>p</i> ^b	Metastasis		AOR (95% CI) ^a	<i>p</i> ^b
	III	IV			T3+T4	T2+T1+T0			Positive	Negative		
<i>Lys</i> ⁷⁵¹ <i>Gln</i>	175	163			276	69			158	187		
AA	95 (54.28)	80 (49.07)	Reference		147 (59.51)	29 (42.02)	Reference		73 (46.20)	103 (55.08)	Reference	
AC	59 (33.71)	56 (34.33)	1.16 (0.71-1.90)	0.54	94 (34.05)	27 (39.13)	0.66 (0.36-1.20)	0.17	58 (36.70)	62 (33.15)	1.46 (0.89-2.40)	0.13
CC	21 (12)	27 (16.56)	1.69 (0.87-3.30)	0.12	35 (23.80)	13 (18.84)	0.49 (0.23-1.06)	0.07	27 (17.08)	22 (11.76)	1.89 (0.96-3.74)	0.06
AC+CC	80 (45.71)	83 (50.92)	1.31 (0.84-2.05)	0.22	129 (46.73)	40 (57.97)	0.62 (0.36-1.06)	0.08	85 (53.79)	84 (44.91)	1.58 (1.01-2.49)	0.04
<i>Asp</i> ³¹² <i>Asn</i>												
GG	76 (43.42)	78 (47.85)	Reference		125 (45.28)	34 (49.27)	Reference		82 (51.89)	77 (41.17)	Reference	
GA	81 (46.28)	73 (44.78)	0.93 (0.58-1.47)	0.75	125 (45.28)	32 (46.37)	1.01 (0.58-1.75)	0.95	66 (41.77)	91 (48.66)	0.73 (0.45-1.16)	0.18
AA	18 (10.28)	12 (7.36)	0.69 (0.29-1.62)	0.40	26 (9.42)	3 (4.34)	2.00 (0.55-7.22)	0.28	10 (6.32)	19 (10.16)	0.60 (0.25-1.44)	0.26
GA+AA	99 (56.57)	85 (52.14)	0.89 (0.57- 1.39)	0.62	151 (54.71)	35 (50.72)	1.10 (0.64-1.88)	0.72	76 (48.10)	110 (58.82)	0.71 (0.45-1.12)	0.14
<i>Arg</i> ¹⁵⁶ <i>Arg</i>												
CC	88 (50.28)	74 (45.39)	Reference		139 (50.36)	28 (40.57)	Reference		71 (44.93)	96 (51.33)	Reference	
CA	69 (39.42)	79 (48.46)	1.29 (0.81-2.05)	0.27	116 (42.02)	36 (52.17)	0.66 (0.38-1.16)	0.15	79 (50)	72 (38.50)	1.40 (0.88-2.23)	0.14
AA	18 (10.28)	10 (6.13)	0.63 (0.26-1.50)	0.30	21 (7.60)	5 (7.24)	0.84 (0.29-2.44)	0.76	8 (3.16)	19 (10.16)	0.42 (0.15-1.15)	0.09
CA+AA	87 (49.71)	89 (54.60)	1.16 (0.74-1.80)	0.30	137 (49.63)	41 (59.42)	0.68 (0.40-1.17)	0.17	88 (55.69)	91 (48.66)	1.20 (0.77-1.88)	0.40

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status.(b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

5.5.5.6 MDR analysis

Non parametric analysis was applied using MDR approach in order to predict the high order interactions of three *XPB* variants (*Lys*⁷⁵¹*Gln*, *Asp*³¹²*Asn*, *Arg*¹⁵⁶*Arg*) and their role in predicting the lung cancer predisposition in North Indians. The MDR analysis was performed on dataset consisted of subject with or without lung cancer depicting average cross validation model (CVC) and average prediction error (Table 5.27; Fig 5.20A). The analysis resulted in best interaction model which includes *XPB Lys*⁷⁵¹*Gln*, *Asp*³¹²*Asn*, *Arg*¹⁵⁶*Arg* with CVC of 100 and prediction error of 0.45 ($p < 0.0001$). A higher CVC and a minimum prediction error are preferred for the best results. The two factor model for evaluating lung cancer risk had CVC of 100 and showing improvised error of 0.44 ($p = 0.0005$). *XPB Lys*⁷⁵¹*Gln* appeared as the best one factor model showing lung cancer risk with CVC of 99 and prediction error of 0.48 ($p = 0.003$). The interaction between the SNPs explained through dendrogram used for cluster analysis. The entropy dendrogram illustrates the type of interaction taking place between the *XPB* variants which are important factors of the models predicted by MDR. The length of the each line connecting these variants signifies the intensity of interaction. The colors of the lines connecting these SNP's indicate towards the type of interaction taking place between them. More synergy was observed between *XPB Asp*³¹²*Asn* and *Arg*¹⁵⁶*Arg* as the smallest line had been connecting them whereas a weak interaction was observed between *XPB Lys*⁷⁵¹*Gln* and other two polymorphic variants.

Table 5.27: Multifactor Dimensionality reduction (MDR) analysis showing interactions of *XPB* variants with lung cancer risk

No. of risk factors	Best interaction model	Cross Validation Consistency (CVC)	Prediction error	<i>p</i> – value
1	<i>XPB Lys</i> ⁷⁵¹ <i>Gln</i>	99/10	0.48	0.003
2	<i>XPB Lys</i> ⁷⁵¹ <i>Gln</i> , <i>XPB Arg</i> ¹⁵⁶ <i>Arg</i>	100/100	0.44	0.0005
3	<i>XPB Lys</i> ⁷⁵¹ <i>Gln</i> , <i>XPB Asp</i> ³¹² <i>Asn</i> , <i>XPB Arg</i> ¹⁵⁶ <i>Arg</i>	100/100	0.45	<0.0001

5.5.5.7 CART analysis

The high order gene-gene interactions with lung cancer risk were analyzed using CART analysis (Table 5.28; Fig 5.20B). Final tree consisting 4 terminal nodes was generated. The root split was present on *XPD Arg¹⁵⁶Arg* suggesting it to be the strongest factor for lung cancer risk amongst other variants. Further subsequent split was on *XPD Lys⁷⁵¹Gln* and *Asp³¹²Asn*. To understand potential mechanism for increased lung cancer risk with SNPS in CART analysis, we also applied logistic regression analysis in the final model selected by CART via evaluating OR and 95%CI. The node with minimum case rate was considered as the reference i.e. node 4 showing case rate of 43.8%. Individuals carrying wild genotype of *XPD Lys⁷⁵¹Gln* and mutant (heterozygous + mutant) genotype of *XPD Arg¹⁵⁶Arg* represent node 4. Subjects with *XPD Asp³¹²Asn (M)*, *Lys⁷⁵¹Gln (M)* and *Arg¹⁵⁶Arg (M)* showed highest risk towards the lung cancer (OR: 1.56; 95%CI: 0.98-2.46; p=**0.05**). Whereas individuals with wild genotype of *Arg¹⁵⁶Arg* possessed a significantly increased risk for the disease (OR: 1.46; 95%CI: 1.03-2.06, p=**0.03**). Node 3 carrying individuals with a combined genotype of *XPD Asp³¹²Asn (W)*, *Lys⁷⁵¹Gln (M)* and *Arg¹⁵⁶Arg (M)* showed almost 2-fold increased risk for lung cancer (OR: 1.89; 95%CI: 0.68-1.89; p=0.61).

Table 5.28: Risk estimates based on CART analysis						
Terminal Node	Genotype for each terminal node	Control (n)	Case (n)	Case Rate ^a	OR (95% CI)	p ^b
Node 1	<i>XPD Arg¹⁵⁶Arg(W)</i>	153	174	53.2	1.46 (1.03-2.06)	0.03
Node 2	<i>XPDAsp³¹²Asn(M)/XPDLys⁷⁵¹Gln(M)/XPD Arg¹⁵⁶Arg(M)</i>	62	51	54.9	1.56 (0.98-2.46)	0.05
Node 3	<i>XPD Asp³¹²Asn(W)/XPD Lys⁷⁵¹Gln(M)/XPD Arg¹⁵⁶Arg(M)</i>	44	39	47.0	1.89 (0.68-1.89)	0.61
Node 4	<i>XPD Lys⁷⁵¹Gln(W)/XPD Arg¹⁵⁶Arg(M)</i>	122	95	43.8	1.0 ^c	
^a case rate is percentage LC cases among all individuals in each terminal node {case / (case+ control) x100}; ^b p<0.05 is considered statistically significant; ^c reference group						

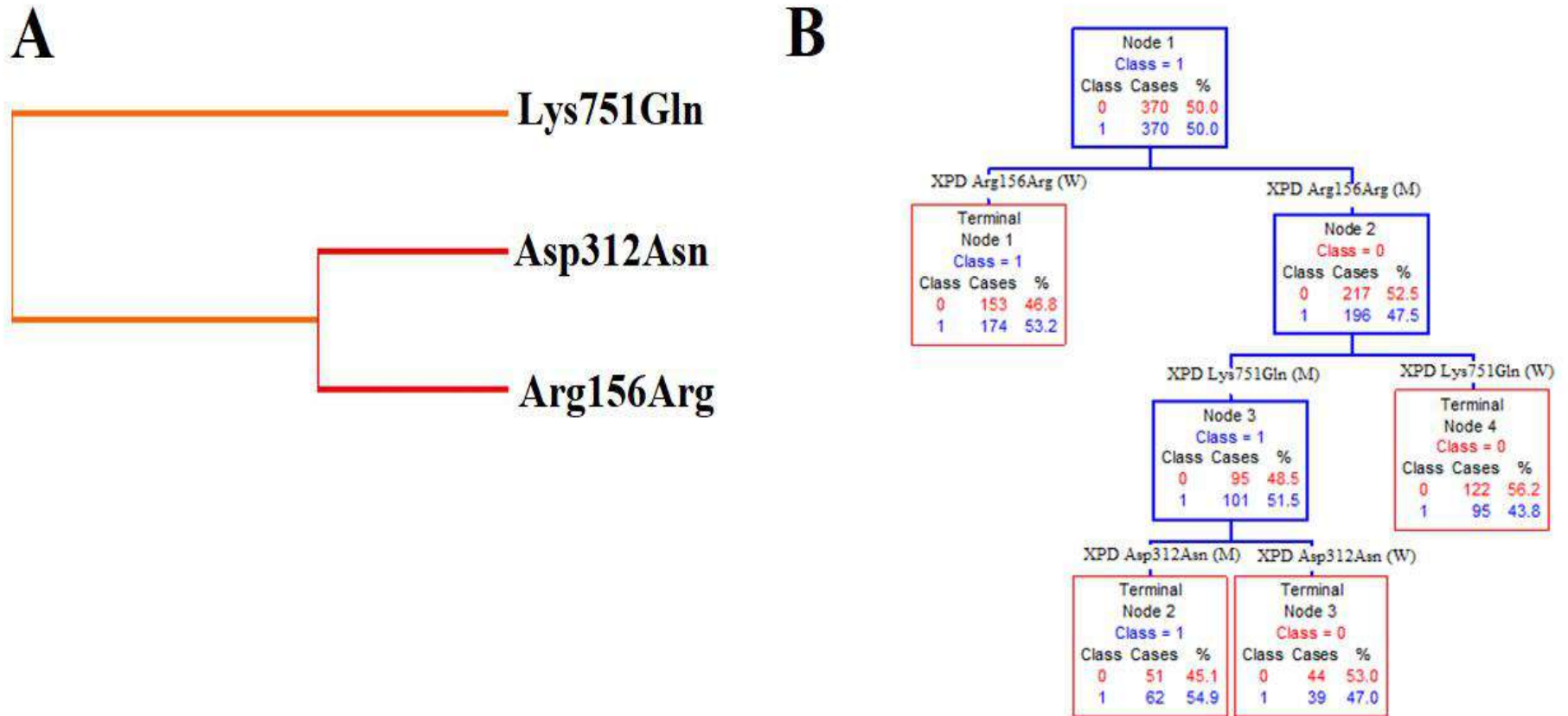


Fig 5.20: (A) Interaction dendrogram gained from the multifactor dimensionality reduction (MDR) showing gene-gene interaction for *XPD Lys⁷⁵¹Gln*, *XPD Asp³¹²Asn* and *XPD Arg¹⁵⁶Arg*. *XPD Asp³¹²Asn* and *XPD Arg¹⁵⁶Arg* show strong synergistic interaction whereas *XPD Lys⁷⁵¹Gln* a weak interaction with other two polymorphism **(B)** Classification and regression tree (CART) analysis depicting high order interaction among three polymorphisms (*Lys⁷⁵¹Gln*, *Asp³¹²Asn* and *Arg¹⁵⁶Arg*) of *XPD*

5.5.5.8 Association of *XPD* SNPs with overall survival of lung cancer patients

The survival analysis for 311 lung cancer patients and their association with *XPD* polymorphism has been depicted in Table 5.29. The cases with *XPD Asp³¹²Asn* (N=311; $\chi^2=1.86$; df=1; p=0.17) and *XPD Arg¹⁵⁶Arg* (N=311; $\chi^2= 0.41$; df=1; p=0.52) polymorphism were in agreement with HWE whereas the cases with *Lys⁷⁵¹Gln* polymorphism were not in agreement with HWE (N=311; $\chi^2= 6.95$; df=1; p=0.008). The wild type genotype for each polymorphism was regarded as the reference *i.e.* AA for *Lys⁷⁵¹Gln*, GG for *Asp³¹²Asn* and CC for *Arg¹⁵⁶Arg*. At first univariate analysis was applied on the study population after that multivariate analysis was performed. For Univariate analysis Kaplan Meier survival analysis was implemented. For multivariate analysis, Cox regression model was applied adjusted with different parameters such age, gender, smoking status, KPS, ECOG, stage and histology. For *Lys⁷⁵¹Gln* polymorphism, variant genotype (CC) was associated with better survival having a MST of 9.23 months, which was slightly higher than the reference genotype (AA) with survival of 7.23 months (log rank p=0.37). The mutant genotype (AA) for *XPD Asp³¹²Asn* had MST of 5.43 months when compared with wild genotype with survival of 9.43 months. *Arg¹⁵⁶Arg* polymorphism showed almost similar survival time was observed for heterozygous (8.63 months) as well as variant genotype (9.46 months) when compared with the reference type (7.13 months) as shown in table 5.30. No significant association was observed between *XPD* polymorphic variants and OS in lung cancer patients.

5.5.5.9 Association of *XPD* SNPs with overall survival in lung cancer histological subtypes

Overall survival in *XPD* polymorphism and their association with different histological subtypes of lung cancer has been explained in Table 5.30 (a, b, c). The genotypic distribution and overall survival of ADCC patients has been tabulated in table 5.30 (a), SQCC patients in table 5.30 (b) and SCLC patients in table 5.30 (c).

In case of *Lys⁷⁵¹Gln* polymorphism, the representation of heterozygous genotype (AC) was less in patients carrying wild type genotype (AA) for ADCC (38.46% vs 47.11%) and SQCC (36.75% vs 52.13%) subtype. ADCC patients carrying homozygous variant (CC) genotype showed a better OS as compared to wild type genotype (9.23 months vs 7.13 months; log rank

P= 0.62). Hazards Ratio obtained from multivariate analysis was 0.80 with insignificant *p* values of 0.28 as shown in table 5.31(a). In SQCC patients, heterozygous (AC) and mutant (CC) genotype suggested a survival period of almost 6.46 months when compared with the reference genotype (GG) (MST=7.13 months) for *Lys*⁷⁵¹*Gln* polymorphism as shown in table 5.31(b). In SCLC subjects carrying CC genotype showed a MST of 12.7 months as compared to subjects carrying the wild type genotype (12.7 vs 5.80 months; log rank P=0.43). On applying Cox regression model, a lower hazards ratio in case of both univariate analysis (HR= 0.61, 95% CI: 0.33-1.14, log rank *p*=0.16; Fig 5.21A) and multivariate analysis (HR¹= 0.49, 95% CI: 0.23-1.03, *p*=**0.05**) was observed as shown in table 5.31(c).

In case of *Asp*³¹²*Asn* polymorphism, ADCC patients having mutant (AA) genotype showed a lower MST when compared with the reference genotype (7.16 vs 9.46; log rank *p*=0.59). SQCC patients showing *Asp*³¹²*Asn* polymorphism showed a poor survival in heterozygous genotype as compared to SQCC patients having wild type genotype (GG) (MST=5.56 months; log rank *p*=0.07; Fig 5.21C) and an increased death probability was also observed in these patients (HR¹= 1.60, 95% CI: 1.02-2.53, *p*=**0.04**) on applying Cox regression model. The MST of SCLC subjects carrying mutant (AA) genotype for *Asp*³¹²*Asn* polymorphism was 1.70 months as compared to 7.23 months as in case of subjects having wild (GG) genotype. The hazards ratio obtained after both univariate (HR= 2.81, 95% CI: 0.66-11.94, *p*=**0.01**; Fig 5.21B) and multivariate analysis (HR¹= 13.38, 95% CI: 3.59-49.81, *p*=**0.0001**) was found to be significantly associated with high risk of death in mutant genotype and a poor prognosis. While for combined genotype (GA+AA), a better survival was observed when compared with reference genotype (GG) in SCLC patients (8.44 vs 7.23 months), a trend of low death ratio was observed but results were not significant (HR¹= 0.72, 95% CI: 0.45-1.16, *p*=0.18).

In *Arg*¹⁵⁶*Arg* polymorphism, no significant association was observed among OS and risk with histological subtypes. ADCC patients showed over presentation of heterozygous genotype (CA) than the wild genotype (52.88% vs 41.34%). All the three genotypes showed OS of almost 9 months when compared with the ADCC subjects with wild type genotype in ADCC subjects (MST= 7 months). In SQCC patients, frequency of heterozygous genotype (CA) was comparatively lower than in SQCC patients having wild genotype (30.76% vs 59.82%) exhibited

a trend of increased death ratio ((HR¹= 1.33, 95% CI: 0.81-2.18, *p*=0.25). Although results obtained were not significant. In SCLC patients showing *Arg*¹⁵⁶*Arg* polymorphism, all the three genotypes *i.e.* heterozygous (MST=8.26 months), mutant (MST=7.23 months), combined (MST=8.26 months) type showed almost similar survival time when compared with the reference type (MST=8.40 months). These results did not show any significant association for *Arg*¹⁵⁶*Arg* polymorphism with the OS on applying multivariate analysis as shown in table 5.30c.

Table 5.29: Association of XPD polymorphism and overall survival analysis in lung cancer patients								
				UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
Genotype	No. n% N=311	Dead n% N=256	Alive n% N=55	Median OS (months)	HR (95% CI) ^a	Log P	HR ¹ (95% CI) ^b	P- value
<i>Lys</i>⁷⁵¹<i>Gln</i>								
AA	152 (48.87)	127 (49.60)	25 (45.45)	7.13	1.00(Reference)		1.00(Reference)	
AC	116 (37.29)	95 (37.10)	21 (38.18)	8.03	0.94(0.72-1.23)	0.68	1.00 (0.76-1.31)	0.97
CC	43 (13.82)	34 (13.280)	9 (16.36)	9.23	0.84(0.58-1.21)	0.37	0.80 (0.54-1.19)	0.28
AC+CC	159 (51.12)	129 (50.39)	30 (54.54)	8.40	0.92(0.71-1.17)	0.50	0.95 (0.74-1.22)	0.95
<i>Asp</i>³¹²<i>Asn</i>								
GG	145 (46.62)	118 (46.09)	26 (47.27)	9.43	1.00(Reference)		1.00(Reference)	
GA	143 (45.65)	118 (46.09)	24 (43.63)	8.03	1.16 (0.90-1.51)	0.22	1.16 (0.89-1.50)	0.25
AA	24 (7.17)	20 (7.81)	4 (7.27)	5.43	1.30 (0.77-2.20)	0.26	1.28 (0.79-2.09)	0.31
GA+AA	166 (53.37)	138 (53.90)	28 (50.90)	7.13	1.18 (0.92-1.51)	0.17	1.16 (0.90-1.49)	0.22
<i>Arg</i>¹⁵⁶<i>Arg</i>								
CC	154(49.51)	127 (49.60)	27 (49.09)	7.13	1.00(Reference)		1.00(Reference)	
CA	133(42.76)	106 (41.40)	27 (49.09)	8.63	0.98 (0.76-1.27)	0.91	1.04 (0.80-1.37)	0.72
AA	24 (7.17)	23 (8.98)	1 (1.81)	9.46	1.18 (0.74-1.90)	0.44	1.25 (0.80-1.98)	0.31
CA+AA	157 (50.48)	129 (50.39)	28 (50.90)	8.80	1.01 (0.79-1.30)	0.88	1.08 (0.83-1.39)	0.54

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

Table 5.30 (a) :Association of XPD polymorphism and overall survival analysis in ADCC patients

Genotype	No. n% N=104	Dead n N=86	Alive n N=18	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (month s)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P-value
<i>Lys⁷⁵¹Gln</i>								
AA	49 (47.11)	41	8	7.13	1.00(Reference)		1.00(Reference)	
AC	40 (38.46)	33	7	9.00	0.89(0.56-1.40)	0.62	0.94 (0.57-1.56)	0.82
CC	15 (14.42)	12	3	9.23	0.85 (0.45-1.57)	0.62	0.68 (0.33-1.38)	0.29
AC+CC	55 (52.88)	45	10	9.36	0.88 (0.51-1.35)	0.56	0.93 (0.59-1.47)	0.76
<i>Asp³¹²Asn</i>								
GG	49 (47.11)	38	11	9.46	1.00(Reference)		1.00(Reference)	
GA	49 (47.11)	43	6	8.80	1.30 (0.84-2.02)	0.22	1.39 (0.88-2.20)	0.15
AA	6 (5.76)	5	1	7.16	1.28 (0.45-3.59)	0.59	1.48 (0.54-4.07)	0.44
GA+AA	55 (52.88)	48	7	7.60	1.31 (0.85-2.00)	0.20	1.41 (0.90-2.21)	0.13
<i>Arg¹⁵⁶Arg</i>								
CC	43 (41.34)	38	5	7	1.00(Reference)		1.00(Reference)	
CA	55 (52.88)	42	13	9	0.80 (0.52-1.27)	0.36	0.89 (0.57-1.40)	0.63
AA	6 (5.76)	6	-	9.90	1.00 (0.42-2.39)	0.98	1.07 (0.43-2.63)	0.87
CA+AA	61 (58.65)	48	13	9.13	0.83 (0.53-1.28)	0.39	0.90 (0.58-1.39)	0.90

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. ADCC=Adenocarcinoma; OS=Overall survival

Table 5.30 (b) :Association of *XPD* polymorphism and overall survival analysis in SQCC patients

Genotype	No. n% N= 117	Dead n N=95	Alive n N=22	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value
<i>Lys</i>⁷⁵¹<i>Gln</i>								
AA	60 (51.28)	48	12	8.63	1.00(Reference)		1.00(Reference)	
AC	43 (36.75)	35	8	7.26	1.74 (0.73-1.77)	0.54	1.12 (0.70-1.78)	0.62
CC	14 (11.96)	12	2	6.46	1.20 (0.61-2.35)	0.56	1.20 (0.63-2.29)	0.57
AC+CC	57 (48.71)	47	10	6.46	1.16 (0.77-1.74)	0.45	1.13 (0.74-1.72)	0.56
<i>Asp</i>³¹²<i>Asn</i>								
GG	58 (49.57)	46	12	9.43	1.00(Reference)		1.00(Reference)	
GA	46 (39.31)	39	7	5.56	1.47 (0.94-2.30)	0.07	1.60 (1.02-2.53)	0.04
AA	13 (11.11)	10	3	6.46	1.15 (0.56-2.37)	0.67	1.04 (0.51-2.10)	0.91
GA+AA	59 (50.42)	49	10	5.56	1.39 (0.92-2.09)	0.10	1.46 (0.15-2.24)	0.08
<i>Arg</i>¹⁵⁶<i>Arg</i>								
CC	70 (59.82)	54	16	6.86	1.00(Reference)		1.00(Reference)	
CA	36 (30.76)	31	5	8.03	1.19 (0.75-1.41)	0.42	1.33 (0.81-2.18)	0.25
AA	11 (9.40%)	10	1	7.26	1.25 (0.60-2.61)	0.50	1.23 (0.61-2.45)	0.55
CA+AA	47 (40.17)	41	6	8.03	1.22 (0.80-1.84)	0.33	1.28 (0.83-1.97)	0.26

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. SQCC=Squamous cell carcinoma; OS=Overall survival

Table 5.30 (c) :Association of XPD polymorphism and overall survival analysis in SCLC patients

Genotype	No. n% N=90	Dead n% N=75	Alive n% N=15	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P-value
<i>Lys⁷⁵¹Gln</i>								
AA	43 (47.77)	38	5	5.80	1.00(Reference)		1.00(Reference)	
AC	43 (47.77)	27	6	8.26	0.82 (0.50-1.34)	0.43	0.81 (0.48-1.36)	0.44
CC	4 (4.44)	10	4	12.7	0.61 (0.33-1.14)	0.16	0.49 (0.23-1.03)	0.05
AC+CC	47 (52.22)	37	10	9.86	0.75 (0.47-1.18)	0.21	0.70 (0.44-1.12)	0.14
<i>Asp³¹²Asn</i>								
GG	38 (42.22)	35	3	7.23	1.00(Reference)		1.00(Reference)	
GA	46 (51.11)	35	11	9	0.68 (0.42-1.10)	0.11	0.62 (0.38-1.02)	0.06
AA	6 (6.67)	5	1	1.70	2.81 (0.66-11.94)	0.01	13.38 (3.59-49.81)	0.0001
GA+AA	52 (57.77)	40	12	8.4	0.76 (0.48-1.21)	0.23	0.72 (0.45-1.16)	0.18
<i>Arg¹⁵⁶Arg</i>								
CC	41 (45.55)	35	6	8.40	1.00(Reference)		1.00(Reference)	
CA	42 (46.67)	33	9	8.26	0.87 (0.54-1.41)	0.58	0.82 (0.49-1.37)	0.47
AA	7 (7.77)	7	-	7.23	1.20 (0.50-2.87)	0.64	1.08 (0.45-2.60)	0.85
CA+AA	49 (54.44)	40	9	8.26	0.91 (0.58-1.44)	0.69	0.85 (0.52-1.40)	0.54

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. SCLC=Small cell lung cancer; OS=Overall survival

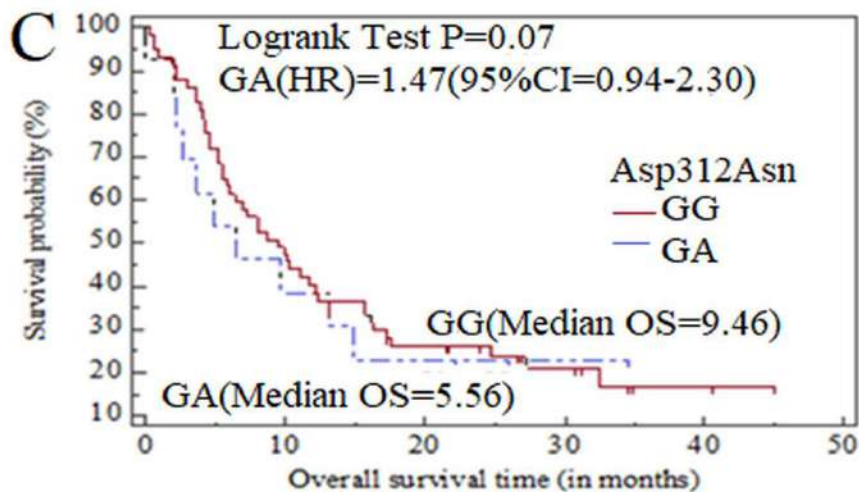
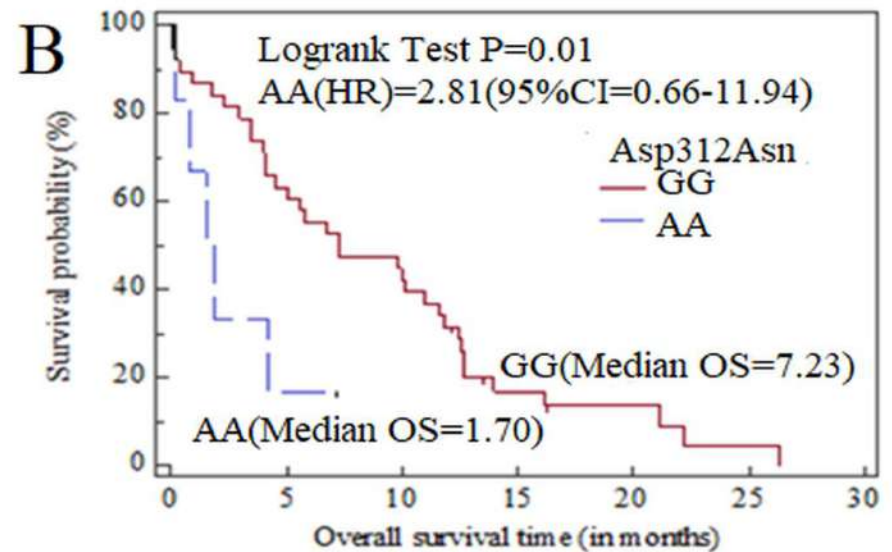
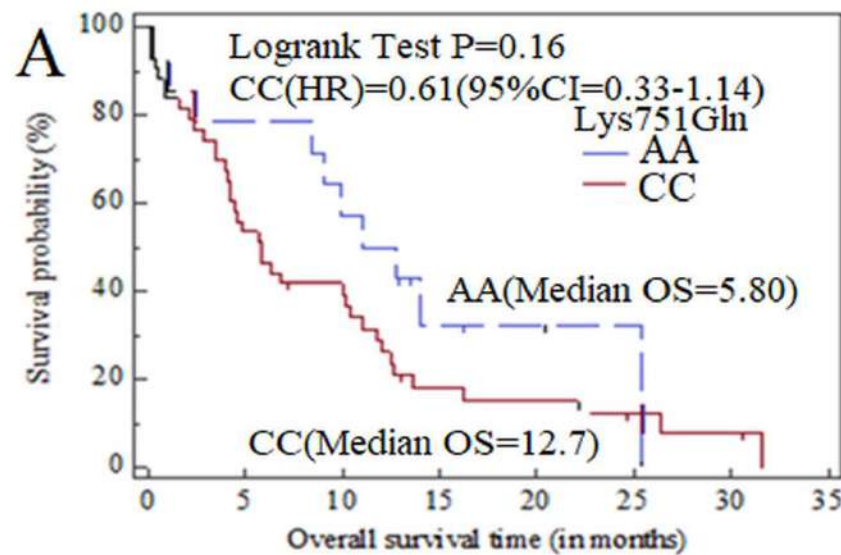


Fig 5.21: Kaplan Meier curves illustrating association between overall survival in patients with **(A)** Mutant genotype (CC) for SCLC showing *XPD Lys⁷⁵¹Gln* polymorphism **(B)** Mutant genotype (AA) for SCLC and **(C)** Heterozygous genotype (GA) depicting *XPD Asp³¹²Asn* polymorphism

5.5.5.10 Association of *XPD* SNPs and Overall survival of lung cancer patients treated with different chemotherapeutic regimen

The survival for patients treated with different chemotherapeutic regimen and their association with *XPD* polymorphism and OS has been studied as shown in Table 5.31. The present study population had been treated with different chemotherapeutic regimen but the analysis was only applied on those regimens with which maximum number of patients was treated in the study. Therefore, the OS of patients treated with three regimens *i.e.* docetaxel cisplatin/carboplatin, irinotecan cisplatin/carboplatin and pemetrexed cisplatin/carboplatin and their association with *XPD* polymorphism has been illustrated in Table 5.31.

As evaluated in table 5.31, for *Lys*⁷⁵¹*Gln* polymorphism, lung cancer patients carrying mutant genotype (CC) and treated with docetaxel cisplatin/carboplatin regimen depicted a poor survival of 5.76 months when compared with the subjects having wild type genotype (AA) having a median survival of 8.63 months. No significant genotypic distribution was observed for *XPD* polymorphism and OS in patients treated with docetaxel cis/carboplatin regimen. For *Arg*¹⁵⁶*Arg* polymorphism, patients with mutant genotype (CC) and treated with docetaxel cisplatin/carboplatin regimen showed a better survival in patients when compared with the reference genotype (AA) (12.3 vs 8.20 months; log rank $P=0.86$). The hazard ratio observed for patients carrying mutant genotype was less but the result was not significant ($HR^1= 1.30$, 95% CI: 0.51-3.27, $p=0.57$). Similarly, the patients with heterozygous genotype for *Arg*¹⁵⁶*Arg* polymorphism showed a high death ratio with poor survival in patients treated with this regimen ($HR^1= 1.81$, 95% CI: 0.94-3.49, $p=0.07$; MST=6.86 months).

Furthermore as shown in table 5.31, the patients harboring mutant genotype (CC) and treated with irinotecan cisplatin/carboplatin regimen depicted a higher survival and better prognosis for *Lys*⁷⁵¹*Gln* polymorphism when compared with the wild genotype (AA) (25.43 vs 5.63 months; log rank $P=0.01$). The Cox regression model indicated a significant trend suggesting a lower mortality rate amongst subjects carrying mutant genotype and have a good survival rate, thus, making it a predictive factor for disease treatment ($HR^1: 0.21$; 95%CI: 0.08-0.57; $p=0.002$; Fig 5.22A). A significant low hazard ratio was observed for the combined genotype (AC+CC) ($HR^1: 0.48$; 95%CI: 0.21-0.92; $p=0.02$, MST=11 months).

For *Asp³¹²Asn* polymorphism, individuals carrying mutant genotype (AA) when treated with irinotecan cis/carboplatin regimen showed a poor survival period when compared with the patients having wild type genotype (1.86 vs 7.28 months; log rank $p=0.01$). A significantly high hazard ratio was also observed for subjects with mutant genotype and treated with irinotecan cisplatin/carboplatin regimen (HR^1 : 16.46; OR: 2.50-108.40; $p=0.03$; Fig 5.22B) when compared with wild genotype, whereas subjects with heterozygous genotype (GA) treated with same regimen showed a better survival with marginally significant protective effect (HR: 0.51; OR: 0.31-1.04; $p=0.05$, MST=10.4 months). *Arg¹⁵⁶Arg* polymorphism did not contributed to the association with the survival of patients treated with irinotecan cisplatin/carboplatin regimen as shown in table 5.31.

The patients treated pemetrexed cisplatin/carboplatin regimen and having *Lys⁷⁵¹Gln* polymorphism depicted a survival time of almost 12 months in heterozygous genotype (AC) in comparison with wild type (AA) which showed a survival time of 9.13 months (log rank $p=0.61$). No significant distribution was observed between *Lys⁷⁵¹Gln* polymorphism and OS in patients treated with pemetrexed cis/carboplatin regimen. In *Asp³¹²Asn* polymorphism, patients carrying heterozygous genotype (GA) for *Asp³¹²Asn* polymorphism when treated with pemetrexed cisplatin/carboplatin regimen had MST of 8.8 months which was lower than the reference genotype (GG) with MST of 12.46 months (log rank $p=0.007$) (Fig 5.22C). On applying Cox regression model, a 2-fold increased death ratio was reported for heterozygous type genotype when compared with the wild type genotype (HR: 2.01, 95%CI: 1.14-3.56; $p=0.01$). When heterozygous and mutant genotype were combined as a single genotype (GA+AA) a high risk was observed in patients treated with the pemetrexed cis/carboplatin regimen showing a mediocre survival (HR: 1.89; 95%CI: 1.08-3.30; $p=0.02$; MST=7.56). The patients carrying mutant genotype for *Arg¹⁵⁶Arg* polymorphism showed a slight better survival than that carrying reference genotype (10.33 vs 9.71 months). Overall there was no significant association that can be established for *XPD Arg¹⁵⁶Arg* and overall survival in lung cancer patients treated with this chemotherapeutic regimen.

TABLE 5.31: Association of XPD polymorphism and overall survival according to regimen

REGIMEN – docetaxel cis/ carboplatin								
Genotype	N=76	Alive N=14	Dead N=62	Median OS(months)	Crude HR ^a (95%CI)	Adjusted HR ¹ (95%CI) ^b	Log rank P	p
<i>Lys⁷⁵¹Gln</i>								
AA	41 (53.94)	6	35	8.63	1.00 (Reference)	1.00 (Reference)		
AC	25 (32.89)	6	19	10.13	0.83 (0.48-1.44)	0.81 (0.44-1.49)	0.52	0.50
CC	10 (13.15)	2	8	5.76	1.18 (0.52-2.66)	1.04 (0.45-2.41)	0.66	0.92
AC+CC	35 (46.05)	8	27	7.56	0.92 (0.56-1.52)	0.88 (0.51-1.50)	0.75	0.65
<i>Asp³¹²Asn</i>								
GG	31 (40.78)	6	25	9.86	1.00 (Reference)	1.00 (Reference)		
GA	35 (46.05)	5	30	7.56	1.34 (0.79-2.28)	1.73 (0.95-3.15)	0.27	0.07
AA	10 (13.15)	3	7	8.21	0.91 (0.40-2.08)	0.83 (0.50-1.39)	0.84	0.50
GA+AA	45 (59.21)	8	37	7.56	1.23 (0.74-2.03)	1.47 (0.83-2.61)	0.41	0.18
<i>Arg¹⁵⁶Arg</i>								
CC	44 (57.89)	10	34	8.20	1.00 (Reference)	1.00 (Reference)		
CA	25 (32.89)	3	22	6.86	1.37 (0.78-2.41)	1.81 (0.94-3.49)	0.23	0.07
AA	7 (9.21)	1	6	12.3	1.08 (0.44-2.63)	1.30 (0.51-3.27)	0.86	0.57
CA+AA	32 (42.10)	4	28	7.75	1.30 (0.78-2.17)	1.58 (0.89-2.80)	0.29	0.11
REGIMEN – Irinotecan cis/carboplatin								
Genotype	N=70	Alive N=17	Dead N=53	Median OS(months)	Crude HR ^a (95%CI)	Adjusted HR ¹ (95%CI) ^b	Logrank P	p
<i>Lys⁷⁵¹Gln</i>								
AA	31 (44.28)	5	26	5.63	1.00 (Reference)	1.00 (Reference)		
AC	26 (37.14)	7	19	8.63	0.64 (0.35-1.15)	0.61 (0.30-1.21)	0.13	0.16
CC	13 (18.57)	5	8	25.43	0.39 (0.20-0.77)	0.21 (0.08-0.57)	0.01	0.002
AC+CC	39 (55.71)	12	27	11	0.54 (0.30-0.95)	0.48 (0.21-0.92)	0.48	0.02
<i>Asp³¹²Asn</i>								
GG	26 (37.14)	4	22	7.28	1.00 (Reference)	1.00 (Reference)		
GA	39 (55.71)	12	27	10.4	0.80 (0.45-1.43)	0.51 (0.31-1.04)	0.45	0.05
AA	5 (7.14)	1	4	1.86	3.29 (0.58-18.54)	16.46 (2.50-108.40)	0.01	0.003
GA+AA	44 (62.85)	13	31	9	0.89 (0.51-1.55)	0.69 (0.38-1.23)	0.69	0.69
<i>Arg¹⁵⁶Arg</i>								
CC	29 (41.42)	7	22	6.73	1.00 (Reference)	1.00 (Reference)		
CA	34 (48.57)	10	24	9	1.02 (0.57-1.83)	0.72 (0.38-1.37)	0.92	0.32
AA	7 (10)	-	7	7.23	1.60 (0.60-4.28)	1.44 (0.52-3.95)	0.26	0.47

CA+AA	41 (58.57)	10	31	9	1.12 (0.65-1.92)	0.86 (0.46-1.59)	0.67	0.64
REGIMEN-Pemetrexed cis/carboplatin								
Genotype	N=69	Alive N=14	Dead N=55	Median OS(months)	Crude HR^a (95% CI)	Adjusted HR¹ (95% CI)^b	Log rank P	p³
<i>Lys⁷⁵¹Gln</i>								
AA	39 (60)	8	31	9.13	1.00 (Reference)	1.00 (Reference)		
AC	21 (32.3)	4	17	12.46	0.86 (0.48-1.53)	0.88 (0.45-1.70)	0.61	0.71
CC	9 (13.84)	2	7	10.33	0.82 (0.37-1.77)	0.67 (0.26-1.70)	0.63	0.40
AC+CC	30 (46.15)	6	24	12.13	0.85 (0.50-1.44)	0.88 (0.45- 1.70)	0.85	0.71
<i>Asp³¹²Asn</i>								
GG	33 (50.76)	11	22	12.46	1.00 (Reference)	1.00 (Reference)		
GA	31 (47.69)	2	29	8.8	2.06 (1.17-3.6)	2.01 (1.14-3.56)	0.007	0.01
AA	5 (7.69)	1	4	7.56	1.70 (0.46-6.25)	1.52 (0.43-5.33)	0.31	0.51
GA+AA	36 (55.38)	3	33	7.56	2.02 (1.18-3.45)	1.89 (1.08-3.30)	0.008	0.02
<i>Arg¹⁵⁶Arg</i>								
CC	32 (49.23)	4	28	9.71	1.00 (Reference)	1.00 (Reference)		
CA	32 (49.23)	10	22	9.01	0.75 (0.43-1.31)	0.96 (0.52-1.76)	0.32	0.90
AA	5 (7.69)	-	5	10.33	1.17 (0.42-3.21)	1.32 (0.78-2.24)	0.73	0.29
CA+AA	37 (56.92)	10	27	9.46	0.79 (0.46-1.35)	1.00 (0.56-1.77)	0.39	0.99

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months and (b)-Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall Survival

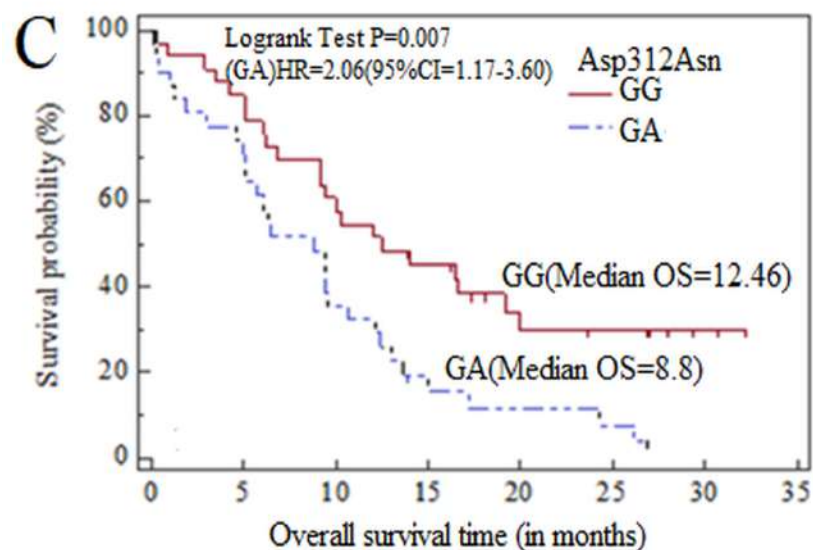
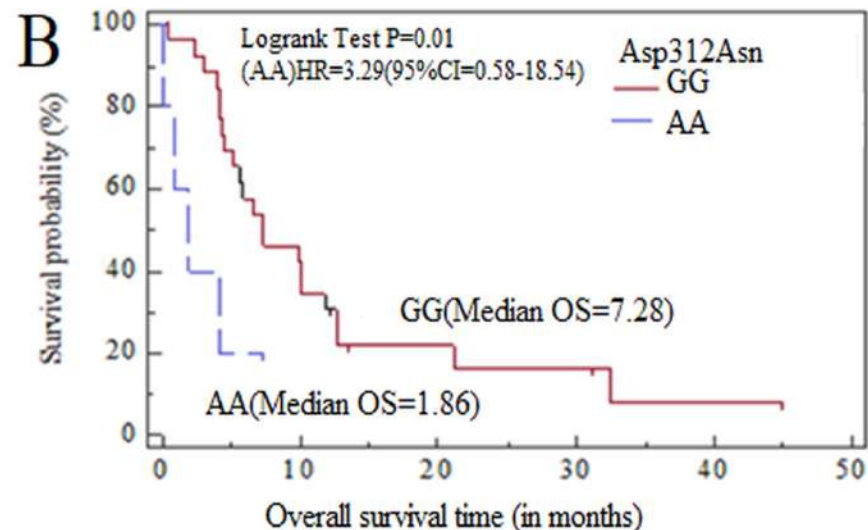
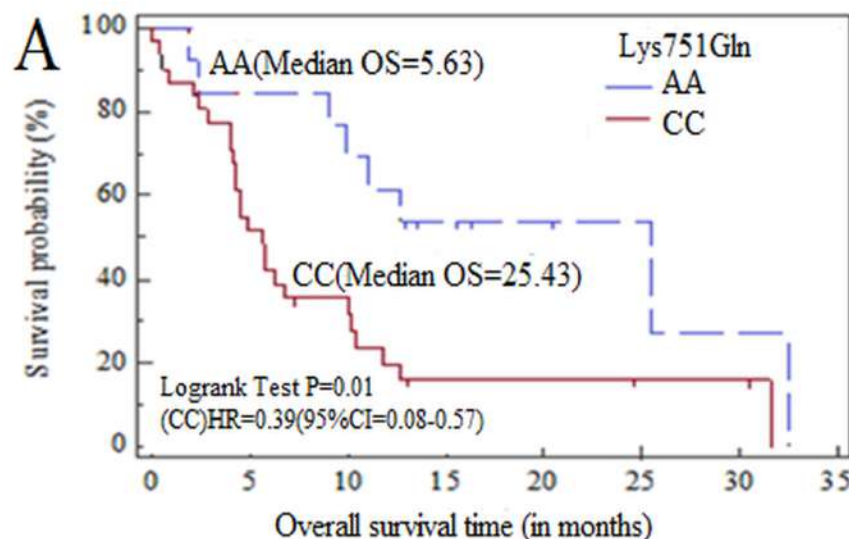


Fig 5.22: Kaplan Meier curves illustrating association between overall survival in lung cancer patients with (A) mutant genotype (CC) treated with irinotecan cisplatin/carboplatin regimen showing *XPD Lys⁷⁵¹Gln* polymorphism (B) mutant genotype (AA) treated with irinotecan cisplatin/ carboplatin regimen depicting *XPD Asp³¹²Gln* polymorphism (C) Heterozygous genotype (GA) treated with pemetrexed cisplatin/carboplatin regimen showing *XPD Asp³¹²Asn* polymorphism

5.5.5.11 Association of *XPB* polymorphism and clinical outcomes

The objective response of the patients towards chemotherapeutic treatment was assessed in relation to the genotype of the *XPB* polymorphic sites which has been summarized in table 5.32. On the basis of objective response using the RECIST criteria, the patients were stratified as responders which included patients showing complete response (CR) and partial response (PR) and non-responders comprising patients showing progressive disease (PD) and stable disease (SD). We assessed responders and non responders in each genotypic category for each polymorphic site. The association of genotype with the objective response was estimated using multivariate logistic regression approach where an adjusted odds ratio was obtained after adjusting for various confounders such as age, gender, smoking, ECOG, KPS, stage, histology. In case of *Lys⁷⁵¹Gln*, the percentage of subjects having combined variant (AC+CC) genotype was higher in case of responders (50%) than non-responders (40%). When further genotypic distribution was observed, responders with mutant genotype (CC) were 16.98% in population when compared with non-responders (10.41%). Therefore, on applying logistic regression analysis mutant genotype (CC) showed a good prognosis or response to chemotherapy when compared with wild genotype (AOR=0.57, 95% CI: 0.22-1.50, $p=0.26$); although result was not significant as shown. The frequency of variant genotype (AA) carrying subjects in case of *Asp³¹²Asn* was more expressed in responders than non responders (10.37% vs 5.20%). Therefore, there was no significant association observed between *Asp³¹²Asn* and clinical response. In case of *Arg¹⁵⁶Arg* polymorphism, the responders were 6.60% and non responder were 8.33%. On the other hand as shown in table 5.33, in case of *Arg¹⁵⁶Arg* polymorphism, there was a difference in the percentage of wild genotype (CC) in responders (53.77%) in comparison with non-responders (45.83%). For heterozygous genotype (CA) representation of responders was less in comparison with non-responders (39.62% vs 45.83%). *Arg¹⁵⁶Arg* polymorphism did not show any significant association observed between the genotype and objective response.

Table 5.32: Genotypic Response of XPD polymorphism to platinum based chemotherapy and their association with lung cancer risk				
Genotype	Responders (CR+PR)(N=106)	Non- Responders (SD+PD)(N=96)	AOR (95% CI)^a	p^b
<i>Lys⁷⁵¹Gln</i>				
AA	53 (50%)	57 (59.37%)	(Reference)	
AC	35 (33.01%)	29 (30.20%)	0.76 (0.39-1.48)	0.42
CC	18 (16.98%)	10 (10.41%)	0.57 (0.22-1.50)	0.26
AC+CC	53 (50%)	39 (40.62%)	0.76 (0.39-1.48)	0.42
<i>Asp³¹²Asn</i>				
GG	51 (48.11%)	40 (41.66%)	(Reference)	
GA	44 (41.50%)	51 (53.12%)	1.64 (0.89-3.03)	0.11
AA	11 (10.37%)	5 (5.20%)	0.79 (0.43-1.47)	0.47
GA+AA	55 (51.88%)	56 (58.33%)	1.43 (0.79-2.57)	0.22
<i>Arg¹⁵⁶Arg</i>				
CC	57 (53.77%)	44 (45.83%)	(Reference)	
CA	42 (39.62%)	44 (45.83%)	1.23 (0.67-2.26)	0.49
AA	7 (6.60%)	8 (8.33%)	1.15 (0.65-2.04)	0.61
CA+AA	49 (46.22%)	52 (54.16%)	1.26 (0.71-2.26)	0.41

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status, histology, stage, KPS and ECOG. (b)- Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. CR-Complete response; PR-Partial response; SD-Stable disease; PD-Progressive disease

5.5.6 XPG

5.5.6.1 XPG variants and its association with lung cancer susceptibility

The distributions of the XPG codon *Phe⁶⁷⁰Leu* and *Asp¹¹⁰⁴His* among cases and controls have been illustrated in Table 5.33 (a). The distribution of the genotypes in controls for both the codons were in agreement with HWE i.e. *Phe⁶⁷⁰Leu* (N=370, $\chi^2=0.05$, df=1, p=0.81) and *Asp¹¹⁰⁴His* (N=370, $\chi^2=1.74$, df=1, p=0.18). Our data revealed that in XPG *Phe⁶⁷⁰Leu* polymorphism heterozygous genotype (CA) was more frequent in controls as compared to cases (36.75% vs 18.10%) whereas the wild genotype (CC) was less represented in controls as compared to cases (55.40% vs 80%). Therefore in case of *Phe⁶⁷⁰Leu* a significant distribution of the genotypes was observed amongst cases and controls ($\chi^2=53.42$, df=2, **p<0.0001**). The MAF for *Phe⁶⁷⁰Leu* allele (A) was higher in controls than cases (26.21% vs 10.94%).

The association between the two polymorphic variants of XPG gene and the risk of lung cancer was further examined after stratifying for potential confounding variables, such as age, sex and

pack-years. The contribution of the two *XPG* polymorphic variants and its association with lung cancer susceptibility was further depicted in Table 5.33(a). Adjusted OR's and 95%CI were calculated using logistic regression analysis. For both polymorphisms the homozygous wild genotype was taken as reference group and the variant genotypes were compared with it. In case of *XPG Phe⁶⁷⁰Leu* polymorphism it was observed that both heterozygous (CA) (OR=0.35; 95%CI: 0.25-0.50; **p<0.0001**) and mutant (AA) (OR=0.18; 95%CI: 0.07-0.13; **p=0.0001**) genotypes were conferring a protective effect towards lung cancer which was found to be significant when compared with the wild type genotype (CC). Similarly when stratified according to histological subtypes (table 5.33b), it was observed that wild genotype (CC) was more expressed in patients diagnosed with SQCC in comparison with heterozygous genotype (82.96% vs 15.55%). Out of 120 ADCC patients, 102 (85%) were having wild genotype, 2 (1.66%) were mutant and 16 (13.33%) were having heterozygous genotype respectively. It was observed that the mutant genotype (AA) showed a significant protective effect in both ADCC (OR=0.14; 95%CI: 0.03-0.61; **p=0.008**) and SQCC (OR=0.14; 95%CI: 0.03-0.65; **p=0.008**) subtypes. Similarly the lung cancer subjects who were heterozygous carriers (CA) also showed a highly protective effect for SQCC subjects (OR: 0.27; 95% CI: 0.16-0.46; **p<0.0001**), ADCC (OR: 0.26; 95%CI: 0.14-0.46; **p<0.0001**) (table 5.33b). The lung cancer patients diagnosed with SCLC subtype showed protective effect in heterozygous carriers (OR: 0.58; 95%CI: 0.35-0.96; **p=0.03**). Similarly, when both the heterozygous and mutant genotypes were combined as a single genotype (CA+AA), a protective effect was observed when compared with the reference genotype (OR: 0.54; 95% CI: 0.33-0.87; **p=0.01**) as shown in table 5.33(b).

For *Asp¹¹⁰⁴His* polymorphism, the mutant genotype (CC) was at higher frequency in controls as compared to the cases (3.57vs 2.97%) whereas heterozygous genotype (GC) was less prevalent in controls as compared to cases (25.13 vs 28.64%). The MAF for controls and cases for *Asp¹¹⁰⁴His* polymorphism was 16.08% and 17.29% respectively as shown in table 5.33a. No significant association was observed for lung cancer risk and *Asp¹¹⁰⁴His* polymorphism. The relationship between *Asp¹¹⁰⁴His* polymorphism and histological subtypes was also explored in table 5.33(b). The wild type genotype (GG) was over represented in comparison with other genotypes for all the histological subtypes *i.e.* 67.40% (SQCC), 72.50%(ADCC) and 64.86%(SCLC) respectively. The mutant genotype was less illustrated in SQCC (1.48%), ADCC (2.50%) and SCLC (5.40%) respectively. The double variant allele (AA) showed a protective

effect in patients with SQCC (OR: 0.45; 95%CI: 0.09-2.12; p=0.31); however the results were not significant. Similarly, there was no significant relationship observed between *XPG* polymorphic variants and risk for lung cancer subtypes.

Table 5.33 (a): Distribution of genotypes of *XPG* polymorphism depicting their association with lung cancer risk

OVERALL LUNG CANCER				
<i>Phe</i> ⁶⁷⁰ <i>Leu</i>	Controls n (%) N =370	Cases n (%) N=370	AOR (95% CI) ^a	<i>p</i> ^b
CC	205 (55.40)	296 (80)	1.00 (Reference)	
CA	136 (36.75)	67 (18.10)	0.35 (0.25-0.50)	<0.0001
AA	29 (7.83)	7 (1.89)	0.18 (0.07-0.43)	0.0001
CA+AA	165 (44.59)	74 (20)	0.32 (0.23-0.45)	<0.0001
C Allele	546 (73.78)	659 (89.05)		
A Allele	194 (26.21)	81 (10.94)		
MAF	26.21	10.94		
OVERALL LUNG CANCER				
<i>Asp</i> ¹¹⁰⁴ <i>His</i>	Controls n (%) N =370	Cases n(%) N=370	AOR (95% CI) ^a	<i>p</i> ^b
GG	264 (71.35)	253 (68.37)	1.00 (Reference)	
GC	93 (25.13)	106 (28.64)	1.19 (0.85-1.67)	0.29
CC	13 (3.51)	11 (2.97)	0.89 (0.38-2.10)	0.80
GC+CC	106 (28.64)	117 (31.62)	1.16 (0.84-1.60)	0.36
G Allele	621 (83.91)	612 (82.70)		
C Allele	119 (16.08)	128 (17.29)		
MAF	16.08	17.29		

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 χ^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 5.33 (b): Distribution of genotypes of <i>XPG</i> polymorphism depicting their association with lung cancer subtypes									
	SQCC			ADCC			SCLC		
<i>Phe</i> ⁶⁷⁰ <i>Leu</i>	Cases n(%) N=135	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N=120	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N=111	AOR (95% CI) ^a	<i>p</i> ^b
CC	112 (82.96)	1.00 (Reference)		102 (85)	1.00 (Reference)		78 (70.27)	1.00 (Reference)	
CA	21 (15.55)	0.27 (0.16-0.46)	<0.0001	16 (13.33)	0.26 (0.14-0.46)	<0.0001	30 (27.02)	0.58 (0.35-0.96)	0.03
AA	2 (1.48)	0.14 (0.03-0.65)	0.01	2 (1.66)	0.14 (0.03-0.61)	0.008	3 (2.70)	0.33 (0.09-1.17)	0.08
CA+AA	23 (17.03)	0.25 (0.15-0.42)	<0.0001	18 (15)	0.23 (0.13-0.41)	<0.0001	33 (29.72)	0.54 (0.33-0.87)	0.01
	SQCC			ADCC			SCLC		
<i>Asp</i> ¹¹⁰⁴ <i>His</i>	Cases n(%) N=135	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N=120	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N=111	AOR (95% CI) ^a	<i>p</i> ^b
GG	91 (67.40)	1.00 (Reference)		87 (72.50)	1.00 (Reference)		72 (64.86)	1.00 (Reference)	
GC	42 (31.11)	1.30 (0.82-2.06)	0.24	30 (25)	1.00 (0.61-1.64)	0.97	33 (29.72)	1.27 (0.76-2.10)	0.34
CC	2 (1.48)	0.45(0.09-2.12)	0.31	3 (2.50)	0.75 (0.20-2.76)	0.67	6 (5.40)	1.27 (0.39-4.08)	0.68
GC+CC	44 (32.59)	1.16 (0.74-1.82)	0.49	33 (27.50)	0.97 (0.60-1.56)	0.92	39 (35.13)	1.28 (0.79-2.06)	0.30

(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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. Abbreviations: ADCC=Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma. All the bold number indicates the significant values.

5.5.6.2 Interactive effect of smoking on the *XPG* polymorphism

Gene-environment interaction was studied further by dividing the study subjects into smokers and non-smokers as illustrated in Table 5.34. The smoking group was then categorized in terms of pack years as light smokers (≤ 25 pack years) and heavy smokers (> 25 years). In case of *Phe*⁶⁷⁰*Leu* polymorphism, smokers carrying a single copy of variant allele *i.e.* heterozygous genotype (CA) were more prevalent in controls than cases (39.11% vs 18.15%) whereas wild genotype (CC) was more presented in cases than controls (79.86% vs 53.87%). On applying logistic regression analysis, a protective effect was observed amongst smokers for all the three variant genotypes *i.e.* heterozygous (CA) (OR: 0.32; 95%CI: 0.22-0.48; **p<0.0001**), mutant (AA) (OR: 0.22; 95%CI: 0.08-0.58; **p=0.02**) and combined genotype (CA+AA) (OR: 0.31; 95%CI: 0.21-0.45; **p<0.0001**), however, the effect was more pronounced in mutant subjects. When smokers were segregated on the basis of pack years: it was observed that heavy smokers carrying both variant alleles (AA) (OR: 0.35; 95%CI: 0.18-0.66; **p= 0.001**) as well as heterozygous genotype (CA) (OR: 0.08; 95%CI: 0.01-0.73; **p=0.02**) exhibited a protective effect for lung cancer when compared with the reference group (CC), however this effect was more pronounced in case of mutant genotype.

For light smokers with *Phe*⁶⁷⁰*Leu* polymorphism, the frequency of subjects carrying mutant genotype (AA) was more prevalent in controls than cases (6.47% vs 3.08%) whereas the wild genotype was more expressed in cases than controls (76.54% vs 51.17%). The subjects carrying heterozygous genotype were more expressed among controls than cases (42.35% vs 20.37%). On applying regression analysis, a protective effect was observed for heterozygous genotype (CA) (OR: 0.32; 95%CI: 0.19-0.55; **p<0.0001**) when compared with the reference genotype (CC). Similarly for *Phe*⁶⁷⁰*Leu* polymorphism, the subjects harboring mutant genotype and who were nonsmokers also expressed a significant protective effect for lung cancer (OR: 0.10; 95%CI: 0.01-0.83; **p=0.03**) when compared with the reference genotype. In addition combined genotype (CA+AA) also showed a significant effect towards the disease (OR: 0.33; 95%CI: 0.20-0.54; **p<0.0001**). Non smokers with combined genotype (CA+AA) showed a higher representation in cases (59.70%) than controls (13.13%) suggesting a protective effect when compared with reference (OR: 0.37; 95%CI: 0.76-0.77; **p=0.008**).

In case of *XPG Asp¹¹⁰⁴His* polymorphism, smokers carrying wild genotype (GG) and having *Asp¹¹⁰⁴His* polymorphism were similarly represented in both controls and cases (70.47% vs 68.97%) and the same representation was also observed for the mutant genotype (CC) (3.69% vs 2.64%). Although there was no significant association observed between *Asp¹¹⁰⁴His* polymorphism and lung cancer in smokers as shown in table 5.35. Heavy smokers carrying mutant genotype (CC) showed similar frequency distribution in both controls and cases (1.98% vs 1.41%). The heterozygous genotype was more prevalent among cases than controls (24.75% vs 29.08%) whereas the wild genotype was more expressed in controls than cases (73.26% vs 69.50%). The subjects carrying mutant genotype (CC) showed a protective effect among heavy smokers when compared with those subjects having wild genotype (GG) (OR: 0.75; 95%CI: 0.10-5.61; p=0.78). The presentation of wild genotype was similar in controls and cases among light smokers (68.82% vs 68.51%). The mutant genotype (CC) was represented slightly higher in controls than cases (4.70% vs 3.70%). Light smokers showed a protective effect in subjects carrying mutant genotype (OR: 0.51; 95%CI: 0.14-1.83; p=0.30), however the results were not significant as shown in table 5.35. Among non smokers heterozygous genotype (GC) was more prominent in cases (29.85%) than controls (23.23%) depicting risk for lung cancer (OR: 1.44; 95%CI: 0.70-2.99; p=0.31); although results observed were not significant.

Table 5.34: Distribution of genotypes of <i>XPG</i> polymorphism according to their smoking status and its association with lung Cancer risk								
	SMOKERS				NON SMOKERS			
<i>Phe</i> ⁶⁷⁰ <i>Leu</i>	Controls n (%) N =271	Cases n (%) N=303	AOR (95% CI) ^a	<i>p</i> ^b	Controls n (%) N =99	Cases n (%) N=67	AOR (95% CI) ^a	<i>p</i> ^b
CC	146 (53.87)	242 (79.86)	1.00 (Reference)		54 (54.54)	57 (85.07)	1.00 (Reference)	
CA	106 (39.11)	55 (18.15)	0.32 (0.22-0.48)	<0.0001	12 (12.12)	30 (44.77)	0.47 (0.21-1.03)	0.06
AA	19 (7.01)	6 (1.98)	0.22 (0.08-0.58)	0.002	1 (1.01)	10 (14.92)	0.10 (0.01-0.83)	0.03
CA+AA	125 (46.12)	61 (20.13)	0.31 (0.21-0.45)	<0.0001	13 (13.13)	40 (59.70)	0.37 (0.17-0.77)	0.008
<i>Asp</i> ¹¹⁰⁴ <i>His</i>								
GG	191 (70.47)	209 (68.97)	1.00 (Reference)		73 (73.73)	44 (65.67)	1.00 (Reference)	
GC	70 (25.83)	86 (28.38)	1.12 (0.76-1.65)	0.54	23 (23.23)	20 (29.85)	1.44 (0.70-2.99)	0.31
CC	10 (3.69)	8 (2.64)	0.54 (0.18-1.57)	0.26	3 (3.03)	3 (4.47)	1.96 (0.37-10.38)	0.42
GC+CC	80 (29.52)	94 (31.02)	1.05 (0.72-1.52)	0.77	26 (26.26)	23 (34.32)	1.52 (0.76-3.03)	0.23
	HEAVY SMOKER				LIGHT SMOKER			
<i>Phe</i> ⁶⁷⁰ <i>Leu</i>	Controls n (%) N =101	Cases n (%) N=141	AOR (95% CI) ^a	<i>p</i> ^b	Controls n (%) N =170	Cases n (%) N=162	AOR (95% CI) ^a	<i>p</i> ^b
CC	60 (59.40)	118 (83.68)	1.00 (Reference)		87 (51.17)	124 (76.54)	1.00 (Reference)	
CA	33 (32.67)	22 (15.60)	0.35 (0.18-0.66)	0.001	72 (42.35)	33 (20.37)	0.32 (0.19-0.55)	<0.0001
AA	8 (7.92)	1 (0.70)	0.08 (0.01-0.73)	0.02	11 (6.47)	5 (3.08)	0.35 (0.11-1.08)	0.06
CA+AA	41(40.59)	23 (16.31)	0.30 (0.16-0.56)	0.0002	83 (48.82)	38 (23.45)	0.33 (0.20-0.54)	<0.0001
<i>Asp</i> ¹¹⁰⁴ <i>His</i>								
GG	74 (73.26)	98 (69.50)	1.00 (Reference)		117 (68.82)	111 (68.51)	1.00 (Reference)	
GC	25 (24.75)	41 (29.08)	1.26 (0.69-2.30)	0.45	45 (26.47)	45 (27.77)	1.03 (0.62-1.71)	0.88
CC	2 (1.98)	2 (1.41)	0.75 (0.10-5.61)	0.78	8 (4.70)	6 (3.70)	0.51 (0.14-1.83)	0.30
GC+CC	27 (26.73)	43 (30.49)	1.22 (0.67-2.19)	0.50	53 (31.17)	51 (31.48)	0.96 (0.59-1.56)	0.88

a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

5.5.6.3 Haplotype analysis

Linkage disequilibrium and haplotype frequencies were analyzed using SHEsis program platform for the two *XPG* SNPs as shown in table 5.35. A significant overall distribution was observed when haplotypic profile of *XPG* variants (*Phe*⁶⁷⁰*Leu* and *Asp*¹¹⁰⁴*His*) were compared (Global $\chi^2=63.31$, $p=1.43\text{e-}013$). The two SNPs were in weak linkage disequilibrium with each other ($D=0.015$, $r^2=0.00$). The G-A haplotype showed a significant protective effect against lung cancer (OR: 0.29; 95%CI: 0.21-0.41; **$p=4.52\text{e-}015$**) as G-A genotype was over presented in more presented in controls (22.6%) than cases (7.9%). Whereas almost 2-fold risk was observed for G-C haplotype (OR: 1.87; 95%CI: 1.49-2.33; **$p=2.93\text{e-}008$**). A-A (3% vs 3.6%) and A-C (14.3% vs 12.5%) haplotype was similarly expressed in both cases and controls. The genotypic distribution was not significant for A-A and A-C haplotypes.

Table 5.35: <i>XPG</i> Haplotypes and their association with lung cancer risk				
<i>XPG Phe670Leu & XPG Asp1104His</i>	Case frequency, n (%)	Control frequency, n(%)	OR (95% CI) ^a	p^b -value
A-A	0.030 (22)	0.036 (27)	0.82 (0.46-1.46)	0.51
A-C	0.143 (106)	0.125(92)	1.17 (0.86-1.57)	0.30
G-A	0.079 (59)	0.226 (167)	0.29 (0.21-0.40)	4.52e-015
G-C	0.748 (553)	0.613 (454)	1.87 (1.49-2.33)	2.93e-008

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding p -values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either haplotypic distribution or frequencies between the cases and controls. Global χ^2 test=63.31, p^b -value = 1.43e-013.

5.5.6.4 Combined effect of *XPG* SNPs on lung cancer risk

To evaluate SNP-SNP interaction, different genotypic combinations of the two *XPG* polymorphic variants were analyzed to find out their association towards risk for lung cancer, as shown in Table 5.36. Our data shows that individuals who were carrying any single variant allele for both polymorphic variants *i.e.* heterozygous carriers (CA+GC) were more presented in controls (20.43%) than cases (9.82%). The patients carrying heterozygous genotype showed a protective effect towards lung cancer development (OR=0.43; 95%CI: 0.24-0.75; **$p=0.003$**). When both the SNP-SNP combinations of heterozygous and mutant genotype considered as a

single genotype a significant decreased risk was observed (OR=0.46; 95%CI: 0.27-0.77; $p=0.003$).

Table 5.36 : Combination of genotypes of <i>XPG</i> polymorphic variants and their relationship with the risk of lung cancer				
Genotypes	Cases, n (%) N = 234	Controls, n(%) N = 186	Adjusted OR (95% CI)^b	<i>p</i>^a
<i>XPG Phe⁶⁷⁰Leu & XPG Asp¹¹⁰⁴His</i>				
0 (CC+GG)	209 (89.31)	147 (79.03)	1.00 (Reference)	
1(CA+GC)	23 (9.82)	38 (20.43)	0.43 (0.24-0.76)	0.003
2(AA+CC)	2 (0.8)	1 (0.5)	-	-
3(1+2)	25 (10.68)	39 (20.96)	0.46 (0.27-0.77)	0.003

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status. (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 3: combined hetero and mutant genotype

5.5.6.5 Relationship of Clinico-pathological features with *XPG* polymorphic variants

Clinico-pathological parameters such as stage, primary tumor extension, lymph node involvement and metastasis were taken into account and analyzed to find out the effect of genetic variations in *XPG* gene on these factors as depicted in Table 5.37. The effect of different genotypes of *XPG* genetic variants on the advancement of these clinical parameters and progression of the disease was assessed using logistic regression approach. The odd of advanced state of each parameter was calculated using wild genotype as the reference group. Odds ratio obtained was adjusted for age, gender and smoking status.

On the basis of clinical stage, patients were stratified into stage III and stage IV as most of the recruited cases were diagnosed with advanced lung cancer. In case of *Phe⁶⁷⁰Leu* polymorphism, the prevalence of heterozygous (CA) genotype carrying subjects were higher in stage III (20%) in comparison with stage IV (14.72%) patients. The double variant genotype (AA) with clinical stage (IV) had a higher representation (82.82%) as compared to subjects with stage III (78.28%). On investigating the relationship of the *Phe⁶⁷⁰Leu* polymorphism with the clinical stage, it was evident this polymorphic site did not contribute towards predicting the stage of tumor (Table 5.37). In case of *Asp¹¹⁰⁴His* polymorphism, the percentage of subjects carrying heterozygous genotype (GC) was 24% for stage III and 32.51% for stage IV. Also, the patients carrying

combined genotype (GC+CC) for *Asp¹¹⁰⁴His* polymorphism with stage IV had a slightly higher frequency (36.19%) than subjects in stage III (28.57%). Logistic regression analysis revealed, a trend towards moderate risk (OR: 1.51; 95%CI: 0.94-2.44; p=0.08) as shown in table 5.37.

The genotypic frequencies have also been computed for the each polymorphic site and compared into two groups on the basis of primary tumor extension. The investigative study of *Phe⁶⁷⁰Leu* polymorphism for its correlation with primary tumor extension suggested that the frequency of subjects harboring the single allele of variant genotype (GC) was slight higher among T3+T4 subgroup (18.84%) than T0+T1+T2 (15.94%). Therefore, no significant association was observed between the polymorphic variant and primary tumor extension (OR: 1.23; 95% CI: 0.60-2.53; p=0.56) as shown in table 5.37. In case of *Asp¹¹⁰⁴His* polymorphism, the heterozygous genotype (GC) was 27.53% represented in T0+T1+T2 tumor size and 29.34% in T3+T4 subgroup. The representation of mutant genotype (CC) was more prevalent in T0+T1+T2 subgroup (7.24%) than T3+T4 (2.17%) subgroup showing a borderline significance for *Asp¹¹⁰⁴His* polymorphism (OR: 0.28; 95%CI: 0.08-1.01; p=**0.05**).

The relation between metastatic status and polymorphic variants is detailed in table 5.37. The genotypic frequencies of variant genotype (AA) for *Phe⁶⁷⁰Leu* polymorphism exhibited no significant difference between the patients who showed metastasis (3.16%) and those who showing absence of metastasis (1.06%). The frequency difference of the heterozygous genotype was higher in subjects without metastasis than subjects having metastasis (20.85% vs 15.18%). Hence, this polymorphic site did not show any association with spread of lung cancer to other organs as shown in table 5.37. In case of *Asp¹¹⁰⁴His* polymorphism, lung cancer subjects with metastasis and heterozygous genotype (GC) were found to have similar frequency with subjects without metastasis (28.48% vs 27.27%). The frequency distribution of heterozygous genotype (GC) was found to be similar in both metastatic groups (29.41% vs 35.44 %). Hence, none of the polymorphism site contributed towards the progression of lung cancer among other organs as shown in table 5.37.

Table 5.37: Relationship of Clinico-pathological features with XPG *Phe*⁶⁷⁰*Leu* and XPG *Asp*¹¹⁰⁴*His* genotypes in lung cancer patients

Variable	Clinical stage		AOR (95% CI) ^a	<i>p</i> ^b	TNM stage		AOR (95% CI) ^a	<i>p</i> ^b	Metastasis		AOR (95% CI) ^a	<i>p</i> ^b
	III	IV			T3+T4	T2+T1+T0			Positive	Negative		
<i>Phe</i>⁶⁷⁰<i>Leu</i>	175	163			276	69			158	187		
CC	137 (78.28)	135 (82.82)	Reference		218 (78.98)	57 (82.60)	Reference		129 (81.64)	146 (78.07)	Reference	
CA	35 (20)	24 (14.72)	0.61 (0.33-1.12)	0.11	52 (18.84)	11 (15.94)	1.23 (0.60-2.53)	0.56	24 (15.18)	39 (20.85)	0.61 (0.33-1.11)	0.10
AA	3 (1.71)	4 (2.45)	1.38 (0.29-6.52)	0.68	6 (2.17)	1 (1.44)	1.47(0.17-12.61)	0.72	5 (3.16)	2 (1.06)	3.17(0.58-17.20)	0.18
CA+AA	38 (21.71)	28 (17.17)	0.67 (0.38-1.18)	0.17	58 (21.01)	12 (17.39)	1.25 (0.62-2.51)	0.51	29 (18.35)	41 (21.92)	0.72 (0.41-1.27)	0.26
<i>Asp</i>¹¹⁰⁴<i>His</i>												
GG	125 (71.42)	104 (63.80)	Reference		189 (68.47)	45 (65.21)	Reference		102 (64.55)	132 (70.58)	Reference	
GC	42 (24)	53 (32.51)	1.55 (0.94-2.54)	0.08	81 (29.34)	19 (27.53)	1.01 (0.55-1.85)	0.95	49 (28.48)	51 (27.27)	1.23 (0.75-2.01)	0.40
CC	8 (4.71)	6 (3.68)	1.28 (0.36-4.46)	0.69	6 (2.17)	5 (7.24)	0.28 (0.08-1.01)	0.05	7 (4.43)	4 (2.13)	1.96 (0.53-7.25)	0.31
GC+CC	50 (28.57)	59 (36.19)	1.51 (0.94-2.44)	0.08	87 (31.520)	24 (34.78)	1.09 (0.60-1.99)	0.75	56 (35.44)	55 (29.41)	1.28 (0.79-2.06)	0.30

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status.(b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

5.5.6.6 Association of *XPG* polymorphic variants and Overall survival (OS) of lung cancer patients

Overall survival analysis was carried for the *XPG* polymorphic sites where a total of 311 subjects were recruited. Median survival time (MST) was estimated for each genotypic group and compared with reference group using Kaplan Meier method and log rank test. Adjusted hazard ratio was estimated using multivariate Cox proportional hazard model. The two variants of *XPG* gene under study negated the role of *XPG* polymorphic variations in predicting the prognosis of the patients as shown in table 5.38. The genotypic distribution in the patients agreed with that estimated from HWE for both *Phe*⁶⁷⁰*Leu* ($\chi^2=1.70$; df=1; p=0.18) and *Asp*¹¹⁰⁴*His* ($\chi^2= 0.06$; df=1; p=0.79) polymorphisms. Homozygous wild genotype was considered as reference for both SNPs: *Phe*⁶⁷⁰*Leu* (CC) and *Asp*¹¹⁰⁴*His* (GG). In case of *Phe*⁶⁷⁰*Leu* polymorphism, the survival period in months of subjects carrying mutant genotype was similar for the persisting genotypes such CC (8.03), CA (8.03) and AA (8.85) respectively. For *Asp*¹¹⁰⁴*His* polymorphism, mutant (TT) genotype had a poor survival time (5.83 months vs 8.80 months, log rank $p=0.13$) as compared to wild (CC) genotype which was not significant and hence, there was no association observed for this genotype with the survival of lung cancer patients. Median survival time (MST) of subjects having combined genotype (GC+CC) for *Asp*¹¹⁰⁴*His* polymorphism showed a better median survival time when compared with the wild genotype (8.80 vs 6.23 months; logrank $p=0.13$). However, the hazards ratio obtained from multivariate analysis lacked statistical significance ($HR^1=1.28$, 95% C.I.: 0.96-1.69, $p=0.08$).

5.5.6.7 Association of *XPG* polymorphic variants and Overall survival (OS) in lung cancer subtypes

For two *XPG* variants, we analyzed the relevance in predicting the prognosis of lung cancer patients with different histological subtype, the data for each histological subtype has been demonstrated in different tables. Results for overall survival analysis of ADCC patients is tabulated in table 5.39 (a), SQCC patients in table 5.39 (b) and SCLC patients in table 5.39 (c).

For *Phe*⁶⁷⁰*Leu*, the percentage of subjects carrying mutant genotype (AA) was 1.92%, 1.70% and 2.22% in lung cancer subjects diagnosed with ADCC, SQCC and SCLC subtype respectively. ADCC patients having mutant (AA) genotype showed a poor survival as compared to ADCC patients having wild (CC) genotype (3.55 months vs 9.23 months). Hazards ratio obtained from

univariate analysis was found to be statistically significant as ADCC patients had a high death ratio (HR= 3.89, 95% C.I.: 0.25-58, logrank $p=0.03$; Fig 5.23A) as shown in table 5.39 (a). However, in case of SQCC patients MST of combined genotype (CA+AA) was the highest (7.23 months) in comparison to mutant type (5.86 months) and heterozygous (6.98 months) but, it did not show any significant association with the mortality rate as hazards ratio in case of both univariate and multivariate were not statistically significant as shown in table 5.39 (b). Further, in case of SCLC patients harboring the mutant genotype (AA) showed median survival time of 22.53 months which was higher than that for wild type (CC) (6.40 months; logrank $P=0.46$) patients. Among patients with combined genotype (CA+AA) a survival of 12.20 months was reported suggesting a decreased mortality rate among SCLC patients when compared with wild genotype (CC) ($HR^1= 0.59$, 95% CI: 0.33-1.05, $p= 0.07$) as shown in table 5.39 (c). Although there was no significant association reported among *Phe*⁶⁷⁰*Leu* genetic variants and overall survival among all the histological subtypes of lung cancer.

In case of *Asp*¹¹⁰⁴*His* polymorphism, the proportion of heterozygous patients (GC) in ADCC, SQCC and SCLC subtype was 25%, 29.5% and 29.05% respectively. ADCC patients carrying the mutant (CC) genotype were few in number but showed almost similar survival of 10.85 months in comparison to wild type (9.36 months). In SQCC patients, a poor survival of 5.36 months was reported in heterozygous (GA) as well as combined genotype when compared with the wild genotype (GG) (8.41 months). The SCLC patients showed a better survival when compared with the reference genotype (12 vs 7.33 months). This difference in MST pertains to a decreased mortality rate ($HR^1=0.97$, 95% C.I.: 0.53-1.77, log rank $P=0.93$), but lacked statistical significance. SCLC patients having mutant (CC) genotype had a poor median survival time of 2.96 months in comparison with wild genotype (GG) (7.33 months). Thus, the SCLC patients carrying mutant genotype (AA) showed a significant mortality rate when compared with the reference genotype (GG) ($HR^1= 2.55$, 95% CI: 0.94-6.87, $p=0.05$; MST=2.96 months; Fig 5.23B) for SCLC as shown in table 39 (c).

Table 5.38: Association of XPG polymorphism and overall survival in Lung cancer risk

Genotype	No. n% N=311	Dead N=256	Alive N=55	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P-value
<i>Phe⁶⁷⁰Leu</i>								
CC	249 (80.06)	205	44	8.03	1.00(Reference)		1.00(Reference)	
CA	56 (18.06)	45	11	8.03	0.89 (0.65-1.22)	0.50	0.94 (0.67-1.31)	0.74
AA	6 (1.92)	6	-	8.85	1.19 (0.49-2.88)	0.66	1.14 (0.49-2.64)	0.74
CA+AA	62 (19.93)	51	11	7.65	0.92 (0.68-1.24)	0.59	0.97 (0.70-1.33)	0.87
<i>Asp¹¹⁰⁴His</i>								
GG	218 (70.09)	175	43	8.80	1.00(Reference)		1.00(Reference)	
GC	84 (27)	73	11	6.23	1.22 (0.92-1.63)	0.13	1.24 (0.93-1.66)	0.13
CC	9 (2.89)	8	1	5.83	1.48 (0.63-3.47)	0.26	1.39 (0.68-2.87)	0.36
GC+CC	93 (29.90)	81	12	6.23	1.25 (0.94-1.64)	0.09	1.28 (0.96-1.69)	0.08

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

Table 5.39 (a): Association of XPG polymorphism and overall survival analysis in ADCC patients

Genotype	No. n% N=104	Dead N=86	Alive N=18	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P-value
<i>Phe⁶⁷⁰Leu</i>								
CC	90 (86.53%)	74	16	9.23	1.00(Reference)		1.00(Reference)	
CA	12 (11.5)	10	2	6.43	1.13 (0.56-2.27)	0.70	1.40 (0.70-2.82)	0.34
AA	2 (1.92)	2	-	3.55	3.89 (0.25-58.95)	0.03	3.37 (0.73-15.57)	0.12
CA+AA	14 (13.46)	12	2	6.43	1.28 (0.65-2.51)	0.41	1.56 (0.82-2.98)	0.17
<i>Asp¹¹⁰⁴His</i>								
GG	76 (73.07)	60	16	9.36	1.00(Reference)		1.00(Reference)	
GC	26 (25)	24	2	6.43	1.40 (0.83-2.37)	0.15	1.45 (0.87-2.42)	0.15
CC	2 (1.92)	2	-	10.85	1.21(0.25-5.65)	0.78	1.11 (0.26-4.76)	0.88
GC+CC	28 (26.92)	25	2	7.13	1.39 (0.84-2.30)	0.15	1.44 (0.88-2.38)	0.14

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

Table 5.39 (b): Association of XPG polymorphism and overall survival analysis in SQCC patients								
				UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
Genotype	No. n% N=117	Dead N=95	Alive N=22	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P-value
<i>Phe⁶⁷⁰Leu</i>								
CC	97 (82.9)	81	16	8.40	1.00(Reference)		1.00(Reference)	
CA	18 (15.38)	12	6	6.98	1.06 (0.64-1.75)	0.80	1.29 (0.75-2.21)	0.35
AA	2 (1.70)	2	-	5.86	1.87 (0.28-12.59)	0.37	1.72 (0.37-7.99)	0.49
CA+AA	20 (17.09)	14	6	7.23	1.10 (0.67-1.80)	0.68	1.32 (0.78-2.23)	0.29
<i>Asp¹¹⁰⁴His</i>								
GG	81 (69.23)	65	16	8.41	1.00(Reference)		1.00(Reference)	
GC	34 (29.05)	29	5	5.36	1.43 (0.89-2.30)	0.10	1.37 (0.84-2.22)	0.19
CC	2 (1.70)	1	1	-	-	-	-	-
GC+CC	36 (30.76)	30	6	5.36	1.37 (0.86-2.18)	0.14	1.32 (0.82-2.14)	0.24

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

Table 5.39 (c): Association of XPG polymorphism and overall survival analysis in SCLC patients								
				UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
Genotype	No. n% N=90	Dead N=75	Alive N=15	Median OS (months)	HR (95% CI)	Log P	HR ¹ (95% CI)	P-value
<i>Phe⁶⁷⁰Leu</i>								
CC	62 (68.88)	50	12	6.40	1.00(Reference)		1.00(Reference)	
CA	26 (28.88)	23	3	9.43	0.60 (0.36-1.00)	0.09	0.59 (0.32-1.09)	0.09
AA	2 (2.22)	2	-	22.53	0.59 (0.19-1.79)	0.46	0.60 (0.14-2.56)	0.50
CA+AA	28 (31.11)	25	3	12.20	0.59 (0.36-0.96)	0.06	0.59 (0.33-1.05)	0.07
<i>Asp¹¹⁰⁴His</i>								
GG	60 (66.67)	49	11	7.33	1.00(Reference)		1.00(Reference)	
GC	25 (27.77)	21	4	12	0.81 (0.49-1.33)	0.42	0.97 (0.53-1.77)	0.93
CC	5 (5.55)	5	-	2.96	2.08 (0.58-7.36)	0.10	2.55 (0.94-6.87)	0.05
GC+CC	30 (33.33)	26	4	9	0.92 (0.57-1.48)	0.75	1.16 (0.66-2.02)	0.59

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

5.5.6.8 XPG polymorphic variants and their association with Overall survival in lung cancer patients treated with different chemotherapy regimens

The overall survival of patients treated with different chemotherapy regimens and their association with XPG polymorphic variants has been illustrated in Table 5.40. Homozygous wild genotype was considered as reference. In case of *Phe*⁶⁷⁰*Leu* polymorphism, it was conferred that lung cancer patients carrying mutant genotype (AA) and treated with docetaxel cisplatin/docetaxel carboplatin regimen showed a better survival as compared to patients carrying the wild type genotype (CC) (MST=12.56 vs 8.03 months, Log rank p=0.12). The number of patients with mutant genotype in the present study was less for this polymorphism. As for three studied regimens only 1.31% study population was found to be having mutant genotype. The patients carrying combined genotype (CA+AA) and treated with docetaxel cisplatin/carboplatin regimen showed a survival of 12.56 months in comparison with the wild type genotype (8.03 months; logrank P=0.12). Similarly the patients having mutant genotype (AA) for the *Phe*⁶⁷⁰*Leu* polymorphic site and treated with irinotecan cisplatin/irinotecan carboplatin regimen showed a better OS and a low death ratio when compared with those patients harboring the wild type genotype (MST=32.5 vs 9 months; HR¹=0.04; 95%CI: 0.001-1.51; p=0.08; Fig 5.23C). For *Phe*⁶⁷⁰*Leu* polymorphism, patients having mutant (AA) and heterozygous (CA) genotype depicted survival of 6.10 months, showing an increased mortality rate in mutant (HR¹= 1.35; 95%CI: 0.15-11.73; p=0.78) and combined (HR¹=1.65; 95%CI: 0.64-4.21; p=0.29) genotype. There was no significant association was observed for pemetrexed cisplatin/carboplatin regimen and OS in lung cancer patients for *Phe*⁶⁷⁰*Leu* polymorphism.

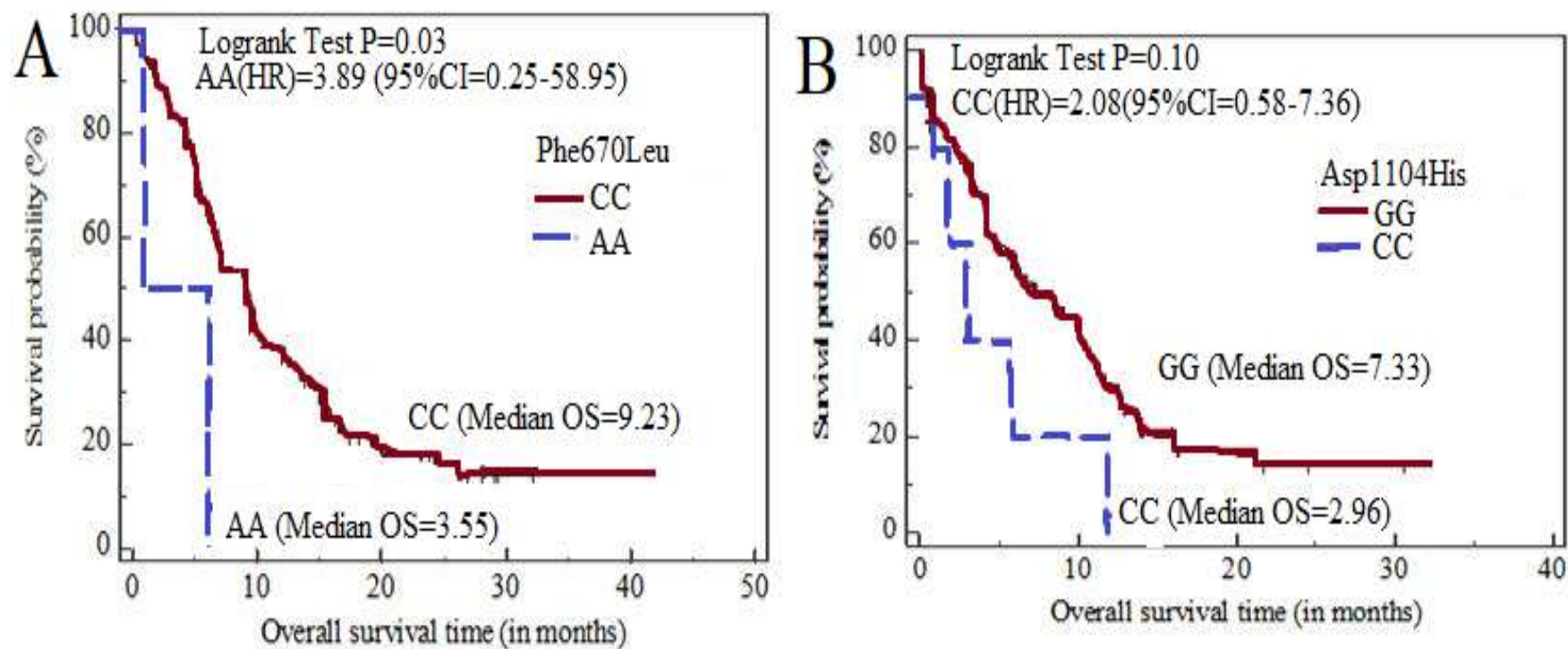
For *Asp*¹¹⁰⁴*His* polymorphism, the frequency distribution of heterozygous subjects (GC) treated with three regimens (docetaxel cis/carboplatin, irinotecan cis/carboplatin, pemetrexed cis/carboplatin) was 23.68%, 25.71% and 27.53% respectively. Individuals treated with docetaxel cis/carboplatin regimen and having combined genotype (GC+CC) showed a better survival of 10.13 months when assessed in comparison with wild genotype (GG) (MST=8.30 months). Although there was no significant association observed between this treatment regimen and overall survival in lung cancer patients. The lung cancer subjects who were treated with irinotecan cisplatin/carboplatin regimen and carrying both the copies of variant alleles (CC) showed a poor OS as compared to subjects receiving same chemotherapeutic regimen and were having wild type genotype (GG) (MST=4.40 vs 9.0 months; logrank p=0.13). A high risk for

mortality was observed for subjects carrying mutant genotype and treated with irinotecan cis/carboplatin regimen ($HR^1=3.01$; 95%CI: 0.92-9.81; $p=0.06$; Fig 5.23D). The mutant genotype (CC) showed almost similar survival period as wild genotype (GG) when treated with pemetrexed cis/carboplatin regimen (10.85 vs 10.33 months; logrank $p=0.63$). For heterozygous patients showing *Asp¹¹⁰⁴His* polymorphism and treated with pemetrexed cis/carboplatin regimen had a poor survival period of 6.43 months, possessing a high death ratio ($HR^1=1.79$; 95%CI: 0.95-3.35; $p=0.07$). Patients having combined genotype (GC+CC) and treated with pemetrexed cis/carboplatin regimen also depicted a 2-fold hazard ratio when compared with the wild genotype ($HR^1=1.83$; 95%CI: 0.97-3.44; $p=0.06$) as shown in table 5.40.

TABLE 5.40: Association of <i>XPG</i> polymorphism and overall survival according to regimen								
REGIMEN – Docetacel cisplatin/Docetacel carboplatin								
Genotype	N=76	Alive	Dead	Median OS(months)	Crude HR (95%CI)	Adjusted HR ¹ (95%CI)	Log rank P	P-value
<i>Phe</i> ⁶⁷⁰ <i>Leu</i>		N=14	N=62					
CC	65 (85.5)	10	55	8.03	1.00 (Reference)	1.00 (Reference)		
CA	10 (13.15)	4	6	9.43	0.53 (0.27-1.04)	0.57 (0.23-1.38)	0.13	0.21
AA	1 (1.31)	0	1	12.56	0.71 (0.13-3.83)	1.04 (0.13-8.35)	0.73	0.96
CA+AA	11 (14.47)	4	7	12.56	0.54 (0.29-1.03)	0.60 (0.26-1.38)	0.12	0.23
<i>Asp</i> ¹¹⁰⁴ <i>His</i>								
GG	56 (73.68)	12	44	8.30	1.00 (Reference)	1.00 (Reference)		
GC	18 (23.68)	1	17	6.73	1.20 (0.66-2.16)	1.00 (0.53-1.90)	0.50	0.97
GC+CC	20 (26.31)	2	18	10.13	0.75 (0.44-1.26)	0.73 (0.37-1.43)	0.28	0.37
REGIMEN – Irinotecan cisplatin/ Irinotecan carboplatin								
Genotype	N=70	Alive	Dead	Median OS(months)	Crude HR (95%CI)	Adjusted HR ¹ (95%CI)	Logrank P	p ³
<i>Phe</i> ⁶⁷⁰ <i>Leu</i>		N=17	N=53					
CC	46 (65.71)	12	34	9	1.00 (Reference)	1.00 (Reference)		
CA	23 (32.85)	5	18	6.73	1.05 (0.59-1.87)	1.35 (0.70-2.59)	0.85	0.35
AA	1 (1.42)	0	1	32.5	0.42 (0.11-1.61)	0.04 (0.001-1.51)	0.34	0.08
CA+AA	24 (34.28)	5	19	7.23	0.98 (0.56-1.72)	1.36 (0.71-2.59)	0.95	0.35
<i>Asp</i> ¹¹⁰⁴ <i>His</i>								
GG	48 (68.57)	12	36	9.00	1.00 (Reference)	1.00 (Reference)		
GA	18 (25.71)	5	13	7.23	1.00 (0.53-1.90)	1.43 (0.67-3.05)	0.97	0.35
CC	4 (5.71)	0	4	4.40	2.14 (0.51-8.96)	3.01 (0.92-9.81)	0.13	0.06
GC+CC	22 (31.42)	5	17	7.23	1.15 (0.63-2.08)	1.65 (0.83-3.29)	0.62	0.14
REGIMEN-Pemetrexed cisplatin/Pemtrexed carboplatin								
Genotype	N=69	Alive	Dead	Median OS(months)	Crude HR (95%CI)	Adjusted HR ¹ (95%CI)	Log rank P	p ³
<i>Phe</i> ⁶⁷⁰ <i>Leu</i>		N=14	N=55					
CC	62 (89.85)	13	49	9.51	1.00 (Reference)	1.00 (Reference)		
CA	6 (8.69)	1	5	5.75	1.46 (0.49-4.29)	1.65 (0.60-4.53)	0.41	0.32
AA	1 (1.44)	0	1	6.10	2.68 (0.10-65.79)	1.35 (0.15-11.73)	0.30	0.78
CA+AA	7 (10.14)	1	6	6.10	1.57 (0.56-4.37)	1.65 (0.64-4.21)	0.28	0.29

Asp¹¹⁰⁴His								
GG	48 (69.56)	12	36	10.33	1.00 (Reference)	1.00 (Reference)		
GC	19 (27.53)	2	17	6.43	1.54 (0.81-2.91)	1.79 (0.95-3.35)	0.13	0.07
CC	2 (2.89)	0	2	10.85	1.41 (0.26-7.47)	1.33 (0.29-6.14)	0.63	0.71
GC+CC	21 (30.43)	2	19	9.13	1.53 (0.83-2.81)	1.83 (0.97-3.44)	0.12	0.06

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months and (b)-Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall Survival



(Continued to next page)

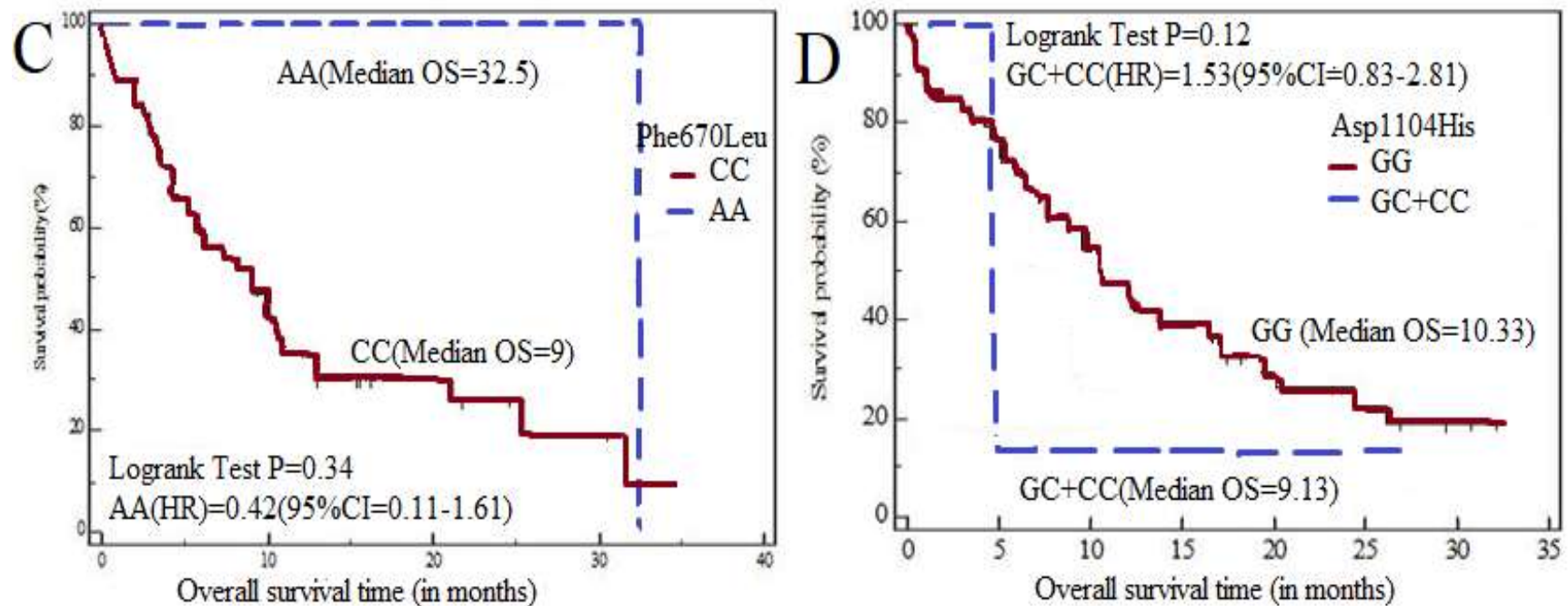


Fig 5.23: Kaplan Meier curve depicting association between (A) *Phe*⁶⁷⁰*Leu* polymorphism and patients having mutant genotype (AA) in ADCC patients (B) *Asp*¹¹⁰⁴*His* polymorphism and patients having mutant genotype (CC) in SCLC patients (C) *Phe*⁶⁷⁰*Leu* polymorphism and patients having mutant genotype (AA) treated with irinotecan-cisplatin/carboplatin regimen (D) *Asp*¹¹⁰⁴*His* polymorphism and patients having combined genotype (GC+CC) treated with pemetrexed-cisplatin/ carboplatin regimen

5.5.6.9 Association of *XPG* polymorphism with clinical response in lung cancer patients

The patients were analyzed for the response to the chemotherapeutic regimen in relation to their genotypes for different *XPG* variants as shown in table 5.41. The patients were divided into responders and non responders on the basis of the RECIST criteria where they are annotated into four groups, CR (complete response), PR (partial response), SD (stable disease) and PD (progressive disease). Responders comprises of CR+PR and non responders consists of SD+PD. Adjusted odds ratio was calculated for each polymorphic site considering wild genotype as the reference. Confounding factors taken into consideration for adjustment were age, gender, smoking, stage, histology, KPS and ECOG. The objective response study was conducted on 202 patients. None of the polymorphic sites of *XPG* i.e. *Phe*⁶⁷⁰*Leu* and *Asp*¹¹⁰⁴*His* gene revealed any significant difference in the frequency of patients showing clinical response. In case of *Phe*⁶⁷⁰*Leu* polymorphism, the representation of patients having CC, CA, AA, CA+AA genotype was almost similar among both responders (78.30%, 19.81%, 1.88% and 21.68%) and non responders (81.25%, 16.66%, 2.08 and 18.75%) respectively. The odds ratio for patients exhibiting mutant genotype was 1.60 (p=0.65). In case of *Asp*¹¹⁰⁴*His* polymorphism, the proportion of subjects having GA+AA genotype was lower (23.58%) among responders as compared to non responders (29.16%) but the difference was not found to be statistically significant. The frequency distribution in *Asp*¹¹⁰⁴*His* polymorphism further showed a higher representation of wild genotype among responders than non responders (76.41% and 68.75% respectively). For GC genotype, 21.69% patients were represented as responder and 26.04% as non responders. The double variant allele (AA) had a very low frequency distribution in both responders as well as non responder (1.88% vs 3.12%). A significant association cannot be observed between *XPG* polymorphism and clinical response in lung cancer patients.

Table 5.41: Genotypic Response of XPG polymorphism to platinum based chemotherapy and their association with lung cancer risk				
Genotype	Responders, n% (CR+PR)(N=106)	Non- Responders, n% (SD+PD)(N=96)	AOR (95% CI)^a	p^b
<i>Phe⁶⁷⁰Leu</i>				
CC	83 (78.30)	78 (81.25)	(Reference)	
CA	21 (19.81)	16 (16.66)	0.79 (0.37-1.69)	0.55
AA	2 (1.88)	2 (2.08)	1.60 (0.19-13.00)	0.65
CA+AA	23 (21.68)	18 (18.75)	0.80 (0.38-1.65)	0.55
<i>Asp¹¹⁰⁴His</i>				
GG	81 (76.41)	66 (68.75)	(Reference)	
GC	23 (21.69)	25 (26.04)	1.20 (0.58-2.67)	0.61
CC	2 (1.88)	3 (3.12)	1.68 (0.24-11.58)	0.59
GC+CC	25 (23.58)	28 (29.16)	0.80 (0.40-1.61)	0.54

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status, histology, stage, KPS and ECOG. (b)- Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. CR-Complete response; PR-Partial response; SD-Stable disease; PD-Progressive disease

5.5.7 XPF

5.5.7.1 XPF polymorphism and their association with lung cancer risk

The genotypic and allelic distribution of *XPF* polymorphic variants *i.e.* *XPF* 673 C>T, *XPF* 11985A>G and *XPF* Arg⁴¹⁵Gln and their association with risk for lung cancer have been represented in Table 5.42. The genotypic distribution in controls for 11985A>G and Arg⁴¹⁵Gln were in agreement with HWE *i.e.* (N=370, $\chi^2=2.35$, df=1, p=0.12) and (N=370, $\chi^2=2.11$, df=1, p=0.14) while controls for *XPF* 673C>T polymorphism was not in agreement with HWE (N=370, $\chi^2=33.77$, df=1, p=0.00). In 673 C>T polymorphism, heterozygous genotype (CT) was more represented in cases as compared to controls (38.10% vs 31.80%) whereas wild genotype was more expressed in controls than cases (53.7% vs 48.9%). MAF for both controls and cases was 0.32. To understand the association between polymorphic variants of *XPF* gene and their role towards lung cancer susceptibility, adjusted OR's and 95%CI were calculated using logistic regression analysis. The analysis was also applied on the histological subtypes of lung cancer. Homozygous wild genotype (CC) was considered as reference group. As shown in table 5.42, our data suggested that lung cancer patients who were carrying heterozygous genotype (CT) for *XPF* 673 C>T polymorphism had a 1.5 fold risk of lung cancer when compared with the wild

genotype (CC) (OR=1.43; 95%CI: 1.02-1.99; **p=0.03**). When segregated on the basis of histological subtypes, the frequency distribution of heterozygous genotype (CT) was 37.5%, 37.07% and 40.54% in ADCC, SQCC and SCLC respectively. Wild type a genotype (CC) in ADCC, SQCC and SCLC was 49.1%, 50.37% and 46.84% respectively. On applying logistic regression analysis, heterozygous genotype (CT) in SQCC patients showed a 1.5-fold increased risk when compared to wild genotype (CC) (OR=1.52; 95%CI: 0.96-2.41; p=0.07). In SCLC patients, a lack of statistical significance was observed for heterozygous genotype (CT) when compared with reference (OR=1.39; 95%CI: 0.85-2.29; p=0.18).

For *11985A>G* polymorphism the wild genotype (AA) was more frequently expressed in controls as compared to cases (61.3% vs 58.10%). Heterozygous genotype (GA) was almost equally distributed in both cases (35.1%) and control (32.16%). Similar frequency distribution was observed for mutant genotype (GG) in control (6.4) and cases (6.75) as depicted in table 5.42. MAF for control and cases was 0.23 and 0.24 respectively. No significant association was observed between *11985A>G* and lung cancer susceptibility as shown in table 5.43. Heterozygous genotype (AG) was less expressed in patients diagnosed with SQCC when compared with wild type genotype (34.07% vs 63.7%). ADCC patients with heterozygous genotype (AG) were also less expressed (36.6%) in comparison with wild genotype (55%). Among SCLC patients, heterozygous genotype was more expressed than the reference genotype (33.3% vs 58.5%). Our data showed that there was no significant relationship was observed for *XPF* polymorphism and lung cancer subtype as shown in table 5.42.

In *Arg⁴¹⁵Gln* polymorphism the wild genotype (GG) was over expressed in cases when compared with controls (53.7% vs 47%). MAF for *Arg⁴¹⁵Gln* polymorphism was 0.30 and 0.25 in cases and controls respectively. No significant association was observed for *Arg⁴¹⁵Gln* polymorphism and lung cancer susceptibility. The wild genotype (GG) was similarly expressed in all the three histological subtypes such as SQCC (51.11%), ADCC (52.5%) and SCLC (57.6%) respectively. After applying logistic linear regression analysis, it was observed that there was no significant association between *Arg⁴¹⁵Gln* polymorphism and lung cancer subtypes as shown in table 5.42.

Table 5.42 : Distribution of genotypes of <i>XPF</i> polymorphism depicting their association with lung cancer risk and its histological subtypes													
	OVERALL LUNG CANCER				SQCC			ADCC			SCLC		
<i>673C>T</i>	Controls n (%) N =370	Cases N (%) N=370	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N=135	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =120	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N=111	AOR (95% CI) ^a	<i>p</i> ^b
CC	192(51.89)	181(48.9)	1.00 (Reference)		68 (50.37)	1.00 (Reference)		59(49.1)	1.0(Reference)		52(46.84)	1.00 (Reference)	
CT	118(31.8)	141(38.1)	1.43(1.02-1.99)	0.03	50(37.07)	1.52(0.96-2.41)	0.07	45(37.5)	1.27(0.79-2.04)	0.31	45(40.54)	1.39(0.85-2.29)	0.18
TT	60(16.2)	48(12.9)	0.92(0.59-1.43)	0.72	17(12.5)	1.01(0.53-1.93)	0.95	16(13.3)	0.89(0.46-1.69)	0.72	14(12.6)	1.00 (0.51-1.98)	0.98
CT+TT	178 (48.1)	189(51.08)	1.25(0.93-1.69)	0.13	67(49.6)	1.34(0.88-2.03)	0.16	61(50.8)	1.13(0.74-1.75)	0.55	59(53.15)	1.28 (0.81-2.01)	0.27
C Allele	516 (69.72)	503(67.97)											
T Allele	238 (32.16)	237(32.07)											
MAF	0.32	0.32											
	OVERALL LUNG CANCER				SQCC			ADCC			SCLC		
<i>11985A>G</i>	Controls n (%) N =370	Cases n(%) N=370	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =135	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =120	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =111	AOR (95% CI) ^a	<i>p</i> ^b
AA	227(61.3)	215(58.10)	1.00 (Reference)		84(62.22)	1.00 (Reference)		66(55)	1.00 (Reference)		65(58.5)	1.00 (Reference)	
GA	119(32.16)	130(35.1)	1.16(0.84-1.59)	0.35	46(34.07)	1.14(0.73-1.79)	0.54	44(36.6)	1.19 (0.75-1.88)	0.44	37(33.3)	1.03(0.63-1.68)	0.88
GG	24(6.4)	25(6.75)	1.08(0.58-2.00)	0.79	5(3.7)	0.58(0.20-1.63)	0.30	10(8.3)	1.59 (0.70-3.64)	0.26	9(8.1)	1.19(0.51-2.79)	0.67
GA+GG	143(38.6)	155(41.8)	1.14(0.84-1.55)	0.36	51(37.7)	1.04(0.68-1.60)	0.84	54(0.45)	1.25 (0.81-1.93)	0.29	46(41.44)	1.06(0.67-1.68)	0.78
A Allele	573(77.42)	560(75.67)											
G Allele	167(22.56)	180(24)											
MAF	0.23	0.24											
	OVERALL LUNG CANCER				SQCC			ADCC			SCLC		
<i>Arg⁴¹⁵Gln</i>	Controls n (%) N =370	Cases n(%) N=370	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =135	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =120	AOR (95% CI) ^a	<i>p</i> ^b	Cases n(%) N =111	AOR (95% CI) ^a	<i>p</i> ^b
GG	174(47.0)	199(53.7)	1.00 (Reference)		69(51.11)	1.00 (Reference)		63(52.5)	1.00 (Reference)		64(57.6)	1.00 (Reference)	
GA	168(45.6)	154(41.6)	0.84(0.62-1.14)	0.27	59(43.7)	0.90(0.61-1.45)	0.80	52(43.3)	0.91(0.59-1.42)	0.70	43(38.7)	0.76(0.48-1.22)	0.26
AA	28(7.5)	17(4.5)	0.59(0.30-1.13)	0.11	7(5.1)	0.62(0.25-1.53)	0.30	5(4.1)	0.56(0.20-1.59)	0.28	4(3.6)	0.46(0.15-1.39)	0.17
GA+AA	196(52.97)	171(46.21)	0.81(0.60-1.09)	0.16	66(48.8)	0.90(0.59-1.36)	0.63	57(47.5)	0.86(0.54-1.36)	0.52	47(42.34)	0.72(0.46-1.14)	0.16
G Allele	516(69.72)	552(74.59)											
A Allele	224(30)	188(25)											
MAF	0.30	0.25											

(a)-AdjustedOdds 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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. Abbreviations: ADCC=Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma. All the bold number indicates the significant value.

5.5.7.2 Effect of smoking on *XPF* polymorphism

Interactive effect of smoking on *XPF* polymorphism (673C>T, 11985 A>T and Arg⁴¹⁵Gln) and their association with lung cancer was assessed as shown in table 5.43. The study population was divided into smokers and non-smokers on the basis of their smoking status (Table 5.43). The smoking group was further dichotomized into light smokers (≤ 25 pack years) and heavy smokers (> 25 years). In case of 673C>T, controls subjects who were smokers and carrying the mutant genotype (TT) were equally distributed to cases (14.39% vs 14.1%). Smokers having wild type genotype (CC) were less prevalent in controls in comparison with cases (28.4% vs 36.3%). The percentage of smokers with heterozygous genotype was 49.5% in cases and 57.19% in controls. Smokers carrying heterozygous (CT) genotype of the 673C>T gene were found to have elevated risk for lung cancer as compared (OR=1.38; 95% CI: 0.94-2.02; $p=0.09$) to smokers having wild type genotype (CC). Among heavy smokers the heterozygous genotype was more presented in cases (33.33%) than controls (24%). Almost a 2-fold risk was observed in those heavy smokers who were carrying of heterozygous genotype (CT) (OR: 1.78; 95%CI: 0.96-3.28; $p=0.06$) as compared to those who were carrying the reference type (CC). Furthermore, it was observed that heavy smokers with the combined (CT+TT) genotype for 673C>T were at significantly higher risk (OR= 1.78, 95%CI=1.02-3.10; **$p=0.03$**) than light smokers showing insignificant risk (OR=1.14, 95%CI=0.73-1.79; $p=0.54$). No relationship was found in non-smokers with 673C>T polymorphism and lung cancer risk as shown in table 5.43.

In case of *XPF* 11985A>G polymorphism, smokers with heterozygous genotype (AG) were more pronounced in cases as compared with controls (35.3% vs 30.9%). Smokers with mutant genotype (GG) were almost equally expressed in both cases and controls (6.2% vs 7.3%). A moderate risk was observed for both heterozygous and combined genotype in 11985A>G polymorphism i.e. AG (OR: 1.23; 95%CI: 0.85-1.78; $p=0.25$) and AG+GG (OR: 1.18; 95%CI: 0.83-1.66; $p=0.34$) for smokers. None of the genotype expressed any association with lung cancer risk in smokers for 11985A>G polymorphism. The mutant genotype (GG) in heavy smokers was almost similarly represented in controls when compared with cases (8% vs 6.3%). Wild type genotype (AA) was more represented in controls as compared to cases (59% vs 54.6%). Among light smokers wild genotype (AA) was similarly present in both cases and controls (61.7% vs 63.15%). The heterozygous genotype also showed equal expression in

controls and cases (29.23% vs 32.09%). A lack of statistical significance was observed for heavy as well as light smokers and lung cancer susceptibility for 11985A>G polymorphism as represented in table 5.43.

For *Arg*⁴¹⁵*Gln* polymorphism, the heterozygous genotype (GA) was almost similarly expressed in cases (42.5%) as well as controls (42.8%) respectively. Smokers with mutant genotype were slightly more expressed in controls (7.74%) than cases (4.9%). Smokers carrying the mutant genotype (AA) were found to be associated with decreased risk for lung cancer (OR=0.70; 95% CI: 0.34-1.44; *p*=0.33) as compared to smokers having wild type genotype (GG). However, results were not significant. In heavy smokers, the representation of controls (47%) was similar to that of cases (49.6%) as shown in table 5.43. A similar trend was observed for heterozygous genotype (AG) illustrating 46% distribution in cases and 46.8% in controls. However, no significant association was found between *Arg*⁴¹⁵*Gln* polymorphism and lung cancer risk in heavy smokers. Among non smokers, heterozygous genotype (GA) was expressed in controls than cases (52.52% vs 37.31%) whereas the wild genotype was more expressed in cases than controls (59.70% vs 40.4%). A significant protective effect among non smokers carrying heterozygous genotype (GA) was observed when compared with the wild genotype (GG) (OR: 0.51; 95%CI: 0.26-0.99; **p=0.04**). Combined genotype (GA+AA) showed protective effect for non smokers with *Arg*⁴¹⁵*Gln* polymorphism (OR: 0.49; 95%CI: 0.25-0.94; **p=0.03**) as shown in table 5.43.

Table 5.43: Distribution of genotypes of <i>XPF</i> polymorphism according to their smoking status and its association with lung cancer risk								
	SMOKERS				NON SMOKERS			
<i>673C>T</i>	Controls n (%) N =271	Cases n (%) N=303	AOR (95% CI) ^a	<i>p</i> ^b	Controls n (%) N =99	Cases n (%) N=67	AOR (95% CI) ^a	<i>p</i> ^b
CC	77(28.4)	110(36.3)	1.00 (Reference)		45(45.45)	31(46.2)	1.00 (Reference)	
CT	155(57.19)	150(49.5)	1.38(0.94-2.02)	0.09	34(34.34)	31(46.2)	1.52(0.75-3.09)	0.24
TT	39(14.39)	43(14.1)	1.12(0.68-1.84)	0.65	20(20.2)	5(7.4)	0.41(0.13-1.28)	0.12
CT+TT	194(71.5)	193(63.6)	1.29(0.91-1.81)	0.14	64(95.5)	36(53.7)	1.10(0.57-2.13)	0.75
<i>11985 A>G</i>								
AA	167(61.6)	177(58.4)	1.00 (Reference)		60(60.6)	38(56.7)	1.00 (Reference)	
AG	84(30.9)	107(35.3)	1.23(0.85-1.78)	0.25	35(35.35)	23(34.32)	1.02(0.51-2.03)	0.94
GG	20(7.3)	19(6.2)	0.93(0.46-1.86)	0.84	4(4.04)	68.9)	2.19(0.49-9.67)	0.30
AG+GG	104(38.3)	126(41.5)	1.18(0.83-1.66)	0.34	39(39.39)	29(43.28)	1.14(0.59-2.20)	0.68
<i>Arg⁴¹⁵Gln</i>								
GG	134(49.44)	159(52.4)	1.00 (Reference)		40(40.4)	40(59.70)	1.00 (Reference)	
GA	116(42.8)	129(42.5)	0.94(0.66-1.33)	0.73	52(52.52)	25(37.31)	0.51(0.26-0.99)	0.04
AA	21(7.74)	15(4.9)	0.70(0.34-1.44)	0.33	7(7.07)	2(2.9)	0.28(0.05-1.50)	0.13
GA+AA	137(50.5)	144(47.52)	0.90(0.64-1.27)	0.56	59(59.59)	27(40.2)	0.49(0.25-0.94)	0.03
	HEAVY SMOKER				LIGHT SMOKER			
<i>673C>T</i>	Controls n (%) N =100	Cases n (%) N=141	AOR (95% CI) ^a	<i>p</i> ^b	Controls n (%) N =171	Casesn (%) N=162	AOR (95% CI) ^a	<i>p</i> ^b
CC	67(67)	76(53.9)	1.00 (Reference)		88(51.4)	74(45.6)	1.00 (Reference)	
CT	24(24)	47(33.33)	1.78(0.96-3.28)	0.06	53(30.9)	63(38.8)	1.24(0.75-2.04)	0.38
TT	9(9)	18(12.76)	1.82(0.74-4.47)	0.19	30(18.7)	25(15.4)	0.97(0.52-1.82)	0.93
CT+TT	33(33)	65(46.09)	1.78(1.02-3.10)	0.03	83(48.5)	88(54.3)	1.14(0.73-1.79)	0.54
<i>11985 A>G</i>								
AA	59(59)	77(54.6)	1.00 (Reference)		108(63.15)	100(61.7)	1.00 (Reference)	
AG	33(33)	55(39.0)	1.22(0.69-2.14)	0.48	50(29.23)	52(32.09)	1.18(0.78-1.93)	0.48
GG	8(8)	9(6.3)	0.78(0.27-2.19)	0.68	13(7.6)	10(6.1)	0.87(0.34-2.24)	0.78
AG+GG	41(41)	64(45.3)	1.14(0.93-2.88)	0.08	63(36.8)	62(38.2)	0.99(0.64-1.53)	0.97
<i>Arg⁴¹⁵Gln</i>								
GG	47(47)	70(49.6)	1.00 (Reference)		86 (50.29)	89(54.9)	1.00 (Reference)	
GA	46(46)	66(46.8)	0.97(0.56-1.76)	0.94	70 (40.93)	63(38.8)	0.91(0.57-1.47)	0.72
AA	7(7)	5(3.5)	0.61(0.18-2.10)	0.44	15 (8.77)	10(6.1)	0.90(0.36-2.23)	0.82
GA+AA	53(53)	71(50.35)	0.93(0.54-1.61)	0.81	85 (49.70)	73(45.06)	0.90(0.57-1.42)	0.67

5.5.7.3 Haplotype analysis of *XPF* variants

We analyzed haplotypes using SHEsis program platform (table 5.44). Our analysis showed that the three SNPs were in weak linkage disequilibrium with each other in this study population: *XPF* 673C>T and *XPF* 11985A>G (D=0.041, $r^2=0.001$); *XPF* 11985A>G and *XPF* Arg⁴¹⁵Gln (D=0.027, $r^2=0.00$); *XPF* Arg⁴¹⁵Gln and *XPF* 673 C>T (D=0.011, $r^2=0.00$). There was no statistically significant difference in the overall haplotype distribution between cases and controls (global $\chi^2=13.24$, $p=0.06$). CAA haplotype showed a protective effect against lung cancer risk (OR: 0.68; 95%CI: 0.51-0.90; $p=0.008$). All other exploratory haplotypes did not show any significant effect. The TGA haplotype was less expressed among controls (12.50%) than cases (22.54%), thus illustrated a close to significant risk in lung cancer patients (OR: 1.80; 95%CI: 0.90-3.67; $p=0.08$). TAA haplotype was more represented in controls (53.26%) than cases (39.35%). All other haplotypes did not represent any significant association with lung cancer risk as they were almost equally distributed in cases and controls as shown in table 5.44.

673C>T, 11985A>G & Arg⁴¹⁵Gln	Case frequency, n (%)	Control frequency, n (%)	OR (95% CI)^a	p^b-value
CAA	0.12 (93.34)	0.17 (129.50)	0.68 (0.51-0.90)	0.008
CAG	0.39 (291.71)	0.36 (269.40)	1.13 (0.92-1.47)	0.23
CGA	0.04(32.77)	0.03 (28.68)	1.14 (0.68-1.91)	0.59
CGG	0.11 (85.19)	0.11 (81.36)	1.05 (0.76-1.45)	0.75
TAA	0.05 (39.35)	0.07 (53.26)	0.72 (0.47-1.10)	0.13
TAG	0.18 (135.60)	0.16 (120.78)	1.15 (0.87-1.50)	0.30
TGA	0.03 (22.54)	0.01 (12.50)	1.80 (0.90-3.67)	0.08
TGG	0.05 (39.50)	0.06 (44.46)	0.88 (0.56-1.37)	0.57

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding p -values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status (b)-Two-sided χ^2 test for either haplotypic distribution or frequencies between the cases and controls. Global χ^2 test=13.24, p^b -value = 0.06

5.5.7.4 Combined effect of *XPF* SNPs on lung cancer risk

The genotypic combinations of the three *XPF* polymorphic variants were also combined with each other and analyzed to assess their association towards lung cancer risk, as shown in table 5.45. Different SNP-SNP combinations were applied for three *XPF* SNPs (673C>T, 11985A>G

and *Arg⁴¹⁵Gln*). Group 0 which composed of wild genotypic combination for all the three SNPs was regarded as the reference. Different dual combinations between the three SNPs were also analyzed. The first combination was between 673C>T and 11985A>G polymorphism, the subject with wild genotype (CC+AA) for both SNPs were more represented in control set than cases (76.54% vs 67.07%) whereas for heterozygous genotypes (CT+AG) cases were more represented than controls (31.09% vs 20.37%). Individuals carrying any single variant allele for 673C>T and 11985A>G polymorphic variants *i.e.* heterozygous carriers (CT+GA) showed almost two fold risk towards lung cancer (OR=1.84; 95%CI: 1.07-3.14; **p=0.02**). As the number of mutant genotype was less therefore, a double variant genotypes (mutants) were combined together with heterozygous genotype and act as a single genotype [(CT+TT) + (GA+AA)] for 673C>T and 11985A>G variants, a similar risk for the disease (OR=1.67; 95%CI: 1.00-2.79; **p=0.04**) was also observed as shown in table 5.45.

For 11985A>G and *Arg⁴¹⁵Gln* polymorphic combination, both genotypic combination groups showed almost equal distribution in cases and controls *i.e.* group 0 (68.36% vs 67.33%) and group 1 (31.63% vs 32.67%). Therefore, no significant relationship can be concluded through this SNP-SNP interaction. The other dual combination which has been applied was between *Arg⁴¹⁵Gln* and 673C>T, the heterozygous genotype (group 1) was slightly less prevalent in cases (62.91%) than controls (64.39%) whereas the variant genotype (group 2) was slightly more expressed in controls (4.54%) than cases (1.98%). A protective trend was observed for this polymorphic combination in variant genotype (OR: 0.58; 95%CI: 0.13-2.51; p=0.47) when compared with the reference group but results were not significant. The only possible triple combination with these SNPs (673C>T, 11985A>G and *Arg⁴¹⁵Gln*) indicates over representation of wild genotype (group 0) in controls (78.68%) than cases (73.41%) while heterozygous genotype (group 1) was more presented in cases (26.58%) than controls (21.31%) as shown in table 5.45. No significant association was observed for triple combination and lung cancer risk.

Table 5.45: Combination of genotypes of *XPF* polymorphic variants and their relationship with the risk of lung cancer

Genotypes	Cases (n)	Controls (n)	Adjusted OR (95% CI) ^a	P ^b
673 C>T, 11985 A>G and Arg⁴¹⁵Gln	N=79	N=61		
0 (CC+AA+GG)	58 (73.41)	48 (78.68)	1.00 (Reference)	
1(CT+AG+GA)	21 (26.58)	13 (21.31)	1.74 (0.74-4.07)	0.19
2(TT+GG+AA)	-	-	-	-
3(1+2)	-	-	-	-
673 C>T and 11985 A>G	N=164	N=162		
0 (CC+AA)	110 (67.07)	124 (76.54)	1.00 (Reference)	
1(CT+AG)	51 (31.09)	33 (20.37)	1.84 (1.07-3.14)	0.02
2 (TT+GG)	3 (1.82)	5 (3.08)	0.69(0.15-3.00)	0.61
3(1+2)	54 (32.92)	38 (23.45)	1.67(1.00-2.79)	0.04
11985 A>G and Arg⁴¹⁵Gln	N=177	N=150		
0 (AA+GG)	121 (68.36)	101 (67.33)	1.00 (Reference)	
1 (AG+GA)	56 (31.63)	49 (32.67)	1.02 (0.63-1.66)	0.90
2 (GG+AA)	-	-	-	-
3(1+2)	-	-	-	-
Arg⁴¹⁵Gln and 673C>T	N=151	N=132		
0 (GG+CC)	95 (62.91)	85 (64.39)	1.00 (Reference)	
1 (GA+CT)	53 (35.09)	41 (31.06)	1.28 (0.76-2.77)	0.34
2 (GG+TT)	3 (1.98)	6 (4.54)	0.58 (0.13-2.51)	0.47
3(1+2)	56 (37.08)	47 (35.60)	1.19 (0.72-1.98)	0.48

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status. (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 3: combined hetero and mutant genotype

5.5.7.5 Relationship of Clinico-pathological features with *XPF* polymorphic variants

Clinico-pathological parameters such as stage, primary tumor extension and metastasis were analyzed to find out the effect of genetic variations in *XPF* gene on these factors as shown in table 5.46. The effect of different genotypes of *XPF* genetic variants and their relationship with advancement of these clinical parameters and progression of the disease was assessed using logistic linear regression. Odds ratio obtained were adjusted for age, gender and smoking status. The wild genotype for each polymorphism was regarded as the reference group.

On the basis of clinical stage the patients were stratified into stage III and stage IV. The genotypic distribution for each polymorphic site of *XPF* with the clinical stage was estimated. According to table 5.46, the wild genotype (CC) depicting 673C>T polymorphism was more represented in stage III (53.71%) than stage IV (47.23%) whereas mutant genotype was equally

distributed in both stage III and stage IV patients (11.4% vs 12.88%). On applying regression analysis no significant association was observed between 673C>T polymorphism and tumor stage. In case of 11985A>G polymorphism, the mutant genotype (GG) was similarly represented in stage III and stage IV patients respectively (6.28% vs 6.74%). The prevalence of subjects carrying heterozygous (AG) genotype was higher in stage III (37.71%) in comparison with stage IV (28.22%) patients. On applying logistic regression analysis, it was found out that cases with heterozygous genotype (AG) showed higher odds ratio in stage IV as compared to stage III (OR: 1.63; 95%CI: 1.05-2.66; p=**0.04**) as shown in table 5.45. It was evident that 11985A>G polymorphism was related with genetic variants that contribute risk in stage IV patients. While in 673C>T polymorphism heterozygous genotype (CT) was less distributed in patients with stage III (34%) than stage IV patients (39.87%). Also, the combined genotype (CT+TT) for 673C>T polymorphism were more represented among stage IV (52.76%) than stage III (46.28%) patients. However, there was a lack of association between 673C>T polymorphism and clinical staging.

The genotypic frequencies for the each polymorphic site were compared into two groups on the basis of primary tumor extension as shown in table 5.46. The investigative study of 673C>T polymorphism for its correlation with primary tumor extension suggested that the frequency of subjects harboring the double allele of variant genotype (TT) was more represented among T3+T4 subgroup (14.49%) than T0+T1+T2 (7.24%). Therefore, no significant association was observed between the polymorphic variant and primary tumor extension (OR: 2.11; 95% CI: 0.77-5.81; p=0.14) as shown in table 5.46. The genotypic frequencies of the patients having variant (GG) genotype for 11985A>G polymorphism were statistically different among T0+T1+T2 (11.59%) and T3+T4 (5.79%) subgroup. Logistic regression analysis showed a decreased risk for mutant genotype (GG) showing T3+T4 extension (OR: 0.68; 95%CI: 0.43-1.10; p=0.11), however results were not significant. Also, in case of Arg⁴¹⁵Gln polymorphism, the frequency of heterozygous patients (GA) was almost similar in T0+T1+T2 (40.57%) and T3+T4 (40.94%) subgroup. The patients with mutant genotype (AA) for Arg⁴¹⁵Gln polymorphism was also equally expressed in T0+T1+T2 subgroup (4.34%) and T3+T4 (4.34%) subgroups showing an insignificant effect for Arg⁴¹⁵Gln polymorphism (OR: 0.93; 95%CI: 0.48-1.82; p=0.84). The analysis showed that none of the polymorphic site was associated with primary tumor extension.

Further the genotypic distribution was analyzed on the basis of distant metastasis. The genotypic frequencies of variant genotype (AA) for 673C>T polymorphism exhibited no significant difference between the patients who showed metastasis (13.29%) and those who showing absence of metastasis (12.69%). The frequency difference of the heterozygous genotype was slightly higher in subjects with metastasis than subjects without metastasis (39.24% vs 36.50%). Hence, this polymorphic site did not show any association with spread of lung cancer to other organs as shown in table 5.46. It was observed that 11985A>G contributed in modulating the metastatic potential of lung tumor. For 11985A>G polymorphism, the patients carrying both the susceptible alleles did not show any significant difference in frequency among M0 (6.87%) and M1 (6.96%) subgroup. For subjects carrying heterozygous genotype (AG) for 11985A>G polymorphism, the percentage of heterozygotes was 39.68% in patients without metastasis and 29.11% in cases showing distant metastasis. Regression analysis showing that patients carrying heterozygous genotype (GA) had a decreased susceptibility of having distant metastasis (OR: 0.59; 95%CI: 0.36-0.95; p=**0.03**). As shown in table 5.46, Such trend was also observed for patients carrying combined genotype (GA+AA) suggesting a decreased risk of having metastasis (OR: 0.60; 95%CI: 0.38-0.95; p=**0.03**). For *Arg⁴¹⁵Gln* polymorphism, it was evident that wild genotype was more frequent in patients with distant metastasis than the patients who lack metastasis (59.49% vs 50.79%) whereas heterozygous genotype (GA) was expressed more in patients without metastasis than patients showed positive metastasis (42.85% vs 37.97%). Since there was no significant difference in the frequency distribution among the two discussed groups, it was evident that *Arg⁴¹⁵Gln* polymorphic site did not show any relation with metastasis or spread of lung cancer to other organs as shown in table 5.46.

Table 5.46: Relationship of Clinico-pathological features with XPF genotypes in lung cancer patients

Variable	Clinical stage		AOR (95% CI) ^a	<i>p</i> ^b	TNM stage		AOR (95% CI) ^a	<i>p</i> ^b	Metastasis		AOR (95% CI) ^a	<i>p</i> ^b
	III	IV			T3+T4	T2+T1+T0			Positive	Negative		
673 C>T	175	163			276	69			158	189		
CC	94 (53.71)	77 (47.23)	Reference		135 (48.91)	36 (52.17)	Reference		75 (47.46)	96 (50.79)	Reference	
CT	61 (34)	65 (39.87)	0.87 (0.53-1.43)	0.60	101 (36.59)	28 (40.57)	0.95 (0.54-1.67)	0.87	62 (39.24)	69 (36.50)	1.00 (0.62-1.63)	0.97
TT	20 (11.4)	21 (12.88)	0.67 (0.33-1.37)	0.28	40 (14.49)	5 (7.24)	2.11 (0.77-5.81)	0.14	21 (13.29)	24 (12.69)	1.22 (0.61-2.43)	0.55
CT+TT	81 (46.28)	86 (52.76)	0.82 (0.52-1.56)	0.39	141 (51.08)	33 (47.82)	1.13 (0.66-1.93)	0.64	83 (52.53)	93 (49.20)	1.06 (0.68-1.65)	0.78
11985 A>G												
AA	98 (56)	106 (65.03)	Reference		163 (59.05)	37 (53.62)	Reference		101 (63.92)	101 (53.43)	Reference	
AG	66 (37.71)	46 (28.22)	1.63 (1.05-2.66)	0.04	97 (35.14)	24 (34.78)	0.88 (0.49-1.57)	0.67	46 (29.11)	75 (39.68)	0.59 (0.36-0.95)	0.03
GG	11 (6.28)	11 (6.74)	1.42 (0.53-3.77)	0.47	16 (5.79)	8 (11.59)	0.68 (0.43-1.10)	0.11	11 (6.96)	13 (6.87)	0.66 (0.25-1.69)	0.38
AG+GG	77 (44)	57 (34.96)	1.59 (1.00-2.53)	0.04	113 (40.94)	32 (46.37)	0.75 (0.50-1.14)	0.18	57 (36.07)	88 (46.56)	0.60 (0.38-0.95)	0.03
Arg⁴¹⁵Gln												
GG	90 (51.42)	96 (58.89)	Reference		150 (54.34)	38 (55.07)	Reference		94 (59.49)	96 (50.79)	Reference	
GA	75 (42.85)	62 (38.03)	1.27 (0.80-2.02)	0.30	113 (40.94)	28 (40.57)	0.99 (0.57-1.72)	0.93	60 (37.97)	81 (42.85)	0.76 (0.48-1.21)	0.26
AA	10 (5.71)	5 (3.06)	1.87 (0.59-5.91)	0.28	13 (4.71)	3 (4.34)	0.93 (0.48-1.82)	0.84	4 (2.53)	12 (6.34)	0.40 (0.12-1.34)	0.14
GA+AA	85 (48.57)	67 (41.10)	1.32 (0.84-2.07)	0.22	126 (45.65)	31 (44.92)	0.98 (0.62-1.56)	0.95	64 (40.50)	93 (49.20)	0.72 (0.46-1.13)	0.15

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status.(b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

5.5.7.6 MDR analysis applied on *XPF* variants

The MDR analysis was performed on a dataset having subjects with or without lung cancer resulting into average cross validation model (CVC) and average prediction error (Table 5.47; Fig 5.24A). MDR tool was used to identify which combination of factors could be the best model to predict high risk subgroups of the provided subjects. The analysis resulted in best one factor model which includes *XPF Arg⁴¹⁵Gln* showing CVC of 100/100 and prediction error of 0.47 ($p=0.06$). A higher CVC and a minimum prediction error are preferred for the best results. *XPF 673C>T, 11985A>G, Arg⁴¹⁵Gln* appeared as the best interaction model with CVC of 100/100 and having prediction error of 0.49 ($p=0.06$). The two factor model (*11985A>G, Arg⁴¹⁵Gln*) for evaluating lung cancer risk had CVC of 99/100 and an improvised prediction error of 0.47 ($p=0.90$). To demonstrate a visualized interaction between the three *XPF* SNPs and lung cancer risk an interaction dendrogram was used (Fig 5.22A). As *Arg⁴¹⁵Gln* and *11985A>G* were connected with the shortest line the more synergy was depicted between the two polymorphisms whereas a weak interaction was observed between *673C>T* and other two polymorphic variants.

Table 5.47: Multifactor Dimensionality reduction (MDR) analysis showing interactions of *XPF* variants with lung cancer risk

No. of risk factors	Best interaction model	Cross Validation Consistency (CVC)	Prediction error	p – value
1	<i>Arg⁴¹⁵Gln</i>	100/100	0.47	0.06
2	<i>11985 A>G, Arg⁴¹⁵Gln</i>	99/100	0.47	0.85
3	<i>673 C>T, 11985 A>G, Arg⁴¹⁵Gln</i>	100/100	0.49	0.06

5.5.7.7 CART analysis applied on *XPF* variants

Higher order SNP-SNP interactions and their association lung cancer risk were analyzed using CART analysis (Table 5.48; Fig 5.24B). In Fig 5.24B, A CART tree comprising of four terminal nodes was observed. The root split takes place at the strongest risk factor suggesting *673C>T* as the most risk factor for lung cancer risk amongst other variants. Further subsequent split was on *XPF Arg⁴¹⁵Gln* and *11985A>G*. To understand potential mechanism for increased lung cancer risk with SNPs in CART analysis, we also applied logistic regression analysis in the final model

selected by CART evaluating OR and 95%CI. In table 5.47, Node 3 was considered as the reference as it has the minimum case rate *i.e.* 33.3%. Individuals carrying wild genotype of *XPF*673C>T and mutant (heterozygous + mutant) genotype of *XPF Arg*⁴¹⁵*Gln* represent node 3. Subjects with *XPF* 673C>T (M) showed highest risk towards the lung cancer (OR: 1.60; 95%CI: 1.18-2.37; p=**0.003**). Whereas Node 2 carrying individuals with *XPF* 11985A>G (W)/*XPF Arg*⁴¹⁵*Gln* (W)/*XPF* 673C>T (W) genotype possessed a significantly increased risk for the disease (OR: 1.44; 95%CI: 1.00-2.08, p=**0.04**). Node 4 has individuals with a combined genotype of *XPF Arg*⁴¹⁵*Gln* (M) and 673C>T (W) showed protective effect for the lung cancer (OR: 0.65; 95%CI: 0.19-2.24; p=0.50).

Table 5.48: Risk estimates for *XPF* variants based on CART analysis

Terminal Node	Genotype for each terminal node	Control (n)	Case (n)	Case Rate ^a	OR (95% CI)	p ^b
Node 1	<i>XPF</i> 673C>T (M)	111	141	56.0	1.60(1.18-2.37)	0.003
Node 2	<i>XPF</i> 11985 A>G (W)/ <i>XPF Arg</i> ⁴¹⁵ <i>Gln</i> (W)/ <i>XPF</i> 673 C>T (W)	102	112	52.3	1.44(1.00-2.08)	0.04
Node 3	<i>XPF</i> 11985 A>G(M)/ <i>XPF Arg</i> ⁴¹⁵ <i>Gln</i> (W)/ <i>XPF</i> 673 C>T(W)	8	4	33.3	1.0 ^c	
Node 4	<i>XPF Arg</i> ⁴¹⁵ <i>Gln</i> (M)/ <i>XPF</i> 673C>T(W)	149	113	43.1	0.65(0.19-2.24)	0.50

^acase rate is percentage LC cases among all individuals in each terminal node {case / (case+ control) x100}; ^bp<0.05 is considered statistically significant; ^c reference group

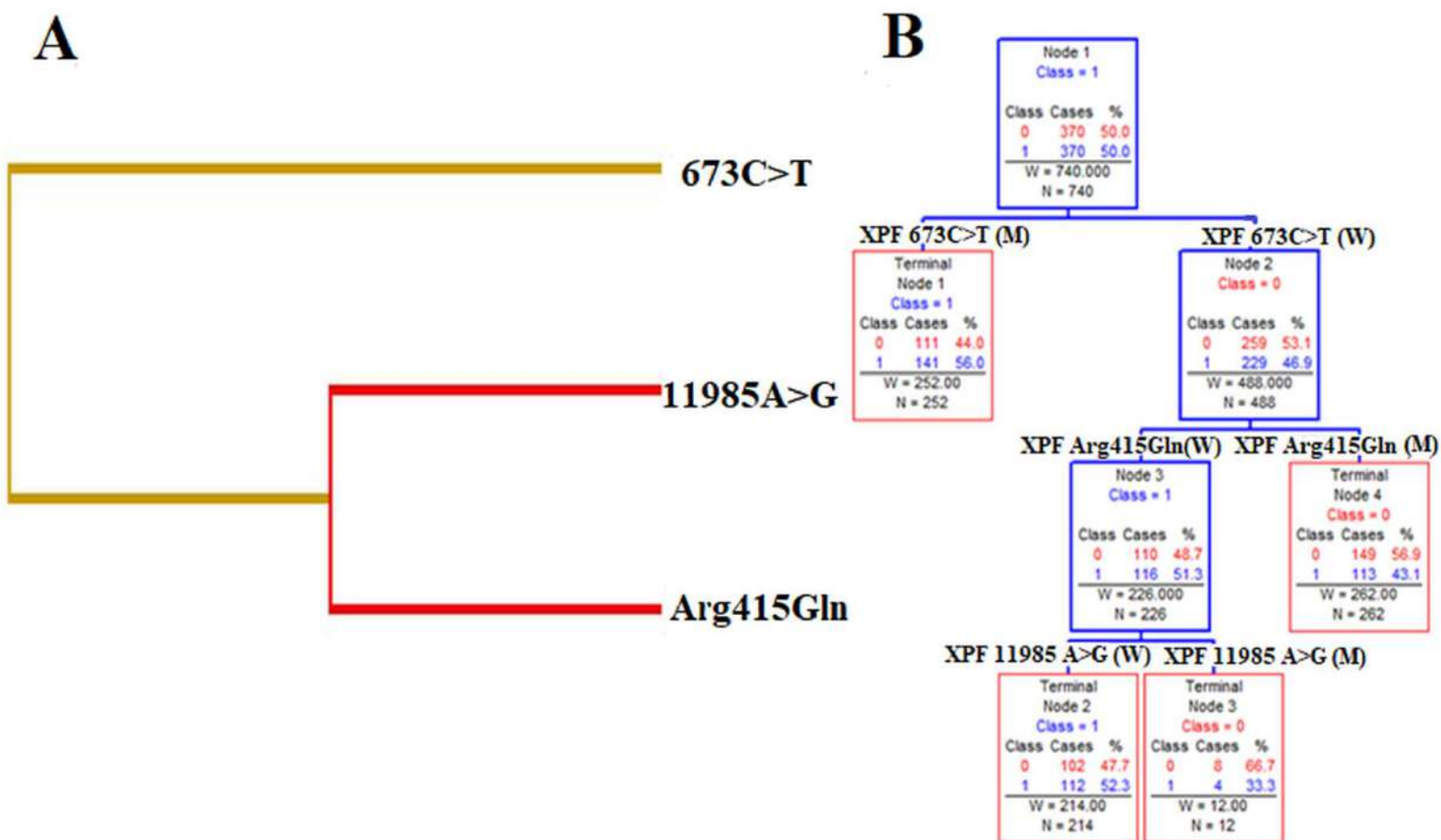


Fig 5.24: (A) Interaction dendrogram gained from the multifactor dimensionality reduction (MDR) showing gene-gene interaction for *XPF* 673C>T, 11985A>G and *Arg*⁴¹⁵Gln **(B)** Classification and regression tree (CART) analysis depicting high order interaction among three polymorphisms of *XPF*

5.5.7.8 *XPF* polymorphism and their association with overall survival in lung cancer patients

Our data also evaluated the role of *XPF* polymorphism on the OS of lung cancer patients as shown in table in table 5.49. Firstly, a univariate analysis was performed by applying Kaplan Meier survival analysis and after that multivariate analysis was applied using Cox regression model. Kaplan Meier survival analysis results in to logrank test and crude hazard ratio. The parameter were further adjusted for age, gender, smoking status, stage, histology, KPS and ECOG using Cox regression analysis resulting into adjusted hazard ratio and p-value. The genotypic distribution in the patients agreed with that estimated from HWE for *11985A>G* ($\chi^2=0.28$; df=1; p=0.59) and *Arg⁴¹⁵Gln* ($\chi^2= 1.45$; df=1; p=0.22) polymorphisms. Genotypic distribution for *673C>T* polymorphism were not in agreement with HWE ($\chi^2= 5.59$; df=1; p=0.01). The wild genotype was more presented in all three polymorphism of *XPF* i.e. *673 C>T* (47.90%), *11985A>G* (59.80%) and *Arg⁴¹⁵Gln* (53.37%) as shown in table 5.49. For *673C>T* polymorphism the homozygous variant genotype (TT) was found in 13.82% lung cancer patients and subjects carrying mutant genotype (TT) showed a survival time of 9.43 months which was higher than that of subjects carrying wild genotype (CC) (MST=9.43 vs 8.06; logrank p=0.36). In case of *11985A>G* polymorphism, patients with mutant genotype (GG) showed similar median survival time in lung cancer cases when compared with wild genotype (AA) (8.40 vs 8.06 months; logrank p=0.17) as shown in table 5.49. Similarly, In case of *Arg⁴¹⁵Gln* polymorphism, the variant genotype (AA) showed a slight poor survival period in comparison with the patients carrying wild genotype (GG) (7.23 vs 8.03 months; logrank p=0.52). Overall none of the *XPF* polymorphic variants depicted any significant relationship with OS in lung cancer patients.

Table 5.49: Association of <i>XPF</i> polymorphism and overall survival analysis in lung cancer patients								
				UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
Genotype	No. n% N=311	Dead n N=27 1	Alive n N=40	Media n OS (mont hs)	HR (95% CI)	Log P	HR ¹ (95% CI)	P-value
673C>T								
CC	149 (47.90)	134	15	8.06	1.00(Reference)		1.00(Reference)	
CT	119(38.26)	101	18	7.33	0.92(0.71-1.20)	0.57	0.90(0.69-1.18)	0.49
TT	43 (13.82)	36	7	9.43	0.84(0.59-1.20)	0.36	0.90(0.60-1.35)	0.63
CT+TT	162 (52.09)	137	25	7.60	0.90(0.71-1.15)	0.40	0.91(0.71-1.16)	0.47
11985A>G								
AA	186 (59.80)	164	22	8.06	1.00(Reference)		1.00(Reference)	
AG	107 (34.4)	95	12	6.76	1.10(0.85-1.43)	0.42	1.16(0.89-1.52)	0.25
GG	18 (5.78)	12	6	8.40	0.66(0.40-1.09)	0.17	0.84(0.62-1.13)	0.25
AG+GG	125 (40.19)	107	18	7.56	1.03(0.80-1.32)	0.79	1.09(0.85-1.41)	0.46
Arg⁴¹⁵Gln								
GG	166 (53.37)	148	18	8.03	1.00(Reference)		1.00(Reference)	
GA	128 (41.15)	109	19	8.05	0.94(0.73-1.20)	0.64	0.94(0.73-1.21)	0.63
AA	17 (5.46)	14	3	7.23	0.83(0.50-1.39)	0.52	0.93(0.69-1.25)	0.66
GA+AA	145 (46.62)	123	22	7.93	0.93(0.73-1.18)	0.56	0.93(0.73-1.19)	0.93

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

5.5.7.9 Association of *XPF* SNPs with overall survival in histological subtypes of lung cancer

Overall survival in *XPF* polymorphism and their association with different histological subtypes of lung cancer are summarized in Table 5.50 (a, b, c). In case of 673C>T polymorphism, the representation of heterozygous genotype (CT) was less in patients in comparison with subjects carrying wild type genotype (CC) for ADCC subjects (37.5% vs 47.15%). ADCC patients carrying heterozygous variant (CT) genotype showed a better OS as compared to wild genotype (8.80 months vs 6.33 months; logrank *p*=0.13). A low hazards ratio was obtained from multivariate analysis and was reported to be significant (HR¹= 0.52, 95% CI: 0.32-0.86, *p*=**0.01**;

Fig 5.25A). A similar trend was observed for combined genotype (CT+TT) suggesting a significantly lower hazard ratio in ADCC patients ($HR^1 = 0.57$, 95% CI: 0.35-0.90, $p = \mathbf{0.01}$; MST=9.43 months). A survival period of 14.26 months was observed for homozygous variant genotype (TT) for 673C>T polymorphism when compared with the reference genotype (14.26 vs 6.33 months; logrank $p = 0.07$). On applying multivariate analysis, reduced hazard ratio was observed in patients carrying mutant genotype (TT) in comparison with wild genotype (HR^1 : 0.50; 95%CI: 0.50-1.13; $p = 0.09$), although results were not significant (table 5.50a). In case of *Arg⁴¹⁵Gln* polymorphism subjects harboring heterozygous genotype (GA) showed a poor survival when compared with the reference genotype (GG) (8.03 vs 9.13 months; logrank $p = 0.56$) depicting an increased mortality risk ($HR^1 = 1.57$, 95% CI: 1.05-2.33, $p = \mathbf{0.02}$; Fig 5.25C) as shown in table 5.50a.

In SQCC patients, for 673C>T polymorphism, the subjects carrying heterozygous genotype (AC) and mutant (CC) genotype had a survival period of 6.66 and 8.03 months respectively when compared with the patients carrying reference genotype (GG) (MST=7.93 months). On applying Kaplan-Meier survival analysis and Cox regression model, no significant association was observed between 673C>T polymorphism and OS in SQCC patients as shown in table 5.50 (b). For 11985A>G polymorphism, the reference genotype (AA) was over represented than the heterozygous genotype (AG) (62.39% vs 34.18%; logrank $p = 0.19$). The mutant genotype (GG) was present in less frequency in SQCC subjects (3.84%). On applying univariate Kaplan-Meier analysis (HR : 0.15; 95%CI: 0.06-0.35; $p = \mathbf{0.02}$), a significant hazard ratio was observed for mutant genotype (GG) when compared with wild type genotype (AA). When the hazard ratio adjusted with confounding factor using Cox regression model (HR^1 : 0.15; 95%CI: 0.02-1.14; $p = 0.06$) the value becomes insignificant. The combined genotype (AG+GG) showed a better survival for subjects 11985A>G polymorphism when compared with wild type genotype (AA) (9.33 vs 6.40 months; logrank $p = 0.06$). For *Arg⁴¹⁵Gln* polymorphism, all genotypes depicted almost similar survival period (GA: 6.86 months; AA: 7.93 months; GA+AA: 7.57) when compared with wild type genotype (7.33 months). No significant association can be suggested for *XPF* variants with OS in SQCC patients (table 5.50b).

Table 5.50 (a): Association of XPF polymorphism and overall survival analysis in ADCC patients

Genotype	No. n% N= 104	Dead n N=86	Alive n N=18	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (month s)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P-value
673C>T								
CC	49 (47.15)	44	5	6.33	1.00(Reference)		1.00(Reference)	
CT	39 (37.5)	31	8	8.80	0.70(0.45-1.11)	0.13	0.52(0.32-0.86)	0.01
TT	16 (15.38)	11	5	14.26	0.56(0.31-0.99)	0.07	0.50(0.50-1.13)	0.09
CT+TT	55 (52.88)	42	13	9.43	0.65(0.42-1.00)	0.04	0.57(0.35-0.90)	0.01
11985A>G								
AA	58 (55.76)	46	12	7.13	1.00(Reference)		1.00(Reference)	
AG	38 (36.53)	34	4	9.23	1.19(0.71-1.75)	0.61	1.12(0.70-1.79)	0.61
GG	8 (7.69)	6	2	8.96	0.88(0.39-1.99)	0.78	0.98(0.62-1.54)	0.94
AG+GG	46 (44.23)	40	6	9.23	0.92(0.71-1.19)	0.72	1.11(0.71-1.75)	0.62
Arg⁴¹⁵Gln								
GG	53 (50.96)	44	9	9.13	1.00(Reference)		1.00(Reference)	
GA	45 (43.26)	38	7	8.03	1.13(0.73-1.75)	0.56	1.57(1.05-2.33)	0.02
AA	6 (5.76)	4	2	9.81	0.75(0.30-1.89)	0.59	1.59(0.82-3.06)	0.16
GA+AA	51 (49.03)	42	9	8.03	1.08(0.70-1.65)	0.70	1.23(0.79-1.92)	0.35

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

Table 5.50 (b): Association of XPF polymorphism and overall survival analysis in SQCC patients

Genotype	No. n% N= 117	Dead n N=99	Alive n N=18	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (month s)	HR (95% CI) ^a	Log P	HR ¹ (95% CI) ^b	P-value
673C>T								
CC	58 (49.57)	50	8	7.93	1.00(Reference)		1.00(Reference)	
CT	44 (37.60)	36	8	6.66	1.01(0.65-1.55)	0.95	1.02(0.65-1.61)	0.90
TT	15 (12.82)	13	2	8.03	1.02(0.55-1.90)	0.92	1.12(0.56-2.20)	0.74
CT+TT	59 (50.42)	49	10	6.86	1.01(0.68-1.50)	0.93	1.02(0.68-1.55)	0.89
11985A>G								
AA	73 (62.39)	66	7	6.40	1.00(Reference)		1.00(Reference)	
AG	40 (34.18)	32	8	8.03	0.75(0.50-1.13)	0.19	0.79(0.50-1.26)	0.33
GG	4 (3.41)	1	3	-	0.15(0.06-0.35)	0.02	0.15(0.02-1.14)	0.06
AG+GG	44 (37.60)	33	11	9.33	0.68(0.45-1.01)	0.06	0.70(0.44-1.10)	0.13

Arg⁴¹⁵Gln								
GG	60 (51.28)	53	7	7.33	1.00(Reference)		1.00(Reference)	
GA	50 (42.73)	40	10	6.86	0.84(0.56-1.27)	0.46	0.75(0.48-1.16)	0.20
AA	7 (5.98)	6	1	7.93	0.85(0.38-1.88)	0.71	0.80(0.51-1.25)	0.34
GA+AA	57 (48.71)	46	11	7.57	0.84(0.57-1.25)	0.40	0.75(0.49-1.15)	0.20

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

Table 5.50 (c) :Association of XPF polymorphism and overall survival analysis in SCLC patients

Genotype	No. n% N= 90	Dead n N=86	Alive n N=4	UNIVARIATE ANALYSIS			MULTIVARIATE ANALYSIS	
				Median OS (month s)	HR (95% CI)	Log P	HR ¹ (95% CI)	P-value
673 C>T								
CC	42 (46.67)	40	2	10.4	1.00(Reference)		1.00(Reference)	
CT	36 (40)	34	2	6.30	1.09(0.68-1.72)	0.70	1.10(0.68-1.78)	0.69
TT	12 (13.33)	12	-	5.10	1.08(0.56-2.09)	0.80	1.65(0.74-3.70)	0.22
CT+TT	48 (53.33)	46	2	6.30	1.09(0.71-1.66)	0.68	1.14(0.73-0.54)	0.54
11985A>G								
AA	55 (61.11)	52	3	11.6	1.00(Reference)		1.00(Reference)	
AG	29 (32.22)	29	-	4	2.40(1.37-4.18)	<0.0001	3.19(1.87-5.44)	<0.0001
GG	6 (6.67)	5	1	4.46	1.61(0.52-4.96)	0.29	2.96(1.00-8.70)	0.04
AG+GG	35 (38.89)	34	1	4.13	2.20(1.33-3.65)	0.0001	2.88(1.75-4.71)	<0.0001
Arg⁴¹⁵Gln								
GG	53 (58.89)	51	2	8.26	1.00(Reference)		1.00(Reference)	
GA	33 (36.67)	31	2	10	0.98(0.63-1.54)	0.95	0.90(0.55-1.47)	0.68
AA	4 (4.44)	4	-	5.80	1.33(0.42-4.22)	0.57	1.52(0.86-2.69)	0.14
GA+AA	37 (41.11)	35	2	9.86	1.02(0.66-1.57)	0.92	0.98(0.61-1.57)	0.95

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall survival

In table 5.50(C), for 673C>T polymorphism, SCLC patients carrying the heterozygous genotype (CT) was less presented when compared with the wild type genotype (40% vs 46.67%). The mutant genotype was showed 13.33% presentation among SCLC patients depicted the lower median survival time of 5.10 months when compared with the SCLC cases carrying wild type genotype (5.10 vs 10.4 months; logrank $p=0.80$). However, no significant association was observed between overall survival in SCLC patients and 673C>T polymorphism as shown in table 5.49(C). In SCLC subjects carrying heterozygous genotype (AG) for 11985 A>G polymorphism showed a MST of 10 months pertaining to a high hazards ratio in case of univariate analysis (HR= 2.40, 95% CI: 0.37-4.18, log rank $p<0.001$; Fig 5.25B) and multivariate analysis (HR¹= 3.19, 95% CI: 1.87-5.44, $p<0.0001$). Similar results were observed for patients carrying combined genotype (AG+GG) suggesting almost 3-fold risk of death for 11985 A>G polymorphism (HR¹= 2.88, 95% CI: 1.75-4.71, $p<0.0001$). In case of Arg⁴¹⁵Gln polymorphism, the heterozygous genotype showed a better survival time when compared with the wild type genotype (10 vs 8.26; logrank $p=0.95$), however results obtained were not significant. Arg⁴¹⁵Gln polymorphism did not show any significant association with OS in SCLC patients.

5.5.7.10 XPF polymorphic variants and their association with Overall survival in lung cancer patients treated with different chemotherapy regimens

The overall survival of patients treated with different chemotherapy regimens and their association with XPF polymorphic variants has been illustrated in Table 5.50. Homozygous wild genotype for all the three polymorphic variants of XPF gene was considered as reference. As evaluated in table 5.51, for 673C>T polymorphism, lung cancer patients carrying heterozygous genotype (CT) treated with docetaxel cisplatin/carboplatin regimen depicted a poor survival of 6.76 months when compared with the wild type genotype (CC) having a median survival of 9.86 months. No significant genotypic distribution was observed for 673C>T polymorphism and OS in patients treated with docetaxel cis/carboplatin regimen. Lung cancer patients with mutant genotype (TT) and treated with irinotecan cisplatin/carboplatin regimen showed a better survival in patients when compared with the reference genotype (AA) (11 vs 7.23 months; log rank $P=0.95$). The hazard ratio observed for patients carrying mutant genotype was less but the result was not significant (HR¹= 0.77, 95% CI: 0.44-1.33, $p=0.72$). Similarly, the patients treated with

pemetrexed cisplatin/carboplatin regimen and having mutant genotype (TT) for 673C>T polymorphism showed a lower probability of risk when compared with the wild type genotype (CC) (HR¹= 0.77, 95% CI: 0.45-1.31, *p*=0.30; MST=12.23 months). However, no significant relationship was observed 673C>T polymorphism and OS in lung cancer treated with different chemotherapy regimen as shown in table 5.51.

In case of 11985A>G polymorphism, lung cancer patients with mutant genotype (GG) and treated with docetaxel cisplatin/docetaxel carboplatin regimen showed a low death ratio as compared to patients who carried the wild type genotype (AA) (HR=0.25; 95%CI: 0.06-1.09; *p*=0.05). Patients having mutant genotype (GG) for XPF 11985A>G polymorphism and treated with irinotecan cisplatin/irinotecan carboplatin regimen showed a shorter OS and a high death ratio when compared with those patients harboring the wild type genotype (AA) (MST=4.46 vs 11.26; HR=4.73; 95%CI: 0.82-27.01; *p*=0.08). Similarly for the same polymorphic variant as above lung cancer subjects who were harboring the heterozygous genotype (AG) and treated with irinotecan cisplatin/irinotecan carboplatin regimen showed a poor OS as compared to subjects having reference genotype (AA). Such patients exhibited a high mortality risk (HR=2.66; 95%CI: 1.45-4.88; *p*=0.001; MST=4.30; Fig 5.23d). Furthermore, Patients having combined genotype (AG+GG) and treated with irinotecan cisplatin/irinotecan carboplatin regimen depicted almost 2-fold hazard ratio when compared with subjects with wild genotype (HR=2.64; 95%CI: 1.46-4.78; ***p*=0.001**; Fig 5.25D).

The lung cancer patients treated with docetaxel cisplatin/carboplatin regimen and having mutant (AA) genotype for Arg⁴¹⁵Gln polymorphism showed a decreased probability of death (HR: 0.45; 95%CI: 0.19-1.09; *p*=0.07); however the results were not significant. Although, lung cancer patients depicted almost similar survival time in subjects carrying heterozygous genotype (GA) and treated with irinotecan cisplatin/carboplatin regimen in comparison with subjects carrying wild genotype (GG) (9.86 vs 9 months; logrank *p*=0.68). Similarly, subjects treated with pemetrexed cisplatin/carboplatin regimen showed similar median survival time for subjects carrying heterozygous genotype (GA) in comparison with wild genotype (GG) for Arg⁴¹⁵Gln polymorphism (9.43 vs 9.56 months; logrank *p*=0.53). There was no significant association was observed between Arg⁴¹⁵Gln polymorphism and OS in lung cancer patients treated with different chemotherapeutic regimens as shown in table 5.51.

TABLE 5.51: Association of XPF polymorphism and overall survival according to regimen

REGIMEN – docetaxel cis/carboplatin								
Genotype	N=76	Dead N=65	Alive N=11	Median OS(months)	Crude HR^a (95%CI)	Adjusted HR¹ (95%CI)^b	Log rank P	p-value
<i>673C>T</i>								
CC	38 (50)	32	6	9.86	1.00 (Reference)	1.00 (Reference)		
CT	31 (40.78)	26	5	6.76	1.18(0.70-1.99)	1.57(0.86-2.88)	0.52	0.14
TT	7 (9.21)	7	-	8.40	1.60(0.61-4.22)	1.34(0.79-2.27)	0.24	0.26
CT+TT	38 (50)	33	5	7.56	1.24(0.76-2.02)	1.51(0.86-2.65)	0.37	0.14
<i>11985A>G</i>								
AA	48 (63.15)	41	7	6.76	1.00 (Reference)	1.00 (Reference)		
AG	23 (30.26)	19	4	8.63	0.84(0.49-1.42)	0.97(0.53-1.79)	0.52	0.93
GG	5 (6.57)	2	3	-	0.31(0.13-0.76)	0.25(0.06-1.09)	0.09	0.05
AG+GG	28 (36.84)	21	7	10.26	0.76(0.43-1.20)	0.77(0.43-1.35)	0.22	0.36
<i>Arg⁴¹⁵Gln</i>								
GG	40 (52.63)	37	3	8.03	1.00 (Reference)	1.00 (Reference)		
GA	32 (42.10)	26	6	7.88	0.78(0.48-1.28)	0.87(0.50-1.51)	0.34	0.63
AA	4 (5.26)	2	2	-	0.30(0.12-0.72)	0.45(0.19-1.09)	0.07	0.08
GA+AA	36 (47.36)	28	8	8.20	0.71(0.43-1.15)	0.78(0.49-1.15)	0.17	0.38
REGIMEN – Irinotecan cisplatin/Irinotecan carboplatin								
Genotype	N=70	Dead N=63	Alive N=7	Median OS(months)	Crude HR^a (95%CI)	Adjusted HR¹ (95%CI)^b	Log rank P	p-value
<i>673C>T</i>								
CC	36 (51.47)	33	3	7.23	1.00 (Reference)	1.00 (Reference)		
CT	27 (38.57)	24	3	8.26	0.98(0.58-1.66)	1.03(0.56-1.89)	0.95	0.91
TT	7 (10)	6	1	11	0.85(0.37-1.96)	0.77(0.44-1.33)	0.72	0.35
CT+TT	34 (48.57)	30	4	9	0.95(0.58-1.56)	0.92(0.54-1.58)	0.85	0.78
<i>11985A>G</i>								
AA	42 (60)	38	4	11.26	1.00 (Reference)	1.00 (Reference)		
AG	25 (35.71)	23	2	4.30	2.34(1.25-4.39)	2.66(1.45-4.88)	0.001	0.001
GG	3 (4.28)	2	1	4.46	1.96(0.28-13.74)	4.73(0.82-27.01)	0.34	0.08

AG+GG	28 (40)	25	3	4.30	2.27(1.25-4.14)	2.64(1.46-4.78)	0.001	0.001
<i>Arg⁴¹⁵Gln</i>								
GG	38 (54.28)	35	3	9	1.00 (Reference)	1.00 (Reference)		
GA	29 (41.42)	25	4	9.86	0.93(0.55-1.55)	0.79(0.42-1.46)	0.78	0.45
AA	3 (4.28)	3	-	7.23	0.79(0.27-2.31)	1.20(0.42-3.43)	0.68	0.72
GA+AA	32 (45.71)	28	4	8.55	0.91(0.55-1.50)	0.73(0.48-1.56)	0.73	0.63
REGIMEN – Pemetrexed cisplatin/pemetrexed-carboplatin								
Genotype	N=69	Dead N=55	Alive N=14	Median OS(months)	Crude HR^a (95% CI)	Adjusted HR¹ (95% CI)^b	Log rank P	p-value
<i>673C>T</i>								
CC	35 (50.72)	30	5	9.46	1.00 (Reference)	1.00 (Reference)		
CT	26 (37.68)	20	6	9.43	0.82(0.47-1.43)	0.62(0.32-1.20)	0.50	0.16
TT	8 (11.59)	5	3	12.23	0.61(0.27-1.37)	0.73(0.36-1.46)	0.30	0.38
CT+TT	34 (49.27)	25	9	9.43	0.77(0.45-1.31)	0.68(0.36-1.25)	0.33	0.78
<i>11985A>G</i>								
AA	42 (60.86)	32	10	10	1.00 (Reference)	1.00 (Reference)		
AG	22 (31.88)	19	3	9.43	1.15(0.64-2.06)	1.17(0.63-2.17)	0.61	0.61
GG	5 (7.24)	4	1	7.6	1.24(0.40-3.88)	1.54(0.49-4.85)	0.67	0.45
AG+GG	27 (39.13)	23	4	9.33	1.17(0.68-2.02)	1.25(0.70-2.23)	0.55	0.43
<i>Arg⁴¹⁵Gln</i>								
GG	37 (53.62)	29	8	9.56	1.00 (Reference)	1.00 (Reference)		
GA	30 (43.47)	25	5	9.43	1.18(0.68-2.02)	1.38(0.76-2.50)	0.53	0.27
AA	2 (2.89)	1	1					
GA+AA	32 (46.37)	26	6	9.38	1.13(0.66-1.93)	1.30(0.72-2.33)	0.63	0.37

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months and (b)-Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS and ECOG. OS=Overall Survival

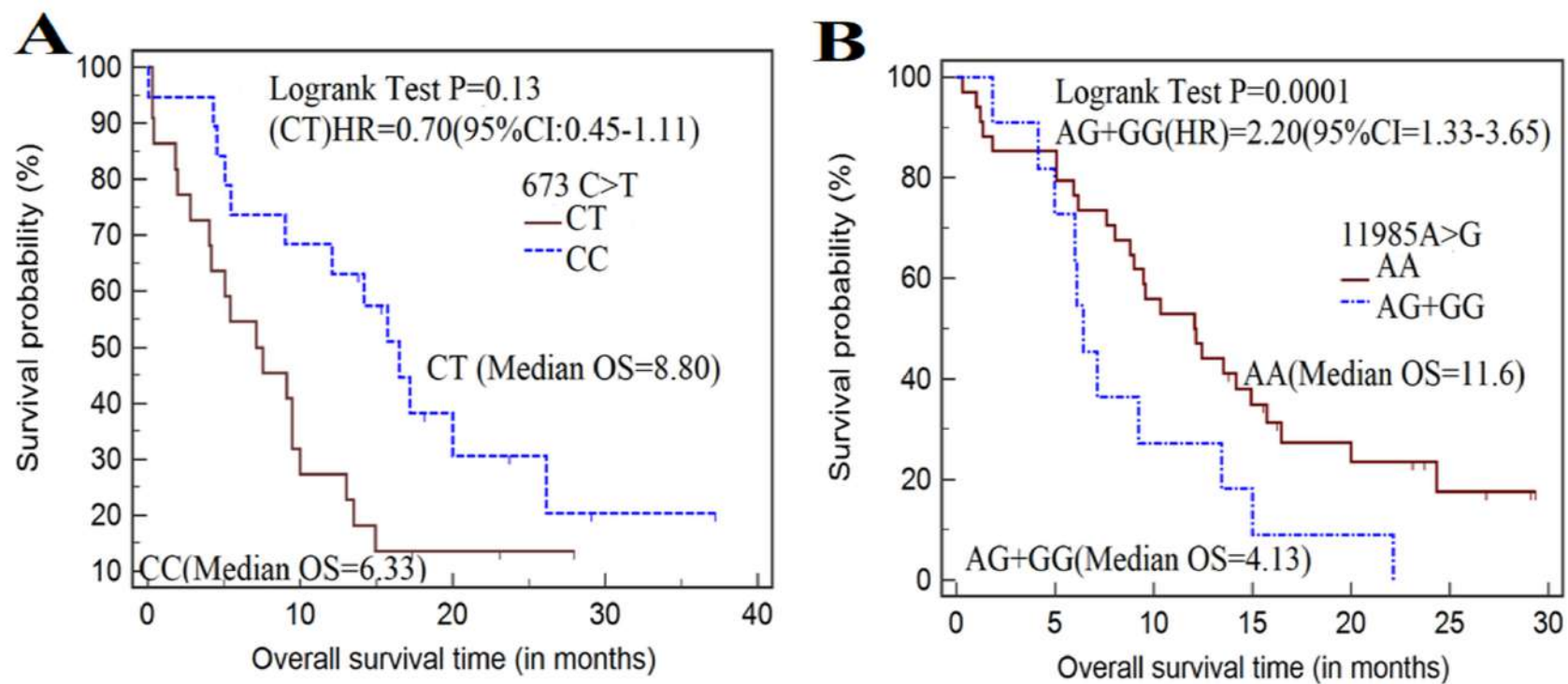


Fig 5.25: Kaplan Meier curves illustrating association between overall survival in patients with (A) Heterozygous genotype (CT) for ADCC showing *XPF* 673C>T polymorphism (B) Heterozygous genotype (AG) for SCLC showing 11985 A>G polymorphism,

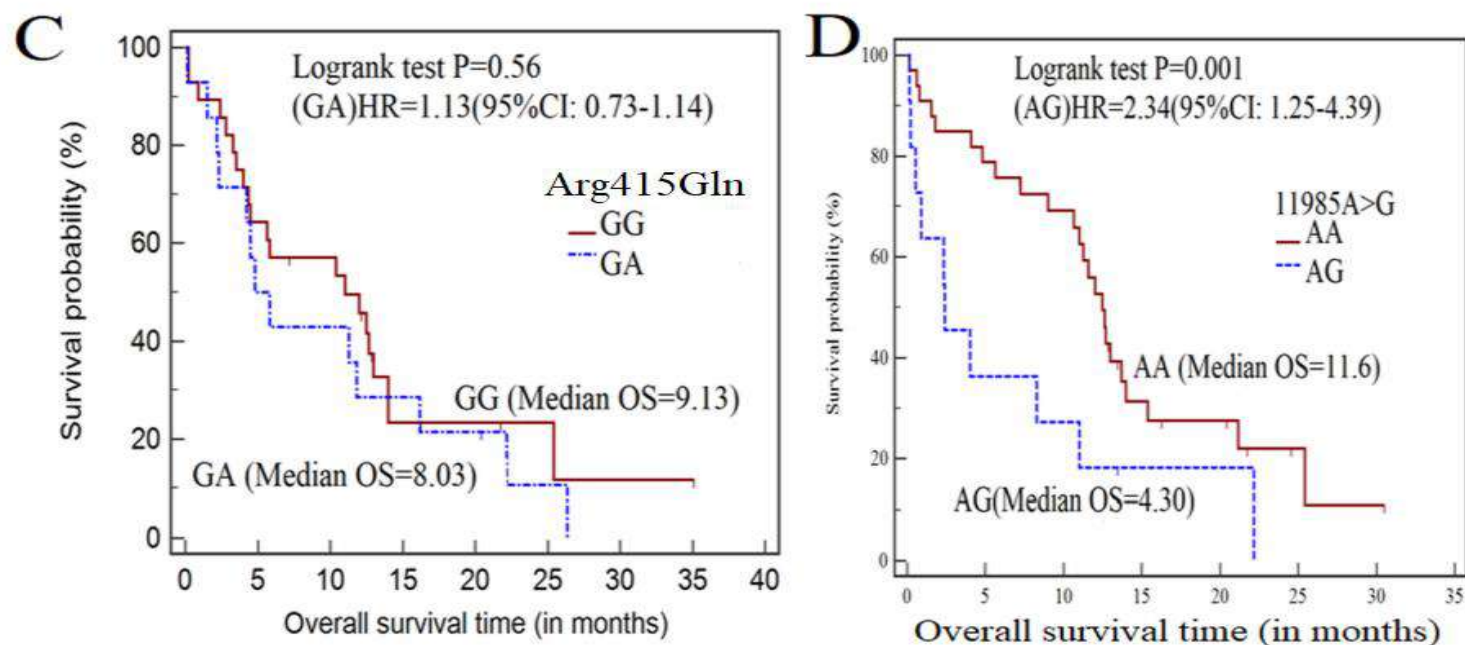


Fig 5.25: Kaplan Meier curves illustrating association between overall survival in patients with (C) Heterozygous genotype (GA) for ADCC in *Arg*⁴¹⁵*Gln* polymorphism (D) Kaplan Meier curves illustrating association between overall survival in lung cancer patients with heterozygous genotype (AG) treated with irinotecancisplatin/carboplatin regimen showing *11985 A>G* polymorphism.

5.5.7.11 Association of *XPF* polymorphism and also their correlation with objective response

The patients were analyzed for the response to the chemotherapeutic regimen in relation to their genotypes for different *XPF* variants as shown in table 5.52. RECIST criteria were applied on the patients and they were divided into four groups as per their response *i.e.* CR (complete response), PR (partial response), SD (stable disease) and PD (progressive disease). Responders comprises of CR+PR and non responders consist of SD+PD. Adjusted odds ratio was calculated for each polymorphic site considering wild genotype as the reference. The odds ratio was adjusted with the confounding factors such as age, gender, smoking, stage, histology, KPS and ECOG. The study was conducted on 202 patients. None of the polymorphic site for 673C>T, 11985A>G and *Arg*⁴¹⁵*Gln* gene revealed a significant difference in the frequency of patients showing clinical response. In case of 673C>T polymorphism, the representation of patients having CC, CA, AA, CA+AA genotype was almost similar among both responders (52.83%, 36.79%, 10.37% and 47.16%) and non responders (46.87%, 42.70%, 10.41% and 53.12%) respectively. The odds ratio for patients exhibiting mutant genotype (CC) for 673C>T was 1.15 (p=0.76). In case of 11985A>G polymorphism, the proportion of subjects having AG+GG genotype was higher (43.39%) among responders as compared to non responders (37.50%) but the difference was not found to be statistically significant. The frequency distribution in 11985A>G polymorphism further showed a higher representation of wild genotype (AA) among non-responders than responders (62.50% vs 56.60%). For *Arg*⁴¹⁵*Gln* polymorphism, 39.62% patients with heterozygous genotype were represented as responder and 40.62% as non responders which were almost similarly represented. The homozygous variant allele (AA) presented a low frequency distribution in both responders as well as non responder (4.71% vs 5.20%). A significant association cannot be observed between *XPF* polymorphic variants and clinical response in lung cancer patients.

Table 5.52: Genotypic Response of <i>XPF</i> polymorphism to platinum based chemotherapy and their association with lung cancer risk				
Genotype	Responders (CR+PR) (N=106)	Non- Responders (SD+PD) (N=96)	AOR (95% CI)^a	p^b
673 C>T				
CC	56 (52.83)	45 (46.87)	(Reference)	
CT	39 (36.79)	41 (42.70)	1.38(0.75-2.53)	0.29
TT	11 (10.37)	10 (10.41)	1.15(0.43-3.03)	0.76
CT+TT	50 (47.16)	51 (53.12)	1.31(0.74-2.32)	0.34
11985 A>G				
AA	60 (56.60)	60 (62.50)	(Reference)	
AG	39 (36.79)	31 (32.29)	0.75(0.41-1.39)	0.37
GG	7 (6.60)	5 (5.20)	0.71(0.20-2.47)	0.59
AG+GG	46 (43.39)	36 (37.50)	0.74(0.41-1.33)	0.32
Arg⁴¹⁵Gln				
GG	59 (55.66)	52 (54.16)	(Reference)	
GA	42 (39.62)	39 (40.62)	1.15(0.63-2.07)	0.63
AA	5 (4.71)	5 (5.20)	1.31(0.32-5.23)	0.70
GA+AA	47 (44.33)	44 (45.83)	1.12(0.63-2.00)	0.68

(a)-Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status, histology, stage, KPS and ECOG. (b)-Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. CR-Complete response; PR-Partial response; SD-Stable disease; PD-Progressive disease

5.6 High order Gene-gene interaction

5.6.1 Distribution of combination of genotypes on different polymorphism and their association with lung cancer risk:

Different genotypic combinations of the various SNP's of the NER pathway genes were evaluated and their association towards lung cancer risk has been elaborated in this study. Firstly we applied combinations as the double combinations. Only significant data has been represented in table 5.52. In this table, group 1 represented wild genotypic combination of two SNPs, group 2 exhibited the heterozygous combination of various polymorphic variants, group 3 represented mutant combinations and group 4 showed both heterozygous+mutant (combined) genotype of two polymorphic variants when applied in combination. The wild type genotype was regarded as the reference for all the combination studies. Adjusted odds ratio and 95%CI was calculated using logistic linear regression. Fdr was further applied to *p*-value in order to avoid false positive results.

As depicted in table 5.53, *XPC Lys⁹³⁹Gln* and *XPB Lys⁷⁵¹Gln* polymorphisms when applied in combination suggested that the overall reference genotype (Group 1) was more presented in controls than in cases (60.27% vs 47.36%), whereas the patients with homozygous variant genotype (group 3) was more illustrated in cases as compared to controls (47.36% vs 38.35%). The heterozygous genotype (Group 2) depicted a higher representation in cases than controls (47.36% vs 38.35%). The heterozygous genotype (Group 2) showed a 2-fold risk for lung cancer when compared with the reference (OR: 1.72; 95%CI: 1.04-2.86; $p=0.03$; $Fdr\ p=0.10$). When heterozygous and the mutant genotype (group 4) were combined together as a single genotype, a 2-fold lung cancer risk was observed when compared with the reference genotype (OR: 1.85; 95%CI: 1.12-3.03; $p=0.01$; $Fdr\ p=0.02$).

XPA G709A and *XPB Lys⁷⁵¹Gln* polymorphisms when evaluated in combination showed a significant protective effect among all the three genotypic combinations. The wild type genotype was more prevalent in cases as compared to controls (65.82% vs 58.27%), whereas 40.39% of the control subjects harbored heterozygous genotype (Group 2) when compared to cases which consisted of only 28.48% of heterozygous genotype. 5.69% of lung cancer cases showed mutant genotype (Group 3) while only 1.32% controls were mutant. Group 2 and Group 4 genotypes both depicted a strong protective effect (OR: 0.60; 95%CI: 0.36-0.98; $p=0.04$; $Fdr\ p=0.12$ and OR: 0.35; 95%CI: 0.22-0.55; $p<0.0001$; $Fdr\ p=0.001$). The mutant genotype (group 3) also showed a significant protective effect when compared with the reference (OR: 0.11; 95%CI: 0.05-0.24; $p<0.0001$; $Fdr\ p=0.001$) as illustrated in table 5.53.

In case of combination of polymorphic variants *XPG Phe⁶⁷⁰Leu* and *XPB Lys⁷⁵¹Gln*, heterozygous genotype (Group 2) were less expressed in cases as compared to controls (15.16% vs 33.33%) where as the wild genotypes was more represented in cases when compared with the controls (83.70% vs 64.32%). It can be stated that patients with single variant allele for both the SNPs (group 2) showed a significant protective effect towards lung cancer (OR: 0.38; 95%CI: 0.22-0.65; $p=0.0005$; $Fdr\ p=0.004$). The combination of both heterozygous and mutant genotype as a single genotype (Group 4) suggested their association towards reduced risk of lung cancer (OR: 0.35; 95%CI: 0.22-0.65; $p=0.0004$; $Fdr\ p=0.002$).

When the association of *XPA* variant *i.e.* *G709A* was studied along with *XPB Arg¹⁵⁶Arg*, it was observed that the heterozygous genotype was more presented in the controls when compared

with the cases (52.55% vs 34.13%) whereas the wild genotype was represented more in cases than controls (62.27% vs 45.98%). After applying logistic regression analysis, it was seen that subjects who were carrying single variant allele for both the SNPs i.e. heterozygous (OR: 0.47; 95%CI: 0.57-1.43; $p=0.002$; $Fdr\ p=0.01$) has a protective effect when compared with wild type genotype as shown in table 5.53. We also observed a protective effect for combined genotype (OR: 0.51; 95%CI: 0.31-0.82; $p=0.005$; $Fdr\ p=0.01$) when compared with reference. On applying on logistic regression on *XPD Arg¹⁵⁶Arg* and *XPG Phe⁶⁷⁰Leu* combination, a significant protective effect was observed for lung cancer in patients carrying heterozygous genotype (group 2) in comparison with homozygous reference genotype (group 1) (OR: 0.32; 95%CI: 0.19-0.54; $p<0.0001$; $Fdr\ p=0.001$).

When genotypes of both *XPC Lys⁹³⁹Gln* and *XPA G709A* polymorphism were analyzed in combination, it was observed that wild type genotype (group 1) had more representation in cases when compared to the controls (55.31% vs 51.26%). Heterozygous genotype (group 2) was more prevalent in controls than cases (46.20% vs 40.42%). Group 3 (mutant) genotype drafted a strong protective effect for lung cancer when compared with the reference genotype (wild) (OR: 0.07; 95%CI: 0.03-0.19; $p<0.0001$; $Fdr\ p=0.001$). Further, the combination of polymorphic variants of *XPC Ala⁴⁹⁹Val* and *XPA G709A* suggested that patients with heterozygous genotype (group 2) showed a significant protective effect towards lung cancer (OR: 0.43; 95%CI: 0.26-0.69; $p=0.0005$; $Fdr\ p=0.004$).

On applying regression analysis on *XPG Phe⁶⁷⁰Leu* and *XPA G709A* combination it was observed that the mutant genotype (group 3) was equally present in both cases and controls (1.01%). The homozygous variant allele (Group 3) suggested its association with reduced risk towards lung cancer for the above polymorphic combination (OR: 0.21; 95%CI: 0.12-0.36; $p<0.0001$; $Fdr\ p=0.001$) as shown in table 5.52. While the individuals with heterozygous genotype (group 2) and represented *XPG Phe⁶⁷⁰Leu* and *XPC Ala⁴⁹⁹Leu* polymorphic combination also emphasized on the decreased risk for the disease (OR: 0.24; 95%CI: 0.14-0.41; $p<0.0001$; $Fdr\ p=0.001$).

In Table 5.52, for *XPC Lys⁹³⁹Gln* and *XPA A23G* polymorphic combination, the individuals with heterozygous genotype (group 2) showed increased risk for the disease (OR: 1.70; 95%CI: 1.06-2.82; $p=0.02$; $Fdr\ p=0.08$) when compared with wild type genotype (group 1). While the

heterozygous genotype (group 2) for *XPA* A23G and *XPG* *Phe*⁶⁷⁰*Leu* polymorphism showed a significant protective effect against the disease (OR: 0.40; 95%CI: 0.23-0.69; $p=0.001$; *Fdr* $p=0.008$). On evaluating *XPF* 673C>T and *XPC* *Lys*⁹³⁹*Gln* in combination, the homozygous variant carriers (group 3) of this polymorphism was more prevalent in controls than cases (62.06% vs 45.98%) whereas the heterozygous genotype (group 2) was more represented in cases as compared to controls (46.71% vs 33.10%). A 2-fold of relative risk of lung cancer was associated with heterozygous genotype (Group 2) (OR: 2.00; 95%CI: 1.20-3.34; $p=0.007$; *Fdr* $p=0.03$) as well as for combined genotype (Group 4) (OR: 2.03; 95%CI: 1.24-3.32; $p=0.004$; *Fdr* $p=0.01$), when group 1 genotype was considered as a reference as shown in table 5.53.

Further it was observed that when *XPG* *Phe*⁶⁷⁰*Leu* and *XPF* 673C>T polymorphism was applied in combination, it contributed in lower risk of having the disease in those patients who were carrying a heterozygous genotype (group 2) when compared to those patients having a wild type genotype (group 1) (OR: 0.65; 95%CI: 0.35-1.14; $p=0.004$; *Fdr* $p=0.01$). When regression analysis was applied on *XPF* 11985A>G and *XPF* *Ala*⁴⁹⁹*Val* combination, it suggested association of this polymorphism with lower risk for lung cancer in individuals carrying heterozygous genotype (Group 2) (OR: 0.95; 95%CI: 0.60-1.51; $p=0.03$; *Fdr* $p=0.10$) as shown in table 5.52. In case of 11985A>G and *XPG* *Phe*⁶⁷⁰*Leu* polymorphism, the patients carrying mutant genotype (group 3) showed a similar frequency distribution in both cases and controls (3.31% and 2.74%), therefore, suggesting an increased lung cancer risk (OR: 1.08; 95%CI: 0.58-1.13; $p=0.02$; *Fdr* $p=0.08$) among those patients who carry mutant genotype (group 3). A protective association was also observed for group 2 (heterozygous) genotype for *XPF* *Arg*⁴¹⁵*Gln* and *XPD* *Lys*⁷⁵¹*Gln* polymorphic combination (OR: 0.69; 95%CI: 0.43-1.11; $p=0.01$; *Fdr* $p=0.10$) when compared with the homozygous reference genotype (group 1).

When *XPA* G709A polymorphism and *XPF* *Arg*⁴¹⁵*Gln* polymorphism applied in combination, wild genotype (Group 1) was more pronounced in cases as compared to the controls (64.77% vs 48.46%). On further applying regression analysis the heterozygous genotype depicted a strong protective effect towards lung cancer when compared with the reference (OR: 0.48; 95%CI: 0.29-1.23; $p=0.003$; *Fdr* $p=0.01$). The combined genotype (Group 4) also showed a significant protective effect for the disease (OR: 0.36; 95%CI: 0.01-0.10; $p=0.001$; *Fdr* $p=0.01$) as shown in table 5.52. In case of *XPG* *Phe*⁶⁷⁰*Leu* and *XPF* *Arg*⁴¹⁵*Gln* polymorphic combination, 86.88% of

lung cancer patients comprises of wild genotype in cases in comparison to 59.32% in controls. The heterozygous genotype (group 2) was more prominent in controls than cases (40.67% vs 13.11%). The patients carrying single variant allele in combination for the two SNPs (Group 2) showed a protective association (OR: 0.23; 95%CI: 0.13-0.40; $p < 0.0001$; *Fdr p* = **0.001**) when compared with the reference genotype (group 1) as mentioned in table 5.53.

Table 5.53: Distribution of genotypes depicting dual combination and their association with lung cancer risk

	OVERALL LUNG CANCER				
	Controls N (%)	Cases N (%)	AOR ^a (95% CI)	<i>P</i> ^b	<i>Fdr p</i>
<i>XPC Lys⁹³⁹Gln & XPD Lys⁷⁵¹Gln</i>	N = 146	N= 133			
0 (Group 1)	88 (60.27)	63(47.36)	1.00 (Reference)		
1 (Group 2)	56 (38.35)	63 (47.36)	1.72(1.04-2.86)	0.03	0.10
2 (Group 3)	2 (1.36)	7 (5.26)	4.26(0.81-22.26)	0.08	0.33
3 (Group 4)	58 (39.72)	70 (52.63)	1.85(1.12-3.03)	0.01	0.02
<i>XPA G709A & XPD Lys⁷⁵¹Gln</i>	N = 151	N= 158			
0 (Group 1)	88 (58.27)	104 (65.82)	1.00 (Reference)		
1 (Group 2)	61 (40.39)	45 (28.48)	0.60(0.36-0.98)	0.04	0.12
2 (Group 3)	2 (1.32)	9 (5.69)	0.11(0.05-0.24)	<0.0001	0.0001
3 (Group 4)	63 (41.72)	54 (34.17)	0.35(0.22-0.55)	<0.0001	0.0001
<i>XPG Phe⁶⁷⁰Leu & XPD Lys⁷⁵¹Gln</i>	N = 171	N= 178			
0 (Group 1)	110 (64.32)	149 (83.70)	1.00 (Reference)		
1 (Group 2)	57 (33.33)	27 (15.16)	0.38(0.22-0.65)	0.0005	0.004
2 (Group 3)	4 (2.33)	2 (1.12)	0.50(0.08-3.01)	0.45	0.61
3 (Group 4)	61 (35.67)	29 (16.29)	0.38(0.22-0.65)	0.0004	0.002

<i>XPD Arg¹⁵⁶Arg & XPA G709A</i>	N = 137	N= 167			
0 (Group 1)	63 (45.98)	104 (62.27)	1.00 (Reference)		
1 (Group 2)	72 (52.55)	57 (34.13)	0.47(0.29-0.76)	0.002	0.01
2 (Group 3)	2 (1.26)	6 (3.59)	1.97(0.37-10.48)	0.42	0.61
3 (Group 4)	74 (54.01)	63 (37.72)	0.51(0.31-0.82)	0.005	0.01
<i>XPD Arg¹⁵⁶Arg & XPG Phe⁶⁷⁰Leu</i>	N = 147	N= 228			
0 (Group 1)	84 (57.14)	144 (63.15)	1.00 (Reference)		
1 (Group 2)	63 (42.85)	84 (36.84)	0.32(0.19-0.54)	<0.0001	0.001
<i>XPC Lys⁹³⁹Gln & XPA G709A</i>	N = 158	N= 141			
0 (Group 1)	81 (51.26)	78 (55.31)	1.00 (Reference)		
1 (Group 2)	73 (46.20)	57 (40.42)	0.82(0.51-1.33)	0.42	0.63
2 (Group 3)	4 (2.53)	6 (4.25)	0.07(0.03-0.19)	<0.0001	0.001
3 (Group 4)	77 (48.73)	63 (44.68)	0.44(0.28-0.68)	0.0003	0.002
<i>XPC Ala⁴⁹⁹Val & XPA G709A</i>	N = 162	N= 156			
0 (Group 1)	67 (38.27)	95 (60.89)	1.00 (Reference)		
1 (Group 2)	95 (58.64)	61 (39.10)	0.43(0.26-0.69)	0.0005	0.004
<i>XPG Phe⁶⁷⁰Leu & XPA G709A</i>	N = 169	N= 197			
0 (Group 1)	101 (59.76)	171 (86.80)	1.00 (Reference)		
1 (Group 2)	66 (39.05)	24 (12.18)	0.39(0.05-3.02)	0.37	0.57
2 (Group 3)	2 (1.01)	2 (1.01)	0.21(0.12-0.36)	<0.0001	0.001
3 (Group 4)	68 (40.23)	26 (13.19)	0.14(0.08-0.25)	<0.0001	0.001
<i>XPG Phe⁶⁷⁰Leu & XPC Ala⁴⁹⁹Val</i>	N = 152	N= 162			
0 (Group 1)	76 (50)	132 (81.48)	1.00 (Reference)		
1 (Group 2)	76 (50)	30 (18.51)	0.24(0.14-0.41)	<0.0001	0.001
<i>XPC Lys⁹³⁹Gln & XPA A23G</i>	N = 151	N= 131			
0 (Group 1)	82 (54.30)	53 (40.45)	1.00 (Reference)		

1 (Group 2)	69 (45.69)	78 (59.54)	1.70(1.06-2.82)	0.02	0.08
<i>XPA A23G & XPG Phe⁶⁷⁰Leu</i>	N = 169	N= 142			
0 (Group 1)	104 (61.53)	113 (79.57)	1.00 (Reference)		
1 (Group 2)	65 (38.46)	29 (20.42)	0.40(0.23-0.69)	0.001	0.008
<i>XPF 673 C>T & XPC Lys⁹³⁹Gln</i>	N = 145	N= 137			
0 (Group 1)	90 (62.06)	63 (45.98)	1.00 (Reference)		
1 (Group 2)	48 (33.10)	64 (46.71)	2.00(1.20-3.34)	0.007	0.03
2 (Group 3)	7 (4.82)	10 (7.29)	2.07(0.70-6.12)	0.18	0.37
3 (Group 4)	55 (37.93)	74 (54.01)	2.03(1.24-3.32)	0.004	0.01
<i>XPF 673 C>T & XPG Phe⁶⁷⁰Leu</i>	N = 137	N= 183			
0 (Group 1)	104 (75.91)	152 (83.60)	1.00 (Reference)		
1 (Group 2)	33 (24.08)	31 (16.93)	0.65(0.35-1.14)	0.004	0.01
<i>XPF 11985 A>G & XPC Ala⁴⁹⁹Val</i>	N = 148	N= 166			
0 (Group 1)	86 (58.10)	99 (59.63)	1.00 (Reference)		
1(Group 2)	62 (41.89)	67 (40.36)	0.95(0.60-1.51)	0.03	0.10
<i>XPF 11985A>G & XPG Phe⁶⁷⁰Leu</i>	N = 182	N= 211			
0 (Group 1)	134 (73.62)	174 (82.46)	1.00 (Reference)		
1 (Group 2)	43 (23.62)	30 (14.21)	0.54(0.32-0.92)	0.02	0.08
2 (Group 3)	5 (2.74)	7 (3.31)	1.08 (0.58-1.13)	0.02	0.08
3 (Group 4)	48 (26.37)	37 (17.53)	0.91(0.26-0.92)	0.01	0.02
<i>XPF Arg⁴¹⁵Gln & XPD Lys⁷⁵¹Gln</i>	N = 159	N= 147			
0 (Group 1)	91 (57.23)	99 (67.34)	1.00 (Reference)		
1 (Group 2)	68 (42.76)	48 (32.65)	0.69 (0.43-1.11)	0.01	0.10
<i>XPF Arg⁴¹⁵Gln & XPA G709A</i>	N = 163	N= 176			
0 (Group 1)	79 (48.46)	114 (64.77)	1.00 (Reference)		

1 (Group 2)	74(45.39)	51 (28.97)	0.48(0.29-0.78)	0.003	0.01
2 (Group 3)	10 (12.65)	11 (6.25)	0.71 (0.56-1.23)	0.12	0.27
3 (Group 4)	84 (51.53)	62 (35.22)	0.36 (0.01-0.10)	0.001	0.01
<i>XPF Arg⁴¹⁵Gln & XPG Phe⁶⁷⁰Leu</i>	N = 177	N= 183			
0 (Group 1)	105 (59.32)	159 (86.88)	1.00 (Reference)		
1(Group 2)	72 (40.67)	24 (13.11)	0.23(0.13-0.40)	<0.0001	0.001

(a)-AdjustedOdds 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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

Abbreviations: Group 1- combination of wild genotypes of any two polymorphism Group 2- heterozygous genotype of two polymorphic variants in combination, Group 3: Mutant genotype in two polymorphic variants Group 4: Heterozygous+mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

5.6.2 Distribution of different combination of genotypes of different NER polymorphism and their association towards lung cancer in relation with histological subtypes:

The combinations were further applied according to the histological subtypes for the above dual combinations as shown in Table 5.54 (a, b, c). Table 5.54(a) showed association of different polymorphic combinations with SQCC, table 5.54(b) represented different polymorphic combination and their relationship towards ADCC risk and table 5.54(c) illustrated correlation of SCLC risk with different polymorphic combination when applied. Only the significant combinatorial values were included in this study.

5.6.2.1 Association of combination genotype on different polymorphism and their association with SQCC:

For *XPC Lys⁹³⁹Gln* and *XPB Lys⁷⁵¹Gln* combination, the subjects who were harboring single variant allele for both the SNPs *i.e.* heterozygous genotypic (group 2) was found to be over presented in cases as compared to controls (51.06% vs 38.35%) whereas wild genotype was more prevalent in controls in comparison with cases (60.27% vs 46.80%). On applying logistic regression analysis, it was observed that heterozygous genotype (group 2) showed almost 2-fold risk for SQCC (OR: 1.83; 95%CI: 0.91-3.69; *p*=**0.04**; *Fdr p*= **0.12**). In case of *G709A* and *XPB*

*Lys*⁷⁵¹*Gln* polymorphic combination, the percentage of the mutant genotype (group 3) was found to be represented more in cases (8.92%) than controls (1.32%) whereas heterozygous genotype (group 2) was more prevalent in controls (40.39%) in comparison with cases (19.64%). The lung cancer patients carrying mutant genotype (group 3). A decreased risk for patients with SQCC and carrying mutant genotype was observed (OR: 0.16; 95%CI: 0.06-0.45; $p=0.005$; *Fdr p*=**0.006**). The heterozygous genotype (group 2) also depicted a decreased risk when compared with reference genotype as shown in table 5.53(a) (OR: 0.34; 95%CI: 0.15-0.76; $p=0.008$; *Fdr p*=**0.04**).

On analyzing *XPG Phe*⁶⁷⁰*Leu* and *XPB Lys*⁷⁵¹*Gln* polymorphic combination, the SQCC cases having reference genotype (group 1) were more expressed than controls (84.28% vs 64.32%). The frequency of heterozygous genotype among controls (33.33%) was higher than cases (14.20%). SQCC patients were associated with the significant protective effect in both Group 2 (heterozygous) (OR: 0.31; 95%CI: 0.14-0.68; $p=0.003$; *Fdr p*=**0.01**) and Group 4 (combined) (OR: 0.32; 95%CI: 0.15-0.69; $p=0.003$; *Fdr p*=**0.01**) genotypes when compared with reference.

*XPB Phe*¹⁵⁶*Leu* in combination with *G709A* polymorphic variant showed over expression of the heterozygous genotype (group 2) in SQCC controls (52.55%) than cases (21.21%) suggesting a significant protective effect (Group 2= OR: 0.31; 95%CI: 0.15-0.64; $p=0.001$; *Fdr p*=**0.007**). SQCC patients having combined genotype also represented a significant effect in comparison with reference genotype (Group 4 = OR: 0.51; 95%CI: 0.31-10.48; $p=0.007$; *Fdr p*=**0.03**).

On evaluating *XPB Arg*¹⁵⁶*Arg* and *XPG C670A* combination, the homozygous wild carriers (group 1) were more prevalent in cases than controls (85.13% vs 45.98%). Heterozygous genotype (group 2) was more frequent in controls as compared to cases (32.64% vs 10.81%). A decreased relative risk of lung cancer was associated with heterozygous genotype (Group 2) (OR: 0.20.; 95%CI: 0.08-0.45; $p=0.0001$; *Fdr p*=**0.001**). The similar results were obtained for SQCC patients with combined genotype (Group 4) (OR: 0.18; 95%CI: 0.08-0.38; $p<0.0001$; *Fdr p*=**0.001**), when group 1 genotype was considered as a reference as shown in table 5.54 (a).

When genotypic variants of *XPC Lys*⁹³⁹*Gln* and *XPA G709A* were applied in combination, wild genotype (group 1) was expressed slightly higher in cases than controls (56.25% vs 51.26%). SQCC patients harboring mutant genotype (group 3) were more prominent in cases as compared

to controls (6.25%vs 2.53%). Heterozygous genotype (group 2) showed protective effect for patients exhibiting double variant allele (OR: 0.11; 95%CI: 0.03-0.40; $p=0.0007$; $Fdr\ p=0.006$). Group 4 (combined) genotype also marked a decreased cancer risk in SQCC patients (OR: 0.41; 95%CI: 0.21-0.79; $p=0.008$; $Fdr\ p=0.03$).

Out of 59 SQCC cases with *XPA G709A* and *XPC Ala⁴⁹⁹Val* polymorphism in combination suggested that the patients with heterozygous genotype (Group 2) showed a minimum risk of having the disease (OR: 0.32; 95%CI: 0.16-0.65; $p=0.001$; $Fdr\ p=0.007$). *XPC Lys⁹³⁹Gln* and *XPA A23G* polymorphism in combination showed a higher representation of combined genotype (group 4) in cases than controls (55% vs 37.93%). The patients with combined genotype (Group 4) showed a 2-fold risk of having SQCC (OR: 2.76; 95%CI: 1.45-5.27; $p=0.002$; $Fdr\ p=0.01$).

SQCC patients consisted of 45% wild (group 1), 47.5% heterozygous (group 2) and 7.5% mutant genotype (group 3) for *XPF 673C>T* and *XPC Lys⁹³⁹Gln* combination. SQCC patients with heterozygous genotype (Group 2) exhibited a 2-fold risk of having the disease when compared with the reference genotype (group 1) (OR: 2.33; 95%CI: 1.08-5.03; $p=0.03$; $Fdr\ p=0.11$). The combined genotype (Group 4) further illustrated risk for squamous cell carcinoma among patients (OR: 2.50; 95%CI: 1.19-5.25; $p=0.01$; $Fdr\ p=0.03$). While *XPF 11985A>G* in combination with *XPG Phe⁶⁷⁰Leu* harbored a protective effect in SQCC patients carrying heterozygous genotype (Group 2) (OR: 0.39; 95%CI: 0.17-0.87; $p=0.02$; $Fdr\ p=0.08$).

Table 5.54 (a): Distribution of genotypes depicting dual combination and their association with SQCC patients					
<i>XPC Lys⁹³⁹Gln</i> & <i>XPB Lys⁷⁵¹Gln</i>	Controls n (%) N = 146	Cases n(%) N= 47	AOR (95% CI) ^a	<i>P^b</i>	<i>Fdr p</i>
0 (Group 1)	88 (60.27)	22 (46.80)	1.00 (Reference)		
1 (Group 2)	56 (38.35)	24 (51.06)	1.83(0.91-3.69)	0.04	0.12
2 (Group 3)	2 (1.36)	1 (2.12)	3.02(0.20-44.95)	0.42	0.72
3(Group 4)	58 (39.72)	25 (53.19)	1.87(0.94-3.73)	0.07	0.15
<i>XPA G709A</i> & <i>XPB Lys⁷⁵¹Gln</i>	N = 151	N =56			
0 (Group 1)	88 (58.27)	40 (71.42)	1.00 (Reference)		
1 (Group 2)	61 (40.39)	11 (19.64)	0.34(0.15-0.76)	0.008	0.04
2 (Group 3)	2 (1.32)	5 (8.92)	0.16(0.06-0.45)	0.0005	0.006
3 (Group 4)	63 (41.72)	16 (28.51)	0.25(0.13-0.50)	0.0001	0.001
<i>XPG Phe⁶⁷⁰Leu</i> & <i>XPB Lys⁷⁵¹Gln</i>	N = 171	N = 70			
0 (Group 1)	110(64.32)	59 (84.28)	1.00 (Reference)		
1(Group 2)	57 (33.33)	10 (14.20)	0.31(0.14-0.68)	0.003	0.01

2 (Group 3)	4 (2.33)	1 (1.4)	0.55(0.05-5.86)	0.62	0.93
3(Group 4)	61 (35.67)	11(15.41)	0.32(0.15-0.69)	0.003	0.01
XPD Arg¹⁵⁶Arg & XPA G709A	N = 137	N = 66			
0 (Group 1)	63 (45.98)	48 (72.72)	1.00 (Reference)		
1(Group 2)	72 (52.55)	14 (21.21)	0.31(0.15-0.64)	0.001	0.007
2 (Group 3)	2 (1.26)	4 (6.06)	2.98(0.47-18.76)	0.24	0.62
3(Group 4)	74 (54.01)	18 (27.27)	0.39(0.20-0.77)	0.007	0.03
XPD Arg¹⁵⁶Arg & XPG Phe⁶⁷⁰Leu	N = 154	N = 73			
0 (Group 1)	84 (54.54)	63 (85.13)	1.00 (Reference)		
1(Group 2)	63 (32.64)	8 (10.81)	0.20(0.08-0.45)	0.0001	0.001
2(Group 3)	7 (3.62)	2 (2.17)	0.37 (0.06-0.42)	0.41	0.70
3 (Group 4)	70 (45.45)	10 (13.69)	0.18(0.08-0.38)	<0.0001	0.001
XPC Lys⁹³⁹Gln & XPA G709A	N = 158	N = 48			
0 (Group 1)	81 (51.26)	27 (56.25)	1.00 (Reference)		
1(Group 2)	73 (46.20)	18 (37.5)	0.72(0.35-1.45)	0.36	0.56
2(Group 3)	4 (2.53)	3 (6.25)	0.11(0.03-0.40)	0.0007	0.006
3(Group 4)	77 (48.73)	21 (43.75)	0.41(0.21-0.79)	0.008	0.03
XPC Ala⁴⁹⁹Val & XPA G709A	N = 162	N = 59			
0 (Group 1)	67 (41.35)	38 (64.40)	1.00 (Reference)		
1(Group 2)	95 (58.64)	21 (35.59)	0.32(0.16-0.65)	0.001	0.007
XPC Lys⁹³⁹Gln & XPA A23G	N = 153	N = 49			
0 (Group 1)	82 (53.59)	16 (32.65)	1.00 (Reference)		
1 (Group 2)	69 (45.09)	29 (59.18)	1.98(0.97-4.05)	0.05	0.15
2 (Group 3)	2 (1.30)	4 (8.16)	0.26(0.08-0.84)	0.02	0.12
3 (Group 4)	71 (46.40)	33 (67.34)	2.76(1.45-5.27)	0.002	0.01
XPF 673C>T & XPC Lys⁹³⁹Gln	N = 145	N = 40			
0 (Group 1)	90 (62.06)	18 (45)	1.00 (Reference)		
1 (Group 2)	48 (33.10)	19 (47.5)	2.33(1.08-5.03)	0.03	0.11
2 (Group 3)	7 (4.82)	3 (7.5)	3.25(0.63-16.73)	0.15	0.47
3(Group 4)	55 (37.93)	22 (55)	2.50(1.19-5.25)	0.01	0.03
XPF 11985A>G & XPG Phe⁶⁷⁰Leu	N = 177	N = 79			
0 (Group 1)	134 (75.7)	70 (88.60)	1.00 (Reference)		
1(Group 2)	43 (24.29)	9 (11.39)	0.39(0.17-0.87)	0.02	0.08

(a)-AdjustedOdds 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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

Abbreviations: Group 1- combination of wild genotypes of any two polymorphism Group 2- heterozygous genotype of two polymorphic variants in combination, Group 3: Mutant genotype in two polymorphic variants Group 4: Heterozygous+mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

Table 5.54 (b): Distribution of genotypes depicting dual combination and their association with ADCC patients					
<i>XPC Lys⁹³⁹Gln & XPD Lys⁷⁵¹Gln</i>	Controls n (%) N = 146	Cases n(%) N= 45	AOR (95% CI) ^a	<i>p</i> ^b	<i>Fdr p</i>
0 (Group 1)	88 (60.27)	17 (37.71)	1.0(Reference)		
1 (Group 2)	56 (38.35)	22 (48.88)	1.84(0.85-3.94)	0.11	0.41
2 (Group 3)	2 (1.36)	6 (13.33)	11.29(2.01-63.39)	0.005	0.02
3(Group 4)	58 (39.72)	28 (62.22)	2.34(1.13-4.84)	0.02	0.07
<i>XPA G709A & XPD Lys⁷⁵¹Gln</i>	N = 151	N = 56			
0 (Group 1)	88 (58.27)	33 (58.92)	1.00 (Reference)		
1 (Group 2)	61 (40.39)	22 (39.28)	0.88(0.45-1.71)	0.72	0.88
2 (Group 3)	2 (1.32)	1 (1.78)	0.04(0.005-0.31)	0.002	0.02
3 (Group 4)	63 (41.72)	23 (41.07)	0.47(0.25-0.88)	0.01	0.05
<i>XPG Phe⁶⁷⁰Leu & XPD Lys⁷⁵¹Gln</i>	N = 167	N = 62			
0 (Group 1)	110 (64.32)	54 (87.09)	1.00 (Reference)		
1(Group 2)	57 (33.33)	8 (12.90)	0.34(0.15-0.79)	0.01	0.08
<i>XPD Arg¹⁵⁶Arg & XPG Phe⁶⁷⁰Leu</i>	N = 147	N = 46			
0 (Group 1)	84 (57.14)	39 (84.78)	1.00 (Reference)		
1(Group 2)	63 (42.86)	7 (15.21)	0.24(0.10-0.58)	0.001	0.01
<i>XPC Lys⁹³⁹Gln & XPA G709A</i>	N = 158	N = 43			
0 (Group 1)	81 (51.26)	19 (44.18)	1.00 (Reference)		
1(Group 2)	73 (46.20)	22 (51.16)	1.22(0.60-2.50)	0.57	0.79
2(Group 3)	4 (2.53)	2 (4.65)	0.10(0.02-0.45)	0.003	0.02
3(Group 4)	77 (48.73)	24 (55.81)	0.64(0.32-1.27)	0.20	0.41
<i>XPC Ala⁴⁹⁹Val & XPA G709A</i>	N = 162	N = 48			
0 (Group 1)	67 (41.35)	28 (58.33)	1.00 (Reference)		
1(Group 2)	95 (58.64)	20 (41.66)	0.48(0.23-0.97)	0.04	0.23
<i>XPF 673 C>T & XPC Lys⁹³⁹Gln</i>	N = 145	N = 49			
0 (Group 1)	90 (62.06)	20 (40.81)	1.00 (Reference)		
1 (Group 2)	48 (33.10)	25 (51.02)	2.26(1.01-4.63)	0.02	0.07
2 (Group 3)	7 (4.82)	4 (8.16)	1.79(0.41-7.74)	0.43	0.68
3(Group 4)	55 (37.93)	29 (59.18)	2.23(1.11-4.47)	0.02	0.07

(a)-AdjustedOdds 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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

Abbreviations: Group 1- combination of wild genotypes of any two polymorphism Group 2- heterozygous genotype of two polymorphic variants in combination, Group 3: Mutant genotype in two polymorphic variants Group 4: Heterozygous+mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

Table 5.54 (c): Distribution of genotypes depicting dual combination and their association with SCLC patients					
<i>XPA G709A & XPD Lys⁷⁵¹Gln</i>	N = 151	N = 43	AOR (95% CI) ^a	<i>p</i> ^b	<i>Fdr p</i>
0 (Group 1)	88 (58.27)	28 (65.11)	1.00 (Reference)		
1 (Group 2)	61 (40.39)	12 (27.90)	0.65(0.30-1.42)	0.28	0.64
2 (Group 3)	2 (1.32)	3 (6.97)	0.14(0.04-0.52)	0.003	0.01
3 (Group 4)	63 (41.72)	15 (34.8)	0.39(0.19-0.80)	0.01	0.07
<i>XPG Phe⁶⁷⁰Leu & XPD Lys⁷⁵¹Gln</i>	N = 171	N = 46			
0 (Group 1)	110 (64.32)	36 (78.26)	1.00 (Reference)		
1(Group 2)	57 (33.33)	9 (19.56)	0.56(0.24-1.31)	0.003	0.01
2 (Group 3)	4 (2.33)	1 (2.17)	1.23(0.11-13.03)	0.86	0.97
3(Group 4)	61 (35.67)	10 (21.73)	0.60(0.26-1.34)	0.003	0.01
<i>XPC Lys⁹³⁹Gln & XPA G709A</i>	N = 158	N = 49			
0 (Group 1)	81 (51.26)	31 (63.26)	1.00 (Reference)		
1(Group 2)	73 (46.20)	17 (34.69)	0.67(0.32-1.40)	0.29	0.64
2(Group 3)	4 (2.53)	1 (2.04)	0.04(0.006-0.34)	0.002	0.01
3(Group 4)	77 (48.73)	18 (36.73)	0.35(0.17-0.71)	0.003	0.03
<i>XPC Lys⁹³⁹Gln & XPA A23G</i>	N = 153	N = 41			
0 (Group 1)	82 (53.59)	18 (43.91)	1.00 (Reference)		
1 (Group 2)	69 (45.09)	19 (46.34)	1.57(0.72-0.94)	0.25	0.64
2 (Group 3)	2 (1.30)	4 (9.75)	0.33(0.10-1.07)	0.06	0.17
3 (Group 4)	71 (46.40)	23 (56.09)	2.01(1.02-3.99)	0.04	0.21

(a)-AdjustedOdds 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 χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

Abbreviations: Group 1- combination of wild genotypes of any two polymorphism Group 2- heterozygous genotype of two polymorphic variants in combination, Group 3: Mutant genotype in two polymorphic variants Group 4: Heterozygous+mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

5.6.2.2 Association of combination genotype on different polymorphism and their association with ADCC:

For *XPC Lys⁹³⁹Gln* and *XPD Lys⁷⁵¹Gln* combination, the heterozygous genotype (group 2) was more presented in cases as compared to controls (48.88% vs 38.35%) whereas wild genotype was more prevalent in controls in comparison with cases (60.27% vs 37.71%). The percentage of mutant genotype (group 3) was higher in cases than controls (13.33% vs 1.36%). On applying logistic regression analysis, mutant genotype (group 3) showed almost 11-fold risk for lung cancer (OR: 11.29; 95%CI: 2.01-63.39; *p*=0.005; *Fdr p* = **0.02**). Among combined genotype for

this polymorphism, 2-fold risk for ADCC was observed (OR: 2.34; 95%CI: 1.13-4.84; $p=0.02$; $Fdr\ p=0.07$) as shown in table 5.54(b).

The percentage of the mutant genotype (group 3) was similar in cases (1.78%) than controls (1.32%) for *XPA G709A* and *XPD Lys⁷⁵¹Gln* polymorphic combination. The wild genotype (group 1) was also similarly expressed in controls (58.27%) and cases (58.92%) as shown in table 5.53(b). A decreased risk for patients with ADCC and carrying mutant genotype was observed for *G709A* and *Lys⁷⁵¹Gln* combination (OR: 0.04; 95%CI: 0.005-0.31; $p=0.002$; $Fdr\ p=0.02$). The combined genotype (group 2) also depicted a decreased risk when compared with reference genotype as shown in table 5.53(b) (OR: 0.47; 95%CI: 0.25-0.88; $p=0.01$; $Fdr\ p=0.05$).

On analyzing *XPG Phe⁶⁷⁰Leu* and *XPD Lys⁷⁵¹Gln* polymorphic combination, the ADCC subjects having reference genotype (group 1) were more expressed in cases than controls (87.09% vs 64.32%). The frequency of heterozygous genotype among controls (33.33%) was higher than cases (12.90%). Regression analysis suggested a significant protective effect in heterozygous genotype (group 2) when compared with reference (group 1) (OR: 0.34; 95%CI: 0.15-0.79; $p=0.01$; $Fdr\ p=0.08$). When *XPC Arg¹⁵⁶Arg* and *XPG Phe⁶⁷⁰Leu* polymorphism applied in combination showed a higher representation of heterozygous genotype (group 2) in controls than cases (42.86% vs 15.21%). The patients depicting heterozygous genotype (Group 4) showed a decreased risk of having ADCC as shown in table 5.53(b) (OR: 0.24; 95%CI: 0.10-0.58; $p=0.001$; $Fdr\ p=0.01$).

XPD Lys⁹³⁹Gln combination with *XPA G709A* polymorphic variant showed over expression of the mutant genotype (group 3) in ADCC cases (4.65%) than controls (2.53%) suggesting a significant protective effect (OR: 0.10; 95%CI: 0.02-0.45; $p=0.003$; $Fdr\ p=0.02$). ADCC patients having heterozygous genotype (group 2) showing *XPC Ala⁴⁹⁹Val* and *XPA G709A* polymorphism also represented a significant effect in comparison with reference genotype (OR: 0.48; 95%CI: 0.23-0.97; $p=0.04$; $Fdr\ p=0.23$).

When genotypic variants of *XPF 673C>T* and *XPC Lys⁹³⁹Gln* were applied in combination, wild genotype (group 1) was more expressed in controls than cases (62.06% vs 40.81%). SQCC patients harboring heterozygous genotype (group 3) were more prominent in controls as compared

to cases (33.10%vs 51.02%). Heterozygous genotype (group 2) showed increased risk for ADCC patients (OR: 2.26; 95%CI: 1.01-4.63; $p=0.02$; $Fdr\ p=0.07$). Group 4 (combined) genotype also marked a 2-fold risk in ADCC patients (OR: 2.23; 95%CI: 1.11-4.47; $p=0.02$; $Fdr\ p=0.07$).

5.6.2.3 Association of combination genotype on different polymorphism and their association with SCLC:

Few combinations have shown a significant association with SCLC susceptibility as mentioned in table 5.53(c). First combination was between *XPA G709A* and *XPB Lys⁷⁵¹Gln*, the percentage of the mutant genotype (group 3) was higher in cases (6.97%) than controls (1.32%). The wild genotype (group 1) was also less expressed in controls (58.27%) in comparison with cases (65.11%) as shown in table 5.54(c). A decreased risk for patients with SCLC and carrying mutant genotype was observed (OR: 0.14; 95%CI: 0.04-0.52; $p=0.003$; $Fdr\ p=0.01$). The combined genotype (group 4) also depicted a decreased risk when compared with reference genotype as shown in table 5.54(c) (OR: 0.39; 95%CI: 0.19-0.80; $p=0.01$; $Fdr\ p=0.07$).

On analyzing *XPG Phe⁶⁷⁰Leu* and *XPB Lys⁷⁵¹Gln* polymorphic combination, the SCLC subjects having reference genotype (group 1) were more expressed in cases than controls (78.26% vs 64.32%). The frequency of heterozygous genotype among controls (33.33%) was higher than cases (19.56%). Regression analysis suggested a significant protective effect in heterozygous genotype (group 2) when compared with reference (group 1) (OR: 0.56; 95%CI: 0.24-1.31; $p=0.003$; $Fdr\ p=0.01$). The patients depicting combined genotype (Group 4) showed a decreased risk of having SCLC as shown in table 5.54(c) (OR: 0.60; 95%CI: 0.26-1.34; $p=0.003$; $Fdr\ p=0.01$).

When genotypic variants of *XPC Lys⁹³⁹Gln* and *G709A* were applied in combination, wild genotype (group 1) was expressed slightly higher in cases than controls (63.26% vs 51.26%). SCLC patients harboring mutant genotype (group 3) were equally represented in cases (2.04%) and controls (2.53%). The combined genotype (group 4) was more prominent among controls than cases (48.76% vs 36.73%). SCLC patients showed a protective effects in subjects exhibiting double variant allele (OR: 0.04; 95%CI: 0.006-0.34; $p=0.002$; $Fdr\ p=0.01$). Group 4 (combined) genotype also marked a decreased cancer risk in SCLC patients when compared with the reference (OR: 0.35; 95%CI: 0.17-0.71; $p=0.003$; $Fdr\ p=0.03$).

XPC Lys⁹³⁹Gln and *XPA A23G* polymorphism in combination constitutes 41 SCLC cases, out of which 43.91% were wild, 46.34% were heterozygous and 9.75% showed mutant genotype respectively. Out of 153 controls, 53.59% were wild, 45.09% had heterozygous and 1.30% were mutant genotypes. The individuals with mutant genotype (group 3) illustrated a decreased risk for the disease (OR: 0.33; 95%CI: 0.10-1.07; $p=0.06$; *Fdr p*=**0.17**). The patients with combined genotype (Group 4) genotype with the above polymorphic combination asserted with increased lung cancer risk when compared with reference as shown in table 5.54(c) (OR: 2.01; 95%CI: 1.02-3.39; $p=0.04$; *Fdr p*=**0.21**).

5.6.3 MDR analysis and lung cancer risk

The data mining tool, MDR was used to identify which combination of factor was the best model to predict high risk subgroups of the subjects (Table 5.55). It consists of CVC, prediction error and p-value obtained from the MDR software for all the number of factors were evaluated (N=1 to 12). *XPG Phe⁶⁷⁰Leu* was the best one factor model with CVC of 100% and prediction error of 0.37 among all the twelve factors ($p=0.0001$). *XPC Lys⁹³⁹Gln*, *XPG Phe⁶⁷⁰Leu* depicted the best two factor model having CVC of 100% and prediction error of 0.37. While *XPC Lys⁹³⁹Gln*, *XPA A23G*, *XPG Phe⁶⁷⁰Leu*, *XPG Asp¹¹⁰⁴His* illustrated prediction error of 0.35 among all the 12 factors with p-value of <0.0001 . The interaction dendrogram was used to define the visualized interaction between the SNPs and lung cancer (Fig 5.26A). Colored lines describe the type of interaction between the interacting SNPs. Red and orange lines indicate the synergy between the

Table 5.55 : Multifactor Dimensionality reduction (MDR) analysis showing interactions of NER variants with lung cancer risk

No. of risk factors	Best interaction model	Cross Validation Consistency (CVC)	Prediction error	<i>p</i> – value
1	<i>XPG Phe⁶⁷⁰Leu</i>	100/100	0.37	0.0001
2	<i>XPC Lys⁹³⁹Gln</i> , <i>XPG Phe⁶⁷⁰Leu</i>	100/100	0.37	0.0001
3	<i>XPC Lys⁹³⁹Gln</i> , <i>XPA A23G</i> , <i>XPG Phe⁶⁷⁰Leu</i>	57/100	0.41	<0.0001
4	<i>XPC Lys⁹³⁹Gln</i> , <i>XPA A23G</i> , <i>XPG Phe⁶⁷⁰Leu</i> , <i>XPG Asp¹¹⁰⁴His</i>	100/100	0.35	<0.0001

SNPs. Yellow suggests independence. Green and blue indicate redundancy or no interaction. The shorter the lines connecting the two SNPs stronger the interaction. In fig 5.24A, *XPA A23G* and *XPG Phe⁶⁷⁰Leu* and *XPA Lys⁹³⁹Gln* polymorphic variants were connected with each other through blue lines suggested no interaction between them. *XPG Asp¹¹⁰⁴His* was further connected with all the three factor through a green lines suggested no interaction among them.

5.6.4 MDR and association with histological subtypes of lung cancer

MDR was further applied to understand association of the NER genes with histological subtypes of lung cancer as represented in Table 5.56, 5.57 and 5.58. *XPG Phe⁶⁷⁰Leu* was detected as the best single factor model in all the three histological subtypes (ADCC, SQCC and SCLC) subjected to 100 percent CVC depicting 0.37 prediction error ($p < 0.0001$). For ADCC patients, *XPG Phe⁶⁷⁰Leu*, *XPA A23G*, *XPB Lys⁷⁵¹Gln* represented best three factor model illustrating a prediction error of 0.37 with 100 CVC ($p < 0.0001$). *XPG Phe⁶⁷⁰Leu*, *XPA A23G* showed a two factor model in case of SQCC patients indicating prediction error of 0.42 with 100 CVC. A four factor model was also observed in case of SQCC subtype i.e. *XPG Phe⁶⁷⁰Leu*, *XPA G709A*, *XPB 11985 A>G*, *XPB Lys⁷⁵¹Gln* with p-value of **0.008**. The CVC and prediction error for this model was 100 and 0.43. *XPG Phe⁶⁷⁰Leu*, *XPA A23G*, *XPB Lys⁷⁵¹Gln*, *XPA G709A*, *XPB Arg⁴¹⁵Gln*, *XPB Arg¹⁵⁶Gln* was depicted as the best six factor model in case of SCLC subtype with minimum prediction error of 0.35 with p-value **0.005** (CVC=100).

The association of the NER genes among histological subtypes was further explained with the help of the dendrograms (Fig 5.26B-D). Colored lines describe the type of interaction between the interacting SNPs. Red and orange lines indicate the synergy between the SNPs. Yellow suggests independence. Green and blue indicate redundancy or no interaction. The shorter the line connecting two SNPs stronger will be the interaction. In fig 5.26B, *XPA A23G* and *XPG Phe⁶⁷⁰Leu* are connected with each other through the shortest line indicating maximum synergy between them. They are further connected with *XPB Lys⁷⁵¹Gln* through a yellow line suggesting a distant association with the two SNPs. In fig 5.26C, it was observed that *XPG Phe⁶⁷⁰Leu*, *XPB Lys⁷⁵¹Gln* and *XPA G709A* were connected through blue lines of which *Phe⁶⁷⁰Leu* and *Lys⁷⁵¹Gln*

are connected with the shortest possible line but did not show any association with each other. A yellow line connects 11985A>G polymorphism with other three suggesting it as the independent factor. Lastly, Fig 5.26D shows a 6-factor model depicting SCLC risk, *XPD Lys⁷⁵¹Gln* and *XPA A23G* showed synergistic effect with each other whereas *XPF Arg⁴¹⁵Gln* polymorphism was independent of other factor.

Table 5.56: Multifactor Dimensionality reduction (MDR) analysis showing gene interaction in NER variants with ADCC				
No. of risk factors	Best interaction model	Cross Validation Consistency (CVC)	Prediction error	p – value
1	<i>XPG Phe⁶⁷⁰Leu</i>	100/100	0.37	<0.0001
3	<i>XPG Phe⁶⁷⁰Leu, XPA A23G, XPD Lys⁷⁵¹Gln</i>	100/100	0.37	<0.0001
Table 5.57: Multifactor Dimensionality reduction (MDR) analysis showing gene interaction in NER variants with SQCC				
No. of risk factors	Best interaction model	Cross Validation Consistency (CVC)	Prediction error	p – value
1	<i>XPG Phe⁶⁷⁰Leu</i>	100/100	0.42	0.004
2	<i>XPG Phe⁶⁷⁰Leu, XPA A23G</i>	100/100	0.42	0.002
4	<i>XPG Phe⁶⁷⁰Leu, XPA G709A, XPF 11985 A>G, XPD Lys⁷⁵¹Gln</i>	100/100	0.43	0.008
Table 5.58: Multifactor Dimensionality reduction (MDR) analysis showing gene interaction in NER variants with SCLC				
No. of risk factors	Best interaction model	Cross Validation Consistency (CVC)	Prediction error	p – value
1	<i>XPG Phe⁶⁷⁰Leu</i>	100/100	0.43	0.002
2	<i>XPG Phe⁶⁷⁰Leu, XPA A23G</i>	100/100	0.42	0.007
3	<i>XPG Phe⁶⁷⁰Leu, XPA A23G, XPF Arg⁴¹⁵Gln</i>	100/100	0.39	0.04
6	<i>XPG Phe⁶⁷⁰Leu, XPA A23G, XPD Lys⁷⁵¹Gln, XPA G709A, XPF Arg⁴¹⁵Gln, XPD Arg¹⁵⁶Arg</i>	100/100	0.35	0.005

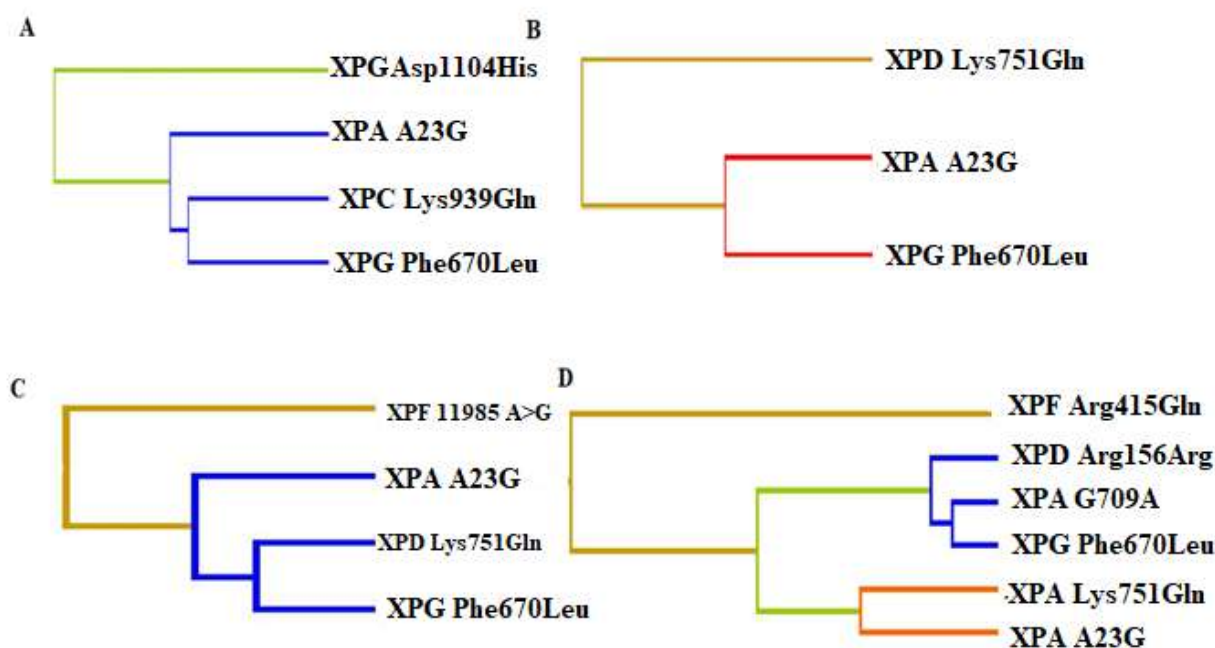


Fig 5.26: MDR depicting the association between the all the NER genes involved in the present study with lung cancer risk (A) a 4-factor model illustrating association with lung cancer risk (B) a 3-factor model depicting ADCC risk (C) a 4-factor model representing SQCC risk (D) a 6-factor model showing SCLC risk

5.6.5 MDR analysis and role of gene-environment interaction in lung cancer risk

We further applied MDR analysis depicting gene-environment interaction in NER variants. Table 5.59 summarizes the CVC and prediction error obtained through MDR for all the number of factors evaluated (N=1 to 13). Smoking status was best one factor model with CVC of 100% and low prediction error of 0.37 among all the 13 factors (Table 5.59; Fig 5.27). The prediction error was statistically significant ($p < 0.001$). The three factor model is the best model including *XPC Ala⁴⁹⁹Val*, *Phe⁶⁷⁰Leu* and smoking with minimum prediction error of 37% and CVC of 100 ($p < 0.001$). This can be very well explained by interaction dendrogram. The minimum distance between *Phe⁶⁷⁰Leu* and smoking was observed but as they are connected through a blue line this

explains a higher degree of redundancy between the two factors. *Ala*⁴⁹⁹*Val* gene did not show any interaction between *Phe*⁶⁷⁰*Leu* and smoking.

Table 5.59: Multifactor Dimensionality reduction (MDR) analysis showing gene-environment interaction in NER variants with lung cancer risk				
No. of risk factors	Best interaction model	Cross Validation Consistency (CVC)	Prediction error	p – value
1	Smoking	100/100	0.37	<0.0001
2	<i>XPG Phe</i> ⁶⁷⁰ <i>Leu</i> , Smoking	100/100	0.37	<0.0001
3	<i>XPC Ala</i> ⁴⁹⁹ <i>Val</i> , <i>XPG Phe</i> ⁶⁷⁰ <i>Leu</i> , Smoking	99/100	0.42	<0.0001

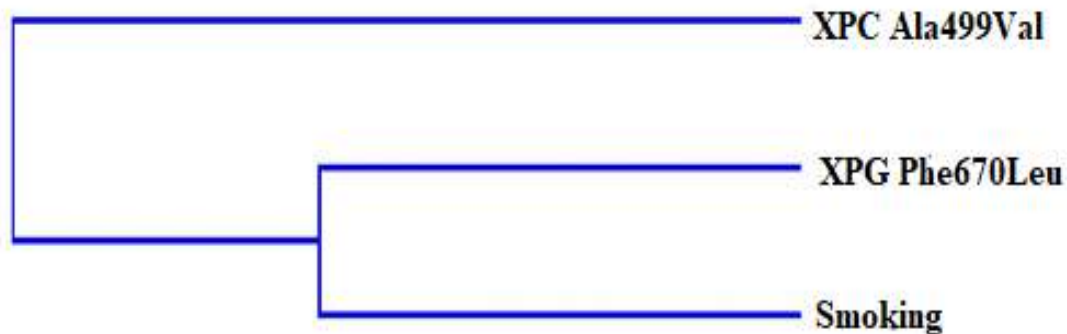


Fig 5.27: MDR graph representing association of gene-environmental factor in NER SNPs

5.6.6 CART analysis

To understand high order gene-gene interactions and their association with lung cancer, we included NER genes in the present study with their polymorphic variants for analysis (Table 5.60). CART is a binary partitioning method. The only split in the decision tree was present on

XPG Phe⁶⁷⁰Leu suggesting it to be the only potential risk factor for the lung cancer risk as compared to other polymorphism. The terminal node with minimum case rate was considered as reference. The terminal node 1 with *XPG Phe⁶⁷⁰Leu* (W) showed almost 3-fold lung cancer risk (OR: 3.23; 95%CI: 2.32-4.50; p=**0.001**).

We further divided population on the basis of histological subtypes. For exploring association of NER variants in combination with ADCC (Table 5.61), the only split in the decision tree was observed on *XPG Phe⁶⁷⁰Leu* suggesting it to be the strongest risk factor. Terminal node 1 showed a 4.5-fold risk of ADCC (OR: 4.56; 95%CI: 2.65-7.85; p=**0.001**). Further, the association of NER variants for SCLC risk was evaluated (Table 5.63; Fig 5.28). The first split was observed at *Phe⁶⁷⁰Leu* marking this polymorphism as the strongest risk factor for SCLC. The ORs for the analysis were calculated by comparing with reference group. *XPD Arg¹⁵⁶Arg* (M), *XPA A23G* (W), *XPF 11985 A>G* (W), *XPA A23G* (W), *XPD Lys⁷⁵¹Gln* (W) was considered as the reference at terminal node 8 as it has the lowest case rate. The terminal node 1 i.e. *XPD A⁷⁵¹C* (M) had a 5-fold risk of having SCLC (OR: 5.35; 95%CI: 2.14-13.36; p=**0.0003**). A 6-fold risk for SCLC was observed for *XPA A23G* (M), *XPD Lys⁷⁵¹Gln* (W) (Node 2) (OR: 6.20; 95%CI: 2.40-16.01; p=**0.0001**). Terminal node 4 (*XPC Ala⁴⁹⁹Val* (W), *XPA A23G* (M), *XPG 11985 A>G* (W), *XPA A23G* (M), *Lys⁷⁵¹Gln* (M)) showed an increased risk for lung cancer (OR: 3.30; 95%CI: 1.52-7.17; p=**0.002**) whereas terminal node 6 showed almost 3-fold increased risk in SCLC patients (OR: 2.73; 95%CI: 1.24-5.97; p=**0.01**). The mutant genotype of *XPF 11985 A>G* (M) in combination with *XPA A23G* (W) and *XPD Lys⁷⁵¹Gln* (W) genotype showed a 5-fold risk for SCLC (Node 3) (OR: 5.25; 95%CI: 1.27-2.62; p=**0.02**). Terminal node 9 which was composed of *XPA A23G* (M), *XPG Phe⁶⁷⁰Leu* (M) showed a decreased SCLC risk (OR: 0.46; 95%CI: 0.15-1.39; p=0.16). *XPD Lys⁷⁵¹Gln* (W), *XPD Arg¹⁵⁶Arg* (M) had a 3.5-fold increased risk among SCLC patients (OR: 3.57; 95%CI: 0.15-1.39; p=**0.03**).

CART analysis was further applied on 135 SQCC patients, there was only a single split was observed at *XPG Phe⁶⁷⁰Leu* stating this polymorphism to be the strongest risk factor for SQCC (Table 5.62). Only a single terminal was observed suggesting a 4-fold risk among the SQCC patients (OR: 4.12; 95%CI: 2.49-6.81; p=**0.001**).

Table 5.60: Risk estimates based on CART analysis for all NER variants						
Terminal Node	Genotype for each terminal node	Control(n)	Case(n)	Case Rate ^a	OR (95% CI)	p ^b
Node 1	<i>XPG Phe⁶⁷⁰Leu (W)</i>	205	296	59.1	3.23 (2.32-4.50)	0.00
Node 2	<i>XPG Phe⁶⁷⁰Leu (M)</i>	165	74	31.0	1 .0 ^c	
^a case rate is percentage LC cases among all individuals in each terminal node {case / (case+ control) x100}; ^b p<0.05 is considered statistically significant; ^c reference group						
Table 5.61: Risk estimates based on CART analysis for all NER variants in ADCC patients						
Terminal Node	Genotype for each terminal node	Control(n)	Case(n)	Case Rate ^a	OR (95% CI)	p ^b
Node 1	<i>XPG Phe⁶⁷⁰Leu (W)</i>	205	102	33.2	4.56 (2.65-7.83)	0.00
Node 2	<i>XPG Phe⁶⁷⁰Leu (M)</i>	165	18	9.80	1 .0 ^c	
^a case rate is percentage LC cases among all individuals in each terminal node {case / (case+ control) x100}; ^b p<0.05 is considered statistically significant; ^c reference group						
Table 5.62: Risk estimates based on CART analysis for all NER variants in SQCC patients						
Terminal Node	Genotype for each terminal node	Control (n)	Case(n)	Case Rate ^a	OR (95% CI)	p ^b
Node 1	<i>XPG Phe⁶⁷⁰Leu (W)</i>	205	112	35.3	4.12 (2.49-6.81)	0.00
Node 2	<i>XPG Phe⁶⁷⁰Leu (M)</i>	165	23	12.2	1 .0 ^c	
^a case rate is percentage LC cases among all individuals in each terminal node {case / (case+ control) x100}; ^b p<0.05 is considered statistically significant; ^c reference group						

Table 5.63: Risk estimates based on CART analysis for all NER variants in SCLC patients						
Terminal Node	Genotype for each terminal node	Control (n)	Case (n)	Case Rate ^a	OR (95% CI)	p-Value ^b
Node 1	<i>XPD Lys⁷⁵¹Gln (M)</i>	13	13	50	5.35 (2.14-13.36)	0.0003
Node 2	<i>XPA A23G (M), XPD Lys⁷⁵¹Gln (W)</i>	13	12	48	6.20 (2.40-16.01)	0.0001
Node 3	<i>XPF11985(M), XPA A23G (W), XPD Lys⁷⁵¹Gln</i>	4	5	55.6	5.24 (1.27-2.62)	0.02
Node 4	<i>XPC Ala⁴⁹⁹Val (W), XPA A23G (M), XPF 11985 (W), XPA A23G (M), XPD Lys⁷⁵¹Gln</i>	25	16	39	3.30 (1.52-7.17)	0.002
Node 5	<i>XPC Ala⁴⁹⁹Val (M), XPA A23G (M), XPF 11985 (W), XPA A23G (M), XPD Lys⁷⁵¹Gln</i>	53	13	19.7	1.49 (0.70-3.18)	0.29
Node 6	<i>XPF 673 C>T (W), XPD Arg156Arg (W), XPA A23G (W), XPF 11985 A>G (W), XPA A23G (W), XPD Lys⁷⁵¹Gln (W)</i>	34	15	30.6	2.73 (1.24-5.97)	0.01
Node 7	<i>XPF 673 C>T (M), XPD Arg156Arg (W), XPA A23G (W), XPF 11985 (W), XPA A23G (W), XPD Lys⁷⁵¹Gln (W)</i>	9	0.0	0	0.00-0.00	0.98
Node 8	<i>XPD Arg¹⁵⁶Arg (M), XPA A23G (W),XPF 11985 (W),XPA A23G (W), XPD Lys⁷⁵¹Gln (W)</i>	54	4	6.9	1.0 ^c	
Node 9	<i>XPA A23G (M),XPG Phe670Leu (M)</i>	10	6	37.5	0.46 (0.15-1.39)	0.16
Node 10	<i>XPD Lys⁷⁵¹Gln (W),XPD Arg¹⁵⁶Ag (M)</i>	155	27	14.8	3.57 (0.15-1.39)	0.03
^a case rate is percentage LC cases among all individuals in each terminal node {case / (case+ control) x100}; ^b p<0.05 is considered statistically significant; ^c reference group						

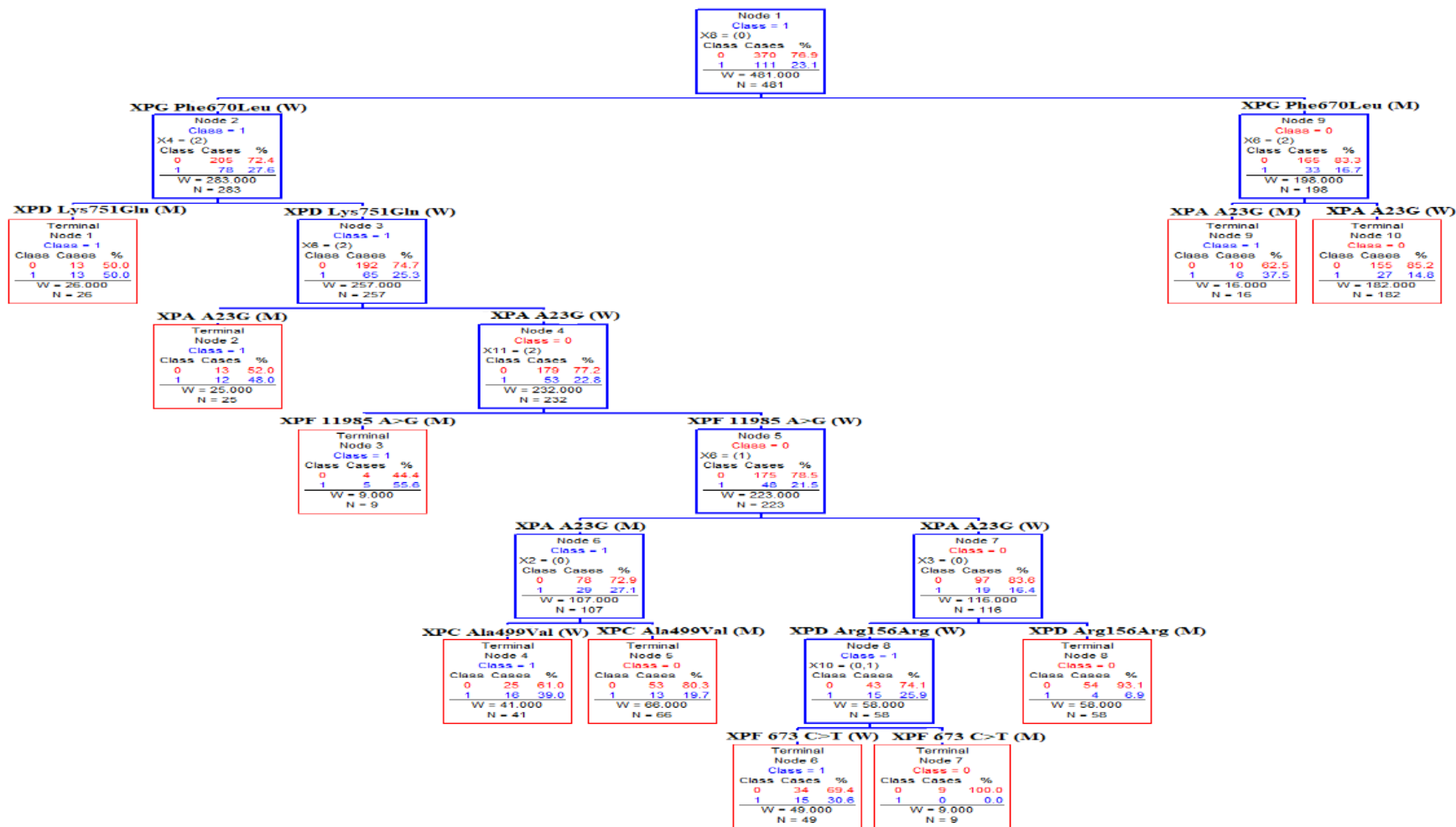


Fig 5.28: Classification and Regression Tree analysis showing high order NER gene-gene interactions in SCLC patient

5.6.7 CART analysis and its association with Smoking parameter

The high order interactions were applied as per the smoking status of an individual in the population (Table 5.64; Fig 5.29). The first split in the decision tree was observed at the smoking status indicating smoking as the main risk factor for lung cancer. CART tree further showed a single split in smokers giving two terminal nodes. Terminal node 1 i.e. *XPG Phe⁶⁷⁰Leu (W)* showed almost no effect for the lung cancer risk (OR: 1.04; 95%CI: 0.45-1.11; p=0.0001). Terminal node 2 consisted of minimum case rate therefore it was regarded as the reference. A parallel CART analysis was applied among non-smokers and we were not able to detect any high order interaction among NER polymorphic variants. A single terminal node was obtained depicting protective effect against the disease (Terminal node 3) (OR: 0.72; 95%CI: 0.46-1.11; p=0.01).

Table 5.64: Risk estimates based on CART analysis among smokers for all NER variants						
Terminal Node	Genotype for each terminal node	Control(n)	Case(n)	Case Rate ^a	OR (95%CI)	p-Value ^b
Node 1	<i>XPG Phe⁶⁷⁰Leu (W)</i>	146	242	62.4	1.04 (0.45-1.11)	0.0001
Node 2	<i>XPG Phe⁶⁷⁰Leu (M)</i>	125	61	32.8	1.0 ^c	
Node 3	Non smoking	67	99	40.4	0.72 (0.46-1.11)	0.01

^acase rate is percentage LC cases among all individuals in each terminal node {case / (case+ control) x100}; ^bp<0.05 is considered statistically significant; ^c reference group

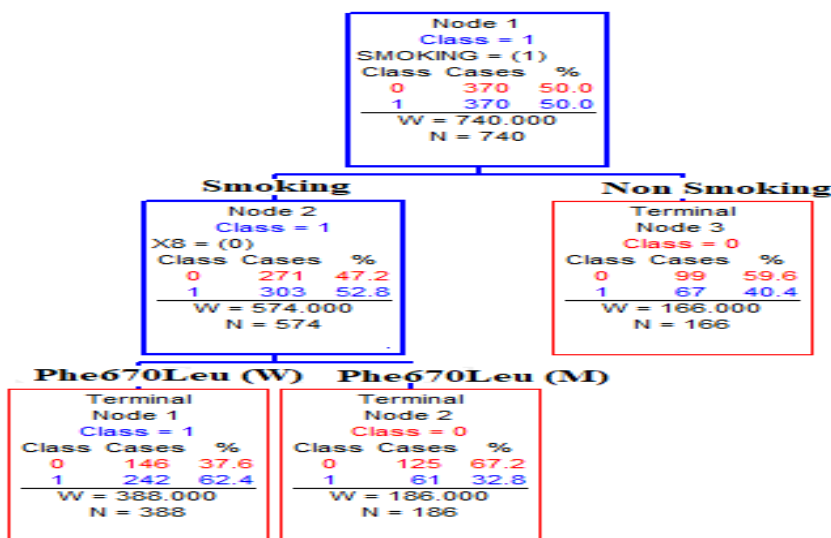


Fig 5.29: CART tree explaining overall NER SNP interaction on smoking

5.6.8 Association of NER genes polymorphic combinations and Overall Survival of Lung cancer patients:

Table 5.65 summarizes the univariate and multivariate analysis for the various polymorphic combinations of NER genes and their association with overall survival lung cancer patients. Only significant combinations are only represented in Table 5.65, other polymorphic combinations without any significant data have not been presented in the study. Median survival time (MST) was calculated using Kaplan Meier statistics and survival curves were obtained for each combination. Genotypic influence on the survival of the patients was examined by using log rank test. Wild genotype was regarded as reference group for each combination. Univariate analysis gave the probability of death (Hazards ratio) whereas multivariate Cox regression analysis provided the adjusted hazards which were adjusted for various covariates such as age, sex, smoking, histology, stage, KPS and ECOG in order to find whether a particular polymorphic combination could be an independent predictor factor. Maximum polymorphic combination did not show any significant association with the OS in lung cancer patients. Only few combinations expressed their association with the survival.

The first combination which was applied between polymorphic combinations *XPF 673C>T* and *XPD Asp³¹²Asn* showed a poor survival in mutant subjects with a borderline significant *p* value (0.06) suggesting high risk of death both in univariate (HR=3.64) and multivariate (HR¹=3.46; 95% CI: 0.94-12.64; *p*=0.12; MST=2.56) analysis. However, mutant and combined variant category was not significantly associated with any such effect. A polymorphic combination of *XPF 673 C>T* and *XPG Asp¹¹⁰⁴His* suggested a poor survival in patients with heterozygous (Group 2) genotype (HR¹=1.51; 95% CI= 0.98-2.34; *p*=0.06; *Fdr p*= 0.12; MST=5.56). While any such effect was not observed for individuals carrying heterozygous in combination with mutant genotype. However, in case of *XPF 11985 A>G* and *XPD Asp³¹²Asn* polymorphism the lung cancer subjects harboring the heterozygous (Group 2) genotype reported with poor MST (5.93 months, log rank *p*-value of 0.05) with hazard ratio of 1.44 although on adjusting with various covariates, no significant change in death probability was observed (HR¹= 1.54, *p*=0.03; *Fdr p*=0.15). Similar effect for the same polymorphism was observed in case of combined (heterozygous+mutant) genotype (HR¹=1.50; 95%CI: 1.02-2.21; *p*=0.04; *Fdr p*=0.10). Further, in case of *XPG Asp¹¹⁰⁴His* and *XPD Arg¹⁵⁶Arg* polymorphic combination, the subjects having the

heterozygous+mutant genotype (Group 4) reported with poor survival time when compared with reference (Group 1) genotype (6.23 vs 7.56). Both univariate (HR: 1.41; $p=0.05$) and multivariate analysis revealed significant 1.5-fold hazards ratio ($HR^1=1.51$; 95%CI=1.02-2.23; $p=0.03$; $Fdr\ p=0.15$). Similarly, heterozygous+mutant (Group 4) genotype was associated with increased hazard ratio with poor survival in *XPG Asp¹¹⁰⁴His* and *XPD Asp³¹²Asn* polymorphic variants ($HR^1: 1.58$; 95%CI: 0.98-2.56; $p=0.06$; $Fdr\ p=0.10$). Other polymorphic combinations exhibited a lack of association with the prognosis of lung cancer in North Indians.

Table 5.65: Association of polymorphic combinations with overall survival in lung cancer patients								
	Overall LC							
<i>XPF673 C>T & XPD Asp³¹²Asn</i>	Dead 110	Alive 25	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P-value	Fdr p
0 (Group 1)	58	14	8.03	1.00(Reference)		1.00(Reference)		
1(Group 2)	49	11	8.26	1.05 (0.72-1.54)	0.78	1.00 (0.66-1.49)	0.99	0.99
2 (Group 3)	3	-	2.56	3.64 (0.43-30.85)	0.01	3.46 (0.94-12.64)	0.06	0.12
3 (Group 4)	52	11	7.56	1.09 (0.75-1.60)	0.62	1.05 (0.70-1.56)	0.80	0.80
<i>XPF673 C>T & XPG Asp¹¹⁰⁴His</i>	Dead 116	Alive 26	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)	P-value	Fdr p
0 (Group 1)	85	20	9	1.00(Reference)		1.00(Reference)		
1(Group 2)	30	3	5.56	1.40 (0.88-2.21)	0.21	1.51 (0.98-2.34)	0.06	0.15
2 (Group 3)	1	3	6.80	0.98 (0.82-1.92)	0.80	0.99 (0.80-1.81)	0.78	0.72
3 (Group 4)	31	6	6.43	1.19 (0.81-1.73)	0.34	1.19 (0.81-1.76)	0.36	0.45
<i>XPF 11985 A>G& XPD Asp³¹²Asn</i>	Dead 116	Alive 23	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)	P-value	Fdr p
0 (Group 1)	68	17	8.03	1.00(Reference)		1.00(Reference)		
1(Group 2)	45	5	5.93	1.44 (0.97-2.15)	0.05	1.54 (1.04-2.28)	0.03	0.15
2 (Group 3)	3	1	7.58	1.01 (0.31-3.24)	0.98	1.18 (0.34-4.06)	0.78	0.72

3 (Group 4)	48	6	7.33	1.40 (0.95-2.07)	0.06	1.50 (1.02-2.21)	0.04	0.10
<i>XPG Asp¹¹⁰⁴His & XPD Arg¹⁵⁶Arg</i>	Dead 122	Alive 21	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)	P- value	Fdr p
0 (Group 1)	89	18	7.56	1.00(Reference)		1.00(Reference)		
1(Group 2)	30	3	6	1.30 (0.83-2.03)	0.20	1.36 (0.86-2.08)	0.18	0.26
2 (Group 3)	3	-	7.1	1.66 (0.85-2.22)	0.45	1.72 (0.86-2.28)	0.22	0.18
3 (Group 4)	33	3	6.23	1.41 (0.95-2.09)	0.05	1.51 (1.02-2.23)	0.03	0.10
<i>XPG Asp¹¹⁰⁴His & XPD Asp³¹²Asn</i>	Dead 114	Alive 24	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)	P- value	Fdr p
0 (Group 1)	80	19	10	1.00(Reference)		1.00(Reference)		
1(Group 2)	32	4	8.06	1.29 (0.83-2.00)	0.21	1.31 (0.85-2.02)	0.21	0.26
2 (Group 3)	2	1	7.23	1.62 (0.80-2.20)	0.36	1.77 (0.83-2.72)	0.20	0.18
3 (Group 4)	34	5	6.1	1.51 (1.01-2.26)	0.02	1.58 (0.98-2.56)	0.06	0.10

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS (karnofsky performance status) and ECOG (Eastern cooperative oncology group). OS=Overall survival.

Group 1- combination of wild genotypes of any two polymorphism Group 2- heterozygous genotype of two polymorphic variants in combination, Group 3: Mutant genotype in two polymorphic variants Group 4: Heterozygous+mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

5.6.9 Association of NER gene polymorphic combinations with overall survival according to histological subtype of Lung cancer patients:

To evaluate the relevance of different NER polymorphic combinations in predicting the survival of patients in different subgroups when stratified according to histology of lung cancer as represented in table 5.66, 5.67 and 5.68. As stated in table 5.66, we found that ADCC patients carrying the heterozygous (Group 2) genotype for *XPC Ala⁴⁹⁹Gln* and *XPD Lys⁷⁵¹Gln* when applied in combination showed a better OS as compared to wild genotype (13.43 months vs 6.73 months). Hazards ratio obtained from multivariate analysis (HR¹ =0.44) with a marginally significant *p* values 0.04 was observed. *XPD Arg¹⁵⁶Arg* and *XPC Ala⁴⁹⁹Val* polymorphic

combination suggested that the individuals carrying heterozygous (Group 2) as well as combined (heterozygous+mutant) genotype were associated lower hazard ratio (HR¹: 0.50; 95% CI: 0.24-1.03; *p*=0.06; *Fdr p*=0.06 and HR¹=0.51; 95%CI: 0.25-1.01;*p*=0.05; *Fdr p*=0.05) for ADCC. ADCC patients harboring heterozygous (Group 2) genotype for *XPG Phe⁶⁷⁰Leu* and *XPC Ala⁴⁹⁹Val* polymorphism exhibited a decreased adjusted hazard ratio (HR¹=0.22, *p*=0.001; *Fdr p*=0.004) with Median survival time of 13.43 months when compared with wild (Group 1) genotype. *XPF 11985A>G* and *XPC Ala⁴⁹⁹Val* polymorphic combination depicted a decreased risk for ADCC for both heterozygous (HR¹=0.43; *p*=0.03; *Fdr p*=0.06) as well as heterozygous+mutant genotype (HR¹=0.44; *p*=0.02; *Fdr p*=0.04). This polymorphic combination showed a better survival of 15 months in both heterozygous and heterozygous+mutant genotype when compared with the reference genotype (MST=7.13). Another polymorphic combination consisted *XPG Asp¹¹⁰⁴His* and *XPA A²³G* polymorphism showed a 2-fold risk among ADCC patients with heterozygous (Group 2= HR¹: 2.32; 95%CI: 1.03-5.22; *p*=0.04; *Fdr p*=0.06) and heterozygous genotype when combined with mutant genotype and represented as a single genotype (Group 4= HR¹: 2.39; 95%CI: 1.19-4.77; *p*=0.01; *Fdr p*=0.03).

In case of SCLC patients (Table 5.67) it was observed that patients having heterozygous (CT+CA) genotype for *XPF673 C>T* and *XPG Phe⁶⁷⁰Leu* polymorphism showed a poor MST as compared to as compared to patients carrying wild type (CC+CC) genotype (4.63 vs 11 months, log rank *p*=0.12). However, on further adjusting with various covariates the adjusted HR was found to be 2.51 (*p*=0.03). Another polymorphic combination which showed almost a 3-fold risk among SCLC patients was *XPF 11985 A>G* and *XPC Ala⁴⁹⁹Val* (Group 2= HR¹= 3.29; 95%CI: 1.53-7.08; *p*=0.002). The heterozygous and mutant genotype (Group 4) when act as a single genotype for the above polymorphism depicted a 2.5-fold risk for SCLC (HR¹= 2.55; 95%CI: 1.32-4.94; *p*=0.005; *Fdr p*=0.01). SCLC patients with *XPF 11985 A>G* and *XPD Arg¹⁵⁶Arg* polymorphism present together showed a 4-fold risk among those who were carrying heterozygous (Group 2= HR¹= 4.84; 95%CI: 1.53-15.30; *p*=0.007; *Fdr p*=0.02) as well as combined genotype (HR¹= 4.79; 95%CI: 1.56- 14.70; *p*=0.006; *Fdr p*=0.01). *11985A>G* polymorphism when work in combination with *Lys⁷⁵¹Gln* suggested an increased SCLC risk in patients carrying combined (heterozygous+mutant) genotype (HR¹= 2.85; 95%CI: 1.18-6.86; *p*=0.01; *Fdr p*=0.01). Other combination which emphasized on the association with increased SCLC risk were *XPF 11985 A>G* with *XPA G⁷⁰⁹A* and *XPF 11985 A>G* with *XPG Phe⁶⁷⁰Leu*.

Out of all the polymorphic combinations applied, For *XPF 11985 A>G* and *XPD Arg¹⁵⁶Arg* polymorphism when applied in combination suggested that SQCC patients (Table 5.68) with heterozygous (Group 2) genotype showed a poor survival time when compared with wild (Group 1) genotype (2.13 vs 5.76, log rank $p=0.02$). Death rate was found to be three times higher in case of these subjects (adjusted HR=3.04, $p=0.04$). For *XPF Arg⁴¹⁵Gln* and *XPG Phe⁶⁷⁰Leu* polymorphic combination, SQCC patients having Group 4 genotype showed a better survival of 32.5 months as compares to wild type (Group 1) genotype having survival period of 6.23 months. Hazards ratio obtained was found to be statistically significant for both the univariate (HR= 0.42; log rank $p=0.05$) and multivariate analysis (HR¹: 0.31; $p=0.02$; *Fdr* $p=0.03$). Furthermore, for *XPG Asp¹¹⁰⁴His* and *XPD Arg¹⁵⁶Arg* polymorphic combination, it was found that SQCC patients possessing heterozygous genotype (Group 1) (MST=5.36) as well as combined genotype (MST=4.3), had a very poor MST as compared subjects with wild (GG+CC) genotype (MST=7.56). The hazards ratio depicted three-fold higher death probability (HR¹ =3.08, $p=0.007$; *Fdr* $p=0.02$) in SQCC patients with combined genotype when compared with SQCC patients carrying wild genotype. SQCC patients having the heterozygous (Group 2) genotype for above polymorphism showed a three-fold higher death risk (adjusted HR=3.08, $p=0.001$; *Fdr* $p= 0.003$). On the other hand, all other polymorphic combination was found to play a negligible role in prognosis of patients with any of the histological subgroup.

Table 5.66: Association of polymorphic combinations NER genes with overall survival in ADCC patients

<i>XPC Ala⁴⁹⁹Val & XPD Lys⁷⁵¹Gln</i>	Dead 31	Alive 11	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	14	2	6.73	1.00(Reference)		1.00(Reference)		
1(Group 2)	16	8	13.43	0.62 (0.29-1.31)	0.18	0.44 (0.19-0.97)	0.04	0.06
2 (Group 3)	1	1						
3 (Group 4)	17	9	12.13	0.60 (0.29-1.24)	0.12	0.48 (0.23-0.99)	0.03	0.05
<i>XPA G709A & XPD Lys⁷⁵¹Gln</i>	Dead 42	Alive 7	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	23	5	6.43	1.00(Reference)		1.00(Reference)		
1(Group 2)	18	2	9.40	0.95 (0.51-1.76)	0.88	1.36 (0.64-2.89)	0.42	0.42
2 (Group 3)	1	-	1.83	4.57 (0.07-289.13)	0.09	52.70 (1.11-249.30)	0.04	0.04
3 (Group 4)	19	2	9.40	1.00 (0.54-1.83)	0.99	1.44 (0.72-2.89)	0.30	0.30
<i>XPD Arg¹⁵⁶Arg & XPC Ala⁴⁹⁹Val</i>	Dead 39	Alive 9	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	18	1	7.56	1.00(Reference)		1.00(Reference)		
1(Group 2)	18	8	15	0.46 (0.22-0.93)	0.01	0.50 (0.24-1.03)	0.06	0.06
2 (Group 3)	3	-						
3 (Group 4)	21	8	13.43	0.46 (0.23-0.93)	0.01	0.51 (0.25-1.01)	0.05	0.05
<i>XPG Phe⁶⁷⁰Leu & XPC Ala⁴⁹⁹Val</i>	Dead 35	Alive 9	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	21	3	8.03	1.00(Reference)		1.00(Reference)		
1(Group 2)	14	5	13.43	0.48 (0.24-0.95)	0.03	0.22 (0.09-0.55)	0.001	0.004
2 (Group 3)	-	1						
3 (Group 4)	14	6	13.43	0.44 (0.22-0.88)	0.01	0.20 (0.08-0.50)	0.0006	0.004
<i>XPF673 C>T & XPC Ala⁴⁹⁹Val</i>	Dead 34	Alive 10	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	21	3	8.03	1.00(Reference)		1.00(Reference)		
1(Group 2)	13	6	13.43	0.48 (0.24-0.95)	0.03	0.22 (0.09-0.55)	0.001	0.004

2 (Group 3)	-	1						
3 (Group 4)	13	7	13.43	0.44 (0.22-0.88)	0.01	0.20 (0.08-0.50)	0.0006	0.004
XPF673 C>T& XPA G709A	Dead 33	Alive 6	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P- value	Fdr p
0 (Group 1)	21	3	5.7	1.00(Reference)		1.00(Reference)		
1(Group 2)	12	2	9.43	0.64 (0.32-1.27)	0.21	0.45 (0.20-1.03)	0.06	0.06
2 (Group 3)								
3 (Group 4)	12	3	9.43	0.56 (0.28-1.12)	0.10	0.43 (0.18-1.01)	0.05	0.05
XPF11985 A>G& XPC Ala⁴⁹⁹Val	Dead 39	Alive 10	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P- value	Fdr p
0 (Group 1)	22	5	7.13	1.00(Reference)		1.00(Reference)		
1(Group 2)	13	4	15	0.57 (0.29-1.10)	0.09	0.43 (0.19-0.95)	0.03	0.06
2 (Group 3)	1	1						
3 (Group 4)	14	5	15	0.60 (0.32-1.13)	0.11	0.44 (0.21-0.91)	0.02	0.04
XPG Asp¹¹⁰⁴His & XPA A23G	Dead 41	Alive 8	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P- value	Fdr p
0 (Group 1)	26	8	12.13	1.00(Reference)		1.00(Reference)		
1(Group 2)	10	-	6.26	1.98 (0.81-4.82)	0.05	2.32 (1.03-5.22)	0.04	0.06
2 (Group 3)	1	-						
3 (Group 4)	11	-	6.43	2.10 (0.99-4.42)	0.01	2.39 (1.19-4.77)	0.01	0.03

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS (karnofsky performance status) and ECOG (Eastern cooperative oncology group). OS=Overall survival.

Group 1- combination of wild genotypes of any two polymorphism Group 2- heterozygous genotype of two polymorphic variants in combination, Group 3: Mutant genotype in two polymorphic variants Group 4: Heterozygous+mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

Table 5.67: Association of polymorphic combinations NER genes with overall survival in SCLC patients

<i>XPF673 C>T& XPG Phe⁶⁷⁰Leu</i>	Dead 32	Alive 7	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	21	7	11	1.00(Reference)		1.00(Reference)		
1(Group 2)	10	-	4.63	1.76 (0.74-4.16)	0.12	2.51 (1.06-5.97)	0.03	0.03
2 (Group 3)	1	-						
3 (Group 4)	11	-	4.46	1.91 (0.82-4.44)	0.06	2.82 (1.20-6.63)	0.01	0.01
<i>XPF11985 A>G& XPC Ala⁴⁹⁹Val</i>	Dead 39	Alive 5	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	21	4	12.63	1.00(Reference)		1.00(Reference)		
1(Group 2)	19	-	4.13	2.49 (1.23-5.02)	0.00 1	3.29 (1.53-7.08)	0.002	0.01
2 (Group 3)	1	1						
3 (Group 4)	20	1	4.25	1.97 (1.08-3.59)	0.01	2.55 (1.32-4.94)	0.005	0.01
<i>XPF 11985 A>G& XPD Arg¹⁵⁶Arg</i>	Dead 28	Alive 5	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	18	5	12.46	1.00(Reference)		1.00(Reference)		
1(Group 2)	9	-	3.43	2.57 (0.92-7.17)	0.01	4.84 (1.53-15.30)	0.007	0.02
2 (Group 3)	1	-						
3 (Group 4)	10	-	3.3	2.69 (1.00-7.20)	0.00 7	4.79 (1.56-14.70)	0.006	0.01
<i>XPF 11985 A>G& XPD Lys⁷⁵¹Gln</i>	Dead 30	Alive 6	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	19	5	10.1	1.00(Reference)		1.00(Reference)		
1(Group 2)	10	1	3.53	2.17 (0.86-5.44)	0.03	2.97 (1.15-7.65)	0.02	0.03
2 (Group 3)	1	-						
3 (Group 4)	11	1	3.53	2.07 (0.87-4.93)	0.04	2.85 (1.18-6.86)	0.01	0.01
<i>XPF 11985 A>G & XPA G709A</i>	Dead 38	Alive 10	Median OS (months)	HR ^a (95% CI)	Log P	HR ¹ (95% CI) ^b	P- value	Fdr p
0 (Group 1)	24	9	12.46	1.00(Reference)		1.00(Reference)		

1(Group 2)	10	1	2.4	2.30 (0.90-5.87)	0.02	3.38 (1.27-9.00)	0.01	0.02
2 (Group 3)	4	-	4.1	2.50 (0.83-4.83)	0.10	3.52 (1.46-8.20)	0.08	0.06
3 (Group 4)	14	1	4.06	2.14 (1.01-4.55)	0.01	3.16 (1.46-6.84)	0.003	0.01
<i>XPF11985 A>G & XPG Phe⁶⁷⁰Leu</i>	Dead 39	Alive 12	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P- value	Fdr p
0 (Group 1)	29	11	12	1.00(Reference)		1.00(Reference)		
1(Group 2)	9	1	5.83	1.78 (0.72-4.36)	0.11	2.44 (1.04-5.72)	0.03	0.03
2 (Group 3)	1	-						
3 (Group 4)	10	1	5.83	1.84 (0.83-4.05)	0.06	2.70 (1.22-5.96)	0.01	0.01

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS (karnofsky performance status) and ECOG (Eastern cooperative oncology group). OS=Overall survival.

Group 1- combination of wild genotypes of any two polymorphism Group 2- heterozygous genotype of two polymorphic variants in combination, Group 3: Mutant genotype in two polymorphic variants Group 4: Heterozygous+mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

Table 5.68: Association of polymorphic combinations NER genes with overall survival in SQCC patients								
<i>XPF 11985 A>G & XPD Arg¹⁵⁶Arg</i>	Dead 40	Alive 8	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P- value	Fdr p
0 (Group 1)	35	7	5.76	1.00(Reference)		1.00(Reference)		
1(Group 2)	5	-	2.13	2.70 (0.65-11.15)	0.02	3.04 (1.03-8.94)	0.04	0.06
2 (Group 3)	-	1						
3 (Group 4)	5	1	3.08	1.45 (0.49-4.29)	0.43	1.75 (0.63-4.84)	0.27	0.27
<i>XPF Arg⁴¹⁵His & XPG Phe⁶⁷⁰Leu</i>	Dead 46	Alive 12	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P- value	Fdr p
0 (Group 1)	41	8	6.23	1.00(Reference)		1.00(Reference)		
1(Group 2)	4	3	17.16	0.51 (0.23-1.13)	0.18	0.41 (0.13-1.22)	0.11	0.11
2 (Group 3)	1	1						

3 (Group 4)	5	4	32.5	0.42 (0.21-0.85)	0.05	0.31 (0.11-0.86)	0.02	0.03
<i>XPG Asp¹¹⁰⁴His</i> & <i>XPD Arg¹⁵⁶Arg</i>	Dead 54	Alive 10	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P- value	Fdr p
0 (Group 1)	39	10	7.56	1.00(Reference)		1.00(Reference)		
1(Group 2)	11	-	5.36	2.24 (0.93-5.42)	0.01	2.78 (1.31-5.85)	0.007	0.02
2 (Group 3)	4	-	3.6	2.06 (1.01-4.02)	0.07	3.32 (1.82-6.22)	0.10	0.08
3 (Group 4)	15	-	4.3	2.22 (1.04-4.72)	0.005	3.08 (1.55-6.13)	0.001	0.003

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS (karnofsky performance status) and ECOG (Eastern cooperative oncology group). OS=Overall survival.

Group 1- combination of wild genotypes of any two polymorphism Group 2- heterozygous genotype of two polymorphic variants in combination, Group 3: Mutant genotype in two polymorphic variants Group 4: Heterozygous+mutant genotype of two polymorphic variants; ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma.

5.6.10 Association of different unfavorable genotypic combinations of NER gene with overall survival of Lung cancer patients:

To understand the cumulative effect of the polymorphic variants of each gene, different genotypic combination of unfavorable genotypes were analyzed for predicting the overall survival of lung cancer patients as illustrated in table 5.59 (Fig 5.30). The Stree software was applied to the data of 311 samples and the results were further confirmed by applying Univariate Kaplan Meier analysis and multivariate Cox regression model. The study group was divided as per the presence of two genotypes *i.e.* wild and heterozygous+mutant. A tree was achieved after running the software which resulted into 9 terminal nodes. Terminal node 9 was regarded as the reference node as it represented a minimum survival period of 3.5 months. *XPG Asp¹¹⁰⁴His* obtained as the root node which further gets split into two daughter nodes. The first daughter node was observed at *XPD Arg¹⁵⁶Arg* and *XPD Lys⁷⁵¹Gln*. These daughter nodes further get split into daughter node and a terminal node and so on. The terminal node 1 consisting of unfavorable genotypes showed a MST of 6.43 months depicting a lower hazard ratio towards the lung cancer risk when applied with univariate (HR: 0.32; *p*=0.0008) and multivariate analysis (HR¹: 0.27;

p=0.001; Fig 5.31A). Terminal node 2 showing *XPG Asp¹¹⁰⁴His (W)*, *XPD Lys⁷⁵¹Gln (W)* genotype showed a mediocre survival of 7.6 months also emphasizing on lower risk of death (HR¹: 0.33; 95%CI: 0.16-0.68; p=0.002) when compared with the reference node. Terminal node 3 depicting *XPG Asp¹¹⁰⁴His (H+M)*, *XPD Arg¹⁵⁶Arg (W)*, *XPA A²³G (H+M)* genotype suggested a better survival of 6 months when compared with the reference. This node also signifies the combination of above genotypes showed a significantly decreased hazard ratio (HR¹: 0.13; 95%CI: 0.11-0.79; p=0.01; Fig 5.31A). Terminal node 4 showed a survival of 4.2 months when compared with the reference node depicting survival period of 3.5 months. Node 4 deciphered *XPG Asp¹¹⁰⁴His (H+M)*, *XPD Arg¹⁵⁶Arg (W)*, *XPA A²³G (W)*, *XPC Lys⁹³⁹Gln (W)* showing hazard ratio of 0.13 with a significant p-value (0.01) (Fig 5.31A). Terminal node 6 represented the best survival time of 11.6 months when compared with the reference node showing a lower hazard ratio in both univariate (HR: 0.24; p<0.0001; Fig 5.31B) as well as multivariate analysis (HR¹: 0.22; p=0.0007). Terminal node 7 comprises of *XPG Asp¹¹⁰⁴His (W)*, *XPD Lys⁷⁵¹Gln (H+M)*, *XPF Arg⁴¹⁵Gln (W)*, *XPF 11985 A>G (W)* genotypes represents a survival time of 10.8 months showing significant protective effect when Cox regression model applied adjusted with various covariate (HR¹: 0.39; 95%CI: 0.16-0.96; p=0.04; Fig 5.31B). The last node observed in the study was terminal node 8 (*XPG Asp¹¹⁰⁴His (W)*, *XPD Lys⁷⁵¹Gln (H+M)*, *XPF Arg⁴¹⁵Gln (W)*, *XPF 11985A>G (H+M)*, *XPD Arg¹⁵⁶Arg (H+M)*) also showed a significant hazard ratio (HR¹: 0.22; 95%CI: 0.04-1.01; p=0.05; MST=9; Fig 5.31B).

Table 5.69: Overall survival in NER SNPs by applying STREE							
Terminal nodes	Dead	Alive	Median OS (months)	HR^a (95% CI)	Log P	HR¹ (95% CI)^b	P-value
Terminal node 1 <i>XPG Asp¹¹⁰⁴His (H+M), XPD Arg¹⁵⁶Arg (W)</i>	38	9	6.43	0.32(0.11-0.93)	0.0008	0.27 (0.12-0.60)	0.001
Terminal node 2 <i>XPG Asp¹¹⁰⁴His(W),XPD Lys⁷⁵¹Gln(W)</i>	85	17	7.6	0.34(0.12-0.97)	0.0007	0.33 (0.16-0.68)	0.002
Terminal node 3 <i>XPG Asp¹¹⁰⁴His (H+M), XPD Arg¹⁵⁶Arg (W), XPA A23G (H+M)</i>	27	2	6	0.44(0.17-1.12)	0.01	0.29 (0.11-0.79)	0.01
Terminal node 4 <i>XPG Asp1104His (H+M), XPD Arg156Arg (W), XPA A23G (W), XPC Lys939Gln (H+M)</i>	8	1	8.3	0.40(0.15-1.08)	0.03	0.13(0.02-0.71)	0.01
Terminal node 5 <i>XPG Asp¹¹⁰⁴His (H+M), XPD Arg¹⁵⁶Arg (W), XPA A23G (W), XPC Lys⁹³⁹Gln (W)</i>	8	-	4.2	0.99(0.39-2.52)	0.99	0.52 (0.11-2.52)	0.42
Terminal node 6 <i>XPG Asp¹¹⁰⁴His (W), XPD Lys⁷⁵¹Gln(H+M), XPF Arg⁴¹⁵Gln (H+M)</i>	38	15	11.6	0.24(0.07-0.82)	<0.0001	0.22 (0.09-0.52)	0.0007
Terminal node 7 <i>XPG Asp¹¹⁰⁴His (W), XPD Lys⁷⁵¹Gln(H+M), XPF Arg⁴¹⁵Gln (W), XPF 11985 A>G (W)</i>	29	9	10.8	0.31(0.10-0.89)	0.0004	0.39 (0.16-0.96)	0.04
Terminal node 8 <i>XPG Asp¹¹⁰⁴His (W), XPD Lys⁷⁵¹Gln(H+M), XPF Arg⁴¹⁵Gln (W), XPF 11985 A>G (H+M), XPD Arg¹⁵⁶Arg (H+M)</i>	13	2	9	0.33(0.12-0.92)	0.003	0.22 (0.04-1.01)	0.05

Terminal node 9 <i>XPG Asp¹¹⁰⁴His (H+M), XPD Lys⁷⁵¹Gln (H+M), XPF Arg⁴¹⁵Gln (W), XPF 11985 A>G (H+M), XPD Arg¹⁵⁶Arg (W)</i>	10	-	3.52	1.00 (Reference)		1.00 (Reference)	

(a)-Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months (b)- Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age , sex, smoking status, stage, KPS (karnofsky performance status) and ECOG (Eastern cooperative oncology group). OS=Overall survival.

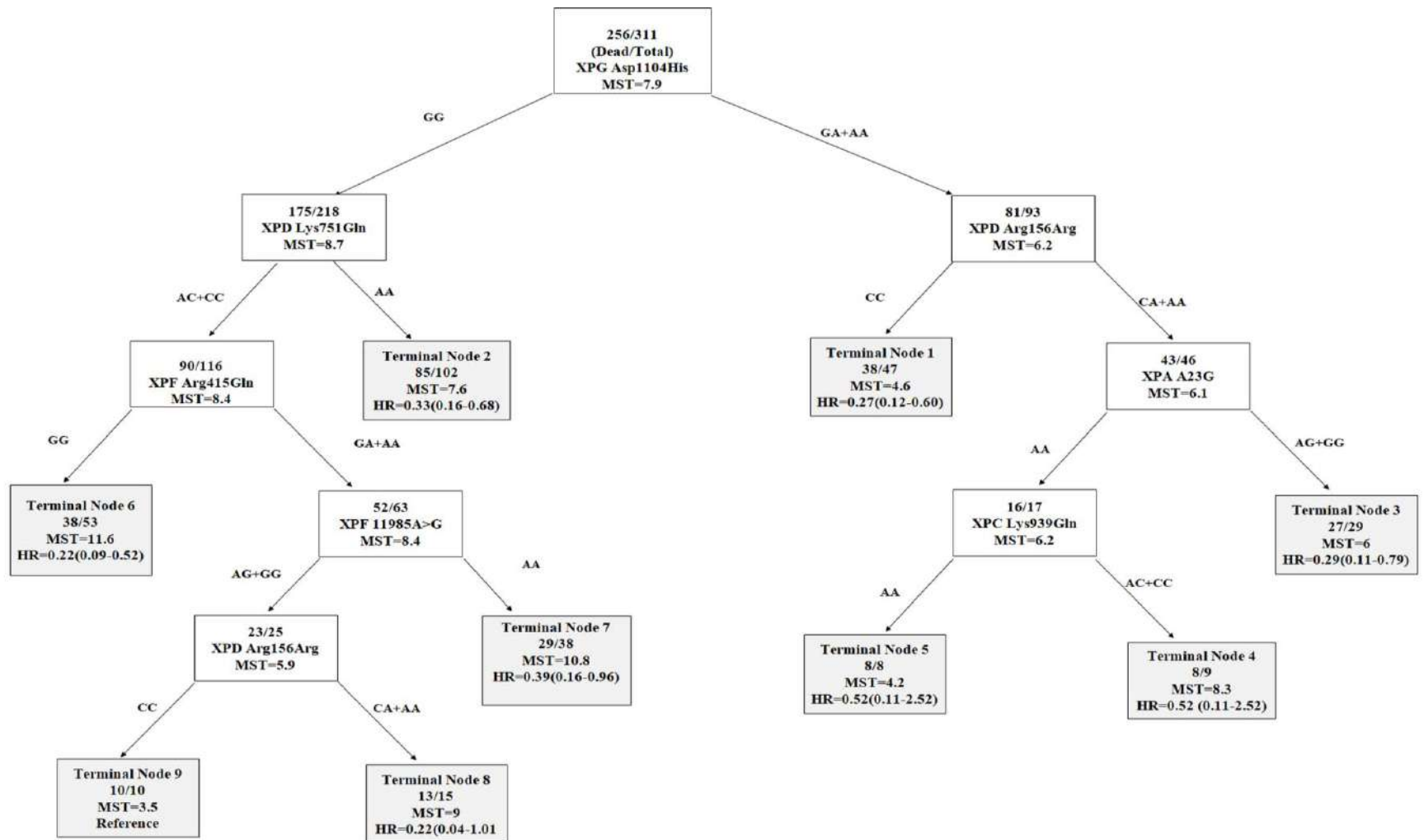


Fig 5.30: Recursive partitioning method (STREE) explaining association of NER SNPs with overall survival in lung cancer patients

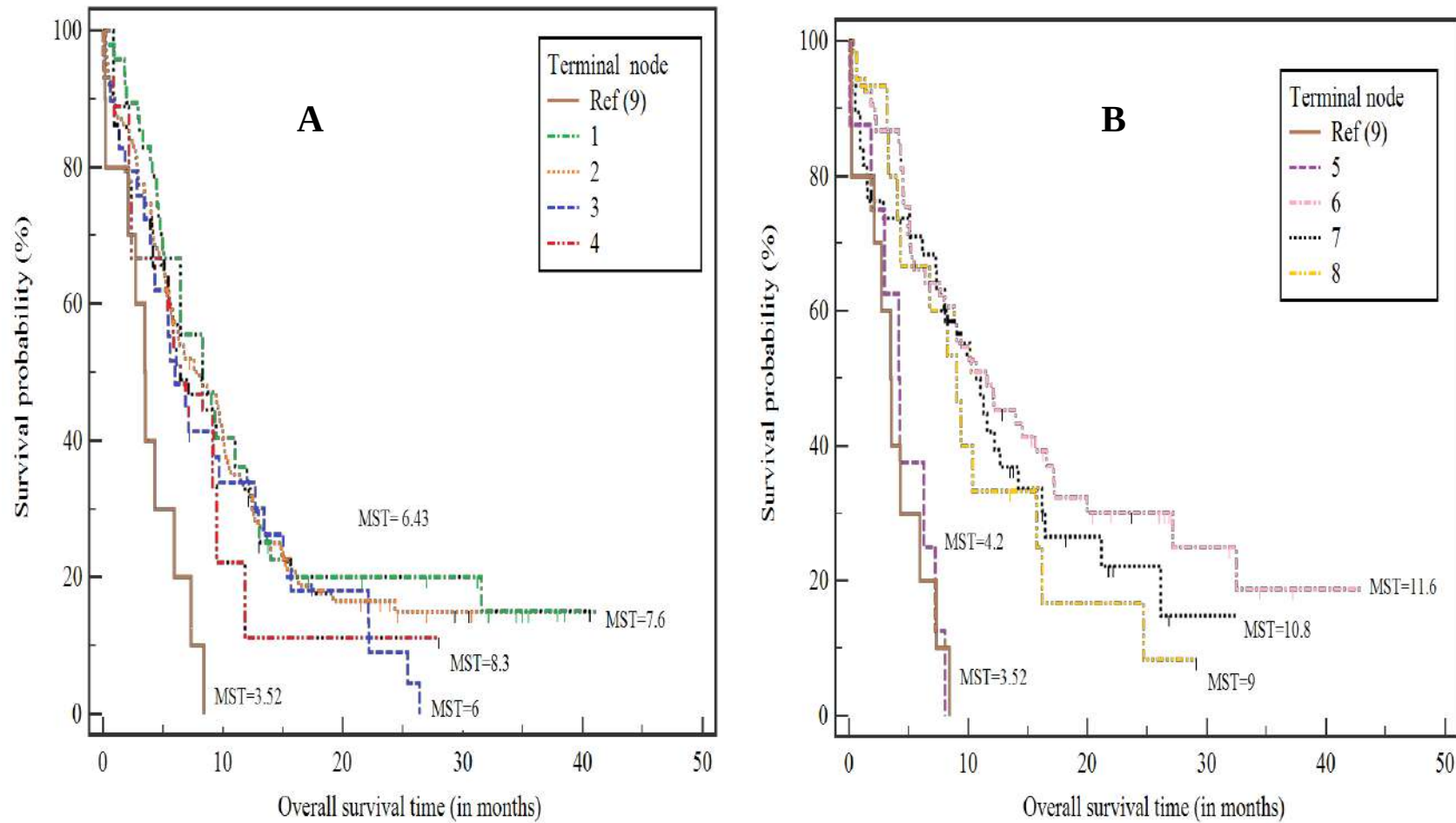


Fig 5.31: Kaplan-Meier graph explaining overall survival in unfavorable genotypic combinations of NER SNPs (a) Terminal nodes 1-4 when compared with reference node (node 9) (b) Terminal nodes 5-8 when compared with reference node (9).

CHAPTER-6

Discussion

Epidemiological studies with large sample size studies provide us insight into relationship between DNA repair genes and lung cancer risk. These studies contribute to the fact that irrespective of the regulatory pathways and environmental variations, single nucleotide polymorphism might influence diseases in gene of interest. These variations add to our sense of understanding of public health domain which occurs due to changes in DNA repair genes.

Human genome is constantly safeguarded by the exogenous as well as endogenous assaults with the help of DNA repair pathways. Any deficiency or single nucleotide polymorphism in DNA repair activity leads to alteration of DNA expression along with suboptimal repair activity. This influences prognosis and survival of lung cancer patients. DNA repair system might behave in a dual manner, where DNA repair capacity (DRC) might show improper repair capacity in removing lesions caused due to mutagens which might increase the chances of having cancer or on the other end might act in positive manner due to reduced rate of removal of anti-cancer drugs from DNA when treated with platinum based chemotherapy. SNPs results in the variations in DRC among individuals. Therefore, these polymorphic variants might help in prognosis of lung cancer patients undergoing chemotherapy.

The present study focuses on the genetic variations in NER genes which alter the repair capacity of DNA resulting into modulation of susceptibility and survival in lung cancer patients. Three major genes *XPC*, *XPG* and *XPA* which are located at different exons and their association with lung cancer predisposition, survival and response were explained in this study. Three SNPs of each *XPD* and *XPF* were also examined for observing lung cancer susceptibility, overall survival and clinical response. The study was conducted in North Indian population and to our knowledge suggested as the first study to be conducted on population of such ethnicity.

6.1 XPC polymorphism and its relationship with lung cancer risk and overall survival

6.1.1 XPC polymorphism and their association with lung cancer risk

In this study, an increased risk for lung cancer was found in patients with *Lys*⁹³⁹*Gln* mutant genotype ($p=0.0007$). The investigations suggested a two-fold higher risk for homozygous variant genotype of *Lys*⁹³⁹*Gln* polymorphism which agreed with the study conducted by Jin *et al.*, 2014 and assessed a higher risk of developing lung cancer in Asian subjects carrying variant allele of *XPC Lys*⁹³⁹*Gln* polymorphism. Raaschou-Nielson *et al.* (2008) further emphasized on an increased risk of lung cancer in variant genotype of *Lys*⁹³⁹*Gln* in Caucasian population. *XPC Lys*⁹³⁹*Gln* suggested to be closely related with the development of lung cancer. Our data was further supported by the studies conducted by Vogel *et al.*, 2005 suggesting association of *Lys*⁹³⁹*Gln* polymorphism with increase lung cancer risk in Finnish population. On the same lines, a study by Hu *et al.*, 2005 was also in agreement with our study emphasizing on the association of *Lys*⁹³⁹*Gln* polymorphism in lung cancer risk. This might be because *XPC* protein binds with the HR23B to form a hetero-dimeric complex which is considered as the crucial step for the detection of wide variety of lesions (Raashou-Nielson *et al.*, 2008) and hence recruit the whole repair machinery to the damaged site (Kusumato *et al.*, 2001). The gene polymorphism in *XPC* especially the nsSNPs, leads to the alteration in the repair capacity of an individual which causes genetic instability and ultimately results into carcinogenesis (Sugusawa *et al.*, 2002). Although there were also few studies which were in disagreement with the present study. A study by Letkova *et al.*, 2013 and Sakoda *et al.*, 2012 conferred association of *Lys*⁹³⁹*Gln* polymorphism with decreased lung cancer risk in Caucasian Slovaks and African Americans which was not in agreement in the present study.

The variant genotype (TT) of another *XPC* variant *i.e.* *Ala*⁴⁹⁹*Val* was found to be demonstrating a strong protective effect in lung cancer patients ($p=0.003$). When analyzed the combined effect of heterozygous and mutant genotype in the polymorphic variants, no significant association was observed towards lung cancer risk. These results suggested that the genetic alteration in *XPC*, interaction with environmental factors, ethnic background, host characteristics or prolonged exposure to the carcinogen could be the possible reasons for this protective effect. Our results were consistent with the studies conducted by Shen *et al.* (2005)

suggesting a protective effect of *XPC Ala⁴⁹⁹Val* towards the risk of lung cancer in Asian population. However, few epidemiological studies reported by Vodicka *et al.* (2004) and Jin *et al.* (2014) differ from our study suggesting no functional significance of *XPC Ala⁴⁹⁹Val* with the risk of lung cancer in European and Korean population. Three studies on Asian population suggested no association of *Ala⁴⁹⁹Val* polymorphism with lung cancer risk (Lee *et al.*, 2005, Shen *et al.*, 2005 and Bai *et al.*, 2007) and were in disagreement with the present study. A study by Hu *et al.*, 2005 and Skoda *et al.*, 2012 also differ from our study as they suggested association of *Ala⁴⁹⁹Val* polymorphism with increased lung cancer risk.

When we divided sample population on the basis of histological subtypes: ADCC, SQCC and SCLC. Our results suggested a significant risk of occurrence of disease in patients with mutant genotype observed for all the three histological subtypes for *Lys⁹³⁹Gln* which includes ADCC (OR: 4.72), SQCC (OR: 3.06) and SCLC (OR: 2.30). Although the reason behind such histology dependent differences in the genetic effect which is conferred by these polymorphism is unknown, but as suggested it might be related with the difference in carcinogenesis pathway among the histological subtypes of lung cancer. Many previous studies suggested that the determination of histological subtype is related with the individual's exposure to the specific initiation agent (Letkova *et al.*, 2013). Therefore, various genetic factors may be responsible for the susceptibility in histological subtypes of lung cancer (Yin *et al.*, 2006; Sakoda *et al.*, 2012).

6.1.2.1 *XPC* polymorphisms and their relationship with smoking as a parameter

Our results also suggested a two-fold increased risk of lung cancer in the homozygous variant (CC) showing *Lys⁹³⁹Gln* polymorphism among both smokers (OR: 2.55) and non-smokers (2.05). A more pronounced risk was observed in CC genotype of heavy smokers (pack years ≥ 25 years) (OR: 3.71) for *Lys⁹³⁹Gln*. The above results strongly suggest that cigarette smoking may not be the major component associated with the initiation of lung cancer but many other unidentified carcinogens may have played the major role in development of the disease. Our study was consistent with the study carried out by Hu *et al.* (2005) which showed more prominent risk in individuals carrying mutant genotype amongst smokers for *Lys⁹³⁹Gln* polymorphism. The studies by Lopez-Cima *et al.*, 2007 and Letkova *et al.*, 2013 further supported our study as they elaborated prevalence of lung cancer in those patients who were smokers and having *XPC Lys⁹³⁹Gln* polymorphism indicating a persistent role of environmental

factor in causing the disease. We observed a strong protective effect in individuals with TT and CT genotypes for *Ala*⁴⁹⁹*Val* polymorphism amongst smokers. These genotypes may have showed their protective effect due to their modulating response towards the carcinogenic exposure. As per the ‘greater apoptosis hypothesis’ as stated by Zhu *et al.* (2006) the apoptosis is usually triggered by exposure to a carcinogen, therefore, a increased protective effect may be observed amongst smokers (Nelson *et al.*, 2002; Bowen *et al.*, 2003).

6.1.2 XPC and haplotypic analysis

Haplotypic analysis showed that when compared with the most abundant A-C haplotype, the A-T haplotype which consist of variant T-allele, conferred a significantly decreased risk of lung cancer. Zhu *et al.* (2006) supported the above results emphasizing on decreased risk of bladder cancer in subjects carrying one variant T allele of *XPC Ala*⁴⁹⁹*Val* and one wild A allele of *Lys*⁹³⁹*Gln*. They further suggested the correlation of the variant T-allele of *Ala*⁴⁹⁹*Val* with the efficient DRC (Qaio *et al.*, 2002; Zhu *et al.*, 2006). Therefore, all the favorable allele which consisted of A-T haplotype are associated with the favorable DRC status and hence confer a decreased risk towards the lung cancer risk. Our data also showed a protective effect for A-T and C-T haplotype while C-C haplotype which consist of the adverse allele showed a slight increased effect towards the lung cancer. According to the ‘greater apoptosis hypothesis’, adverse allele have a less efficient DRC and it may or may not show the apoptotic effect (Bowen *et al.*, 2003; Zhang *et al.*, 2014).

6.1.3 XPC polymorphism and overall survival

The survival analysis of 311 patients was carried out using Kaplan-Meier survival analysis and Cox regression analysis suggesting no significant association of *XPC* variants in both univariate Kaplan Meier survival analysis and multivariate Cox regression analysis. Our results were in concordance with the study conducted by Matakideu *et al.* (2006), as they suggested no significant association of *Lys*⁹³⁹*Gln* variant between stages of diagnosis in Lung cancer patients. They further emphasized on an increased mean survival for mutant genotype in *XPC Lys*⁹³⁹*Gln* variant and differ from the findings of the present study. The clinical response is an important parameter which is related with the clinical trials efficacy. The patients suffering from the lung cancer undergoes platinum based chemotherapy. Many population based studies supported a

relationship between polymorphism involved in DNA repair and decreased chemotherapy response. The polymorphism in the DNA repair pathway genes such as *XPC* alters the DNA repair activities and may influence response in patients towards the drug which may affect patient's survival and disease outcomes (Li *et al.*, 2013; Tiesco *et al.*, 2013; Xu *et al.*, 2013). DNA repair mechanism is a complex process and understanding it along with the heterogeneity of tumor cells makes it more complicated for us to evaluate the effect of genetic polymorphism on the repair mechanism of DNA as well as response of patients to chemotherapy due to these repair mechanism. The present study did not show any significant association of the *XPC* genetic variants with the patients treated with the platinum based chemotherapy. This means our results do not support the possibility that *XPC* genetic polymorphism involved in DNA repair capacity may affect treatment response and did not have any influence on patients survival. Our results were in agreement with the findings of Matakideu *et al.* (2006) as they suggested no significant association between the lung cancer patients with platinum based chemotherapy and both the *XPC* genetic variants. Our study was in disagreement with the study by Kang *et al.*, 2012 as they suggested association of DNA repair gene *XPC* with clinical outcome for Epithelial ovarian cancer treated with platinum based chemotherapy. The possible reason behind such results may be the low sample size of the study or the ethnic background of the population.

6.2 XPA polymorphisms and their association with lung cancer predisposition and Overall Survival

6.2.1 XPA polymorphism and lung cancer risk

Lung cancer patients were observed to have 2-fold risk of acquiring risk in subjects with variant genotype (GG) (**p=0.0007**) showing A23G polymorphism. A less pronounced risk was observed for combined genotype (AG+GG) (**p=0.03**) when compared with variant genotype (GG). The results in the present study were in accordance with the studies conducted on East Asians (Qian *et al.* 2011), Caucasians (Butkiewicz *et al.* 2004), Mixed Americans (Wu *et al.* 2003), Danish (Vogel *et al.* 2005) and African American population (Wu *et al.* 2003). Lou *et al.*, 2014 suggested that A allele of rs1800975 (A23G polymorphism) might be a risk factor for lung cancer in the East Asians and hence supported our study. A study by Kiyohara *et al.* 2010 also

explained a strong relationship of A23G polymorphism with increased risk of lung cancer and therefore, was in agreement with our results. The study carried by Wang *et al.* (2007) reported association of GG allele with the increased risk of oral cancer. On the above lines, the study by Pan *et al.* (2009) observed an association of homozygous variant genotype with increased risk of esophageal cancer. Lou *et al.*, 2014 also suggested that variant genotype might be a danger factor for lung cancer in the East Asians population having A23G polymorphism and therefore, was in agreement with our study. Three more studies on Caucasian population suggested correlation of double variant allele (GG) with increased risk of colorectal cancer (Jelonek *et al.* 2010, Gil *et al.* 2011, Hansen *et al.* 2007). Studies conducted on Polish (Jelonek *et al.* 2010) and population of mixed ethnicity (Crew *et al.* 2007) showed risk for breast cancer in subjects carrying mutant genotype (GG) and was in agreement with the present study. It was observed that few studies conducted on East Asians (Park *et al.* 2002, Zeng *et al.* 2013) and Caucasians (Sakoda *et al.* 2012) contradict the present study and suggested protective effect for A23G polymorphism. A study by Zheng *et al.*, 2013 differ from our results as they emphasized that no association was observed between XPA A23G polymorphism and lung cancer susceptibility among Asian population. Further, studies by De Ruyk *et al.*, 2007 and Raaschou-Nielson *et al.*, 2008 suggested no association between XPA polymorphism and lung cancer in Caucasian population and was in disagreement with the present study. This varied response from population to population might be due to the ethnic difference in the genetic background and the environment they reside in could be plausible reason behind such inconsistency.

The other polymorphism of XPA *i.e.* G709A polymorphism expressed a strong protective effect for heterozygous (GA) (***p*=0.001**) and combined (GA+AA) genotype (***p*=0.004**). Arg228Gln (G709A) is a rare polymorphism which has been considered as a XPA mutation resulted out of DNA repair defects (Porter *et al.* 2005). Arg228Gln is situated in exon VI which barely defines the TFII H binding domain of XPA. Any change in binding affinity with TFII H may lead to improved NER kinetics (Pleschke *et al.* 2000). G709A lies within agreement with a poly (ADP-ribose) covering 215-237 amino acid residues and any change in the non-conservative Arg228Gln disrupts this conserved sequence (Pleschke *et al.* 2000). In a study by Pleschke *et al.* (2000) they suggested that formation of poly (ADP-ribose) takes place in response to DNA damage which on binding with XPA inhibits NER pathway and any disruption in poly (ADP-

ribose) binding site might prevent the inhibition of *XPA*. Hence, result into improved DNA repair.

We also evaluated association of *XPA* polymorphism with lung cancer histological subtypes. Histological stratification analysis for A23G polymorphism projected mutant (GG) genotype as a highest risk factor among all the three histological subtypes. However the risk for this genotype was more prominent in SCLC patients (**p=0.0007**). This difference might be due to differences in carcinogenesis pathway involved in histological subtypes of lung cancer (Marchand *et al.* 1998, Gu *et al.* 1999). A meta-analytic study conducted by Lou *et al.* (2014) observed concordance with the present study and showed correlation of double variant allele towards risk for SQCC and ADCC. A study on Caucasian (Butkiewicz *et al.* 2004) and Chinese (Qian *et al.* 2011) population also supported the finding in which our results where both the studies have shown risk for the variant genotype in SQCC subjects. However, a study carried out by Park *et al.* (2002) on Korean population was in disagreement with the present study suggesting a protective effect for *XPA*23G polymorphism. G709A polymorphism showed a significant protective effect for heterozygous genotype in SQCC (**OR=0.45; p=0.001**) and SCLC (**OR=0.60; p=0.03**) patients. A significant protective effect in subjects with combined genotype (GA+AA) for SQCC was observed (**p=0.004**). The functional significance of G709A polymorphism is still unknown as any prior study on this polymorphism has not been reported.

6.2.1.2 *XPA* polymorphism and its association with lung cancer risk regarding smoking as an important parameter

Smokers depicted a 3-fold risk for lung cancer (**p=0.0002**) for double variant allele (GG) for A23G polymorphism. When further segregated on the basis of pack years, the variant genotype (GG) showed adverse risk for acquiring the disease (**p=0.0004**) amongst light smokers (pack years ≤ 25). Our results were in agreement with the study conducted on Danish (Raaschou-Nielsen *et al.* 2008), Korean (Qian *et al.* 2011) and Chinese (Wu *et al.* 2003) population suggesting a high risk for acquiring lung cancer for subjects carrying GG genotype. Although our findings are bit contradictory as already known that homozygous variant allele (GG) had 13% higher DNA repair capacity than homozygous wild allele (AA) (Wu *et al.* 2011). As lung cancer patients with poor DNA repair capacity were compared with healthy controls and this risk might be modulated by the smoking status (Spitz *et al.* 2001).

6.2.2 XPA polymorphism and overall survival in lung cancer patients

The overall survival in lung cancer patients and its association with *XPA* polymorphism was evaluated using univariate Kaplan and multivariate Cox regression model. Our results were in agreement with the study carried on Asian (Su *et al.* 2015), Caucasian (Shen *et al.* 2007) and Polish (Butkiewicz *et al.* 2011) population suggesting no association between A23G polymorphism and lung cancer risk. *XPA* A23G polymorphism located in 5'UTR region might influence transcription of *XPA* which plays an important role in NER pathway. This substitution could also remove a CpG site that interferes with the methylation process of *XPA* promoter as well as regulation of *XPA* expression (Candiloro *et al.*, 2009). The *XPA* allele could also be in linkage disequilibrium with allele of adjacent gene and might be susceptible to survival outcomes. Further when the lung cancer patients were treated with different chemotherapy regimens, no significant association was observed between *XPA* polymorphism in the present study and overall survival in lung cancer patients treated with those regimens. A study by Monzo *et al.*, 2005 suggested association of GG genotype of A23G polymorphism as the most important factor in chemoresistance and depicts its relationship with overall survival, but it differs from the results of the present study. There was also no relationship was observed among *XPA* polymorphic variants and clinical response. A study by Carles *et al.* 2006 differs from the present study and found association of A/G genotype with poor response to radiotherapy in patients with stage I to II head-and-neck cancer. Clinical impact on prognosis for *XPA* polymorphism was also demonstrated in patients with cancers of the bladder and colorectum by Gu *et al.*, 2005 and hence was in disagreement with the present study.

6.3 XPD polymorphisms and its role in lung cancer risk and overall survival

6.3.1 XPD polymorphisms and their association with lung cancer risk

Overall our study exhibited no significant risk towards lung cancer for *XPD* Asp³¹²Asn and *XPD* Arg¹⁵⁶Arg genotypes, however on the contrary the *XPD* Lys⁷⁵¹Gln mutant genotype (CC) was found to be strongly associated towards risk for lung cancer (p=0.01). C variant allele shows lung cancer risk as suggested by Xing *et al* (2002), Sturgis *et al* (1998) and Zhou *et al* (2002), which was similar to the present study. A study on Danish (Vogel *et al.*, 2006), Chinese (Liang *et al.*, 2003), Finish (Misra *et al.*, 2003) and Norwegian (Zienolddinny *et al.*, 2006) populations

have further supported our data, where the results also documented a positive association of mutant genotype (CC) towards risk for lung cancer. Vogel *et al* (2006) and Qiao *et al* (2002) further emphasized on the presence of lower DRC in lung cancer patients with mutant *Gln/Gln* (CC) genotype. When segregated on the basis of histological subtypes, *Lys*⁷⁵¹*Gln* polymorphism have a pronounced effect towards susceptibility for all the three subtypes of lung cancer *i.e.* ADCC (p=0.03), SQCC (p=0.008) and SCLC (p=0.02). Previous studies conducted on Chinese (Yin *et al.*, 2005) and Spanish (Lopez-Cima *et al.*, 2007) populations suggesting a similar risk towards ADCC subjects with variant genotype for *Lys*⁷⁵¹*Gln* polymorphism.

A lack of association for the *XPB Asp*³¹²*Asn* polymorphism towards risk for lung cancer was observed in our study which was supported by Lopez-Cima *et al* (2007) and Yu *et al* (2004). On the contrary some epidemiological studies have supported an association for the *G312A* polymorphism towards lung cancer (Strugis *et al.*, 1998; Xing *et al.*, 2002; Hou *et al.*, 2002). When elaborated on basis of histology the *Asp*³¹²*Asn* polymorphic variant was not associated with any of the histological subtypes of lung cancer. Hou *et al* (2002) and Qiao *et al* (2002) reported that the two common variant alleles (located in codons 751 and 312) of the *XPB* gene might lead to reduced repair of DNA adducts, and promotion of tumor formation.

*Arg*¹⁵⁶*Arg* polymorphism did not show any significant risk for lung cancer risk in North Indians. However, *XPB Arg*¹⁵⁶*Arg* heterozygous genotype (CA) exhibited a strong protective effect among SQCC patients (p=0.04). Yin *et al* (2005), Strugis *et al* (1998) and Catana *et al* (2012) showed similarity with the present study whereas studies on Danish (Dybdahl *et al.*, 1999), Caucasian (Dybdahl *et al.*, 1999) and American populations (Caggana *et al.*, 2001) varied by emphasizing on a significant risk for basal cell carcinoma and gliomas. *XPB Arg*¹⁵⁶*Arg* polymorphism does not show any amino acid change and hence, the enzymatic function of this protein is unlikely to be affected by the mutation, this polymorphic substitution would rather alter rate of translation by changing codon usage or by reducing *XPB* protein level of the gene leading to mRNA instability (Dybdahl *et al.*, 1999). Catana *et al* (2002) illustrated significant risk for ADCC subtype with the *XPB Arg*¹⁵⁶*Arg* variant genotype while there was no association was observed for SQCC patients.

6.3.1.2 MDR and CART analysis for *XPB*

The non-parametric methods like MDR and CART analysis were used to evaluate SNP-SNP interaction in *XPB* variants. A significantly strong synergistic interaction between *XPB* *Asp³¹²Asn* and *Arg¹⁵⁶Arg* consistent with logistic regression analysis was found in MDR analysis. *XPB* *Lys⁷⁵¹Gln*, *Asp³¹²Asn* and *Arg¹⁵⁶Arg* assessed as the best three factor risk model while *Lys⁷⁵¹Gln* emerged as the highest risk factor. *Mei et al* (2014) showed consistence with our results depicting a close association of *Lys⁷⁵¹Gln* polymorphism with lung cancer risk using MDR analysis. CART analysis was also applied on our data showing first split at *Arg¹⁵⁶Arg* followed by next split at *Asp³¹²Asn* and *Lys⁷⁵¹Gln*. CART analysis further distributes risk group present on different terminal nodes and risk is calculated for every node which help us to evaluate risk factor as mentioned in the previous studies conducted on NER genes (*Chen et al.*, 2007).

6.3.2 *XPB* polymorphisms and their association with overall survival

Polymorphic variants of *XPB* do not suggest any association with OS in lung cancer patients. Earlier studies conducted on Asian (*Provinco et al.*, 2012; *Liu et al.*, 2011) and Spanish (*Liu et al.*, 2011) subjects proposed *Lys⁷⁵¹Gln* polymorphism as the genetic determinant for survival in lung cancer patients treated with platinum based chemotherapy. When segregated on the basis of histology a protective effect was observed for individuals with SCLC carrying mutant (CC) genotype (HR: 0.49; p=0.04) for *XPB* *Lys⁷⁵¹Gln* polymorphism. Our results were supported by *Li et al* (2013), *Ludovini et al* (2011) and *Mathiaux et al* (2011) who also did not observe any association between OS and *XPB* *Lys⁷⁵¹Gln* polymorphism in NSCLC patients of Caucasian ethnicity. *Yao et al* (2009) and *Zhang et al* (2012) further validated that there was no association between *Lys⁷⁵¹Gln* polymorphism and OS in Asian subjects diagnosed with NSCLC. In the present study, there was no significant relationship observed between OS in lung cancer patients and *XPB* *Asp³¹²Asn* polymorphism. Studies by *Li et al* (2010) and *Wu et al* (2012) were consistent with our study suggesting no significant relationship between *Asp³¹²Asn* polymorphism and OS in lung cancer patients of Asian and Caucasian ethnicity. When differentiated on the basis of histological subtypes, for *Asp³¹²Asn* polymorphism, SCLC patients

with AA genotype (mutant) were associated with poor survival of 1.7 months and with a high death rate among diseased subjects (HR: 13.38; p=**0.0001**).

6.3.2.1 XPD and overall survival in lung cancer patients treated with different regimens

Most cancer patients with specific type of cancer and stage receive almost same treatment. The study also analyzes role of *XPD* polymorphic variants with OS in patients treated with different chemotherapy regimens. Patients treated with irinotecan-cisplatin/irinotecan-carboplatin regimen demonstrated a trend of better response (25.2 months) with extremely low hazard ratio (HR: 0.21; p=**0.002**) for patients carrying mutant genotype (CC) for *Lys⁷⁵¹Gln* polymorphism when compared with subjects carrying wild genotype receiving same treatment. The subjects carrying mutant genotype (AA) with *Asp³¹²Asn* polymorphism exhibited a poor survival (1.86 months) with high hazard ratio (HR: 16.46; p=**0.003**) when treated with irinotecan cisplatin/irinotecan carboplatin regimen. As per the results analyzed for irinotecan treatment in previous studies, they have demonstrated efficacy in the treatment of small cell lung cancer. It also appears to show promising activity in NSCLC patients producing a response rate upto 32% (Chen *et al.*, 2009). When administered in combination with cisplatin/carboplatin drug, a response rate of 25%-56% was detected in patients of American ethnicity who received treatment in stage II and III of the disease and thus showing median survival time of 9-12 months (Kinoshita *et al.*, 2004). A study in Japanese population reported a response rate of 36% for advanced NSCLC patients treated with irinotecan cisplatin/irinotecan carboplatin regimen (Fukuda *et al.*, 2004). DNA repair gene like *ERCC2* has been majorly associated with response in other platinum based chemotherapeutic agents (Pillot *et al.*, 2006). Therefore, no previous study has been reported on association of NER gene *XPD* with OS for lung cancer patients treated with irinotecan cisplatin/irinotecan carboplatin regimen except a single study on colorectal carcinoma. Artac *et al* (2010) emphasized on association of *Gln/Gln* genotype with a high risk of mortality in colorectal carcinoma patients who were treated with the above regimen and hence it does not support the findings in the present study. Lung cancer patients carrying heterozygous genotype (GA) for the *Asp³¹²Asn* polymorphism and treated with pemetrexed cisplatin/pemetrexed carboplatin regimen showed a median survival (8.8 months) with significant two fold hazard ratio (p=**0.01**). Pemetrexed is a multitargeted antifolate drug which has been approved as first line treatment for patients with NSCLC in combination with cisplatin/carboplatin and work as a

single agent for relapsed chemotherapy or chemotherapy refractory NSCLC after platinum containing chemotherapy (Metro *et al.*, 2011). Pemetrexed undergoes intracellular activation which is essential for its anti-proliferative activity. This regimen is reported to be associated with effective treatment of patients diagnosed with NSCLC (Metro *et al.*, 2011).

6.4 XPG polymorphisms and their association with lung cancer pre-disposition and overall survival

6.4.1 XPG polymorphisms and their association with lung cancer susceptibility

XPG acts as an endonuclease to participate in two incision steps in NER pathway, dysfunction of which leads to the excision repair deficiency. *XPG* cleaves the DNA strand at the 3' side of the damaged site, and stabilizes the DNA repair complex to the damaged DNA with *ERCC1/XPF* complex by making 5' incision. Many *XPG* gene polymorphisms can change the ability of the encoded enzyme to repair the DNA damage. Genetic variants may modulate susceptibility to lung cancer through complex gene-gene and gene-environment interactions. Evidence in the present study elaborates the role of *XPG* gene towards lung cancer resistance. *Phe*⁶⁷⁰*Leu* conferred a protective effect towards lung cancer in North Indian patients. A significant protective effect was observed in patients carrying heterozygous (OR=0.35, **p<0.0001**) and mutant (OR=0.18, **p=0.0001**) genotypes for lung cancer. As far our knowledge this SNP has not been much explored for analyzing its association with lung cancer. A study on Finnish population reported a lack of association of *Phe*⁶⁷⁰*Leu* polymorphism with breast cancer risk (Kumar *et al.*, 2003). Only a single study illustrating role of *XPG Phe*⁶⁷⁰*Leu* polymorphism and its association with cancer has been reported till now.

On the contrary for the *Asp*¹¹⁰⁴*His* polymorphism earlier studies have exhaustively studied its role in tumor genesis. It is located outside the conserved domains of the gene, the Asp to His amino acid change results in functional difference in the *XPG* gene at codon 1104 (Klungland *et al.*, 1999). Our results emphasize on the a lack association of *Asp*¹¹⁰⁴*His* polymorphism towards lung cancer. A study conducted by Shen *et al* (2005) showed similarity with our results and suggested no significant association between *Asp*¹¹⁰⁴*His* polymorphism and

lung cancer. Lavanya *et al* (2015) also reported no association of *Asp¹¹⁰⁴His* polymorphism towards breast cancer susceptibility in Indian population. On the same lines, studies conducted in SCCHN and gastric cancer also reported a lack of association with *Asp¹¹⁰⁴His* polymorphism (An *et al.*, 2007; Molina *et al.*, 2008). But there are many studies which were in disagreement with the present study concluding association of *Asp¹¹⁰⁴His* polymorphism with cancer susceptibility. Kumar *et al* (2003) depicted a strong association of mutant genotype (AA) showing *Asp¹¹⁰⁴His* with lung cancer risk among Europeans. A study on African American (Cheng *et al.*, 1998) and Asian (Cheng *et al.*, 1998) population have also conferred a positive association of *Asp¹¹⁰⁴His* polymorphism towards lung cancer susceptibility. Another study in Korean lung cancer patients depicted a strong association of the mutant genotype (*His/His*) for conferring a protective effect towards lung cancer (Jeon *et al.*, 2003).

Histological analysis exhibited a prominent protective effect for the double variant allele (AA) in both ADCC (OR=0.14; **p=0.008**) and SQCC (OR=0.14; **p=0.008**) subtypes. For *Asp¹¹⁰⁴His* polymorphism no such relationship can be established with histological subtypes of lung cancer. Various evidences predicted histological subtypes of the lung cancer are determined by the initiating agent to which an individual is exposed (Hoffman *et al.*, 1993). As polyaromatic hydrocarbons present in tobacco smoke have a strong relationship with ADCC, SQCC and SCLC, this result might have been due to the fact that *XPG* plays an important role in DNA damage repair caused by these polyaromatic hydrocarbons. This lack of statistical significance in case of *Asp¹¹⁰⁴His* polymorphism might be due to low sample size or difference in ethnicities.

6.4.1.2 *XPG* polymorphism and their association with lung cancer risk in smokers

Level of carcinogen exposure defines its relationship with genetic susceptibility to lung cancer (Jeon *et al.*, 2003). Therefore, association of *XPG* variants with lung cancer susceptibility was examined amongst smokers. Smokers who were lung cancer patients and carrying the mutant allele for *Phe⁶⁷⁰Leu* polymorphism were observed to be protected towards disease ($p<0.0001$). For *XPG Phe⁶⁷⁰Leu* polymorphism, a protective effect was observed amongst smokers for all the three variant genotypes *i.e.* heterozygous (CA) (OR: 0.32; **p<0.0001**), mutant (AA) (OR: 0.22; **p=0.02**) and combined genotype (CA+AA) (OR: 0.31; **p<0.0001**). When smokers were dichotomized into two groups based on the pack years of smoking, a protective effect was observed in both the light and heavy smokers for *Phe⁶⁷⁰Leu* variant. There are possible reasons

behind such findings. At high doses of carcinogen exposure the genetic differences in risk may be smaller until environmental factor overpowers the genetic predisposition (Shiromani *et al.*, 1994). Such results may be attributable to the chance due to low sample size (Hoffman *et al.*, 2003). The study by Jeon *et al* (2003) supported the present study and suggested a protective effect amongst smokers for *XPG* polymorphism. Whereas He *et al* (2012) showed disagreement with the study by emphasizing on association of *XPG* variants with high risk of lung cancer amongst smokers.

6.4.2 *XPG* polymorphisms and their association with overall survival

There was no significant association observed for both *XPG* variants (*Phe*⁶⁷⁰*Leu*, *Asp*¹¹⁰⁴*His*) and OS. When categorized according to histological subtype, a poor survival for SCLC patients with mutant genotype (AA) was reported showing 2.5-fold hazard ratio for *Asp1104His* polymorphism. We further evaluated association between *XPG* variants and OS according to chemotherapeutic regimen. It is known that selection of chemotherapeutic regimen could improve response and hence overall survival of cancer patients. A longer survival period of 32.5 months was observed in patients having mutant genotype (AA) and treated with irinotecan-cisplatin/irinotecan-carboplatin regimen for *Phe*⁶⁷⁰*Leu* polymorphism. A lower death rate was observed in mutant subjects (AA) treated with the above regimen for *Phe*⁶⁷⁰*Leu* variant of *XPG* (HR=0.04). Whereas for *Asp*¹¹⁰⁴*His* polymorphism, a higher death rate was observed in patients carrying mutant genotype (CC) treated with irinotecan-cisplatin/irinotecan-carboplatin regimen (HR=3.01; p=0.06). Irinotecan consists of SN-38 active metabolite and interacts with topoisomerase I and forms a stable complex with DNA (Fukuda *et al.*, 2004). Irinotecan has always demonstrated its efficacy in SCLC treatment but it also show a promising effect in NSCLC patients producing a response rate to 32% (Fukuda *et al.*, 2004). No previous study has been reported till now on association of *XPG* gene and OS for lung cancer patients treated with irinotecan-cisplatin/irinotecan-carboplatin regimen. Pemetrexed is a multitargeted antifolate drug which has been approved as first line treatment for patients with NSCLC in combination with cisplatin/carboplatin and work as a single agent for relapsed chemotherapy or chemotherapy refractory NSCLC after platinum containing chemotherapy (Metero *et al.*, 2011). *XPG Asp*¹¹⁰⁴*His* also showed a higher mortality rate in patients carrying combined genotype (GA+AA) treated with pemetrexed-cisplatin/pemetrexed-carboplatin (HR=1.83; p=0.06).

DNA repair is essential for safeguarding the genome stability and integrity and *XPG* protein is required for making incision at the 3' site of DNA damage during NER pathway (Wakasugi et al., 1997). NER is associated with resistance to platinum agents, and reduced NER might be related with the increased platinum agent sensitivity. Therefore, *XPG* polymorphism may affect the clinical outcomes of cancer patients. Previous studies have suggested association of *XPG* polymorphism with predictive clinical response of various cancers (Zhou et al., 2015). The current study did not show any association of *XPG* polymorphism with clinical outcomes. Our study was supported by Bartolucci et al (2009), Sun et al (2009), Butkiewicz et al (2011) and Zhang et al (2012) who suggested no significant association between *XPG* polymorphism and clinical outcomes in NSCLC patients. A study on Chinese population predicted association of *XPG* polymorphism towards prognosis of NSCLC patients (Metro et al., 2011).

6.5 *XPF* polymorphisms and their association lung cancer susceptibility and overall survival

6.5.1 *XPF* polymorphisms and their role in lung cancer predisposition

XPF polymorphic variants and their role in different cancers has not been much explored in prior studies. An earlier study on Caucasian population illustrated a strong association of 673C>T variants with increased bladder risk which was found to be concordant with the present study. Studies by Yu et al. (2010) and Shao et al. (2008) illustrated association of 673C>T polymorphism with decreased lung cancer risk and therefore vary from the findings of our study. Liu et al. (2015) negated any association of 673C>T variants towards esophageal cancer risk and was in disagreement with this study. Other polymorphic variant of *XPF* i.e. 11985 A>G has been least explored till date and to our knowledge, no prior data on 11985A>G polymorphism and its role in cancer risk has been reported. A single study by Liu et al. (2015) has been reported with does not show any role of 11985 A>G, rs254942 polymorphism towards lung cancer risk and hence, supported our results.

*Arg*⁴¹⁵*Gln* polymorphism has been the most widely studied *XPF* SNP. The present study suggested no significant relationship between *Arg*⁴¹⁵*Gln* genotypes and lung cancer. An earlier

study by Krupa *et al.* (2001) was in agreement with our results suggesting a significant association between *Arg*⁴¹⁵*Gln* polymorphism and laryngeal cancer in Polish population. Another study on Spanish (Milne *et al.*, 2006) and American (Rajaraman *et al.*, 2008) population further elaborated inconsistent association of *Arg*⁴¹⁵*Arg* polymorphism with breast cancer risk and support our findings. Study by Shi *et al.* (2015) emphasized on the strong association of *Arg*⁴¹⁵*Arg* polymorphism with increased lung cancer risk and hence, deviates from the present study.

6.5.1.2 XPF polymorphism and association with lung cancer among smokers

Many previous studies have revealed a significant gene-environment interaction resulting in pathogenesis of lung cancer. NER pathway is important in modulating lung cancer susceptibility (Sancar *et al.*, 2004). NER is the primary mechanism for repairing helix distorting bulky DNA adducts which are generated due to cigarette smoke (DeMarini *et al.*, 1995; Asmani *et al.*, 1997). A NER protein plays an important role in repairing large DNA lesions and adducts (Torres-Ramos *et al.*, 2000). Genetic variants present in the promoter region of *XPF* might lead to substrate activation caused due to cigarette smoke and act as contributor to cancer susceptibility. The present study emphasize on the protective effect shown by *Arg*⁴¹⁵*Arg* polymorphism among non-smokers. Further no association was observed for smokers in *XPF* polymorphisms and lung cancer risk except for 673C>T, rs316028 polymorphism suggesting an increased lung cancer risk in heavy smokers (**p=0.03**). Yu *et al.* (2000) suggested that there was no interaction between 673C>T polymorphism and smoking status while Shao *et al.* (2008) focused on association of *XPF* genotypic variants with increased lung cancer risk among smokers. Hence, both the above studies were in disagreement with the present study. A study by Liu *et al.* (2015) further suggested a decreased risk of esophageal cancer among non-smokers showing 673C>T polymorphism and therefore was in disagreement with present study.

6.5.2 *XPF* polymorphisms and their role in overall survival of lung cancer patients

XPF protein in NER pathway play an essential role in removing platinum based compounds. *XPF* binds with *ERCC1* to form a complex and then binds with the DNA strands during DNA repair pathway (Alexander *et al.*, 2012). On considering the crucial rate limiting factor in NER, *XPF* genes lower expression might act as a genetic susceptibility factor in cancer development and a prognosis risk factor after chemotherapy in several cancers such as lung cancer (Alexander *et al.*, 2012). Therefore, when evaluated role of *XPF* polymorphism in the survival of lung cancer patients treated with platinum based chemotherapy, there was no significant association between *XPF* polymorphic variants and OS observed for the present study. ADCC patients with 673 C>T polymorphism have a low hazard ratio showing survival period of 14 months in patients with variant genotype (TT). A poor survival of 4 months was predicted in SCLC patients with 11985 A>G polymorphism for each with includes heterozygous (AG), mutant (GG) and heterozygous+mutant (AG+GG) type against the reference genotype (AA) (MST=11.6 months). ADCC patients showing *Arg*⁴¹⁵*Asn* polymorphism and having heterozygous genotype (GA) were associated with a moderate survival of 8.03 months.

6.5.2.1 *XPF* and its role in overall survival in lung cancer patients treated with different chemotherapy regimens

DNA repair genes plays pivotal role in carcinogenesis and platinum function (Tsuruo *et al.*, 1988). A shorter survival period of 4.30 months was observed in patients with heterozygous genotype (GA) when treated with irinotecan-cisplatin/carboplatin regimen for 11985 A>G polymorphism. A higher death rate was observed in all the genotypic variants of 11985 A>G (rs254942) polymorphism treated with the above regimen for i.e. heterozygous (HR=2.66), mutant (HR=4.73) and heterozygous+mutant (HR=2.64). Interpretation regarding results could be a complicated issue as it might emphasize on the fact that genotype conferred decreased DNA repair capacity, reaching at any particular conclusion might be immature. These polymorphism might not be functional but could be a marker of a specific haplotypes of section of chromosome 16, where other DNA repair genes also reside. These haplotypes might be associated with the differential activity because of variations in DNA repair genes (Fukuda *et al.*, 2004). This study

might generate a necessary data to move forward to decide which chemotherapy could be individualized as per the genetic constitution of a patient. The careful consideration of the repair pathway involved in the metabolism of commonly used chemotherapy drugs would be the most critical part and provides the basis for treatment of advanced lung cancer cases which includes combination of cytotoxic agents.

We can develop a much clearer understanding of genetic variations among individuals and their role in defining treatment tolerance as well as response and its overall outcome. The present study showed no significant association between *XPF* variants and clinical response. A study on British population denied any association of *XPF* polymorphism with platinum based chemotherapy response (Fareed *et al.*, 2012). A study by Shi *et al.* (2015) and Huang *et al.* (2014) also agreed with the present study expressing no significant association between *XPF* polymorphisms and clinical outcome in NSCLC patients.

6.6 Gene-gene and gene-environment interaction

Our results have emphasized upon the existence of gene-gene interactions in NER pathway modulating risk of lung cancer. Dual combinations between polymorphic variants *XPC Lys⁹³⁹Gln* and *XPB Lys⁷⁵¹Gln* were associated with increased lung cancer risk. While combinations like *XPF Arg⁴¹⁵His* and *XPG Phe⁶⁷⁰Leu*, *XPF Arg⁴¹⁵His* and *XPB Lys⁷⁵¹Gln*, *XPF 11985 A>G* and *XPG Phe⁶⁷⁰Leu*, *XPF 11985A>G* and *XPC Ala⁴⁹⁹Val*, *XPF 673C>T* and *XPG Phe⁶⁷⁰Leu*, *XPF 673C>T* and *XPC Lys⁹³⁹Gln*, *XPG Phe⁶⁷⁰Leu* and *XPC Ala⁴⁹⁹Val* showed a protective effect towards the disease. Individual contribution of these polymorphism on the lung cancer risk can be modest but when these polymorphism act together in different combinations they might affect lung cancer risk through gene-gene and gene-smoking interaction. Logistic regression analysis is the commonly used method for modeling a relationship between the predictor factor such as genotypes and clinical outcome (Templeton, 2000). But logistic regression unlike other parametric statistical analysis is less practical when it comes to dealing with high dimensional data (Schlichting and Pigliucci, 1998). When high order interactions are being modeled there supposed to be many contingency table cells which have no observation present in them resulting into large coefficient estimates and standard error. The only solution to this problem will be to collect large number of sample and then perform robust estimation of interaction effects (Hosmer and Lemeshow, 2000). This magnitude might often result into a prohibitive response. Another

possible solution to this problem is to develop a new statistical and computational method that has better approach in detecting multi-locus effect in small sample size. To overcome this issue, MDR was used to characterize and detect high order gene-gene and gene-environment interactions among case-control studies of relatively small sample size (Hosmer and Lemeshow, 2000). Therefore, this combinatorial partitioning method was used where a data reduction was applied for the explanatory analysis of qualitative traits.

6.7 MDR and CART analysis in genetic variants of NER

MDR is the method which has been applied to reduce the dimensionality caused by simultaneously fitting multiple genes and their interaction into models (Concato *et al.*, 1993). In MDR, model did not detect any interaction effect between the SNPs but the main effect of one SNP is detected. MDR method also enables the choice of the best model with interaction effect with other SNPs. *XPG Phe⁶⁷⁰Leu* as the best one factor model in the present study ($p=0.0001$). The study of complex interactions was more explanatory in our study of the NER genes. *XPC Lys⁹³⁹Gln*, *XPA A23G*, *XPG Phe⁶⁷⁰Leu*, *XPG Asp¹¹⁰⁴His* was analyzed as the best four factor model ($p=0.0001$). When differentiated on the basis of the histological subtype, *XPG Phe⁶⁷⁰Leu* was the judged as the risk factor in the all the three histological subtype. Among SCLC patients *XPG Phe⁶⁷⁰Leu*, *XPA A23G*, *XPB Lys751Gln*, *XPA G709A*, *XPD Arg⁴¹⁵Gln*, *XPD Arg¹⁵⁶Arg* was the best six factor model with CVC of 100 and $p=0.005$. When gene-environmental relationship was analyzed by applying MDR, it illustrated smoking as the primary factor related with lung cancer risk ($p<0.0001$). The SNPs detected in these methods *i.e.* MDR and CART are more likely to be functional. This reveals that our results did not occur by chance but might be due to any biological links between genes present within the genetic pathway (Briollais *et al.*, 2007). The interactions detected in MDR are more diverse and less influenced by the main SNPs effects. Overall, our results suggested that use of multiple statistical approaches rather than a single methodology could be the best strategy to elucidate complex gene interaction and have generally different patterns. MDR method has some advantages such as it facilitates the simultaneous detection and characterization of multiple genetic loci. The other possible advantage of this method is that it is a non parametric statistical approach which is relying upon generalized linear model (Peduzzi *et al.*, 1996). Another big advantage of this method is that it follows no particular genetic model that means no mode of inheritance has been defined;

therefore, it can be directly applied upon case-control studies (Ripley *et al.*, 1996). The most important advantage of this method is that false-positive results are reduced due to multiple testing are minimized (Namkung *et al.*, 2009). The major limitation of this method are that it becomes computationally complicated when more than 10 polymorphism are evaluated (Brosh and Rotter, 2009). Secondly, MDR models are difficult to interpret (Ripley *et al.*, 1996).

Along with MDR method, CART algorithm was also applied for the adjustment of continuous properties of phenotypes. CART is a model free and non parametric method produces different results when compared with the linear mixed model (Hsu *et al.*, 2008; Chen *et al.*, 2014). *Phe*⁶⁷⁰*Leu* was identified as the prominent risk factor in both the methods. Although CART analysis did not reveal any high order interactions among 12 SNPs except a single split observed at *Phe*⁶⁷⁰*Leu*, giving a single node and making this polymorphism as the highest risk factor. The follow up logistic regression suggested that the markers did not show any multiplicative interaction effects among the studied polymorphism. When CART analysis was applied on histological subtypes, Only SCLC population depicted *XPD Lys*⁷⁵¹*Gln* as the individual risk factor for the disease. *XPD* gene encodes 5'-3' DNA helicase having single stranded DNA dependent ATPase activity (Samson *et al.*, 2011). It is an important basal transcription factor IIH (TFIIH) that has dual function that is transcription and DNA damage repair. *XPD* acts as the p53 binding site to the TFIIH, thus also plays pivotal role in apoptosis (Namkung *et al.*, 2009). Therefore, carrier of *XPD Lys*⁷⁵¹*Gln* variant allele were associated with almost 5-fold risk while *XPD Lys*⁷⁵¹*Gln* (W) in combination with *XPA A23G* depicts a 6-fold risk for SCLC (p=0.0001). Node 4 that consist of *XPC Ala*⁴⁹⁹*Val* (W), *XPA A23G* (M), *XPF 11985 A>G* (W), *XPA A23G* (M), *XPD Lys*⁷⁵¹*Gln* (W) almost show a 3-fold risk for the disease (p=0.02). The odds ratio subjected to have three, five and six risk allele were more than 1.40. Since etiology of complex disease like lung cancer involves multiple gene including various biological pathways as well as environmental factors. Effect of individual polymorphism will always be moderate and might have a limited predictive value in the assessment of disease. A combined effect of multiple risk variants might provide a more precise effect on risk group and could also suggest future direction to association studies.

Smoking is one of the main risk factor in lung cancer which was validated by the data mining tools in our study (Zhou *et al.*, 2002). In CART first split and in MDR best one factor model both

depicted smoking as the predominant risk factor. When logistic regression analysis was applied, it was suggested that *XPG Phe⁶⁷⁰Leu (M)* was associated with decreased risk for the disease. This represents that gene-smoking interactions might modify functions of NER genes. A parallel CART analysis on non-smokers was also applied; however we were not able to detect any high order interactions among the NER polymorphism. These results proved to be biologically plausible as NER pathway mainly takes part in repairing bulky DNA adducts caused due to smoking. Among non smokers, major damage is probably not caused because of bulky DNA adducts but due to base damages and DNA breaks, which includes base excision repair and double strand break repair (Chen *et al.*, 2007).

6.8 Gene-gene interaction and Overall survival

In the present study, the dual combinations of genotypes of 12 NER SNPs and their impact on overall survival in North Indian population was observed. The unfavorable genotypic combinations for 12 SNPs were also applied using Recursive partitioning method (STREE). The study implicated no significant association between overall survival in lung cancer patients and genotypic combinations. When sample combinations were applied on histological subtypes a prominent indication of association with poor survival in SCLC patients was observed. *XPF673 C>T* and *XPG Phe⁶⁷⁰Leu*, *XPF 11985A>G* and *XPC Ala⁴⁹⁹Val*, *XPF11985 A>G* and *XPB Arg¹⁵⁶Arg*, *XPF 11985 A>G* and *XPA G709A*, *XPF 11985 A>G* and *XPG Phe⁶⁷⁰Leu* were the dual combination which were related with the SCLC risk. Among SCLC patients when different combinations were applied among genotypes of different genes, a significant hazard ratio depicting a poor prognosis was observed. *XPF 11985A>G* and *XPB Arg¹⁵⁶Arg* shows a weaker survival among SCLC patients showing survival of only 3.3 months. The same polymorphism illustrated 4.8-fold hazard ratio indicating a high risk of death among SCLC patients. Although reason behind these histology dependent difference which has been conferred by these polymorphism are still unknown, but it might be associated with the difference in their carcinogenesis pathway among histological subtypes of lung cancer. Many studies suggested the histological subtype determination is related with individual's exposure to a scientific initiation agent. Therefore, various genetic factors might be responsible for such results (Wu *et al.*, 2012).

The simultaneous monitoring of association of large number of genes holds a great promise for better understanding of cancer biology as well as for development of accurate tumor treatment. We applied recursive partitioning tree based analysis for understanding the survival of overall survival of NER SNPs in lung cancer patients. The selection of optimal subset of genes has two related problems: first determining the number of genes to be selected and also determination of fact that which genes belong to the set. The partitioning method is able to incorporate these features in selection as a part of learning algorithm and also able to resolve these issues. Therefore, recursive partitioning method (STREE) was applied on survival data. The best results were observed at terminal node 6 showing *XPG Asp¹¹⁰⁴His (W)*, *XPD Lys⁷⁵¹Gln (H+M)*, *XPF Arg⁴¹⁵Gln (H+M)* associated with lower hazard ratio ($p=0.0007$) showing survival of 11.6 months when compared with the reference. The maximum nodes obtained through recursive partitioning method were associated with lower hazard ratio. The other terminal node which suggested a better survival was node 7 consisting of sample population consisting of *XPG Asp¹¹⁰⁴His (W)*, *XPD Lys⁷⁵¹Gln (H+M)*, *XPF Arg⁴¹⁵Gln (W)*, *XPF 11985 A>G (W)* genotypes with survival period of 10.8 months. The knowledge of correct genotype and their response to the treatment helps in supported future treatment endeavors. Although not much work has been done till now towards understanding the effect of combinatorial studies on survival and we are first one to perform such study in North Indian population.

The present study has several limitations that should be considered. First, there are chances of selection bias and information bias in the current data, all lung cancer patients were selected from a single hospital and therefore these cases may not be representative of patients from other population. The biasness could be reduced by adjusting the final analysis for gender, age, smoking status, ECOG, KPS and histology which could minimize the selection bias. Secondly, we have conducted investigations with a limited sample size.

CHAPTER-7

Conclusions

Findings from the present study

- *XPC* gene particularly *XPC Lys⁹³⁹Gln* modulates the risk of lung cancer and other polymorphic variant *XPC Ala⁴⁹⁹Val* might be the cause of the lung cancer but majorly shows the protective effect towards the risk of lung cancer in North Indian population.
- The survival analysis suggests no significant association between the *XPC* variants and survival of lung cancer patients. It further reveals that *XPC Lys⁹³⁹Gln* and genotypic *XPC Ala⁴⁹⁹Val* are not associated with an effective response to platinum based chemotherapy.
- *XPA A23G* polymorphism might be associated with increased lung cancer risk in variant genotype in the North Indian population. *XPA G709A* polymorphism modulates protective effect in North Indian population.
- *XPD Lys⁷⁵¹Gln* associated with increased risk of lung cancer whereas no significant association was observed for *XPD Asp³¹²Asn* and *XPD Arg¹⁵⁶Arg*.
- MDR analysis also revealed association of *XPD Lys⁷⁵¹Gln* polymorphism with increased risk of lung cancer.
- *ERCC2* functional SNPs (*Lys⁷⁵¹Gln* and *Asp³¹²Asn*) independently affect the OS of patients treated with irinotecan cisplatin/carboplatin regimen and may act as an important predictive parameter in survival of lung cancer patients.
- Our study provides statistical evidence that *XPG Phe⁶⁷⁰Leu* was associated with protective effect in lung cancer and its histological subtypes. *Asp¹¹⁰⁴His* polymorphism does not represent any association with lung cancer susceptibility.
- No significant association was observed for *XPG* variants with OS and clinical outcomes in North Indian population.

- *XPF* 673 C>T polymorphism might be associated with increased lung cancer risk. *11985A>G* and *Arg⁴¹⁵Gln* polymorphism does not show any association with lung cancer risk.
- The genetic variants of 673C>T polymorphism were associated with a better survival in ADCC patients while *11985A>G* polymorphism was associated with poor survival in SCLC patients.
- Lung cancer patients treated with irinotecan cisplatin/irinotecan carboplatin regimen indicated their association with shorter survival time and high risk of death in *11985A>G*, polymorphism.
- The combined analysis showed that individuals carrying multiple copies of risk variants have decreased risk for lung cancer. MDR analysis suggested prevalence of gene-gene interaction as well as gene-environment interaction. *XPG Phe⁶⁷⁰Leu* was predicted as the highest risk factor. CART analysis did not confirm gene-gene interaction but gene-smoking interaction was observed for NER SNPs.
- The combinatorial studies suggested that there was no association observed between polymorphic combinations of NER genes with overall survival in lung cancer patients. When applied on histological subtypes, the polymorphic were associated with poor survival showing a high hazardous effect in SCLC patients.
- Recursive partitioning method (STREE) when applied in combination among unfavorable genotypes emphasized on association of NER polymorphism combination with lower hazard ratio predicting them to be good prognosis factor.

In conclusion this is the first Indian study to highlight association of DNA repair genes altogether on such large size and also their correlation with clinico-pathological factors. The stratified analysis was also performed for smoking and histology to find out high risk posing subgroups. Subjects with single allelic variant for initial NER genes *XPC Lys⁹³⁹Gln* and *XPB Lys⁷⁵¹Gln* were observed to be positively correlated with lung cancer predisposition. Our study is the first attempt in DNA repair genes for lung cancer in Indian population to make use of multi-factorial approach (CART and MDR) to explore the complex interaction between the DNA repair genes and their relationship towards lung cancer risk. This is the first study in the North Indian region to analyze the relationship of SNPs with overall survival and clinical outcomes. We also analyzed the stratified clinical outcomes with genotype for the genetic variants in the

study. The present findings for overall and stratified groups might help in development of biomarkers both for diagnosis of lung cancer and its prognosis.

7.1 Future prospects of the study

DNA repair enzymes play an important part in removal of DNA damage induced by different carcinogens and anticancer drug. It is very likely that SNPs in DNA repair genes may play an important part in cancer susceptibility. These polymorphisms in the DNA repair pathways might serve as a platform for detecting and predicting individual susceptibility for lung cancer. It might also help in revealing the prognosis or the disease. This study would also help in construction of reliable SNP and haplotype analyses on different ethnic population like Indian which could help in predictive genotype-drug response in patients. Some remarkable results were observed in the present study when lung cancer patients treated with different chemotherapy regimens *i.e.* irinotecan-cisplatin/carboplatin regimen, pemetrexed-cisplatin/carboplatin regimen and their association with *XPD* polymorphism was analyzed. It is known that selection of chemotherapeutic regimen could improve response and hence overall survival of cancer patients. Personalized chemotherapy as per the molecular marker could help to improve the response rate as well as survival of lung cancer patients. This study in future can facilitate assessment of genetic variations of different polymorphisms and help patients to make therapeutic decisions for individualized therapy in advanced lung cancer patients.

REFERENCES

- Abbasi, R., Ramroth, H., Becher, H., Dietz, A., Schmezer, P., *et al.* (2009) Laryngeal cancer risk associated with smoking and alcohol consumption is modified by genetic polymorphisms in ERCC5, ERCC6 and RAD23B but not by polymorphisms in five other nucleotide excision repair genes. *Int J Cancer* 125; 1431–1439.
- Agalliu, I., Kwon, E. M., Salinas, C. A., Koopmeiners, J. S., Ostrander, E. A., *et al.* (2010). Genetic variation in DNA repair genes and prostate cancer risk: results from a population-based study. *Cancer Causes Control* 21; 289–300.
- Ahmad, A., Robinson, A. R., Duensing, *et al.* (2008). ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Mol Cell Biol.* 28; 5082–5092.
- Aihara, T., Kim, Y., Takatsuka, Y. (2002). Phase II study of weekly docetaxel in patients with metastatic breast cancer. *Ann Oncol.* 13; 286–92.
- Al-Abd, A. M., Alamoudi, A. J., Abdel-Naim, A. B., Neamatallah, T. A. and 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.
- Alexander, B. M., Wang, X. Z., Niemierko, A., *et al* (2012). DNA repair biomarkers predict response to neo-adjuvant chemoradiotherapy in esophageal cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 83; 164–171.
- An, J., Liu, Z., Hu, Z., Li, G., Wang, L., Sturgis, E. M., El-Naggar, A. K., Spitz, M. R. and Wei, Q. (2007). Potentially Functional Single Nucleotide Polymorphisms in the Core Nucleotide Excision Repair Genes and Risk of Squamous Cell Carcinoma of the Head and Neck. *Cancer Epidemiol Biomarkers Prev* 16(8).
- Ankathil, R. (2010). Tobacco, genetic susceptibility and lung cancer. *Tobacco Use Insights* 3, TUI–S2819.
- Ansari, J., Palmer, D., Rea, D. and Hussain, S. (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.

- Ardizzoni, A., Boni, L., Tiseo, M., Fossella, F. V., Schiller, J. H., Paesmans, M., et al. (2007). Cisplatin- versus carboplatin-based chemotherapy in first-line treatment of advanced non-small-cell lung cancer: an individual patient data meta-analysis. *J Natl Cancer Inst* 99; 847–857.
- Artac, M., Bozuk, H., Pehlival, S., Akcan, S., Pehlivan, M., et al. (2010). The value of XPD and XRCC1 genotypes polymorphism to predict clinical outcome in metastatic colorectal carcinoma patients with irinotecan based regime. *J. Cancer Res. Clin. Oncol.* 136(6); 803-809.
- Asahina, H., Kuraoka, I., Shirakawa, M., Morita, E. H., Miura, N., Miyamoto, I., Ohtsuka, E., Okada, Y. and Tanaka, K. (1994). The XPA protein is a zinc me tall oprotein with an ability to recognize various kinds of DNA damage. *Mutat. Res.* 315; 229-237.
- Bahl, A., Sharma, D. N., Julka, P. K., Rath, G. K. (2006). Chemotherapy related toxicity in locally advanced non-small cell lung cancer. *J Cancer Res Ther* 2; 14–6.
- Bai, Y., Xu, L., Yang, X., Hu, Z., Yuan, J., et al. (2007). Sequence variations in DNA repair gene XPC is associated with lung cancer risk in a Chinese population: a case-control study. *BMC Cancer* 7; 81.
- Bala, S., Gundeti, S., Linga, V. G., Maddali, L. S., Digumarti, R. R., Uppin, S. G. (2016). Clinicopathological features and outcomes in advanced nonsmall cell lung cancer with tailored therapy. *Indian Journal of Medical and Paediatric Oncology : Official Journal of Indian Society of Medical & Paediatric Oncology* 37(4); 242-250
- Bănescu, C., Iancu, M., Trifa, A. P., Dobreanu, M., Moldovan, V. G., Duicu, C., Tripon, F., Crauciuc, A., Skypnyk, C., Bogliș, A., Lazar, E. (2016). Influence of XPC, XPD, XPF, and XPG gene polymorphisms on the risk and the outcome of acute myeloid leukemia in a Romanian population. *Tumour Biol.* 37(7); 9357-66.
- Bartolucci, R., Wei, J., Sanchez, J. J., et al. (2009). XPG mRNA expression levels modulate prognosis in respected non-small-cell lung cancer in conjunction with BRCA1 and ERCC1 expression. *Clin. Lung Cancer* 10; 47-52.
- Batty, D. P., and Wood, R. D. (2000). Damage recognition in nucleotide excision repair of DNA. *Gene (Amst.)* 241; 193–204.

- Bau, D.T., Fu, Y.P., Chen, S.T., Cheng, T.C., Yu, J.C., Wu, P.E. and Shen, C.Y. (2004). Breast cancer risk and the DNA double-strand break end-joining capacity of nonhomologous end-joining genes are affected by BRCA1. *Cancer Res.* 64; 5013-5019.
- Behera, D. (2007). New approach to the treatment of lung cancer: the molecular targeted therapy. *INDIAN JOURNAL OF CHEST DISEASES AND ALLIED SCIENCES* 49(3); 149.
- Behera, D. (2012). Lung cancer in India. *Medical Update* Vol 22; Pg 401-407.
- Behera, D. and Balamugesh, T. (2004). Lung cancer in India. *Indian Journal of Chest Diseases and Allied Sciences* 46; 269-282.
- Benhamou, S., Sarasain, A (2002). ERCC2/XPD gene polymorphism and cancer risk. *Mutagenesis* 6; 463-469.
- Berman, D. W. and Crump, K. S. (2008). Update of potency factors for asbestos-related lung cancer and mesothelioma. *Critical reviews in toxicology* 38(sup1), 1–47.
- Bernard-Gallon, D., Bosviel, R., Delort, L., Fontana, L., Chamoux, A. et al. (2008). DNA repair gene ERCC2 polymorphisms and associations with breast and ovarian cancer risk. *Molecular Cancer* 7; 36.
- Bernstein, C., Bernstein, H., Payne, C. M. and Garewal, H. (2002). DNA repair/proapoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutat. Res.* 511; 145-178.
- Bernstein, C., Prasad, A. R., Nfonsam, V., Bernstein, H. (2013). DNA Damage, DNA Repair and Cancer. *New Research Directions in DNA Repair*.
- Bernstein, H., Payne, C. M., Bernstein, C., Garewal, H., Dvorak, K. (2008). Cancer and aging as consequences of un-repaired DNA damage. *In: New Research on DNA Damages*, 1; 1–47.
- Blankenburg, S., Konig, I. R., Moessner, R., et al. (2005). Assessment of 3 xeroderma pigmentosum group C gene polymorphisms and risk of cutaneous melanoma: a case-control study. *Carcinogenesis* 26; 1085–90.
- Booton, R., Ward, T., Heighway, J., Taylor, P., Power, F., et al. (2006). Xeroderma pigmentosum group D haplotype predicts for response, survival, and toxicity after platinum-based chemotherapy in advanced nonsmall cell lung cancer. *Cancer* 106; 2421–2427.

- Bowden, N. A., Ashton, K. A., Vilain, R. E., Avery, K., Iejda, K. A., Davey, R. J., Murray, H. C., Budden, T., Braye, S. G., Zhang, X. D., Hersey, P. et al. (2013). Regulators of global genome repair do not respond to DNA damaging therapy but correlate with survival in melanoma. *PLoS ONE* 8e; 70424.
- Briollais, L., Wang, Y., Rajendram, Y., Onay, V., Shi, E., Knight, J., et al. (2007). Methodological issues in detecting gene-gene interactions in breast cancer susceptibility: a population-based study in Ontario. *BMC medicine* 5; 22.
- Brosh, R. and Rotter, V. (2009). When mutants gain new powers: news from the mutant p53 field. *Nat Rev Cancer* 9; 701–713.
- Brueckner, F. and Cramer, P. (2007). DNA photodamage recognition by RNA polymerase II. *FEBS Lett.* 581; 2757-2760.
- Bunick, C. G., Miller, M. R., Fuller, B. E., Fanning, E., Chazin, W. J. (2006). Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. *Biochemistry* 45; 14965–14979.
- Butkiewicz, D., Popanda, O., Risch, A., Edler, L., Dienemann, H., Schulz, V., et al. (2004). Association between the risk for lung adenocarcinoma and a (–4) G-to-A polymorphism in the XPA gene. *Cancer Epidemiol Biomarkers Prev.* 13; 2242–6.
- Butkiewicz, D., Rusin, M., Enewold, L., Shields, P. G., Chorazy, M. M., Harris, C. C. (2001). Genetic polymorphisms in DNA repair genes and risk of lung cancer. *Carcinogenesis* 22; 593–597.
- Butkiewicz, D., Rusin, M., Harris, C. C., Chorazy, M. (2000). Identification of four single nucleotide polymorphisms in DNA repair genes: XPA and XPB (ERCC3) in Polish population. *Hum Mutat.* 15; 577–8.
- Butkiewicz, D., Rusin, M., Sikora, B. et al. (2011). An association between DNA repair gene polymorphisms and survival in patients with resected non-small cell lung cancer. *Mol. Biol. Rep.* 38; 5231-5241.

- Butkiewicz, D., Drosik, A., Suwiński, R., Krześniak, M., Rusin, M., Kosarewicz, A., Rachtan, J., Matuszczyk, I., Gawkowska-Suwińska, M. (2012). Influence of DNA repair gene polymorphisms on prognosis in inoperable non-small cell lung cancer patients treated with radiotherapy and platinum-based chemotherapy *Int J Cancer* 1; 131(7)-E1100-8.
- Cabanero, M., Sangha, R., Sheffield, B., Sukhai, M., Pakkal, M., Kamel-Reid, S., Karsan, A., Ionescu, D., Juergens, R., Butts, C. et al. (2017). Management of egfr-mutated nonsmall- cell lung cancer: practical implications from a clinical and pathology perspective. *Current Oncology* 24(2), 111.
- Canbay, E., Cakmakoglu, B., Zeybek, U., Sozen, S., Cacina, C., et al. (2011). Association of APE1 and hOGG1 polymorphisms with colorectal cancer risk in a Turkish population. *Curr Med Res Opin.* 27; 1295–1302.
- Carles, J., Monzo, M., Amat, M., et al. (2006). Singlenucleotide polymorphisms in base excision repair, nucleotide excision repair, and double strand break genes as markers for response to radiotherapy in patients with Stage I to II headand- neck cancer. *Int J Radiat Oncol Biol Phys* 66; 1022–1030.
- Carlsten, C., Sagoo, G.S., Frodsham, A.J., Burke, W., Higgins, J. P. (2008). Glutathione Stransferase *M1* (*GSTM1*) polymorphisms and lung cancer: a literature-based systematic HuGE review and meta analysis. *American Journal of Epidemiology* 167; 759–774.
- Catana, A., Popp, A. R., Pop, M., Porojan, M. D., Petrisor, F. M., Pop, I. V. (2012). Genetic Polymorphism of DNA repair gene ERCC2/XPD (Arg156Arg) (A22541C) and Lung Cancer Risk in Northern Romania. *RevistaRomana de Medicina de Laborator* 20; 2-4.
- Chang, A. (2011). Chemotherapy, chemoresistance and the changing treatment landscape for nsclc. *Lung cancer* 71(1); 3–10.
- Chang, J. S., Wrensch, M. R., Hansen, H. M., Sison, J. D., Aldrich, M. C., et al. (2008). Nucleotide excision repair genes and risk of lung cancer among San Francisco Bay Area Latinos and African Americans. *Int J Cancer* 123; 2095–2104.

- Chang, J. S., Wrensch, M. R., Hansen, H. M., Sison, J. D., Aldrich, M. C., *et al.* (2008). Nucleotide excision repair genes and risk of lung cancer among San Francisco Bay Area Latinos and African Americans. *Int J Cancer* 123; 2095–2104.
- Chen, S., Tang, D., Xue, K., *et al.* (2002). DNA repair gene XRCC1 and XPD polymorphisms and risk of lung cancer in a Chinese population. *Carcinogenesis* 23; 1321–5.
- Chen, M., Kamat, A. M., Huang, M., Grossman, H. B., Dinney, C. B., Lerner, S. P., Wu, X., Gu, J. (2007). High-order interactions among genetic polymorphisms in nucleotide excision repair pathway genes and smoking in modulating bladder cancer risk. *Carcinogenesis* 10(28); 2160–2165.
- Chen, Z., *et al.* (2007). Attenuated expression of xeroderma pigmentosum group C is associated with critical events in human bladder cancer carcinogenesis and progression. *Cancer Res.* 67; 4578–4585.
- Chen, G., Huynh, M., Fehrenbacher, L., West, H., Lara, P. N. Jr., Yavorkovsky, L. L., Russin, M., Goldstein, D., Gandara, D., Lau, D. (2009) Phase II trial of irinotecan and carboplatin for extensive or relapsed small-cell lung cancer. *J Clin Oncol.* 20-27(9); 1401-4.
- Cheng, H. B., Xie, C., Zhang, R. Y., *et al.* (2013). Xeroderma pigmentosum complementation group F polymorphisms influence risk of glioma. *Asian Pac J Cancer Prev* 14; 4083-7.
- Cheng, H., Qin, Q., Sun, X., Li, F., Sun, N., *et al.* (2013). Predictive Effect of XPA and XPD Polymorphisms on Survival of Advanced NSCLC Patients Treated with Platinum-based Chemotherapy: A Three-dimensional (3-D), Polyacrylamide Gel-Based DNA Microarray Method. *Technol Cancer Res Treat.*
- Cheng, L., Eicher, S. A., Guo, Z., Hong, W. K., Spitz, M. R., Wei, Q. (1998). Reduced DNA capacity in head and neck cancer patients. *Cancer Epidemiol Biomarkers Prev.*, 7; 465-468.

- Cheng, L., Spitz, M. R., Hong, W. K., *et al.* (2000). Reduced expression levels of nucleotide excision repair genes in lung cancer: a case-control analysis. *Carcinogenesis* 21; 1527-1530.
- Citterio, E., Van, D. B. V., Schnitzler, G., Kanaar, R., Bonte, E., Kingston, R.E., Hoeijmakers, J.H. and Vermeulen, W. (2000). ATP –dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Mol. Cell Biol.* 20; 7643-7653.
- Clancy, S. (2008). DNA damage & repair: mechanisms for maintaining DNA integrity. *Nature Education* 1(1); 103.
- Cleaver, J.E., Lam, E.T. and Revet, I. (2009). Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat. Rev. Gent.* 10; 756-768.
- Cohen, S. M. and Arnold, L. L. (2010). Chemical carcinogenesis. *Toxicological Sciences* 120(suppl 1); S76–S92.
- Coin, F., Oksenysh, V., and Egly, J.M. (2007). Distinct roles for the XP B/p52 and XPD /p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol. Cell* 26; 245-256.
- Coin, F., Oksenysh, V., Mocquet, V., Groh, S., Blattner, C. and Egly, J. M. (2008). Nucleotide excision repair driven by the dissociation of CA K from TFIIH. *Mol. Cell* 31; 9-20.
- Coin, F., Proietti De Santis, L., Nardo, T., Zlobinskaya, O., Stefanini, M. and Egly, J.M. (2006). p8/TTD -A as a repair-specific TFIIH subunit. *Mol. Cell* 21; 215-226.
- Colby, T. V. (1995). Tumors of the lower respiratory tract. *Atlas of Tumor Pathology*, 309-317.
- Concato, J., Feinstein, A. R., Holford, T. R. (1993). The risk of determining risk with multivariable models. *Ann Intern Med.* 118; 201–210.
- Control, U. D. and Prevention (2010). *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General*, National Center for Chronic Disease Prevention and Health Promotion (US); Office on Smoking and Health (US).
- Cooper, W. A., Lam, D. C., OToole, S. A. and Minna, J. D. (2013). Molecular biology of lung cancer. *Journal of thoracic disease* 5(Suppl 5); S479.

- Costa, R. M., Chigancas, V., Galhardo Rda, S., Carvalho, H. and Menck, C.F. (2003). The eukaryotic nucleotide excision repair pathway. *Biochimie* 85; 1083-1099.
- Crew, K. D., Gammon, M. D., Terry, M. B., Fang, F. Z., Zablotska, L. B., *et al.* (2007). Polymorphisms in nucleotide excision repair genes, polycyclic aromatic hydrocarbon-DNA adducts, and breast cancer risk. *Cancer Epidemiology Biomarkers and Prevention* 16; 2033–2041.
- Crew, K.D., Gammon, M.D., Terry, M.B., Zhang, F.F., Zablotska, L.B., Agrawal, M., Shen, J., Long, C.M., Eng, S.M., Sagiv, S.K., Teitelbaum, S.L., Neugut, A.I., Santella, R.M. (2007). Polymorphisms in nucleotide excision repair genes, polycyclic aromatic hydrocarbon-DNA adducts, and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 16(10); 2033–2041.
- Cruz, C. S. D., Tanoue, L. T. and Matthay, R. A. (2011). Lung cancer: epidemiology, etiology, and prevention. *Clinics in chest medicine* 32(4); 605–644.
- Cui, Y., Morgenstern, H., Greenland, S., Tashkin, D. P., Mao, J., *et al.* (2006). Polymorphism of Xeroderma Pigmentosum group G and the risk of lung cancer and squamous cell carcinomas of the oropharynx, larynx and esophagus. *Int J Cancer* 118; 714–720.
- Dalavanga, Y. (2017). The principles governing the lung structure and function. *Anatomy Physiology & Biochemistry International Journal* 2.
- Datkhile, K., Patil, M., Khamkar, T., Durgawale, P. P. (2016). Association between genetic polymorphisms of DNA repair genes (XPC, XPG) and head and neck cancer susceptibility in Indian population: A hospital based case-control study. *European Journal of Pharmaceutical and Medical Research*, 408-415.
- David-Beabes, G. L., Lunn, R. M. and London, S. J. (2001). No association between the XPD (Lys751Gln) polymorphism or the XRCC3 (Thr241Met) polymorphism and lung cancer risk. *Cancer Epidemiol Biomarkers Prev.* 10; 911–912.
- de Groot, P. and Munden, R. F. (2012). Lung cancer epidemiology, risk factors, and prevention. *Radiologic Clinics*, 50(5); 863–876.
- de Laat, W. L., Jaspers, N. G. J., Hoeijmakers, J. H. J. (1999). Molecular mechanism of nucleotide excision repair. *Genes Dev*, 13; 768–85.

- de Laat, W. L., Sijbers, A. M., Odijk, H., Jaspers, N.G. and Hoeijmakers, J.H. (1998). Mapping of interaction domains between human repair proteins ERCC 1 and XPF. *Nucleic Acids Res.* 26; 4146-4152.
- De Ruyck, K., Szaumkessel, M., De Rudder, I., Dehoorne, A., Vral, A., Claes, K., Velghe, A., Van Meerbeeck, J., Thierens, H. (2007). Polymorphisms in base-excision repair and nucleotide-excision repair genes in relation to lung cancer risk. *Mutat Res.* 631(2); 101-10.
- Debniak, T., Scott, R. J., Huzarski, T., Byrski, T., Masojc, B., van de Wetering, T., Serrano-Fernandez, P., *et al.* (2006). XPD common variants and their association with melanoma and breast cancer risk. *Breast Cancer Res Treat* 98; 209–215.
- DeLoia, J. A., Bhagwat, N. R., Darcy, K. M., *et al.* (2012). Comparison of *ERCC1/XPF* genetic variation, mRNA and protein levels in women with advanced stage ovarian cancer treated with intraperitoneal platinum. *Gynecol. Oncol.* 126; 448–454.
- Ding, D., Zhang, Y., Yu, H., Guo, Y., Jiang, L., He, X., Ma, W., Zheng, W. (2012). Genetic variation of XPA gene and risk of cancer: a systematic review and pooled analysis. *Int J Cancer* 15; 131(2);488-96.
- Doherty, J. A., Weiss, N. S., Fish, S., *et al.* (2011). Polymorphisms in nucleotide excision repair genes and endometrial cancer risk. *Cancer Epidemiol Biomarkers Prev* 20; 1873-82.
- Dong, Z., Guo, W., Zhou, R., *et al.* (2008). Polymorphisms of the DNA repair gene XPA and XPC and its correlation with gastric cardiac adenocarcinoma in a high incidence population in North China. *J Clin Gastroenterol.* 42(8); 910–915.
- Donnadieu, N., Paesmans, M. and Sculier, J. P. (1991). Chimiothérapie du cancer bronchique non à petites cellules: méta-analyse de la littérature en fonction de l'extension de la maladie. *Rev Mal Respir* 8; 197–204.
- Downward, J. (2003). Targeting ras signalling pathways in cancer therapy. *Nature Reviews Cancer* 3(1); 11
- Dybdahl M, Vogel U, Frentz G, Wallin H, Nexø BA. Polymorphisms in the DNA repair gene XPD: correlations with risk and age at onset of basal cell carcinoma. *Cancer Epidemiol. Biomark. Prev.* 8, 77–81 (1999).

- Elaine, N., Marieb, K. H. (2012). Human Anatomy & Physiology. *Pearson*.
- Emmert, S., Schneider, T. D., Khan, S. G. *et al.* (2001). The human XPG gene: gene architecture, alternative splicing and single nucleotide polymorphisms. *Nucleic Acids Res.* 29; 1443-1452.
- Emmert, S., Schneider, T. D., Khan, S. G., Kraemer, K. H. (2001). The human XPG gene: gene architecture, alternative splicing and single nucleotide polymorphisms. *Nucleic Acids Res.* 29; 1443-1452.
- Empey, D. W. (1978). Diseases of the respiratory system. introduction: structure and function of the lungs. *British medical journal*, 1(6113); 631.
- Enzlin, J. H. and Scharer, O. D. (2002). The active site of the DNA repair endonuclease XPF -ERCC 1 forms a highly conserved nuclease motif. *EMBO J.* 21; 2045-2053.
- Fanning, E., Klimovich, V. and Nager, A. R. (2006). A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res.* 34; 4126-4137.
- Fareed, K. R., Al-Attar, A., Soomro, I. N. *et al.* (2012). Tumour regression and ERCC1 nuclear protein expression predict clinical outcome in patients with gastro-oesophageal cancer treated with neoadjuvant chemotherapy. *Br. J. Cancer* 102; 1600–1607.
- Feng, X. X., Duan, P. F., Wang, L. B., Zhang, J. B., Lu, Z. X. (2008). Study on the relationship between polymorphisms of XPA gene and susceptibility of esophageal cancer. *Zhonghua Liu Xing Bing Xue Za Zhi* 29(9); 930–933.
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. and Bray, F. (2012). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer*, 136; E359–E386.
- Figl, A., Scherer, D., Nagore, E., *et al.* (2010). Single-nucleotide polymorphisms in DNA-repair genes and cutaneous melanoma. *Mutat Res* 702; 8-16.
- Fleck, O. (2004). DNA repair pathways. *The Molecular Biology of Schizosaccharomyces pombe: Genetics, Genomics and Beyond*, ed. E. Richard, Springer, pp. 101-115.
- Fleiss, J. L. and 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.

- Fleming, N. D., Agadjanian, H., Nassanian, H., Miller, C. W., et al. (2012). Xerodermapigmentosum complementation group C single-nucleotide polymorphisms in the nucleotide excision repair pathway correlate with prolonged progression-free survival in advanced ovarian cancer. *Cancer*, 118; 689-697.
- Fleming, N. D., Agadjanian, H., Nassanian, H., Miller, C. W., Orsulic, S., Karlan, B. Y., Walsh, C. S. (2012). Xeroderma pigmentosum complementation group C single-nucleotide polymorphisms in the nucleotide excision repair pathway correlate with prolonged progression-free survival in advanced ovarian cancer. *Cancer* 1; 118(3).
- Flurina, C., Camenisch, U., Fei, J., Kaczmarek, N., Mathieu, N., Naegeli, H. (2010). Dynamic two-stage mechanism of versatile DNA damage recognition by xeroderma pigmentosum group C protein. *Mutation Research* 685; 21–28.
- Fong, K., Sekido, Y., Gazdar, A. and Minna, J. (2003). Lung cancer 9: Molecular biology of lung cancer: clinical implications. *Thorax* 58(10); 892–900.
- Forgacs, E., Zöschbauer-Müller, S., Oláh, E. and Minna, J. D. (2001). Molecular genetic abnormalities in the pathogenesis of human lung cancer. *Pathology Oncology Research* 7(1); 6–13.
- Fousteri, M., Vermeulen, W., van Zeeland, A. A. and Mullenders, L. H. (2006). Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Mol. Cell* 23; 471-482.
- Friedberg, E. C., Aguilera, A., Gellert, M., Hanawalt, P.C., Hays, J.B., Lehmann, A. R., Lindahl, T., Lowndes, N., Sarasin, A., and Wood, R. D. (2006). DNA repair: from molecular mechanism to human disease. *DNA Repair (Amst)* 5; 986-996.
- Fukuda, M., Oka, M., Soda, H. et al. (2004). Phase II study of irinotecan combined with carboplatin in previously untreated non-small-cell lung cancer. *Cancer Chemother. Pharmacol.* 154; 73–577.
- Funai, K., Yokose, T., Ishii, G. I., Araki, K., Yoshida, J., Nishimura, M., Nagai, K., Nishiwaki, Y. and Ochiai, A. (2003). Clinicopathologic characteristics of peripheral squamous cell carcinoma of the lung. *The American journal of surgical pathology*, 27(7); 978–984.

- Fuss, J. O., Cooper, P. K. (2006). DNA repair: dynamic defenders against cancer and aging. *PLoS Biology*. 4(6), e203.
- Gandara, D. R., Kawaguchi, T., Crowley, J., Moon, J., Furuse, K., *et al.* (2009). Japanese-US common-arm analysis of paclitaxel plus carboplatin in advanced non-small-cell lung cancer: a model for assessing population-related pharmacogenomics. *J Clin Oncol* 27; 3540–3546.
- Garcia-Closas, M., Malats, N., Real, F. X., Welch, R., Kogevinas, M., *et al.* (2006). Genetic variation in the nucleotide excision repair pathway and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev.* 15; 536–542.
- Gavande, N. S., VanderVere-Carozza, P. S., Hinshaw, H. D., Jalal, S. I., Sears, C. R., Pawelczak, K. S., Turchi, J. J. (2016). DNA repair targeted therapy: The past or future of cancer treatment. *Pharmacol. Ther.* 160; 65–83.
- Gawe lek, E., Drozdowska, B. and Fuchs, A. (2017). Radon as a risk factor of lung cancer. *Przegląd epidemiologiczny*, 71(1); 90–98.
- Gelboin, H. V. (1980). Benzo [alpha] pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes. *Physiological reviews* 60(4); 1107–1166.
- Gil, J., Ramsey, D., Stembalska, A., *et al.* (2012). The C/A polymorphism in intron 11 of the XPC gene plays a crucial role in the modulation of an individual's susceptibility to sporadic colorectal cancer. *Mol Biol Rep.* 39(1); 527–534.
- Gil, J., Ramsey, D., Stembalska, A., *et al.* (2012). The C/A polymorphism in intron 11 of the XPC gene plays a crucial role in the modulation of an individual's susceptibility to sporadic colorectal cancer. *Mol Biol Rep* 39; 527-34.
- Gil, J., Ramsey, D., Stembalska, A., Karpinski, P., Pesz, K. A., *et al.* (2011). The C/A polymorphism in intron 11 of the XPC gene plays a crucial role in the modulation of an individual's susceptibility to sporadic colorectal cancer. *Mol Biol Rep.*
- Global Adult Tobacco Survey (GATS) India: 2009-2010. Available from: http://www.searo.who.int/LinkFiles/Regional_Tobacco_Surveillance_System_GATS_India.pdf. published by IIPS, Mumbai and funded by the Ministry of Health and Family Welfare, GOI 2010 .

- Globocan (2012). Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. GLOBOCAN. 2013; 2012 v1. 0. Available from: <http://globocan.iarc.fr>.
- Goodsell, D. S. (1999). The molecular perspective: the ras oncogene. *Stem cells* 17(4); 235–236.
- Gu, J., Zhao, H., Dinney, C. P., et al. (2005). Nucleotide excision repair gene polymorphisms and recurrence after treatment for superficial bladder cancer. *Clin Cancer Res* 11; 1408–1415.
- Guo, W., Zhou, R. M., Wan, L. L., et al. (2008). Polymorphisms of the DNA repair gene xeroderma pigmentosum groups A and C and risk of esophageal squamous cell carcinoma in a population of high incidence region of North China. *J Cancer Res Clin Oncol*. 134(2); 263–270.
- Gurubhagavatula, S., Liu, G., Park, S., Zhou, W., Su, L., et al. (2004). XPD and XRCC1 genetic polymorphisms are prognostic factors in advanced non-small-cell lung cancer patients treated with platinum chemotherapy. *J Clin Oncol* 22; 2594–2601.
- Hackshaw, A. K., Law, M. R. and Wald, N. J. (1997). The accumulated evidence on lung cancer and environmental tobacco smoke. *Bmj* 315(7114); 980–988.
- Hall, J., Hashibe, M., Boffetta, P., et al. (2007). The association of sequence variants in DNA repair and cell cycle genes with cancers of the upper aerodigestive tract. *Carcinogenesis* 28(3); 665–671.
- Harms, C., Salama, A. S., Sierra-Torres, C. H., Cajas-Salazar, N. and William, W. A. (2004). Polymorphisms in DNA repair genes, chromosome aberrations, and lung cancer. *Environmental and Molecular mutagenesis* 44(1); 74-82.
- Han, J., Colditz, G. A., Liu, J. S. and Hunter, D. J. (2005). Genetic variation in XPD, sun exposure, and risk of skin cancer. *Cancer epidemiology biomarker and prevention* 14(6); 1539-44.
- Han, J., Haiman, C., Niu, T., et al. (2009). Genetic variation in DNA repair pathway genes and premenopausal breast cancer risk. *Breast Cancer Res Treat.* 115; 613–22.
- Hansen, R. D., Sorensen, M., Tjonneland, A., et al. (2007). XPA A23G, XPC Lys939Gln, XPD Lys751Gln and XPD Asp312Asn polymorphisms, interactions with smoking, alcohol and dietary factors, and risk of colorectal cancer. *Mutat Res.* 619(1–2):68–80.

- Harth, V., Schafer, M., Abel, J., Maintz, L., Neuhaus, T., et al. (2008). Head and neck squamous-cell cancer and its association with polymorphic enzymes of xenobiotic metabolism and repair. *J Toxicol Environ Health* A71; 887–897.
- 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. (2012). Lung carcinogenesis by tobacco smoke. *International journal of cancer* 131(12); 2724–2732.
- Hermanson-Miller, I. L. and Turchi, J.J. (2002). Strand-specific binding of RPA and XPA to damaged duplex DNA. *Biochemistry* 41; 2402–2408.
- Hirschhorn, J. N., Lohmueller, K., Byrne, E., Hirschhorn, K., (2002). A comprehensive review of genetic association studies. *Genet. Med.* 4(2); 45–61.
- Hoeijmakers, J. H. (2009). DNA damage, aging, and cancer. *N Engl J Med.* 361 (15); 1475–1485.
- Hoeijmakers, J. H., Egly, J. M., Vermeulen, W (1996). TFIIH: a key component in multiple DNA transactions. *Curr. Opin. Genet. Dev.* 26; 26–33.
- Hosmer, D. W. and Lemeshow, S. (2000). Applied logistic regression. John Wiley & Sons, New York.
- Houtgraaf, J. H., Versmissen, J. and van der Giessen, W. J. (2006). A concise review of dna damage checkpoints and repair in mammalian cells. *Cardiovascular Revascularization Medicine* 7(3); 165–172.
- Hsu, F. C., Lindström, S., Sun, J., Wiklund, F., Chen, S. H., Adami, H. O., Turner, A. R. et al. (2008). A multigenic approach to evaluating prostate cancer risk in asystematic replication study. *Cancer Genet Cytogenet.* 183(2); 94–98.
- Hu, Z., Wang, Y., Wang, X., Liang, G., Miao, X., Xu, Y., Tan, W., Wei, Q., Lin, D., Shen, H. (2005). DNA repair gene XPC genotypes/haplotypes and risk of lung cancer in a Chinese population. *Int. J. Cancer* 115; 478–483.

- Huang, S.J., Shi, G.Y., Liu, L.G. (2014). Role of ERCC1 variants in response to chemotherapy and clinical outcome of advanced non-small cell lung cancer. *Tumor Biol.* 35(5); 4023–4029.
- Huang, W. Y., Berndt, S. I., Kang, D., *et al.* (2006). Nucleotide excision repair gene polymorphisms and risk of advanced colorectal adenoma: XPC polymorphisms modify smoking-related risk. *Cancer Epidemiol Biomarkers Prev* 15; 306-11.
- Hung, R. J., Christiani, D. C., Risch, A., Popanda, O., Haugen, A., *et al.* (2008). International Lung Cancer Consortium: pooled analysis of sequence variants in DNA repair and cell cycle pathways. *Cancer Epidemiol Biomarkers Prev* 17; 3081–3089.
- Hussain, S. K., Mu, L. N., Cai, L., Chang, S. C., Park, S. L., Oh, S. S., *et al.* (2009). Genetic variation in immune regulation and DNA repair pathways and stomach cancer in China. *Cancer Epidemiol Biomarkers Prev* 18; 2304–2309.
- Hyde, L. and Hyde, C. I. (1974). Clinical manifestations of lung cancer. *Chest* 65(3); 299-306.
- Ibarrola-Villava, M., Pena-Chilet, M., Fernandez, L. P., Aviles, J. A., Mayor, M., *et al.* (2011). Genetic polymorphisms in DNA repair and oxidative stress pathways associated with malignant melanoma susceptibility. *European Journal of Cancer*.
- Isla, D., Sarries, C., Rosell, R., Alonso, G., Domine, M., *et al.* (2004). Single nucleotide polymorphisms and outcome in docetaxel-cisplatin-treated advanced non-small-cell lung cancer. *Ann Oncol* 15; 1194–1203.
- Jelonek, K., Gdowicz-Klosok, A., Pietrowska, M., *et al.* (2010). Association between single-nucleotide polymorphisms of selected genes involved in the response to DNA damage and risk of colon, head and neck, and breast cancers in a Polish population. *J Appl Genet.* 51(3); 343–352.
- Jelonek, K., Gdowicz-Klosok, A., Pietrowska, M., *et al.* (2010). Association between single- nucleotide polymorphisms of selected genes involved in the response to DNA damage and risk of colon, head and neck, and breast cancers in a Polish population. *J Appl Genet.* 51(3); 343–352.

- Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., Forman, D. (2011). Global cancer statistics. *CA Cancer J. Clin.* 20(61); 69-90.
- Jeon, H. S., Kim, K. M., Park, S. H. et al. (2003). Relationship between XPG codon 1104 polymorphism and risk of primary lung cancer. *Carcinogenesis* 24; 1671-1681 .
- Jeon, H. S., Kim, K. M., Park, S. H., Lee, S. Y., Choi, J. E., Lee, G. Y., et al. (2003). Relationship between XPG codon 1104 polymorphism and risk of primary lung cancer. *Carcinogenesis* 24; 1671-1681
- Ji, Y. B., Tae, K., Lee, Y. S., Lee, S.H., et al. (2010). XPD polymorphisms and risk of Squamous cell Carcinoma of the head and neck in a Korean sample. *Clin. Exp. Otorhinolaryngol.* 3; 42-47.
- Jin, B., Dong, Y., Zhang, X., Wang, H., Han, B. (2014). Association of XPC Polymorphisms and Lung Cancer Risk: A Meta-Analysis. *Plos one* 9(4); e93937.
- Jindal, S., Malik, S. et al. (1988). Smoking index-a measure to quantify cumulative smoking exposure. *Lung India* 6(4); 195.
- Joerger, M., Burgers, S. A., Baas, P., Smit, E. F., Haitjema, T. J., et al. (2012). Germline polymorphisms in patients with advanced nonsmall cell lung cancer receiving first-line platinum-gemcitabine chemotherapy: a prospective clinical study. *Cancer* 118(9); 2466–2475.
- Jorgensen, T. J., Visvanathan, K., Ruczinski, I., Thuita, L., Hoffman, S., Helzlsouer, K. J. (2007). Breast cancer risk is not associated with polymorphic forms of xeroderma pigmentosum genes in a cohort of women from Washington County, Maryland. *Breast Cancer Res Treat* 101; 65–71.
- Jorgensen, T. J., Visvanathan, K., Ruczinski, I., Thuita, L., Hoffman, S., et al. (2007). Breast cancer risk is not associated with polymorphic forms of xeroderma pigmentosum genes in a cohort of women from Washington County, Maryland. *Breast Cancer Res Treat.* 101:65–71.
- Joshi, A. D., Corral, R., Siegmund, K. D., et al. (2009). Red meat and poultry intake, polymorphisms in the nucleotide excision repair and mismatch repair pathways and colorectal cancer risk. *Carcinogenesis* 30; 472-9.
- Kadara, H., Scheet, P., Wistuba, I. I. and Spira, A. E. (2016). Early events in the molecular pathogenesis of lung cancer. *Cancer Prevention Research* 9(7); 518–527.

- Kalikaki, A., Kanaki, M., Vassalou, H., Souglakos, J., Voutsina, A., Georgoulas, V., Mavroudis, D. (2009). DNA repair gene polymorphisms predict favorable clinical outcome in advanced non-small-cell lung cancer. *Clin Lung Cancer* 10(2); 118–123.
- Kang, S., Sun, H. Y., Zhou, R. M., Wang, N., Hu, P., Li, Y. (2013). DNA Repair Gene Associated with Clinical Outcome of Epithelial Ovarian Cancer Treated with Platinum-based Chemotherapy. *Asian Pacific J Cancer Prev* 14 (2); 941-946.
- Kannouche, P. and Stary. A. (2003). Xeroderma pigmentosum variant and error-prone DNA polymerases. *Biochimie* 85; 1123-1132.
- Karahalil, B., Bohr, V., Wilson, D. (2012). Impact of DNA polymorphisms in key DNA base excision repair proteins on cancer risk. *Human and Experimental Toxicology*, 31(10); 981–1005.
- Kertat, K., Rosdahl, I., Sun, X. F., et al. (2008). The Gln/Gln genotype of XPD codon 751 as a genetic marker for melanoma risk and Lys/Gln as an important predictor for melanoma progression: A case control study in the Swedish population. *Oncol Rep* 20; 179-83.
- Key Statistics for Lung Cancer. (2017, January 2). Retrieved October 31, 2017, from <https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html>
- Khan, S. G., Metter, E. J., Tarone, R. E., Bohr, V. A., Grossman, L., Hedayati, M., Bale, S. J., Emmert, S., Kraemer, K. H. (2000). A new xeroderma pigmentosum group C poly(AT) insertion/deletion polymorphism. *Carcinogenesis*, 21; 1821–5.
- Kim, J.K., Patel, D. and Choi, B.S. (1995). Contrasting structural impacts induced by cis-syn cyclobutane dimer and (6-4) adduct in DNA duplex decamers: implication in mutagenesis and repair activity. *Photochem. Photobiol.* 62; 44-50.
- Kirschbaum, M. H. and 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.
- Kohlhasse, S., Bogdanova, N. V., Schürmann, P., Bermisheva, M., Khusnutdinova, E., Antonenkova, N., et al. (2014). Mutation Analysis of the *ERCC4/FANCD1* Gene in Hereditary Breast Cancer. *PLoS ONE* 9(1); e85334.

- Kong, F. J., Xu, W. S., Wang, H. L., et al. (2008). Association of the mono-nucleotide polymorphism ha DNA repair gone XPD with the risk ef lung cancer. *Prac J Med Pharm* 10; 1161-1163.
- Kornguth, D. G., Garden, A. S., Zheng, Y., et al. (2005). Gastrostomy in oropharyngeal cancer patients with *ERCC4 (XPF)* germline variants. *Int. J. Radiat. Oncol. Biol. Phys.* 62; 665–671.
- Kosaka, T., Yatabe, Y., Endoh, H., Kuwano, H., Takahashi, T. and Mitsudomi, T. (2004). Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer research* 64(24); 8919–8923.
- Krupa, R., Kasznicki, J., Gajęcka, M. et al. (2001). Polymorphism of the DNA polymorphism of the DNA repair genes XRCC1 and ERC4 are not associated with smoking and drinking dependent larynx cancer in a polish population. *Exp Oncol.*, 33 (1); 55-56.
- Krupa, R., Kasznicki, J., Gajecka, M., et al. (2011). Polymorphisms of the DNA repair genes XRCC1 and ERCC4 are not associated with smoking- and drinking-dependent larynx cancer in a Polish population. *Exp Oncol* 33; 55-6.
- Kumar, R., Hoglund, L., Zhao, C., Forsti, A., Snellman, E., et al. (2003). Single nucleotide polymorphisms in the XPG gene: determination of role in DNA repair and breast cancer risk. *Int J Cancer* 103; 671–675.
- Kumar, R., Hoglund, L., Zhao, C., Forsti, R., Snellman, E., Hemminki, K. (2003). Single nucleotide polymorphisms in the XPG gene: determination of role in the DNA repair and breast cancer risk. *Int J Cancer*, 103; 671–675.
- Kusumoto, R., Masutani, C., Sugasawa, K. *et al.* (2007). Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. *Mutat Res*, 485; 219–27.
- Lam, T. K., Gallicchio, L., Lindsley, K., Shiels, M., Hammond, E., Tao, X. G., Chen, L., Robinson, K. A., Caulfield, L. E., Herman, J. G. *et al.* (2009). Cruciferous vegetable consumption and lung cancer risk: a systematic review. *Cancer Epidemiology and Prevention Biomarkers*, 18(1); 184–195.
- Latimer, K. M. and Mott, T. F. (2015). Lung cancer: diagnosis, treatment principles, and screening. *American family physician* 91(4).

- Lavanya, J., Kumar, V. J., Sudhakar, N., Prathap, S. (2015). Analysis of DNA repair genetic polymorphism in breast cancer population. *Int J Pharm Bio Sci* 6(3B); 966 – 973.
- Lavanya, J., Kumar, V. J., Sudhakar, N. et al. (2015). Analysis of DNA repair genetic polymorphism in breast cancer population. *Int. J. Pharm. Bio.Sci.* 6(3); 966 – 973.
- Lee, G. Y., Jang, J. S., Lee, S. Y., Jeon, H. S., Kim, K. M., et al. (2005). XPC polymorphisms and lung cancer risk. *Int J Cancer* 115; 807–813.
- Letkova, L., Matakova, T., Musak, L., Sarlinova, M., Krutakova, M., Slovakova, P., Kavcova, E., Jakusova, V., Janickova, M., Drgova, A., Berzinec, P., Halasova, E. (2013). DNA repair genes polymorphism and lung cancer risk with the emphasis to sex differences. *Mol Biol Rep* 40; 5261–5273.
- Li, F., Sun, X., Sun, N., Qin, S., Cheng, H., et al. (2010). Association between polymorphisms of ERCC1 and XPD and clinical response to platinum-based chemotherapy in advanced non-small cell lung cancer. *Am J Clin Oncol* 33; 489–494.
- Li, L., Lu, X., Peterson, C., Legerski, R., Cinteracts, P. (1997). XPC with both HHR23B and HHR23A in vivo. *Mutat. Res.* 383; 197–203.
- Li, X. D., Han, J. C., Zhang, Y. H., Li, H. B., Wu, X. Y. (2013). Common variations of DNA repair genes are associated with response to platinum-based chemotherapy in NSCLCs. *Asian Pac. J. Cancer Prev.* 14; 145–148.
- Liang, G., Xing, D., Miao, X., Tan, W., Yu, C., Lu, W., Lin, D. (2003). Sequence variations in the DNA repair gene XPD and risk of lung cancer in a Chinese population. *Int. J. Cancer* 105; 669–673.
- Liu, J., Zhang, Z., Cao, X. L., et al. (2012). XPA A23G polymorphism and susceptibility to cancer: a meta-analysis. *Mol Biol Rep* 39; 6791-6799.
- Liu, F., Killian, J. K., Yang, M., Walker, R. L., Hong, J. A., Zhang, M. and Rao, M. (2010). Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene* 29(25); 3650-3664.
- Liu, L., Yuan, P., Wu, C., Zhang, X., et al. (2011). Assessment of XPD Lys751Gln and XRCC1 T-77C polymorphisms in advanced non-small-cell lung cancer patients treated with platinum-based chemotherapy. *Lung Cancer* 73; 110–115.

- Liu, P., Siciliano, J., White, B., *et al.* (1993). Regional mapping of human DNA excision repair gene ERCC4 to chromosome 16p13.13-p13.2. *Mutagenesis* 8; 199–205.
- Liu, Y., Cao, L., Chang, J., *et al.* (2015). XPF-673C > T polymorphism effect on the susceptibility to esophageal cancer in Chinese population. *PLoS ONE* 9(4); e94136 (2015).
- Ljungman, M. and Zhang, F. (1996). Blockage of RNA polymerase as a possible trigger for u.v.. light-induced apoptosis. *Oncogene* 13; 823-831.
- Loeb, K. R. and Loeb, L. A. (2000). Significance of multiple mutations in cancer. *Carcinogenesis* 21(3); 379–385.
- Lopez-Cima, M. F., Gonzalez-Arriaga, P., Garcia-Castro, L., Pascual, T., Marron, M. G., *et al.* (2007). Polymorphisms in XPC, XPD, XRCC1, and XRCC3 DNA repair genes and lung cancer risk in a population of northern Spain. *BMC Cancer* 7; 162.
- Lou, Y., Li, R., Zhang, Y., Zhong, R., *et al.* (2014). XPA gene rs1800975 single nucleotide polymorphism and lung cancer risk : a metanalysis. *Tumor Biol.* 35; 6607-6617.
- Lovatt, T., Alldersea, J., Lear, J. T., Hoban, P. R., Ramachandran, S., Fryer, A. A., *et al.* (2005). Polymorphism in the nuclear excision repair gene ERCC2/XPD: association between an exon 6-exon 10 haplotype and susceptibility to cutaneous basal cell carcinoma. *Hum Mutat.* 25; 353–359.
- Ludovini, V., Floriani, I., Pistola, L., Minotti, V., Meacci, M., *et al.* (2011). Association of cytidine deaminase and xeroderma pigmentosum group D polymorphisms with response, toxicity, and survival in cisplatin/gemcitabine-treated advanced non-small cell lung cancer patients. *J. Thorac. Oncol.* 6; 2018–2026.
- Lung Cancer (2012).URL: www.globocan.iarc.fr/Pages/fact-sheets-cancer.aspx
- Lung cancer (2017). Retrieved October 31, 2017, from <https://ghr.nlm.nih.gov/condition/lung-cancer>
- Lung Cancer (2017). URL: <https://ghr.nlm.nih.gov/condition/lung-cancer>
- Lung cancer statistics UK (2015). URL: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer>

- Lunn, R. M., Helzlsouer, K. J., Parshad, R., Umbach, D. M., Harris, E. L., Sanford, K. K., Bell, D. A. (2000). XPD polymorphisms: effects on DNA repair proficiency. *Carcinogenesis* 21(4):551-5
- Ma, H., Yu, H., Liu, Z., Wang, L. E., *et al.* (2012). Polymorphisms of XPG/ERCC5 and risk of squamous cell carcinoma of the head and neck. *Pharmacogenet. Genomics* 22; 50-57.
- Majumder, M., Sikdar, N., Ghosh, S. and Roy, B. (2007). Polymorphisms at XPD and XRCC1 DNA repair loci and increased risk of oral leukoplakia and cancer among NAT2 slow acetylators. *Int. J. Cancer* 120; 2148-2156.
- Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C. and Boffetta, P. (2016). Risk factors for lung cancer worldwide. *European Respiratory Journal* ERJ-00359.
- Malik, P. S. and Raina, V. (2015). Lung cancer: Prevalent trends & emerging concepts. *The Indian journal of medical research* 141(1); 5.
- Manelyte, L., Kim, Y. I. T., Smith, A. J., Smit, R. M., Savery, N. J. (2010). Regulation and rate enhancement during transcription-coupled DNA repair. *Mol. Cell* 40 (5); 714-724.
- Marin, M. S., Lopez-Cima, M. F., Garcia-Castro, L., Pascual, T., Marron, M. G., *et al.* (2004) Poly (AT) polymorphism in intron 11 of the XPC DNA repair gene enhances the risk of lung cancer. *Cancer Epidemiol Biomarkers Prev* 13; 1788– 1793.
- Martin, O. V., Shialis, T., Boobis, A. R., Voulvoulis, N., Lester, J. N., Scrimshaw, M. D. (2008). Defective Spermatogenesis: Martin *et al.* Respond. *Environmental Health Perspectives* 116(8); A332.
- Massion, P. P. and Carbone, D. P. (2003). The molecular basis of lung cancer: molecular abnormalities and therapeutic implications. *Respiratory research* 4(1); 12.
- Matakideu, A., Eisen, T., Bridle, H., Houlsten, S. R. (2006). Nucleotide excision repair polymorphism modulate overall lung cancer survival and responsiveness to platinum based chemotherapy agents. *Jr of Clinical Oncology* 24; 10004.
- Mathiaux, J., Le Morvan, V., Pulido, M., Jougon, J., Begueret, H., *et al.* (2011). Role of DNA repair gene polymorphisms in the efficiency of platinum-based adjuvant chemotherapy for non-small cell lung cancer. *Mol. Diagn. Ther* 15; 159–166.

- Matsunaga, T., Park, C. H., Bessho, T., Mu, D. and Sancar, A. (1996). Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. *J. Biol. Chem.* 271; 11047-11050.
- Matullo, G., Dunning, A. M., Guarrera, S., Baynes, C., Polidoro, S., Garte, S., *et al.* (2006). DNA repair polymorphisms and cancer risk in non-smokers in a cohort study. *Carcinogenesis* 27; 997–1007.
- Matullo, G., Guarrera, S., Carturan, S., Peluso, M., Malaveille C, Davico L, *et al.* (2001). DNA repair gene polymorphisms, bulky DNA adducts in white blood cell and bladder cancer in a case-control study. *Int. J. Cancer* 92; 562–567.
- Matutani, C., Araki, M., Sugawara, K., van der Spek, P.J., Yamada, A., Uchida, A., Maekawa, T., Bootsma, D., Hoeijmakers, J. H. J., Hanaoka, F (2006). Identification and characterization of XPC binding domain of hHR23B. *Mol Cell Biol* 17; 6915–23.
- McCarthy, W. J., Meza, R., Jeon, J. and Moolgavkar, S. H. (2012). Lung cancer in never smokers: Epidemiology and risk prediction models. *Risk Analysis* 32(s1).
- McGlynn, P. and Lloyd, R.G. (2002). Recombinational repair and restart of damaged replication forks. *Nat. Rev. Mol. Cell Biol.* 3; 859-870.
- McWilliams, R. R., Bamlet, W. R., Cunningham, J. M., Goode, E. L., de Andrade, M., *et al.* (2008). Polymorphisms in DNA repair genes, smoking, and pancreatic adenocarcinoma risk. *Cancer Res* 68; 4928–4935
- Mechanic, L. E., Millikan, R. C., Player, J., de Cotret, A. R., Winkel, S., Worley, K., Heard, K., Heard, K., Tse, C. K., Keku, T. (2006). Polymorphisms in nucleotide excision repair genes, smoking and breast cancer in African Americans and whites: a population-based case-control study. *Carcinogenesis* 27; 1377–1385.
- Mechanic, L. E., Millikan, R. C., Player, J., de Cotret, A. R., Winkel, S., *et al.* (2006). Polymorphisms in nucleotide excision repair genes, smoking and breast cancer in African Americans and whites: a population-based case-control study. *Carcinogenesis* 27; 1377–1385.
- Melville, A. and Eastwood, A. (1998). Management of lung cancer. *Quality in health care: QHC* 7(3); 170.

- Metro, M., Chiari, R., Mare, M., Giannarelli, D., Tofanetti, F. R., Minotti, V., et al. (2011). Carboplatin plus pemetrexed for platinum-pretreated, advanced non-small cell lung cancer: a retrospective study with pharmacogenetic evaluation. *Cancer Chemother. Pharmacol.* 68(6); 1405-12.
- Miller, K., Karagas, M. R., Kraft, P., et al. (2006). XPA, haplotypes, and risk of basal and squamous cell carcinoma. *Carcinogenesis* 27; 1670–1675.
- Ming-Shiean, H., Yu, J. C., Wang, H. W., et al. (2010). Synergistic effects of polymorphisms in DNA repair genes and endogenous estrogen exposure on female breast cancer risk. *Ann Surg Oncol.* 17; 760–771.
- 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* 33(3); 139-145.
- Misra, R. R., Ratnasinghe, D., Tangrea, J. A., et al. (2003). Polymorphisms in the DNA repair genes XPD, XRCC1, XRCC3, and APE/ref-1, and the risk of lung cancer among male smokers in Finland. *Cancer Lett* 191(2); 171—8.
- Mittal, R. D. and Mandal, R. K. (2012). Genetic variation in nucleotide excision repair pathway genes influence prostate and bladder cancer susceptibility in north Indian population. *Ind.J. Hum. Genet.* 18; 47–55.
- Miyamoto, I., Miura, N., Niwa, H., Miyazaki, J. and Tanaka, K. (1992). Mutational analysis of the structure and function of the xeroderma pigmentosum group A complementing protein. Identification of essential domains for nuclear localization and DNA excision repair. *J. Biol. Chem.* 267; 12182-12187.
- Mocquet, V., Laine, J. P., Riedl, T., Yajin, Z., Lee, M. Y. and Egly, J. M. (2008). Sequential recruitment of the repair factors during NER : the role of XPG in initiating the resynthesis step. *EM - BO J.* 27; 155-167.
- Molina, J. R., Yang, P., Cassivi, S. D. (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 83; 584-94.
- Moller, P., Wallin, H., Dybdahl, M., Frentz, G. and Nexø, B. A. (2000). Psoriasis patients with basal cell carcinoma have more repair-mediated DNA strandbreaks after UVC damage in lymphocytes than psoriasis patients without basal cell carcinoma. *Cancer Lett.* 151; 187–192.

- Monzo, M., Brunet, S., Urbano-Ispizua, A., Navarro, A., Perea, G., Esteve, J., et al. (2006). CETLAM Genomic polymorphisms provide prognostic information in intermediate-risk acute myeloblastic leukemia. *Blood* 107; 12.
- Moreno, V., Gemignani, F., Landi, S., et al. (2006). Polymorphisms in genes of nucleotide and base excision repair: risk and prognosis of colorectal cancer. *Clin Cancer Res* 12; 2101-8.
- Morita, R., Nakane, S., Shimada, A., et al. (2010). Molecular mechanisms of the whole DNA repair system: a comparison of bacterial and eukaryotic systems.
- Müller-Hermelink, H. K., Engel, P., Kuo, T. T., Ströbel, P., Marx, A., Harris, N. L., & Harris, C. C. (2004). Pathology and Genetics, Tumours of the lung, pleura, thymus and heart. *World Health Organization Classification of Tumors, IARC Press, Lyon, France*, 146-147.
- Nagaishi, C. and Nagasawa, N. (1972). Functional anatomy and histology of the lung. *University Park Press*.
- Namkung, J., Kim, K., Yi, S., et al. (2009). New evaluation measures for multifactor dimensionality reduction classifiers in gene-gene interaction analysis. *Bioinformatics* 25; 338–345.
- Narter, K. F., Ergen, A., Agachan, B., Gormus, U., Timirci, O., et al. (2009). Bladder cancer and polymorphisms of DNA repair genes (XRCC1, XRCC3, XPD, XPG, APE1, hOGG1). *Anticancer Res.* 29; 1389–1393.
- *New cancer cases in India by 2020* (2015). URL: <http://icmr.nic.in/icmrsql/archive/2016/7.pdf>
- 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 cases with surgical specimens. *The American Journal of Surgical Pathology* 26(9); 1184-1197.
- Nishikawa, A., Mori, Y., Lee, I. S., Tanaka, T. and Hirose, M. (2004). Cigarette smoking, metabolic activation and carcinogenesis. *Curr Drug Metab.* 5(5); 363-73
- Noronha, V., Pinninti, R., Patil, V. M., Joshi, A. and Prabhash, K. (2016). Lung cancer in the indian subcontinent. *South Asian Journal of Cancer* 5(3); 95.

- Omiecinski, C. J., Vanden Heuvel, J. P., Perdew, G. H. and Peters, J. M. (2010). Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. *Toxicological Sciences* 120(suppl 1); S49–S75.
- Pabalan, N., Francisco-Pabalan, O., Sung, L., *et al.* (2010). Meta-analysis of two ERCC2 (XPD) polymorphisms, asp312asn and lys751gln, in breast cancer. *Breast Cancer Res Treat* 124; 531-41.
- Palli, D., Polidoro, S., D’Errico, M., *et al.* (2010). Polymorphic DNA repair and metabolic genes: a multigenic study on gastric cancer. *Mutagenesis* 25(6); 569–575.
- Pani, B. and Nudle, E. (2017). Mechanistic insights into transcription coupled DNA repair. *DNA repair* 56; 42-50.
- Pardini, B., Naccarati, A., Novotny, J., Smerhovsky, Z., Vodickova, L., *et al.* (2008). DNA repair genetic polymorphisms and risk of colorectal cancer in the Czech Republic. *Mutat Res.* 638; 146–153.
- Parikh, P. M., Ranade, A. A., Govind, B., *et al.* (2016). Lung cancer in India: Current status and promising strategies. *South Asian Journal of Cancer* 5(3); 93-95.
- Park, J. Y., Lee, S. Y., Jeon, H. S., *et al.* (2002). Lys751Gln polymorphism in the DNA repair gene XPD and risk of primary lung cancer. (Letter). *Lung Cancer* 36; 15–16.
- Park, J.Y. (2002) Genetic susceptibility to lung cancer in Koreans. *Cancer Res Treatment* 34(1); 8–12.
- Park, J.Y., Park, S.H., Choi, J.E., Lee, S.Y., Jeon, H.S., Cha, S.I., Kim, C.H., Park, J.H., Kam, S., Park, R.W., Kim, I.S., Jung, T.H. (2002). Polymorphisms of the DNA repair gene xeroderma pigmentosum group A and risk of primary lung cancer. *Cancer Epidemiol Biomarkers Prev* 11(10); 993–997.
- Parkin, D. M., Whelan, S. L., Ferlay, J., Teppo, L., & Thomas, D. B. (2002). Cancer incidence in five continents. Vol. VIII. *IARC Scientific Publication*, No. 155. Lyon, France: International Agency for Research on Cancer.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol.* 49; 1373–1379.

- Phillips, D. H. (2002). Smoking-related dna and protein adducts in human tissues. *Carcinogenesis* 23(12); 1979–2004.
- Pillot, G. A., Read, W. L., Hennenfent, K. L., Pharm, D., et al. (2006). A Phase II Study of Irinotecan and Carboplatin in Advanced Non-small Cell Lung Cancer with PharmacogenomicAnalysis: Final Report. *Journal of Thoracic Oncology* 1(9).
- Popanda, O., Schattenberg, T., Phong, C. T., Butkiewicz, D., et al. (2004). Specific combinations of DNA repair gene variants and increased risk for non-small cell lung cancer. *Carcinogenesis* 25(12); 2433-41.
- Porter, P. C., Isabel Mellon, J., States, C. (2005). XP-A cells complemented with Arg228Gln and Val234Leu polymorphic XPA alleles repair BPDE-induced DNA damage better than cells complemented with the wild type allele. *DNA Repair* 4; 341–349.
- Povey, J. E., Darakhshan, F., Robertson, K., Bisset, Y., Mekky, M., et al. (2007). DNA repair gene polymorphisms and genetic predisposition to cutaneous melanoma. *Carcinogenesis* 28; 1087–1093.
- Provencio, M., Camps, C., Cobo, M., et al. (2012). Prospective assessment of XRCC3, XPD and Aurora kinase A single-nucleotide polymorphisms in advanced lung cancer. *Cancer Chemother. Pharmacol.* 70; 883–890.
- Przybylowska, K., Kabzinski, J., Sygut, A., Dziki, L., Dziki, A. and Majsterek, I. (2013). An association selected polymorphisms of xrcc1, ogg1 and mutyh gene and the level of efficiency oxidative dna damage repair with a risk of colorectal cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 745; 6–15.
- Qian, B., Zhang, H., Zhang, L., Zhou, X., Yu, H., Chen, K. (2011). Association of genetic polymorphisms in DNA repair pathway genes with non-small cell lung cancer risk. *Lung Cancer* 73; 138–46.
- Raaschou-Nielsen, O., Sorensen, M., Overvad, K., Tjonneland, A., Vogel, U. (2008). Polymorphisms in nucleotide excision repair genes, smoking and intake of fruit and vegetables in relation to lung cancer. *Lung Cancer* 59; 171–179
- RadiologyInfo.org (2017), ‘Lung cancer treatment’.

- Rajaraman, P., Bhatti, P., Doody, M. M. *et al.* (2008). Nucleotide excision repair polymorphisms may modify ionizing radiation-related breast cancer risk in US radiologic technologists. *Int. J. Cancer* 123, 2713–2716.
- Rajaraman, P., Bhatti, P., Doody, M. M., Simon, S. L., Weinstock, R. M., et al. (2008). Nucleotide excision repair polymorphisms may modify ionizing radiation-related breast cancer risk in US radiologic technologists. *International Journal of Cancer* 123; 2713–2716.
- Rajaraman, P., Hutchinson, A., Wichner, S., Black, P. M., Fine, H. A., et al. (2010). DNA repair gene polymorphisms and risk of adult meningioma, glioma, and acoustic neuroma. *Neuro Oncol* 12; 37–48.
- Rami-Porta, R., Crowley, J. J. and Goldstraw, P. (2009). Review the revised tnm staging system for lung cancer. *Ann Thorac Cardiovasc Surg.* 15(1); 5.
- Reardon, J. T. and Sancar, A. (2002). Molecular anatomy of the human excision nuclease assembled at sites of DNA damage. *Mol. Cell. Biol.* 22; 5938-5945.
- Reddy, K. S. and Gupta, P. C. (2004). Report on Tobacco Control in India (New Delhi, India) New Delhi, India: Ministry of Health and Family Welfare.
- Ren. S., Zhou, S., Wu, F., Zhang, L., Li, X., et al. (2012). Association between polymorphisms of DNA repair genes and survival of advanced NSCLC patients treated with platinum-based chemotherapy. *Lung Cancer* 75; 102–109.
- Riedl, et al. 2003. The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J* 22 ; 5293-5303.
- Riely, G. J., Kris, M. G., Rosenbaum, D., Marks, J., Li, A., Chitale, D. A., Nafa, K., Riedel, E. R., Hsu, M., Pao, W. *et al.* (2008). Frequency and distinctive spectrum of kras mutations in never smokers with lung adenocarcinoma. *Clinical Cancer Research* 14(18); 5731–5734.
- Ripley, B. D. (1996). Pattern recognition and neural networks. *Cambridge University Press, Cambridge.*
- Rivlin, N., Brosh, R., Oren, M. and Rotter, V. (2011). Mutations in the p53 tumor suppressor gene: important milestones at the various steps of tumorigenesis. *Genes & cancer* 2(4); 466–474.

- Rom, W. N., Hay, J. G., Lee, T. C., Jiang, Y. and Tchou-Wong, K.-M. (2000). Molecular and genetic aspects of lung cancer. *American journal of respiratory and critical care medicine* 161(4); 1355–1367.
- Roos, W. P. and Kaina, B. (2006). Dna damage-induced cell death by apoptosis. *Trends in molecular medicine* 12(9); 440–450.
- 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. *Archives of toxicology* 88(8); 1479–1490.
- Rouissi, K., Ouerhani, S., Hamrita, B., Bougatef, K., Marrakchi, R., *et al.* (2011). Smoking and Polymorphisms in Xenobiotic Metabolism and DNA Repair Genes are Additive Risk Factors Affecting Bladder Cancer in Northern Tunisia. *Pathol Oncol Res.*
- Ryu, J. S., Hong, Y. C., Han, H. S., Lee, J. E., Kim, S., *et al.* (2004). Association between polymorphisms of ERCC1 and XPD and survival in non-small-cell lung cancer patients treated with cisplatin combination chemotherapy. *Lung Cancer* 44; 311–316.
- Sak, S. C., *et al.* (2005). The polyAT, intronic IVS11-6 and Lys939Gln XPC polymorphisms are not associated with transitional cell carcinoma of the bladder. *Br. J. Cancer*, 92; 2262–2265.
- Sakiyama, T., Kohno, T., Mimaki, S., Ohta, T., Yanagitani, N., *et al.* (2005). Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. *Int J Cancer* 114; 730–737.
- Sakoda, L. C., Loomis, M. M., Doherty, J. A., Julianto, L., Barnett, M. J., Neuhausser, M. L., Thornquist, M. D., Weiss, N. S., Goodman, G. E., Chen, C. (2012). Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers.
- Sakoda, L. C., Loomis, M. M., Doherty, J. A., Julianto, L., Barnett, M. J., *et al.* (2012). Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers. *Int J Mol Epidemiol Genet* 3; 1–17.
- Sakoda, L. C., Loomis, M. M., Doherty, J. A., Julianto, L., Barnett, M. J., Neuhausser, M. L., *et al.* (2012). Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers. *Int J Mol Epidemiol Genet.* 3; 1–17.

- Sakoda, L. C., Loomis, M. M., Doherty, J. A., Julianto, L., Barnett, M. J., *et al.* (2012). Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers. *Int J Mol Epidemiol Genet*, 3; 1–17.
- Samson, M., Singh, S. S., Rama, R., *et al.* (2011). XPD Lys751 increases the risk of breast cancer. *Oncol Lett*, 2; 155–159.
- Sancar, A. (1996). DNA excision repair. *Annu. Rev. Biochem.* 65; 43-81.
- Santos, L. S., Gomes, B. C., Gouveia, R., *et al.* (2013). The role of CCNH Val270Ala (rs2230641) and other nucleotide excision repair polymorphisms in individual susceptibility to well-differentiated thyroid cancer. *Oncol Rep* 30; 2458-66
- Santos, P. A. R. d., Rocha, R. S. d., Pipkin, M., Silveira, M. L. d., Cypel, M., Rios, J. O. and 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(6); 647–654.
- Sanyal, S., Festa, F., Sakano, S., Zhang, Z., Steineck, G., *et al.* (2004). Polymorphisms in DNA repair and metabolic genes in bladder cancer. *Carcinogenesis* 25; 729–734.
- Sanyal, S., Verdier, P. J., Steineck, G., Larsson, P., Onelo, E., *et al.* (2007). Polymorphisms in XPD, XPC and the risk of death in patients with urinary bladder neoplasms. *Acta Oncologica* 46; 3141.
- Sarasin, A. and Sary, A. (2007). New insights for understanding the transcription-coupled repair pathway. *DNA Repair (Amst)* 6; 265-269.
- Sarker, A.H., Tsutakawa, S.E., Kostek, S., Ng, C., Shin, D.S., Peris, M., Campeau, E., Tainer, J.A., Nogales, E. and Cooper, P.K. (2005). Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH : insights for transcription-coupled repair and Cockayne Syndrome. *Mol. Cell* 20, 187-198.
- Sasco, A., Secretan, M. and Straif, K. (2004). Tobacco smoking and cancer: a brief review of recent epidemiological evidence. *Lung cancer* 45; S3–S9.
- Schlichting, C. D. and Pigliucci, M. (1998). Phenotypic evolution: a reaction norm perspective. Sinauer Associates, Sunderland, MA.

- Seker, H., *et al.* (2001). Functional significance of XPD polymorphic variants: attenuated apoptosis in human lymphoblastoid cells with the XPD 312 Asp/Asp genotype. *Cancer Res.*
- Shao, M., Ma, H., Wang, Y. *et al.* (2008). Polymorphisms in excision repair cross-complementing group 4 (ERCC4) and susceptibility to primary lung cancer in a Chinese Han population. *Lung Cancer* 60; 332–339.
- Shen, B., Nolan, J. P., Sklar, L. A., Park, M. S. (1996). Essential amino acids for substrate binding and catalysis of human flap endonuclease. *J. Biol. Chem.*, 271: 9173– 6.
- Shen, J., Desai, M., Agrawal, M., Kennedy, D. O., Senie, R. T., Santella, R. M., Terry, M. B. (2006). Polymorphisms in nucleotide excision repair genes and DNA repair capacity phenotype in sisters discordant for breast cancer. *Cancer Epidemiol Biomarkers Prev* 15; 1614–1619.
- Shen, M. R., Jones, I. M., Mohrenweiser, H. (1998). Non conservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res.*, 58; 604–8.
- Shen, M. R., Jones, I. M., Mohrenweiser, H. (1998). Non conservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res.* 58; 604–8.
- Shen, M. R., Jones, I. M., Mohrenweiser, H. (1998). Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res.*, 58; 604–608.
- Shen, M., Berndt, S. I., Rothman, N., Demarini, D. M., Mumford, J. L., *et al.* (2005). Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. *Int J Cancer* 116; 768–773.
- Shi, Q., Guan, M., Wang, Y., *et al.* (2018). Survival analysis of patients with advanced non-small cell lung cancer receiving tyrosine kinase inhibitor (TKI) treatment: A multi-center retrospective study. *Thoracic Cancer* 9(2); 278-283.
- Shi, T. Y., He, J., Qiu, L. X., *et al.* (2012). Association between XPF polymorphisms and cancer risk: a meta-analysis. *PLoS One* 7; e38606.

- Shi, T. Y., He, J., Qiu, L. X., *et al.* (2012). Association between XPF polymorphisms and cancer risk: a meta-analysis. *PLoS One*, 7; e38606.
- Sholl, M. (2016). The molecular pathology of lung cancer. *Elsevier* 9(3); 353–78.
- Shrivastav, M., De Haro, L. P. and Nickoloff, J. A. (2008). Regulation of dna doublestrand break repair pathway choice. *Cell research* 18(1); 134.
- Smith, T. R., Levine, E. A., Freimanis, R. I. *et al.* (2008) Polygenic model of DNA repair genetic polymorphisms in human breast cancer risk. *Carcinogenesis*, 29; 2132–2138.
- Smith, T. R., Levine, E. A., Freimanis, R. I., Akman, S. A., Allen, G.O., *et al.* (2008). Polygenic model of DNA repair genetic polymorphisms in human breast cancer risk. *Carcinogenesis* 29; 2132–2138.
- Song, X. *et al.* (2013). Variants in nucleotide excision repair core genes and susceptibility to recurrence of squamous cell carcinoma of the oropharynx. *Int J Cancer*, 133; 695–704.
- Spitz, M. R., Wu, X., Wang, Y., Wang, L. E., *et al.* (2001). Modulation of nucleotide excision capacity by XPD polymorphisms in lung cancer patients. *Cancer Res.* 61; 1354–7.
- Staresinic, L., Fagbemi, A. F., Enzlin, J. H., Gourdin, A. M., Wijgers, N., Dunand-Sauthier, I., Giglia-Mari, G., Clarkson, S. G., Vermeulen, W. and Scharer, O. D. (2009). Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J.* 28; 1111-1120.
- Steck, S. E., Butler, L. M., Keku, T., *et al.* (2014). Nucleotide excision repair gene polymorphisms, meat intake and colon cancer risk. *Mutat Res Fundam Mol Mech Mutagen* 762; 24-31.
- Steenland, K., Mannetje, A., Boffetta, P., Stayner, L., Attfield, M., Chen, J., Dosemeci, M., DeKlerk, N., Hnizdo, E., Koskela, R. *et al.* (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.
- Stern, M. C., *et al.* (2009). Polymorphisms in DNA Repair Genes, Smoking, and Bladder Cancer Risk: Findings from the International Consortium of Bladder Cancer. *Cancer Res.*, 69; 6857–6864.

- Strom, S. S., Estey, E., Otschoorn, U. M. and Garcia-Manero, G. (2010). Acute myeloid leukemia outcome: role of nucleotide excision repair polymorphisms in intermediate risk patients. *Leuk Lymphoma* 51; 598–605.
- Sturgis EM, Zheng R, Castillo EJ, Eicher SA, Chen M, Strom SS, *et al.* XPD/ERCC2 polymorphisms and risk of head and neck cancer: a case-control analysis. *Carcinogenesis* 21, 2219–2223 (1998).
- Sturgis, E. M., Zheng, R., Castillo, E. J., Eicher, S. A., Chen, M., Strom, S. S., *et al.* (1998). XPD/ERCC2 polymorphisms and risk of head and neck cancer: a case-control analysis. *Carcinogenesis* 21; 2219–2223.
- Su, Y., Zhang, H., Xu, F., Kong, J., Yu, H., Qian, B. (2015). DNA repair gene polymorphisms in relation to nonsmall cell lung cancer survival. *Cell Physiol Biochem* 36; 1419–1429.
- Sugawara, K., Okamoto, T., Shimizu, Y., Masutani, C. I., Wai, S. and Hanaoka, F. (2001). A multistep damage recognition mechanism for global genomic nucleotide excision repair. *GenesGenes Dev.* 15; 50
- Sugimura, T., Kumimoto, H., Tohnai, I., Fukui, T., Matsuo, K., *et al.* (2006). Gene-environment interaction involved in oral carcinogenesis: Molecular epidemiological study for metabolic and DNA repair gene polymorphisms. *Journal of Oral Pathology and Medicine* 35; 11–18.
- Sugimura, T., Kumimoto, H., Tohnai, I., Fukui, T., Matsuo, K., Tsurusako, S., Mitsudo, K., Ueda, M., Tajima, K. and Ishizaki, K. (2006). Gene-environment interaction involved in oral carcinogenesis: molecular epidemiological study for metabolic and DNA repair gene polymorphisms. *J. Oral. Pathol. Med.* 35; 11-18.
- Sugiyama, E., Kaniwa, N., Kim, S. R., *et al.* (2007). Pharmacokinetics of gemcitabine in Japanese cancer patients: the impact of a cytidine deaminase polymorphism. *J Clin Oncol* 25; 32– 42.
- Sullivan, I., Salazar, J., Majem, M., Pallarés, C., Del Río, E., Páez, D., Baiget, M., Barnadas, A. (2014). Pharmacogenetics of the DNA repair pathways in advanced non-small cell lung cancer patients treated with platinum-based chemotherapy. *Cancer Lett.* 353; 160–6.

- Sun, X., Li, F., Sun, N. et al. (2009). Polymorphisms in XRCC1 and XPG and response to platinum-based chemotherapy in advanced non-small cell lung cancer patients. *Lung Cancer* 65; 230-236.
- Sun, Z. et al. (2015). Genotype-phenotype correlation of xeroderma pigmentosum in a Chinese Han population. *The British journal of dermatology* 172(4):1096–102.
- Templeton, A. R. (2000). Epistasis and complex traits. In: Wade M, Brodie B III, Wolf J (eds) Epistasis and evolutionary process. *Oxford University Press*, 41–57
- Tamura, T., Kurishima, K., Nakazawa, K., Kagohashi, K., Ishikawa, H., Satoh, H. and Hizawa, N. (2015). Specific organ metastases and survival in metastatic non-small-cell lung cancer. *Molecular and clinical oncology* 3(1), 217–221.
- Tantin, D. (1998). RNA polymerase II elongation complexes containing the Cockayne syndrome group B protein interact with a molecular complex containing the transcription factor IIH components xeroderma pigmentosum B and p62. *J. Biol. Chem.* 273; 27794-27797.
- Tapias, A., Auriol, J., Forget, D., Enzlin, J. H., Scharer, O. D., Coin. F., Coulombe, B. and Egly, J. M. (2004). Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *J. Biol Chem.* 279; 19074-19083.
- Taxak, N. and Bharatam, P. V. (2014). Drug metabolism a fascinating link between chemistry and biology.
- TerryDesJardins. (2008). Cardiopulmonary Anatomy Physiology Essentials of Respiratory Care. *Thomson Delmar Learning*.
- Thomas, A., Rajan, A. and Giaccone, G. (2012). Tyrosine kinase inhibitors in lung cancer. *Hematology/Oncology Clinics* 26(3); 589–605.
- Three-Year Report of Population Based Cancer Registries 2012-2014. (March, 2016). Retrieved October 31, 2017, from http://www.ncrpinidia.org/ALL_NCRP_REPORTS/PBCR_REPORT_2012_2014/index.htm
- Tibaldi, C., Giovannetti, E., Vasile, E., Mey, V., Laan, A. C., et al. (2008). Correlation of CDA, ERCC1, and XPD polymorphisms with response and survival in gemcitabine/cisplatin-treated advanced non-small cell lung cancer patients. *Clin Cancer Res* 14; 1797–1803.

- Tomasini, P., Barlesi, F., Mascaux, C. and Greillier, L. (2016). Pemetrexed for advanced stage nonsquamous non-small cell lung cancer: latest evidence about its extended use and outcomes, *Ther Adv Med Oncol* 8(3); 198–208.
- Tomescu, D., Kavanagh, G., Ha, T., Campbell, H., Melton, D. W. (2001). Nucleotide excision repair gene XPD polymorphisms and genetic predisposition to melanoma. *Carcinogenesis* 22; 403–408.
- Tornaletti, S., Maeda, L.S., Lloyd, D.R., Reines, D. and Hanawalt, P.C. (2001). Effect of thymine glycol on transcription elongation by T7 RNA polymerase and mammalian RNA polymerase II. *J. Biol. Chem.* 276; 45367-45371.
- Travis et al. (2015). The 2015 World Health Organization Classification of Lung Tumors Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *Journal of Thoracic Oncology*.
- Travis, W. D. and Harris, C. (2004). Pathology and genetics of tumours of the lung, pleura, thymus and heart. *Feance: IARC Press*.
- Truglio, J. J., Croteau, D. L., Van Houten, B., Kisker, C. (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical Reviews* 106 (2); 233–252.
- Tsodikov, O. V., Enzlin, J. H., Scharer, O. D. *et al.* (2005). Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc Natl AcadSci.*, 102; 11236–11241.
- Tsuruo, T., Matsuzaki, T., Matsushita, M., *et al.* (1988). Antitumor effect of CPT-11, a new derivative of camptothecin, against pleiotropic drug resistant tumors in vitro and in vivo. *Cancer Chemother. Pharmacol.* 21; 71–74 .
- Tu, J., Inthavong, K., Ahmadi, G. (2013). Computational Fluid and Particle Dynamics in the Human Resoiratory System.
- *Type of Tumors* (2016). URL: <http://pathology.jhu.edu/pc/BasicTypes1.php?area=ba>
- Uchida, A., Sugasawa, K., Masutani, C., Dohmae, N., Araki, M., Yokoi, M., Ohkuma, Y., Hanaoka, F. (2002). The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH. *DNA Repair* 1; 449–461.

- Vaezi1, A., Wang, A. Z., Buch, S., Gooding, W., Wang, L., Seethala, R. R., Weaver, D. T., et al. (2011). XPF expression correlates with clinical outcome in squamous cell carcinoma of the head and neck. *Clin Cancer Res.* 15: 17(16); 5513–5522.
- Vincent, M. D. (2009). Optimizing the management of advanced non-small-cell lung cancer: a personal view. *Current Oncology* 16(4); 9-21.
- Vinolas, N., Provencio, M., Reguart, N., Cardenal, F., Alberola, V., et al. (2011). Single nucleotide polymorphisms in MDR1 gen correlates with outcome in advanced non-small-cell lung cancer patients treated with cisplatin plus vinorelbine. *Lung Cancer* 71; 191–198.
- Vodicka, P., Kumar, R., Stetina, R., Sanyal, S., Soucek, P., Haufröid, V., Dusinska, M., Kuricova, M., Zamecnikova, M., Musak, L., Buchancova, J., Norppa, H., et al. (2004). Genetic polymorphisms in DNA repair genes and possible links with DNA repair rates, chromosomal aberrations and single-strand breaks in DNA. *Carcinogenesis* 25; 757–63.
- Vogel, U., Overvad, K., Wallin, H., Tjønneland, A., Nexø, B.A. & Raaschou-Nielsen, O. (2005). Combinations of polymorphisms in XPD, XPC and XPA in relation to risk of lung cancer. *Cancer Lett. Vol.* 222 (1); 67-74.
- Volker, M., Mone, M. J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J. H., van Driel, R., van Zeeland, A. A., Mullenders, L. H. (2001). Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell* 8(1); 213–224.
- Wakasugi, M., and Sancar, A. (1999). Order of assembly of human DNA repair excision nuclease. *J. Biol. Chem.* 274; 18759–18768.
- Wakasugi, M., Reardon, J. T., Sancar, A. (1997). The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair. *J. Biol. Chem.* 272; 16030–4.
- Walter, J. B. and Pryce, D. M. (1955). The histology of lung cancer. *Thorax* 10(2); 107.
- Wang, F., Zhang, S. D., Xu, H. M., et al. (2016). XPG rs2296147 T>C polymorphism predicted clinical outcome in colorectal cancer. *Oncotarget* 7(10); 11724-11732.
- Wang, L. E., Li, C., Strom, S. S., Goldberg, L. H., Brewster, A., et al. (2007). Repair capacity for UV light induced DNA damage associated with risk of nonmelanoma skin cancer and tumor progression. *Clin Cancer Res.* 13; 6532–6539.

- Wang, X. J., Xu, C. A., Zhang, Y., *et al.* (2012). Relationship between ERCC1 and XPD genetic polymorphism and sensitivity of platinum-based chemotherapy in advanced non-small cell lung cancer. *Chin Gen Pract.* 15; 1010-1014.
- Wang, X. F., Liu, S., Shao, Z. K. (2013). Effects of polymorphisms in nucleotide excision repair genes on glioma risk in a Chinese population. *Gene* 529; 317-20.
- Wang, Y. G., Xing, D. Y., Tan, W., Wang, L. J., Tang, P. Z., *et al.* (2003). Poly(AT) polymorphism in DNA repair gene XPC and lung cancer risk. *Zhonghua Zhong Liu Za Zhi* 25; 555–557.
- Weaver, D. A., Crawford, E. L., Warner, K. A., Elkhairi, F., Khuder, S. A. and Willey, J. C. (2005). ABCC5, ERCC2, XPA and XRCC1 transcript abundance levels correlate with cisplatin chemoresistance in non-small cell lung cancer cell lines. *Molecular Cancer* 4; 18.
- Wei, Q., Cheng, L., Amos, C. I., *et al.* (2000). Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. *J Natl Cancer Inst.* 92; 1764-1772.
- Weissferdt, A., & Moran, C. A. (2012). Diagnostic Pathology of Pleuropulmonary Neoplasia. *Springer Science & Business Media*. Major cancer types by deaths worldwide (2015). Statistic. Retrieved October 31, 2017, from <https://www.statista.com/statistics/288580/number-of-cancer-deaths-worldwide-by-type/>
- Wiencke, J. K., Kelsey, K. T., Varkonyi, A., Semey, K., Wain, J. C., Mark, E. and Christiani, D. C. (1995). Correlation of dna adducts in blood mononuclear cells with tobacco carcinogen-induced damage in human lung. *Cancer research* 55(21); 4910–4914.
- Wiesmüller, L., Ford, J. M., Schiestl, R. H. (2002). DNA Damage, Repair, and Diseases. *Journal of Biomedicine and Biotechnology* 2(2); 45.
- Williamson, S. K., Crowley, J. J., Lara, P. N., McCoy, J., *et al.* (2005). Phase III Trial of Paclitaxel Plus Carboplatin With or Without Tirapazamine in Advanced Non–Small-Cell Lung Cancer: Southwest Oncology Group Trial S0003 23(36).
- Wistuba, I. I., Mao, L. and Gazdar, A. F. (2002). Smoking molecular damage in bronchial epithelium. *Oncogene* 21(48); 7298.

- Wood, R. D. (1999). DNA damage recognition during nucleotide excision repair in mammalian cells. *Biochemie* 81; 39-44.
- Wu, F., Zhang, J., Liu, Y., Zheng, Y., Hu, N. (2013). HIF1alpha genetic variants and protein expressions determine the response to platinum based chemotherapy and clinical outcome in patients with advanced NSCLC. *Cell Physiol Biochem* 32; 1566-1576.
- Wu, W., Li, H., Wang, H., Zhao, X., Gao, Z., Qiao, R., Zhang, W., Qian, J., Wang, J., Chen, H., Wei, Q., Han, B., Lu, D. (2012). Effect of polymorphisms in XPD on clinical outcomes of platinum-based chemotherapy for Chinese non-small cell lung cancer patients. *PLoS One* 7; e33200.
- Wu, X., Gu, J., Grossman, H. B., Amos, C. I., Etzel, C., *et al.* (2006). Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. *Am J Hum Genet.* 78:464–479.
- Wu, X., Zhao, H., Wei, Q., Amos, C. I., Zhang, K., Guo, Z., Qiao, Y., Hong, W. K., Spitz, M. R. (2003). XPA polymorphism associated with reduced lung cancer risk and a modulating effect on nucleotide excision repair capacity. *Carcinogenesis* 24(3); 505–509.
- Wu, X., Zhao, H., Wei, Q., Amos, C.I., Zhang, K., Guo, Z., Qiao, Y., Hong, W.K., Spitz, M.R. (2003). XPA polymorphism associated with reduced lung cancer risk and a modulating effect on nucleotide excision repair capacity. *Carcinogenesis* 24(3); 505–509.
- Wyss, A. B., Herring, A. H., Avery, C. L., *et al.* (2013). Single-nucleotide polymorphisms in nucleotide excision repair genes, cigarette smoking, and the risk of head and neck cancer. *Cancer Epidemiol Biomarkers Prev* 22; 1428-45.
- Xing, D., Tan, W., Wei, Q., Lin, D. (2002). Polymorphisms of the DNA repair gene XPD and risk of lung cancer in a Chinese population. *Lung Cancer* 38; 123–9.
- Xue, J., Yang, S., Seng, S. (2014). Mechanisms of Cancer Induction by Tobacco-Specific NNK and NNN. *Cancers* 6(2);1138-1156.
- Yan, D., Liang, X. H., Ding, W., Xu, X. J., Wang, X. Y. (2018). Contribution of DNA repair xeroderma pigmentosum group D genotypes to pancreatic cancer risk in the Chinese Han population. *Genetics and Molecular Biology* 41(1); 18-26.
- Yan, Y., Liang, H., Light, M., Li, T., Deng, Y., Li, M., Li, S. and Qin, X. (2014). XPD Asp312Asn and Lys751Gln polymorphisms and breast cancer susceptibility: a meta-analysis. *Tumour Biol* 35; 1907-1915.

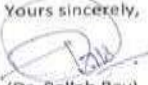
- Yang, L. M., Li, X. H., Bao, C. F. (2012). Glutathione S-transferase P1 and DNA polymorphisms influence response to chemotherapy and prognosis of bone tumors. *Asian Pac J Cancer Prev* 13; 5883-5886.
- Yao, C. Y., Huang, X. E., Li, C., Shen, H. B., Shi, M. Q., et al. (2009). Lack of influence of XRCC1 and XPD gene polymorphisms on outcome of platinum-based chemotherapy for advanced non small cell lung cancers. *Asian Pac J Cancer Prev* 10; 859–864.
- Yin, J., Li, J., Ma, Y., Guo, L., Wang, H., Vogel, U. (2005). The DNA repair gene ERCC2/XPD polymorphism Arg156Arg (A22541C) and the risk of lung cancer in a Chinese population. *Cancer Lett* 223; 219-226.
- Yin, M., Yan, J., Voutsina, A., Tibaldi, C., Christiani, D. C., et al. (2011). No evidence of an association of ERCC1 and ERCC2 polymorphisms with clinical outcomes of platinum-based chemotherapies in non-small cell lung cancer: a meta-analysis. *Lung Cancer* 72; 370–377.
- Yin, Q. H., Liu, C., Hu, J. B., et al (2013). XPD Lys751Gln and Asp312Asn polymorphisms and gastric cancer susceptibility: a meta-analysis of case-control studies. *Asian Pac J Cancer Prev* 14; 231-6.
- Yu, D. K., Wu, C., Tan, W., Lin, D. X. (2010). Functional XPF polymorphisms associated with lung cancer susceptibility in a Chinese population. *Front. Med. China* 1; 82–89.
- Yu, H., Liu, Z., Huang, YJ, Yin M, Wang LE, et al. (2011). Association between single nucleotide polymorphisms in *ERCC4* and risk of squamous cell carcinoma of the head and neck submitted.
- Yuli , Y., Zhe, S., Xia, W., Siqing, L., Zhenxuan, W., Yu-Hua, Z., Bing, S., Jun-Wei, C. (2013). XPG is a novel biomarker of clinical outcome in advanced non-small-cell lung cancer. *Pak J Med Sci.* 29; 762–767.
- Yuli, Y., Zhe, S., Xia, W., Siqing, L., Zhenxuan, W., Yu-Hua, Z., et al. (2013). XPG is a novel biomarker of clinical outcome in advanced non-small-cell lung cancer. *Pak J Med Sci* 29(3); 762–7.
- 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.

- Zeinolddiny, S., Skang, V., Landvik, N. E. (2009). The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. *Carcinogenesis* 30(8); 1368–71.
- Zhan, P., Wang, Q., Wei, S. Z., Wang, J., Qian, Q., Yu, L. K., et al. (2010). ERCC2/XPD Lys751Gln and Asp312Asn gene polymorphism and lung cancer risk: a meta-analysis involving 22 case–control studies. *J Thorac Oncol.* 5; 1337–45.
- Zhang, T., Sun, J., Lv, M., Zhang, L., Wang, X., Ren, J. C., Wang, B. (2013). XPG is predictive gene of clinical outcome in advanced non-small-cell lung cancer with platinum drug therapy. *Asian Pac J Cancer Prev.* 14; 701–705.
- Zhang, X. J., Liu, P. and Zhu F. (2014). Polymorphisms of DNA repair-related genes with susceptibility and prognosis of prostate cancer. *Genet Mol Res* 13; 4419-4424.
- Zhang, Y., Rohde, L. H., Wu, H. (2009). Involvement of nucleotide excision and mismatch repair mechanisms in double strand break repair. *Current Genomics* 10(4); 250–258.
- Zhang, Y., Sheng, J., Kang, S., Fang, W., Yan, Y., Hu, Z., Hong, S., Wu, X., Qin, T., Liang, W., Zhang, L. (2014). Patients with exon 19 deletion were associated with longer progression-free survival compared to those with L858R mutation after first-line EGFR-TKIs for advanced non-small cell lung cancer: a meta-analysis. *PLoS One* 9(9); e107161.
- Zhang, Y., Wang, X., Zhang, W. and Gong, S. (2013). An association between XPC Lys939Gln polymorphism and the risk of bladder cancer: a metaanalysis. *Tumor Biology* 34(2); 973-982.
- Zhang, Z. Y., Tian, X., Wu, R., Liang, Y., Jin, X. Y. (2012). Predictive role of ERCC1 and XPD genetic polymorphisms in survival of Chinese non-small cell lung cancer patients receiving chemotherapy. *Asian Pac J Cancer Prev* 13; 2583–2586.
- Zheng, H., Kang, M.F. (2013). Association between XPA A23G polymorphism and XPG His1104Asp polymorphism in lung cancer. *Guang Dong Medical Journal* 34; 413–6.
- Zhi, Y., et al. (2012). Interaction between Polymorphisms of DNA Repair Genes Significantly Modulated Bladder Cancer Risk. *Int. J. Med. Sci.* 9; 498–505.
- Zhou, L., Lu, Y., Yang, G. and Wu, J. (2014). Quantitative assessment of the association between XPC Lys939Gln polymorphism and cutaneous melanoma risk. *Tumour Biol* 35; 1427- 1432.

- Zhou, W., Liu, G., Miller, D. P., Thurston, S. W., Xu, L. L., Wain, J. C., et al. (2002). Gene–environment interaction for the ERCC2 polymorphisms and cumulative cigarette smoking exposure in lung cancer. *Cancer Res.* 62; 1377–1381.
- Zhu, C., et al. (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* 19(4); 556-66.
- Zhu, M. L., Hua, R. X., Zheng, L. (2013). Associations between polymorphisms of the XPC gene and lung cancer susceptibility: a meta-analysis. *Tumour Biol.* [Epub ahead of print].
- Zhu, M. L., Shi, T. Y., Hu, H. C., He, J., et al. (2012). Polymorphisms in the ERCC5 gene and risk of esophageal squamous cell carcinoma (ESCC) in Eastern Chinese populations. *PLoS One* 7; e41500.
- Zhu, X., Sun, X. C., Chen, B., Sun, N., et al. (2010). XPC Lys939Gln polymorphism is associated with the decreased response to platinum based chemotherapy in advanced non-small-cell lung cancer. *Chinese Medical Journal* 123(23); 3427-3432.
- Zhu, Y., Lai, M., Yang, H., Lin, J., Huang, M., Grossman, H. B., Dinney, C. P., Wu, X. (2007). Genotypes, haplotypes and diplotypes of XPC and risk of bladder cancer. *Carcinogenesis* 28; 698–703.
- Zhu, Y., et al. (2008). Genotypes, haplotypes and diplotypes of XPC and risk of bladder cancer. *Carcinogenesis* 28; 698–703.
- Zhu, Y., Spitz, M. R., Lei, L., Mills, G. B., Wu, X. (2001). A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter enhances lung cancer susceptibility. *Cancer Res.* 61; 7825–9.
- Zienolddiny, S., Campa, D., Lind, H., Ryberg, D., Skaug, V., Stangeland, L., Phillips, D. H., Canzian, F., Haugen, A. (2006). Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis* 27(3); 560–567.
- Zilfou, J. T. and Lowe, S. W. (2009). Tumor suppressive functions of p53. *Cold Spring Harbor perspectives in biology* 1(5); a001883.
- Zou, H. Z., Zhao, Y. Q. (2015). XPG polymorphisms are associated with prognosis of advanced non-small cell lung cancer treated with platinum-based doublet chemotherapy. *Genet Mol Res.* 14; 500–506.

Appendix

A.1 Ethics committee report

Institute Ethics Committee Postgraduate Institute of Medical Education and Research, Chandigarh	
Prof. KS Chugh <i>Chairman</i>	Prof. Pallab Ray <i>Convener</i>
	No: PGI/IEC/2014/..... Date: Aug 12, 2014
Chairman Prof. KS Chugh	
Members Justice VK Jhanji / Justice MR Agnihotri Dr BN Goswami / Dr Virendra Kumar Dr Girish Varshney Dr Promila Vasudeva Prof. Savita Malhotra Prof. R Kochhar Prof. Savita Kumari Prof. Vanita Suri Prof. S Radhika Prof. Ashim Das Prof. RK Dhiman Dr Jayashree M Dr L Kaman Dr JS Thakur Dr Nusrat Shafiq	Dr. Navneet Singh ✓ Department of Pulmonary Medicine, PGIMER, Chandigarh Title: Project entitled "Association of single Nucleotide Polymorphisms in the DNA Repair Genes and Survival of Advanced Non-Small-Cell Lung cancer patients treated with platinum-Based Doublet Chemotherapy." Project Number: P-106 Dear Dr Singh, We received five copies of your above mentioned project proposal, which has been assigned the project number P-106 . The project was discussed in the meeting of the Institute Ethics committee held on 09.11.2011. After due consideration and discussion, your project was CONDITIONALLY APPROVED . Subsequently we received the clarification on the issue raised by the Committee and now your project is APPROVED . It is understood that the study will be conducted under your direction strictly as per the submitted protocol. Any deviations from the approved protocol and study documents must be sent to the Ethics Committee for re-approval. Any adverse reaction or condition noted during the study period should be reported to the Ethics Committee immediately. A report of the study should be sent to the Ethics Committee yearly. All future correspondence with respect to this project must mention the project number . Kind Regards Yours sincerely,  (Dr. Pallab Ray) Copy to: Professor-in-Charge, Medical Education Cell, PGIMER, Chandigarh.
Convener Prof. Pallab Ray	
Contact Ph: +91 172 2755157 Email: iecpgi@gmail.com	

Office: Department of Medical Microbiology, Second Floor, Research Block A, PGIMER, Chandigarh.

A.2 PATIENT INFORMATION SHEET

PATIENT INFORMATION SHEET

Department of Biotechnology,
Thapar University &
Dept. of Pulmonary Medicine, PGIMER,
Chandigarh

Investigator: Dr. Siddharth Sharma, Ph.D

Co-Investigator: Dr. Ashutosh Agarwal, M.D. D.M

Co-Investigator: Dr. Navneet Singh, M.D. D.M

Name of Patient_____

Address _____

(Rural/Urban)

C.R. Number_____

Date of _____

Sampling _____

Type of _____

cancer_____

Stage

Sex

Age (in years)_____

Education_____

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 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 a 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_____

Patient Consent Form

Name of participant:_____

Name of the Principal (Co-)_____

Investigators: Name of 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 **Association of single nucleotide polymorphism in DNA repair genes and survival of Advanced Non-Small-Cell Lung cancer patients treated with platinum-Based doublet chemotherapy**. I have read and understood consent form and the information provided to me.

1. I have had the consent document explained to me.
2. I have been explained about the nature of the study.
3. My rights and responsibilities have been explained to me by the investigator.
4. I have the right to not be a subject in the above study and am not willing to give sample.
5. I have informed the investigator of all the treatments I am taking or have taken in the past _____. months including any desi (alternative) treatments.
6. Patient data and confidentiality has been explained to me.
7. 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.
8. My identity will be kept confidential if my data are publicly presented.
9. I have had my questions answered to my satisfaction.
10. 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 form, 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 _____

A.3

1. **0.5M EDTA:** Dissolved 9.306 g of disodium salt of EDTA in 20 ml of deionised water, and then adjusted the pH to 8.0 by 1 M sodium hydroxide. Sterilized the solution by autoclaving.
2. **10% SDS:** Dissolved 1 g of SDS in 10 ml of deionised water.
3. **100 mM Tris-Cl (pH 8.0):** Dissolved 0.32 g 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.
4. **10 mg/ml Proteinase K:** Dissolved 10 mg Proteinase K in 1 ml of double distilled water. Sterilized the solution by autoclaving.
5. **1 mg/ml BSA:** Dissolved 100 mg of BSA in 100 ml of deionised sterile water and kept at 4 °C overnight.
6. **5 M Sodium chloride (NaCl):** Dissolved 5.85 g of sodium chloride in 20ml of deionised water. Sterilized the solution by autoclaving.
7. **5X TBE buffer:** Dissolved 54 g of Tris base and 27.5 g of boric acid in 980 ml of double distilled water and then added 20 ml of 0.5 EDTA. Sterilized the solution by autoclaving.
8. **Ethidium Bromide (10 mg/ml):** Dissolved 1 g of ethidium bromide in 100 ml of water. Mixed the solution properly.
9. **Magnesium chloride (MgCl₂) (100 mM):** Dissolved 0.41 gms of MgCl₂ in 20 ml of deionised water and sterilized by autoclaving.
10. **Sucrose (1 M):** Dissolved 3.41 g of sucrose in 10 ml of deionised water and sterilized by autoclaving.
11. **TE buffer (pH 8.0):** Added 1 ml of 100 mM Tris-Cl (pH 8.0) and 200 µl of 0.5 M EDTA solution to 8.8 ml of deionised water. Sterilized the solution.
12. **Triton X- 100 (10%):** Took 100 µl of TritonX-100 and mixed with 900 µl of deionised water and mixed properly.

XPC Polymorphism and Risk for Lung Cancer in North Indian Patients Treated with Platinum Based Chemotherapy and Its Association with Clinical Outcomes

Shweta Lawania¹ · Navneet Singh² · Digamber Behera² · Siddharth Sharma¹

Received: 14 May 2016 / Accepted: 8 May 2017
© Arányi Lajos Foundation 2017

Abstract Xeroderma pigmentosum complementation group C plays an important role in the human repair system. As reported in previous studies its polymorphism are associated with lung cancer susceptibility. The purpose of this study is to investigate the association of *XPC* gene with lung cancer susceptibility, overall response and clinical outcomes amongst North Indians. A hospital based study of 370 lung cancer cases and 370 healthy controls was conducted and genotypes were determined using PCR-RFLP assay. Results were assessed using logistic linear regression adjusted for age, sex and smoking status. Survival analysis was conducted using Kaplan-Meier survival analysis and Cox regression analysis. The treatment outcomes of 167 lung cancer patients treated with platinum based chemotherapy were evaluated. The mutant genotypic variant of *XPC* Lys⁹³⁹Gln has been associated with elevated risk of lung cancer (OR: 2.30; 95% CI: 1.41–3.73; $p=0.0007$) whereas *XPC* Ala⁴⁹⁹Val showed a highly protective effect (OR: 0.25; 95% CI: 0.10–0.63; $p=0.003$). The mutant genotype of *XPC* Lys⁹³⁹Gln presented a higher risk of developing lung cancer in heavy smokers (OR: 3.71; 95% CI: 1.46–9.45; $p=0.005$). The survival analysis presented that heterozygous genotype showed least survival in comparison with mutant genotype in *XPC* Ala⁴⁹⁹Val genetic variant whereas no significant association was observed in *XPC* Lys⁹³⁹Gln. In conclusion, *XPC* Lys⁹³⁹Gln is associated with

significant risk towards the lung cancer whereas on contrary *XPC* Ala⁴⁹⁹Val shows a protective effect.

Keywords Lung cancer · Xeroderma pigmentosum group C; PCR-RFLP; overall survival · Platinum based chemotherapy

Abbreviations

ADCC	(Adenocarcinoma)
HR	(Hazard ratio)
MST	(Mean survival time)
ns	(Non synonymous single nucleotide polymorphism)
OR	(Odds ratio)
SCLC	(Small cell lung cancer)
SQCC	(Squamous cell carcinoma)
XPC	(Xeroderma pigmentosum group C)

Introduction

Lung cancer remains the most common cause for the cancer related deaths worldwide [1]. In India the incidence of lung cancer is increasing and in 2012, 1.8 million people were diagnosed with lung cancer [2]. The increase in the consumption of cigarette smoking among the population is found to be the major cause of the disease [3]. Although evidence does suggest that only a small fraction of smokers (usually <20%) develop the disease, however it has also been postulated that individual susceptibility towards lung cancer can also be attributed to genetic variation and possible gene-environment interaction [4]. This susceptibility may result from inherited polymorphism in genes which control the carcinogen metabolism as well as the DNA repair mechanism [5].

✉ Siddharth Sharma
siddharthsharma.phd@thapar.edu

¹ Department of Biotechnology, Thapar University, Punjab 147002, India

² Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research (PGIMER), Sector 14, Chandigarh, India

DNA repair mechanism plays a crucial role in protecting genome from the damage caused by tobacco smoke and radiation [6]. Tobacco smoke consists of many carcinogens that leads to DNA bulky adduct, crosslinks, oxidative damage and DNA strand breaks. The major repair pathway which eliminates these lesions caused by smoking is the nucleotide excision repair (NER) pathway [7]. NER is controlled by the sequential assembly of repair proteins at the site of lesion. It is a complex pathway which requires thirty different proteins to carry out a multistep cut and patch mechanism [8]. DNA repair Capacity (DRC) of the NER pathway is responsible for maintaining the normal cellular functions and varies in general population [9]. An important key enzyme involved in the NER pathway is Xeroderma pigmentosum complementation group C gene (*XPC*). *XPC* gene encodes a 940 amino-acid protein involved in DNA damage recognition [10]. Two common polymorphic sites of the *XPC* gene have been investigated in different forms of cancer: A-to-C transition in exon 15 resulting in a lysine-to-glutamine transition at position 939 (Lys⁹³⁹Gln; rs2228001), located in the interaction domain with transcription factor IIH (TFIIH) and the second one involves a C-to-T transition in exon 8 which results in the substitution of alanine for valine at codon 499 (Ala⁴⁹⁹Val) in the interaction domain with hHRAD23 protein. Both these polymorphic sites are suggested to be in linkage disequilibrium with each other [11].

Previous studies have suggested an association of *XPC* polymorphisms with the occurrences of various forms of cancer such as gastric, head and neck, breast, renal, esophageal, colorectal, adenoma, oral, pancreatic and lung cancer [12, 13] but the results were inconclusive. These findings suggested that the effects of both polymorphisms remain ambiguous, and further investigations are needed to resolve these discrepancies.

Therefore, in the present study we have investigated the association of the two polymorphic sites in the *XPC* gene towards susceptibility for lung cancer (LC) and histological sub-types, furthermore we have also evaluated the role of both these SNPs towards the clinical outcome (overall survival) and response to platinum based chemotherapy treatment.

Material and Methods

Study Subjects

The current study recruited 370 lung cancer patients and 370 healthy controls. This study was approved by the Institutional Ethics Committee board of the Post Graduate Institute of medical education and research (PGIMER), Chandigarh. From each subject 5 ml of blood sample was collected for polymorphism studies. All subjects were genetically unrelated ethnic North Indians. Only lung cancer patients who were histopathologically confirmed were selected. The controls selected were hospital based and were not diagnosed with any cancer

but visited the hospital for routine health check-up. All the patients and controls were obtained from the department of pulmonary medicine, Post Graduate Institute of medical education and research (PGIMER), Chandigarh. All the histopathologically recruited patients were diagnosed as having non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). There were no age, gender, smoking, histological or TNM stage restrictions, but patients with a prior history of cancer were excluded from this study. All the subjects were interviewed after obtaining informed consent from them. A detailed questionnaire regarding demographic and life time history of smoking were obtained from all of the subjects. Information regarding tobacco habits included smoking of cigarette or beedi (a native cigarette-like stick of coarse tobacco hand-rolled in a dry tembuhumi leaf). As an indication of cumulative smoking exposure, pack-years were calculated by the following formula: [(cigarettes or beedis per day/20) X years smoked]. While medical information of cases, including histology, TNM classification, clinical staging, primary tumor size, involvement of lymph node and metastasis were obtained from medical records of the hospital. The follow-up data of the patients were obtained telephonically using the information of the patient as mentioned in the clinical records.

Genotyping of *XPC* A⁹³⁹C and C⁴⁹⁹T Polymorphisms

Genomic DNA was extracted from the leukocyte pellet obtained from blood sample according to the protocol of Sodhi et al. [14]. The *XPC* exon 9 C⁴⁹⁹T and *XPC* exon 15 A⁹³⁹C polymorphism was detected using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) technique. For the A⁹³⁹C polymorphism PCR amplification was carried out using the following set of primers *XPC*⁹³⁹ (F)-5-ACC AGC TCT CAA GCA GAA GC-3 and *XPC*⁹³⁹ (R)-CTG CCT CAG TTT GCC TTC TC-3(33). In the case of *XPC* exon 9 C⁴⁹⁹T polymorphism the primers used were: *XPC*⁴⁹⁹ (F) 5-TAA GGA CCC AAG CTT GCC CG-3; and *XPC*⁴⁹⁹ (R) 5-CCC ACT TTT CCT CCT GCT CAC AG-3(33). The amplification for both sites was carried out in 25 µl volume with approximately 50 ng of DNA, 0.5 µM of each primer, 200 µM of dNTP, 1X PCR buffer, 1.5 mM of MgCl₂, 100 µg BSA and IU of Taq Polymerase (DNAzyme II DNA Polymerase, Thermo Scientific). The PCR profile consisted of melting step of 94 °C for 5 min and followed by 30 cycles of 94 °C for 45 s, 56 °C for 30s and 72 °C for 45 s for the *XPC* A⁹³⁹C A⁹³⁹C polymorphism. In case of *XPC* C⁴⁹⁹T, the annealing temperature of 63 °C was used whereas all others conditions were same as that of *XPC* A⁹³⁹C. After confirmation of successful PCR amplification by 1.5% agarose gel electrophoresis, PCR products of *XPC* A⁹³⁹C (281 bp) and *XPC* A⁴⁹⁹T (152 bp) were digested overnight by 5 U of *PvuII* and *SacI* (New England Biolabs, MA, USA) at 37 °C

overnight. In case of *XPC* A⁹³⁹C the resulting fragments were separated on 2.5% agarose gel electrophoresis. The variant C allele had a *PvuII* restriction site that resulted in bands of 150 and 131 bp and the wild-type A allele lacked the restriction site and therefore produced a single 281 bp band. For the *XPC* A⁴⁹⁹T gene the digestion products were separated on a 6% polyacrylamide gel and stained with silver staining. The wild-type (152 bp) C allele produces fragments of 131 and 21 bp and the mutant (T) allele produces a single 152 bp fragment. For both the *XPC* polymorphisms, 15% of the samples were also repeated using the same assay. These results were all 100% concordant.

Statistical Analysis

Differences in distribution of certain demographic variables such as age, sex, smoking status, pack-years smoked were calculated using the Chi-square tests (χ^2 test) for the categorical data and Student's *t*-test for continuous variables. The frequencies of the *XPC* genotypes and alleles between the cases and controls were evaluated by using the Pearson's χ^2 test. Multivariate logistic regression was used to evaluate the associations between *XPC* variants and risk for lung cancer by computing the odds ratio (OR) and their 95% CI after adjustment for age, sex and smoking status. To compare the observed and expected genotypic frequencies among the control and case subjects Hardy-Weinberg equilibrium was tested by a goodness-of-fit χ^2 test. In addition to overall association analysis, stratified analysis was performed to estimate risk for subgroups formed on the basis of smoking status, tumor histology and clinic-pathological features. Haplotype analysis for the both SNPs was performed and the *D'* value and *r*² was calculated with the SHEsis software (25) (<http://202.120.31.177/myAnalysis.php>). The associations between overall survival and genotype were estimated using the Kaplan-Meier method and Logrank test. Survival time was calculated from the date of cancer diagnosis to the date of death. Univariate and multivariable Cox proportional hazards regression models were performed to estimate crude hazard ratio (HR) or adjusted HR and their 95% confidence intervals (CIs). The Cox regression model was also used to determine factors predictive of cancer prognosis, with a significant level of *P* < 0.05 for entering and *P* > 0.10 for removal of the variables. Tumor response was evaluated every 4 cycles of chemotherapy according to Response Evaluation Criteria in Solid Tumor Group (RECIST). Patients show complete response (CR) or partial response (PR) were defined as “responders”, and patients with stable disease (SD) or progressive disease were defined as “non-responders”. All *p* values were two sided, and a *p*-value of <0.05 was considered statistically significant. All the statistical analyses were performed with MedCalc Statistical Software version 15.11.4 (MedCalc Software bvba, Ostend, Belgium; <https://www.medcalc.org>; 2015).

Results

Demographic Characteristics of Lung Cancer Cases and Healthy Controls

The current study was a hospital based case-control study which evaluated the potential role of *XPC* polymorphism and its risk associated with lung cancer. A total of 740 subjects were recruited which included 370 cases and 370 healthy controls. The subjects various demographic characteristics such as age, gender, smoking status, pack years, histological subtypes, TNM staging and other clinical parameters are summarized in Table 1. Mean age of lung cancer cases was

Table 1 Demographic characteristics among cases and control patients in North Indian population

Variable	Cases, n (%) N = 370	Controls, n (%) N = 370	<i>p</i> ^a – value
Age (years)	58.11 ± 10.44	53.83 ± 10.18	<0.0001
Mean ± SD			
Range			
Gender			0.91
Male	319(86.21%)	319(86.21%)	
Female	51(13.78%)	51(13.78%)	
Smoking Status			0.0205
Smokers	303(81.89%)	276(74.59%)	
Non – Smokers	67(18.10%)	94(25.40%)	
Pack Years			
Mean ± SD	24.97 ± 35.04	17.88 ± 19.45	<0.0001
Histological Types			
SQCC	135(36.48%)		
ADCC	120(32.43%)		
SCLC	111(30%)		
Others	4(1.08%)		
Unknown	0(0.0)		
TNM Staging			
I	3(0.81%)		
II	10(2.70%)		
III	175(45.94%)		
IV	163(44.05%)		
Unclassified	19(5.13%)		
Tumor Size			
Tx	11(2.97%)		
T1	16(4.32%)		
T2	42(11.35%)		
T3	85(22.97%)		
T4	191(51.62%)		
Unknown	25(6.75%)		
Lymph Node Involvement			
N0	49(16.3)		
N1	36(12)		
N2	111(37)		
N3	72(24)		
UNKNOWN	32(10.7)		
Metastasis			
M0	182(49.18%)		
M1	163(44.05%)		
Unknown	25(6.75%)		

All *p* values are two – sided. *p* < 0.05 was considered statistically significant

SD Standard Deviation, *n* total number of case patients or control subjects

^a *p*-values were derived from Pearson Chi-square test except age; Student *t*-test was used for age

58.11 ± 10.44 (range 29–86), whereas mean age of controls was 53.83 ± 10.18 (range 22–83). There were 319 (86.21%) males and 51 (13.78%) females in both the control and case groups. Hence, no significant difference had been observed between the distribution of males and females, suggesting adequate matching ($p = 0.91$). In the present study, 81.89% cases were smokers and 18.10% were non-smokers when compared with controls having 74.59% of smokers and 25.4% non-smokers. Smoking was further elaborated in terms of pack years. Pack years were significantly higher amongst cases in comparison with controls (24.97 ± 35.04 vs 17.88 ± 19.45, $p < 0.0001$). From the total of 370 cases, 135 (36.48%) patients were classified with Squamous cell carcinoma (SQCC), 120 (32.43%) with Adenocarcinoma (ADCC) and 111 (30%) with SCLC. Out of 370 cases, TNM staging data were available for 351 (94.86%) patients (Stage I: 3 (0.81%), II: 10 (2.7%), III: 175 (45.94%), IV: 163 (44.05%). Lung cancer subjects with T1 and T2 tumor size constituted 4.32% and 11.35% while T3 and T4 were 22.97% and 51.62% of the total subjects. The frequencies for lymph node involvement classified as: N0 with 16.3% of the patients whereas N1, N2, N3 with 12%, 37%, and 24% respectively. The available data includes 49.18% cases which do not have any metastasis (M0) while 44.05% cases predict involvement of distant metastasis (M1).

Genotypic and Allelic Distribution of XPC Lys⁹³⁹Gln and XPC Ala⁴⁹⁹Val

The allelic and genotypic frequencies of two XPC polymorphic variants, Lys⁹³⁹Gln and Ala⁴⁹⁹Val are shown in Table 2. Genotypic frequencies of XPC Lys⁹³⁹Gln among controls ($n = 370$; $\chi^2 = 0.002$, $df = 1$, $p = 0.95$) and cases ($n = 370$, $\chi^2 = 0.222$, $df = 1$, $p = 0.63$) were in Hardy-Weinberg equilibrium, suggesting no sample biasness for control and cases. However, the genotypic frequencies for the other variant, XPC Ala⁴⁹⁹Val for both control ($n = 370$, $\chi^2 = 23.2$, $df = 1$, $p = 0.00$) and cases ($n = 370$, $\chi^2 = 32.71$, $df = 1$, $p = 0.00$) were not in Hardy-Weinberg Equilibrium, suggesting population stratification, migration and random mating within the sample set. The frequency of the wild genotype of XPC A⁹³⁹C gene (AA) in lung cancer subjects was found to be less represented as compared to the controls (35.94 vs 44.86%). However as shown in Table 2 the frequency of the mutant genotype for XPC A⁹³⁹C was over represented in cases as compared to controls, thus overall there was a significant difference among the genotypic distribution between cases and controls ($\chi^2 = 9.07$, $df = 2$, $p = 0.01$). Minor allelic frequency (MAF) calculated for XPC A⁹³⁹C in controls was 0.33 whereas in cases it was 0.40, suggesting a significant difference among the sample population. In case of XPC variant, C⁴⁹⁹T the frequency of the mutant genotype (TT) was found to be more

represented in the control group as compared to the cases (6.21 vs 1.89%), overall there was a significant difference in the distribution of genotypes between the controls and cases ($\chi^2 = 12.22$, $df = 2$, $p = 0.002$). Furthermore, the MAF calculated for control and cases was 0.34, 0.28 respectively, suggesting that both the variant allele was found to be more pronounced in the control than cases.

Table 2 also showed the lung cancer risk related to the XPC polymorphisms. Adjusted ORs and 95% CIs were calculated using the logistic regression analysis using the more common genotype as the reference group. Homozygous wild genotype AA (A⁹³⁹C) and CC (C⁴⁹⁹T) were considered as reference genotypes with which the variant genotypes were compared. In case of the XPC Lys⁹³⁹Gly polymorphism, it was observed that individuals carrying both the mutant alleles (CC) had a 2-fold elevated risk for lung cancer which was found to be significant (OR = 2.30; 95% CI: 1.41–3.73; $p = 0.0007$). Whereas in subjects with heterozygote genotypes (AC), the risk was less pronounced as compared to the mutant genotype but was reported to be significant (OR = 1.44; 95% CI: 1.04–1.99; $p = 0.02$). When both the genotypes were combined as a single genotype (AC + CC) and compared with the reference genotype a significant association towards lung cancer risk was also observed (OR = 1.60; 95% CI: 1.18–2.17; $p = 0.002$).

Lung cancer cases were further segregated into histological subgroups in order to evaluate the risk of susceptibility amongst these subgroups. As indicated in Table 2 the subjects carrying the mutant genotype (CC) for XPC Lys⁹³⁹Gly polymorphism were at higher risk for all sub-groups of lung cancer as compared to subjects with heterozygous genotypes, however amongst the different histological sub-types; it was observed that subjects with mutant genotype had a 4 and 3-fold increased risk for ADCC (OR = 4.72; 95% CI: 2.16–10.32; $p < 0.0001$) and SQCC (OR = 3.06, 95% CI: 1.58–5.92; $p = 0.009$) which was found to be highly significant.

For XPC Ala⁴⁹⁹Val genotype it was observed that individuals with mutant (TT) genotype had the least risk for lung cancer (OR = 0.25; 95% CI: 0.10–0.63; $p = 0.003$) indicating a protective effect towards the development of the disease. However when stratified according to histological sub-types; ADCC subjects with mutant genotype showed the lowest risk for lung cancer (OR = 0.12, 95% CI: 0.15–0.96; $p = 0.04$). However no such significant association was observed in any other sub-types of lung cancer.

For the current study, we analyzed haplotypes using SHEsis program platform (Table 3). The two SNPs were in linkage disequilibrium in this study population ($D = 0.039$, $r^2 = 0.00$). The haplotypes were composed of two coding SNPs. There was a statistically significant difference in the overall haplotype distribution between cases and controls (Global χ^2 test = 18.811, $p = 0.0003$). We considered the

Table 2 Distribution of genotypes of XPC polymorphism depicting their association with Lung cancer risk and its histological subtypes

Overall lung cancer			SQCC			ADCC			SCLC		
XPC	Controls n (%)	Cases N (%)	AOR (95% CI) ^b	<i>p</i> ^a	Cases n(%)	AOR (95% CI) ^b	<i>p</i> ^a	Cases n(%)	AOR (95% CI) ^b	<i>p</i> ^a	Cases n(%)
lys ⁹³⁹ gly	N = 370	N = 370			N = 135			N = 120			N = 111
AA	166(44.86)	133(35.94)	1.00 (Reference)		42(31.11)	1.00 (Reference)		39(32.5)	1.00 (Reference)		51(45.9)
AC	164(44.32)	174(47.02)	1.44 (1.04–1.99)	0.02	68(50.37)	1.81(1.14–2.88)	0.011	59(49.1)	2.15(2.28–3.63)	0.0039	46(41.4)
CC	40(10.8)	63(17.02)	2.30 (1.41–3.73)	0.0007	25(18.51)	3.06(1.58–5.92)	0.0009	22(18.3)	4.72 (2.16–10.3)	<0.0001	14(12.6)
AC + CC	204(55.13)	237(64.05)	1.60 (1.18–2.17)	0.002	93(68.88)	2.03(1.31–3.15)	0.0014	81(67.5)	0.86(0.47–1.56)	0.063	60 (54.0)
A Allele	496(67.03)	440(59.46)									
C Allele	244(32.97)	300(40.54)									
MAF	0.33	0.40									
XPC Ala ⁴⁹⁹ Val											
CC	138(37.29)	169(45.67)	1.00 (Reference)		62(45.92)	1.00 (Reference)		58(48.3)	1.00 (Reference)		47(42.34)
CT	209(56)	194(52.43)	0.76(0.56–1.04)	0.088	70(51.85)	0.69 (0.45–1.05)	0.82	61(50.8)	0.749(0.48–1.1)	0.192	61(54.9)
TT	23(6.21)	7 (1.89)	0.25(0.105–0.63)	0.003	3 (2.2)	0.28 (0.07–0.98)	0.39	1 (0.83)	0.12(0.15–0.96)	0.04	3 (2.7)
TC + TT	232(62.02)	201(54.32)	0.72(0.53–0.97)	0.034	73 (54.07)	0.64 (0.42–0.98)	0.84	62 (51.6)	0.70 (0.45–1.08)	0.113	64 (57)
C Allele	485 (65.54)	532 (71.89)									
T Allele	255 (34.45)	208(28.10)									
MAF	0.34	0.28									

^a Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls^b 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
ADCC Adenocarcinoma, SQCC Squamous cell carcinoma, SCLC Small cell lung carcinoma

Table 3 XPC Haplotypes and their association with lung cancer risk

XPC Lys ⁹³⁹ Gln & XPC Ala ⁴⁹⁹ Val	Case frequency, n (%)	Control frequency, n (%)	OR (95% CI) ^a	<i>p</i> -value ^b
A-C	0.418 (309)	0.449 (331)	0.883 (0.719–1.08)	0.236
A-T	0.176 (130)	0.226 (167)	0.735 (0.56–0.94)	0.018
C-C	0.301 (222)	0.207 (153)	1.65 (1.301–2.09)	3.31e-005
C-T	0.105 (77)	0.119 (87)	0.866 (0.626–1.19)	0.383

^a Two-sided χ^2 test for either haplotypic distribution or frequencies between the cases and controls^b Global χ^2 test = 18.811, *p*-value=0.0003

individuals with 939A-499C haplotype as the reference group for OR estimations. The C-C haplotype was associated with increased risk of lung cancer (OR = 1.65; 95%CI: 1.30–2.09; *p* = 3.31e-005). Among individuals having the A-T haplotype there was a reduced risk of lung cancer (OR = 0.73; 95%CI: .56–.94; *p* = 0.018). All other haplotypes (CT) did not show any significant effect.

Interactive Effect of Smoking on the Genetic Variants and Its Association with Lung Cancer Susceptibility

Table 4 shows the correlation between *XPC* polymorphic variants A⁹³⁹C and C⁴⁹⁹T and smoking and its association with lung cancer risk. The sample population is divided into two subgroups: smokers and non-smokers. The smokers were further divided into heavy and light smokers. The patients who

had smoked less than or equal to 25 pack years are grouped under light smokers whereas more than 25 pack years are grouped as heavy smokers. Smokers carrying the mutant (CC) genotype of the *XPCA*⁹³⁹C gene were found to have 2-fold elevated risk for lung cancer as compared (OR = 2.55; 95% CI: 1.46–4.45; *p* = **0.001**) to smokers having the wild type genotype (AA), a similar trend was also seen in subjects who were hetero-zygotes (OR = 1.53; 95%CI: 1.06–2.20; *p* = 0.02), however the effect was less pronounced. Furthermore, it was observed that heavy smokers with the mutant (CC) genotype for Lys⁹³⁹Gly were at significantly higher risk (OR = 3.71, 95%CI = 1.46–9.45; *p* = **0.005**) than light smokers (OR = 2.18, 95%CI = 1.10–4.28; *p* = 0.02). In case of the C⁴⁹⁹T genotype, no such association was observed between smoking and associated risk towards lung cancer with the any genotypic combinations of the C⁴⁹⁹T gene.

Table 4 Distribution of genotypes of XPC polymorphism according to their smoking status and its association with lung cancer risk

Smokers					Non smokers				
XPC Lys ⁹³⁹ Gln	Controls n (%)	Cases n (%)	AOR (95% CI) ^b	<i>p</i> ^a	Controls n (%)	Cases n (%)	AOR (95% CI) ^b	<i>p</i> ^a	
	<i>N</i> = 276	<i>N</i> = 303			<i>N</i> = 94	<i>N</i> = 67			
AA	131 (47.4)	111 (36.6)	1.00 (Reference)		35	22 (32.8)	1.00 (Reference)		
AC	119 (43.1)	142 (46.8)	1.53 (1.06–2.2)	0.021	45	32 (47.7)	1.25 (0.59–2.61)	0.55	
CC	26 (9.42)	50 (16.5)	2.55 (1.46–4.45)	0.001	14	13 (19.4)	2.05 (0.76–5.56)	0.15	
AC + CC	145 (52.5)	50 (16.5)	1.71 (1.21–2.42)	0.002	59	45 (67.16)	1.40 (0.70–2.79)	0.32	
XPC Ala ⁴⁹⁹ Val									
AA	156 (56.52)	167 (55.11)	1.00 (Reference)		38 (40.4)	36 (53.7)			
AC	20 (7.26)	174 (57.42)	0.809 (0.571–1.14)	0.23	53 (56.3)	31 (46.2)	0.614 (0.316–1.19)	0.150	
CC	100 (36.23)	7 (2.3)	0.334 (0.133–0.835)	0.02	3 (3.19)		0.0 (0.0–0.0)	0.993	
AC + CC	120 (43.47)	181 (59.73)	0.758 (0.537–1.14)	0.116	56 (59.5)	31 (46.2)	0.587 (0.302–1.13)	0.115	
HEAVY SMOKER					LIGHT SMOKER				
XPC Lys ⁹³⁹ Gln	Controls n (%)	Cases n (%)	AOR (95% CI) ^b	<i>p</i> ^a	Controls n (%)	Cases n (%)	AOR (95% CI) ^b	<i>p</i> ^a	
	<i>N</i> = 104	<i>N</i> = 141			<i>N</i> = 174	<i>N</i> = 162			
AA	47 (45.14)	45 (31.91)	1.00 (Reference)		85 (48.85)	66 (40.74)	1.00 (Reference)		
AC	49 (47.11)	69 (48.93)	1.54 (0.87–2.72)	0.13	71 (40.80)	73 (45.06)	1.24 (0.77–1.97)	0.36	
CC	8 (7.69)	27 (19.14)	3.71 (1.46–9.45)	0.005	18 (10.34)	23 (14.19)	2.18 (1.10–4.28)	0.02	
AC + CC	57 (54.80)	96 (68.08)	1.84 (1.07–3.16)	0.02	89 (51.14)	96 (59.25)	1.43 (0.943–2.17)	0.09	
XPC Ala ⁴⁹⁹ Val									
CC	39 (37.5)	65 (46.09)	1.00 (Reference)		62 (35.63)	68 (41.97)	1.00 (Reference)		
CT	55 (52.8)	73 (51.77)	0.952 (0.49–1.82)	0.88	102 (58.62)	90 (55.55)	0.86 (0.54–1.36)	0.52	
TT	10 (9.61)	3 (2.12)	0.299 (0.05–1.50)	1.43	10 (5.74)	4 (2.46)	0.51 (0.14–1.77)	0.29	
CT + TT	65 (62.5)	76 (53.90)	0.882 (0.96–1.66)	0.69	112 (64.36)	94 (58.02)	0.82 (0.52–1.30)	0.41	

^a Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls^b Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status.

Combined Effect of both Polymorphic Variant XPC Lys⁹³⁹Gln and XPC Ala⁴⁹⁹Val on Lung Cancer Risk

Different genotypic combinations of two *XPC* variants were analyzed to find out their association with lung cancer risk in order to evaluate gene-gene interaction between the two *XPC* polymorphic variants (Table 5). A 2-fold risk of lung cancer was observed in subjects carrying either variant alleles (AC + CT) for both the polymorphic variants, however this effect was not significant (OR = 1.90; 95%CI: 0.91–3.94; $p = 0.446$). No association was observed in other combined *XPC* genotypes in the current study.

Relationship of Clinic-Pathological Features and Variant Genotypes in Lung Cancer Patients

To determine association between the two polymorphic variants of *XPC*, (A⁹³⁹C, C⁴⁹⁹T) and clinic-pathological parameters we divided patients on the basis of stage of cancer (III vs IV), primary tumor extension (T3 + T4 vs T1 + T2) and metastatic status (positive vs negative) (Table 6). No significant association was observed between staging/tumor metastasis and *XPC* variants. However, higher frequency of *XPC* A⁹³⁹C heterozygous genotype (AC) was observed in patients with T3/T4 tumor as compared to patients with T1/T2 tumor, therefore a 3.5 fold risk for lung cancer was observed for subject with T3/T4 tumor extension and heterozygous genotype of the *XPC* A⁹³⁹C as compared to patients with T1/T2 tumor size (OR = 3.54, 95%CI: 1.71–7.51, $p = 0.0006$). No association was observed for *XPC* C⁴⁹⁹T genetic variant and clinic pathological features.

Associations of XPC SNPs with Overall Survival of Patients with Lung Cancer and Specific Histological sub-Types

Univariate analyses using Kaplan-Meier curves and log-rank tests was performed so to evaluate the association of the *XPC* polymorphism with overall survival of patients as shown in Table 7. The distribution of observed genotypes in the patients agreed with those estimated from the Hardy-Weinberg

equilibrium ($\chi^2 = 0.07$; $P = 0.78$; for *XPC* Lys⁹³⁹Gln). Since the frequency of the mutant genotype in case of *XPC* C⁴⁹⁹T was less, we combined the heterozygous and mutant genotype as a single group and compared the OS with the wild genotype. As shown in Table 7 the variant genotypes for *XPC* Ala⁴⁹⁹Val were associated with shorter survival as compared to the wild type genotype (median survival time, 6.23 months versus 7.20 months, log-rank $p = 0.44$ Fig. 1a). In the multivariate Cox proportional hazards regression analysis, no significant association was observed between the *XPC* Ala⁴⁹⁹Val mutant genotype and overall lung cancer survival (HR = 0.89, 95% CI = 0.66–1.19). However when stratified according to histology ADCC patients with the variant genotype had a longer survival as compared to ADCC patients with wild type genotype (MST, 10.33 months versus 6.70 months log-rank $p = 0.03$, Fig. 1c). No such association was observed neither for SQCC (Fig. 1b) nor for SCLC (Fig. 1d) patients. Furthermore, in the multivariate Cox model as shown in Table 7, ADCC subjects with CT/TT genotype of *XPC* Ala⁴⁹⁹Val were significantly associated with better overall survival (adjusted HR = 0.45, 95% CI = 0.25–0.82, $p = 0.0096$), compared with the CC genotype. Thus this could be considered as an independent prognostic factor for ADCC patients receiving chemotherapy. Other clinical features were all seen to have no apparent effect in influencing the prognostic outcomes of lung cancer patients receiving chemotherapy except ECOG status (Table 8). There might be potential association of ECOG status with the prognosis of ADCC patients (HR = 1.43, 95%CI = 1.06–1.93; $p = 0.01$). In case of the *XPC* Lys⁹³⁹Gln both the univariate and multivariate cox regression models did not show any significant association between the variant genotypes and overall survival (Table 7, Fig. 2a-d).

Association between XPC Polymorphisms and Clinical Benefit

XPC gene polymorphisms and chemotherapy response in lung cancer patients treated with first-line platinum-based drug along with taxol based drug or combination chemotherapy for ≥ 6 cycles are shown in Table 9. The association between responders (platinum-sensitive) and non-responders

Table 5 Combination of genotypes of XPC polymorphic variants and their relationship with the risk of lung cancer

Genotypes	Cases, n (%) $N = 141$	Controls, n (%) $N = 156$	Adjusted OR (95% CI) ^b	p -value ^a
XPC Lys ⁹³⁹ Gln & XPC Ala ⁴⁹⁹ Val				
0 (AA + CC)	56	64	1.00 (Reference)	
1 (AC + CT)	85	91	1.90 (0.918–3.94)	0.446
2 (CC + TT)	-	1	-	-
1–2	85	92	1.29 (0.823–2.049)	0.260

^a Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

^b Adjusted Odds ratios, 95% confidence intervals and their corresponding p -values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status.

0 wild genotype, 1 heterozygote genotype, 2 mutant genotype, 3 combined hetero and mutant genotype

Table 6 Relationship of clinicopathological features with XPC Lys⁹³⁹Gln and XPC Ala⁴⁹⁹Val genotypes in lung cancer patients

Variable	Clinical stage		AOR ^b (95% CI)	<i>p</i> ^a value	TNM stage		AOR ^b (95% CI)	<i>p</i> ^a value	Metastasis		AOR ^b (95% CI)	<i>p</i> ^a value
	III	IV			<i>T3</i> + <i>T4</i>	<i>T2</i> + <i>T1</i> + <i>T0</i>			<i>Positive</i>	<i>Negative</i>		
<i>XPCLys</i> ⁹³⁹ <i>Gln</i>	175	163			276	50			188	189		
AA	59	64	Reference		96	30	Reference		74	67	Reference	
AC	82	75	0.80 (0.48–1.15)	0.39	131	12	3.54(1.71–7.35)	0.0006	81	87	0.79 (0.50–1.27)	0.34
CC	34	25	0.60 (0.30–1.17)	0.13	49	9	1.71(0.74–3.93)	0.20	33	35	0.59 (0.10–3.29)	0.48
AC + CC	141	139	0.75 (0.47–1.19)	0.23	180	21	1.93(1.08–3.43)	0.02	114	122	1.12 (0.71–1.75)	0.61
<i>XPCAla</i> ⁴⁹⁹ <i>Val</i>												
CC	82	73	Reference		125	36	Reference		69	86	Reference	
CT	91	86	0.91 (0.53–1.57)	0.74	147	29	0.91 (0.53–1.57)	0.74	84	99	1.13 (0.72–1.77)	0.58
TT	2	4	0.59 (0.10–3.29)	0.55	5	2	0.59 (0.10–3.29)	0.55	3	4	1.07 (0.22–5.01)	0.92
CT + TT	93	90	0.85 (0.49–1.44)	0.58	152	31	0.85 (0.49–1.44)	0.55	87	103	1.13 (0.72–1.77)	0.05

^a Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype

^b Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status

1 heterozygote genotype, 2 mutant genotype, 3 combined hetero and mutant genotype

Table 7 Overall survival analysis using in genotypes of XPC polymorphic variants and their association with lung cancer risk

Genotype	No.	Censored (Alive)	Uncensored (Dead)	Median overall survival MST(months)	Crude HR ^a (95%CI)	Adjusted HR ^b (95%CI)	Log rank P	<i>p</i> ³
<i>Lys</i> ⁹³⁹ Gln								
Lys/Lys	86	61	25	7.20	1.00(Reference)	1.00(Reference)		
Lys/Gln	123	90	33	5.93	1.00(0.73–1.39)	1.06(0.76–1.49)	0.79	0.70
Gln/Gln	41	32	9	6.23	1.14(0.74–1.78)	1.01(0.85–1.32)	0.79	0.93
<i>Ala</i> ⁴⁹⁹ Val								
Ala/Ala	108	82	26	6.40	1.00(Reference)	1.00(Reference)		
Ala/Val + Val/Val	142	101	41	6.73	0.89(0.66–1.19)	0.84(0.63–1.10)	0.44	0.26
ADCC								
<i>Lys</i> ⁹³⁹ Gln								
Lys/Lys	24	18	6	8.53	1.00(Reference)	1.00(Reference)		
Lys/Gln	45	32	13	7.56	1.13(0.64–2.01)	1.01(0.54–1.91)	0.90	0.95
Gln/Gln	13	9	4	9.30	1.06(0.48–2.34)	1.07(0.71–1.60)	0.90	0.72
<i>Ala</i> ⁴⁹⁹ Val								
Ala/Ala	39	32	7	6.70	1.00(Reference)	1.00(Reference)		
Ala/Val + Val/Val	43	27	16	10.33	0.58(1.02–2.88)	0.45 (0.25–0.82)	0.03	0.0096
SCLC								
<i>Lys</i> ⁹³⁹ Gln								
Lys/Lys	30	22	8	6.73	1.00(Reference)	1.00(Reference)		
Lys/Gln	31	22	9	8.26	0.99 (0.55–1.79)	0.99(0.52–1.87)	0.99	0.97
Gln/Gln	8	7	1	3.43	1.05(0.44–2.50)	1.24(0.43–3.56)	0.99	0.68
<i>Ala</i> ⁴⁹⁹ Val								
Ala/Ala	26	18	8	4.46	1.00(Reference)	1.00(Reference)		
Ala/Val + Val/Val	43	33	10	7.23	0.95(0.53–1.69)	0.85(0.45–1.57)	0.85	0.60
SQCC								
<i>Lys</i> ⁹³⁹ Gln								
Lys/Lys	31	21	10	7.20	1.00(Reference)	1.00(Reference)		
Lys/Gln	47	36	11	5.43	0.97(0.56–1.65)	1.09(0.61–1.97)	0.91	0.75
Gln/Gln	19	15	4	6.86	1.09 (0.55–2.10)	0.85(0.38–1.87)	0.91	0.69
<i>Ala</i> ⁴⁹⁹ Val								
Ala/Ala	42	31	11	6.86	1.00(Reference)	1.00(Reference)		
Ala/Val + Val/Val	55	41	14	5.26	1.18 (0.53–1.34)	0.99(0.62–1.60)	0.47	0.99

^a Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

^b Hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Kaplan Meier survival analysis after adjusting for remission and survival in months and ^b Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, kps and ecog

0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype

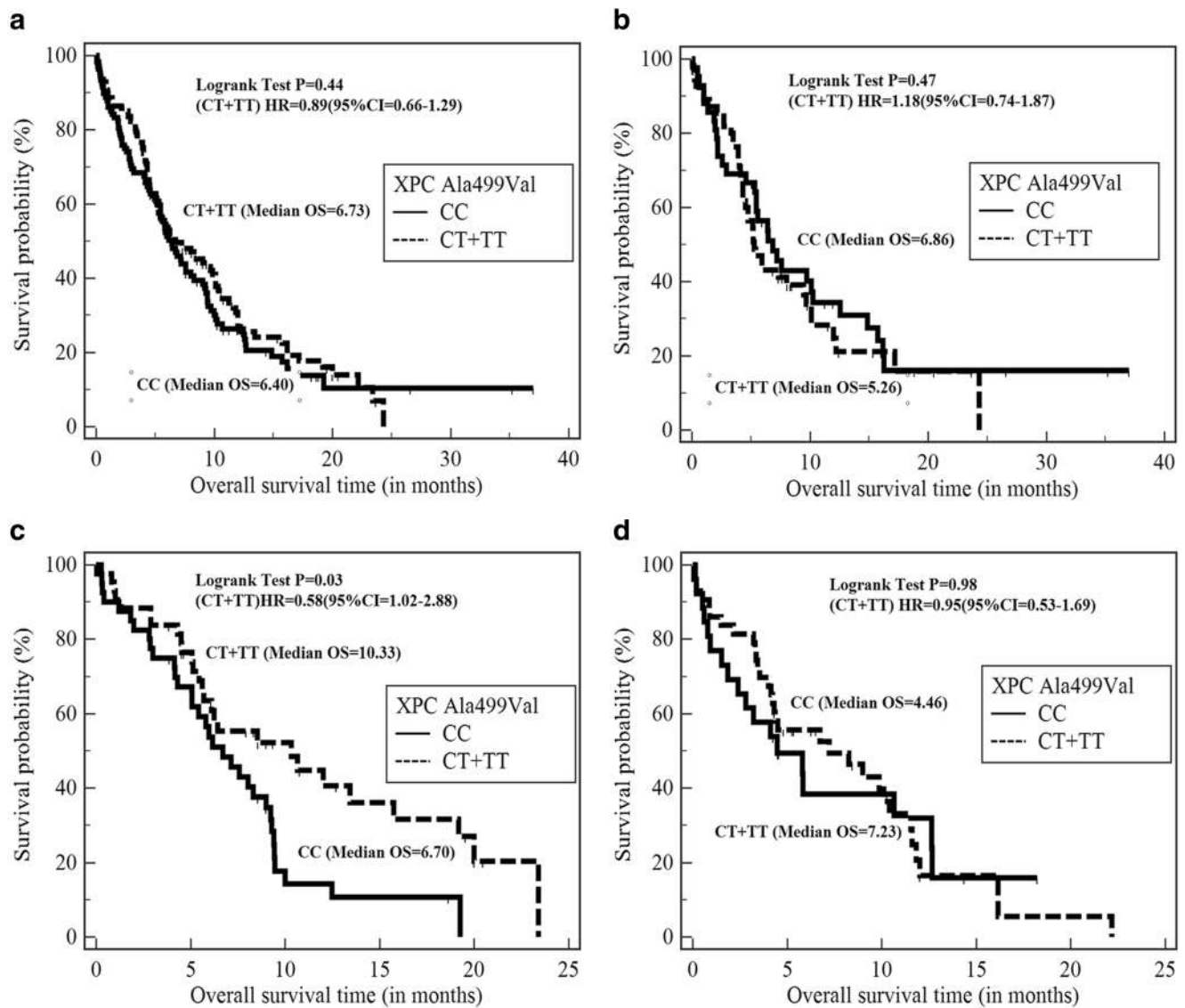


Fig. 1 The overall survival in lung cancer patients is represented in form of Kaplan Meier survival curve. The survival probability of different genotypes present in the sample population has been compared with the overall survival time in months for *XPC* Ala⁴⁹⁹Val Polymorphism. **(a)** showed overall survival in *XPC* Ala⁴⁹⁹Val, the dotted line represents the survival of the CT + TT (double genotype) with the median survival time of 6.73 months while the dark black line shows the overall survival of the wild genotype with median survival time of 6.49 months with log rank test value $P = 0.44$. **(b)** shows the Kaplan Meier survival curve for the

patients diagnosed with Squamous cell cancer for *XPC* Ala499Val polymorphism. The dotted line depicts the survival of the CT + TT (Mean survival time = 5.26 months) genotype while black line depicts the survival of wild genotype (6.86 months). **(c)** shows the Kaplan Meier survival curve for the patients diagnosed with Adenocarcinoma for *XPC* Ala499 Val polymorphism. The figure showed a better survival for CT + TT genotype depicting Mean survival time of 10.33 months. **(d)** showed the Kaplan Meier curve for patients diagnosed with Small cell lung cancer for *XPC* Ala499Val polymorphism.

(platinum-resistant) and the panel of *XPC* polymorphisms was assessed by a multivariate logistic regression model adjusted for age, stage, smoking, stage and histology. Each genotype was compared with the wild genotypes. The sample population having total 167 lung cancer patients were divided into two groups: good responders (CR + PR) and poor responders (SD + PD). There was no statistically significant difference between responders and non-responders and the XP genotypes. The lack of statistical significance can be accounted towards the low sample size.

Discussion

In NER pathway, *XPC* is one of the major gene which is related to recognition of damage, forms an open complex and takes part in DNA repair [15]. The NER pathway primarily involved in removal of large lesions and thus considered as the important defense mechanism for the structurally unrelated lesions [16]. Many past studies have emphasized on the relation of DNA adduct levels with the development of lung cancer [17]. Functional SNPs which are basically present in

Table 8 Cox regression survival analysis adjusted for Ala⁴⁹⁹Val variant in ADCC patients

<i>N</i> = 80 Censored = 23 Uncensored = 60 Covariates	b	SE	Wald	<i>P</i> ^a -value	Adjusted OR ^b (Exp b)	95% CI ^b of Exp (b)
Age	0.013	0.015	0.797	0.371	1.013	0.983to1.044
Sex	-0.240	0.445	0.291	0.589	0.786	0.328to1.881
Smoking	0.389	0.393	0.980	0.322	1.475	0.682to3.190
Stage	0.110	0.214	0.266	0.605	1.117	0.733to1.700
KPS	-0.442	0.305	2.103	0.146	0.642	0.353to1.168
ECOG	0.361	0.153	5.545	0.0185	1.435	1.062to1.938
XPC Ala ⁴⁹⁹ Val	-0.785	0.303	6.715	0.096	0.455	0.251to0.825

The values in bold represented significant *p*-values. Whereas non-significant values are not represented in bold

^a Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

^b Adjusted hazard ratios, 95% confidence intervals and their corresponding *p*-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, kps and ecog

0 wild genotype, 1 heterozygote genotype, 2 mutant genotype, 1 combined hetero and mutant genotype

protein coding region alter the amino acid sequence which leads to the alteration of protein function. The *XPC* Lys⁹³⁹Gln and Ala⁴⁹⁹Val polymorphism cause amino acid substitution in the functional domain which alter the DNA repair capacity and its susceptibility towards the lung cancer.

In this hospital-based case-control study, we investigated the association of two polymorphisms within the *XPC* gene i.e. C⁴⁹⁹T and A⁹³⁹C towards risk for the lung cancer and also evaluated the clinical outcomes in a North Indian population. When we analyzed each variant of *XPC* gene separately, an increased risk of lung cancer was found in patients with *XPC* A⁹³⁹C mutant genotype (*p* = **0.0007**), similarly when the results were analyzed for histological subgroups, the variant genotype was found to be significantly associated with all the histology's of lung cancer. On the contrary the variant genotype of *XPC* C⁴⁹⁹T was found to be less represented in lung cancer patients thus suggesting a strong protective effect (*p* = **0.003**). Furthermore, we analyzed the combined effect of both the polymorphic variants, no significant association was observed towards risk for lung cancer.

The investigations in the present study have shown two fold higher risk for the homozygous variant of *XPC* Lys⁹³⁹Gln gene in North Indian population which agreed with the results of Jin et al. [18] who also assessed a higher risk of developing lung cancer in Asian subjects carrying variant allele of *XPC* Lys⁹³⁹Gln polymorphism. The above study was further supported by Nielson et al. [16] who also emphasized on an increased risk of lung cancer in variant genotype of A⁹³⁹C in Caucasian population. It can be suggested that the variant form of *XPC* A⁹³⁹C might be closely related towards increased susceptibility towards the development of lung cancer. This may be because *XPC* protein binds with the HR23B to form a heterodimeric complex which is considered as the crucial step for the detection of wide variety of lesions [16] and hence recruit the whole repair machinery to the damaged site [19, 20]. It has also been suggested that polymorphisms in

XPC might lead to the alteration in the repair capacity of an individual which causes genetic instability and may ultimately results into carcinogenesis [21].

Further, the present study suggested a strong protective effect for the lung cancer in *XPC* C⁴⁹⁹T, particularly in the mutant (TT) genotype (OR:0.25). The results suggested that the genetic alteration in *XPC*, interaction with environmental factors, ethnic background, host characteristics or prolonged exposure to the carcinogen may be the possible reasons for this protective effect. Our results were consistent with the studies conducted by Shen et al. [22] suggesting a protective effect of *XPC* C⁴⁹⁹T towards the risk of lung cancer in Asian population. However, studies reported by Vodicka et al. [23] and Jin et al. [18] differ from the present study suggesting no functional significance of *XPC* Ala⁴⁹⁹Val towards risk of lung cancer in European and Korean population. However, the difference in findings may be due to ethnic differences or low sample size with limited statistical power.

However, when the sample population was stratified into different histological subtypes: ADCC, SQCC and SCLC, our results suggested a significant risk of occurrence of disease in patients with mutant genotype observed for all the three histological subtypes for *XPC* Lys⁹³⁹Gln which included ADCC (OR: 4.72), SQCC (OR: 3.06) and SCLC (OR: 2.30). Although the reasons behind such histology dependent difference which is conferred by these polymorphism is unknown, but as suggested it may be related with the difference in carcinogenesis pathway among the histological subtypes of lung cancer. Many previous studies suggested that the determination of histological subtype is related with the individual's exposure to the specific initiation agent [24, 25]. Therefore, various genetic factors may be responsible for the susceptibility in histological subtypes of lung cancer [26, 27].

We also evaluated the role of gene-gene of the SNPs of the *XPC* gene towards modulating the risk for lung cancer [28]. A study by Letkova et al. [29] reported that when the genotypes of

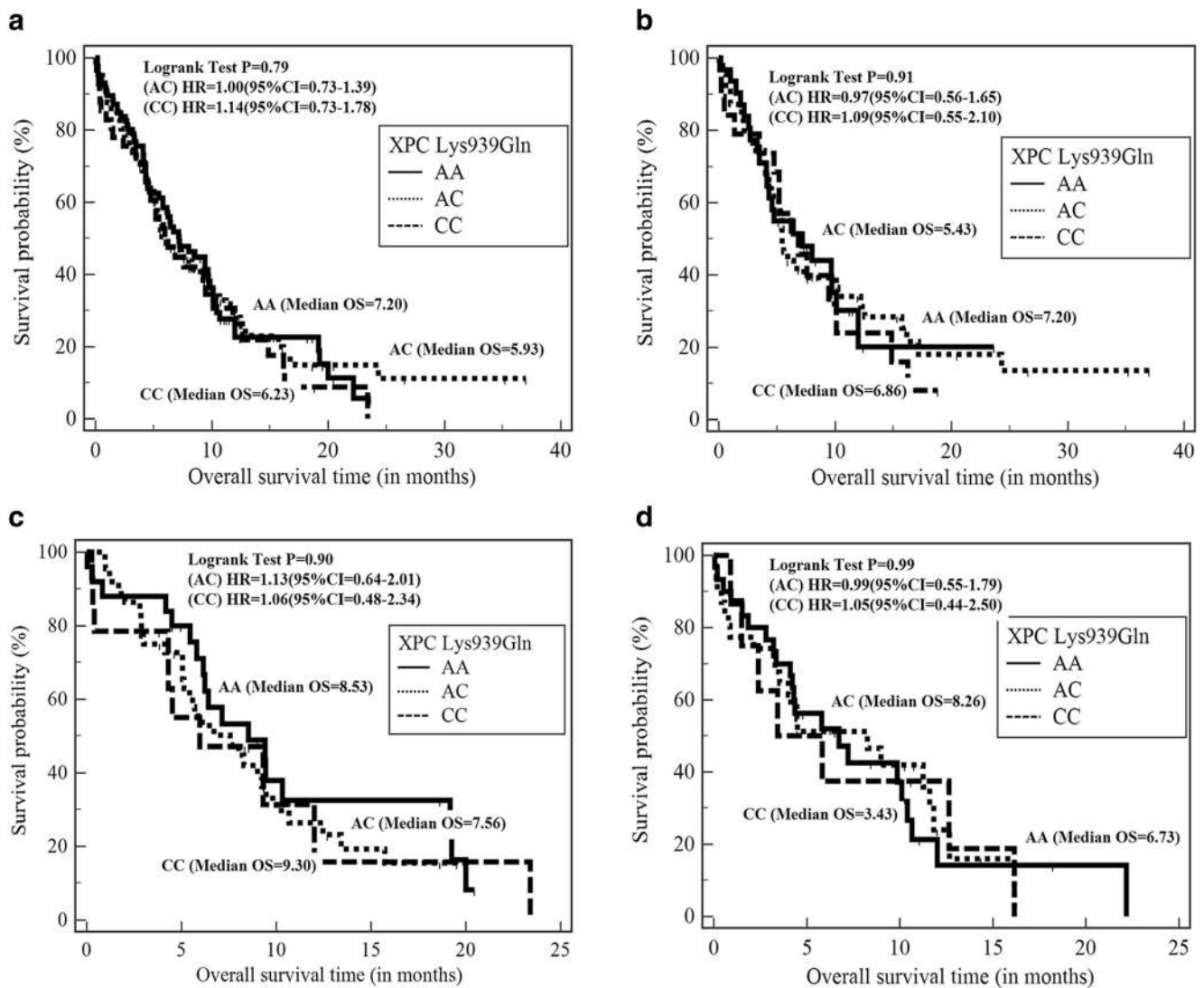


Fig. 2 The overall survival in lung cancer patients is represented in form of Kaplan Meier survival curve. The survival probability of different genotypes present in the sample population has been compared with the overall survival time in months for *XPC* Lys⁹³⁹Gln polymorphism. **(a)** showed the overall survival of *XPC* Lys⁹³⁹Gln, the dotted line represents the survival of the AC genotype (Median survival time = 5.93 months), wedged line shows the survival of CC genotype (Mean survival time = 6.23 months) while the dark black line shows the overall

survival of the wild genotype (Mean survival time = 7.20 months) with median survival time of 6.49 months with log rank test value $P=0.79$. **(b)** shows the Kaplan Meier survival curve for the patients diagnosed with Squamous cell cancer for *XPC* Lys⁹³⁹Gln polymorphism. **(c)** shows the Kaplan Meier survival curve for the patients diagnosed with Adenocarcinoma for *XPC* Lys⁹³⁹Gln polymorphism. **(d)** showed the Kaplan Meier curve for patients diagnosed with Small cell lung cancer for *XPC* Lys⁹³⁹Gln polymorphism

both the polymorphic variants were combined as a single genotype, they did not show any significant association towards the risk of lung cancer. Our results were in concordance with the above study suggesting no significant risk associated with double variant genotype of both *XPC* variants for lung cancer.

In the present study, the analysis based on the haplotypes showed that when compared to the most abundant A-C haplotype, the A-T haplotype which consist of variant T-allele, conferred a significantly decreased risk of lung cancer. Our results were in accordance with the Zhu et al. [28] who observed decreased risk of bladder cancer (OR:0.64) in subjects carrying one variant T allele of *XPC* Ala⁴⁹⁹Val genotype and

one wild A allele of *Lys*⁹³⁹Gln. They further suggested the correlation of the variant T-allele of Ala⁴⁹⁹Val with the efficient DRC [28, 30]. Therefore, the favorable allele which consisted of A-T haplotype has been found to be associated with the favorable DRC status and hence confer a decreased risk towards the lung cancer risk. Our data also showed a protective effect for A-T and C-T haplotype while C-C haplotype which consist of the adverse allele showed a slight increased effect towards the lung cancer. The possible reason behind such effect in case of adverse allele is due to their less efficient DRC which it may or may not lead to a apoptotic effect [31, 32].

Table 9 Genotypic Response of XPC polymorphism to platinum based chemotherapy and their association with lung cancer risk

Genotype	Good responders (CR + PR)(N = 89)	Poor responders (SD + PD)(N = 78)	AOR ^b (95% CI)	P ^a - value
XPC Lys ⁹³⁹ Gln				
AA	30(33.70%)	26(29.21%)	(Reference)	
AC	43(48.31%)	36(46.15%)	1.15(0.55–2.38)	0.40
CC	16(17.97%)	16(20.51%)	0.97(0.61–1.50)	0.29
AC + CC	59(66.2%)	52(66.67%)	1.11(0.56–2.19)	0.40
XPC Ala ⁴⁹⁹ Val				
CC	35(39.32%)	30(38.46%)	0.89(0.48–1.66)	0.73
CT	52(58.42%)	47(60.25%)	(Reference)	
TT	2(2.24%)	1(1.28%)	0.84(0.43–1.62)	0.44
CT + TT	54(60.67%)	48(61.5%)	1.31(0.35–4.87)	0.31
			0.85(0.44–1.69)	0.42

^a Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

^b Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status, histology, stage, kps and ecog

0 wild genotype, 1 heterozygote genotype, 2 mutant genotype, 3 combined hetero and mutant genotype

We then analyzed the effect of *XPC* polymorphism on smoking. Smoking is the widely known predominant risk factor for lung cancer. Our results suggested a two fold increase in the risk of lung cancer in the homozygous variant of *Lys*⁹³⁹*Gln* for both smokers (OR:2.55) and non-smokers. A more pronounced risk was observed in mutant (CC) genotype of heavy smokers (OR:3.71) for *XPC Lys*⁹³⁹*Gln*. As a strong risk for lung cancer was observed in both smokers and non-smokers, it can be concluded that cigarette smoking may not be the major component associated with the initiation of lung cancer but many other unidentified carcinogens may have played the major role in development of the disease. The above study showed its consistence with the study carried out by Hu et al. [33] which showed more prominent risk in individuals carrying mutant genotype amongst smokers for *Lys*⁹³⁹*Gln* polymorphism. Whereas the present study further showed a strong protective effect in individuals with *TT* and *CT* genotypes for *Ala*⁴⁹⁹*Val* polymorphism amongst smokers. These genotypes may have showed their protective effect due to their modulating response towards the carcinogenic exposure. As per the 'greater apoptosis hypothesis' as stated by Zhu et al. [28] the apoptosis is usually triggered by exposure to a carcinogen, therefore, a increased protective effect may be observed amongst smokers [31, 32].

The survival analysis of 250 patients was carried out using Kaplan-Meier survival analysis and Cox regression analysis. The investigations in the present study suggested no significant association of *Lys*⁹³⁹*Gln* variant in both univariate Kaplan Meier survival analysis and multivariate Cox regression analysis. Our results were in concordance with the study conducted by Matakideu et al. [34], as they suggested no significant association of *Lys*⁹³⁹*Gln* variant between histology and stage of diagnosis in lung cancer patients. They further emphasized on an increased mean survival for mutant

genotype in *XPC*⁹³⁹*Gln* variant, which differs from the findings of the present study. These limitations may be due to small sample size, inadequate statistical power, different ethnic background and lack of evaluation of the effects of multiple related genes in same pathways.

As reported in previous studies that platinum based chemotherapeutics are the most common regimens used for the treatment of lung cancer patients and genetic factors are recognized as the important parameters for the efficacy of individuals towards the drugs [35, 36]. Many population based studies supported a relationship between polymorphism involved in DNA repair and decreased chemotherapy response. A poor survival has been observed in NSCLC patients which is partly due to resistance shown by the drugs. It has been suggested that the genetic variants as well as protein expression might affect the response to platinum based chemotherapy as well as its prognosis [37, 38]. Platinum based chemotherapy shows its anti-cancer activity by formation of bulky platinum-DNA adducts which result in the destabilization of the DNA helix leading to the blockage of replication which further results in inhibition of transcription process. These adducts are repaired by the NER-BER pathways and plays an important role in evaluating the sensitivity of the tumor cells to the platinum based treatment [39].

The polymorphism in the DNA repair pathway genes such as *XPC* alters the DNA repair activities and may influence response in patients towards the drug which may affect patient's survival and disease outcomes [40–42]. DNA repair mechanism is a complex process and understanding it along with the heterogeneity of tumor cells makes it more complicated for us to evaluate the effect of genetic polymorphism on the repair mechanism of DNA as well as response of patients to chemotherapy due to these repair mechanism. The present study does not show any significant association of the *XPC* genetic variants with the

patients treated with the platinum based chemotherapy. This means our results does not support the possibility that *XPC* genetic polymorphism involved in DNA repair capacity may affect treatment and do not have any influence on patients survival. Our results were in agreement with the findings of Matakideu et al. [34] as they suggested no significant association between the lung cancer patients with platinum based chemotherapy and both the *XPC* genetic variants. The possible reason behind such results may be the low sample size of the study or the ethnic background of the population.

In Conclusion, our study suggested that 2 exonic variants of *XPC* gene particularly *XPC* A⁹³⁹C modulates the risk of lung cancer in North Indian population. The survival analysis conducted in present study suggests no significant association between the *XPC* variants and survival of lung cancer patients. The study further reveals that *XPC* Lys⁹³⁹Gln and genotypic *XPC* Ala⁴⁹⁹Val are not associated with an effective response to platinum based chemotherapy.

Acknowledgements We would like to express our gratitude to all the subjects who participated in this current study. This work was supported by grant from the Indian Council of Medical Research, New Delhi, India. (Grant No. 5/13/126/2011/NCD-III).

Compliance with Ethical Standards

Conflict of Interest None.

References

1. Lyon (2003) IARC: tobacco smoke and involuntary smoking. IARC
2. Malik PS, Raina V (2015) Lung cancer: prevalent trends & emerging concepts. Indian J Med Res 141(1):5–7
3. Corrao MA, Guindon GE, Cokkinides V, Sharma N (2000) Building the evidence base for global tobacco control. Bull World Health Organ 78:884–890
4. Parkin DM, Pisani P, Lopez AD, Masuyer E (1994) At least one in seven cases of cancer is caused by smoking: global estimates for 1985. Int J Cancer 59:494–504
5. Shields PG, Harris CC (2000) Cancer risk and low-penetrance susceptibility genes in gene–environment interactions. J Clin Oncol 18:2309–2315
6. Lindahl T, Wood RD (1999) Quality control by DNA repair. Science 286:1897–1905
7. Sancar A (1995) DNA repair in humans. Annu Rev Genet 29:69–105
8. de Laat WL, Jaspers NGJ, Hoeijmakers JHJ (1993) Molecular mechanism of nucleotide excision repair. Genes Dev 13:768–785
9. Wei Q, Spitz MR (1997) The role of DNA repair capacity in susceptibility to lung cancer: a review. Cancer Metastasis Rev 16:295–307
10. Matutani C, Araki M, Sugawara K, van der Spek PJ, Yamada A, Uchida A, Maekawa T, Bootsma D, Hoeijmakers JHJ, Hanaoka F (1997) Identification and characterization of XPC binding domain of HR23B. Mol Cell Biol 17:6915–6923
11. Khan SG, Metter EJ, Tarone RE, Bohr VA, Grossman L, Hedayati M, Bale SJ, Emmert S, Kraemer KH (2000) A new xeroderma pigmentosum group C poly (AT) insertion/deletion polymorphism. Carcinogenesis 21:1821–1825
12. Sanyal S, Festa F, Sakano S, Zhang Z, Steineck G, Norming U, Wijkstrom H, Larsson P, Kumar R, Hemminki K (2004) Polymorphisms in DNA repair and metabolic genes in bladder cancer. Carcinogenesis 25:729–734
13. Forsti A, Angelini S, Festa F, Sanyal S, Zhang Z, Grzybowski E, Pamula J, Pekala W, Zientek H, Hemminki K, Kumar R (2004) Single nucleotide polymorphisms in breast cancer. Oncol Rep 11: 917–922
14. Sodhi KK, Bahl C, Singh N, Behera D, Sharma S (2015) Functional genetic variants in pre-miR-146a and 196a2 genes are associated with risk of lung cancer in North Indians. Future Oncol 11(15): 2159–2173
15. Sugawara K, Ng JM, Masutani C, Iwai S, van der Spek PJ et al (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol Cell 2: 223–232
16. Raaschou-Nielsen O, Sorensen M, Overvad K, Tjonneland A, Vogel U (2008) Polymorphisms in nucleotide excision repair genes, smoking and intake of fruit and vegetables in relation to lung cancer. Lung Cancer 59:171–179
17. Bak H, Autrup H, Thomsen BL, Tjonneland A, Overvad K et al (2008) Bulky DNA adducts as risk indicator of lung cancer in a Danish case-cohort study. Int J Cancer 1189:1618–1622
18. Jin B, Dong Y, Zhang X, Wang H, Han B (2014) Association of XPC polymorphisms and lung cancer risk. A meta-analysis. Plos one 9(4):e93937
19. Kusumoto R, Masutani C, Sugawara K et al (2007) Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. Mutat Res 485:219–227
20. Yokoi M, Masutani C, Maekawa T, Sugawara K, Ohkuma Y, Hanaoka F (2000) The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. J Biol Chem 275:9870–9875
21. Sugawara K, Shimizu Y, Iwai S, Hanaoka F (2002) A molecular mechanism for DNA damage recognition by the xeroderma pigmentosum group C protein complex. DNA Repair 1:95–107
22. Shen M, Berndt SI, Rothman N, Demarini DM, Mumford JL et al (2005) Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. Int J Cancer 116:768–773
23. Vodicka P, Kumar R, Stetina R, Sanyal S, Soucek P, Hauftroid V, Dusinska M, Kuricova M, Zamecnikova M, Musak L, Buchancova J, Norppa H et al (2004) Genetic polymorphisms in DNA repair genes and possible links with DNA repair rates, chromosomal aberrations and single-strand breaks in DNA. Carcinogenesis 25:757–763
24. Deutsch-Wenzel RP, Brune H, Grimmer G, Dettbarn G, Misfeld J (1983) Experimental studies in rat lungs on the carcinogenicity and dose-response relationships of eight frequently occurring environmental polycyclic aromatic hydrocarbons. J Natl Cancer Inst 71:539–544
25. Hoffman (1993) Cigarette smoking and adenocarcinoma of the lung: the relevance of nicotine-derived nitrosamines. J Smoking Relat Disord 4:165–190
26. Xing D, Tan W, Wei Q, Lin D (2002) Polymorphisms of the DNA repair gene XPD and risk of lung cancer in a Chinese population. Lung Cancer 38:123–129
27. Yin J, Vogel U, Ma Y, Guo L, Wang H, Qi R (2006) Polymorphism of the DNA repair gene ERCC2 Lys751Gln and risk of lung cancer in a northeastern Chinese population. Cancer Genet Cytogenet 169:27–32
28. Zhu Y, Lai M, Yang H, Lin J, Huang M, Grossman HB, Colin PD, Wu X (2006) Genotypes, haplotypes and diplotypes of XPC and risk of bladder cancer. Carcinogenesis 28(3):698–703
29. Letkova L, Matakova T, Musak L, Sarlinova M, Krutakova M, Slovackova P, Kavcova E, Jakusova V, Janickova M, Drgova A, Berzinec P, Halasova E (2013) DNA repair genes polymorphism and lung cancer risk with the emphasis to sex differences. Mol Biol Rep 40:5261–5273

30. Qiao Y, Spitz MR, Guo Z, Mohammad H, Lawrence G, Kenneth KH, Wei Q (2002) Rapid assessment of repair of ultraviolet DNA damage with a modified host-cell reactivation assay using a luciferase reporter gene and correlation with polymorphisms of DNA repair genes in normal human lymphocytes. *Mutat Res* 509:165–174
31. Nelson HH, Kelsey KT, Mott LA, Karagas MR (2002) The XRCC1 Arg399Gln polymorphism, sunburn, and non-melanoma skin cancer: evidence of gene-environment interaction. *Cancer Res* 62:152–155
32. Bowen AR, Hanks AN, Allen SM, Alexander A, Diedrich MJ, Grossman D (2003) Apoptosis regulators and responses in human melanocytic and keratinocytic cells. *J Invest Dermatol* 120:48–55
33. Hu Z, Wang Y, Wang X, Liang G, Miao X, Xu Y, Tan W, Wei Q, Lin D, Shen H (2005) DNA repair gene XPC genotypes/haplotypes and risk of lung cancer in a Chinese population. *Int J Cancer* 115: 478–483
34. Matakideu A, Eisen T, Bridle H, Houlsten S. R. (2006) Nucleotide excision repair polymorphism modulate overall lung cancer survival and responsiveness to platinum based chemotherapy agents. *Jr of Clinical Oncology* 24: 10004.
35. Vicent G, Mar T (2008) Influence of DNA damage and repair upon the risk of treatment related leukemia. *Leuk Lymphoma* 49:204–217
36. Gao R, Price DK, Sissung T, Reed E, Figg WD (2008) Ethnic disparities in Americans of European descent versus Americans of African descent related to polymorphic ERCC1, ERCC2, XRCC1, and PARP1. *Mol Cancer Ther* 7:1246–1250
37. Shi C, Qian J, Ma M, Zhang Y, Han B (2014) Notch 3 protein, not its gene polymorphism, is associated with the chemotherapy response and prognosis of advanced NSCLC patient. *Cell Physiol Biochem* 34:743–752
38. Wu F, Zhang J, Liu Y, Zheng Y, Hu N (2013) HIF1 α genetic variants and protein expressions determine the response to platinum based chemotherapy and clinical outcome in patients with advanced NSCLC. *Cell Physiol Biochem* 32: 1566–1576.
39. Sullivan I, Salazar J, Majem M, Pallares C, Del Rio E, Paez D, Baiget M, Barnadas (2014) A: pharmacogenetics of the DNA repair pathways in advanced non-small cell lung cancer patients treated with platinum-based chemotherapy. *Cancer Lett* 353: 160–166.
40. Tiseo M, Bordi P, Bortesi B, Boni L, Boni C, Baldini E, Grossi F, Recchia F, Zanelli F, Fontanini G, Naldi N, Campanini N, Azzoni C, Bordi C, Ardizzoni A (2013) ERCC1/BRCA1 expression and gene polymorphisms as prognostic and predictive factors in advanced NSCLC treated with or without cisplatin. *Br J Cancer* 108:1695–1703
41. Xu TP, Shen H, Liu LX, Shu YQ (2013) Association of ERCC1-C118T and -C8092A polymorphisms with lung cancer risk and survival of advanced-stage non-small cell lung cancer patients receiving platinum-based chemotherapy: a pooled analysis based on 39 reports. *Gene* 526:265–274
42. Li XD, Han JC, Zhang YJ, Li HB, Wu XY (2013) Common variations of DNA repair genes are associated with response to platinum-based chemotherapy in NSCLCs. *Asian Pac J Cancer Prev* 14:145–148

RESEARCH ARTICLE

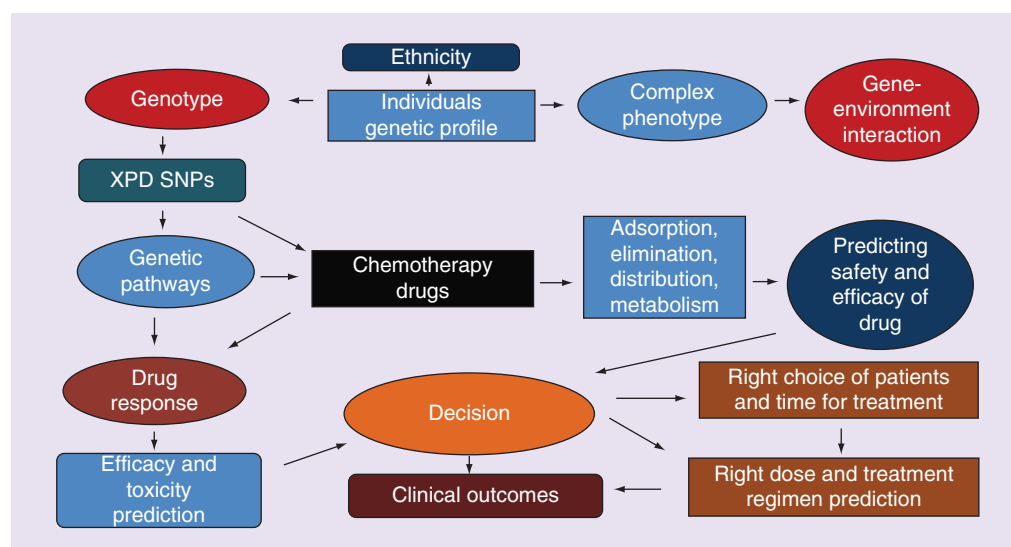
For reprint orders, please contact: reprints@futuremedicine.com

Xeroderma pigmentosum complementation group D polymorphism toward lung cancer susceptibility survival and response in patients treated with platinum chemotherapy

Shweta Lawania¹, Navneet Singh², Digamber Behera² & Siddharth Sharma^{*1}

Aim: The study investigated role of xeroderma pigmentosum complementation group D (XPD) single nucleotide polymorphisms in modulating lung cancer risk and its association with overall survival and clinical outcomes. **Methods:** XPD polymorphisms were detected using Polymerase Chain Reaction Restriction Fragment Length Polymorphism (PCR-RFLP). **Results:** CC genotype of A751C polymorphism was associated with an increased lung cancer risk ($p = 0.01$). Classification and Regression Tree (CART) analysis depicted C156A as the major contributing factor. Patients having CC, treated with irinotecan-cisplatin/carboplatin regimen showed a better survival (median survival time = 25.2) whereas a poor survival was for XPDG312A. Similarly, patients treated with pemetrexed and carrying heterozygous genotype of G312A polymorphism had a poor survival ($p = 0.01$). **Conclusion:** A751C and G312A act as a predictive marker in lung cancer patients treated with platinum chemotherapy. These findings might facilitate therapeutic decisions for individualized therapy in lung cancer patient.

First draft submitted: 3 May 2017; Accepted for publication: 9 June 2017; Published online: 16 October 2017



¹Department of Biotechnology, Thapar University, Patiala, Punjab 147002, India

²Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research (PGIMER), Sector 14, Chandigarh, India

*Author for correspondence: siddharthsharma.phd@thapar.edu

KEYWORDS

• chemotherapy regimen
• overall survival • platinum-based chemotherapy • *XPD* polymorphism

DNA repair systems play a critical role in maintaining genomic integrity and protecting it against cancer causing mutations [1]. Defective DNA repair capacity may lead to cancer susceptibility via oncogene activation, tumor suppressor gene inactivation or heterozygosity [2]. Lung cancer is a polygenic disease mainly caused by tobacco smoke but still only $\leq 20\%$ of smokers develop the disease [3]. Cigarette smoking induces the formation of bulky DNA adducts, cross-links and DNA strand breaks resulting into DNA damage which is mainly repaired by nucleotide excision repair (NER) pathway [4].

Xeroderma pigmentosum complementation group D (*XPD*)/excision repair cross complementation group 2 (*ERCC2*) is an ATP-dependent DNA helicase. It is essential for transcription and NER activity [5]. Many SNPs are located on *XPD* locus but only three SNPs present on codon 156, 312 and 751 are studied as they have high allele frequency. *XPD Arg¹⁵⁶Arg* (rs238406) polymorphic variant located in exon 6 represents a silent polymorphism leading to C→A transition, resulting in an Arg→Arg substitution at codon 156 [6]. Another SNP of the *XPD* gene is present in codon 312 (rs1799793) of exon 10, resulting into G→A substitution, which changes aspartate (Asp) into asparagine (Asn) [7]. The third SNP present on codon 751 (rs13181) and located at exon 23 of *XPD* gene is characterized by A→C substitution resulting in change of lysine (Lys) to glutamine (Gln) [7]. Various epidemiological studies have suggested association of *XPD* polymorphism with different types of cancers such as head and neck [8], melanoma [9], basal cell carcinoma [10], lung [11] and bladder cancer [12], although the results reported were inconsistent.

Chemotherapy has been a major process for lung cancer treatment [13]. From the available chemotherapy regimens, platinum-based doublets have been associated with improved overall survival (OS). Platinum-based compounds such as cisplatin forms inter- as well as intra-DNA strands leading to bulky adduct formation resulting into double helix destabilization and hence inhibition of DNA replication [14]. The suboptimal DNA repair present in the tumor may result into decreased platinum-DNA adduct formation and therefore better clinical response which ultimately leads to improved survival of cancer patients [15]. DNA repair is a composite process and is carried out by a series of DNA repair pathways such as NER. NER pathway is responsible for the repair of majority of

platinum-DNA adducts followed by restoration of DNA segment. SNPs of NER genes modulate repair capacity which may contribute to individual's variations toward chemotherapeutic regimen.

Therefore, a case-control study was conducted to examine association of three polymorphic sites of *XPD* gene (*A751C*, *G312A*, *C156A*) toward lung cancer susceptibility and its histological subtypes. The interactive effect of smoking on *XPD* variants and its association with lung cancer was analyzed. High-order SNP-SNP interaction was investigated using multifactor dimensionality reduction (MDR) and Classification and Regression Tree (CART) analysis. We have further assessed the role of three SNPs toward OS of patients treated with platinum-based chemotherapy and clinical outcomes.

Materials & methods

• Study population

The present study was a hospital-based case-control study conducted for 370 cases and 370 healthy controls who were registered in the Department of Pulmonary Medicine of Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh. The Ethical Committee Board of PGIMER had reviewed and approved the study. The patients included in the study were selected without any age, gender, smoking, histology or tumor-node metastasis (TNM) restrictions. Patients with any prior cancer history were eluded. Healthy controls were selected among individuals who visited hospital for routine check-up and were not having cancer. Written consent from each individual was obtained. A detailed questionnaire about demographic and lifetime smoking history for each case and control was completed by a trained interviewer. Smoking exposure was calculated using following formula: (cigarette or beedis [Indian cigarette] per day/20 × years smoked). The details regarding TNM staging, regimen of treatment and number of chemotherapeutic cycles were obtained from the patient's medical records present in the hospital. The follow-up data were obtained telephonically from the information present in their medical records.

• Genotyping of *XPD* variants

Genomic DNA was isolated from 5-ml venous blood sample using phenol chloroform extraction protocol by Sodhi *et al.* [16]. Genotyping was performed using PCR-RFLP method as described

Table 1. Demographic characteristics among cases and control patients in North Indian population.

Variable	Cases, n (%) n = 370	Controls, n (%) n = 370	p [†] -value
Age (years)			
Mean ± SD	58.11 ± 10.44	53.83 ± 10.18	<0.0001
Range			
Gender			
Male	319 (86.21%)	319 (86.21%)	0.91
Female	51 (13.78%)	51 (13.78%)	
Smoking status			
Smokers	303 (81.89%)	271 (73.24%)	0.0063
Nonsmokers	67 (18.10%)	99 (26.75%)	
Pack years			
Mean ± SD	24.97 ± 35.04	17.88 ± 19.45	<0.0001
Histological types			
SQCC	135 (36.48%)		
ADCC	120 (32.43%)		
SCLC	111 (30%)		
Others	4 (1.08%)		
Unknown	0 (0.0)		
TNM staging			
I	3 (0.81%)		
II	10 (2.70%)		
III	175 (45.94%)		
IV	163 (44.05%)		
Unclassified	19 (5.13%)		
Tumor size			
Tx	11 (2.97%)		
T1	16 (4.32%)		
T2	42 (11.35%)		
T3	85 (22.97%)		
T4	191 (51.62%)		
Unknown	25 (6.75%)		
Lymph node involvement			
N0	46 (12.43%)		
N1	36 (9.72%)		
N2	157 (42.43%)		
N3	106 (28.64%)		
Unknown	25 (6.75%)		
Metastasis			
M0	187 (50.54%)		
M1	158 (42.70%)		
Unknown	25 (6.75%)		

[†]p-values were derived from Pearson's chi-square test except age; Student's t-test was used for age. All p-values are two-sided. p < 0.05 was considered statistically significant.

ADCC: Adenocarcinoma; n: Total number of case patients or control subjects; SD: Standard deviation; SQCC: Squamous cell carcinoma; TNM: Tumor-node metastasis.

in previous studies [2,17]. For all the polymorphisms, 15% samples were repeated twice using same procedure and the reproducibility of results was 100%.

• Chemotherapeutic treatment

All the patients in the present study received platinum-based chemotherapy and the

chemotherapy drugs that included cisplatin or carboplatin in combination with either docetaxel, irinotecan or pemetrexed. This combination chemotherapy has docetaxel (75 mg/m²), pemetrexed (500 mg/m²) or irinotecan (75 mg/m²) administered as a 1-h infusion which was followed by cisplatin 70 mg/m² iv. infused for over 3 h. All

Table 2. Distribution of genotypes of Xeroderma pigmentosum complementation group D polymorphism depicting their association with lung cancer risk and its histological subtypes.

Overall lung cancer									
SQCC					ADCC				
XPD A751C	Controls n (%)	Cases n (%)	AOR (95% CI) [†]	p [‡]	Cases n (%)	AOR (95% CI) [†]	p [‡]	Cases n (%)	AOR (95% CI) [†]
	n = 370	n = 370			n = 135			n = 120	
AA	203 (54.86)	187 (50.54)	1.00 (Reference)		70 (51.85)	1.00 (Reference)		60 (50)	1.00 (Reference)
AC	139 (37.56)	130 (35.13)	1.03–1.43	0.81	48 (35.55)	0.94 (0.60–1.48)	0.80	42 (35)	1.02 (0.64–1.63)
CC	28 (7.5)	53 (14.32)	1.96 (1.17–3.27)	0.01	17 (12.59)	1.85 (0.91–3.76)	0.008	18 (15)	2.07 (1.05–4.08)
AC + CC	167 (45.13)	183 (49.45)	1.21 (0.90–1.63)	0.20	65 (48.14)	1.09 (0.71–1.65)	0.67	60 (50)	1.23 (0.80–1.89)
A allele	545 (73.64)	504 (68.10)							
C allele	195 (26.35)	236 (31.89)							
MAF	0.26	0.32							
XPDG312A									
GG	176 (47.56)	172 (46.48)	1.00 (Reference)		66 (48.88)	1.00 (Reference)		56 (46.66)	1.00 (Reference)
GA	167 (45.13)	169 (45.67)	1.06 (0.78–1.44)	0.69	54 (40)	0.86 (0.55–1.33)	0.51	58 (48.33)	1.19 (0.77–1.84)
AA	27 (7.29)	29 (7.83)	1.17 (0.65–2.10)	0.58	15 (11.11)	1.62 (0.78–3.37)	0.19	6 (5)	0.89 (0.33–2.34)
GA + AA	194 (52.43)	198 (53.51)	1.07 (0.79–1.45)	0.62	69 (51.11)	0.75 (0.48–1.16)	0.20	64 (53.33)	1.15 (0.75–1.76)
G allele	519 (70.13)	513 (69.32)							
A allele	221 (29.86)	227 (30.67)							
MAF	0.30	0.31							
XPDC156A									
CC	153 (41.35)	179 (48.37)	1.00 (Reference)		75 (55.55)	1.00 (Reference)		48 (40)	1.00 (Reference)
CA	185 (50)	167 (45.13)	0.82 (0.60–1.13)	0.23	47 (34.81)	0.63 (0.40–0.99)	0.04	64 (53.33)	1.04 (0.67–1.63)
AA	32 (8.64)	29 (7.83)	0.83 (0.47–1.46)	0.52	13 (9.62)	1.07 (0.49–2.34)	0.84	8 (66.67)	0.73 (0.31–1.73)
CA + AA	217 (58.64)	196 (52.97)	0.83 (0.61–1.12)	0.23	60 (44.44)	0.69 (0.45–1.05)	0.08	72 (60)	1.00 (0.65–1.54)
C allele	491 (66.35)	525 (70.94)							
A allele	249 (33.64)	225 (30.40)							
MAF	0.34	0.30							

[†]Adjusted ORs, 95% CIs and their corresponding p-values were calculated using logistic regression analysis after adjusting for age, gender and smoking status.

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

ADCC: Adenocarcinoma; AOR: Adjusted odds ratio; CI: Confidence interval; OR: Odds ratio; SCLC: Small cell lung carcinoma; SQCC: Squamous cell carcinoma; XPD: Xeroderma pigmentosum complementation group D.

Table 3. Distribution of genotypes of xeroderma pigmentosum complementation group D polymorphism according to their smoking status and its association with lung cancer risk.

<i>XPD A751C</i>	Smokers				Nonsmokers			
	Controls n (%) n = 263	Cases n (%) n = 303	AOR (95% CI) ^a	p [†]	Controls n (%) n = 107	Cases n (%) n = 67	AOR (95% CI) [†]	p ^b
AA	144 (54.75)	152 (50.16)	1.00 (Reference)		59 (55.14)	35 (52.28)	1.00 (Reference)	
AC	100 (38.02)	106 (34.98)	1.06 (0.73–1.53)	0.73	39 (36.44)	24 (35.82)	1.00 (0.50–1.99)	0.98
CC	19 (7.22)	45 (14.85)	2.11 (1.15–3.86)	0.01	9 (8.41)	8 (11.94)	1.51 (0.52–4.33)	0.44
AC + CC	119 (32.16)	151 (49.83)	1.23 (0.87–1.74)	0.22	48 (44.85)	32 (47.76)	1.12 (0.59–2.11)	0.71
<i>XPD G312A</i>								
GG	125 (47.52)	134 (44.22)	1.00 (Reference)		51 (47.66)	38 (56.71)	1.00 (Reference)	
GA	119 (45.24)	141 (46.53)	1.12 (0.79–1.61)	0.50	48 (44.85)	28 (41.79)	0.77 (0.40–1.47)	0.43
AA	19 (7.22)	28 (9.24)	1.56 (0.82–2.98)	0.17	8 (7.76)	1 (1.49)	0.20 (0.02–1.73)	0.14
GA + AA	138 (52.47)	169 (55.77)	1.19 (0.84–1.68)	0.30	56 (52.33)	29 (43.28)	0.69 (0.36–1.31)	0.26
<i>XPD C156A</i>								
CC	117 (44.48)	145 (47.85)	1.00 (Reference)		36 (33.64)	29 (43.28)	1.00 (Reference)	
CA	127 (48.28)	133 (43.89)	0.88 (0.62–1.27)	0.51	58 (54.20)	34 (50.74)	0.70 (0.36–1.37)	0.30
AA	19 (7.22)	25 (8.25)	1.12 (0.58–2.16)	0.72	13 (12.14)	4 (5.97)	0.39 (0.11–1.38)	0.14
CA + AA	146 (55.51)	158 (52.14)	0.92 (0.65–1.29)	0.63	71 (66.35)	38 (56.71)	0.64 (0.33–1.23)	0.18
<i>XPD A751C</i>	Heavy smoker				Light smoker			
	Controls n (%) n = 96	Cases n (%) n = 141	AOR (95% CI) [†]	p [†]	Controls n (%) n = 167	Cases n (%) n = 162	AOR (95% CI) [†]	p ^b
AA	52 (54.16)	70 (49.64)	1.00 (Reference)		92 (55.08)	82 (50.61)	1.00 (Reference)	
AC	37 (38.54)	50 (35.46)	1.62 (0.87–3.00)	0.12	63 (37.72)	56 (34.56)	1.01 (0.62–1.64)	0.94
CC	7 (7.29)	21 (14.89)	2.31 (0.90–5.94)	0.08	12 (7.18)	24 (14.81)	1.93 (0.87–4.27)	0.10
AC+CC	44 (45.83)	71 (50.35)	1.77 (1.00–3.13)	0.04	75 (44.91)	106 (65.43)	1.17 (0.74–1.83)	0.48
<i>XPD G312A</i>								
GG	38 (39.58)	70 (49.64)	1.00 (Reference)		87 (52.09)	64 (39.50)	1.00 (Reference)	
GA	50 (52.08)	60 (42.55)	0.64 (0.36–1.12)	0.12	69 (41.31)	81 (50)	1.65 (1.02–2.65)	0.03
AA	8 (8.33)	11 (7.80)	0.78 (0.28–2.16)	0.64	11 (6.58)	18 (11.11)	2.45 (1.05–5.69)	0.03
GA + AA	58 (60.41)	71 (50.35)	0.76 (0.43–1.32)	0.33	80 (47.90)	99 (61.11)	2.03 (1.28–3.23)	0.02
<i>XPD C156A</i>								
CC	49 (51.04)	67 (47.51)	1.00 (Reference)		68 (40.71)	78 (48.14)	1.00 (Reference)	
CA	42 (43.75)	61 (43.26)	1.05 (0.60–1.84)	0.84	85 (50.89)	72 (44.44)	0.78 (0.49–1.25)	0.31
AA	5 (5.20)	13 (9.21)	1.43 (0.46–4.44)	0.52	14 (8.38)	12 (7.40)	0.84 (0.36–1.97)	0.69
CA + AA	47 (48.95)	80 (56.73)	1.11 (0.65–1.91)	0.68	99 (59.28)	84 (51.85)	0.79 (0.50–1.25)	0.32

[†]Adjusted ORs, 95% CIs and their corresponding p-values were calculated logistic regression analysis after adjusting for age, gender and pack years of smoking. The smoking data consist of smokers (0.3–100 pack years) and nonsmokers (0 pack years). On stratification the basis of pack years: light smokers (≤25 pack years) and heavy smokers (>25 years).

^bTwo-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.

CI: Confidence interval; OR: Odds ratio; XPD: Xeroderma pigmentosum complementation group D.

patients received standard folate and vitamin B₁₂ supplementation. In our study, 76 patients were administered docetaxel + cisplatin/docetaxel + carboplatin, 70 patients were treated with irinotecan + cisplatin/irinotecan + carboplatin and 69 patients received pemetrexed + cisplatin/pemetrexed + carboplatin. Premedication included dexamethasone, grisetron and ranitidine. All the chemotherapeutic drugs were administered intravenously, and treatment cycles were repeated every

3–4 weeks. As per normal protocol followed at our institute, four cycles of chemotherapy were administered before performing tumor response assessment. In presence of unacceptable toxicity or clinic-radiological suggestion of disease progression, tumor response assessment was performed prior to completion of four cycles and chemotherapy was stopped, if indicated. Patients with an objective response to treatment received two additional cycles (maximum of six cycles).

Table 4. Relationship of xeroderma pigmentosum complementation group D polymorphism with overall survival of lung cancer patients.

Overall ILC																ADCC					
XPD A751C	Dead 256	Alive 55	Media OS	HR (95% CI)	Log p	HR (95% CI)	p-value	XPD A751C	Dead 86	Alive 18	Median OS	HR (95% CI)	Log p	HR' (95% CI)	p-value						
AA	127	25	7.13	1.00 (Reference)		1.00 (Reference)		AA	41	8	7.13	1.00 (Reference)		1.00 (Reference)							
AC	95	21	8.03	0.94 (0.72–1.23)	0.68	1.00 (0.76–1.31)	0.97	AC	33	7	9.00	0.89 (0.56–1.40)	0.62	0.94 (0.57–1.56)	0.82						
CC	34	9	9.23	0.84 (0.58–1.21)	0.37	0.80 (0.54–1.19)	0.28	CC	12	3	9.23	0.85 (0.45–1.57)	0.62	0.68 (0.33–1.38)	0.29						
AC + CC	129	30	8.40	0.92 (0.71–1.17)	0.50	0.95 (0.74–1.22)	0.95	AC + CC	45	10	9.36	0.88 (0.51–1.35)	0.56	0.93 (0.59–1.47)	0.76						
SCLC																SQCC					
XPD A751C	Dead 75	Alive 15	Median OS	HR (95% CI)	Log p	HR' (95% CI)	p-value	XPD A751C	Dead 95	Alive 22	Median OS	HR (95% CI)	Log p	HR' (95% CI)	p-value						
AA	38	5	5.80	1.00 (Reference)		1.00 (Reference)		AA	48	12	8.63	1.00 (Reference)		1.00 (Reference)							
AC	27	6	8.26	0.82 (0.50–1.34)	0.43	0.81 (0.48–1.36)	0.44	AC	35	8	7.26	1.74 (0.73–1.77)	0.54	1.12 (0.70–1.78)	0.62						
CC	10	4	12.7	0.61 (0.33–1.14)	0.16	0.49 (0.23–1.03)	0.05	CC	12	2	6.46	1.20 (0.61–2.35)	0.56	1.20 (0.63–2.29)	0.57						
AC + CC	37	10	9.86	0.75 (0.47–1.18)	0.21	0.70 (0.44–1.12)	0.14	AC + CC	47	10	6.46	1.16 (0.77–1.74)	0.45	1.13 (0.74–1.72)	0.56						
Overall ILC																ADCC					
XPD G312A	Dead 256	Alive 55	Median OS	HR (95% CI)	Log p	HR' (95% CI)	p-value	XPD G312A	Dead 86	Alive 18	Median OS	HR (95% CI)	Log p	HR' (95% CI)	p-value						
GG	118	26	9.43	1.00 (Reference)		1.00 (Reference)		GG	38	11	9.46	1.00 (Reference)		1.00 (Reference)							
GA	118	24	8.03	1.16 (0.90–1.51)	0.22	1.16 (0.89–1.50)	0.25	GA	43	6	8.80	1.30 (0.84–2.02)	0.22	1.39 (0.88–2.20)	0.15						
AA	20	4	5.43	1.30 (0.77–2.20)	0.26	1.28 (0.79–2.09)	0.31	AA	5	1	7.16	1.28 (0.45–3.59)	0.59	1.48 (0.54–4.07)	0.44						
GA + AA	138	28	7.13	1.18 (0.92–1.51)	0.17	1.16 (0.90–1.49)	0.22	GA + AA	48	7	7.60	1.31 (0.85–2.00)	0.20	1.41 (0.90–2.21)	0.13						
SCLC																SQCC					
XPD G312A	Dead 75	Alive 15	Median OS	HR (95% CI)	Log p	HR' (95% CI)	p-value	XPD G312A	Dead 95	Alive 22	Median OS	HR (95% CI)	Log p	HR' (95% CI)	p-value						
GG	35	3	7.23	1.00 (Reference)		1.00 (Reference)		GG	46	12	9.43	1.00 (Reference)		1.00 (Reference)							
GA	35	11	9	0.68 (0.42–1.10)	0.11	0.62 (0.38–1.02)	0.06	GA	39	7	5.56	1.47 (0.94–2.30)	0.07	1.60 (1.02–2.53)	0.04						
AA	5	1	1.70	2.81 (0.66–11.94)	0.01	13.38 (3.59–49.81)	0.0001	AA	10	3	6.46	1.15 (0.56–2.37)	0.67	1.04 (0.51–2.10)	0.91						
GA + AA	40	12	8.4	0.76 (0.48–1.21)	0.23	0.72 (0.45–1.16)	0.18	GA + AA	49	10	5.56	1.39 (0.92–2.09)	0.10	1.46 (0.15–2.24)	0.08						

*Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype.

[†]Hazard ratios, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted HRs, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.

DDCC: Adenocarcinoma; CI: Confidence interval; HR: Hazard ratio; SCLC: Small cell lung carcinoma; SQCC: Squamous cell carcinoma; XPD: Xeroderma pigmentosum complementation group D.

[†]Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype.
[‡]Hazard ratios, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted HRs, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.
ADCC: Adenocarcinoma; CI: Confidence interval; HR: Hazard ratio; SCLC: Small cell lung carcinoma; SQCC: Squamous cell carcinoma; XPD: Xeroderma pigmentosum complementation group D.

Table 4. Relationship of xeroderma pigmentosum complementation group D polymorphism with overall survival of lung cancer patients (cont.).

Overall LLC										ADCC					
XPB C156A	Dead 256	Alive 55	Median OS (months)	HR (95% CI)	Log p	HR ¹ (95% CI)	p-value	XPB C156A	Dead 86	Alive 18	Median OS (months)	HR (95% CI)	Log p	HR ¹ (95% CI)	p-value
CC	127	27	7.13	1.00 (Reference)		1.00 (Reference)		CC	38	5	7	1.00 (Reference)		1.00 (Reference)	
CA	106	27	8.63	0.98 (0.76–1.27)	0.91	1.04 (0.80–1.37)	0.72	CA	42	13	9	0.80 (0.52–1.27)	0.36	0.89 (0.57–1.40)	0.63
AA	23	1	9.46	1.18 (0.74–1.90)	0.44	1.25 (0.80–1.98)	0.31	AA	6	–	9.90	1.00 (0.42–2.39)	0.98	1.07 (0.43–2.63)	0.87
CA + AA	129	28	8.80	1.01 (0.79–1.30)	0.88	1.08 (0.83–1.39)	0.54	CA + AA	48	13	9.13	0.83 (0.53–1.28)	0.39	0.90 (0.58–1.39)	0.90
SCLC								SQCC							
XPB C156A	Dead 75	Alive 15	Median OS (months)	HR (95% CI)	Log p	HR ¹ (95% CI)	p-value	XPB C156A	Dead 95	Alive 22	Median OS (months)	HR (95% CI)	Log p	HR ¹ (95% CI)	p-value
CC	35	6	8.40	1.00 (Reference)		1.00 (Reference)		CC	54	16	6.86	1.00 (Reference)		1.00 (Reference)	
CA	33	9	8.26	0.87 (0.54–1.41)	0.58	0.82 (0.49–1.37)	0.47	CA	31	5	8.03	1.19 (0.75–1.41)	0.42	1.33 (0.81–2.18)	0.25
AA	7	–	7.23	1.20 (0.50–2.87)	0.64	1.08 (0.45–2.60)	0.85	AA	10	1	7.26	1.25 (0.60–2.61)	0.50	1.23 (0.61–2.45)	0.55
CA + AA	48	13	8.26	0.91 (0.58–1.44)	0.69	0.85 (0.52–1.40)	0.54	CA+AA	41	6	8.03	1.22 (0.80–1.84)	0.33	1.28 (0.83–1.97)	0.26
*Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype.															
†Hazard ratios, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted HRs, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.															
ADCC: Adenocarcinoma; CI: Confidence interval; HR: Hazard ratio; SCLC: Small cell lung carcinoma; SQCC: Squamous cell carcinoma; XPD: Xeroderma pigmentosum complementation group D.															

¹Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype. 1: combined hetero and mutant genotype.

²Hazard ratios, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted HRs, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.

ADCC: Adenocarcinoma; CI: Confidence interval; HR: Hazard ratio; SCLC: Small cell lung carcinoma; SQCC: Squamous cell carcinoma; XPB: Xeroderma pigmentosum complementation group D.

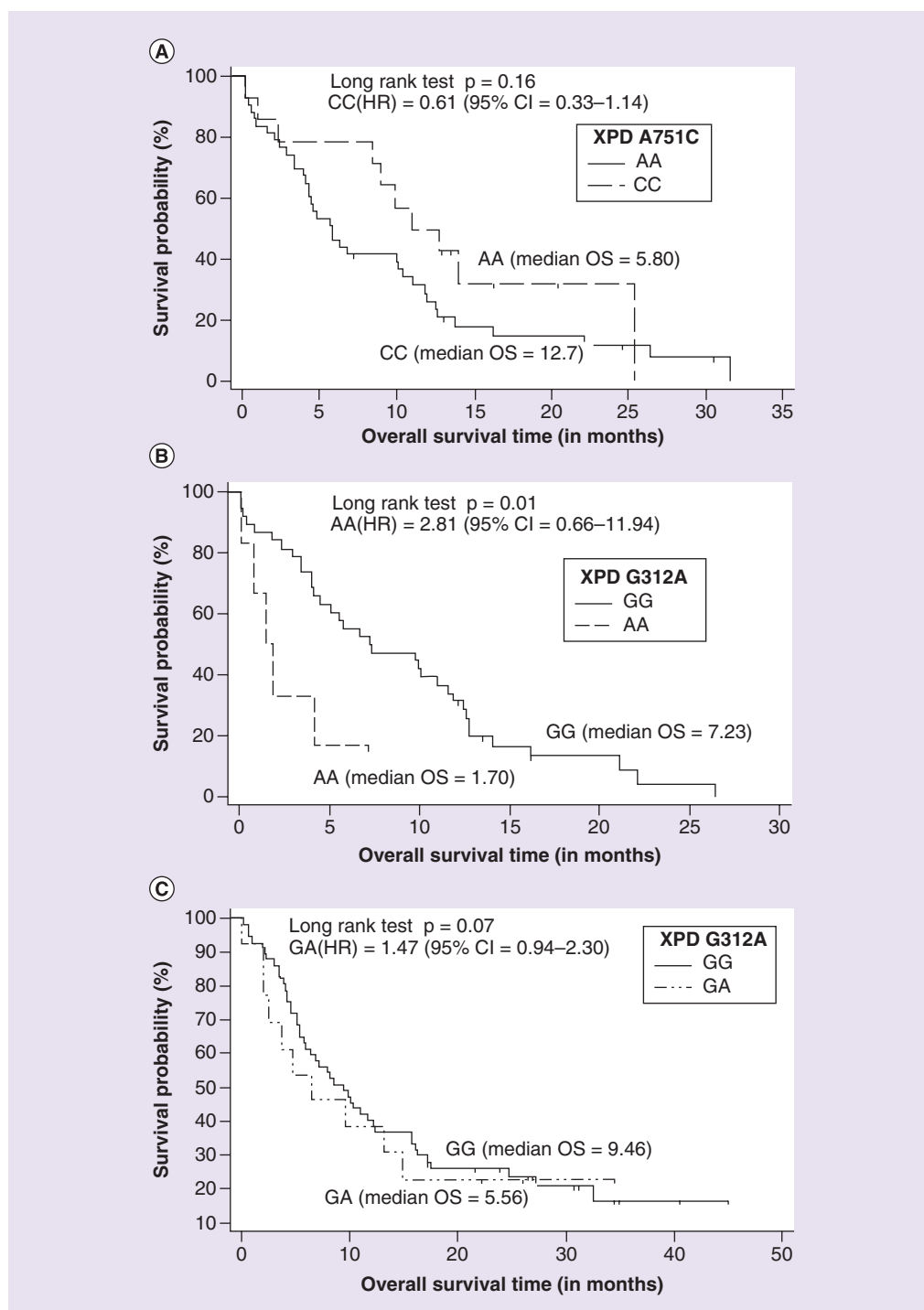


Figure 1. Kaplan Meier curves illustrating association between overall survival in patients with (A) Mutant genotype (CC) for SCLC showing XPD Lys751Gln polymorphism (B) Mutant genotype (AA) for SCLC and (C) Heterozygous genotype (GA) depicting XPD Asp312Asn polymorphism.

• Power calculations

To calculate the sample size required to replicate the association between polymorphic variants and lung cancer risk, we conducted power

analysis. Power calculations were performed using CaTS program (<http://www.sph.umich.edu/csg/abecasis/CaTS/>). We estimated that 370 controls and 370 cases would provide >99%

Table 5. Association of Xeroderma pigmentosum complementation group D polymorphism and overall survival according to regimen.

Regimen – doce cis/doce carb								
Genotype (N = 76)	Number	Alive	Dead	Median OS (months)	Crude HR (95% CI)	Adjusted HR ¹ (95% CI)	log-rank p	p ³
XPD A751C								
AA	41	6	35	8.63	1.00 (Reference)	1.00 (Reference)		
AC	25	6	19	10.13	0.83 (0.48–1.44)	0.81 (0.44–1.49)	0.52	0.50
CC	10	2	8	5.76	1.18 (0.52–2.66)	1.04 (0.45–2.41)	0.66	0.92
AC + CC	35	8	27	7.56	0.92 (0.56–1.52)	0.88 (0.51–1.50)	0.75	0.65
XPD G312A								
GG	31	6	25	9.86	1.00 (Reference)	1.00 (Reference)		
GA	35	5	30	7.56	1.34 (0.79–2.28)	1.73 (0.95–3.15)	0.27	0.07
AA	10	3	7	8.21	0.91 (0.40–2.08)	0.83 (0.50–1.39)	0.84	0.50
GA + AA	45	8	37	7.56	1.23 (0.74–2.03)	1.47 (0.83–2.61)	0.41	0.18
XPD C156A								
CC	44	10	34	8.20	1.00 (Reference)	1.00 (Reference)		
CA	25	3	22	6.86	1.37 (0.78–2.41)	1.81 (0.94–3.49)	0.23	0.07
AA	7	1	6	12.3	1.08 (0.44–2.63)	1.30 (0.51–3.27)	0.86	0.57
CA + AA	32	4	28	7.75	1.30 (0.78–2.17)	1.58 (0.89–2.80)	0.29	0.11
Regimen – irino cis/irino carb								
Genotype (N = 70)	Number	Alive	Dead	Median OS(months)	Crude HR (95% CI)	Adjusted HR ¹ (95% CI)	log-rank P	p ³
XPD A751C								
AA	31	5	26	5.63	1.00 (Reference)	1.00 (Reference)		
AC	26	7	19	8.63	0.64 (0.35–1.15)	0.61 (0.30–1.21)	0.13	0.16
CC	13	5	8	25.43	0.39 (0.20–0.77)	0.21 (0.08–0.57)	0.01	0.002
AC + CC	39	12	27	11	0.54 (0.30–0.95)	0.48 (0.21–0.92)	0.48	0.02
XPD G312A								
GG	26	4	22	7.28	1.00 (Reference)	1.00 (Reference)		
GA	39	12	27	10.4	0.80 (0.45–1.43)	0.51 (0.31–1.04)	0.45	0.05
AA	5	1	4	1.86	3.29 (0.58–18.54)	16.46 (2.50–108.40)	0.01	0.003
GA + AA	44	13	31	9	0.89 (0.51–1.55)	0.69 (0.38–1.23)	0.69	0.69
XPD C156A								
CC	29	7	22	6.73	1.00 (Reference)	1.00 (Reference)		
CA	34	10	24	9	1.02 (0.57–1.83)	0.72 (0.38–1.37)	0.92	0.32
AA	7	–	7	7.23	1.60 (0.60–4.28)	1.44 (0.52–3.95)	0.26	0.47
CA	41	10	31	9	1.12 (0.65–1.92)	0.86 (0.46–1.59)	0.67	0.64
Regimen – pem cis/pem carb								
Genotype (N = 69)	Number	Alive	Dead	Median OS (months)	Crude HR (95% CI)	Adjusted HR ¹ (95% CI)	Log rank P	p ³
XPD A751C								
AA	39	8	31	9.13	1.00 (Reference)	1.00 (Reference)		
AC	21	4	17	12.46	0.86 (0.48–1.53)	0.88 (0.45–1.70)	0.61	0.71
CC	9	2	7	10.33	0.82 (0.37–1.77)	0.67 (0.26–1.70)	0.63	0.40
AC + CC	30	6	24	12.13	0.85 (0.50–1.44)	0.88 (0.45–1.70)	0.85	0.71

¹Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotypes.

²HRs, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted HRs, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.

carb: Carboplatin; CI: Confidence interval; cis: Cisplatin; doce: Docetaxel; HR: Hazard ratio; irino: Irinotecan; pem: Pemetrexed; XPD: Xeroderma pigmentosum complementation group D.

Table 5. Association of Xeroderma pigmentosum complementation group D polymorphism and overall survival according to regimen. (cont).

XPD G312A								
GG	33	11	22	12.46	1.00 (Reference)	1.00 (Reference)		
GA	31	2	29	8.8	2.06 (1.17–3.6)	2.01 (1.14–3.56)	0.007	0.01
AA	5	1	4	7.56	1.70 (0.46–6.25)	1.52 (0.43–5.33)	0.31	0.51
GA + AA	36	3	33	7.56	2.02 (1.18–3.45)	1.89 (1.08–3.30)	0.008	0.02
XPD C156A								
CC	32	4	28	9.71	1.00 (Reference)	1.00 (Reference)		
CA	32	10	22	9.01	0.75 (0.43–1.31)	0.96 (0.52–1.76)	0.32	0.90
AA	5	-	5	10.33	1.17 (0.42–3.21)	1.32 (0.78–2.24)	0.73	0.29
CA + AA	37	10	27	9.46	0.79 (0.46–1.35)	1.00 (0.56–1.77)	0.39	0.99

[†]Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotypes.

[‡]HRs, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted HRs, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.

carb: Carboplatin; CI: Confidence interval; cis: Cisplatin; doce: Docetaxel; HR: Hazard ratio; irino: Irinotecan; pem: Pemetrexed; XPD: Xeroderma pigmentosum complementation group D.

power to confirm the findings reported in previous studies [2,17]. When applied on risk allele frequency of 14% cases and 7% control subjects for CC genotype showing A751C polymorphism we would have 80% power to detect an odds ratio (OR) of 1.96, assuming the dominant model at p = 0.05 level of statistical significance. When applied on AC genotype for A751C polymorphism has 80% power to detect 1.03 OR with p = 0.05 level of significance.

• Statistical analysis

Chi-square test has been used to compare demographic features like sex, smoking status, whereas for continuous variables, in other words, age and pack-year smoked Student’s t-test was applied. Hardy–Weinberg equilibrium (HWE) was applied to differentiate between observed and expected genotypic frequencies of controls and cases applying goodness of χ^2 test. Allelic and genotypic frequencies of cases and controls were evaluated using Pearson’s χ^2 test. To estimate the association between XPD variants and lung cancer risk, OR and 95% confidence interval (CI) adjusted for age, sex and smoking status were calculated by applying logistic regression analysis. OS has been calculated for lung cancer patients from the first day of chemotherapeutic administration till the death or last follow-up. HWE for each polymorphism was calculated using Pearson’s chi-square test. Univariate Kaplan–Meier method was applied to estimate survival distribution and log-rank test for each SNP and its relation with lung cancer. Median survival time (MST) among censored

samples was calculated. The multivariate Cox proportional hazard model was applied to assess the independent effect of each polymorphism on OS when adjusted for age, gender, smoking status, stage, histology, Karnofsky performance score (KPS) and Eastern Cooperative Oncology Group (ECOG). Further, relationship between OS and XPD polymorphic variants when categorized as per chemotherapeutic regimen was evaluated using Kaplan–Meier method and Cox hazard analysis method. Tumor response was evaluated using response evaluation criteria in solid tumors for patients treated with greater than or equal to six chemotherapy cycles. Patients showing complete response or partial response were regarded as responders whereas those showing stable disease or progressive disease were considered as nonresponders. All the p-values were two-sided and p < 0.05 was referred to as statistically significant. All the statistical analyses were carried out using MedCalc Statistical Software version 15.11.4 (MedCalc Software bvba, Ostend, Belgium).

Furthermore, MDR which is a nonparametric genetic model method that is used to recognize interaction models resulting in characterization of gene–gene interactions was also applied. This model overcomes the limitations of logistic regression method. The multilocus genotypes are arranged in two groups depicting high and low risk, hence, reducing the genotype predictors from n dimensions to one. Cross-validation and permutation testing define the status of the disease. Of all the genotypic models generated, the combinations with high cross-validation consistency

(CVC) and testing accuracy are considered the best interaction models. Logistic regression analysis was used to calculate the combined effect of the variables. MDR analysis was carried out using MDR software package version 0.5.1 available online (www.epistasis.org). Finally, CART was

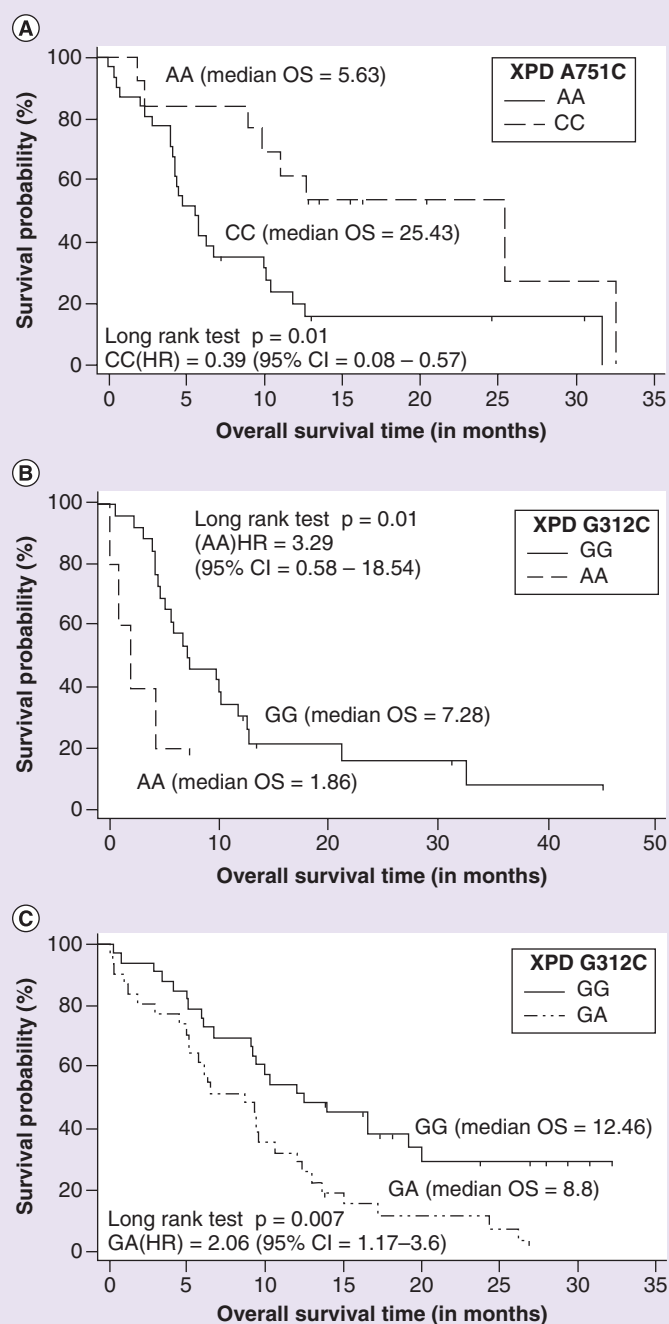


Figure 2. Kaplan Meier curves illustrating association between overall survival in lung cancer patients with: **(A)** mutant genotype (CC) treated with irinotecan cisplatin/irinotecan carboplatin regimen showing XPD Lys751Gln polymorphism; **(B)** mutant genotype (AA) treated with irinotecan cisplatin/irinotecan carboplatin regimen depicting XPD Asp312Asn polymorphism; **(C)** heterozygous genotype (GA) treated with pemetrexed cisplatin/pemetrexed carboplatin regimen showing XPD Asp312Asn polymorphism.

Table 6. Multifactor dimensionality reduction analysis showing interactions of xeroderma pigmentosum complementation group D variants with lung cancer risk.

Number of risk factors	Best interaction model	Cross-validation consistency	Prediction error	p-value
1	<i>XPD A751C</i>	8/10	0.48	0.003
2	<i>XPD A751C, XPD C156A</i>	10/10	0.44	0.0005
3	<i>XPD A751C, XPD G312A, XPD C156A</i>	10/10	0.45	<0.0001

XPD: Xeroderma pigmentosum complementation group D.

conducted to evaluate high-order SNP–SNP interactions and to further examine interactions among the selected polymorphism logistic regression analysis was applied using CART software (8.0, Salford Systems, CA, USA). CART is a data mining tool based on binary recursive partitioning method which partitions the data used for analysis into terminal nodes producing a decision tree used to detect high risk group. The most significant predictor splits the sample into sub-nodes until the difference becomes insignificant. The node with lowest case rate was considered as reference for evaluating risk for lung cancer by calculating OR and 95% CI.

Results

• Demographic characteristics of *XPD* polymorphism

Demographic characteristics of both the cases and controls have been illustrated in Table 1. The study consisted of 370 cases and 370 controls. The mean age for controls and cases were 53.83 ± 10.44 and 58.83 ± 10.18, respectively (p < 0.0001). Cases (81.99%) were reported as smokers and 18.10% as nonsmokers, whereas 73.24% controls were smokers and 26.75% were registered as nonsmokers. Furthermore when divided in terms of pack years, it was observed that a significant higher number of pack years was observed for cases as compared with controls (24.97 ± 35.04 vs 17.88 ± 19.45, p < 0.0001). When stratified according to histological subtypes, from the total number of cases: 135 (36.38%) were squamous cell carcinoma (SQCC), 120 (32.48%) were adenocarcinoma (ADCC) and 111 (30%) were small cell lung carcinoma (SCLC).

• *XPD* polymorphic variants & their association with lung cancer susceptibility

The allelic and genotypic distribution of the three *XPD* variants, in other words, *A751C*, *G312A* and *C156A* and their association toward lung cancer is presented in Table 2. The genotypic frequencies of *XPD A751C* ($\chi^2 = 0.38$, df = 1; p = 0.53) and *XPD G312A* ($\chi^2 = 2.21$, df = 1; p = 0.13) in control

groups were in agreement with HWE. The power analysis rate was evaluated for the present study and the genetic power of significance was estimated as 80%. We further assessed the association between the *XPD* variants and lung cancer risk as shown in Table 2. Adjusted ORs and 95% CIs were calculated using logistic linear regression analysis. Wild-type genotype was regarded as the reference group for all the three *XPD* SNPs. For *XPD A751C* polymorphism, the analysis showed a strong association between the mutant genotype (CC) and lung cancer susceptibility (OR: 1.96; 95% CI: 1.17–3.27; p = 0.01), whereas heterozygous genotype did not show any significant association (OR: 1.03; 95% CI: 0.75–1.43; p = 0.81). When segregated according to histological subtypes, elevated risk was reported for all the three histological subtypes: SQCC (OR: 1.85; 95% CI: 0.91–3.76; p = 0.008), ADCC (OR: 2.07; 95% CI: 1.05–4.08; p = 0.03) and SCLC (OR: 2.27; 95% CI: 1.12–4.63; p = 0.02). Furthermore for *C156A*, the heterozygous genotype (CA) showed a protective effect in subjects with SQCC (OR: 0.63; 95% CI: 0.40–0.99; p = 0.04) when compared with mutant genotype.

• Haplotype analysis of *XPD* variants

We analyzed haplotypes using SHEsis program platform (Supplementary Table 1). All the three SNPs were in weak linkage disequilibrium with each other in this study population: *XPD A751C* and *G312A* (D = 0.019, r² = 0.00); *XPD G312A*, *XPD C156A* (D = 0.035, r² = 0.01); *XPD A751C* and *C156A* (D = 0.065, r² = 0.004). There was no statistically significant difference in the overall haplotype distribution between cases and controls (global $\chi^2 = 9.73$; p = 0.20). All the exploratory haplotypes did not show any significant effect toward lung cancer susceptibility.

• Effect of smoking on the *XPD* polymorphism & their association with lung cancer risk

We further investigated whether the three *XPD* variants in the present study could modify the

association between exposure to smoking and lung cancer risk (Table 3). We observed that for *XPD A751C* polymorphism, lung cancer subjects who were smokers and having *CC* genotype (mutant) were strongly associated toward lung cancer susceptibility (OR: 2.11; 95% CI: 1.75–3.86; $p = 0.01$) as compared with the subjects carrying reference genotype (*AA*). On the contrary, no such association was observed for the

other two *XPD* polymorphisms. Furthermore, we carried out association studies between pack years smoked and *XPD* gene polymorphism as shown in Table 4. Individuals carrying mutant (*CC*) and heterozygous genotype (*AC*) when combined as a single genotype were associated with increased risk for lung cancer among heavy smokers for *A751C* polymorphism (OR: 1.77; 95% CI: 1.00–3.13; $p = 0.04$). Whereas an

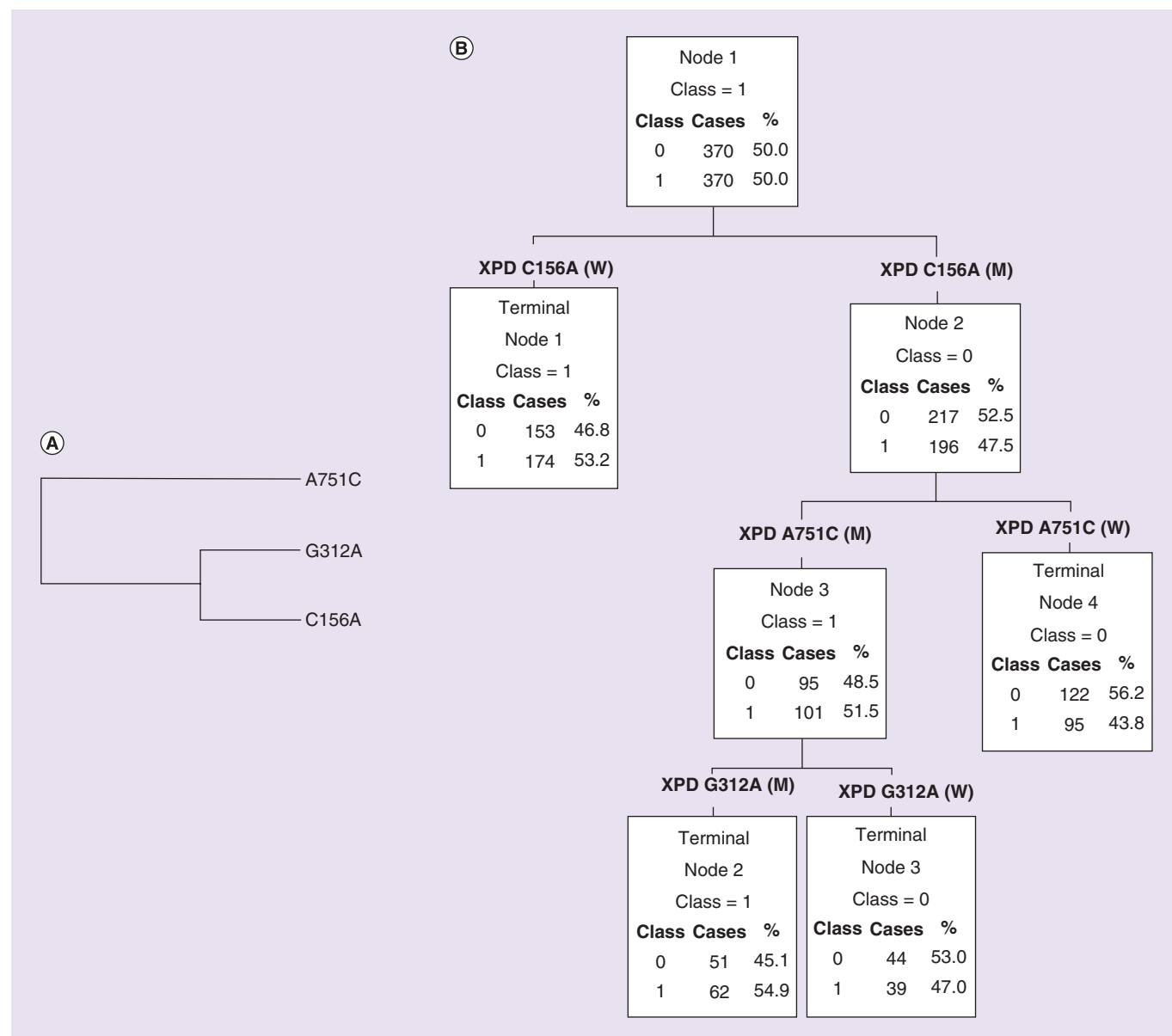


Figure 3. Interaction dendrogram gained from the multifactor dimensionality reduction (MDR) showing gene-gene interaction for *XPD Lys751Gln*, *XPD Asp312Asn* and *XPD Arg156Arg*. *XPD Asp312Asn* and *XPD Arg156Arg* show strong synergistic interaction whereas *XPD Lys751Gln* a weak interaction with other two polymorphism. **(B)** Classification and regression tree (CART) analysis depicting high order interaction among three polymorphisms (*Lys751Gln*, *Asp312Asn* and *Arg156Arg*) of *XPD*.

Table 7. Risk estimates based on Classification and Regression Tree analysis.

Terminal node	Genotype for each terminal node	Control (n)	Case (n)	Case rate [†]	OR (95% CI)	p-Value [‡]
Node 1	<i>XPD C156A(W)</i>	153	174	53.2	1.46 (1.03-2.06)	0.03
Node 2	<i>XPD G312A(M)/XPD A751C(M)/XPD C156A(M)</i>	62	51	54.9	1.56 (0.98-2.46)	0.05
Node 3	<i>XPD G312A(W)/XPD A751C(M)/XPD C156A(M)</i>	44	39	47.0	1.89 (0.68-1.89)	0.61
Node 4	<i>XPD A751C(W)/XPD C156A(M)</i>	122	95	43.8	1.0 [§]	

The bold values represent statistically significant data.
[†]case rate is percentage LC cases among all individuals in each terminal node (case/[case + control] x100).
[‡]p < 0.05 is considered statistically significant.
[§]Reference group.
OR: Odds ratio; XPD: Xeroderma pigmentosum complementation group D.

increased risk for the disease was reported for light smokers and all the genotypic variants (*AA*, *GA*, *GA + AA*) of *G312A* when compared with the reference genotype *GG* (*AA* genotype: OR: 2.45; 95% CI: 1.05–5.69; p = 0.03, *GA* genotype: OR: 1.65; 95% CI: 1.02–2.65; p = 0.03, *GA + AA* genotype: OR: 2.03; 95% CI: 1.28–3.23, p = 0.02).

• **Combinatorial effect of polymorphic variants of *XPD* on lung cancer risk**
To study SNP–SNP interaction, different genotypic combinations of *XPD* variants (*A751C*, *G312A* and *C156A*) were evaluated to predict their relation to lung cancer risk (Supplementary Table 2). However, no significant combinational association was observed between the *XPD* variants and lung cancer risk.

• **Relationship of *XPD* polymorphic variants with OS in lung cancer patients**
The survival analysis for 311 lung cancer patients and their association with *XPD* polymorphism has been depicted in Table 4. The cases with *XPD G312A* (N = 311; $\chi^2 = 1.86$; df = 1; p = 0.17) and *XPD C156A* (N = 311; $\chi^2 = 0.41$; df = 1; p = 0.52) polymorphism were in agreement with HWE. For *XPD A751C* polymorphism, the variant genotype (*CC*) was associated with better survival for lung cancer patients (9.23 vs 7.23 months). In case of *XPD G312A* polymorphism, it was observed that subjects carrying the mutant genotype (*AA*) had a shorter OS (MST = 5.43 months; HR: 1.30; log-rank test p = 0.26) as compared with subjects carrying the wild genotype (MST = 9.43 months). When stratified on the basis of histological subtypes, for *XPD A751C* polymorphism, it was observed that SCLC patients carrying mutant genotype (*CC*) illustrated a better survival when compared with the wild-type genotype (*AA*) (MST = 12.7 vs 5.80). In the Cox regression hazard model,

when adjusted for tumor stage, KPS, ECOG, smoking status, age, gender, we found hazard ratio (HR) was lower for SCLC subjects carrying homozygous variant genotype (*CC*) for *XPD A751C* polymorphism (HR: 0.49; 95% CI: 0.23–1.03; p = 0.05; MST: 12.7 months; Figure 1A), as compared with the wild genotype. On the contrary, for *XPD G312A* polymorphism a poor OS of 1.70 months was observed in SCLC subjects carrying the mutant genotype and thus showing a 13-fold HR for the disease which was found to be highly significant (HR: 13.38; 95% CI: 3.59–49.81; p = 0.0001; Figure 1B). However in case of SQCC, subjects with heterozygous genotype (*GA*) showed a low death ratio (HR: 1.60; 95% CI: 1.02–2.53; p = 0.04; MST = 5.56; Figure 1C) for the *G312A* polymorphism. *XPD C156A* polymorphism did not show any significant association with OS in lung cancer patients.

• **Association of *XPD* SNPs & OS of lung cancer patients treated with different chemotherapeutic regimens**
The survival of patients treated with different chemotherapeutic regimens and their association with *XPD* polymorphism has been depicted in Table 5. As shown in Figure 2A, for *XPD A751C* polymorphism, subjects carrying the mutant type genotype (*CC*) and treated with irinotecan cisplatin/irinotecan carboplatin regimen were found to be associated with better survival as compared with wild genotype (*AA*) (MST: 25.43 vs 5.63; HR: 3.29; 95% CI: 0.58–18.54; log-rank p = 0.01). After applying Cox regression model, a better prognosis with low death rate was observed for the above subjects, thus making it a predictive factor for disease treatment (HR: 0.21; 95% CI: 0.08–0.57; p = 0.002). Similarly, a significant low HR was also observed for the combined genotype (*AC + CC*) for irinotecan cisplatin/irinotecan carboplatin treatment regimen (HR: 0.48; 95%

CI: 0.21–0.92; $p = 0.02$, MST = 11 months). In *XPD G312A* polymorphism, individuals carrying mutant genotype (*AA*) and treated with irinotecan cisplatin/irinotecan carboplatin regimen showed high death rate, depicting poor survival (HR: 16.46; OR: 2.50–108.40; $p = 0.003$; MST: 1.86 months, **Figure 2B**) when compared with wild genotype (MST: 7.28 months). Furthermore, lung cancer patients treated with pemetrexed cisplatin/pemetrexed carboplatin chemotherapeutic agent and harboring single allele (*GA*) for *XPD G312A* polymorphism had a poor OS (MST = 8.8, log-rank test $p = 0.007$; **Figure 2C**) as compared with subjects with wild-type (*AA*) genotype (MST = 12.46). On applying Cox regression model, a twofold increased death rate was reported for heterozygous type genotype when compared with the wild-type genotype (HR: 2.01, 95% CI: 1.14–3.56; $p = 0.01$). No significant association was observed for *XPD C156A* and OS in lung cancer patients treated with any chemotherapeutic regimen.

• Association of *XPD* polymorphism & clinical outcomes

XPD gene polymorphism and its association with clinical outcomes were evaluated using multivariate logistic regression (**Supplementary Table 3**). The 202 patients were treated with platinum-based drug given in combination with taxol-based drug for greater than or equal to six cycles were selected. There was no significant association observed between *XPD* polymorphism and response rate. Furthermore, clinical response in lung cancer patients was evaluated when stratified according to their histological subtypes using multivariate logistic regression (**Supplementary Table 4**). The SQCC patients with mutant genotype (*CC*) and showing *A751C* polymorphism depicted a significant response (OR: 0.16; 95% CI: 0.02–0.87; $p = 0.03$).

• MDR analysis

The MDR analysis was performed on a dataset that consisted of subjects with or without lung cancer depicting average CVC model and average prediction error (**Table 6**). The analysis resulted in best interaction model which includes *XPD A751C*, *G312A*, *C156A* with CVC of 10/10 and prediction error of 0.45 ($p < 0.0001$). A higher CVC and a minimum prediction error are preferred for the best results. The two factor model (*XPD A751C*, *C156A*) for evaluating lung cancer risk had CVC of 10/10 and showing improvised

prediction error of 0.44 ($p = 0.0005$). *XPD Lys⁷⁵¹Gln* appeared as the best one factor model showing lung cancer risk with CVC of 8/10 and prediction error of 0.48 ($p = 0.003$). The interaction dendrogram was used to demonstrate the visualized interaction between the *XPD* SNPs (*XPD A751C*, *G312A* and *C156A*) and lung cancer risk (**Figure 3A**). More synergy was observed between *XPD G312A* and *C156A* as the shortest line had been connecting them whereas a weak interaction was observed between *XPD A751C* and other two polymorphic variants.

• CART analysis

High-order SNP–SNP interactions with lung cancer risk were analyzed using CART analysis (**Table 7**). Final tree consisting of four terminal nodes was generated (**Figure 3B**). The root split was present on *XPD C156A* suggesting it to be the strongest factor for lung cancer risk among other variants. Further subsequent split was on *XPD A751C* and *G312A*. To understand the potential mechanism for increased lung cancer risk with SNPs in CART analysis, we also applied logistic regression analysis in the final model selected by CART evaluating OR and 95% CI. The terminal node with minimum case rate was considered as the reference, in other words, node 4 showing case rate of 43.8%. Individuals carrying wild genotype of *XPD A751C* and mutant (heterozygous + mutant) genotype of *XPD C156A* represent node 4. Subjects with *XPD G312A* (*M*), *A751C* (*M*) and *C156A* (*M*) showed highest risk toward the lung cancer (OR: 1.56; 95% CI: 0.98–2.46; $p = 0.05$). Whereas individuals with wild genotype of *C156A* possessed a significantly increased risk for the disease (OR: 1.46; 95% CI: 1.03–2.06, $p = 0.03$). Node 3 carrying individuals with a combined genotype of *XPD G312A* (*W*), *A751C* (*M*) and *C156A* (*M*) showed almost twofold increased risk for the lung cancer (OR: 1.89; 95% CI: 0.68–1.89; $p = 0.61$).

Discussion

In this study, we assessed the role of NER gene *XPD* toward lung cancer susceptibility, defining its association with OS and clinical outcomes in lung cancer patients undergoing platinum-based doublet chemotherapy. Overall, our study reveals no significant risk toward lung cancer for both *XPD G312A* and *XPD C156A* genotypes, however on the contrary, the *XPD A751C* mutant genotype (*CC*) was found to be strongly associated toward risk for lung cancer

($p = 0.01$). Our data is consistent with the study conducted by Xing *et al.* [18], Sturgis *et al.* [8] and Zhou *et al.* [19], where the authors also have suggested a positive association of variant C allele toward lung cancer risk. A study conducted on Danish [20], Chinese [21], Finish [22] and Norwegian [23] populations have further fortified our data, where the results also documented a positive association of mutant genotype (CC) toward risk for lung cancer. Furthermore, Vogel *et al.* [10] and Qiao *et al.* [24] found an association of the mutant *Gln/Gln* (CC) genotype with lower DNA repair capacity. This SNP present at 751 amino acid of the *XPB* gene is crucially important in terms of *XPB* activity as it is located in interactive domain, in other words, its helicase activator p44 is present inside basal transcription factor IIH complex [25]. The *XPB 751Gln/Gln* genotype was also found to be associated with a higher frequency of chromatid breaks as compared with wild genotype and therefore, related to high risk for carcinogenesis [26]. Furthermore, our results also revealed that when segregated on the basis of histological subtypes the *Lys⁷⁵¹Gln* polymorphism was found to have a pronounced effect toward susceptibility for all the three subtypes of lung cancer, in other words, ADCC ($p = 0.03$), SQCC ($p = 0.008$) and SCLC ($p = 0.02$). Similar findings have also been reported in Chinese [2] and Spanish [27] populations suggesting a risk toward ADCC subjects for the variant genotype of *A751C* polymorphism.

The present study reports a lack of association for the *XPB G312A* polymorphism toward risk for lung cancer. This is in line to that reported by Lopez-Cima *et al.* [27] and Yu *et al.* [28]. On the contrary, some epidemiological studies have supported an association for the *G312A* polymorphism toward lung cancer [8,18,23]. When stratified on basis of histology the *G312A* polymorphic variant did not show any significant association with any of the histological subtypes of lung cancer. Hou *et al.* [29] and Qiao *et al.* [24] reported that the two common variant alleles (located on codons 751 and 312) of the *XPB* gene might lead to reduced repair of DNA adducts, and promotion of tumor formation. *XPB G312A* is present in highly conserved seven motif DNA helicase domain of Rec Q family [10]. The *XPB G312A* variants alter the folding of the target protein which may affect the sites as well as function of the downstream sites of protein interaction [10].

In the present study, no significant association was observed between *C156A* polymorphism and lung cancer risk in North Indians. However, *XPB C156A* heterozygous genotype (CA) was found to have a strong protective effect in SQCC patients ($p = 0.04$). Very few studies have been reported on this polymorphism. Our results show similarities with the studies conducted by Yin *et al.* [2], Strugis *et al.* [8] and Catana *et al.* [30] whereas studies done on Danish [31], Caucasian [31] and American populations [32] depicted a significant risk for basal cell carcinoma and gliomas. *XPB C156A* polymorphism does not show any amino acid change and hence, the enzymatic function of this protein is unlikely to be affected by the mutation, this polymorphic substitution would rather alter the rate of translation by changing codon usage or by reducing *XPB* protein level of the gene leading to mRNA instability [31]. Catana *et al.* [30] reported a significant risk for ADCC subtype with the *XPB C156A* variant genotype and no association was observed for SQCC patients.

To evaluate high-order SNP–SNP interactions, MDR and CART analysis were performed suggesting the existence of a significant interaction for the three SNPs. The interaction dendrogram showed a significantly strong synergistic interaction between *XPB G312A* and *C156A* and were validated by logistic regression analysis present in MDR analysis. *XPB A751C*, *G312A* and *C156A* assessed as the best three factor risk model while *A751C* emerged as the highest risk factor. Our results were concordant with the study conducted by Mei *et al.* [33] who depicted close association of *A751C* polymorphism with lung cancer risk using MDR analysis. CART analysis was also applied on our data showing first split at *C156A* followed by next split at *G312A* and *A751C*. CART analysis further distributes risk group present on different terminal nodes and risk is calculated for every node which help us to evaluate risk factor as mentioned in the previous studies conducted on NER genes [34].

We further evaluated association of *XPB* polymorphism with OS in lung cancer patients using survival analysis. In the present study, no significant association was observed between *XPB A751C* genotypes and OS. Our results were concordant with the study conducted on Asian [35,36] and Spanish [36] subjects suggesting *A751C* polymorphism might not be a genetic determinant for survival in lung cancer patients

treated with platinum-based chemotherapy. When segregated on the basis of histology, a protective effect was observed for individuals with SCLC carrying mutant (*CC*) genotype (hazard ratio [HR]: 0.49; $p = 0.04$) for *XPB A751C* polymorphism. Whereas no significant results were observed between OS of patients with other histological subtypes (i.e. ADCC, SQCC) and *XPB A751C* polymorphism. Our results were supported by Li *et al.* [37], Ludovini *et al.* [38] and Mathiaux *et al.* [39] who also did not observe any association between OS and *XPB A751C* polymorphism in non-small cell lung cancer (NSCLC) patients of Caucasian ethnicity. Yao *et al.* [40] and Zhang *et al.* [41] further validated that there was no association between *A751C* polymorphism and OS in Asian subjects diagnosed with NSCLC. In the present study, there was no significant relationship observed between OS in lung cancer patients and *XPB G312A* polymorphism. Studies by Li *et al.* [42] and Wu *et al.* [43] were consistent with our study suggesting no significant relationship between *G312A* polymorphism and OS in lung cancer patients of Asian and Caucasian ethnicity. When differentiated on the basis of histological subtypes, for *G312A* polymorphism, SCLC patients with *AA* genotype (mutant) were associated with poor survival of 1.7 months and with a high death rate among diseased subjects (HR: 13.38; $p = 0.0001$).

We also evaluated the association of *XPB* polymorphic variants and OS according to chemotherapeutic regimens. Most cancer patients with specific type of cancer and stage received almost the same treatment. It has been observed that the treatment works well only in few patients. The knowledge of genetic differences in patients and their tumors could explain this varied response in patients toward the drug. A targeted treatment will target cancer specific genes and proteins that might result in better understanding of patient's response toward a specific drug. Therefore, the patients treated with irinotecan-cisplatin/irinotecan-carboplatin regimen demonstrated a trend of better response (25.2 months) with extremely low HR (HR: 0.21; $p = 0.002$) for patients carrying mutant genotype (*CC*) for *A751C* polymorphism when compared with subjects carrying wild genotype receiving same treatment. For *XPB G312A* polymorphism, the subjects carrying mutant genotype (*AA*) were exhibiting a poor survival (1.86 months) with high HR (HR: 16.46; $p = 0.003$) when treated

with irinotecan cisplatin/irinotecan carboplatin regimen. These results might play a supportive role in personalized chemotherapy of cancer patients. The prior knowledge about individual's genetic makeup and growth of its tumor will be helpful in interpreting its prevention, screening and treatment strategies. These analyses suggest the possible relationship between *XPB* polymorphism and specific chemotherapeutic response. Irinotecan is a camptothecin derived compound whose SN-38 active metabolite interacts with the topoisomerase I via stabilizing the complex formed with DNA [45]. Due to this novel cytotoxic mechanism, irinotecan differs from other chemotherapeutic agents [44]. Irinotecan has always demonstrated efficacy in the treatment of small cell lung cancer. It also appears to show promising activity in NSCLC patients producing a response rate upto 32% [44]. When administered in combination with cisplatin/carboplatin drug, a response rate of 25–56% was detected in patients of American ethnicity who received treatment in stages II and III of the disease and thus showing median survival time of 9–12 months [45]. A study in Japanese population reported a response rate of 36% for advanced NSCLC patients treated with irinotecan cisplatin/irinotecan carboplatin regimen [44]. DNA repair gene like *ERCC2* has been majorly associated with response in other platinum-based chemotherapeutic agents [46]. Therefore, no previous study has been reported on association of NER gene *XPB* with OS for lung cancer patients treated with irinotecan cisplatin/irinotecan carboplatin regimen except a single study on colorectal carcinoma. Artac *et al.* [47] emphasized on association of *Gln/Gln* genotype with a high risk of mortality in colorectal carcinoma patients who were treated with the above regimen and hence it does not support the findings in the present study. Lung cancer patients carrying heterozygous genotype (*GA*) for the *G312A* polymorphism and treated with pemetrexed cisplatin/pemetrexed carboplatin regimen showed a median survival (8.8 months) with significant twofold HR ($p = 0.01$). Pemetrexed is a multitargeted antifolate drug which has been approved as first-line treatment for patients with NSCLC in combination with cisplatin/carboplatin and work as a single agent for relapsed chemotherapy or chemotherapy refractory NSCLC after platinum containing chemotherapy [48]. Pemetrexed undergoes intracellular activation which is essential for its antiproliferative activity. This regimen is

reported to be associated with effective treatment of patients diagnosed with NSCLC [48].

The present study has few limitations that might influence the study. The maximum patients reported to PGIMER were diagnosed in advance stages of tumor. The patients who used to visit PGIMER for treatment were majorly from rural parts of Northern India. These patients who resided in rural areas did not report

to the hospital regularly as it was difficult for them to commute from those areas on a regular basis. Whereas some patients discontinued the treatment from the hospital before or after completing six chemotherapy cycles. So, due to lack of clinical data, it was difficult to conduct progression-free survival studies on the present study population.

SUMMARY POINTS

Association of *XPB* gene (*A751C*, *G312A*, *C312A*) with lung cancer susceptibility & its histological subtypes

- The mutant genotype (CC) was associated with an increased lung cancer risk ($p = 0.01$) and also with its all histological subtypes, in other words, adenocarcinoma ($p = 0.03$), squamous cell carcinoma ($p = 0.008$) and small cell lung carcinoma (SCLC) ($p = 0.02$) for xeroderma pigmentosum complementation group D (*XPB*) *A751C* polymorphism.

Interactive effect of smoking on *XPB* variants & its association with lung cancer

- Smokers carrying the *A751C* variant allele were associated with twofold increased risk of lung cancer ($p = 0.01$). In case of *G312A* polymorphism, we found that light smokers (pack years ≤ 25) carrying the mutant genotype (AA) were found to be strongly associated toward lung cancer susceptibility ($p = 0.03$).

Multifactor dimensionality reduction & Classification and Regression Tree analysis was conducted to investigate high-order SNP–SNP interaction

- *XPB A751C*, *XPB G312A*, *XPB C156A* were analyzed as the best one factor model using multifactor dimensionality reduction analysis and were associated with increased risk for lung cancer ($p \leq 0.0001$). Classification and Regression Tree analysis depicted *C156A* as the highest risk group associated toward lung cancer susceptibility.

The role of three *XPB* SNPs toward overall survival of patients treated with platinum-based chemotherapy was analyzed along with clinical response

- *XPB* variants did not show any association with overall survival in lung cancer patients.
- When stratified on the basis of histological subtypes, for *XPB A751C* polymorphism, it was observed that SCLC patients carrying mutant genotype (CC) illustrated a better survival when compared with the wild-type genotype (AA) (median survival time = 12.7 vs 5.80). In the Cox regression hazard model when adjusted for tumor stage, KPS, ECOG, smoking status, age, gender, we found hazard ratio (HR) was lower for SCLC subjects carrying homozygous variant genotype (CC) for *XPB A751C* polymorphism (HR: 0.49; 95% CI: 0.23–1.03; $p = 0.05$; median survival time: 12.7 months; [Figure 1A](#)), as compared with wild genotype.
- No significant association was observed between *XPB* polymorphism and clinical response in lung cancer patients.

Relationship between overall survival & *XPB* polymorphic variants when categorized as per chemotherapeutic regimen was evaluated using Kaplan–Meier method & Cox hazard analysis method

- In a subgroup analysis based on chemotherapeutic regimens, it was observed that patients administered irinotecan-platinum, the *XPB A751C* mutant genotype (CC) was significantly related to better OS as compared with wild-type genotype (AA) (25.43 vs 5.63, log rank $p = 0.01$). The Cox model revealed a significant effect (HR = 0.21, 95% CI = 0.08–0.57; $p = 0.002$), whereas a poor survival was reported for subjects who were carrying the mutant genotype (AA) for *XPB G312A* polymorphism and treated with same regimen as above in comparison with wild-type genotype (GG) (1.86 vs 7.28, log rank $p = 0.01$).
- Patients treated with pemetrexed-platinum doublets and carrying the heterozygous genotype of *G312A* polymorphism had a poor survival (HR = 2.01, 95% CI = 1.14–3.56; $p = 0.01$).

Conclusion

- *XPB* SNPs *G312A* (SCLC), *A751C*, *G312A* (irinotecan–platinum doublets group) and *XPB G312A* (pemetrexed-platinum doublets group) were associated with OS and might act as a predictive marker in lung cancer patients treated with platinum-based doublet chemotherapy.

Conclusion

In conclusion, the findings of our study indicated association of *XPD A751C* with increased risk of lung cancer whereas no significant association was observed for *XPD A312C* and *XPD C156A*. MDR analysis also revealed association of *A751C* polymorphism with increased risk of lung cancer. *ERCC2* functional SNPs (*A156C* and *G312A*) independently affect the OS of patients treated with irinotecan cisplatin/irinotecan carboplatin regimen and may act as an important predictive parameter in survival of lung cancer patients.

Future perspective

This study has provided insight into the association of *XPD* polymorphism with lung cancer susceptibility, OS and clinical response in North Indian patients which were not explored before the present study. Some remarkable results were observed when lung cancer patients treated with different chemotherapy regimens, in other words, irinotecan-cisplatin/irinotecan-carboplatin regimen, pemetrexed-cisplatin/pemetrexed-carboplatin regimen and their association with *XPD* polymorphism was analyzed. It is known that selection of chemotherapeutic regimen could improve response and hence OS of cancer

patients. Personalized chemotherapy as per the molecular marker could help to improve the response rate as well as survival of lung cancer patients. This study in future can facilitate assessment of genetic variations of *XPD/ERCC2* and help patients to make therapeutic decisions for individualized therapy in advanced lung cancer patients.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/fon-2017-0211

Financial & competing interests disclosure

This work was supported by grant from the Indian Council of Medical Research, New Delhi, India (5/13/126/2011/ NCD-III). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Acknowledgements

We would like to express our gratitude to all the subjects who participated in the current study.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

- Winsey S, Haldar NA, Marsh HP, Bunce M, Marshall SE, Harris AL. A variant within the DNA repair gene XRCC3 is associated with the development of melanoma. *Cancer Res.* 60, 5612–5616 (2010).
- Yin J, Li J, Ma Y, Guo L, Wang H, Vogel U. The DNA repair gene ERCC2/XPB polymorphism Arg156Arg (A22541C) and the risk of lung cancer in a Chinese population. *Cancer Lett.* 223, 219–226 (2005).
- Shows Arg¹⁵⁶Arg polymorphism as well as depicts a detailed study about the relationship between Arg¹⁵⁶Arg polymorphism and lung cancer risk. The primers for Arg156Arg polymorphism are taken from this study.
- Chen Z. Medical genetics. *People Hygienic Press Beijing* 1, 262–263 (2001).
- Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair gene and associations with cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 11, 1513–1530 (2002).
- Hoeijmakers JH, Egly JM, Vermeulen W. TFIIH: a key component in multiple DNA transactions. *Curr. Opin. Genet. Dev.* 26, 26–33 (1996).
- Benhamou S, Sarasain A. ERCC2/SPB gene polymorphism and cancer risk. *Mutagenesis* 6, 463–469 (2002).
- Shen MR, Jones IM, Mohrenweiser H. Non conservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res.* 58, 604–8 (1998).
- Sturgis EM, Zheng R, Castillo EJ *et al.* XPB/ERCC2 polymorphisms and risk of head and neck cancer: a case-control analysis. *Carcinogenesis* 21, 2219–2223 (1998).
- Tomescu D, Kavanagh G, Ha T, Campbell H, Melton DW. Nucleotide excision repair gene XPB polymorphisms and genetic predisposition to melanoma. *Carcinogenesis* 22, 403–408 (2001).
- Vogel U, Hedayati M, Dybdahl M, Grossman L, Nexø BA. Polymorphisms of the DNA repair gene XPB: correlation with risk of basal cell carcinoma revisited. *Carcinogenesis* 22, 899–904 (2001).
- Butkiewicz D, Rusin M, Enewold L, Shields PG, Chorazy MM, Harris CC. Genetic polymorphisms in DNA repair genes and risk of lung cancer. *Carcinogenesis* 22, 593–597 (2001).
- Matullo G, Guarrera S, Carturan S *et al.* DNA repair gene polymorphisms, bulky DNA adducts in white blood cell and bladder cancer in a case-control study. *Int. J. Cancer* 92, 562–567 (2001).
- Non-Small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised trials. *BMJ* 311, 899–909 (1995).
- Johnson SW, Stevenson JP, O'Dwyer PJ, DeVita VT, Hellman S, Rosenberg SA. Cisplatin and its analogues. In: *Cancer: Principles and Practice of Oncology (6th Edition)*. DeVita VT, Hellman S, Rosenberg SA (Eds). Lipincott Williams & Wilkins, PA, USA 376–388 (2001).

- 15 Bosken CH, Wei Q, Amos CI *et al.* An analysis of DNA repair as a determinant of survival in patients with non-small-cell lung cancer. *J. Natl Cancer Inst.* 94, 1091–1099 (2001).
- 16 Sodhi KK, Bahl C, Singh N, Behera D, Sharma S. Functional genetic variants in pre-miR-146a and 196a2 genes are associated with risk of lung cancer in North Indians. *Future Oncol.* 11(15), 2159–2173 (2015).
- 17 Li J, Hassan MM, Bondy ML, Abbruzzese JL, Evans DB, Li D. The XPD Asp312Asn and Lys751Gln polymorphism, corresponding haplotype and pancreatic cancer risk. *Cancer Lett.* 245, 61–68 (2007).
- From this study, the primers for the current study have been selected for *Asp³¹²Asn* and *Lys⁷⁵¹Gln* polymorphism. A detailed mode of mechanism regarding xeroderma pigmentosum complementation group D (XPD) Asp312Asn and XPD Lys751Gln is mentioned in this study.
- 18 Xing D, Tan W, Wei Q, Lin D. Polymorphisms of the DNA repair gene *XPB* and risk of lung cancer in a Chinese population. *Lung Cancer* 38, 123–9 (2002).
- 19 Zhou W, Liu G, Miller DP *et al.* Gene–environment interaction for the ERCC2 polymorphisms and cumulative cigarette smoking exposure in lung cancer. *Cancer Res.* 62, 1377–1381 (2002).
- 20 Vogel U, Nexø BA, Tjønneland A, Wallin H, Hertel O, Raaschou-Nielsen O. *ERCC1*, *XPB* and *RAI* mRNA levels in lymphocytes are not associated with lung cancer risk in a prospective study of Danes. *Mutat. Res.* 593, 88–96 (2006).
- 21 Liang G, Xing D, Miao X, Tan W, Yu C, Lu W, Lin D. Sequence variations in the DNA repair gene *XPB* and risk of lung cancer in a Chinese population. *Int. J. Cancer* 105, 669–673 (2003).
- 22 Misra RR, Ratnasinghe D, Tangrea JA *et al.* Polymorphisms in the DNA repair genes *XPB*, *XRCC1*, *XRCC3*, and *APE/ref-1*, and the risk of lung cancer among male smoking in Finland. *Cancer Lett.* 191, 171–178 (2003).
- 23 Zienoldiny S, Campa D, Lind H *et al.* Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis* 27(3), 560–567 (2006).
- 24 Qiao Y, Spitz MR, Shen H *et al.* Modulation of repair of ultraviolet damage in the host-cell reactivation assay by polymorphic *XPB* and *XPB/ERCC2* genotypes. *Carcinogenesis* 23, 2959 (2002).
- 25 Seroz T, Perez C, Bergmann E, Bradsher J, Egly JM. The regulatory subunit of the *XPB*/RAD3 helicase, plays a crucial role in transcriptional activity of TFIIH. *J. Biol. Chem.* 275, 33260–33266 (2000).
- 26 Spitz MR, Wu X, Wang Y *et al.* Modulation of nucleotide excision capacity by *XPB* polymorphisms in lung cancer patients. *Cancer Res.* 61, 1354–7 (2001).
- Explains the mechanism of action of *XPB* genes. It also gives detailed knowledge of interactive effect of smoking with *XPB* genes and its association with lung cancer susceptibility. The study also defines association of *XPB* polymorphisms and their association with lung cancer histological subtypes.
- 27 López-Cima MF, González-Arriaga P, García-Castro L *et al.* Polymorphisms in *XPB*, *XPB*, *XRCC1*, and *XRCC3* DNA repair genes and lung cancer risk in a population of Northern Spain. *BMC Cancer* 7, 162 (2007).
- 28 Yu HP, Wang XL, Sun X *et al.* Polymorphism in the DNA repair gene *XPB* and susceptibility to esophageal squamous cell carcinoma. *Cancer Genet. Cytogenet.* 154, 10–15 (2004).
- 29 Hou SM, Falt S, Angelini S *et al.* The DNA variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. *Carcinogenesis* 23, 599–603 (2002).
- 30 Catana A, Popp AR, Pop M, Porojan MD, Petrisor FM, Pop IV. Genetic polymorphism of DNA repair gene *ERCC2/XPB* (Arg156Arg) (A22541C) and lung cancer risk in Northern Romania. *Revista Romana de Medicina de Laborator* 20, 2–4 (2012).
- 31 Dybdahl M, Vogel U, Frentz G, Wallin H, Nexø BA. Polymorphisms in the DNA repair gene *XPB*: correlations with risk and age at onset of basal cell carcinoma. *Cancer Epidemiol. Biomark. Prev.* 8, 77–81 (1999).
- 32 Caggana M, Kilgallen J, Conroy JM *et al.* Associations between *ercc2* polymorphisms and gliomas. *Cancer Epidemiol. Biomark. Prev.* 10, 355–360 (2001).
- 33 Mei C, Hou M, Guo S *et al.* Polymorphisms in DNA repair genes of *XRCC1*, *XPB*, *XPB* and associations with lung cancer risk in Chinese people. *Thoracic Cancer* 5, 232–242 (2014).
- This is the one of the few studies conducted on high order gene interactions using multifactor dimensionality reduction and Classification and Regression Tree analysis and is the first study to explain the association of *XPB* genes with lung cancer risk using these data mining tools.
- 34 Chen M, Kamat AM, Huang M *et al.* High-order interactions among genetic polymorphisms in nucleotide excision repair pathway genes and smoking in modulating bladder cancer risk. *Carcinogenesis* 28(10), 2160–2165 (2007).
- 35 Provencio M, Camps C, Cobo M *et al.* Prospective assessment of *XRCC3*, *XPB* and Aurora kinase A single-nucleotide polymorphisms in advanced lung cancer. *Cancer Chemother. Pharmacol.* 70, 883–890 (2012).
- 36 Liu L, Yuan P, Wu C *et al.* Assessment of *XPB* Lys751Gln and *XRCC1* T-77C polymorphisms in advanced non-small-cell lung cancer patients treated with platinum-based chemotherapy. *Lung Cancer* 73, 110–115 (2011).
- 37 Li XD, Han JC, Zhang YH, Li HB, Wu XY. Common variations of DNA repair genes are associated with response to platinum-based chemotherapy in NSCLCs. *Asian Pac. J. Cancer Prev.* 14, 145–148 (2013).
- Explains the association between *XPB* polymorphism and OS in patients treated with platinum-based chemotherapy.
- 38 Ludovini V, Floriani I, Pistola L *et al.* Association of cytidine deaminase and xeroderma pigmentosum group D polymorphisms with response, toxicity, and survival in cisplatin/gemcitabine-treated advanced non-small cell lung cancer patients. *J. Thorac. Oncol.* 6, 2018–2026 (2011).
- 39 Mathiaux J, Le Morvan V, Pulido M *et al.* Role of DNA repair gene polymorphisms in the efficiency of platinum-based adjuvant chemotherapy for non-small cell lung cancer. *Mol. Diagn. Ther.* 15, 159–166 (2011).
- 40 Yao CY, Huang XE, Li C *et al.* Lack of influence of *XRCC1* and *XPB* gene polymorphisms on outcome of platinum-based chemotherapy for advanced non small cell lung cancers. *Asian Pac. J. Cancer Prev.* 10, 859–864 (2009).
- 41 Zhang ZY, Tian X, Wu R, Liang Y, Jin XY. Predictive role of *ERCC1* and *XPB* genetic polymorphisms in survival of Chinese non-small cell lung cancer patients receiving chemotherapy. *Asian Pac. J. Cancer Prev.* 13, 2583–2586 (2012).
- 42 Li F, Sun X, Sun N *et al.* Association between polymorphisms of *ERCC1* and *XPB* and clinical response to platinum-based chemotherapy in advanced non-small cell lung cancer. *Am. J. Clin. Oncol.* 33, 489–494 (2010).
- 43 Wu W, Li H, Wang H *et al.* Effect of polymorphisms in *XPB* on clinical outcomes of platinum-based chemotherapy for Chinese

- non-small cell lung cancer patients. *PLoS ONE* 7, e33200 (2012).
- 44 Tsuruo T, Matsuzaki T, Matsushita M *et al.* Antitumor effect of CPT-11, a new derivative of camptothecin, against pleiotropic drug resistant tumors *in vitro* and *in vivo*. *Cancer Chemother. Pharmacol.* 21, 71–74 (1988).
- 45 Kinoshita A, Fukuda M, Soda H *et al.* Phase II study of irinotecan combined with carboplatin in previously untreated non-small-cell lung cancer. *Br. J. Cancer* 94(9), 1267–1271 (2004).
- **First study which depicts association between XPD polymorphism and OS in patients treated with irinotecan–cisplatin/irinotecan–carboplatin regimen. It gives detailed description of mechanism of irinotecan with DNA repair genes and also explains its impact on lung cancer patients when infused in combination with the platinum-based drug.**
- 46 Pillot GA, Read WL, Hennenfent KL *et al.* A Phase II study of irinotecan and carboplatin in advanced non-small cell lung cancer with pharmacogenomic analysis: final report. *J. Thorac. Oncol.* 1(9), 2006.
- 47 Artac M, Bozuk H, Pehlival S *et al.* The value of XPD and XRCC1 genotypes polymorphism to predict clinical outcome in metastatic colorectal carcinoma patients with irinotecan based regime. *J. Cancer Res. Clin. Oncol.* 136(6), 803–809 (2010).
- 48 Metro M, Chiari R, Mare M *et al.* Carboplatin plus pemetrexed for platinum-pretreated, advanced non-small cell lung cancer: a retrospective study with pharmacogenetic evaluation. *Cancer Chemother. Pharmacol.* 68(6), 1405–1412 (2011).
- **Important study illustrating role of pemetrexed-cisplatin/pemetrexed-carboplatin regimen in cancer patients and their impact on their survival. It also depicts mechanism of action of pemetrexed in DNA repair.**

XPF polymorphism toward lung cancer susceptibility and survival in patients treated with platinum-based chemotherapy

Shweta Lawania¹, Siddharth Sharma^{*1}, Navneet Singh² & Digamber Behera²

¹Department of Biotechnology, Thapar University, Punjab 147002, India

²Department of Pulmonary Medicine, Postgraduate Institute of Medical Education & Research (PGIMER), Sector 14, Chandigarh, India

*Author for correspondence: siddharthsharma.phd@thapar.edu

Aim: To evaluate the association of three *XPF* polymorphic variants (673 C>T, 11985 A>G, G415A) with lung cancer, overall survival and clinical response in North Indians. **Methods:** Genotyping was performed using PCR-restriction fragment length polymorphism. **Results:** A total of 673 C>T polymorphism was associated with 1.5-fold increased lung cancer risk for heterozygous genotype (CT; $p = 0.03$). Adenocarcinoma patients with 673 C>T polymorphism carrying heterozygous genotype (CT) had a lower hazard ratio ($p = 0.01$). Classification and regression tree analysis predicted *XPF* 673 C>T (M) as the strongest risk factor for the lung cancer ($p = 0.003$). For 11985 A>G polymorphism, lung cancer subjects treated with irinotecan cisplatin/carboplatin regimen having heterozygous genotype (AG) was associated with high mortality risk ($p = 0.0001$). **Conclusion:** 673 C>T polymorphism was associated with increased lung cancer risk.

First draft submitted: 25 October 2017; Accepted for publication: 4 January 2018; Published online: 9 May 2018

Keywords: lung cancer • overall survival • platinum-based chemotherapy • xeroderma pigmentosum group F

Lung cancer is the most common cancer as it accounts for 19% of cancer-related deaths worldwide and the overall 5-year survival rate of lung cancer patients in advanced stages is only 5% in developing countries, while it is 15% in developed countries [1]. It has been predicted that the increase in lung cancer related deaths might be partly attributable to smoking habits [2]. But few studies have claimed an increase in the incidence rate of lung cancer in females as well as non-smokers [2].

Eukaryotic cells undergo constant mutagenic assault from endogenous and exogenous agents. There are multiple environmental factors like diet, infection, lifestyle factors which majorly includes tobacco smoking that leads to formation of oxidized bases, bulky DNA adducts, strand breaks that causes DNA damage [3]. Human system has evolved various repair pathways which restores the damaged DNA and its genomic integrity. Nucleotide excision repair (NER) pathway is the most well-studied pathway which is responsible for repairing bulky DNA lesions. Its general mechanism consists of five steps: recognition of DNA damage (which is performed by XPC-hHR23B complex in global genome NER pathway); unwinding of DNA by the TFIIH complex which also includes XPD and verification of lesion at both ends; double incision at both the ends of the lesion by the XPG and ERCC1–XPF complex excision of the nucleotide and resynthesis of DNA fragment by DNA polymerase; and ligation of residual nick by DNA ligase [4].

In the NER system one of the important gene is *XPF* which is also known as *ERCC4*. It is located on chromosome 16p13.2 and consists of 11 exons which span approximately 28.2 kb [5]. The main function of *XPF* gene is to make 5'-incision during NER process. *XPF* protein is composed of 916 amino acids and an ERCC4 domain which belong to nuclease family [6]. ERCC1 binds to form a complex with *XPF* which acts like a structure-specific endonuclease responsible for 5' primer incision during DNA excision repair [7]. This complex also plays an important role in DNA interstrand crosslink as well as double strand breaks removal [7].

There are 580 single nucleotide polymorphisms (SNPs) in the *XPF* gene reported till date but only few have shown susceptibility to cancers like breast, endometrium and colorectal. Three polymorphisms are being studied in this study, in other words, 673 C>T, 11985 A>G and Arg⁴¹⁵ Gln. Arg⁴¹⁵ Gln (rs1800067) results in an arginine-to-glutamine transition at codon 415 and this transition affects protein interaction, hence hampers *XPF/ERCC1* complex activity during NER pathway and alters genetic susceptibility toward cancer [8]. Another polymorphism is *XPF* 673 C>T (rs316028) which suggests a cytosine-to-thymine transition, thus affecting the transcriptional activity of the gene. Third and the least studied polymorphism is *XPF* 11985 A>G (rs254942) depicting alanine-to-glycine transition. Very few epidemiological studies have been reported on *XPF* variants till now. A study by Smith *et al.* suggested association of *XPF* G415A polymorphism with increased risk toward breast cancer [9]. On the same lines, a study on Polish population documented relationship of *XPF* G415A, rs1800067 polymorphism with laryngeal cancer [10]. On the contrary Liu *et al.* suggested association of 673 C>T, rs316028 polymorphism with decreased esophageal cancer risk [11]. The effect of these three *XPF* polymorphisms (673 C>T, 11985 A>G and G415A) on lung cancer risk, overall survival and response was not much explored in the North Indian population. Therefore, these SNPs were preferred in the present study.

Platinum-based doublet chemotherapy has been commonly used for the treatment of cancer patients but still survival rate of cancer patients is very low [12]. Platinum compounds bind with DNA forming platinum–DNA adduct which may lead to formation of interstrand cross-links and bulky distortions [12]. Many mechanisms play an important role in the repair of DNA damage caused by these platinum compounds including NER pathway which may influence anticancer effect of these agents [13]. The studies have emphasized the association of various polymorphisms in NER pathway gene which may lead to alteration in the repair efficiency of DNA which may influence treatment response and cancer survival [13].

Therefore, in the present study, we conducted a case–control study to examine the association of *XPF* polymorphism (673 C>T, 11985 A>G, G415A) toward risk for lung cancer and its histological subtypes. The effect of smoking on *XPF* variants and its relation with lung cancer was also evaluated. We further assessed role of *XPF* variants toward overall survival of patients treated with platinum-based chemotherapy and their response rates. High order SNP–SNP interactions were also evaluated using MDR and CART analysis.

Materials & methods

Study population

A study of 370 lung cancer cases and 370 controls who were enrolled in the Department of Pulmonary Medicine of Post Graduate Institute of Medical Education and Research, Chandigarh. The study was approved by the Ethics Committee board of Post Graduate Institute of Medical Education and Research. The study has patients without any age, gender, smoking, histology and Tumor, Node, Metastasis (TNM) staging restrictions. The case patients included those patients which were newly diagnosed, were previously untreated and histopathologically confirmed with lung cancer. The patients came to hospital for check-up and were not having cancer preferred as controls. A written consent from each and every individual was obtained. A trained interviewer fills up a detailed questionnaire for each case and control regarding demographics and lifetime of smoking history. Details about histology, TNM classification, clinical staging, regimen of treatment and number of chemotherapeutic cycles were obtained from the medical records of each patient available in the hospital. The follow-up data of patients were collected telephonically from the information mentioned in the medical records.

Genotyping of *XPF* variants

Genomic DNA was isolated using phenol chloroform extraction protocol as mentioned by Sodhi *et al.* [14]. *XPF* polymorphisms were detected using PCR–restriction fragment length polymorphism assay as mentioned in previous studies [11,15]. A total of 15% samples were repeated for each polymorphism and results obtained were 100% concordant.

Chemotherapeutic treatment

All patients were treated with platinum-based doublet chemotherapy. The chemotherapy drugs composed of a platinum-based drug (cisplatin or carboplatin) provided along with a taxol based drug in other words, docetaxel, irinotecan or pemetrexed. Docetaxel is a taxane, however irinotecan is a topoisomerase I poison and pemetrexed is an antifolate. These drugs were administered as an 1-h infusion, in other words, docetaxel 75, 500 mg/m² pemetrexed or 75 mg/m² irinotecan, followed by infusion of 70 mg/m² a platinum-based drug for 3 h. The

chemotherapy was repeated for every 3–4 weeks and tumor response was observed after four chemotherapy cycles. In case of unacceptable toxicity chemotherapy treatment was stopped. If an objective response was reported then patients were treated with two additional chemotherapy cycles (maximum of six cycles).

Statistical analysis

To compare demographic features like sex and smoking status, χ^2 test was applied. For continuous variables such as age and pack years smoked Student's t-test has been used. To differentiate between observed and expected genotypic frequencies of controls and cases Hardy–Weinberg equilibrium (HWE) was applied. Allelic and genotypic frequencies of cases and controls were evaluated using Pearson's χ^2 test. Logistic regression analysis was used to evaluate odds ratio (OR) and 95% CI adjusted for age, sex and smoking status to predict association between *XPF* variants and lung cancer risk. Kaplan–Meier survival analysis was applied to estimate survival distribution and log-rank test for each SNP and its relation with lung cancer. The multivariate Cox proportional hazard model was applied to assess the independent effect of each polymorphism on overall survival (OS) when adjusted for age, gender, smoking status, stage, histology, Karnofsky performance scale (KPS) and Eastern Cooperative Oncology Group (ECOG). Further relationship between OS and *XPF* polymorphic variants when categorized as per chemotherapeutic regimen was evaluated using Kaplan–Meier method and Cox hazard analysis method. Tumor response was evaluated using response evaluation criteria in solid tumors for patients treated with ≥ 6 chemotherapy cycles. Patients showing complete response or partial response were regarded as responders whereas showing stable disease or progressive disease were considered as nonresponders. All the p-values were two sided and $p < 0.05$ was referred as statistically significant. All the statistical analysis was carried out using Medcalc Statistical Software version 15.11.4 (MedCalc Software bvba, Ostend, Belgium).

A nonparametric method multifactor dimensionality reduction (MDR) was applied using MDR software package version 0.5.1 available online (www.epistasis.org), it results in interaction models used for characterization of gene–gene interactions. The aim of this method was to find main factor and make combinations with multiple factors such as genotypes which were frequently associated with cases then controls. The multilocus genotypes were arranged in two groups depicting high and low risk, hence reducing the genotype predictors from n dimensions to one. Cross validation consistency frequency (CVC) and permutation testing defines the status of the disease. The combinations with high CVC and testing accuracy are considered the best interaction model from all the models that have been generated. We further applied classification and regression tree analysis (CART) was applied to evaluate high order SNP–SNP interaction using CART software (8.0, Salford systems, CA, USA). It is a binary recursive-partitioning tool which results in a decision tree to identify the subgroups of subjects with high risk. The recursive partitioning algorithm starts at the first node and then applies statistical hypothesis testing approach to determine first locally optimal split and then each subsequent split of the dataset, with each multiplicity adjusted p-value to control tree growth ($p < 0.05$). The process continues until the terminal nodes keep on splitting to reach subsequently significant nodes or minimum size of subjects in analysis.

Result

Demographic characteristics of *XPF* variants

Demographic characteristics of *XPF* polymorphisms and its association risk of lung cancer had been represented in Table 1. The study had 370 cases and 370 controls depicting a significant difference in the mean age among cases and controls as 58.11 ± 10.44 , 53.83 ± 10.18 ($p < 0.0001$) respectively. A higher representation of smokers in cases was observed in the present study as compared with controls (81.89 vs 73.29%). When cumulative smoking dose was calculated in terms of pack years, it was observed that a significantly higher pack years was associated with cases in comparison with the controls (29.47 ± 35.04 vs 17.88 ± 19.45 ; $p = 0.0001$). The lung cancer cases were further classified into different histological subtypes: 36.48% were squamous cell carcinoma patients, 32.48% patients were detected with adenocarcinoma (ADCC) and 30% were having small-cell lung cancer (SCLC). Similarly, the cases were categorized according to TNM staging: 0.81% patients were diagnosed at stage I, 2.70% at stage II, 45.94 and 44.05% at stage III and IV respectively of the lung cancer. The frequencies of lymph node involvement was divided as 12.43, 9.72, 42.43, 28.64% for N0, N1, N2, N3 respectively. A total of 42.70% individuals were having metastasis (M1) whereas 50.54% patients did not show any metastasis (M0). A total of 311 lung cancer patients on which survival analysis was performed were classified on basis of KPS and ECOG: 123 (39.54%) patients were having KPS range from 100 to 90 while 138 (44.37%) had KPS score between 80 and 70, and only 14.46% patients were reported with KPS ≤ 60 . Similarly on ECOG scale, 47.58% (148) of the cases had

Table 1. Demographic and clinical characteristics among cases and control patients in North Indian population.

Variable	Cases, n (%) n = 370	Controls, n (%) n = 370	p-value [†]
Age (years): – Mean ± SD	58.11 ± 10.44	53.83 ± 10.18	<0.0001 [‡]
Gender: – Male – Female	319 (86.21%) 51 (13.78%)	319 (86.21%) 51 (13.78%)	0.91
Smoking status: – Smokers – Non-smokers	303 (81.89%) 67 (18.10%)	271 (73.24%) 99 (26.75%)	0.0063 [‡]
Pack years: – Mean ± SD	24.97 ± 35.04	17.88 ± 19.45	<0.0001 [‡]
Histological types (n = 370): – SQCC – ADCC – SCLC – Others – Unknown	135 (36.48%) 120 (32.43%) 111 (30%) 4 (1.08%) 0 (0.0)		
Tumor size (n = 370): – Tx – T1 – T2 – T3 – T4 – Unknown	11 (2.97%) 16 (4.32%) 42 (11.35%) 85 (22.97%) 191 (51.62%) 25 (6.75%)		
Metastasis (n = 370): – M0 – M1 – Unknown	187 (50.54%) 158 (42.70%) 25 (6.75%)		
TNM staging (n = 370): – I – II – III – IV – Unclassified	3 (0.81%) 10 (2.70%) 175 (45.94%) 163 (44.05%) 19 (5.13%)		
Lymph node involvement (n = 370): – N0 – N1 – N2 – N3 – Unknown	46 (12.43%) 36 (9.72%) 157 (42.43%) 106 (28.64%) 25 (6.75%)		
KPS (n = 311): – 100–90 – 80–70 – ≤60 – Unknown	123 (39.54%) 138 (44.37%) 45 (14.46%) 5 (1.60%)		
ECOG (n = 311): – 0–1 – 2–4 – Unknown	148 (47.58%) 158 (50.80%) 5 (1.60%)		
Regimen (n = 311): – Doce–cis/doce–carb – Irino–cis/irino–carb – Peme–cis/peme–carb – Others/unknown	76 (24.43%) 70 (22.50%) 69 (22.18%) 96 (30.86%)		

[†]p-values were derived from Pearson χ^2 test except age; Student's t-test was used for age. All p-values are two-sided.
[‡]Statistically significant p-values ($p < 0.05$ was considered statistically significant).

ADCC: Adenocarcinoma; Carb: Carboplatin; Cis: Cisplatin; Doce: Docetaxel; ECOG: Eastern Cooperative Oncology Group; Irino: Irinotecan; KPS: Karnofsky performance scale; n: Total number of case patients or control subject; Peme: Pemetrexed; SCLC: Small-cell lung carcinoma; SD: Stable disease; SQCC: Squamous cell carcinoma; TNM: Tumor, Node, Metastasis.

an ECOG value of 0–1 whereas 50.80% (158) had an ECOG value of 2–4. When cases were distributed as per their chemotherapy regimen, only those regimens were preferred for the study with which maximum number of patients was treated. Out of them 76 were treated with docetaxel–cisplatin/docetaxel–carboplatin regimen, 70 with irinotecan–cisplatin/irinotecan–carboplatin regimen and 69 with pemetrexed–cisplatin/pemetrexed–carboplatin regimen.

XPF polymorphism & their association with lung cancer risk

The genotypic and allelic distribution of *XPF* polymorphic variants in other words, *XPF* 673 C>T, *XPF* 11985 A>G and *XPF* Arg⁴¹⁵Gln and their association with lung cancer risk have been represented in Table 2. The genotypic distribution in controls for *XPF* 11985 A>G, rs254942 and *G415A*, rs1800067 were in agreement with HWE in other words, ($n = 370$, χ^2 df = 2.35, df = 1; $p = 0.12$) and ($n = 370$, χ^2 df = 2.11, df = 1; $p = 0.14$) whereas controls of *XPF* 673 C>T, rs316028 polymorphism was not in accordance with the HWE ($n = 370$, χ^2 df = 28.61, df = 1; $p = 0.00$). In *XPF* 673 C>T, rs316028 polymorphism, heterozygous genotype (CT) was more represented in cases as compared with controls (38.10 vs 31.80%) whereas the wild genotype (CC) was more expressed among controls when compared with the cases (53.7 vs 48.9%). We also found that the mutant genotype (TT) was more pronounced in controls than cases (16.2 vs 12.9%). The minor allele frequency (MAF) for 673 C>T, rs316028 polymorphism was 0.32 for both cases and controls.

For 11985 A>G, rs254942 polymorphism the wild genotype (AA) was more frequently expressed in controls as compared with cases (61.3 vs 58.10%). The double variant genotype (AG) was slightly more represented in cases than controls (35.1 vs 32.16%). In *G415A*, rs1800067 polymorphism, Out of 370 cases 53.7% were homozygous wild-type (GG), 41.6% were heterozygous and 4.7% were homozygous variant genotype. Overall, there was a significant difference among the genotypic distribution between cases and controls was observed for 11985 A>G, rs254942 polymorphism ($\chi^2 = 0.832$, df = 2; $p = 0.03$) whereas for 673 C>T, rs316028 polymorphism there was no significant overall genotypic distribution observed among cases and controls ($\chi^2 = 4.16$, df = 2; $p = 0.07$). Controls having *G415A*, rs1800067 polymorphism presented 47% wild, 45.6% heterozygous and 7.5% mutant genotype. However, overall there was no significant genotypic distribution observed for *G415A*, rs1800067 polymorphism ($\chi^2 = 4.97$, df = 2; $p = 0.08$). A total of 11,985 A>G polymorphism showed 0.23, 0.24 MAF in controls and cases while MAF for *G415A* polymorphism was 0.30 and 0.25 in cases and controls respectively.

To understand association between polymorphic variants of *XPF* gene and their role toward lung cancer susceptibility, adjusted OR's and 95% CI were calculated using logistic regression analysis. The analysis was also applied on the histological subtypes of lung cancer. Homozygous wild genotype was considered as reference group. A 1.5-fold of lung cancer risk was observed for heterozygous genotype (CT) for 673 C>T, rs316028 polymorphism when compared with the wild genotype (CC) (OR: 1.43; 95% CI: 1.02–1.99; $p = 0.03$). Similarly, when compared with CC genotype, a high lung cancer risk for heterozygous genotype (CT) was observed in squamous cell carcinoma patients (OR: 1.52; 95% CI: 0.96–2.41; $p = 0.07$). A total of 11,985 A>G (rs254942) and *G415A* (rs1800067) polymorphism did not show any significant results with lung cancer risk.

Effect of smoking on XPF polymorphism

Interactive effect of smoking on *XPF* polymorphism and their association with lung cancer was assessed. The study population was divided into smokers and non-smokers on the basis of their smoking status (Table 3). The smoking group was further dichotomized into light smokers (≤ 25 pack years) and heavy smokers (> 25 years). There was no significant relationship observed between smoking in 11985 A>G, rs254942 polymorphism and lung cancer susceptibility. 673 C>T genotypes also did not show any association with smokers but a significant twofold risk was observed in individuals carrying heterozygous and mutant genotype (CT + TT) in heavy smokers (OR: 1.78; 95% CI: 1.02–3.10; $p = 0.03$). For *G415A*, rs1800067 polymorphism a significant protective effect among non-smokers carrying heterozygous genotype (GA) was observed when compared with the wild genotype (GG; OR: 0.51; 95% CI: 0.26–0.99; $p = 0.04$). Combined genotype (GA + AA) also showed a protective effect for non-smokers with *G415A*, rs1800067 polymorphism (OR: 0.49; 95% CI: 0.25–0.94; $p = 0.03$).

Combined effect of XPF SNPs on lung cancer risk

The genotypic combinations of the three *XPF* polymorphic variants were analyzed to assess their association with lung cancer, as shown in Supplementary Table 1. Individuals carrying any single variant allele for 673 C>T (rs316028) and 11985 A>G (rs254942) polymorphic variants in other words, heterozygous carriers (CT + AG) showed almost twofold risk toward lung cancer (OR: 1.84; 95% CI: 1.07–3.14; $p = 0.02$). When double variant genotypes were combined together as a single genotype ([CT + TT] + [AG + AA]) for 673 C>T and 11985 A>G variants also showed a high risk for the disease (OR: 1.67; 95% CI: 1.00–2.79; $p = 0.04$), however any other combination between the SNPs of *XPF* did not show any significant association with the lung cancer risk.

Table 2. Distribution of genotypes of XPF polymorphism depicting their association with lung cancer risk and its histological subtypes

XPF SNPs	Overall lung cancer				SQCC				ADCC				SCLC			
	Controls n (%) n = 370	Cases n (%) n = 370	AOR (95% CI) [†]	p-value [‡]	Cases n (%) n = 135	AOR (95% CI) [†]	p-value [‡]	Cases n (%) n = 120	AOR (95% CI) [†]	p-value [‡]	Cases n (%) n = 111	AOR (95% CI) [†]	p-value [‡]	Cases n (%) n = 111	AOR (95% CI) [†]	p-value [‡]
673 C > T (rs316028)																
CC	199 (53.7)	181 (48.9)	1.00 (Reference)		68 (50.37)	1.00 (Reference)		59 (49.1)	1.0 (Reference)		52(46.84)	1.00 (Reference)		52(46.84)	1.00 (Reference)	
CT	118 (31.8)	141 (38.1)	1.43 (1.02–1.99)	0.03 [§]	50(37.07)	1.52 (0.96–2.41)	0.07	45 (37.5)	1.27 (0.79–2.04)	0.31	45 (40.54)	1.39 (0.85–2.29)	0.18	45 (40.54)	1.39 (0.85–2.29)	0.18
TT	60 (16.2)	48 (12.9)	0.92 (0.59–1.43)	0.72	17 (12.5)	1.01 (0.53–1.93)	0.95	16 (13.3)	0.89 (0.46–1.69)	0.72	14 (12.6)	1.00 (0.51–1.98)	0.98	14 (12.6)	1.00 (0.51–1.98)	0.98
CT + TT	178 (48.1)	189 (51.08)	1.25 (0.93–1.69)	0.13	67 (49.6)	1.34 (0.88–2.03)	0.16	61 (50.8)	1.13 (0.74–1.75)	0.55	59 (53.15)	1.28 (0.81–2.01)	0.27	59 (53.15)	1.28 (0.81–2.01)	0.27
C Allele	516 (69.72)	503 (67.97)														
T Allele	238 (32.16)	237 (32.07)														
MAF	0.32	0.32														
11985 A > G (rs254942)																
AA	227 (61.3)	215 (58.10)	1.00 (Reference)		86 (63.7)	1.00 (Reference)		66 (55)	1.00 (Reference)		65 (58.5)	1.00 (Reference)		65 (58.5)	1.00 (Reference)	
GA	119 (32.16)	130 (35.1)	1.16 (0.84–1.59)	0.35	46 (34.07)	1.14 (0.73–1.79)	0.54	44 (36.6)	1.19 (0.75–1.88)	0.44	37 (33.3)	1.03 (0.63–1.68)	0.88	37 (33.3)	1.03 (0.63–1.68)	0.88
GG	24 (6.4)	25 (6.75)	1.08 (0.58–2.00)	0.79	5 (3.7)	0.58 (0.20–1.63)	0.30	10 (8.3)	1.59 (0.70–3.64)	0.26	9 (8.1)	1.19 (0.51–2.79)	0.67	9 (8.1)	1.19 (0.51–2.79)	0.67
GA + GG	143 (38.6)	155 (41.8)	1.14 (0.84–1.55)	0.36	51 (37.7)	1.04 (0.68–1.60)	0.84	54 (0.45)	1.25 (0.81–1.93)	0.29	46 (41.44)	1.06 (0.67–1.68)	0.78	46 (41.44)	1.06 (0.67–1.68)	0.78
A Allele	573 (77.42)	560 (75.67)														
G Allele	167 (22.56)	180 (24)														
MAF	0.23	0.24														
G415A (rs1800067)																
GG	174 (47.0)	199 (53.7)	1.00 (Reference)		69 (51.11)	1.00 (Reference)		63 (52.5)	1.00 (Reference)		64 (57.6)	1.00 (Reference)		64 (57.6)	1.00 (Reference)	
GA	168 (45.6)	154 (41.6)	0.84 (0.62–1.14)	0.27	59 (43.7)	0.90 (0.61–1.45)	0.80	52 (43.3)	0.91 (0.59–1.42)	0.70	43 (38.7)	0.76 (0.48–1.22)	0.26	43 (38.7)	0.76 (0.48–1.22)	0.26
AA	28 (7.5)	17 (4.5)	0.59 (0.30–1.13)	0.11	7 (5.1)	0.62 (0.25–1.53)	0.30	5 (4.1)	0.56 (0.20–1.59)	0.28	4 (3.6)	0.46 (0.15–1.39)	0.17	4 (3.6)	0.46 (0.15–1.39)	0.17
GA + AA	196 (52.97)	171 (46.21)	0.81 (0.60–1.09)	0.16	66 (48.8)	0.90 (0.59–1.36)	0.63	57 (47.5)	0.86 (0.54–1.36)	0.52	47 (42.34)	0.72 (0.46–1.14)	0.16	47 (42.34)	0.72 (0.46–1.14)	0.16
G Allele	516 (69.72)	552 (74.59)														
A Allele	224 (30)	188 (25)														
MAF	0.30	0.25														

† Adjusted odds ratios; 95% CIs and their corresponding p-values were calculated logistic regression analysis after adjusting for age, gender and smoking status.

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

§ Significant p-value.

ADCC: Adenocarcinoma; AOR: Adjusted odds ratio; MAF: Minor allele frequency; SCLC: Small-cell lung carcinoma; SQCC: Squamous cell carcinoma.

Table 3. Distribution of genotypes of XPF polymorphism according to their smoking status and its association with lung cancer risk.

SNP & genotype	Smokers				Non-smokers			
	Controls n (%) n = 271	Cases n (%) n = 303	AOR (95% CI) [†]	p-value [‡]	Controls n (%) n = 99	Cases n (%) n = 67	AOR (95% CI) [†]	p-value [‡]
673 C>T (rs316028)								
CC	77 (28.4)	110 (36.3)	1.00 (Reference)		45 (45.45)	31 (46.2)	1.00 (Reference)	
CT	155 (57.19)	150 (49.5)	1.38 (0.94–2.02)	0.09	34 (34.34)	31 (46.2)	1.52 (0.75–3.09)	0.24
TT	39 (14.39)	43 (14.1)	1.12 (0.68–1.84)	0.65	20 (20.2)	5 (7.4)	0.41 (0.13–1.28)	0.12
CT + TT	194 (71.5)	193 (63.6)	1.29 (0.91–1.81)	0.14	64 (95.5)	36 (53.7)	1.10 (0.57–2.13)	0.75
11985 A>G (rs254942)								
AA	167 (61.6)	177 (58.4)	1.00 (Reference)		60 (60.6)	38 (56.7)	1.00 (Reference)	
AG	84 (30.9)	107 (35.3)	1.23 (0.85–1.78)	0.25	35 (35.35)	23 (34.32)	1.02 (0.51–2.03)	0.94
GG	20 (7.3)	19 (6.2)	0.93 (0.46–1.86)	0.84	4 (4.04)	68.9	2.19 (0.49–9.67)	0.30
AG + GG	104 (38.3)	126 (41.5)	1.18 (0.83–1.66)	0.34	39 (39.39)	29 (43.28)	1.14 (0.59–2.20)	0.68
G415A (rs1800067)								
GG	134 (49.44)	159 (52.4)	1.00 (Reference)		40 (40.4)	40 (59.70)	1.00 (Reference)	
GA	116 (42.8)	129 (42.5)	0.94 (0.66–1.33)	0.73	52 (52.52)	25 (37.31)	0.51 (0.26–0.99)	0.04 [§]
AA	21 (7.74)	15 (4.9)	0.70 (0.34–1.44)	0.33	7 (7.07)	2 (2.9)	0.28 (0.05–1.50)	0.13
GA + AA	137 (50.5)	144 (47.52)	0.90 (0.64–1.27)	0.56	59 (59.59)	27 (40.2)	0.49 (0.25–0.94)	0.03 [§]
	Heavy smoker				Light smoker			
673 C>T (rs316028)	Controls n (%) n = 100	Cases n (%) n = 141	AOR (95% CI) [†]	p-value [‡]	Controls n (%) n = 171	Cases n (%) n = 162	AOR (95% CI) [†]	p-value [‡]
CC	67 (77)	76 (53.9)	1.00 (Reference)		88(51.4)	74 (45.6)	1.00 (Reference)	
CT	24 (24)	47 (33.33)	1.78 (0.96–3.28)	0.06	53(30.9)	63 (38.8)	1.24 (0.75–2.04)	0.38
TT	9 (9)	18 (12.76)	1.82 (0.74–4.47)	0.19	30(18.7)	25 (15.4)	0.97 (0.52–1.82)	0.93
CT + TT	33(33)	65 (46.09)	1.78 (1.02–3.10)	0.03 [§]	83(48.5)	88 (54.3)	1.14 (0.73–1.79)	0.54
11985 A>G (rs254942)								
AA	59 (59)	77 (54.6)	1.00 (Reference)		108 (63.15)	100 (61.7)	1.00 (Reference)	
AG	34 (34)	55 (39.0)	1.22 (0.69–2.14)	0.48	50 (29.23)	52 (32.09)	1.18 (0.78–1.93)	0.48
GG	8 (8)	9 (6.3)	0.78 (0.27–2.19)	0.68	13 (7.6)	10 (6.1)	0.87 (0.34–2.24)	0.78
AG + GG	42 (42)	64 (45.3)	1.14 (0.93–2.88)	0.08	63 (36.8)	62 (38.2)	0.99 (0.64–1.53)	0.97
G415A (rs1800067)								
GG	47 (47)	70 (49.6)	1.00 (Reference)		86 (50.29)	89 (54.9)	1.00 (Reference)	
GA	46 (46)	66 (46.8)	0.97 (0.56–1.76)	0.94	70 (40.93)	63 (38.8)	0.91 (0.57–1.47)	0.72
AA	7 (7)	5 (3.5)	0.61 (0.18–2.10)	0.44	14 (8.18)	10 (6.1)	0.90 (0.36–2.23)	0.82
GA + AA	53 (53)	71 (50.35)	0.93 (0.54–1.61)	0.81	84 (49.12)	73 (45.06)	0.90 (0.57–1.42)	0.67

[†] Adjusted odds ratios, 95% CIs and their corresponding p-values were calculated logistic regression analysis after adjusting for age, gender and smoking status.

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

[§] Significant p-value

AOR: Adjusted odds ratio.

Relationship of clinic-pathological features with XPF polymorphic variants

Clinic-pathological parameters such as stage, primary tumor extension, lymph-node involvement and metastasis were analyzed to find out the effect of genetic variations in *XPF* gene on these factors as shown in Supplementary Table 2. The effect of different genotypes of *XPF* genetic variants and their relationship with advancement of these clinical parameters and progression of the disease was assessed using logistic linear regression. The wild genotype for each polymorphism was regarded as the reference group. Odds ratios obtained were adjusted for age, gender and smoking status. For *11985 A>G*, rs254942 polymorphism, an increased risk for heterozygous genotype (AG) was observed in stage IV patients when compared with stage III patients (OR: 1.63; 95% CI: 1.05–2.66; $p = 0.04$). Combined genotype (AG + GG) was also associated with higher risk for stage IV patients in *11985 A>G*, rs254942 polymorphism (OR: 1.59; 95% CI: 1.00–2.53; $p = 0.04$). A decreased risk was observed in patients with distant metastasis when compared with the metastatic patient (OR: 59; 95% CI: 0.36–0.95; $p = 0.03$) showing *11985 A>G*, rs254942 polymorphism. The similar trend was observed for heterozygous + mutant patients suggesting a

protective effect for patients without metastasis (OR: 0.60; 95% CI: 0.38–0.95; $p = 0.03$). There was no significant distribution for clinic-pathological features in genotype variants of *G415A* (rs1800067) and *673 C>T* (rs316028) polymorphism.

Association of *XPF* polymorphic variants & OS of lung cancer patients

A survival analysis for 311 lung cancer cases was conducted and its association with the polymorphic variants of the *XPF* has been assessed in Table 4. Association between OS and *XPF* polymorphic variants was applied using a univariate analysis expressed as Kaplan–Meier curve and log-rank test followed by a multivariate Cox regression model adjusted for various factors like age, gender, smoking, histology, stage, KPS and ECOG. The genotypic distribution in the patients agreed with that estimated from HWE for *11985 A>G* ($\chi^2 = 0.28$; df = 1; $p = 0.59$) and *G415A* ($\chi^2 = 1.45$; df = 1; $p = 0.22$) polymorphisms. No significant association was observed between all the three *XPF* variants and OS of all the lung cancer cases. We also analyzed patients on the basis of their histological subtypes and evaluated their OS. For *673 C>T*, rs316028 polymorphism, it was observed that patients with ADCC and harboring the mutant genotype (AA) showed a longer survival period of 14.26 months when compared with the ADCC patients expressing the wild-type genotype (CC) and having a median survival time (MST) of 6.33 months. Similarly ADCC patients with heterozygous (CT) genotype had a better OS as compared with the ADCC subjects carrying the wild-type genotype (MST: 8.80 vs 6.33, log-rank $p = 0.13$). On applying the Cox regression model, a lower death ratio when compared with the wild genotype (hazard ratio [HR]: 0.52; 95% CI: 0.32–0.86; $p = 0.01$; Figure 1A) was observed for ADCC subjects. No significant association was observed for the other two lung cancer histology. For the other SNP of *XPF* gene in other words, *11,985 A>G*, rs254942 polymorphism, it was observed that SCLC patients carrying single copy of the variant allele (AG) had a poor OS as compared with subjects with the wild-type genotype (AA) (MST = 4.0 vs 11.6, log-rank $p = 0.001$) as shown in Figure 1B. Such patients also had a very poor prognosis as the death ratio associated was high (HR: 2.40; 95% CI: 1.37–4.18, $p = 0.001$). On applying the Cox regression model, it was observed that there was an increase in the (HR: 3.19; 95% CI: 1.87–5.144; $p = 0.001$).

In case of the *XPF G415A* (rs1800067) polymorphism ADCC patients with heterozygous genotype (GA) showed a 1.5-fold hazard ratio with the moderate survival among cases (HR: 1.57; 95% CI: 1.05–2.33; $p = 0.02$; MST = 8.03 months; Figure 1C). The Cox regression analysis has been depicted for all the SNPs and their association with lung cancer and its histological subtypes have been illustrated in Supplementary Table 4 (a–l).

Association of *XPF* polymorphic variants & their association with overall survival in lung cancer patients treated with different chemotherapy regimens

The overall survival of patients treated with different chemotherapy regimens and their association with *XPF* polymorphic variants has been illustrated in Table 5. Homozygous wild genotype for all the three polymorphic variants of *XPF* gene was considered as reference. In case of *11985 A>G*, rs254942 polymorphism, lung cancer patients with mutant genotype (GG) and treated with docetaxel–cisplatin/docetaxel–carboplatin regimen showed a low death ratio with those patients who expressed the wild-type genotype (AA) in study population (HR: 0.25; 95% CI: 0.06–1.09; $p = 0.05$). Patients having mutant genotype (GG) for *XPF 11985 A>G*, rs254942 polymorphism and treated with irinotecan–cisplatin/irinotecan–carboplatin regimen showed a shorter OS and a high death ratio when compared with those patients harboring the wild-type genotype (MST = 4.46 vs 11.26; HR: 4.73; 95% CI: 0.82–27.01; $p = 0.08$). Similarly for the same polymorphic variant as above lung cancer subjects who were harboring the heterozygous genotype (AG) and treated with irinotecan cisplatin/irinotecan carboplatin regimen showed a poor OS as compared with subjects having reference genotype (AA). Such patients exhibited a high mortality risk (HR = 2.66; 95% CI: 1.45–4.88; $p = 0.001$; MST = 4.30; Figure 1D). Furthermore, Patients having combined genotype (AG + GG) and treated with irinotecan–cisplatin/irinotecan–carboplatin regimen depicted almost twofold HR when compared with subjects with wild genotype (HR: 2.64; 95% CI: 1.46–4.78; $p = 0.001$).

MDR analysis

The MDR analysis was performed on a dataset having subjects with or without lung cancer resulting into average CVC and average prediction error (Table 6 & Figure 2A). MDR tool was used to identify which combination of factors could be the best model to predict high risk subgroups of the provided subjects. The analysis resulted in best one factor model which includes *XPF G415A* showing CVC of 100/100 and prediction error of 0.47 ($p = 0.06$). A higher CVC and a minimum prediction error are preferred for the best results. *XPF 673 C>T*, *11985 A>G*, *G415A*

Table 4. Relationship of XPF polymorphism with overall survival of lung cancer patients.

XPF SNPs		Overall LC							ADCC							
		Dead 271	Alive 40	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value	673 C>T (rs316028)	Dead 86	Alive 18	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value
673 C>T (rs316028)		134	15	8.06	1.00 (Reference)		1.00 (Reference)		CC	44	5	6.33	1.00 (Reference)		1.00 (Reference)	
CT		101	18	7.33	0.92 (0.71–1.20)	0.57	0.90 (0.69–1.18)	0.49	CT	31	8	8.80	0.70 (0.45–1.11)	0.13	0.52 (0.32–0.86)	0.01
TT		36	7	9.43	0.84 (0.59–1.20)	0.36	0.90 (0.60–1.35)	0.63	TT	11	5	14.26	0.56 (0.31–0.99)	0.07	0.50 (0.50–1.13)	0.09
CT + TT		137	25	7.60	0.90 (0.71–1.15)	0.40	0.91 (0.71–1.16)	0.47	CT + TT	42	13	9.43	0.65 (0.42–1.00)	0.04	0.57 (0.35–0.90)	0.01
SQCC																
673 C>T (rs316028)		Dead 86	Alive 4	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value	673 C>T (rs316028)	Dead 99	Alive 18	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value
		40	2	10.4	1.00 (Reference)		1.00 (Reference)		CC	50	8	7.93	1.00 (Reference)		1.00 (Reference)	
CT		34	2	6.30	1.09 (0.68–1.72)	0.70	1.10 (0.68–1.78)	0.69	CT	36	8	6.66	1.01 (0.65–1.55)	0.95	1.02 (0.65–1.61)	0.90
TT		12	-	5.10	1.08 (0.56–2.09)	0.80	1.65 (0.74–3.70)	0.22	TT	13	2	8.03	1.02 (0.55–1.90)	0.92	1.12 (0.56–2.20)	0.74
CT + TT		46	2	6.30	1.09 (0.71–1.66)	0.68	1.14 (0.73–0.54)	0.54	CT + TT	49	10	6.86	1.01 (0.68–1.50)	0.93	1.02 (0.68–1.55)	0.89
ADCC																
11985 A>G (rs254942)		Dead 271	Alive 40	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value	11985 A>G (rs254942)	Dead 86	Alive 18	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value
		164	22	8.06	1.00 (Reference)		1.00 (Reference)		AA	46	12	7.13	1.00 (Reference)		1.00 (Reference)	
AG		95	12	6.76	1.10 (0.85–1.43)	0.42	1.16 (0.89–1.52)	0.25	AG	34	4	9.23	1.19 (0.71–1.75)	0.61	1.12 (0.70–1.79)	0.61
GG		12	6	8.40	0.66 (0.40–1.09)	0.17	0.84 (0.62–1.13)	0.25	GG	6	2	8.96	0.88 (0.39–1.99)	0.78	0.98 (0.62–1.54)	0.94
AG + GG		107	18	7.56	1.03 (0.80–1.32)	0.79	1.09 (0.85–1.41)	0.46	AG + GG	40	6	9.23	9.23 (1.07–1.65)	0.72	1.11 (0.71–1.75)	0.62
SQCC																
11985 A>G (rs254942)		Dead 86	Alive 4	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value	11985 A>G (rs254942)	Dead 99	Alive 18	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value
		52	3	11.6	1.00 (Reference)		1.00 (Reference)		AA	66	7	6.40	1.00 (Reference)		1.00 (Reference)	
AG		29	-	4	2.40 (1.37–4.18)	<0.0001 [§]	3.19 (1.87–5.44)	<0.0001 [§]	AG	32	8	8.03	0.75 (0.50–1.13)	0.19	0.79 (0.50–1.26)	0.33
Two-sided χ ² -test for either genotype distribution or allelic frequencies between the cases and controls.0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype. †Hazard ratios, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months; adjusted hazard ratios, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG. §Significant p-value.																

†Two-sided χ^2 -test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype.

‡Hazard ratios, 95% CIs and their corresponding p-values were calculated by Kaplan-Meier survival analysis after adjusting for remission and survival in months; adjusted hazard ratios, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.

§Significant p-value.

ADCC: Adenocarcinoma; CI: Confidence interval; ECOG: Eastern Cooperative Oncology Group; HR: Hazard ratio; KPS: Karnofsky performance scale; OS: Overall survival.

Table 4. Relationship of XPF polymorphism with overall survival of lung cancer patients. (cont.).															
XPF SNPs		Overall LC						ADCC							
673 C>T (rs316028)	Dead 271	Alive 40	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value	673 C>T (rs316028)	Dead 86	Alive 18	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value
GG	5	1	4.46	1.61 (0.52–4.96)	0.29	2.96 (1.00–8.70)	0.04 [§]	GG	1	3	-	0.15 (0.06–0.35)	0.02	0.15 (0.02–1.14)	0.06
AG + GG	34	1	4.13	2.20 (1.33–3.65)	0.0001 [§]	2.88 (1.75–4.71)	<0.0001 [§]	AG + GG	33	8	9.33	0.68 (0.45–1.01)	0.06	0.70 (0.44–1.10)	0.13
Overall LC															
G415A (rs1800067)	Dead 271	Alive 40	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value	G415A rs1800067	Dead 86	Alive 18	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value
GG	148	18	8.03	1.00 (Reference)		1.00 (Reference)		GG	44	9	9.13	1.00 (Reference)		1.00 (Reference)	
GA	109	19	8.05	0.94 (0.73–1.20)	0.64	0.94 (0.73–1.21)	0.63	GA	38	7	8.03	1.13 (0.73–1.75)	0.56	1.57 (1.05–2.33)	0.02
AA	14	3	7.23	0.83 (0.50–1.39)	0.52	0.93 (0.69–1.25)	0.66	AA	4	2	9.81	0.75 (0.30–1.89)	0.59	1.59 (0.86–3.06)	0.16
GA + AA	123	22	7.93	0.93 (0.73–1.18)	0.56	0.93 (0.73–1.19)	0.60	GA + AA	42	9	8.03	1.08 (0.70–1.65)	0.70	1.23 (0.79–1.92)	0.35
SCLC															
G415A (rs1800067)	Dead 86	Alive 4	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value	G415A rs1800067	Dead 99	Alive 18	Median OS (months)	HR (95% CI)	Log p [†]	HR [‡] (95% CI)	p-value
GG	51	2	8.26	1.00 (Reference)		1.00 (Reference)		GG	53	7	7.33	1.00 (Reference)		1.00 (Reference)	
GA	31	2	10	0.98 (0.63–1.54)	0.95	0.90 (0.55–1.47)	0.68	GA	40	10	6.86	0.84 (0.56–1.27)	0.46	0.75 (0.48–1.16)	0.20
AA	4	-	5.80	1.33 (0.42–4.22)	0.57	1.52 (0.86–2.69)	0.14	AA	6	1	7.93	0.85 (0.38–1.88)	0.71	0.80 (0.51–1.25)	0.34
GA + AA	35	2	9.86	1.02 (0.66–1.57)	0.92	0.98 (0.61–1.57)	0.95	GA + AA	46	11	7.57	0.84 (0.57–1.25)	0.40	0.75 (0.49–1.15)	0.20
†Two-sided χ^2 -test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype.															
‡Hazard ratios, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months; adjusted hazard ratios, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.															
§Significant p-value.															
ADCC: Adenocarcinoma; CI: Confidence interval; ECOG: Eastern Cooperative Oncology Group; HR: Hazard ratio; KPS: Karnofsky performance scale; OS: Overall survival.															

Table 5. Association of XPF polymorphism and overall survival according to regimen.

Regimen – docetaxel cis/carboplatin								
Genotype (n = 76)	No.	Dead	Alive	Median OS (months)	Crude HR (95% CI)	Adjusted HR [†] (95% CI)	Log-rank p	p-value [‡]
673 C>T (rs316028)								
CC	38	32	6	9.86	1.00 (Reference)	1.00 (Reference)		
CT	31	26	5	6.76	1.18 (0.70–1.99)	1.57 (0.86–2.88)	0.52	0.14
TT	7	7	–	8.40	1.60 (0.61–4.22)	1.34 (0.79–2.27)	0.24	0.26
CT + TT	38	33	5	7.56	1.24 (0.76–2.02)	1.51 (0.86–2.65)	0.37	0.14
11985 A>G (rs254942)								
AA	48	41	7	6.76	1.00 (Reference)	1.00 (Reference)		
AG	23	19	4	8.63	0.84 (0.49–1.42)	0.97 (0.53–1.79)	0.52	0.93
GG	5	2	3	–	0.31 (0.13–0.76)	0.25 (0.06–1.09)	0.09	0.05 [§]
AG + GG	28	21	7	10.26	0.76 (0.43–1.20)	0.77 (0.43–1.35)	0.22	0.36
G415A (rs1800067)								
GG	40	37	3	8.03	1.00 (Reference)	1.00 (Reference)		
GA	32	26	6	7.88	0.78 (0.48–1.28)	0.87 (0.50–1.51)	0.34	0.63
AA	4	2	2	–	0.30 (0.12–0.72)	0.45 (0.19–1.09)	0.07	0.08
GA + AA	36	28	8	8.20	0.71 (0.43–1.15)	0.78 (0.49–1.15)	0.17	0.38
REGIMEN – irinotecan cis/carboplatin								
Genotype (n = 70)	No.	Dead	Alive	Median OS (months)	Crude HR (95% CI)	Adjusted HR [†] (95% CI)	Log-rank p	p-value [‡]
673 C>T (rs316028)								
CC	36	33	3	7.23	1.00 (Reference)	1.00 (Reference)		
CT	27	24	3	8.26	0.98 (0.58–1.66)	1.03 (0.56–1.89)	0.95	0.91
TT	7	6	1	11	0.85 (0.37–1.96)	0.77 (0.44–1.33)	0.72	0.35
CT + TT	34	30	4	9	0.95 (0.58–1.56)	0.92 (0.54–1.58)	0.85	0.78
11985 A>G (rs254942)								
AA	42	28	14	11.26	1.00 (Reference)	1.00 (Reference)		
AG	25	23	2	4.30	2.34 (1.25–4.39)	2.66 (1.45–4.88)	0.001	0.001 [§]
GG	3	2	1	4.46	1.96 (0.28–13.74)	4.73 (0.82–27.01)	0.34	0.08
AG + GG	28	25	3	4.30	2.27 (1.25–4.14)	2.64 (1.46–4.78)	0.001	0.001 [§]
G415A (rs1800067)								
GG	38	35	3	9	1.00 (Reference)	1.00 (Reference)		
GA	29	25	4	9.86	0.93 (0.55–1.55)	0.79 (0.42–1.46)	0.78	0.45
AA	3	3	–	7.23	0.79 (0.27–2.31)	1.20 (0.42–3.43)	0.68	0.72
GA + AA	32	28	4	8.55	0.91 (0.55–1.50)	0.73 (0.48–1.56)	0.73	0.63
REGIMEN – pemetrexed cis/carboplatin								
Genotype (n = 69)	No.	Dead	Alive	Median OS (months)	Crude HR (95% CI)	Adjusted HR [†] (95% CI)	Log-rank p	p-value [‡]
673 C>T (rs316028)								
CC	35	30	5	9.46	1.00 (Reference)	1.00 (Reference)		
CT	26	20	6	9.43	0.82 (0.47–1.43)	0.62 (0.32–1.20)	0.50	0.16
TT	8	5	3	12.23	0.61 (0.27–1.37)	0.73 (0.36–1.46)	0.30	0.38
CT + TT	34	25	9	9.43	0.77 (0.45–1.31)	0.68 (0.36–1.25)	0.33	0.78
11985 A>G (rs254942)								
AA	42	32	10	10	1.00 (Reference)	1.00 (Reference)		
AG	22	19	3	9.43	1.15 (0.64–2.06)	1.17 (0.63–2.17)	0.61	0.61
GG	5	4	1	7.6	1.24 (0.40–3.88)	1.54 (0.49–4.85)	0.67	0.45
AG + GG	27	23	4	9.33	1.17 (0.68–2.02)	1.25 (0.70–2.23)	0.55	0.43

[†]Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype.

[‡]Hazard ratios, 95% CI and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted hazard ratios, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, kps and ecog.

[§]Significant p-value.

CI: Confidence interval; Cis: Cisplatin; OS: Overall survival.

Table 5. Association of XPF polymorphism and overall survival according to regimen (cont.).

Regimen – docetaxel cis/carboplatin								
Genotype (n = 76)	No.	Dead	Alive	Median OS (months)	Crude HR (95% CI)	Adjusted HR [†] (95% CI)	Log-rank p	p-value [‡]
G415A (rs1800067)								
GG	37	29	8	9.56	1.00 (Reference)	1.00 (Reference)		
GA	30	25	5	9.43	1.18 (0.68–2.02)	1.38 (0.76–2.50)	0.53	0.27
AA	2	1	1					
GA + AA	32	26	6	9.38	1.13 (0.66–1.93)	1.30 (0.72–2.33)	0.63	0.37

[†]Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype.

[‡]Hazard ratios, 95% CI and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted hazard ratios, 95% CIs and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, kps and ecog.

[§]Significant p-value.

CI: Confidence interval; Cis: Cisplatin; OS: Overall survival.

Table 6. Multifactor dimensionality reduction analysis showing interactions of XPF variants with lung cancer risk.

No. of risk factors	Best interaction model	CVC	Prediction error	p-value
1	<i>XPF G415A</i>	100/100	0.47	0.06
2	<i>XPF 11985 A>G, XPF G415A</i>	99/100	0.47	0.85
3	<i>XPF 673 C>T, XPD 11985 A>G, XPF G415A</i>	100/100	0.49	0.06

CVC: Cross validation consistency.

Table 7. Risk estimates based on classification and regression tree analysis.

Terminal node	Genotype for each terminal node	Control (n)	Case (n)	Case rate [†]	OR (95% CI)	p-value
Node 1	<i>XPF 673 C>T (M)</i>	111	141	56.0	1.60 (1.18–2.37)	0.003 [‡]
Node 2	<i>XPF 11985 A>G (W)/XPF G415A(W)/XPF 673 C>T (W)</i>	102	112	52.3	1.44 (1.00–2.08)	0.04 [‡]
Node 3	<i>XPF 11985 A>G (M)/XPF G415A(W)/XPF 673 C>T (W)</i>	8	4	33.3	1.0 [§]	
Node 4	<i>XPF Arg415Gln (M)/XPF 673 C>T (W)</i>	149	113	43.1	0.65 (0.19–2.24)	0.50

[†]Case rate is percentage LC cases among all individuals in each terminal node (case/[case + control] × 100).

[‡]p < 0.05 is considered statistically significant.

[§]Reference group.

appeared as the best interaction model with CVC of 100/100 and having prediction error of 0.49 ($p = 0.06$). The two factor model (*11985 A>G, G415A*) for evaluating lung cancer risk had CVC of 99/100 and an improved prediction error of 0.47 ($p = 0.90$). To demonstrate a visualized interaction between the three *XPF* SNPs and lung cancer risk an interaction dendrogram was used. As *G415A* and *11985 A>G* were connected with the shortest line the more synergy was depicted between the two polymorphisms whereas a weak interaction was observed between *673 C>T*, rs316028 and other two polymorphic variants.

CART analysis

Higher order SNP–SNP interactions and their association lung cancer risk were analyzed using CART analysis (Table 7 & Figure 2B). Final tree comprises of four terminal nodes. The root split takes place at the strongest risk factor suggesting *XPF673* as the most risk factor for lung cancer risk among other variants. Further subsequent split was on *XPF G415A*, rs1800067 and *11985 A>G*, rs254942. To understand potential mechanism for increased lung cancer risk with SNPs in CART analysis, we also applied logistic regression analysis in the final model selected by CART evaluating OR and 95% CI. Node 3 was considered as the reference as it has the minimum

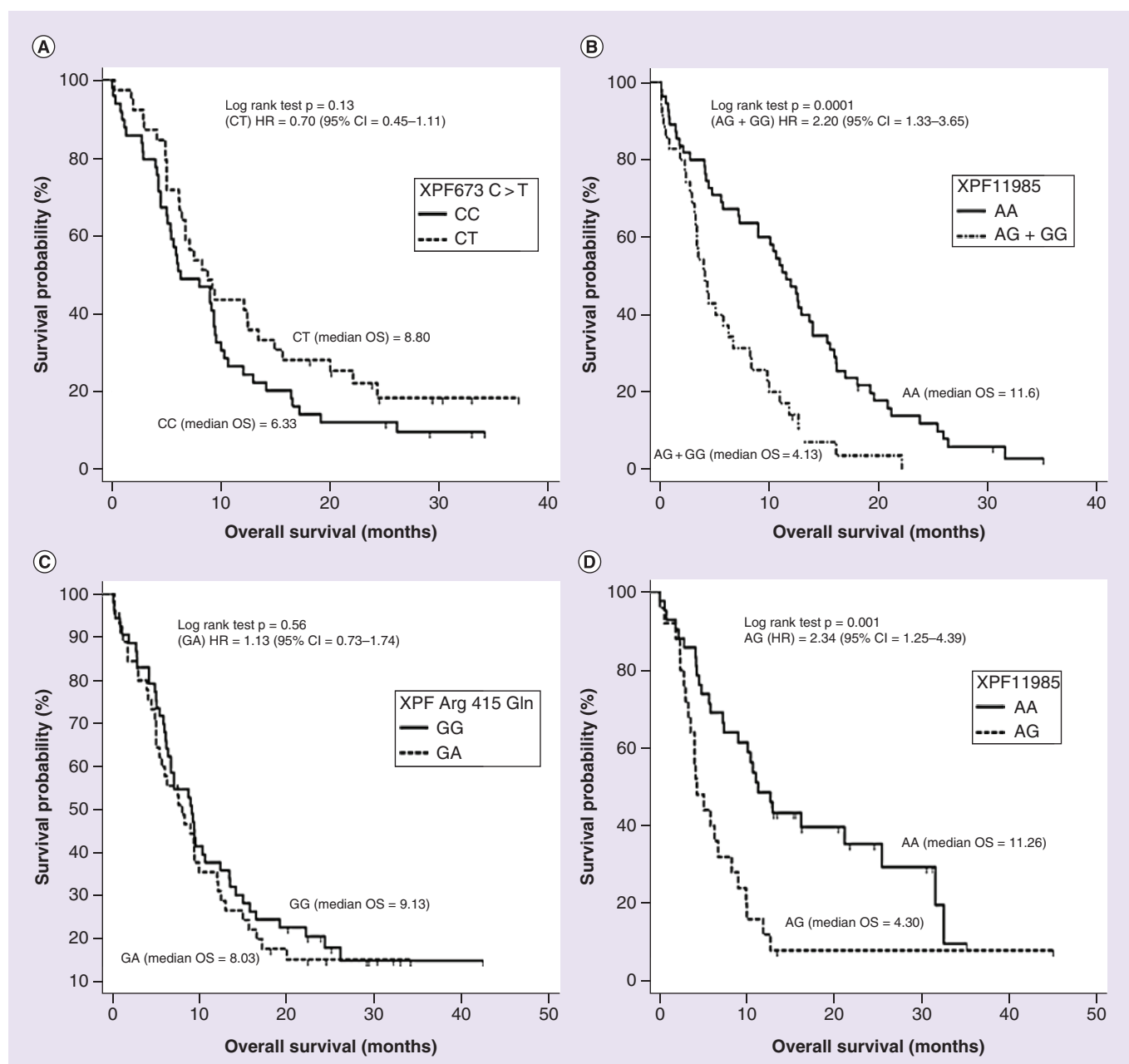


Figure 1. Kaplan–Meier curves and the association between overall survival in patients with the XPF genotype. Kaplan–Meier curves illustrating association between overall survival in patients with (A) heterozygous genotype (CT) for ADCC showing *XPF* 673 C>T polymorphism (B) heterozygous genotype (AG) for SCLC showing 11985 A>G polymorphism and (C) heterozygous genotype (GA) for ADCC in *G415A* polymorphism (D) Kaplan–Meier curves illustrating association between overall survival in lung cancer patients with heterozygous genotype (AG) treated with irinotecan–cisplatin/irinotecan–carboplatin regimen showing 11985 A>G polymorphism. ADCC: Adenocarcinoma; SCLC: Small-cell lung carcinoma.

case rate in other words, 33.3%. Individuals carrying wild genotype (*W*) of *XPF*673 C>T, rs316028 and mutant (heterozygous + mutant) genotype (*M*) of *XPF* *G415A*, rs1800067 represent node 3. Subjects with *XPF* 673 C>T (*M*) showed highest risk toward the lung cancer (OR: 1.60; 95% CI: 1.18–2.37; $p = 0.003$). Whereas node 2 carrying individuals with *XPF* 11985 A>G (*W*)/*XPF* Arg415Gln (*W*)/*XPF* 673 C>T (*W*) genotype possessed a significantly increased risk for the disease (OR: 1.44; 95% CI: 1.00–2.08; $p = 0.04$). Node 4 has individuals with a combined genotype of *XPF* *G415C* (*M*) and 673 C>T (*W*) showed protective effect for the lung cancer (OR: 0.65; 95% CI: 0.19–2.24; $p = 0.50$).

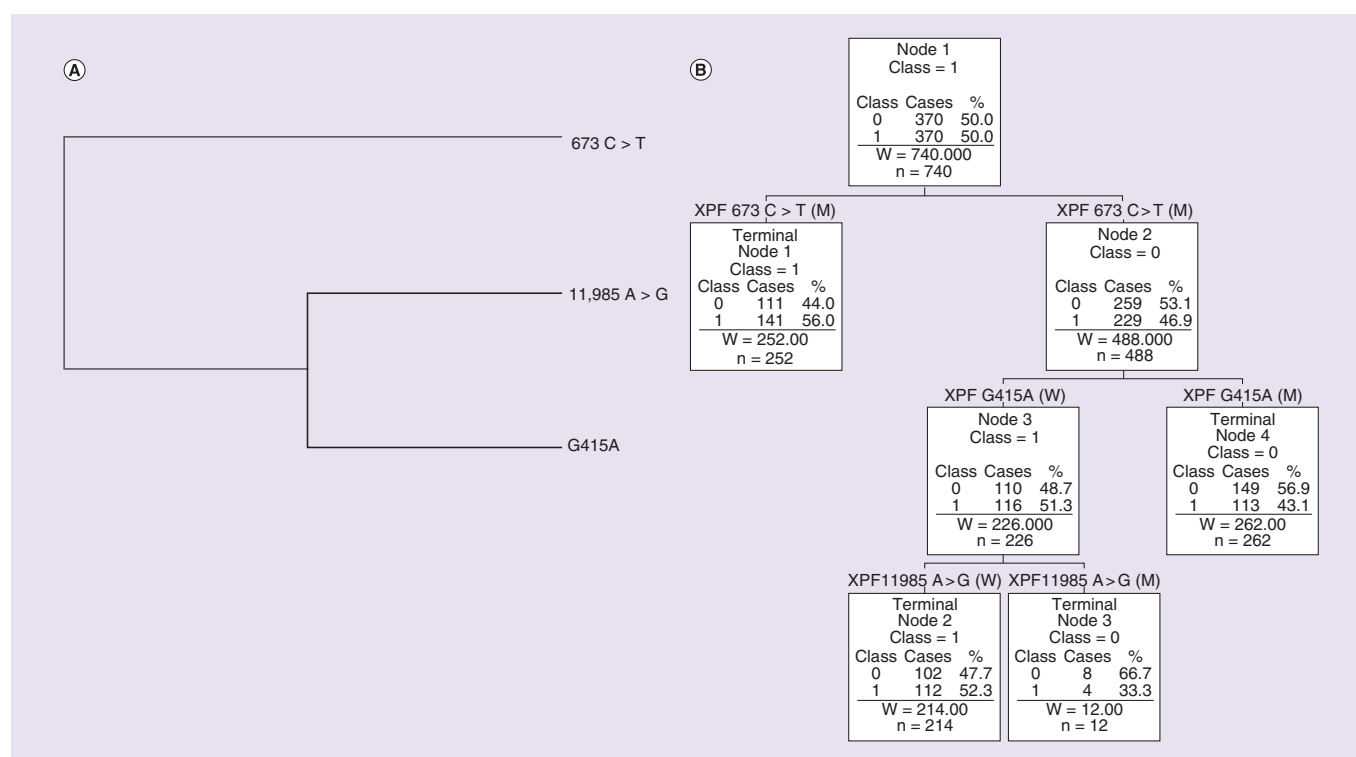


Figure 2. Interaction dendrogram and CART analysis for XPF 673 C>T, 11985 A>G and G415A polymorphisms. (A) Interaction dendrogram gained from the MDR showing gene-gene interaction for XPF 673 C>T, 11985 A>G and G415A. (B) CART analysis depicting high order interaction among three polymorphisms of XPF.

CART: Classification and regression tree; MDR: Multifactor dimensionality reduction.

Association of XPF polymorphism & clinical outcomes

To evaluate the clinical response in patients treated with platinum-based chemotherapy and its association with XPF polymorphism, the analysis using multivariate logistic regression adjusted for factors such as age, gender, smoking, stage, histology, KPS and ECOG was applied. The 202 lung cancer cases were divided as responders (complete response + partial response) and nonresponders (stable disease + progressive disease). No significant association was concluded between XPF polymorphism and clinical response as shown in Supplementary Table 3.

Discussion

In this study, we investigated association of three XPF SNPs (673 C>T, 11985 A>G, 415 G>A) with lung cancer risk, overall survival and its clinical response with different chemotherapy regimens in North Indian patients. Our data shows that the subjects with variant 673 C>T genotype was found to be associated with increased lung cancer risk compared with wild-type genotype. For 11985 A>G, rs254942 and G415A, rs1800067 polymorphism there was no significant difference observed between lung cancer cases and controls.

Very few studies have been reported on the role of the SNPs in XPF gene and their association with different cancers. A study on Caucasian population found a strong association of 673 C>T, rs316028 polymorphism toward increased risk for bladder cancer and was concordant with the present study [16]. On the contrary a few studies as reported by Yu *et al.* [17] and Shao *et al.* [18] illustrated association of 673 C>T, rs316028 polymorphism toward decreased lung cancer risk and were in disagreement with the present study. Another study by Liu *et al.* [11] negated any association of 673 C>T, rs316028 variant toward esophageal cancer risk and thus did not support this study. This discrepancy in results might be due to difference in genetic background among different study populations. To our knowledge, no prior data on 11985 A>G, rs254942 polymorphism and its role in cancer risk has been reported. Only a single study by Liu *et al.* [11] suggested no role of 11985 A>G, rs254942 polymorphism toward lung cancer risk which was in accordance with our results.

The most studied polymorphism of *XPF* gene is *G415A*, rs1800067. The present study suggested no significant relationship between *G415A*, rs1800067 genotypes and lung cancer was observed. Krupa *et al.* [10] were in agreement with our results as no significant association was observed between *G415A*, rs1800067 polymorphism and larynx cancer in Polish population. A study on Spanish [19] and American [20] populations have reported inconsistent association of *G415A*, rs1800067 polymorphism with breast cancer risk and support our findings. The study conducted by Shi *et al.* [16] deviates from the present study and suggests strong association of *G415A*, rs1800067 polymorphism toward increased lung cancer risk. There was no significant relationship derived for *XPF* variants and lung cancer risk in the present study.

Many previous studies have revealed a significant gene–environment interaction resulting in pathogenesis of lung cancer. NER pathway is important in modulating lung cancer susceptibility [21]. NER is the primary mechanism for repairing helix distorting bulky DNA adducts which are generated due to cigarette smoke [22,23]. A NER protein plays an important role in repairing large DNA lesions and adducts [24]. Genetic variants present in the promoter region of *XPF* might lead to substrate activation caused due to cigarette smoke and might contribute to different susceptibility to cancers. In the present study, a protective effect was observed in *G415A*, rs1800067 polymorphism amongst non-smokers. There was no significant association observed among smokers for *XPF* polymorphisms and lung cancer risk except for *673C>T*, rs316028 polymorphism suggesting an increased lung cancer risk in heavy smokers ($p = 0.03$). A study conducted by Yu *et al.* [17] emphasized on the fact that there was no interaction between *673 C>T*, rs316028 polymorphism and smoking status while Shao *et al.* [18] showed that *XPF* genotypic variants were associated with increased lung cancer risk among smokers. Hence, both the above studies were in disagreement with the present study. A study by Liu *et al.* [11] suggested a decreased risk of esophageal cancer among non-smokers of *XPF 673 C>T*, rs316028 polymorphism and therefore vary from our findings.

XPF is an essential protein in NER pathway which is necessary for removing platinum-based compounds. *XPF* binds with the DNA strands and after forming a complex with *ERCC1*, it takes part in the DNA repair pathway [25]. On considering the crucial rate limiting factor in NER, *XPF* genes lower expression might act as a genetic susceptibility factor in cancer development and a prognosis risk factor after chemotherapy in several cancers such as lung cancer [25]. Therefore, in the present study, we evaluated role of *XPF* polymorphism in the survival of lung cancer patients treated with platinum-based chemotherapy. There was no significant association between *XPF* polymorphic variants and OS. ADCC patients with *673 C>T*, rs316028 polymorphism were associated with low-HR with a survival period of 14 months in patients with mutant genotype (TT). It can be predicted that *11985 A>G* polymorphism was associated with poor survival in SCLC patients with minimum survival time of 4 months for each genotypic variant that is heterozygous (AG), mutant (GG) and heterozygous + mutant (AG + GG) type against the reference genotype (AA; MST = 11.6 months). ADCC patients showing *G415A*, rs1800067 polymorphism and having heterozygous genotype (GA) were associated with a moderated survival of 8.03 months.

DNA repair genes play a pivotal role in carcinogenesis and platinum function [26]. DNA repair can be termed as a double edged sword because a decrease DNA repair is associated with increased risk of developing cancer, although it might improve survival in patients which were priority diagnosed with cancer when treated with platinum-based drugs. A shorter survival period of 4.30 months was observed in patients with heterozygous genotype (GA) and treated with irinotecan–cisplatin/irinotecan–carboplatin regimen for *11985 A>G*, rs254942 polymorphism. A higher death rate was observed in all the genotypic variants of *11985 A>G* (rs254942) polymorphism treated with the above regimen for in other words, heterozygous (HR = 2.66), mutant (HR = 4.73) and heterozygous + mutant (HR = 2.64). Interpretation regarding results could be a complicated issue as it might emphasize the fact that genotype conferred decreased DNA repair capacity, reaching at any particular conclusion might be immature. These polymorphism may not be functional but could be a marker of a specific haplotypes of section of chromosome 16, where other DNA repair genes also reside. These haplotypes might be associated with the differential activity because of variations in DNA repair genes [27]. This study might generate necessary data to move forward to decide which chemotherapy could be individualized as per the genetic constitution of a patient. The careful consideration of the repair pathway involved in the metabolism of commonly used chemotherapy drugs would be the most critical part and provides the basis for treatment of advanced lung cancer cases which includes combination of cytotoxic agents. We can develop a much clearer understanding of genetic variations among individuals and their role in defining treatment tolerance as well as response and its overall outcome.

As *XPF* is an essential part of the NER pathway, many previous studies have suggested its role in recombination repair, mismatch repair, immunoglobulin class switching but its main function is to identify damaged sites in NER

pathway [28]. It has been believed that *XPF* polymorphism might influence chemotherapy by modulating its function [29]. In the present study, there was no significant association observed between *XPF* variants and clinical response. A study reported on the British population showing association of *XPF* polymorphism with platinum-based chemotherapy response suggesting no significant association with the clinical response [30]. A study by Shi *et al.* [31] and a study by Huang *et al.* [32] were also in accordance with the present study suggesting no significance between *XPF* polymorphisms and clinical outcome in non-small-cell lung cancer patients. They were not able to identify any relationship for *XPF* polymorphism as the predictor of response.

The present study has few limitations: the age and smoking status differences between cases and controls in the present study showed a statistically significant difference in cases and controls. Age and smoking status both are important factors which are related with the lung cancer susceptibility. Differences in the distribution of the age groups of the smokers and the duration of regular heavy smoking contribute to the quantitative variations causing age-specific lung cancer risks.

Conclusion

This case-control study indicated that 673 C>T, rs316028 polymorphism might be associated with increased lung cancer risk. 11985 A>G, rs254942 and G415A, rs1800067 polymorphism does not show any association with lung cancer risk. The genetic variants of 673 C>T, rs316028 polymorphism were associated with a better survival in ADCC patients while 11985 A>G, rs254942 polymorphism was associated with poor survival in SCLC patients. Lung cancer patients treated with irinotecan-cisplatin/irinotecan-carboplatin regimen indicated their association with shorter survival time and high risk of death in 11985 A>G, rs254942 polymorphism.

Future perspective

To our knowledge, this is the first study conducted on North Indian population suggesting role of *XPF* polymorphism with lung cancer, OS and clinical response. Some remarkable results were observed depicting association of *XPF* gene with lung cancer risk. The present study has emphasized upon association of 673 C>T, rs316028 polymorphism with strong risk of causing lung cancer. Lung cancer patients treated with different chemotherapy regimens such as irinotecan-cisplatin/irinotecan-carboplatin regimen and their association with *XPF* polymorphism marked its association with poor survival in lung cancer patients with 11985 A>G, rs254942 polymorphism. It is known that selection of chemotherapeutic regimen as per the molecular marker could improve response rate and hence overall survival of cancer patients. This study in future can be regarded to facilitate assessment of genetic variations of *ERCC4* and help patients to make therapeutic decisions for individualized therapy in advanced lung cancer patients. As the present study is performed on a small sample size, for further authentication of the results we need large sample studies that can confirm functional significance of *XPF* polymorphism and its role in cancer susceptibility, OS and clinical outcomes. The present study can be used as reference for those large sample studies.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/fon-2017-0569

Acknowledgements

The authors would like to express their gratitude to all the subjects who participated in this current study.

Financial & competing interests disclosure

This work was supported by grant from the Indian Council of Medical Research, New Delhi, India. (Grant No. 5/13/126/2011/NCD-III). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Summary points

- The findings of the present study suggest that heterozygous genotype ($p = 0.03$) poses an increased risk for *XPF* 673 C>T (rs316028) polymorphism.
- When stratified according to histological subtypes, heterozygous genotype (CT) showed a 1.5-fold risk for squamous cell carcinoma ($p = 0.07$). 11985 A>G (rs254942) and G415A (rs1800067) polymorphism did not show any significant association with lung cancer susceptibility.
- 673 C>T (rs316028) genotypes showed twofold risk in individuals carrying heterozygous and mutant genotype (CT + TT) and were heavy smokers ($p = 0.03$).
- For G415A (rs1800067) polymorphism a significant protective effect among non-smokers carrying heterozygous genotype (GA) was observed when compared with the wild genotype (GG) ($p = 0.04$). Combined genotype (GA + AA) also showed a protective effect for non-smokers with G415A, rs1800067 polymorphism ($p = 0.03$).
- 673 C>T (rs316028) and 11985 A>G (rs254942) polymorphic variants with heterozygous carriers (CT + AG) have almost twofold lung cancer risk ($p = 0.02$). When double variant genotypes were combined together as a single genotype [(CT + TT) + (AG + AA)] for 673 C>T and 11985 A>G, rs254942 variants also showed a high risk for the disease ($p = 0.04$).
- When analyzed on the basis of lung cancer histological subtypes and evaluated their overall survival (OS), 673 C>T (rs316028) polymorphism showed adenocarcinoma (ADCC) patients with heterozygous (CT) genotype had a better OS ($p = 0.01$).
- For 11985 A>G, rs254942 polymorphism, it was observed that small-cell lung carcinoma patients carrying single copy of the variant allele (AG) had a poor OS as compare with subjects with the wild-type genotype (AA) (MST = 4.0 vs 11.6, log-rank $p = 0.001$). Such patients also had a very poor prognosis as the death ratio associated was high (HR = 3.19, $p = 0.001$).
- In case of the *XPF* G415A (rs1800067), ADCC patients with heterozygous genotype (GA) showed a 1.5-fold hazard ratio with the moderate survival amongst cases ($p = 0.02$).
- For 11985 A>G (rs254942) polymorphic variant, lung cancer subjects who were harboring the heterozygous genotype (AG) and treated with irinotecan–cisplatin/irinotecan–carboplatin regimen exhibited a high mortality risk (HR: 2.66; 95% CI: 1.45–4.88; $p = 0.001$; MST = 4.30). Furthermore, patients having combined genotype (AG + GG) and treated with irinotecan–cisplatin/irinotecan–carboplatin regimen depicted almost twofold hazard ratio when compared with subjects with wild genotype (HR: 2.64; 95% CI: 1.46–4.78; $p = 0.001$).
- *XPF* G415A (rs1800067) with cross validation consistency of 100/100 and prediction error of 0.47 ($p = 0.06$) analyzed as the best one factor model associated with lung cancer risk. Classification and regression tree analysis depicted 673 C>T (rs316028) as the highest risk group associated toward lung cancer susceptibility.
- In conclusion, our data provides statistical evidence that 673 C>T (rs316028) polymorphism might be associated with increased lung cancer risk. The genetic variants of 673 C>T (rs316028) polymorphism were associated with a better survival in ADCC patients. Patients treated with irinotecan–cisplatin/irinotecan–carboplatin regimen indicated their association with shorter survival time and high risk of death in 11985 A>G, rs254942 polymorphism.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Malik PS, Raina V. Lung cancer: prevalent trends & emerging concepts. *Indian J. Med. Res.* 141(1), 5–7 (2015).
- **It is an important study explaining prevailing lung cancer trends in Indian population.**
2. Mohan A, Latifi AN, Guleria R. Increasing incidence of adenocarcinoma lung in India: following the global trend? *Indian J. Cancer* 53(1), 92–95 (2016).
3. Johnson-Thompson MC, Guthrie J. Ongoing research to identify environmental risk factors in breast carcinoma. *Cancer (Phila.)* 88, 1224–1229 (2000).
4. Petrusseva IO, Evdokimov AN, Lavrik AN. Molecular mechanism of global genome nucleotide excision repair. *Acta Naturae* 6(1), 23–24 (2014).
5. Liu P, Siciliano J, White B *et al.* Regional mapping of human DNA excision repair gene ERCC4 to chromosome 16p13.13–p13.2. *Mutagenesis* 8, 199–205 (1993).
6. Tsodikov OV, Enzlin JH, Schärer OD *et al.* Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc. Natl Acad. Sci.* 102, 11236–11241 (2005)
- **Gives a detail insight of the structure as well as function of XPF genes.**
7. Ahmad A, Robinson AR, Duensing *et al.* ERCC1–XPF endonuclease facilitates DNA double-strand break repair. *Mol. Cell Biol.* 28, 5082–5092 (2008).
8. Shen MR, Jones IM, Mohnenweiser H. Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res.* 58, 604–608 (1998).

9. Smith TR, Levine EA, Freimanis RI *et al.* Polygenic model of DNA repair genetic polymorphisms in human breast cancer risk. *Carcinogenesis* 29, 2132–2138 (2008).
10. Krupa R, Kasznicki J, Gajęcka M *et al.* Polymorphism of the DNA polymorphism of the DNA repair genes XRCC1 and ERCC4 are not associated with smoking and drinking dependent larynx cancer in a polish population. *Exp. Oncol.* 33(1), 55–56 (2001).
11. Liu Y, Cao L, Chang J *et al.* XPF-673C > T polymorphism effect on the susceptibility to esophageal cancer in Chinese population. *PLoS ONE* 9(4), e94136 (2015).
- **This is the only study which has reported association of 673 C>T and 11985 A>G polymorphism with lung cancer. The primers for XPF 673 C>T and 11985 A>G polymorphism are taken from this study.**
12. Wood RD, Mitchell M, Sgouros J *et al.* Human DNA repair genes. *Science* 291, 1284–1289 (2001).
13. Fleming ND, Agadjanian H, Nassanian H *et al.* Xerodermapigmentosum complementation group C single-nucleotide polymorphisms in the nucleotide excision repair pathway correlate with prolonged progression-free survival in advanced ovarian cancer. *Cancer* 118, 689–697 (2012).
14. Sodhi KK, Bahl C, Singh N, Behera D, Sharma S. Functional genetic variants in pre-miR-146a and 196a2 genes are associated with risk of lung cancer in North Indians. *Future Oncol.* 11(15), 2159–2173 (2015).
15. Lavanya J, Kumar VJ, Sudhakar N, Prathap S. Analysis of DNA repair genetic polymorphism in breast cancer population. *Int. J. Pharm. Bio. Sci.* 6(3), 966–973 (2015).
- **The primers for XPF G415A polymorphism are taken from this study. A detailed mode of mechanism regarding G415A has been mentioned in this study.**
16. Shi TY, He J, Qiu LX *et al.* Association between XPF polymorphisms and cancer risk: a meta-analysis. *PLoS ONE* 7, e38606 (2012).
- **Explains the mechanism of action of XPF genes. It also gives detail knowledge of interactive effect of smoking with XPF genes and its association with cancer susceptibility. The study also defines association of XPF polymorphisms and their association with lung cancer histological subtypes.**
17. Yu DK, Wu C, Tan W, Lin DX. Functional XPF polymorphisms associated with lung cancer susceptibility in a Chinese population. *Front. Med. China* 1, 82–89 (2010).
18. Shao M, Ma H, Wang Y *et al.* Polymorphisms in excision repair cross-complementing group 4 (ERCC4) and susceptibility to primary lung cancer in a Chinese Han population. *Lung Cancer* 60, 332–339 (2008).
19. Milne RL, Ribas G, Gonzalez-Neira A *et al.* ERCC4 associated with breast cancer risk: a two-stage case–control study using high-throughput genotyping. *Cancer Res.* 66, 9420–9427 (2006).
20. Rajaraman P, Bhatti P, Doody MM *et al.* Nucleotide excision repair polymorphisms may modify ionizing radiation-related breast cancer risk in US radiologic technologists. *Int. J. Cancer* 123, 2713–2716 (2008).
21. Sancar A, Lindsey-Boltz LA, Unsal-Kaçmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* 73, 39–85 (2004).
22. DeMarini DM, Shelton ML, Levine JG. Mutation spectrum of cigarette smoke condensate in *Salmonella*: comparison to mutations in smoking-associated tumors. *Carcinogenesis* 16, 2535–2542 (1995).
23. Asami S, Manabe H, Miyake J *et al.* Cigarette smoking induces an increase in oxidative DNA damage, 8-hydroxydeoxyguanosine, in a central site of the human lung. *Carcinogenesis* 18, 1763–1766 (1997).
24. Torres-Ramos CA, Johnson RE, Prakash L, Prakash S. Evidence for the involvement of nucleotide excision repair in the removal of a basic sites in yeast. *Mol. Cell Biol.* 20, 3522–3528 (2000).
25. Alexander BM, Wang XZ, Niemierko A *et al.* DNA repair biomarkers predict response to neo-adjuvant chemoradiotherapy in esophageal cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 83, 164–171 (2012).
26. Tsuruo T, Matsuzaki T, Matsushita M *et al.* Antitumor effect of CPT-11, a new derivative of camptothecin, against pleiotropic drug resistant tumors *in vitro* and *in vivo*. *Cancer Chemother. Pharmacol.* 21, 71–74 (1988).
27. Fukuda M, Oka M, Soda H *et al.* Phase II study of irinotecan combined with carboplatin in previously untreated non-small-cell lung cancer. *Cancer Chemother. Pharmacol.* 154, 73–577 (2004).
- **This is the first study which depicts association between a gene polymorphism and overall survival in patients treated with irinotecan–cisplatin/irinotecan–carboplatin regimen. It gives detail description of mechanism of irinotecan with DNA repair genes and also explains its impact on lung cancer patients when infused in combination with the platinum-based drug.**
28. Kornguth DG, Garden AS, Zheng Y *et al.* Gastrostomy in oropharyngeal cancer patients with ERCC4 (XPF) germline variants. *Int. J. Radiat. Oncol. Biol. Phys.* 62, 665–671 (2005).
29. DeLoia JA, Bhagwat NR, Darcy KM *et al.* Comparison of ERCC1/XPF genetic variation, mRNA and protein levels in women with advanced stage ovarian cancer treated with intraperitoneal platinum. *Gynecol. Oncol.* 126, 448–454 (2012).
30. Fareed KR, Al-Attar A, Soomro IN *et al.* Tumour regression and ERCC1 nuclear protein expression predict clinical outcome in patients with gastro-oesophageal cancer treated with neoadjuvant chemotherapy. *Br. J. Cancer* 102, 1600–1607 (2012).
- **One of the important study depicting role of XPF polymorphism and its association with clinical outcomes.**

31. Shi ZH, Shi GY, Liu LG. Polymorphisms in *ERCC1* and *XPF* gene and response to chemotherapy and overall survival of non-small cell lung cancer. *Int. J. Clin. Exp. Pathol.* 8(3), 3132–3137 (2015).
- **Important study illustrating association between *XPF* polymorphism and clinical outcomes in non-small-cell lung cancer patients.**
32. Huang SJ, Shi GY, Liu LG. Role of *ERCC1* variants in response to chemotherapy and clinical outcome of advanced non-small cell lung cancer. *Tumor Biol.* 35(5), 4023–4029 (2014).

Association of *XPA* Polymorphisms Towards Lung Cancer Susceptibility and its Predictive Role in Overall Survival of North Indians

Shweta Lawania¹ · Navneet Singh² · Digambar Behera² · Siddharth Sharma¹ 

Received: 25 November 2017 / Accepted: 20 February 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract The present study investigated the role of Xeroderma pigmentosum group A (*XPA*) polymorphism (A23G and G709A) with lung cancer risk and its association with overall survival in North Indians. 370 cases and 370 controls were investigated to evaluate association between *XPA* polymorphism (A23G and G709A) with lung cancer risk using logistic regression analysis. A follow-up study was also conducted for 291 lung cancer cases illustrating correlation between overall survival in lung cancer patients and *XPA* variants. GG genotype showed an increased lung cancer risk ($p = 0.0007$) for A23G polymorphism whereas G709A polymorphism was associated with significant protective effect in heterozygous (AG) subjects ($p = 0.001$). When stratified according to smoking status an increased risk for lung cancer was observed for GG genotype in A23G polymorphism ($p = 0.0002$). A poor survival in females carrying variant genotype (GG) was observed ($p = 0.001$; MST = 4.16 months) for A23G polymorphism. Adenocarcinoma patients with heterozygous genotype showed an increased hazard ratio ($p = 0.02$) for A23G polymorphism. G709A was associated with a reduced hazard ratio marking a better survival among mutant females (HR 0.17; $p = 0.05$; MST = 18.63 months). It can be concluded that A23G polymorphism might contribute to increased lung cancer risk in North Indian population emphasizing on poor survival among females. G709A polymorphism might result in protective effect in lung cancer subjects. The present

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10528-018-9850-0>) contains supplementary material, which is available to authorized users.

✉ Siddharth Sharma
siddharthsharma.phd@thapar.edu

¹ Department of Biotechnology, Thapar University, Patiala, Punjab 147002, India

² Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research (PGIMER), Sector 14, Chandigarh, India

study had a low sample size but it could act as reference for the large sample studies in future.

Keywords Xeroderma pigmentosum A · Lung cancer · Overall survival · Platinum-based chemotherapy

Introduction

Lung cancer is considered as the most commonly diagnosed malignancy in terms of external carcinogen exposure leading to cancer-related deaths globally (Jemal et al. 2011). The survival in lung cancer patients has improved slowly but prognosis still remains poor (only 10% in terms of 5-year survival rate) (Sant et al. 2009). The relative risk for lung cancer is 20-fold higher among smokers than never smokers (Khuder 2001). In India, 90% of the lung cancer cases are attributed due to smoking but only a minor fraction (< 15%) actually develops the disease as genetic constitution of an individual plays an important role in defining its susceptibility (Behera 2012).

Tobacco smoke which consist of carcinogenic substances such as hydrocarbons, heterocyclic amines, and nitrosamine induces DNA adduct formation in an individual when exposed to these xenobiotic substances (Khuder 2001). These genotoxins cause mutations which ultimately lead to cancer formation. These bulky, helix-distorting adducts which are primarily induced by chemical carcinogen present in cigarette smoke are repaired by Nucleotide excision repair (NER) pathway (Wood et al. 2001). The human NER pathway eliminates the damaged DNA. Hence, restoring DNA by safeguarding its genomic integrity which is essential for the survival of cell and therefore prevents cancer (Batty and Wood 2000). The Xeroderma pigmentosum group A (*XPA*) gene maps at chromosome 9 is a zinc finger DNA-binding protein associated with excision of wide range of lesions such as benzo[α]pyrene-induced DNA bulky adducts (Sancer et al. 2004). *XPA* protein mediates the recognition of DNA damage by interacting with replication protein A, transcription factor II H, and excision repair cross-complementing group 1–xeroderma pigmentosum group F (Wu et al. 2003). The *XPA* gene is involved in both global genome and transcription-coupled pathways, hence, interacts with many repair factors which take part in DNA repair process (Wu et al. 2003). Therefore, any variation that would occur in *XPA* gene has potential to influence the protein function which might subsequently affect the DNA repair capacity (DRC). *XPA* gene has the ability to interact with the repair factors, therefore, considering its strong affinity towards the damaged DNA, it is likely to play a central role in the correct positioning of the repair machinery around the lesion. *XPA* plays an important role in studying tumor genesis and chemotherapy as it might act as a rate-limiting factor that regulates the entire activity of nucleotide excision repair (Kang et al. 2011).

The most commonly studied *XPA* polymorphism is (A $\rightarrow$ G) present at position 23 in the 5' non-coding region with four nucleotides present in the upstream of the start codon of the gene (Butkiewicz et al. 2000). Numerous studies on *XPA*

A23G polymorphism have suggested the protective effect for the G-allele against the lung cancer risk and homozygous carriers of G-allele shows 13% higher DNA repair capacity than A-allele carriers in Asian population (Wu et al. 2003). Although other studies on Caucasians (Butkiewicz et al. 2004), East Asian (Wu et al. 2003), and American African (Wu et al. 2003) population differ from these studies suggesting association of variant genotype with increased risk for lung cancer. The studies conducted on the association of *XPA* A23G polymorphism with cancer are reported with contradictorily different results. The other known polymorphism of *XPA* is codon 228 present in exon 6 associated with *Arg228Gln* (G709A, rs1805160) substitution (Butkiewicz et al. 2000). The functional significance of these polymorphism is still unknown (Hirschhorn et al. 2002).

Therefore, in the present study we conducted a case–control study to investigate the association of two polymorphic sites in the *XPA* gene towards the lung cancer and its histological subtypes. We further evaluated the role of both these SNPs towards overall survival and response to platinum-based chemotherapy treatment.

Materials and Methods

Study Population

We conducted a hospital-based case–control study which included 370 lung cancer cases and 370 healthy controls registered at Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh. The Institutional Ethics Committee board of PGIMER, Chandigarh had reviewed and approved the study. The current study included patients without any age, gender, smoking, histology, or TNM stage restrictions but patients with any previous cancer history were avoided. Healthy volunteers were randomly selected who visited hospital for the routine check-up and were not diagnosed with cancer. A written consent from each subject was obtained. A detailed questionnaire for each case and control was completed by trained interviewer regarding demographic and lifetime smoking history. Information regarding histology, TNM classification, clinical staging, and chemotherapeutic regimen were collected from medical records of the patients from the hospital. The follow-up data of patients were telephonically obtained using information mentioned in their clinical records.

Follow-Up and Response Assessment

A total of 291 lung cancer subjects were followed up for survival through telephonic conversations. Every 2 months until death or the end of the study period patients were called and inquired about health. Primarily we assessed overall survival and for that we considered survival time from the date of diagnosis till the date of last follow-up or death due to other reasons.

Genotyping of *XPA* Variants

Genomic DNA was isolated from the 5 ml peripheral blood lymphocytes obtained from subjects using proteinase-k digestion and phenol–chloroform extraction protocol by Sodhi et al. 2015. *XPA* polymorphism was detected using PCR–RFLP assay as prescribed in the previous studies (Park et al. 2002) (Fig. 1b). For both polymorphisms, 15% samples were repeated using same procedure and 100% concordant results were obtained.

Statistical Analysis

Chi-square test (χ^2 test) was applied to compare cases and controls for certain demographic characteristics such as age, sex, smoking status, and pack years smoked whereas student's *t* test was used for continuous variables. The Hardy–Weinberg equilibrium (HWE) was tested to compare observed and expected genotypic frequencies among controls and lung cancer patients using goodness of fit χ^2 test. Pearson χ^2 test was applied to evaluate significant difference in allelic and genotypic frequencies of *XPA* between cases and controls. Odds ratio (OR) and their 95% confidence interval (CI) adjusted for age, sex, and smoking status were calculated using Multivariate logistic regression. Haplotype analysis was performed for both the SNPs using SHEsis software (<http://202.120.31.177/myAnalysis.php>). Univariate Kaplan–Meier method was applied to evaluate the association between overall survival (OS) and logrank test. Multivariate Cox regression model adjusted for age, gender, smoking, stage, tumor histology, Karnofsky performance status (KPS), and eastern cooperative oncology group (ECOG) was applied to estimate adjusted hazard ratio (HR) and their 95% confidence interval (CI). The Cox regression model

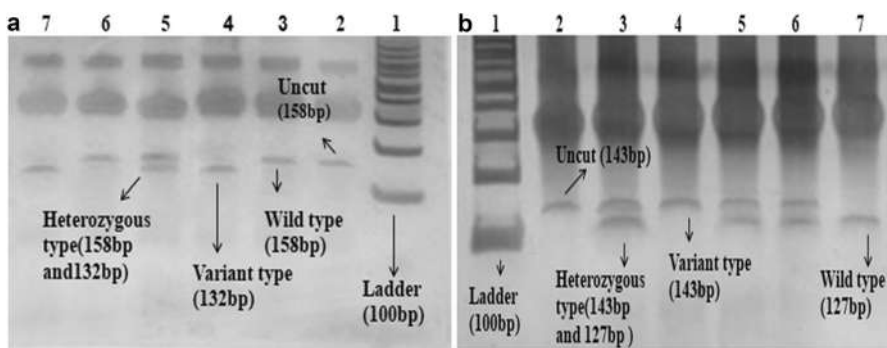


Fig. 1 Restriction digestion pattern of *XPA* variants **a** A23G polymorphism (wild type (AA): 158 bp; variant type (GG): 132 and 26 bp (the 26 bp product was too small to resolve accurately); heterozygous type (AG): 158, 132, and 26 bp. Lane 1: marker; Lane 2: Uncut; Lane 3, 6: AA genotype; Lane 4, 7: GG genotype; Lane 5: AG genotype) **b** G709A polymorphism (wild type (GG): 127 and 16 bp (16 bp product was too small to resolve accurately); variant type (AA): 143 bp; heterozygous type (GA): 143, 127, and 16 bp. Lane 1: marker; Lane 2: Uncut; Lane 3, 5, 6: GA genotype; Lane 4: AA genotype; Lane 7: GG genotype)

was also used as a predictive factor for prognosis of cancer with $p < 0.05$ as significant value.

Results

Demographic Characteristics

Demographic characteristics depicting the role of *XPA* polymorphism and its association with lung cancer risk have been summarized in Table 1. Mean age of cases and control were 58.11 ± 10.44 (range 29–86) and 58.83 ± 10.18 (range 22–83). There were 319 (86.21%) males and 51 (13.78%) females in both diseased group and healthy controls ($p = 0.91$). In the present study, a high prevalence among smokers was observed in cases as compared to controls (81.89 vs. 73.24%, $p = 0.006$). When stratified in terms of pack years a significantly higher pack years was observed to be associated for cases when compared with controls (24.97 ± 35.04 vs. 17.88 ± 19.45 , $p < 0.0001$). When classified on the basis of histological subtypes, out of 370 cases, 135 (36.48%) were having squamous cell carcinoma (SQCC), 120 (32.43%) were diagnosed with adenocarcinoma (ADCC), and 111 (30%) with small cell lung carcinoma (SCLC).

XPA Variants and Lung Cancer Susceptibility

The allelic and genotypic distribution of two *XPA* variants, i.e., *A23G* and codon 228 are shown in Table 2. The controls for both the SNPs i.e., *XPA A23G* ($N = 370$, $\chi^2 = 2.84$, $df = 1$, $p = 0.09$) and codon 228 polymorphism ($N = 370$, $\chi^2 = 2.55$, $df = 1$, $p = 0.1$) were in accordance with HWE. Table 2 shows the association of lung cancer risk with *XPA* polymorphic variants. Adjusted OR's and 95% CI were calculated using logistic regression analysis. For both the *XPA* polymorphism the homozygous wild genotype was regarded as the reference group with which variant genotypes were compared. Regression analysis showed that subjects with variant genotype (GG) for *A23G* polymorphism possessed a 2.5-fold elevated risk for the disease (OR 2.56; 95% CI 1.49–4.39; $p = 0.0007$). When both heterozygous and variant genotypes were combined and used as a single genotype (AG+GG), a slight risk was observed (OR 1.37; 95% CI 1.02–1.35; $p = 0.03$). The study group was further comprehended into histological subgroups to evaluate severity of risk towards the lung cancer. The subjects carrying double variant allele (GG) for *A23G* polymorphism were at higher risk of developing lung cancer in all the subgroups. A significant twofold risk was observed for individuals with SQCC (OR 2.58; $p = 0.01$) whereas a significant threefold risk for SCLC patients was observed (OR 3.76; $p = 0.0007$). For *XPA* codon 228 polymorphism, the heterozygous genotype (GA) showed a highly significant protective effect (OR 0.54; 95% CI 0.38–0.78; $p = 0.001$). When both heterozygous and variant genotype combined as a single genotype (GA+AA) and compared with the reference genotype (GG), a significant protective effect was observed (OR 0.64; 95% CI 0.47–0.87; $p = 0.004$). When

Table 1 Demographic characteristics among cases and control patients in North Indian population

Variable	Cases [<i>n</i> (%)] <i>N</i> = 370	Controls [<i>n</i> (%)] <i>N</i> = 370	<i>p</i> ^a value
Age (years)			
Mean ± SD	58.11 ± 10.44	53.83 ± 10.18	< 0.0001
Range			
Gender			
Male	319 (86.21%)	319 (86.21%)	0.91
Female	51 (13.78%)	51 (13.78%)	
Smoking status			
Smokers	303 (81.89%)	271 (73.24%)	0.0063
Non-smokers	67 (18.10%)	99 (26.75%)	
Pack years			
Mean ± SD	24.97 ± 35.04	17.88 ± 19.45	< 0.0001
Histological types (<i>N</i> = 370)			
SQCC	135 (36.48%)		
ADCC	120 (32.43%)		
SCLC	111 (30%)		
Others	4 (1.08%)		
Unknown	0 (0.0)		
Tumor size (<i>N</i> = 370)			
Tx	11 (2.97%)		
T1	16 (4.32%)		
T2	42 (11.35%)		
T3	85 (22.97%)		
T4	191 (51.62%)		
Unknown	25 (6.75%)		
Metastasis (<i>N</i> = 370)			
<i>M</i> ₀	187 (50.54%)		
<i>M</i> ₁	158 (42.70%)		
Unknown	25 (6.75%)		
TNM staging (<i>N</i> = 370)			
I	3 (0.81%)		
II	10 (2.70%)		
III	175 (45.94%)		
IV	163 (44.05%)		
Unclassified	19 (5.13%)		
Lymph node involvement (<i>N</i> = 370)			
<i>N</i> ₀	46 (12.43%)		
<i>N</i> ₁	36 (9.72%)		
<i>N</i> ₂	157 (42.43%)		
<i>N</i> ₃	106 (28.64%)		
Unknown	25 (6.75%)		
Regimen (<i>N</i> = 291)			
Docetaxel–cisplatin/carboplatin	70 (24.05%)		

Table 1 (continued)

Variable	Cases [n (%)] N = 370	Controls [n (%)] N = 370	<i>p</i> ^a value
Irinotecan–cisplatin/carboplatin	65 (22.33%)		
Pemetrexed–cisplatin/carboplatin	64 (21.99%)		
Others/unknown	92 (31.61%)		

Highly significant values are given bold

SD standard deviation, *n* total number of case patients or control subjects

^a*p* values were derived from Pearson χ^2 square test except age; student *t* test was used for age. All *p*^a values are two-sided. *p* < 0.05 was considered statistically significant

stratified according to histology, individuals carrying heterozygous genotype (GA) showed a significant protective effect for SQCC (OR 0.45; 95% CI 0.28–0.73; *p* = 0.001) and SCLC (OR 0.60; 95% CI 0.37–0.97; *p* = 0.03).

Haplotype Analysis of XPA Variants

SNPs in the current study were in weak linkage disequilibrium for the study population (*D* = 0.081, *r*² = 0.001) (Table 3). Comparison of overall haplotypic profile showed a statistically significant distribution between cases and control group (Global χ^2 = 14.73, *p* = 0.002). The subjects carrying A–A haplotype showed a less pronounced protective effect (OR 0.71; 95% CI 0.55–0.92; *p* = 0.01). The haplotypes with both variant allele (G–G) were associated with increased risk of lung cancer (OR 1.42; 95% CI 1.12–1.80; *p* = 0.003). All other haplotypes did not show any significant effect.

Interactive Effect of Smoking on the XPA Polymorphism and Their Association with Lung Cancer Risk

The study group was divided into smoking and non-smoking group considering their smoking status (Table 4). The smokers were further divided into light and heavy smokers as per the number of pack years they smoked. The individual who had smoked ≤ 25 years were categorized as light smokers whereas individuals who has smoked > 25 years were defined as heavy smokers. For A23G polymorphism, it was observed that smokers with variant genotype (GG) and combined (AG+GG) exhibited a elevated risk for the lung cancer when compared to reference group (AA) for A23G polymorphism (OR 3.61; 95% CI 1.82–7.04; *p* = 0.0002; OR 1.40; 95% CI 0.99–1.98; *p* = 0.05). Whereas the subjects harboring the variant genotype (GG) and were light smokers expressed a sevenfold increased susceptibility (OR 6.78; 95% CI 2.36–19.49; *p* = 0.0004) towards lung cancer. In G709A polymorphism, the heterozygous genotype (GA) and combined genotype (GA+AA) displayed a protective effect towards the disease for subjects who were smokers as compared to non-smokers (OR 0.54; 95% CI 0.38–0.78; *p* = 0.001; OR 0.68; 95%

Table 2 Distribution of genotypes of XPA polymorphism depicting their association with Lung cancer risk and its histological subtypes

	Controls [n (%)] N = 370	Overall lung cancer			SQCC			ADCC			SCLC		
		Cases [N (%)] N = 370	Adjusted odds ratio (95% CI) ^a	P ^b	Cases [n (%)] N = 135	Adjusted odds ratio (95% CI) ^a	P ^b	Cases [n (%)] N = 120	Adjusted odds ratio (95% CI) ^a	P ^b	Cases [n (%)] N = 111	Adjusted odds ratio (95% CI) ^a	P ^b
XPA A23G													
AA	181 (48.91)	146 (39.45)	1.00 (Reference)		52 (38.51)	1.00 (Reference)		53 (44.16)	1.0 (Reference)		40 (36.03)	1.00 (Reference)	
AG	165 (44.59)	170 (45.94)	1.21 (0.88–1.65)	0.233	67 (49.62)	1.23 (0.79–1.92)	0.33	50 (41.66)	0.93 (0.59–1.46)	0.76	53 (47.74)	1.37 (0.84–2.24)	0.20
GG	24 (6.48)	54 (14.59)	2.56 (1.49–4.39)	0.0007	16 (11.85)	2.58 (1.17–5.70)	0.018	17 (14.16)	2.01 (0.97–4.16)	0.05	18 (16.21)	3.76 (1.75–8.09)	0.0007
AG+GG	189 (51.08)	224 (30.27)	1.37 (1.02–1.85)	0.03	83 (61.48)	1.36 (0.89–2.08)	0.147	67 (55.83)	1.24 (0.80–1.90)	0.32	71 (48.37)	1.61 (1.02–2.56)	0.04
A-Allele	527 (71.21)	462 (62.43)											
G-Allele	213 (28.78)	278 (37.56)											
MAF	0.29	0.38											
XPA G709A													
GG	177 (47.83)	211 (57.02)	1.00 (Reference)		82 (60.74)	1.00 (Reference)		62 (51.66)	1.00 (Reference)		63 (56.75)	1.00 (Reference)	
GA	167 (45.13)	121 (32.70)	0.54 (0.38–0.78)	0.001	37 (27.40)	0.45 (0.28–0.73)	0.001	46 (38.33)	0.74 (0.47–1.16)	0.19	38 (34.23)	0.60 (0.37–0.97)	0.03
AA	26 (7.02)	38 (10.27)	0.97 (0.51–1.86)	0.94	16 (11.85)	1.21 (0.58–2.49)	0.60	12 (10)	0.99 (0.44–2.19)	0.98	10 (9.00)	0.92 (0.40–2.11)	0.84
GA+AA	293 (79.18)	159 (42.97)	0.64 (0.47–0.87)	0.004	53 (39.25)	0.56 (0.36–0.86)	0.007	58 (48.33)	0.78 (0.51–1.20)	0.26	48 (43.24)	0.65 (0.41–1.02)	0.06
G-Allele	521 (70.40)	543 (73.37)											
A-Allele	219 (29.59)	197 (26.62)											
MAF	0.30	0.27											

Highly significant values are given bold

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 by logistic regression analysis after adjusting for age, gender, and smoking status

^b Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

CI 0.43–0.85, $p = 0.004$). A highly protective effect was also observed for heterozygous genotype (GA) for subjects who smoked > 25 pack years (OR 0.49; 95% CI 0.29–0.84; $p = 0.01$).

Relationship of Clinic-Pathological Features with *XPA* Polymorphic Variants

To evaluate association of two polymorphic variants of *XPA* (*A23G*, *G709A*) on various clinic-pathological features, we segregated patients according to clinical stage (III vs. IV), primary tumor extension (T3 + T4 vs. T0 + T1 + T2), and metastatic status (positive vs. negative) (Supplementary Table 1). No significant association was observed between tumor stage/metastasis and *XPA* polymorphic variants. However, for *A23G* variant, heterozygous genotype (AG) was highly represented in patients with T3 + T4 as compared to patients with T1 + T2 tumor stage, therefore, almost twofold risk was observed for subjects carrying heterozygous genotype (AG) having T3 + T4 tumor stage with T0 + T1 + T2 stage as a reference (OR 1.80; 95% CI 0.99–3.28; $p = 0.05$). No statistically significant association was observed for tumor staging and genotypes of *G709A* polymorphism.

Association of *XPA* SNPs with Overall Survival of Lung Cancer Patients

Overall survival analysis of 291 lung cancer patients was conducted as shown in Table 5. First, univariate analysis was performed using Kaplan–Meier curves and logrank test and then multivariate Cox hazard proportional analysis were performed to evaluate the association of *XPA* polymorphism with overall survival of patients. The distribution of observed genotypes in the patients agreed with those estimated from HWE ($\chi^2 = 0.06$; df = 1; $p = 0.79$) for *A23G* polymorphic variant. The homozygous wild type (AA (*A23G*) and GG (*G709A*) was considered as reference. For *A23G* polymorphism, heterozygous genotype (AG) was found to be associated with shorter survival period when compared with wild genotype (AA) (6.76 vs. 7.56 months, logrank test $P = 0.98$). Whereas for codon 228 polymorphism a better survival was observed in heterozygous genotype (GA) as compared to the reference genotype (GG) (8.20 vs. 6.43 months). No significant association was observed between *XPA* polymorphism and overall survival in both Kaplan–Meier analysis and Cox regression model. When study population was segregated on the basis of histological subtypes, no significant association was observed between *G709A* polymorphism but for *A23G* polymorphism, ADCC subjects with heterozygous genotype (AG) showed a significant hazard ratio (HR 5.61; 95% CI 1.28–24.46; $p = 0.02$) (Fig. 2a).

We also analyzed OS after classifying the lung cancer patients on the basis of their gender. In *XPA A23G*, the female subjects carrying variant genotype (GG) showed poor survival when compared with wild genotype (AA) (HR 4.35; 95% CI 0.63–29.79; logrank test $P = 0.002$, MST = 4.16 vs. 9.46 months) (Fig. 2b). After Cox regression analysis, a significant trend suggesting association of variant genotype (GG) with increased mortality in females was observed when compared to heterozygous + variant genotype (AG+GG) for *A23G* polymorphism (HR¹ 278.58;

Table 3 XPA haplotypes and their association with lung cancer risk

XPA A23G & XPAG709A	Case frequency [n (%)]	Control frequency [n (%)]	OR (95% CI) ^a	<i>p</i> ^b value
A–A	0.170 (126)	0.222 (164)	0.71 (0.55–0.92)	0.01
A–G	0.455 (336)	0.490 (363)	0.86 (0.70–1.06)	0.16
G–A	0.097 (71)	0.074 (55)	1.33 (0.92–1.92)	0.12
G–G	0.279 (207)	0.214 (158)	1.42 (1.12–1.80)	0.003

Highly significant values are given bold

^aTwo-sided χ^2 test for either haplotypic distribution or frequencies between the cases and controls

^bGlobal χ^2 test = 14.73, *p* value = 0.002

95% CI 9.95–7797.68; *p* = 0.001 vs. HR¹ 2.10; 95% CI 0.82–5.39; 95%; *p* = 0.12). Whereas for codon 228, a very high survival rate for females with variant genotype (AA) was observed as compared to males (18.63 vs. 6.86 months). A good prognosis towards lung cancer had been observed suggesting a better survival in female patients carrying variant genotype (HR 0.17; 95% CI 0.02–1.02; *p* = 0.05) (Fig. 2c).

We further evaluated OS of the patients on the basis of their smoking status and their association with XPA polymorphism. For *G709A*, a significant association was observed for heterozygous genotype (GA) suggesting a better survival (HR 0.38; 95% CI 0.17–0.81; *p* = 0.01, MST = 9.46 months) when compared with reference (Fig. 2d). The subjects with combined genotype (GA+AA) were significantly associated with longer survival (HR 0.39; 95% CI 0.18–0.83; *p* = 0.01) when compared with the wild genotype (GG). We also evaluated the association of XPA SNPs and overall survival of lung cancer patients treated with different chemotherapeutic regimen but was not able to obtain any significant results (Supplementary Table 2).

Discussion

In this study we have analyzed the association of two XPA SNPs (*A23G*, *G709A*) towards susceptibility for lung cancer and overall survival in lung cancer patients. The subjects with variant genotype (GG) showed a significant twofold risk of acquiring lung cancer (*p* = 0.0007) for XPA *A23G* polymorphism. When variant genotype combined with the heterozygous genotype (AG+GG), risk for lung cancer was observed (*p* = 0.03) which was less pronounced as compared to variant genotype (GG). Our results were in concordance with the studies conducted on East Asians (Qian et al. 2011), Caucasians (Butkiewicz et al. 2004), Mixed Americans (Wu et al. 2003), Danish (Vogel et al. 2005), and African American population (Wu et al. 2003). A meta-analysis (Kiyohara et al. 2010) also reported a strong association of XPA *A23G* polymorphism with increased risk of lung cancer and was in agreement with our study. *A23G* is a single nucleotide substitution (A → G) placed at the 5' non-coding region is associated with gene expression regulation via transcriptional and post-transcriptional control mechanism (Butkiewicz et al. 2000). This polymorphism could affect the structure of mRNA and its stability which might influence the

Table 4 Distribution of genotypes of XPA polymorphism according to their smoking status and its association with Lung Cancer risk

	Smokers				Non-smokers			
	Controls [n (%)] N = 271	Cases [n (%)] N = 303	AOR (95% CI) ^a	p ^b	Controls [n (%)] N = 99	Cases [n (%)] N = 67	AOR (95% CI) ^a	p ^b
XPA A23G								
AA	125 (46.12)	116 (38.28)	1.00 (Reference)		56 (56.56)	30 (44.77)	1.00 (Reference)	
AG	132 (48.70)	142 (46.86)	1.17 (0.82–1.68)	0.37	33 (33.33)	28 (41.79)	1.49 (0.75–2.98)	0.25
GG	14 (5.1)	45 (14.8)	3.61 (1.82–7.04)	0.0002	10 (10.10)	9 (13.43)	1.39 (0.48–4.02)	0.54
AG+GG	146 (39.40)	187 (50.54)	1.40 (0.99–1.98)	0.05	43 (43.43)	37 (55.22)	1.46 (0.77–2.78)	0.23
XPA G709A								
GG	131 (48.33)	175 (57.75)	1.00 (Reference)		46 (46.46)	36 (53.73)	1.00 (Reference)	
GA	121 (44.64)	96 (31.68)	0.54 (0.38–0.78)	0.001	46 (46.46)	25 (37.31)	0.60 (0.33–1.27)	0.21
AA	19 (7.01)	32 (10.56)	0.97 (0.51–1.86)	0.94	7 (7.07)	6 (8.95)	0.84 (0.24–2.87)	0.78
GA+AA	130 (47.97)	128 (42.24)	0.68 (0.43–0.85)	0.004	53 (53.53)	31 (46.2)	0.69 (0.36–1.31)	0.25
Heavy smoker								
	Controls [n (%)] N = 108	Cases [n (%)] N = 157	AOR (95% CI) ^a	p ^b	Controls [n (%)] N = 163	Cases [n (%)] N = 146	AOR (95% CI) ^a	p ^b
XPA A23G								
AA	45 (41.66)	57 (36.30)	1.00 (Reference)		80 (49.07)	59 (40.41)	1.00 (Reference)	
AG	54 (50)	77 (49.04)	1.12 (0.65–1.92)	0.68	5 (3.06)	65 (44.52)	1.20 (0.74–1.96)	0.45
GG	9 (8.33)	23 (14.64)	1.87 (0.79–4.54)	0.16	78 (47.85)	22 (15.06)	6.78 (2.36–19.49)	0.0004
AG+GG	63 (58.33)	100 (63.69)	1.23 (0.73–2.07)	0.43	83 (50.92)	87 (59.58)	1.52 (0.95–2.43)	0.07
XPA G709A								
GG	48 (44.44)	94 (59.87)	1.00 (Reference)		82 (50.30)	81 (55.47)	1.00 (Reference)	
GA	54 (50)	50 (31.84)	0.49 (0.29–0.84)	0.01	68 (7.97)	46 (31.50)	0.59 (0.35–0.98)	0.04

Table 4 (continued)

	Heavy smoker			Light smoker				
	Controls [<i>n</i> (%)] <i>N</i> = 108	Cases [<i>n</i> (%)] <i>N</i> = 157	AOR (95% CI) ^a	<i>p</i> ^b	Controls [<i>n</i> (%)] <i>N</i> = 163	Cases [<i>n</i> (%)] <i>N</i> = 146	AOR (95% CI) ^a	<i>p</i> ^b
AA	6 (5.55)	13 (8.28)	0.80 (0.27–2.39)	0.70	13 (7.97)	19 (13.01)	1.17 (0.52–2.64)	0.69
GA+AA	60 (55.55)	63 (40.12)	0.53 (0.31–0.89)	0.01	81 (49.69)	65 (44.52)	0.68 (0.42–1.09)	0.11

Highly significant values are given bold

^aAdjusted odds ratios, 95% confidence intervals, and their corresponding *p* values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status

^bTwo-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls

Table 5 Relationship of XPA polymorphism with overall survival of patients

Overall LC															
Overall LC								Overall LC							
XPA A23G	Dead 224	Alive 67	Median OS (months)	HR ^a (95% CI)	Log P	HR ^b (95% CI)	p ^c value	XPA G709A	Dead 224	Alive 67	Median OS (months)	HR ^a (95% CI)	Log P	HR ^b (95% CI)	p ^c value
AA	90	28	7.56	1.00 (Reference)		1.00 (Reference)		GG	120	43	6.43	1.00 (Reference)		1.00 (Reference)	
AG	102	31	6.76	0.99 (0.75–1.31)	0.98	0.94 (0.70–1.25)	0.68	GA	82	18	8.20	1.03 (0.77–1.36)	0.83	1.01 (0.75–1.36)	0.91
GG	32	8	7.13	1.14 (0.75–1.74)	0.53	1.07 (0.71–1.62)	0.73	AA	22	6	7.00	0.85 (0.55–1.31)	0.49	0.82 (0.51–1.69)	0.40
AG+GG	134	67	7.00	1.02 (0.78–1.74)	0.83	0.99 (0.75–1.30)	0.95	GA+AA	104	24	8.03	0.98 (0.76–1.28)	0.93	0.93 (0.71–1.22)	0.64
ADCC															
XPA A23G	Dead 73	Alive 23	Median OS (months)	HR ^a (95% CI)	Log P	HR ^b (95% CI)	p ^c value	XPA G709A	Dead 73	Alive 23	Median OS (months)	HR ^a (95% CI)	Log P	HR ^b (95% CI)	p ^c value
AA	31	13	9.43	1.00 (Reference)		1.00 (Reference)		GG	36	15	6.43	1.00 (Reference)		1.00 (Reference)	
AG	35	7	6.43	1.55 (0.95–2.52)	0.07	5.61 (1.28–24.46)	0.02	GA	35	6	8.53	1.06 (0.66–1.69)	0.78	1.03 (0.61–1.72)	0.90
GG	10	3	5.13	1.62 (0.75–3.50)	0.16	1534 (0.018–130,983,727.43)	0.20	AA	5	2	7.00	0.71 (0.31–1.65)	0.47	0.62 (0.22–1.69)	0.35

SCLC

SCLC				SCLC				SCLC							
XPA	Dead 62	Alive 20	Median OS (months)	HR ^a (95% CI)	Log P	HR ^b (95% CI)	<i>p</i> ^c value	XPA	Dead 62	Alive 20	Median OS (months)	HR ^a (95% CI)	Log P	HR ^b (95% CI)	<i>p</i> ^c value
A23G								G709A							
AA	26	6	6.73	1.00 (Ref- erence)		1.00 (Ref- erence)		GG	34	13	7.23	1.00 (Ref- erence)		1.00 (Ref- erence)	
AG	29	10	5.80	0.79 (0.46– 1.35)	0.37	0.71 (0.35– 1.45)		GA	22	5	4.50	1.06 (0.61– 1.82)	0.82	1.19 (0.67– 2.11)	0.53
GG	7	4	11.60	0.50 (0.24– 1.04)	0.08	1.26 (0.67– 2.35)	0.45	AA	6	2	5.80	0.96 (0.41– 2.28)	0.94	1.02 (0.39– 2.67)	0.95
AG+GG	36	14	6.73	0.72 (0.42– 1.22)	0.19	0.89 (0.56– 1.39)	0.61	GA+AA	28	7	5.80	1.04 (0.63– 1.72)	0.85	1.08 (0.64– 1.81)	0.75
Males															
Males								Males							
XPA	Dead 194	Alive 60	Median OS (months)	HR ^a (95% CI)	Log P	HR ^b (95% CI)	<i>p</i> ^c value	XPA	Dead 194	Alive 60	Median OS (months)	HR ^a (95% CI)	Log P	HR ^b (95% CI)	<i>p</i> ^c value
A23G								G709A							
AA	76	24	6.23	1.00 (Ref- erence)		1.00 (Ref- erence)		GG	103	40	6.23	1.00 (Ref- erence)		1.00 (Ref- erence)	
AG	90	29	7.00	0.93 (0.68– 1.27)	0.68	2.10 (0.82– 5.39)	0.45	GA	72	16	8.26	1.03 (0.76– 1.40)	0.81	1.00 (0.73– 1.36)	0.99
GG	28	7	7.26	1.03 (0.66– 1.59)	0.88	1.00 (0.65– 1.56)	0.91	AA	19	4	6.86	0.78 (0.57– 1.50)	0.78	0.85 (0.51– 1.41)	0.54

Table 5 (continued)

Males								Males							
XPA A23G	Dead 194	Alive 60	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value	XPA G709A	Dead 194	Alive 60	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value
AG+GG	118	36	7.13	1.03 (0.77– 1.38)	0.79	0.99 (0.68– 1.20)	0.58	GA+AA	91	20	8.03	1.01 (0.76– 1.34)	0.93	1.06 (0.78– 1.44)	0.68
Females								Females							
XPA A23G	Dead 30	Alive 7	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value	XPA G709A	Dead 30	Alive 7	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value
AA	14	4	9.46	1.00 (Ref- erence)		1.00 (Ref- erence)		GG	17	3	6.73	1.00 (Ref- erence)		1.00 (Ref- erence)	
AG	12	2	4.50	1.69 (0.74– 3.81)	0.15	1.43 (0.52– 3.95)	0.48	GA	10	2	4.50	0.96 (0.41– 2.10)	0.92	1.23 (0.40– 3.78)	0.71
GG	4	1	4.16	4.35 (0.63– 29.79)	0.002	278.58 (9.95– 7797–68)	0.001	AA	3	2	18.63	0.55 (0.20– 1.56)	0.29	0.17 (0.02– 1.02)	0.05
AG+GG	16	3	4.50	1.92 (0.91– 4.04)	0.05	2.10 (0.82– 5.39)	0.12	GA+AA	13	4	8.8	0.84 (0.41– 1.72)	0.62	0.61 (0.24– 1.52)	0.29

Table 5 (continued)

Smokers										
XPA A23G		Smokers								
Dead 183	Alive 56	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value	XPA G709A	Dead 183	Alive 56	Median OS (months)
73	21	6.23	1.00 (Ref- erence)		1.00 (Ref- erence)		GG	98	37	7.23
AG	82	7.00	0.92 (0.67–1.27)	0.63	0.87 (0.63–1.20)	0.42	GA	65	13	6.23
										1.17 (0.85–1.61)
GG	28	6	1.11 (0.71–1.75)	0.61	1.14 (0.73–1.78)	0.55	AA	20	6	7.00
										0.85 (0.53–1.34)
AG+GG	110	35	0.96 (0.71–1.30)	0.83	0.94 (0.69–1.27)	0.71	GA+AA	85	19	6.86
										1.07 (0.80–1.44)
Non-smokers										
XPA A23G		Non-smokers								
Dead 41	Alive 11	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value	XPA G709A	Dead 41	Alive 11	Median OS (months)
17	7	9.03	1.00 (Ref- erence)		1.00 (Ref- erence)		GG	22	6	5.83
AG	20	6.73	1.41 (0.73–2.69)	0.28	1.32 (0.62–2.80)	0.46	GA	17	5	9.46
										0.59 (0.31–1.12)
GG	4	2	0.32 (0.32–2.85)	0.94	1.14 (0.19–6.71)	0.88	AA	2	0	6.23
										1.04 (0.23–4.59)
Non-smokers										
XPA A23G		Non-smokers								
Dead 41	Alive 11	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value	XPA G709A	Dead 41	Alive 11	Median OS (months)
17	7	9.03	1.00 (Ref- erence)		1.00 (Ref- erence)		GG	22	6	5.83
AG	20	6.73	1.41 (0.73–2.69)	0.28	1.32 (0.62–2.80)	0.46	GA	17	5	9.46
										0.59 (0.31–1.12)
GG	4	2	0.32 (0.32–2.85)	0.94	1.14 (0.19–6.71)	0.88	AA	2	0	6.23
										1.04 (0.23–4.59)
Non-smokers										
XPA A23G		Non-smokers								
Dead 41	Alive 11	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value	XPA G709A	Dead 41	Alive 11	Median OS (months)
17	7	9.03	1.00 (Ref- erence)		1.00 (Ref- erence)		GG	22	6	5.83
AG	20	6.73	1.41 (0.73–2.69)	0.28	1.32 (0.62–2.80)	0.46	GA	17	5	9.46
										0.59 (0.31–1.12)
GG	4	2	0.32 (0.32–2.85)	0.94	1.14 (0.19–6.71)	0.88	AA	2	0	6.23
										1.04 (0.23–4.59)

Table 5 (continued)

Non-smokers										Non-smokers					
XPA A23G	Dead 41	Alive 11	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value	XPA G709A	Dead 41	Alive 11	Median OS (months)	HR ^a (95% CI)	Log <i>P</i>	HR ^b (95% CI)	<i>p</i> ^c value
AG+GG	24	6	8.03	0.77 (0.42– 1.43)	0.41	1.36 (0.65– 2.84)	0.41	GA+AA	19	5	9.43	0.62 (0.33– 1.16)	0.11	0.39 (0.18– 0.83)	0.01

Highly significant values are given bold

^aHazard ratios, 95% confidence intervals, and their corresponding log rank *p* values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months

^bTwo-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: variant genotype, 1: combined hetero and variant genotype

^cAdjusted hazard ratios, 95% confidence intervals, and their corresponding *p* values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, kps, and ecog

mRNA binding with translational factors. In addition many studies regarding site-directed mutagenesis have manifested that any mutation in *XPA*s functional domain could partially or completely disable its function (Park et al. 2002). Therefore, this might be the reason that *A23G* polymorphism represents risk for the lung cancer in North Indian population. As cancer is a composite disease, it could also be possible that the same polymorphism might show a different response for cancer risk across different tumor size and population of different ethnic background (Hirschhorn et al. 2002).

For *XPA* codon 228 polymorphism, a strong protective effect was observed for heterozygous (GA) ($p = 0.001$) and combined (GA+AA) genotype ($p = 0.004$) at codon 228 polymorphism. *Arg228Gln* (*G709A*) is a rare polymorphism which has been considered as a *XPA* mutation resulted out of DNA repair defects (Porter et al. 2005). *Arg228Gln* is situated in exon VI which barely defines the TFII H binding domain of *XPA*. Any change in binding affinity with TFII H may lead to improved NER kinetics (Pleschke et al. 2000). *G709A* lies within agreement with a poly (ADP-ribose) covering 215–237 amino acid residues and any change in the non-conservative *Arg228Gln* disrupts this conserved sequence (Pleschke et al. 2000). Pleschke et al. 2000 suggested that formation of poly (ADP-ribose) takes place in

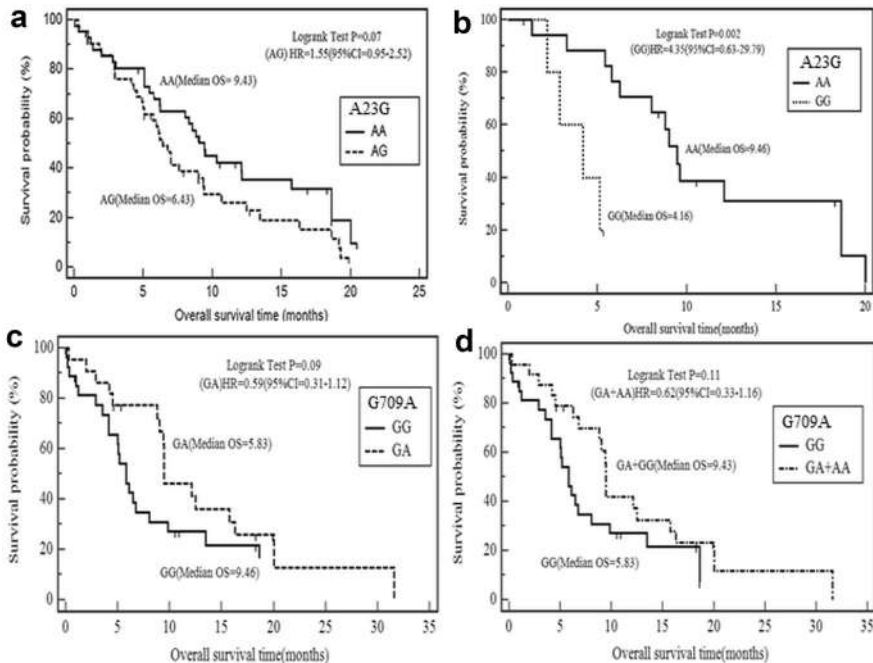


Fig. 2 Kaplan–Meier curve depicting association between *XPA* A23G polymorphism and **a** Heterozygous genotype (AG) in ADCC patients **b** variant genotype (GG) in females. Kaplan–Meier curve depicting association between *XPA* G709A polymorphism and **c** variant genotype (AA) in females **d** heterozygous genotype (GA) in non-smokers

response to DNA damage which on binding with *XPA* inhibits NER pathway and any disruption in poly (ADP-ribose)-binding site might prevent the inhibition of *XPA*. Hence, result into improved DNA repair. A study by Cleaver and States (1997) suggested that the patients with mutations were associated with early onset of a disease resulting in the inactivation of *XPA*. These mutations were commonly present in the DNA-binding region of the protein. Patients with delayed onset of a disease have mutations that allow expression of partially functional *XPA* or are point mutations in exon VI. This might be the reason that GA shows a protective effect than the variant AA genotype for *G709A* polymorphism.

We further evaluated association of *XPA* polymorphism and histological subtypes. Histological stratification analysis suggested that GG genotype was presented as a higher risk factor for all the three histological subtypes. However in this study risk was more pronounced for SCLC ($p = 0.0007$) for *A23G* polymorphism. This difference might be due to the differences in the carcinogenesis pathway involved in histological subtypes of lung cancer (Gu et al. 1999). A study on Caucasian (Butkiewicz et al. 2004) and Chinese (Qian et al. 2011) population also supported our findings where variant genotype showed risk in SQCC and ADCC subjects.

Further we also evaluated the role of SNP-smoking status and its association with lung cancer risk. Smokers with variant genotype (GG) in *XPA A23G* polymorphism depicted a threefold risk for lung cancer ($p = 0.0002$). When further segregated on the basis of pack years, the variant genotype (GG) showed adverse risk for acquiring the disease ($p = 0.0004$) among light smokers (pack years ≤ 25). Our results were in agreement with the study conducted on Danish (Raaschou-Nielsen et al. 2008), Korean (Qian et al. 2011), and Chinese (Wu et al. 2003) population suggesting a high risk for acquiring lung cancer for subjects carrying GG genotype. Although our findings are a bit contradictory as it is already known that homozygous variant allele (GG) had 13% higher DNA repair capacity than homozygous wild allele (AA) (Wu et al. 2003). Therefore, lung cancer patients with poor DNA repair capacity were compared with healthy controls and this risk might be modulated by the smoking status (Spitz et al. 2001).

Further, the univariate Kaplan and multivariate Cox regression model was applied to evaluate the overall survival of lung cancer patients and its association with *XPA* polymorphism. In *XPA A23G* polymorphism, a significant risk with moderate survival for heterozygous genotype (AG) in ADCC patients was observed (MST = 6.43 vs. 9.46, $p = 0.02$) when compared with wild genotype whereas a shorter survival period with increased risk was observed for females with variant genotype GG (MST = 4.16 months, $p = 0.002$). Our results were in agreement with the study carried on Asian (Su et al. 2015), Caucasian (Shen et al. 2007), and Polish (Butkiewicz et al. 2011) population suggesting association of variant genotype (GG) with increased risk of lung cancer and poor survival in non-smoking females. *XPA A23G* polymorphism located in 5'UTR region might influence transcription of *XPA* which plays an important role in NER pathway. This substitution could also remove a CpG site that interferes with the methylation process of *XPA* promoter as well as regulation of *XPA* expression (Candiloro and Dobrovic 2009). The *XPA* allele could also be in linkage disequilibrium with allele of adjacent gene and might be susceptible to survival outcomes. A study on *XPA (A23G)* polymorphism had concluded

that its association with the alteration of TP53 mutation in lung cancer which leads to an aggressive phenotype might be the reason for poor survival (Cho et al. 2011). Whereas the findings of the present study suggested a significant protective effect in females with variant genotype (AA) showing *G709A* polymorphism with a longer survival period when compared with reference (MST-18.63 vs. 6.73, $p = 0.05$). A protective effect among smokers of heterozygous (GA) and combined (GA+AA) genotype was also observed with better survival period for *XPA228* (MST-9.46 (GA), 9.43(GA+AA), $p = 0.01$). As suggested in past studies, *G709A* polymorphism reported with better survival in the population of two continents (Butkiewicz et al. 2001; Mellon et al. 2002). Porter et al. (2005) suggested that this polymorphism might behave in hyperactive state for UV defect repair depending upon the *XPA* complementation which supposed to be sub-optimally low.

In conclusion, *XPA A23G* polymorphism might be associated with increased lung cancer risk in variant genotype in the North Indian population integrated with poor survival in female lung cancer patients. Whereas *XPA G709A* polymorphism could modulate the protective effect in North Indian population incorporated with moderate survival among non-smokers.

Funding This work was supported by Grant from the Indian Council of Medical Research, New Delhi, India (Grant No. 5/13/126/2011/NCD-III).

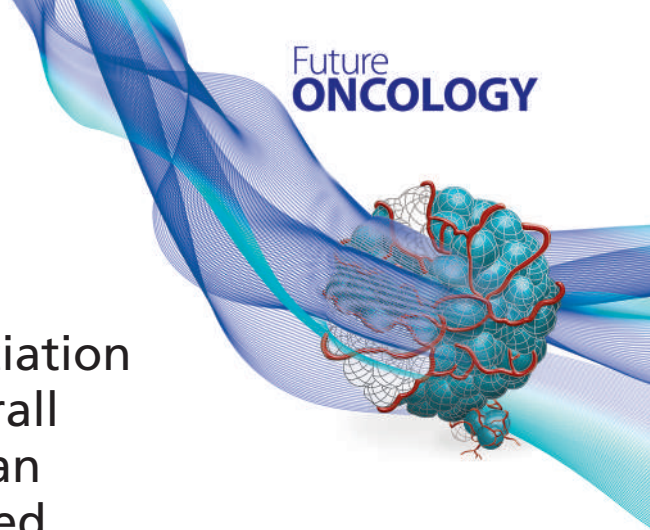
Compliance with Ethical Standards

Conflict of interest All authors declare that they have no conflict of interest.

References

- Batty DP, Wood RD (2000) Damage recognition in nucleotide excision repair of DNA. *Gene* 241:193–204
- Behera D (2012) Lung cancer in India. *Med Update* 22:401–406
- Butkiewicz D, Rusin M, Harris CC, Chorazy M (2000) Identification of four single nucleotide polymorphisms in DNA repair genes: *XPA* and *XPB* (ERCC3) in Polish population. *Hum Mutat* 15:577–578
- Butkiewicz D, Rusin M, Enewold L, Shields PG, Chorazy M, Harris CC et al (2001) Genetic polymorphisms in DNA repair genes and risk of lung cancer. *Carcinogenesis* 22:593–597
- Butkiewicz D, Popanda O, Risch A, Edler L, Dienemann H, Schulz V et al (2004) Association between the risk for lung adenocarcinoma and a (–4) G-to-A polymorphism in the *XPA* gene. *Cancer Epidemiol Biomarkers Prev* 213:2242–2246
- Butkiewicz D, Rusin M, Zena Sikora B, Lach A, Mieczysław C (2011) An association between DNA repair gene polymorphisms and survival in patients with resected non-small cell lung cancer. *Mol Biol Rep* 38:5231–5241
- Candiloro IL, Dobrovic A (2009) Detection of MGMT promoter methylation in normal individuals is strongly associated with the T allele of the rs16906252 MGMT promoter single nucleotide polymorphism. *Cancer Prev Res* 2:862–867
- Cho S, Kim MJ, Choi YY, Yoo SS, Lee WK, Lee EJ et al (2011) Associations between polymorphisms in DNA repair genes and TP53 mutations in non-small cell lung cancer. *Lung Cancer* 73:25–31
- Cleaver JE, States JC (1997) The DNA damage-recognition problem in human and other eukaryotic cells: the *XPA* damage binding protein. *Biochem J* 328:1–12
- Gu J, Spitz MR, Yang F, Wu X (1999) Ethnic differences in poly (ADPribose) polymerase pseudogene genotype distribution and association with lung cancer risk. *Carcinogenesis* 20:1465–1469

- Hirschhorn JH, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. *Genet Med* 4(2):45–61
- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA Cancer J Clin* 20(61):69–90
- Kang T, Reardon J, Sancar A (2011) Regulation of nucleotide excision repair activity by transcriptional and post-transcriptional control of the XPA protein. *Nucleic Acids Res* 39(8):3176–3187
- Khuder SA (2001) Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer* 31:139–148
- Kiyohara C, Takayama K, Nakanishi Y (2010) Lung cancer risk and genetic polymorphism in DNA repair pathways: metaanalysis. *J Nucl Acid* 2010:701760
- Mellon I, Hock T, Reid R, Porter PC, States JC (2002) Polymorphisms in the human xeroderma pigmentosum group A gene and their impact on cell survival and nucleotide excision repair. *DNA Repair* 1:531–546
- Park JY, Park SH, Choi JE, Lee SY, Jeon HS, Cha SI et al (2002) Polymorphisms of the DNA repair gene xeroderma pigmentosum group A and risk of primary lung cancer. *Cancer Epidemiol Biomarkers Prev* 11(10):993–997
- Pleschke JM, Kleczkowska HE, Strohm M, Althaus FR (2000) Poly(ADP-ribose) binds to specific domains in DNA damage checkpoint proteins. *J Biol Chem* 275:40974–40980
- Porter PC, Mellon I, States JC (2005) XP-A cells complemented with Arg228Gln and Val234Leu polymorphic XPA alleles repair BPDE-induced DNA damage better than cells complemented with the wild type allele. *DNA Repair* 4:341–349
- Qian B, Zhang H, Zhang L, Zhou X, Yu H, Chen K (2011) Association of genetic polymorphisms in DNA repair pathway genes with non-small cell lung cancer risk. *Lung Cancer* 73:138–146
- Raaschou-Nielsen O, Sørensen M, Overvad K, Tjønneland A, Vogel U (2008) Polymorphisms in nucleotide excision repair genes, smoking and intake of fruit and vegetables in relation to lung cancer. *Lung Cancer* 59:171–179
- Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz A, Linn S (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* 73:39–85
- Sant M, Allemani C, Santaquilani M, Knijn A, Marchesi F, Capocaccia R (2009) Survival of cancer patients diagnosed in 1995-1999: results and commentary. *Eur J Cancer* 45:931–991
- Shen J, Deininger P, Hunt JD, Zhao H (2007) 8-Hydroxy-2'-deoxyguanosine (8-OH-dG) as a potential survival biomarker in patients with nonsmall-cell lung cancer. *Cancer* 109:574–580
- Sodhi KK, Bahl C, Singh N, Behera D, Sharma S (2015) Functional genetic variants in pre-miR-146a and 196a2 genes are associated with risk of lung cancer in North Indians. *Future Oncol* 1(15):2159–2173
- Spitz MR, Wu X, Wang Y, Wang LE, Shete S, Amos CI (2001) Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. *Cancer Res* 6:1354–1357
- Su Y, Zhang H, Xu F, Kong J, Yu H, Qian B (2015) DNA repair gene polymorphisms in relation to non-small cell lung cancer survival. *Cell Physiol Biochem* 36:1419–1429
- Vogel U, Overvad K, Wallin H, Tjønneland A, Nexø BA, Nielsen Raaschou- (2005) Combinations of polymorphisms in XPD, XPC and XPA in relation to risk of lung cancer. *Cancer Lett* 222(1):67–74
- Wood RC, Mitchell M, Sgouros J, Lindahl T (2001) Human DNA repair genes. *Science* 22:1284–1288
- Wu X, Zhao H, Wei Q, Amos CI, Zhang K, Guo Z et al (2003) XPA polymorphism associated with reduced lung cancer risk and a modulating effect on nucleotide excision repair capacity. *Carcinogenesis* 24(3):505–509



XPG polymorphisms and their association with lung cancer susceptibility, overall survival and response in North Indian patients treated with platinum-based doublet chemotherapy

Shweta Lawania¹, Navneet Singh², Digambar Behera² & Siddharth Sharma^{*,1}

¹Department of Biotechnology, Thapar University, Punjab 147002, India

²Department of Pulmonary Medicine, Post Graduate Institute of Medical & Educational Research (PGIMER), Sector 14, Chandigarh, India

*Author for correspondence: siddharthsharma.phd@thapar.edu

Aim: The study investigates association of XPG polymorphism with lung cancer susceptibility, overall survival and clinical outcomes in North Indian population. **Results:** A significant protective effect was observed for 2228959 C/A polymorphism with lung cancer and its histological subtypes. An increased hazard ratio (HR) was observed in 17655 G/C variant among small-cell lung carcinoma patients with mutant genotype (HR: 2.55; $p = 0.05$). Individuals treated with irinotecan-cisplatin/carboplatin regimen showed a longer survival time (HR¹: 0.04; median survival time [MST]: 32.5 months). Subjects treated with pemetrexed-cisplatin/carboplatin regimen were associated with higher mortality rate in lung cancer patients (HR¹: 1.83; MST: 9.13 months). **Conclusion:** 2228959 C/A polymorphism contributes to protective effect in lung cancer patients. 2228959 C/A polymorphism might be associated with favorable prognosis in lung cancer risk.

First draft submitted: 30 May 2018; Accepted for publication: 19 September 2018; Published online: 7 December 2018

Keywords: chemotherapy regimen • clinical outcomes • lung cancer susceptibility • overall survival • XPG polymorphism

Lung cancer is regarded as a disease completely related with the external environmental carcinogen exposure and smoking is the main causative factor, which is closely related to the onset of lung cancer in 80–90% of patients [1]. Lung cancer is a result of interactions between external environmental factors and internal genetic factors [1]. When exposed to same environmental conditions only a few individual develop the disease suggesting role of individual's genetic constitution in defining its susceptibility toward the disease [2]. DNA repair system plays an important role in safeguarding the genome from the malign effects of cancer-causing agent such as tobacco smoke [3]. Nucleotide excision repair (NER) is the crucial DNA repair pathway with majorly counteracts with the consequences induced by the tobacco carcinogens such as benzo [a] pyrenediol epoxide [3]. NER system first recognizes the DNA damage, incises the lesion, removes the damage containing oligonucleotide and resynthesizes the fragment [4]. Defective NER pathway may lead to a rare disease like xeroderma pigmentosum (XP), which is characterized by UV-sensitivity in humans. There are seven complementation groups from XPA to XPG that has been identified so far [4]. XPG gene identified as ERCC5, maps at chromosome 13q32-33 and encodes 1186 amino acid structure specific endonuclease which plays essential role in two incision steps in NER pathway [5]. The protein shows extensive homology with two other nucleases, in other words, RAD2 (a yeast DNA repair protein) and bacteriophage T4 RNase H. These enzymes cleave single to double strand DNA at junctions forming a flap or bubble like structure [6]. XPG catalysis incision approximately at five nucleotide from 3' site of the damage and also involved nonenzymatically in subsequent 5' incision by XPF/ERCC1 heterodimer during NER process [7,8].

Two polymorphic variants of *XPG* has been identified, in other words, *Phe*⁶⁷⁰*Leu* and *Asp*¹¹⁰⁴*His* but functional significance of these polymorphisms are still unknown. *Phe*⁶⁷⁰*Leu* polymorphism (rs2228959) is located on exon 9 leads to C→A transition (C2207A), thus resulting in phenylalanine (Phe) to Leucine (Leu) amino acid change [9]. *Asp*¹¹⁰⁴*His* (rs17655) is present at exon 15, results in G→C substitution (G3507C), which changes aspartate (Asp) to histidine (His) amino acid [7]. As *XPG* is a rare polymorphism, very few epidemiological studies have been reported till now. Emmert *et al.* [10] stated association of 17655 G/C with increased breast cancer risk whereas exon 9 polymorphism was not detected in the study population. Hussain *et al.* [11] further reported association of *XPG* exon 15 polymorphism with reduced gastric cancer risk. A study on non-Hispanic white population documented weak association of *XPG* single nucleotide polymorphisms (SNPs) with small cell carcinoma of head and neck cancer [12] while Jeon *et al.* [7] suggested association of 17655 G/C polymorphism with the increased lung cancer risk.

For lung cancer patients, platinum-based doublet chemotherapy has been the most commonly used treatment. However, the response rate among patients is only 20% [1]. The patients having similar tumor stage and histological subtype show different chemotherapy response along with different prognosis suggesting role of molecular marker in predicting the treatment sensitivity and clinical outcomes [13]. Platinum compounds bind to DNA forming platinum-DNA adducts which result into interstrand crosslinks and bulky distortions. Many mechanisms play important role in repair of DNA damage caused by these platinum compounds including NER pathway which may influence anticancer effect of these agents [13]. Few studies have emphasized on the association of various polymorphisms in NER-pathway gene that may lead to alteration in the repair efficiency of DNA, which may influence treatment response and cancer survival [13]. Zhang *et al.* [14] and Yuli *et al.* [15] found *XPG* polymorphism as the useful prognostic factor in non-small-cell lung cancer (NSCLC) patients treated with platinum-based chemotherapy, whereas few studies have indicated that *XPG* genes are the potential predictor for the clinical outcomes of the cancer patients such as ovarian cancer, melanoma, osteocarcinoma, gastric cancer, lung cancer and bladder cancer [16,17].

Therefore, in the present study, we conducted a case-control study to examine the association of two *XPG* polymorphic variants (*Phe*⁶⁷⁰*Leu* and *Asp*¹¹⁰⁴*His*) toward risk for lung cancer and its histological subtypes. We accessed the role of the two polymorphic variants of *XPG* toward clinic-pathological parameters in lung cancer and their association with *XPG* polymorphism. We further assessed role of *XPG* variants toward overall survival (OS) of patients treated with platinum-based chemotherapy and their response rates.

Material & methods

Study population

A hospital-based study having 370 cases and 370 controls was registered in the Department of Pulmonary Medicine of Post Graduate Institute of Medical Education and Research, Chandigarh, India. The study was reviewed and approved by the Ethics Committee board of Post Graduate Institute of Medical Education and Research. The present study has patients without any age, gender, smoking, histology and Tumor, Node, Metastasis (TNM) staging restrictions. Patients with prior cancer history were avoided. All lung cancer patients were confirmed histopathologically and were diagnosed with NSCLC and small-cell lung carcinoma (SCLC). The patients who were not diagnosed with any cancer and came to hospital for routine checkup were selected as healthy controls. A written consent from each and every individual was obtained. A detailed questionnaire was filled by the trained interviewer for each case and control regarding demographics and lifetime of smoking history. Details regarding histology, TNM classification, clinical staging, regimen of treatment and number of chemotherapeutic cycles were retrieved from the medical records of each patient from the hospital. The follow-up data of patients were telephonically obtained by utilizing the information present in the medical records.

Genotyping of *XPG* variants

Genomic DNA was isolated using phenol chloroform extraction protocol as mentioned by Sodhi *et al.* [18]. Two *XPG* polymorphisms, in other words, 2228959 C/A and 17655 G/C were detected using PCR-restriction fragment length polymorphism assay as previously described (Figure 1A & B) [7,9]. Fifteen percent samples were repeated for each polymorphism and results obtained were 100% concordant.

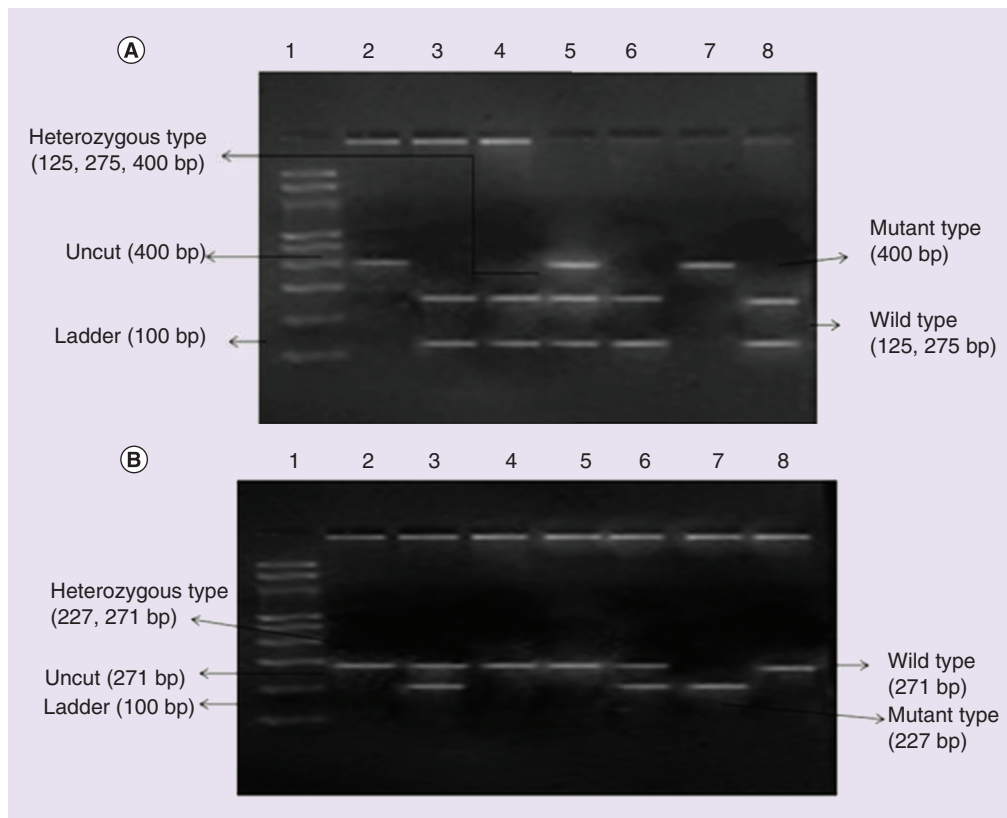


Figure 1. Restriction digestion pattern of XPG variants. (A) Restriction digestion pattern of 2228959 C/A polymorphism: wild-type (CC): 125, 275 bp; mutant-type (AA): 400 bp; heterozygous-type (CA): 125 bp, 275 bp and 400 bp. Lane 1: Marker; Lane 2: Uncut; Lane 3, 4, 6, 8: CC genotype; Lane 7: AA genotype; Lane 5: CA genotype. **(B)** Restriction digestion pattern of 17655 G/C polymorphism: wild-type (GG): 271 bp; mutant-type (CC): 227 bp (33 bp product was too small to resolve accurately); heterozygous-type (GC): 227 and 271 bp. Lane 1: Marker; Lane 2: Uncut; Lane 3, 6: GC genotype; Lane 7: CC genotype; Lane 4, 5, 8: GG genotype. bp: Base pair.

Chemotherapeutic treatment

In the present study, all patients received platinum-based chemotherapy. The chemotherapy drugs consist of cisplatin or carboplatin given in combination with taxol-based drug, in other words, docetaxel, irinotecan or pemetrexed. This included 75 mg/m² docetaxel, 500 mg/m² pemetrexed or 75 mg/m² irinotecan administered as an hour infusion, followed by infusion of 70 mg/m² cisplatin or carboplatin for 3 h. The treatment was repeated for 3–4 weeks and tumor response was assessed after four cycles of chemotherapy. Chemotherapy was stopped in case of unacceptable toxicity. If an objective response was reported then patients were treated with two additional chemotherapy cycles (maximum of six cycles).

Follow-up & response assessment

Study subjects were followed up for survival through telephonic conversations. Every 2 months until death or the end of the study period patients were called and inquired about health. Primarily we assessed OS and for that we considered survival time from the date of diagnosis till the date of last follow-up or death due to other reasons.

Statistical analysis

χ^2 test was used to compare demographic features like age, sex, smoking status and pack year smoked, whereas for continuous variables, Student's t-test has been used. To differentiate between observed and expected genotypic frequencies of controls and cases, Hardy–Weinberg equilibrium (HWE) was applied. Allelic and genotypic frequencies of cases and controls were evaluated using Pearson's χ^2 -test. Association between XPG variants and lung cancer risk was evaluated using odds ratio (OR) and 95% CI adjusted for age, sex and smoking status.

by applying logistic regression analysis. Haplotype analysis for the SNPs was applied using SHEsis software (<http://202.120.31.177/myAnalysis.php>) to calculate linkage disequilibrium and haplotype frequency estimation. Univariate Kaplan–Meier method was applied to estimate survival distribution and log-rank test for each SNP and its relation with lung cancer. Median survival time (MST) among censored samples was calculated. The multivariate Cox proportional hazard model was applied to assess the independent effect of each polymorphism on OS when adjusted for age, gender, smoking status, stage, histology, Karnofsky performance status (KPS) and Eastern Cooperative Oncology Group (ECOG). Further, relationship between OS and *XPG* polymorphic variants when categorized as per chemotherapeutic regimen was evaluated using Kaplan–Meier method and Cox hazard analysis method. Tumor response was evaluated using Response Evaluation Criteria in Solid Tumors for patients treated with ≥ 6 chemotherapy cycles. Patients showing complete response or partial response were regarded as responders, whereas patients showing stable disease or progressive disease were considered as nonresponders. All the p-values were two sided and $p < 0.05$ was referred as statistically significant. All the statistical analysis was carried out using Medcalc Statistical Software version 15.11.4 (MedCalc Software bvba, Ostend, Belgium).

Result

Demographic characteristics

Table 1 depicted the demographic characteristics of *XPG* polymorphisms and its association risk of lung cancer. The study consists of 370 cases and 370 controls. Mean age among cases and controls were 58.11 ± 10.44 and 58.83 ± 10.18 ($p < 0.0001$). All the bold values in the text represent the significant p-values. When stratified on basis of gender, the number of males and females were equally represented in both groups, in other words, 319 (86.21%) males and 51 (13.78%, $p = 0.91$) females. Our demographic data also showed a higher representation of smokers in cases as compared with controls (81.89 vs 73.29%). When classified in terms of pack years, a significantly higher pack year was associated with cases as compared with controls (29.47 ± 35.04 vs 17.88 ± 19.45 ; $p = 0.0001$). When the total lung cancer cases stratified on the basis of histological subtype, 135 (36.48%) were squamous cell carcinoma patients, 120 (32.48%) patients were diagnosed with adenocarcinoma and 111 (30%) were having SCLC patients. Similarly when the cases were distributed according to TNM staging: out of 370 cases, 0.81% were diagnosed at stage I; 2.70% with stage II; 45.94% and 44.05% at stage III and IV, respectively. The frequencies of lymph node involvement were divided as 12.43, 9.72, 42.43 and 28.64% for N0, N1, N2 and N3, respectively. The total data included 42.70% individuals having metastasis (M1) while 50.54% individuals did not show any metastasis (M0).

XPG variants & its association with lung cancer susceptibility

The distributions of the *XPG* codon *Phe*⁶⁷⁰*Leu* (rs2228959) and *Asp*¹¹⁰⁴*His* (rs17655) among cases and controls have been illustrated in Table 2. The distribution of the genotypes in controls for both the codons were in agreement with HWE, in other words, 2228959 C/A ($n = 370$; $\chi^2 = 0.05$; $df = 1$; $p = 0.81$) and 17655 G/C ($n = 370$; $\chi^2 = 1.74$; degree of freedom [df] = 1; $p = 0.18$). Our data revealed that in 2228959 C/A polymorphism heterozygous genotype (CA) was more frequent in controls as compared with cases (36.75 vs 18.10%), whereas the wild genotype (CC) was less represented in controls as compared with cases (55.40 vs 80%). Therefore in case of 2228959 C/A a significant distribution of the genotypes was observed among cases and controls ($\chi^2 = 53.42$; $df = 2$; $p < 0.0001$). The MAF for 2228959 C/A allele (A) was higher in controls than cases (26.21 vs 10.94). For 17655 G/C, the mutant genotype (CC) was at higher frequency in controls as compared with the cases (3.57 vs 2.97%), whereas heterozygous genotype (GC) was less represented in controls as compared with cases (25.13 vs 28.64%). The MAF for controls and cases for 17655 G/C polymorphism are 16.08, 17.29, respectively, as shown in Table 2.

The association between the two polymorphic variants of *XPG* gene and the risk of lung cancer was further examined after stratifying for potential confounding variables, such as age, sex and pack-years. The contribution of the two *XPG* polymorphic variants and its association with lung cancer susceptibility was further depicted in Table 2. Adjusted OR's and 95% CI were calculated using logistic regression analysis. For both polymorphisms the homozygous wild genotype was taken as reference group and the variant genotypes were compared with it. In case of 2228959 C/A polymorphism, it was observed that both heterozygous (OR: 0.35, 95% CI: 0.25–0.50; $p < 0.0001$) and mutant (OR: 0.18, 95% CI: 0.07–0.13; $p = 0.0001$) genotypes were conferring a protective effect when compared with the wild-type genotype (CC). Similarly, when stratified according to histological subtypes, it was observed that a protective effect was observed in NSCLC patients having mutant genotype (AA; OR: 0.14,

Table 1. Demographic characteristics among cases and control patients in North Indian population.

Variable	Cases, n (%) n = 370	Controls, n (%) n = 370	p-value [†]
Age (years):			
– Mean $\pm$ SD	58.11 $\pm$ 10.44	53.83 $\pm$ 10.18	<0.0001 [‡]
– Range			
Gender:			
– Male	319 (86.21%)	319 (86.21%)	0.91
– Female	51 (13.78%)	51 (13.78%)	
Smoking status:			
– Smokers	303 (81.89%)	271 (73.24%)	0.0063 [‡]
– Nonsmokers	67 (18.10%)	99 (26.75%)	
Pack years:			
– Mean $\pm$ SD	24.97 $\pm$ 35.04	17.88 $\pm$ 19.45	<0.0001 [‡]
Histological types (n = 370):			
– SQCC	135 (36.48%)		
– ADCC	120 (32.43%)		
– SCLC	111 (30%)		
– Others	4 (1.08%)		
– Unknown	0 (0.0)		
Tumor size (n = 370):			
– Tx	11 (2.97%)		
– T1	16 (4.32%)		
– T2	42 (11.35%)		
– T3	85 (22.97%)		
– T4	191 (51.62%)		
– Unknown	25 (6.75%)		
Metastasis (n = 370):			
– M0	187 (50.54%)		
– M1	158 (42.70%)		
– Unknown	25 (6.75%)		
ECOG (n = 311):			
– 0–1	148 (47.58%)		
– 2–4	158 (50.80%)		
– Unknown	5 (1.60%)		
TNM staging (n = 370):			
– I	3 (0.81%)		
– II	10 (2.70%)		
– III	175 (45.94%)		
– IV	163 (44.05%)		
– Unclassified	19 (5.13%)		
Lymph node involvement (n = 370):			
– N0	46 (12.43%)		
– N1	36 (9.72%)		
– N2	157 (42.43%)		
– N3	106 (28.64%)		
– Unknown	25 (6.75%)		
KPS (n = 311):			
– 100–90	123 (39.54%)		
– 80–70	138 (44.37%)		
– $\leq$ 60	45 (14.46%)		
– Unknown	5 (1.60%)		
Regimen (n = 311):			
– Doce-cis/doce-carb	76 (24.43%)		
– Irino-cis/irino-carb	70 (22.50%)		
– Peme-cis/peme-carb	69 (22.18%)		
– Others/unknown	96 (30.86%)		

[†] p-values were derived from Pearson χ^2 test except age; Student t-test was used for age. All p-values are two-sided. p < 0.05 was considered statistically significant.

[‡] Significant p-values.

ADCC: Adenocarcinoma; ECOG: Eastern Cooperative Oncology Group; KPS: Karnofsky performance score; SCLC: Small-cell lung carcinoma; SD: Standard deviation; SQCC: Squamous cell carcinoma; n: Total number of case patient or control subject; TNM: Tumor, Node, Metastasis.

95% CI: 0.04–0.42; p = 0.0004). Similarly when compared with CC genotype, the lung cancer subjects who were heterozygous carriers (CA) also showed a protective effect for NSCLC (OR: 0.27, 95% CI: 0.18–0.41; p < 0.0001) as well as SCLC (OR: 0.58, 95% CI: 0.35–0.96; p = 0.03) subtypes, respectively. For *rs17655 G/C* polymorphism, no significant association was observed for lung cancer risk and its histological subtypes.

Table 2. Distribution of genotypes of *XPG* polymorphism depicting their association with lung cancer risk and its histological subtypes.

Genotype	Controls, n (%) n = 370	Overall lung cancer			NSCLC			SCLC		
		Cases, n (%) n = 370	AOR (95% CI) [‡]	p-value [†]	Cases, n (%) n = 255	AOR (95% CI) [‡]	p-value [†]	Cases, n (%) n = 111	AOR (95% CI) [‡]	p-value [†]
2228959 C/A										
CC	205 (55.40)	296 (80)	1.00 (Reference)		214 (83.92)	1.00 (Reference)		78 (70.27)	1.00 (Reference)	
CA	136 (36.75)	67 (18.10)	0.35 (0.25–0.50)	<0.0001 [§]	37 (14.50)	0.27 (0.18–0.41)	<0.0001 [§]	30 (27.02)	0.58 (0.35–0.96)	0.03 [§]
AA	29 (7.83)	7 (1.89)	0.18 (0.07–0.43)	0.0001 [§]	4 (1.56)	0.14 (0.04–0.42)	0.0004 [§]	3 (2.70)	0.33 (0.09–1.17)	0.08
CA + AA	165 (44.59)	74 (20)	0.32 (0.23–0.45)	<0.0001 [§]	41 (7.85)	0.25 (0.51–2.13)	<0.0001 [§]	33 (29.72)	0.54 (0.33–0.87)	0.01 [§]
C Allele	546 (73.78)	659 (89.05)								
A Allele	194 (26.21)	81 (10.94)								
MAF	26.21	10.94								
HWE	$\chi^2 = 0.05$; df = 1; p = 0.81 [§]									
176555 G/C										
GG	264 (71.35)	253 (68.37)	1.00 (Reference)		178 (69.80)	1.00 (Reference)		72 (64.86)	1.00 (Reference)	
GC	93 (25.13)	106 (28.64)	1.19 (0.85–1.67)	0.29	72 (28.23)	1.16 (0.80–1.68)	0.43	33 (29.72)	1.27 (0.76–2.10)	0.34
CC	13 (3.51)	11 (2.97)	0.89 (0.38–2.10)	0.80	5 (1.96)	0.61 (0.21–1.79)	0.37	6 (5.40)	1.27 (0.39–4.08)	0.68
GC + CC	106 (28.64)	117 (31.62)	1.16 (0.84–1.60)	0.36	77 (30.19)	1.10 (0.51–2.13)	0.76	39 (35.13)	1.28 (0.79–2.06)	0.30
G Allele	621 (83.91)	612 (82.70)								
C Allele	119 (16.08)	128 (17.29)								
MAF	16.08	17.29								
HWE	$\chi^2 = 1.74$; df = 1; p = 0.18 [§]									
The wild type genotype for SNPs was regarded as reference in the table.										
†Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls.										
‡Adjusted odds ratios, 95% confidence intervals and their corresponding p-values were calculated logistic regression analysis after adjusting for age, gender and smoking status.										
§Significant p-value.										
NSCLC: Non-small-cell lung cancer.										

Haplotype analysis

Linkage disequilibrium and haplotype frequencies were analyzed using SHEsis program platform for the two *XPG* SNPs (Supplementary Table 1). A significant overall distribution was observed when haplotypic profile of *XPG* variants were compared (Global $\chi^2 = 63.31$; p = 1.43e-013). The two SNPs were in weak linkage disequilibrium with each other (D: 0.015; r^2 : 0.00). The G-A haplotype showed a significant protective effect against lung cancer (OR: 0.29, 95% CI: 0.21–0.41; p = 4.52e-015), whereas almost twofold risk was observed for G-C haplotype (OR: 1.87, 95% CI: 1.49–2.33; p = 2.93e-008).

Interactive effect of smoking on the *XPG* polymorphism

Gene–environment interaction was studied further by dividing the study subjects into smokers and nonsmokers (Table 3). The study population was divided into two groups on the basis of their smoking status, in other words, smokers and nonsmokers. The smoking group was then categorized in terms of pack years as light smokers (≤ 25 pack years) and heavy smokers (> 25 years). For rs2228959 C/A polymorphism, a protective effect was observed among smokers for all the three variant genotypes, in other words, heterozygous (CA; OR: 0.32, 95% CI: 0.22–0.48; p < 0.0001), mutant (AA; OR: 0.22; 95% CI: 0.08–0.58; p = 0.02) and combined genotype (CA + AA; OR: 0.31, 95% CI: 0.21–0.45; p < 0.0001). When smokers were segregated on the basis pack years, it was observed that heavy smokers carrying both mutant alleles (AA; OR: 0.35, 95% CI: 0.18–0.66; p = 0.001) as well as heterozygous carriers (CA; OR: 0.08, 95% CI: 0.01–0.73; p = 0.02) exhibited a protective effect for lung cancer

Table 3. Distribution of genotypes of *XPG* polymorphism according to their smoking status and its association with lung cancer risk.

Genotype	Smokers				Nonsmokers			
	Controls, n (%) n = 271	Cases, n (%) n = 303	AOR (95% CI) [‡]	p-value [†]	Controls, n (%) n = 99	Cases, n (%) n = 67	AOR (95% CI) [‡]	p-value [†]
2228959 C/A								
CC	146 (53.87)	242 (79.86)	1.00 (Reference)		54 (54.54)	59 (88.05)	1.00 (Reference)	
CA	106 (39.11)	55 (18.15)	0.32 (0.22–0.48)	<0.0001 [§]	12 (12.12)	30 (44.77)	0.47 (0.21–1.03)	0.06 [§]
AA	19 (7.01)	6 (1.98)	0.22 (0.08–0.58)	0.002 [§]	1 (1.01)	10 (14.92)	0.10 (0.01–0.83)	0.03 [§]
CA + AA	125 (46.12)	61 (20.13)	0.31 (0.21–0.45)	<0.0001 [§]	13 (13.13)	40 (59.70)	0.37 (0.17–0.77)	0.008 [§]
17655G/C								
GG	191 (70.47)	209 (68.97)	1.00 (Reference)		73 (73.73)	44 (65.67)	1.00 (Reference)	
GC	70 (25.83)	86 (28.38)	1.12 (0.76–1.65)	0.54	23 (23.23)	20 (29.85)	1.44 (0.70–2.99)	0.31
CC	10 (3.69)	8 (2.64)	0.54 (0.18–1.57)	0.26	3 (3.03)	3 (4.47)	1.96 (0.37–10.38)	0.42
GC + CC	80 (29.52)	94 (31.02)	1.05 (0.72–1.52)	0.77	26 (26.26)	23 (34.32)	1.52 (0.76–3.03)	0.23
	Heavy smoker				Light smoker			
2228959 C/A	Controls, n (%) n = 101	Cases, n (%) n = 141	AOR (95% CI) [‡]	p-value [†]	Controls, n (%) n = 170	Cases, n (%) n = 162	AOR (95% CI) [‡]	p-value [†]
CC	60 (59.40)	118 (83.68)	1.00 (Reference)		87 (51.17)	124 (76.54)	1.00 (Reference)	
CA	33 (32.67)	22 (15.60)	0.35 (0.18–0.66)	0.001 [§]	72 (42.35)	33 (20.37)	0.32 (0.19–0.55)	<0.0001 [§]
AA	8 (7.92)	1 (0.70)	0.08 (0.01–0.73)	0.02 [§]	11 (6.47)	5 (3.08)	0.35 (0.11–1.08)	0.06 [§]
CA + AA	41 (40.59)	23 (16.31)	0.30 (0.16–0.56)	0.0002 [§]	83 (48.82)	38 (23.45)	0.33 (0.20–0.54)	<0.0001 [§]
17655G/C								
GG	74 (73.26)	98 (69.50)	1.00 (Reference)		117 (68.82)	111 (68.51)	1.00 (Reference)	
GC	25 (24.75)	41 (29.08)	1.26 (0.69–2.30)	0.45	45 (26.47)	45 (27.77)	1.03 (0.62–1.71)	0.88
CC	2 (1.98)	2 (1.41)	0.75 (0.10–5.61)	0.78	8 (4.70)	6 (3.70)	0.51 (0.14–1.83)	0.30
GC + CC	27 (26.73)	43 (30.49)	1.22 (0.67–2.19)	0.50	53 (31.17)	51 (31.48)	0.96 (0.59–1.56)	0.88

[†]Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. All values in bold represent significant p-values. The term reference indicates wild genotype of the SNPs in the table.
[‡]Adjusted odds ratios, 95% CI and their corresponding p-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status.
[§]Significant p-value.
AOR: Adjusted odds ratio.

when compared with the reference group (CC). For light smokers with 2228959 C/A polymorphism, a protective effect was observed for heterozygous genotype (CA; OR: 0.32, 95% CI: 0.19–0.55; $p < 0.0001$) when compared with the reference genotype (CC). In addition, combined genotype (CA + AA) also showed a significant effect toward the disease (OR: 0.33, 95% CI: 0.20–0.54; $p < 0.0001$). Similarly for 2228959 C/A polymorphism, the subjects harboring mutant genotype and who were nonsmokers also expressed a significant protective effect for lung cancer (OR: 0.10, 95% CI: 0.01–0.83; $p = 0.03$) when compared with the reference genotype. In case of 17655 G/C polymorphism, no significant association was observed between lung cancer risk and smoking status.

Relationship of clinic-pathological features with *XPG* polymorphic variants

To understand the association between *XPG* SNPs (2228959 C/A, 17655 G/C) with various clinic-pathological features, we divided patients according to clinical stage (III vs IV), primary tumor extension (T3 + T4 vs T0 + T1 + T2) and metastasis extension (positive vs negative) (Supplementary Table 2). Only for rs17655 G/C variant, lung cancer subjects with T3 + T4 tumor and harboring the mutant genotype (CC) showed a protective effect as compared with those with T0 + T1 + T2 having mutant genotype (OR: 0.28, 95% CI: 0.08–1.01; $p = 0.05$).

Association of *XPG* polymorphic variants & OS of lung cancer patients

For 311 lung cancer patients, a survival analysis in relation to the polymorphic variants of the *XPG* gene was conducted as shown in Table 4. To evaluate the association between OS and *XPG* polymorphic variants, first, a univariate analysis was applied using Kaplan–Meier curve and log-rank test followed by a multivariate Cox regression model. The genotypic distribution in the patients agreed with that estimated from HWE for both 2228959 C/A

Table 4. Relationship of *XPG* polymorphism with overall survival of patients.

Genotype	Overall LC						
	Dead 256	Alive 55	Median OS (months)	HR (95% CI) [†]	Log p	HR [‡] (95% CI) [‡]	p-value [‡]
2228959 C/A							
CC	205	44	8.03	1.00 (Reference)		1.00 (Reference)	
CA	45	11	8.03	0.89 (0.65–1.22)	0.50	0.94 (0.67–1.31)	0.74
AA	6	–	8.85	1.19 (0.49–2.88)	0.66	1.14 (0.49–2.64)	0.74
CA + AA	51	11	7.65	0.92 (0.68–1.24)	0.59	0.97 (0.70–1.33)	0.87
17655 G/C							
GG	175	43	8.80	1.00 (Reference)		1.00 (Reference)	
GC	73	11	6.23	1.22 (0.92–1.63)	0.13	1.24 (0.93–1.66)	0.13
CC	8	1	5.83	1.48 (0.63–3.47)	0.26	1.39 (0.68–2.87)	0.36
GC + CC	81	12	6.23	1.25 (0.94–1.64)	0.09	1.28 (0.96–1.69)	0.08
NSCLC							
2228959 C/A	Dead 181	Alive 40	Median OS (months)	HR (95% CI)	Log p	HR [‡] (95% CI)	p-value
CC	155	32	7.93	1.00 (Reference)		1.00 (Reference)	
CA	22	8	8.16	0.76 (0.51–1.15)	0.24	0.76 (0.48–1.20)	0.25
AA	4	–	9.33	1.04 (0.37–2.87)	0.93	0.99 (0.36–2.74)	0.99
CA + AA	26	8	8.3	0.79 (0.54–1.17)	0.27	0.81 (0.55–1.24)	0.34
17655 G/C							
GG	125	32	9	1.00 (Reference)		1.00 (Reference)	
GC	53	7	5.9	1.42 (1.00–2.02)	0.03 [§]	1.41 (1.00–1.91)	0.04 [§]
CC	3	1	10.85	0.93 (0.30–2.82)	0.90	0.75 (0.23–2.39)	0.63
GC + CC	56	8	6	1.38 (0.98–1.94)	0.04 [§]	1.36 (0.97–1.91)	0.06
SLCLC							
	Dead 75	Alive 15	Median OS (months)	HR (95% CI)	Log p	HR [‡] (95% CI)	p-value
2228959 C/A							
CC	50	12	6.40	1.00 (Reference)		1.00 (Reference)	
CA	23	3	9.43	0.60 (0.36–1.00)	0.09	0.59 (0.32–1.09)	0.09
AA	2	–	22.53	0.59 (0.19–1.79)	0.46	0.60 (0.14–2.56)	0.50
CA + AA	25	15	12.20	0.59 (0.36–0.96)	0.06	0.59 (0.33–1.05)	0.07
17655 G/C							
GG	49	11	7.33	1.00 (Reference)		1.00 (Reference)	
GC	21	4	12	0.81 (0.49–1.33)	0.42	0.97 (0.53–1.77)	0.93
CC	5	–	2.96	2.08 (0.58–7.36)	0.10	2.55 (0.94–6.87)	0.05 [§]
GC + CC	26	4	9	0.92 (0.57–1.48)	0.75	1.16 (0.66–2.02)	0.59

The term reference indicates wild genotype of the SNPs in the table.

[†]Hazard ratios, 95% CIs and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and [‡]Adjusted hazard ratios, 95% confidence intervals and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.

[‡]Two-sided χ^2 test for either genotypic distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 1: combined hetero and mutant genotype.

[§]Significant p-values.

ECOG: Eastern cooperative oncology group; HR: Hazard ratio; KPS: Karnofsky performance status; NSCLC: Non-small-cell lung cancer; OS: Overall survival.

($\chi^2 = 1.70$; df = 1; p = 0.18) and 17655 G/C ($\chi^2 = 0.06$; df = 1; p = 0.79) polymorphisms. Homozygous wild genotype was considered as reference for both SNPs: 2228959 C/A (CC) and 17655 G/C (GG). We further divided patients on the basis of their histological subtypes and evaluated their OS. In case of 2228959 C/A polymorphism, the lung cancer subtypes did not show any significant association with OS. For SCLC patients, a longer survival was observed among mutant subjects (AA) as compared with patients with wild genotype (CC; MST: 22.53 vs 6.40 months, log-rank p = 0.50). When heterozygous genotype combined with the mutant genotype and used as a single genotype (CA + AA), a protective effect was observed when compared with the reference genotype (MST: 12.20 months; hazard ratio [HR]: 0.59, 95% CI: 0.33–1.05; p = 0.07).

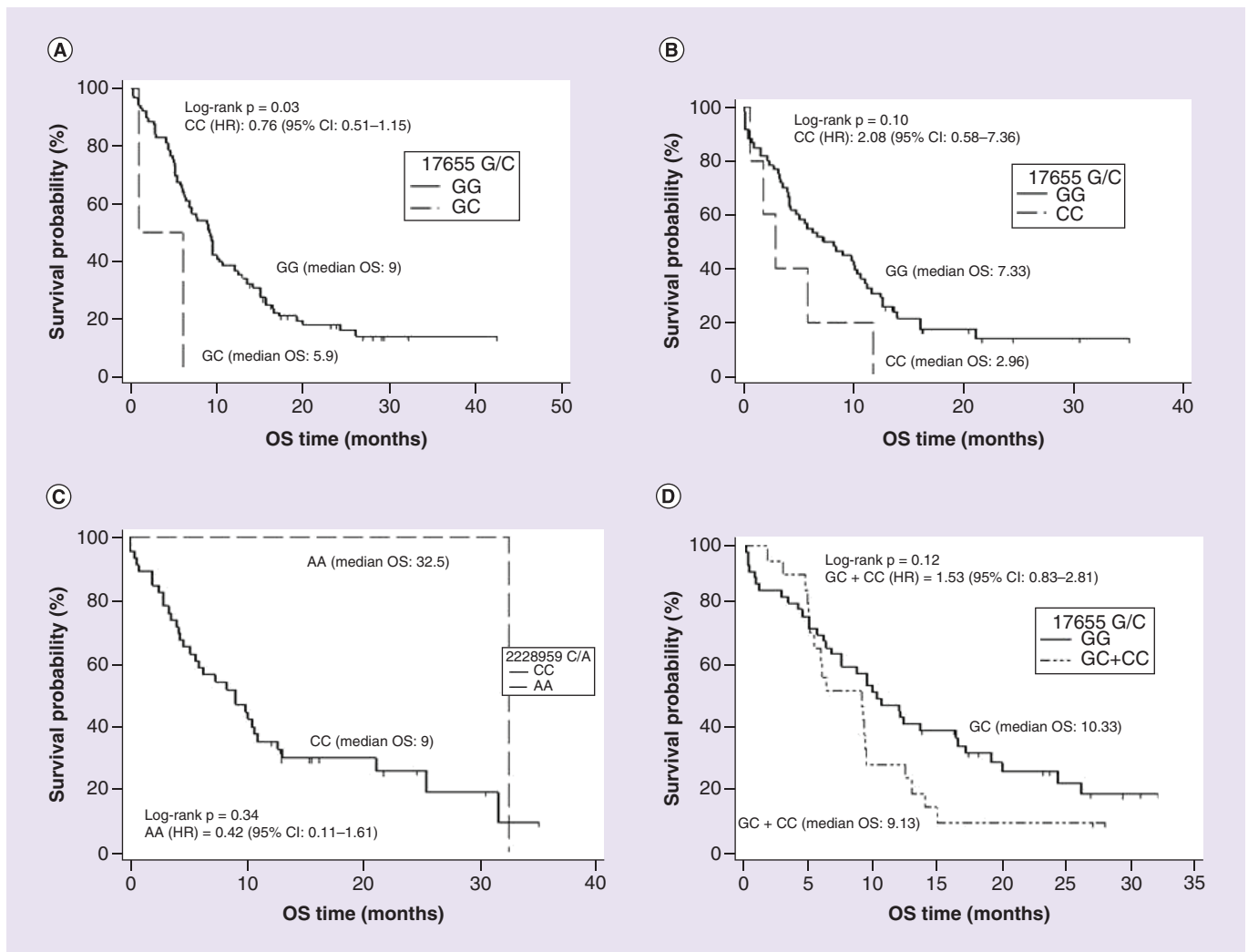


Figure 2. Kaplan–Meier curve. Curve depicting association between (A) 17655 G/C polymorphism and heterozygous genotype (GC) in NSCLC patients (B) 17655 G/C polymorphism and mutant genotype (CC) in SCLC patients (C) 2228959 C/A polymorphism and patients having mutant genotype (AA) treated with irinotecan-cisplatin/carboplatin regimen (D) 17655 G/C polymorphism and patients having combined genotype (GC + CC) treated with pemetrexed-cisplatin/carboplatin regimen. HR: Hazard ratio; NSCLC: Non-small-cell lung cancer; OS: Overall survival.

For 17655 G/C polymorphism, subjects with mutant genotype (AA) had shorter survival period as compared with the patients carrying wild genotype (GG; MST: 5.83 vs 8.80 months, log-rank $p = 0.26$). No significant association was observed between 17655 G/C polymorphism and OS in both Kaplan–Meier analysis and Cox regression model. For 17655 G/C polymorphism, when segregated according to histological subtypes, NSCLC patients with heterozygous genotype (GC) were associated with increased mortality when compared with reference genotype (HR¹: 1.41, 95% CI: 1.00–1.99; $p = 0.04$; MST: 5.9 months; Figure 2A). When Cox regression model was applied for the lung cancer patients carrying the combined genotype (GC + CC), an increased HR was observed although results were not significant (HR¹: 1.36, 95% CI: 0.97–1.91; $p = 0.06$; MST: 6 months). Among SCLC patients, a poor survival was observed for mutant subjects when compared with the reference genotype (2.96 vs 7.33 months) (Figure 2B). The SCLC patients carrying mutant genotype (CC) showed 2.5-fold HR when Cox multivariate analysis was applied (HR¹: 2.55, 95% CI: 0.94–6.81; $p = 0.05$).

Table 5. Association of *XPG* polymorphism and overall survival according to regimen

Regimen	No.	Alive	Dead	Median OS (months)	Crude HR (95% CI)	Adjusted HR ¹ (95% CI) [†]	Log-rank p	p-value ^{3‡}
REGIMEN – docetacel cisplatin/carboplatin								
Genotype (n = 76)								
<i>2228959 C/A:</i>								
– CC	65	10	55	8.03	1.00 (Reference)	1.00 (Reference)		
– CA	10	4	6	9.43	0.53 (0.27–1.04)	0.57 (0.23–1.38)	0.13	0.21
– AA	1	0	1	12.56	0.71 (0.13–3.83)	1.04 (0.13–8.35)	0.73	0.96
– CA + AA	11	4	7	12.56	0.54 (0.29–1.03)	0.60 (0.26–1.38)	0.12	0.23
<i>rs17655 G/C:</i>								
– GG	56	12	44	8.30	1.00 (Reference)	1.00 (Reference)		
– GC	18	1	17	6.73	1.20 (0.66–2.16)	1.00 (0.53–1.90)	0.50	0.97
– GC + CC	20	2	18	10.13	0.75 (0.44–1.26)	0.73 (0.37–1.43)	0.28	0.37
REGIMEN – irinotecan cisplatin/carboplatin								
Genotype (n = 70)								
<i>2228959 C/A:</i>								
– CC	46	12	34	9	1.00 (Reference)	1.00 (Reference)		
– CA	23	5	18	6.73	1.05 (0.59–1.87)	1.35 (0.70–2.59)	0.85	0.35
– AA	1	0	1	32.5	0.42 (0.11–1.61)	0.04 [§] (0.001–1.51)	0.34	0.08
– CA + AA	24	5	19	7.23	0.98 (0.56–1.72)	1.36 (0.71–2.59)	0.95	0.35
<i>17655 G/C:</i>								
– GG	48	12	36	9.00	1.00 (Reference)	1.00 (Reference)		
– GC	18	5	13	7.23	1.00 (0.53–1.90)	1.43 (0.67–3.05)	0.97	0.35
– CC	4	0	4	4.40	2.14 (0.51–8.96)	3.01 (0.92–9.81)	0.13	0.06
– GC + CC	22	5	17	7.23	1.15 (0.63–2.08)	1.65 (0.83–3.29)	0.62	0.14
REGIMEN – pemetrexed cisplatin/carboplatin								
Genotype (n = 69)								
<i>2228959 C/A:</i>								
– CC	62	13	49	9.51	1.00 (Reference)	1.00 (Reference)		
– CA	6	1	5	5.75	1.46 (0.49–4.29)	1.65 (0.60–4.53)	0.41	0.32
– AA	1	0	1	6.10	2.68 (0.10–65.79)	1.35 (0.15–11.73)	0.30	0.78
– CA + AA	7	1	6	6.10	1.57 (0.56–4.37)	1.65 (0.64–4.21)	0.28	0.29
<i>17655 G/C:</i>								
– GG	48	12	36	10.33	1.00 (Reference)	1.00 (Reference)		
– GC	19	2	17	6.43	1.54 (0.81–2.91)	1.79 (0.95–3.35)	0.13	0.07
– CC	2	0	2	10.85	1.41 (0.26–7.47)	1.33 (0.29–6.14)	0.63	0.71
– GC + CC	21	2	19	9.13	1.53 (0.83–2.81)	1.83 (0.97–3.44)	0.12	0.06

The term reference indicates wild genotype of the SNPs in the table.

[†]HR, 95% CI and their corresponding p-values were calculated by Kaplan–Meier survival analysis after adjusting for remission and survival in months and adjusted hazard ratios (HR¹), 95% CI and their corresponding p-values were calculated by Cox regression models adjusted for age, sex, smoking status, stage, KPS and ECOG.

[‡]Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls (p-value). $p < 0.05$ was considered as statistically significant.

0: Wild genotype; 1: Heterozygote genotype; 2: Mutant genotype; 1: Combined hetero- and mutant genotype; ECOG: Eastern Cooperative Oncology Group; HR: Hazard ratio; KPS: Karnofsky performance status; OS: Overall survival.

Association of *XPG* polymorphic variants & their association with OS in lung cancer patients treated with different chemotherapy regimens

The OS of patients treated with different chemotherapy regimens and their association with *XPG* polymorphic variants has been illustrated in Table 5. Homozygous wild genotype was considered as reference. In case of *2228959 C/A* polymorphism, it was conferred that lung cancer patients carrying mutant genotype (AA) and treated with docetacel cisplatin/carboplatin regimen showed a better survival as compared with patients carrying the wild-type genotype (CC; MST: 12.56 vs 8.03 months, log-rank $p = 0.12$). Similarly the patients having mutant genotype

(AA) and treated with irinotecan cisplatin/carboplatin regimen showed a longer OS and a low death ratio when compared with those patients harboring the wild-type genotype (MST: 32.5 vs 9; HR: 0.04, 95% CI: 0.001–1.51; $p = 0.08$) (Figure 2C).

For 17655 G/C polymorphism, the lung cancer subjects who were treated with irinotecan cisplatin/carboplatin regimen and carrying both the copies of mutant alleles (CC) showed a poor OS as compared with subjects receiving same chemotherapeutic regimen but who were having wild-type genotype (GG; MST: 4.40 vs 9.0 months) and a high risk for mortality with a threefold HR (HR: 3.01, 95% CI: 0.92–9.81; $p = 0.06$). Furthermore, In case of 17655 G/C polymorphism, patients having combined genotype (GC + CC) and treated with pemetrexed-cisplatin/pemetrexed-carboplatin regimen depicted almost twofold HR when compared with the wild genotype (HR: 1.83, 95% CI: 0.97–3.44; $p = 0.06$) (Figure 2D).

Association of XPG polymorphism & clinical outcomes

The relationship between XPG polymorphism and clinical outcomes was assessed by applying multivariate logistic regression adjusted for factors such as age, gender, smoking, stage, histology, KPS and ECOG (Supplementary Table 3). A total of 202 lung cancer patients were divided into two groups, in other words, responders (complete response + partial response) and nonresponders (stable disease + progressive disease). There was no significant association observed between XPG polymorphism and clinical outcomes.

Discussion

A widespread interest has been generated in studying DNA sequence variants, such as SNPs, as they play a potential role in human diseases [4]. Every individual differs in their capacity of repairing damaged DNA and furthermore SNPs present in DNA repair genes might be responsible for interindividual differences for repairing damaged DNA [3]. Therefore, this study is focused on the association of XPG variants (2228959 C/A, 17655 G/C) toward susceptibility for lung cancer in North Indian patients. We also evaluated relationship of these XPG variants with OS and clinical response toward platinum-based doublet chemotherapy.

In present study, we found that 2228959 C/A polymorphism was found to be strongly associated for conferring a protective effect toward lung cancer. As far our knowledge is concerned very few studies have been reported for this SNP and its association for lung cancer. Kumar *et al.* [9] have reported a lack of association for 2228959 C/A polymorphism toward breast cancer risk in Finnish population. This is the only study reported on 2228959 C/A polymorphism explaining its association toward cancer.

On the contrary most of the studies have investigated the association of 17655 G/C polymorphism toward tumorigenesis. Although located outside the conserved domains of the gene, the Asp to His amino acid change results in functional difference in the XPG gene at codon 1104 [19]. Our data suggest a lack association of 17655 G/C polymorphism toward lung cancer. Our results were concordant with the study conducted by Shen *et al.* [20] who have suggested no significant association between 17655 G/C polymorphism and lung cancer. Similarly, Lavanya *et al.* [21] have also reported no association of 17655 G/C polymorphism toward breast cancer susceptibility in Indian population. On the same lines, studies conducted in SCCHN and gastric cancer also reported a lack of association with 17655 G/C polymorphism [11,12]. However, some studies have pointed out an association between the 17655 G/C polymorphism and susceptibility for cancer. Kumar *et al.* [9] reported that mutant genotype (CC) of 17655 G/C polymorphism was found to be strongly associated toward risk for lung cancer in European population. A study on African-Americans [22] and Asians [22] have also conferred a positive association of 17655 G/C polymorphism toward lung cancer susceptibility. Study done on Korean lung cancer patients has [7] depicted an association of the mutant genotype (*His/His*) for conferring a protective effect toward lung cancer. It is biologically plausible to evaluate association between XPG polymorphism and lung cancer risk, as XPG plays an important role in NER pathway, which is responsible for repair of DNA lesions majorly induced by genotoxins like tobacco smoke. For repair, NER normally requires interaction between DNA repair gene products including XPG [23]. These genes are located on C-terminus of XPG, which is essential for interaction between XPB, XPD, p62 and p44 subunits of transcription factor IIH inside incision complex in NER pathway [23]. The C-terminal extremities are conserved in character than sequence consisting of basic amino acids [23]. Therefore, an amino acid change could alter the binding with other proteins present in incision complex and hence alter NER capacity. Another possible reason behind such function may be that these polymorphisms are present in linkage disequilibrium with another polymorphism that alters their gene function, or they are present in adjacent proximity of another susceptible gene.

When lung cancers were stratified by histological subtypes, our data suggest that 2228959 C/A showed a significant protective effect for both NSCLC and SCLC subtypes. For 17655 G/C polymorphism, no association with any histological subtypes of lung cancer was observed. Various evidences suggest that histological subtypes of the lung cancer are determined by the initiating agent to which an individual is exposed [24]. As polyaromatic hydrocarbons present in tobacco smoke have a strong relationship with adenocarcinoma, squamous cell carcinoma and SCLC, this result might have been due to the fact that *XPG* plays an important role in DNA damage repair caused by these polyaromatic hydrocarbons. However, this lack of statistical significance in case of 17655 G/C polymorphism might be due to low sample size or difference in ethnicities.

Level of carcinogen exposure defines its relationship with genetic susceptibility to lung cancer [7]. Therefore, association of *XPG* variants with lung cancer susceptibility was examined among smokers. Smokers who were lung cancer patients and carrying the mutant allele for 2228959 C/A polymorphism were observed to be protected toward disease ($p < 0.0001$). When smokers were dichotomized into two groups based on the pack years of smoking, a protective effect was observed in both the light and heavy smokers for 2228959 C/A variant. There are possible reasons behind such findings. At high doses of carcinogen exposure, the genetic differences in risk may be smaller until environmental factor overpowers the genetic predisposition [23]. Such results may be attributable to the chance due to low sample size. The study by Jeon *et al.* [7] supported the present study and suggested a protective effect among smokers for *XPG* polymorphism, whereas He *et al.* [25] showed disagreement with the study by emphasizing on association of *XPG* variants with high risk of lung cancer among smokers.

We further conducted a survival analysis to elaborate role of *XPG* polymorphism and OS in lung cancer patients. There was no significant association observed for both *XPG* variants (2228959 C/A, 17655 G/C) and OS. When categorized according to histological subtype, a poor survival for SCLC patients with mutant genotype (CC) was reported showing 2.5-fold HR for rs17655 G/C polymorphism. We further evaluated association between *XPG* variants and OS according to chemotherapeutic regimen. It is known that selection of chemotherapeutic regimen could improve response and hence OS of cancer patients. A longer survival period of 32.5 months was observed in patients having mutant genotype (AA) and treated with irinotecan-cisplatin/carboplatin regimen for 2228959 C/A polymorphism. A lower death rate was observed in mutant subjects (AA) treated with the above regimen for 2228959 C/A variant of *XPG* (HR: 0.04). For 17655 G/C polymorphism, a higher death rate was observed in patients carrying mutant genotype (AA) treated with irinotecan-cisplatin/carboplatin regimen (HR: 3.01; $p = 0.06$). Irinotecan consists of SN-38 active metabolite and interacts with topoisomerase I and forms a stable complex with DNA [26]. Irinotecan has always demonstrated its efficacy in SCLC treatment but it also show a promising effect in NSCLC patients producing a response rate to 32% [26]. No previous study has been reported till now on association of *XPG* gene and OS for lung cancer patients treated with irinotecan-cisplatin/carboplatin regimen. A higher mortality rate in patients carrying combined genotype (GA C AA) treated with pemetrexed-cisplatin/carboplatin was observed for 17655 G/C polymorphism (HR: 1.83; $p = 0.06$).

DNA repair is essential for safeguarding the genome stability and integrity and *XPG* protein is required for making incision at the 3' site of DNA damage during NER pathway [5]. NER is associated with resistance to platinum agents, and reduced NER might be related with the increased platinum agent sensitivity. Therefore, *XPG* polymorphism may affect the clinical outcomes of cancer patients. Previous studies have suggested association of *XPG* polymorphism with predictive clinical response of various cancers [27]. The current study did not show any association of *XPG* polymorphism with clinical outcomes. Our study was supported by Zienolddinny *et al.* [28], Bartolucci *et al.* [29], Sun *et al.* [30], Butkiewicz *et al.* [31] and Zhang *et al.* [14] suggested no significant association between *XPG* polymorphism and clinical outcomes in NSCLC patients. A study on Chinese population predicted association of *XPG* polymorphism toward prognosis of NSCLC patients [27].

The study had several limitations that should be considered. First, there are chances of selection bias and information bias in the current data; all lung cancer patients were selected from a single hospital, and therefore these cases may not be representative of patients from other population. The biasness could be reduced by adjusting the final analysis for gender, age, smoking status, ECOG, KPS and histology which could minimize the selection bias. Second, we have investigated only two SNPs in *XPG* gene with limited sample size.

Conclusion

Our study provides statistical evidence that 2228959 C/A was associated with protective effect in lung cancer and its histological subtypes. For 17655 G/C polymorphism, no association with lung cancer susceptibility was observed. No significant association was observed for *XPG* variants with OS and clinical outcomes in North Indian

population. Therefore, large sample studies are required for further confirming functional significance of *XPG* polymorphism and its role in cancer susceptibility, OS and clinical outcomes.

Future perspective

This study has provided insight about the association of *XPG* polymorphism with lung cancer susceptibility, OS and clinical response in North Indian patients that were not explored before the present study. Some remarkable results were observed depicting association of *XPG* gene with lung cancer risk. The present study has stressed upon association of 2228959 C/A polymorphism with strong protective effect against lung cancer risk as well as its histological subtypes (i.e., NSCLC and SCLC) while 17655 G/C polymorphism was associated with insignificant lung cancer risk. Lung cancer patients treated with different chemotherapy regimens, in other words, irinotecan-cisplatin/irinotecan-carboplatin regimen, pemetrexed-cisplatin/pemetrexed-carboplatin regimens and their association with *XPG* polymorphism was also analyzed. It is known that selection of chemotherapeutic regimen as per the molecular marker could improve response rate and hence OS of cancer patients. This study in future can be regarded to facilitate assessment of genetic variations of *ERCC5* and help patients to make therapeutic decisions for individualized therapy in advanced lung cancer patients. As the present study is performed on a small sample size, for further authentication of the results, we need large sample studies that can confirm functional significance of *XPG* polymorphism and its role in cancer susceptibility, OS and clinical outcomes. The present study can be used as reference for those large sample studies.

Summary points

- The findings of the present study suggest that both heterozygous ($p < 0.0001$) and mutant ($p = 0.0001$) genotypes possess a significant protective effect for 2228959 C/A polymorphism. When stratified according to histological subtypes, non-small-cell lung cancer patients showed protective effect for both heterozygous ($p < 0.0001$) and mutant ($p = 0.0004$) genotypes. 17655 G/C polymorphism did not show any significant association with lung cancer susceptibility.
- For 2228959 C/A polymorphism, a protective effect was observed among smokers for all the three variant genotypes, in other words, heterozygous genotype (CA; $p < 0.0001$), mutant genotype (AA; $p = 0.02$) and combined genotype (CA + AA; $p < 0.0001$).
- When smokers were segregated on the basis pack years, heavy smokers with mutant (AA; $p = 0.001$) as well as heterozygous (CA; odds ratio [OR]: 0.08, 95% CI: 0.01–0.73; $p = 0.02$) genotype exhibited a protective effect for lung cancer. Light smokers with rs2228959 C/A polymorphism showed a protective effect for heterozygous genotype (CA; OR: 0.32, 95% CI: 0.19–0.55; $p < 0.0001$). rs17655 G/C polymorphism did not show any significant association between lung cancer risk and individuals smoking status.
- For rs17655 G/C polymorphism, mutant genotype (CC) showed a protective effect having T3 + T4 tumor stage with T0 + T1 + T2 stage as a reference (OR: 0.28, 95% CI: 0.08–1.01; $p = 0.05$).
- A weak linkage disequilibrium was observed between the rs2228959 C/A and rs17655 G/C ($D = 0.015$; $r^2 = 0.00$). G-A haplotype was associated with protective effect against lung cancer (OR: 0.29, 95% CI: 0.21–0.41; $p = 4.52 \times 10^{-15}$) while G-C haplotype showed a significant risk (OR: 1.87, 95% CI: 1.49–2.33; $p = 2.93 \times 10^{-8}$).
- *XPG* variants did not show any association with overall survival in lung cancer patients. When stratified according to histological subtypes, for rs2228959 C/A polymorphism, SCLC patients had a longer survival for mutant subjects (AA) as compared with patients with wild genotype (CC; 22.53 vs 6.40). For rs17655 G/C polymorphism, non-small-cell lung cancer patients showed increased death ratio in individuals carrying heterozygous genotype (hazard ratio [HR¹]: 1.41; $p = 0.04$) while SCLC patients carrying mutant genotype (CC) showed a poor survival with 2.5-fold HR when Cox multivariate analysis was applied (HR¹: 2.55, 95% CI: 0.94–6.81; $p = 0.05$; median survival time = 2.96).
- In a subgroup analysis based on chemotherapeutic regimens, there was no significant association observed for overall survival and *XPG* polymorphic variants when treated with different chemotherapy regimens. Patients with mutant genotype (AA) and treated with irinotecan cisplatin/carboplatin regimen showed a significant protective effect with longer survival for rs2228959 C/A polymorphism (HR¹: 0.04, 95% CI: 0.001–1.51; $p = 0.08$; median survival time: 32.5). Patients having combined genotype (GC + CC) and treated with pemetrexed-cisplatin/carboplatin regimen depicted almost twofold HR for rs17655 G/C polymorphism (HR¹: 1.83, 95% CI: 0.97–3.44; $p = 0.06$).
- *XPG* variants did not show any association with clinical response.
- Our study provides statistical evidence that rs2228959 C/A was associated with protective effect in lung cancer and its histological subtypes. *XPG* polymorphism might be associated with unfavorable prognosis in lung cancer risk; for that, further large sample studies are required to confirm the data.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: <https://www.futuremedicine.com/doi/suppl/10.2217/fon-2018-0408>

Acknowledgements

The authors would like to express our gratitude to all the subjects who participated in this current study.

Financial & competing interests disclosure

This work was supported by grant from the Indian Council of Medical Research, New Delhi, India (Grant Number 5/13/126/2011/NCD-III). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

- Jemal A, Siegel R, Xu J *et al.* Cancer statistics. *CA Cancer J. Clin.* 60(5), 277–300 (2010).
- Shields PG, Harris CC. Cancer risk and low-penetrance susceptibility genes in gene–environment interactions. *J. Clin. Oncol.* 18, 2309–2315 (2000).
- Wood RC, Mitchell M, Sgouros J *et al.* Human DNA repair genes. *Science* 291, 1284–1288 (2001).
- deLaat WL, Jaspers NG, Hoeijmakers JH. Molecular mechanism of nucleotide excision repair. *Genes Dev.* 13, 768–785 (1999).
- Wakasugi M, Reardon JT, Sancar A. The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair. *J. Biol. Chem.* 272, 16030–16034 (1997).
- Shen B, Nolan JP, Sklar LA *et al.* Essential amino acids for substrate binding and catalysis of human flap endonuclease. *J. Biol. Chem.* 271, 9173–9176 (1996).
- Jeon HS, Kim KM, Park SH *et al.* Relationship between XPG codon 1104 polymorphism and risk of primary lung cancer. *Carcinogenesis* 24, 1671–1681 (2003).
- **The primers for the current study has been selected for rs17655 G/C polymorphism. A detailed mode of mechanism regarding rs17655 G/C has been mentioned in this study. This study also explains relationship between rs17655 G/C polymorphism and lung cancer risk.**
- Cheng L, Eicher SA, Guo Z *et al.* Reduced DNA capacity in head and neck cancer patients. *Cancer Epidemiol. Biomarkers Prev.* 7, 465–468 (1998).
- Kumar R, Hoglund L, Zhao C *et al.* Single nucleotide polymorphisms in the XPG gene: determination of role in the DNA repair and breast cancer risk. *Int. J. Cancer* 103, 671–675 (2003).
- **It is the only study performed on XPG Phe670Leu polymorphism till now. The primer for the rs2228959 C/A polymorphism was selected from the above study. This study also explains a mechanism of action of rs2228959 C/A and rs17655 G/A polymorphism.**
- Emmert S, Schneider TD, Khan SG *et al.* The human XPG gene: gene architecture, alternative splicing and single nucleotide polymorphisms. *Nucleic Acids Res.* 29, 1443–1452 (2001).
- **This is one of the very first study which explained the structure and function of the XPG gene. This study also explained association of XPG gene with breast cancer risk.**
- Hussain SK, Mu LN, Cai L *et al.* Genetic variation in immune regulation and DNA repair pathways and stomach cancer in China. *Cancer Epidemiol. Biomarkers Prev.* 18, 2304–2309 (2009).
- An J, Liu Z, Hu Z *et al.* Potentially functional single nucleotide polymorphisms in the core nucleotide excision repair genes and risk of squamous cell carcinoma of the head and neck. *Cancer Epidemiol. Biomarkers Prev.* 16, 1633–1638 (2007).
- Molina JR, Yang P, Cassivi SD *et al.* Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* 83, 584–594 (2008).
- Zhang L, Gao G, Li X *et al.* Association between single nucleotide polymorphisms (SNPs) and toxicity of advanced non-small-cell lung cancer patients treated with chemotherapy. *PLoS ONE* 7, e48350 (2012).
- Yuli Y, Zhe S, Xia W *et al.* XPG is a novel biomarker of clinical outcome in advanced non-small-cell lung cancer. *Pak. J. Med. Sci.* 29(3), 762–767 (2013).
- Caiola E, Porcu L, Fruscio R *et al.* DNA-damage response gene polymorphisms and therapeutic outcomes in ovarian cancer. *Pharmacogenomics J.* 13(2), 159–172 (2013).

17. Biason P, Hattinger CM, Innocenti F *et al.* Nucleotide excision repair gene variants and association with survival in osteosarcoma patients treated with neoadjuvant chemotherapy. *Pharmacogenomics J.* 12(6), 476–483 (2012).
18. Lawania S, Singh N, Behera D *et al.* XPD polymorphism towards lung cancer susceptibility, overall survival and clinical response. *Future Oncol.* 13(22), 1945–1965 (2017).
- **This is similar kind of study performed on other gene. It might help in understanding the mechanism and procedure for this study as well as act as base for studies in patients treated in different regimens.**
19. Klungland A, Hoss M, Gunz D *et al.* Base excision repair of oxidative DNA damage activated by XPG protein. *Mol. Cell.* 3, 33–42 (1993).
20. Shen M, Berndt SI, Rothman N *et al.* Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. *Int. J. Cancer* 116, 768–773 (2005).
21. Lavanya J, Kumar VJ, Sudhakar N *et al.* Analysis of DNA repair genetic polymorphism in breast cancer population. *Int. J. Pharm. Bio. Sci.* 6(3), (B) 966–973 (2015).
- **It is the only study that explains the association of XPG gene with breast cancer risk in Indian population.**
22. Cheng L, Eicher SA, Guo Z *et al.* Reduced DNA repair capacity in head and neck cancer patients. *Cancer Epidemiol. Biomarkers Prev.* 7, 465–468 (1998).
23. Shiomi T, Harada Y, Saito T *et al.* An ERCC5 gene with homology to yeast RAD2 is involved in group G xeroderma pigmentosum. *Mutat. Res.* 314, 167–175 (1994).
24. Hoffman D, Rivenson A, Murphy SE *et al.* Cigarette smoking and adenocarcinoma of the lung: the relevance of nicotine-derived nitrosamines. *J. Smoking Relat. Disord.* 4, 165–190 (1993).
25. He J, Qiu LX, Wang MY *et al.* Polymorphisms in the XPG gene and risk of gastric cancer in Chinese populations. *Hum. Genet.* 131, 1235–1244 (2012).
26. Fukuda M, Oka M, Soda H *et al.* Phase II study of irinotecan combined with carboplatin in previously untreated non-small-cell lung cancer. *Cancer Chemother. Pharmacol.* 154, 573–577 (2004).
- **This is the first study that depicts association between nucleotide excision repair polymorphism and overall survival in patients treated with irinotecan-cisplatin/carboplatin regimen. It gives detail description of mechanism of irinotecan with DNA repair genes and also explains its impact on lung cancer patients when infused in combination with the platinum-based drug.**
27. Zou HZ, Zhao YQ. XPG polymorphisms are associated with prognosis of advanced non-small cell lung cancer treated with platinum-based doublet chemotherapy. *Genet. Mol. Res.* 14(1), 500–506 (2015).
- **This study explains role of XPG variants in overall survival and clinical response in non-small-cell lung cancer patients.**
28. Zienoldiny S, Campa D, Lind H *et al.* Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis* 27, 560–567 (2006).
29. Bartolucci R, Wei J, Sanchez JJ *et al.* XPG mRNA expression levels modulate prognosis in respected non-small-cell lung cancer in conjunction with BRCA1 and ERCC1 expression. *Clin. Lung Cancer* 10, 47–52 (2009).
30. Sun X, Li F, Sun N *et al.* Polymorphisms in XRCC1 and XPG and response to platinum-based chemotherapy in advanced non-small cell lung cancer patients. *Lung Cancer* 65, 230–236 (2009).
31. Butkiewicz D, Rusin M, Sikora B *et al.* An association between DNA repair gene polymorphisms and survival in patients with respected non-small cell lung cancer. *Mol. Biol. Rep.* 38, 5231–5241 (2011).
- **It is the important study explaining role of XPG polymorphism and overall survival in non-small-cell lung cancer patients.**

