

Prognostic significance of *MGMT* & *ERCC1* gene polymorphism in lung cancer patients treated with platinum-based doublet chemotherapy

A Dissertation submitted in partial fulfillment of the requirement for the award
of degree of
Master of Technology
In
Biotechnology



Submitted by

Anu Gupta

Roll No.: 601604002

Under the guidance of

Dr. Siddharth Sharma

Associate Professor

**DEPARTMENT OF BIOTECHNOLOGY
THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY
PATIALA-147004
July-2018**

DECLARATION

I hereby declare that work done in this seminar report entitled, "Prognostic significance of *MGMT* & *ERCC1* gene polymorphism in lung cancer patients treated with platinum-based doublet chemotherapy" submitted towards partial fulfilment of requirement for award of **Master of Technology** degree in Biotechnology in Biotechnology Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of work carried out by me under the supervision and guidance of **Dr. Siddharth Sharma, Associate Professor of Biotechnology Department, Thapar Institute of Engineering and Technology, Patiala.**

This matter embodied in this report has not been submitted in part or full to any other university or institute for the award of any degree.



Anu Gupta

Roll No. 601604002

M.Tech (Biotech)

Date: 04-09-2018

Place: Patiala

CERTIFICATE

This is to certify that the dissertation entitled “Prognostic significance of *MGMT* and *ERCC1* gene polymorphism in lung cancer patients treated with platinum-based doublet chemotherapy” submitted for the degree of Master of Technology in the subject of Biotechnology at Thapar Institute of Engineering and Technology (TIET), Patiala is a bonafide work carried out by **Ms. Anu Gupta** under my supervision and guidance.

To the best of my knowledge, the matter embodied in this dissertation has not been submitted to any other university/institute for award of any Degree or Diploma



Dr. Siddharth Sharma

Associate Professor

Department of Biotechnology

Thapar Institute of Engineering and Technology, Patiala

ACKNOWLEDGEMENT

I want to express my deepest gratitude and sincere thanks to the following people without whom my thesis could not have been possible. I thank the Almighty for His blessings in the completion of the project.

First and foremost, I am sincerely grateful from the bottom of my heart to my supervisor and guide, **Dr. Siddharth Sharma** for his continuous support, trust, motivation, patience, enthusiasm, and knowledge which helped me to successfully carry out this project. I am thankful to him for resolving every problem, guidance and friendly advice in every phase of the project which helped in shaping this thesis perfectly.

I express my thanks to **Dr. Moushumi Ghosh**, Professor and Head, Department of Biotechnology, Thapar University, Patiala, for providing with the best laboratory facilities.

I express my deepest thanks to **Kinshuk Sharma** and **Harleen Walia** who despite of her completion of project, was constantly resolving all my doubts and encouraging me at every step.

I am deeply thankful to my labmates and my dear friends, **Ms Kirti Rai and Divya Alhuwalia** for being with me through thick and thin.

I was fortunate enough to avail help and assistance from department's laboratory staff, **Mr. Ram Nawal, Mr. Lallan, Mr. Surinder, Mr. Prabhat and Mr. Phoolchand**.

A special thanks to my M.Tech and M.Sc friends for their direct or indirect support and motivation throughout the project.

I am sincerely thankful to all of you.



Anu Gupta

(601604002)

ABSTRACT

Lung cancer is the leading cause of death worldwide. Non-small cell lung cancer (NSCLC) contributes about 80-85 % of the lung cancer and rest is contributed by small cell lung cancer (SCLC). It has been reported that genetic variation in DNA repair genes may affect the chemotherapeutic response among non-small cell lung cancer patients. *MGMT* gene and *ERCC1* gene plays an important role in DNA direct reversal pathway and Nucleotide Excision repair pathway respectively. Present study focuses on the chemotherapeutic response to platinum based drug in lung cancer patients of North India related to genetic variants of *MGMT* and *ERCC1*. It includes a total of 214 patients treated with platinum based chemotherapy (docetaxel, irinotecan or pemetrexed in combination with cisplatin/carboplatin). Prognosis was observed for *MGMT* (rs12917) and *ERCC1* (rs3212986) individually and in combination. PCR-RFLP methodology was used to classify the patients into wild type, heterozygous type and mutant type genotype. The genotypic distribution of *MGMT* and *ERCC1* (individually as well as in combination) falls within Hardy-Weinberg equilibrium. Univariate analysis showed a non-significant association of genetic polymorphism in *MGMT* (rs12917) and *ERCC1* (rs3212986) (when analysed individually and in combination) with overall survival on the basis of parameters like genotype, smoking status, pack per years, gender, regimen and histology. Also when analysed by logistic regression method, a non-significant relationship of genetic polymorphism in *MGMT* (rs12917) and *ERCC1* (rs3212986) (when analysed individually and in combination) with the tumor stage, primary tumor extension, metastasis, lymph node involvement, ECOG, KPS and clinical response among lung cancer patients was also observed. Furthermore genetic polymorphism in *MGMT* (rs12917) and *ERCC1* (rs3212986) (when analysed individually and in combination) was associated with heavy smokers and not associated with chemotherapeutic regimens. In-silico analysis showed that *ERCC1* gene polymorphism present outside of protein domain and was tolerated in nature; while results were contradictory and inconsistent for *MGMT* that according to some analysis this was observed inside of protein domain and deleterious in nature whereas according to other software was present outside of protein domain and its effect was neutral in nature. In conclusion large sample sizes are needed to evaluate the effect of these polymorphism on overall survival.

Keywords: NSCLC, SCLC, *MGMT*, *ERCC1*, Hardy-Weinberg equilibrium, Univariate and multivariate analysis, logistic regression analysis, Prognosis.

TABLE OF CONTENTS

Sr. No.	Content	Page No.
1	Declaration	i
2	Certificate	ii
3	Acknowledgement	iii
4	Abstract	iv
5	Abbreviations used	x-xi
6	List of tables	xii-xiii
7	List of figures	xiv-xviii
Chapter 1	Introduction	1-2
Chapter 2	Review of literature	3-26
2.1	Cancer	
2.1.1	Introduction	3
2.1.2	Types of cancer	3
2.1.3	Cancer statistics	4
2.1.4	Risk factors of cancer	5-6
2.2	Lung cancer	
2.2.1	Introduction	7
2.2.2	Classification of lung cancer	7
2.2.3	Statistics of lung cancer	7-9
2.2.4	Lung cancer stages	9
2.2.5	Lung cancer risk factors	9
2.2.5.1	Lung cancer and smoking in India	10
2.3	DNA damage and DNA repair system	10-11
2.4	<i>MGMT</i>	11
2.4.1	Introduction	11
2.4.2	Protein structure	12
2.4.3	Variants of <i>MGMT</i>	13
2.4.4	Leu84Phe (rs12917)	14
2.4.5	<i>MGMT</i> and chemotherapy	14
2.4.6	Relation of <i>MGMT</i> with cancer	15
2.5	<i>ERCC1</i>	
2.5.1	Introduction	16

2.5.2	Protein structure	16
2.5.3	NER pathway	17
2.5.3.1	Role of <i>ERCC1</i> in NER pathway	17-18
2.5.4	<i>ERCC1</i> variants	18
2.5.5	Relation of overall survival with <i>ERCC1</i>	18-19
2.6	Platinum based chemotherapy	19-21
2.6.1	Cisplatin and lung cancer	21-22
2.7	Gaps in study	23
2.8	Computational analysis of deleterious SNP of MGMT and ERCC1	24-26
Chapter 3	Objectives	27
Chapter 4	Material and Methods	
4.1	Sample collection	28
4.2	In-vitro analysis	
4.2.1	Genomic DNA isolation	28-32
4.2.2	Polymerase chain reaction	32-33
4.2.3	Restriction fragment length polymorphism of <i>MGMT</i> and <i>ERCC1</i>	33-34
4.3	Statistical analysis	35
4.4	Bioinformatics methods	35
4.4.1	Prediction of structural impact of nsSNP on protein by SIFT software	35
4.4.2	Prediction of functional effect of amino acid substitution by Provean software	36
4.4.3	Prediction of deleterious SNP by Polyphen2 software	36
4.4.4	To predict nsSNP impact on stability of protein by I-Mutant software	37
4.4.5	Prediction of stability of protein on the basis of change in energy by MuPro software	38
4.4.6	To anticipate whether the mutation is disease related or neutral	38-39
4.4.7	To predict whether mutation is related to disease or not by SNP & GO software	39
4.4.8	Prediction of effect of mutations on folding of protein and protein-protein interactions by ELASPIC software	39-40
4.4.9	Prediction of deleterious SNP with the help of PANTHER	40-41

4.4.10	Prediction of harmful mutation by MutPred Software	42
4.4.11	Prediction of spatial arrangement of amino acid substitution on protein models and structures with the help of Mutation 3D software	43
4.4.12	Prediction of interaction of gene with other genes with the help of Genemania	43-44
Chapter 5	Results	
5.1	Genotyping	
5.1.1	DNA isolation	45
5.1.2	Amplification of <i>MGMT</i> and <i>ERCC1</i> gene	45-46
5.1.3	Restriction digestion of <i>MGMT</i> and <i>ERCC1</i> gene	47-48
5.2	Statistical analysis	
5.2.1	Distribution of <i>MGMT</i> and <i>ERCC1</i> gene genotype	48-49
5.2.2	Demographic distribution of lung cancer patients for <i>MGMT</i> and <i>ERCC1</i> gene	49-54
5.2.3	Demographic distribution of lung cancer patients for combined effect of <i>MGMT</i> and <i>ERCC1</i> gene	55-59
5.2.4	Overall survival analysis by Kaplan-Meier and cox regression in lung cancer patients	60-64
5.2.5	Overall survival analysis by Kaplan Meier and cox regression in lung cancer patients on the basis of histology	
5.2.5.1	Overall survival analysis among SCLC cases	65-69
5.2.5.2	Overall survival analysis among ADCC cases	69-74
5.2.5.3	Overall survival analysis among SQCC cases	74-79
5.2.6	Overall survival analysis by Kaplan Meier and cox regression in lung cancer patients on the basis of smoking habits (light smokers, heavy smokers)	
5.2.6.1	Overall survival analysis on the basis of light smokers	79-81
5.2.6.2	Overall survival analysis on the basis of heavy smokers	82-84
5.2.7	Overall survival analysis by Kaplan Meier and cox regression in lung cancer patients the basis of smoking status (smokers, non-smokers)	
5.2.7.1	Overall survival analysis among smokers	84-86
5.2.7.2	Overall survival analysis among non-smokers	86-88

5.2.8	Overall survival of lung cancer patients treated with different chemotherapeutic regimens	
5.2.8.1	Overall survival of lung cancer patients treated with docetaxel + cisplatin/carboplatin and its association with <i>MGMT</i> and <i>ERCC1</i> genotype	88-90
5.2.8.2	Overall survival of lung cancer patients treated with irinotecan + cisplatin / carboplatin and its association with <i>MGMT</i> and <i>ERCC1</i> genotype	90-92
5.2.8.3	Overall survival of lung cancer patients treated with pemetrexed in combination with cisplatin/carboplatin and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype	92-94
5.2.9	Overall survival of lung cancer patients and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype on the basis of gender	
5.2.9.1	Overall survival of female lung cancer patients and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype	94-96
5.2.9.2	Overall survival of male lung cancer patients having <i>MGMT</i> and <i>ERCC1</i> genotype	96-98
5.2.10	Logistic regression analysis of lung cancer patients on the basis of tumor stage and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype	98-100
5.2.11	Logistic regression analysis of lung cancer patients on the basis of primary tumor extension and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype	100-102
5.2.12	Logistic regression analysis of lung cancer patients on the basis of lymph node invasion and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype	102-103
5.2.13	Logistic regression analysis of lung cancer patients on the basis of metastasis and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype	104-105
5.2.14	Logistic regression analysis of lung cancer patients on the basis of ECOG and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype	105-107
5.2.15	Logistic regression analysis of lung cancer patients on the basis of KPS and its relationship with <i>MGMT</i> and <i>ERCC1</i> genotype	107-108
5.2.16	Logistic regression analysis of lung cancer patients on the basis of clinical response and its relationship with <i>MGMT</i> and <i>ERCC1</i>	109-110

	genotype	
5.3	Computational analysis of gene polymorphism in <i>MGMT</i> and <i>ERCC1</i> gene by using multiple bioinformatics tools	110-130
Chapter 6	Discussion	131-135
Chapter 7	Conclusion	136-137
	References	138-145

ABBREVIATIONS USED

APS	Ammonium per sulfate
BER	Base excision repair
BSA	Bovine serum albumin
DNA	Deoxyribonucleic acid
DRR	DNA direct repair pathway
DSBs	Double strand breaks
EDTA	Ethylene diamine tetra acetic acid
<i>ERCC1</i>	Excision repair cross complementing enzyme 1
<i>ERCC2</i>	Excision repair cross complementing enzyme 2
<i>ERCC3</i>	Excision repair cross complementing enzyme 3
<i>ERCC4</i>	Excision repair cross complementing enzyme 4
<i>ERCC5</i>	Excision repair cross complementing enzyme 5
HR	Homologous recombination
ICMR	Indian council of medical research
<i>MGMT</i>	O ⁶ –methylguanine–DNA methyl transferase
MMEJ	Micro-homology mediated end joining
MMR	Mis-match repair
NaCl	Sodium chloride
NCRP	National cancer registry programme
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NSCLC	Non-small cell lung cancer
PAHs	Polycyclic Aromatic Hydrocarbons
PGIMER	Post Graduate Institute Medical Education and Research
POLYPHEN2	Polymorphism Phenotyping v2
PROVEAN	Protein Variation Effect Analyzer
PY	Pack year
RPA	Replication protein A
SCLC	Small cell lung cancer
SDS	Sodium dodecyl sulfate
SEER	Surveillance, epidemiology and end result program

SIFT	Separating intolerant from tolerant
SNP	Single nucleotide polymorphism
TBE	Tris-borate
TEMED	Tetramethylethylenediamine
TFIIH	Transcription factor II H
TNM	Tumor, node, metastasis
TSNA	Tobacco specific nitrosamine
US	United state
XPA	Xeroderma pigmentosum complementation group A
XPC	Xeroderma pigmentosum complementation group C
XPF	Xeroderma pigmentosum complementation group F
XPG	Xeroderma pigmentosum complementation group G

LIST OF TABLES

Table No.	Title	Page No.
2.1	Variants of <i>MGMT</i>	13
2.2	Bioinformatics tool with URL	26
4.1	Preparation of washing buffer	29
4.2	Preparation of lysis buffer	29
4.3	Preparation of Tris-EDTA buffer	29
4.4	Preparation of TBE buffer and 6X loading dye	31
4.5	Reagents used in PCR	33
4.6	Temperature and time conditions for PCR	33
4.7	Components of 6% polyacrylamide gel	34
5.1	Demographic distribution of lung cancer patients for <i>MGMT</i> and <i>ERCC1</i>	49
5.2	Demographic distribution of lung cancer patients for combined effect of <i>MGMT</i> and <i>ERCC1</i>	55
5.3	Overall survival analysis by Kaplan-Meier and cox regression in lung cancer patients	60
5.4	Overall survival analysis on the basis of SCLC	65
5.5	Overall survival analysis on the basis of ADCC	69
5.6	Overall survival analysis on the basis of SQCC	74
5.7	Overall analysis on the basis of Light Smokers	79
5.8	Overall survival analysis on the basis of Heavy Smokers	82
5.9	Overall survival on the basis of smokers	84
5.10	Overall survival analysis on the basis of non-smokers	86
5.11	Overall survival analysis on the basis of regimen docetaxel + cisplatin/carboplatin	88
5.12	Overall survival analysis on the basis of regimen irinotecan + cisplatin/carboplatin	90
5.13	Overall analysis on the basis of regimen pemetrexed + cisplatin/carboplatin	92
5.14	Overall survival analysis among female lung cancer patients	94
5.15	Overall survival analysis among male lung cancer patients	96
5.16	Logistic regression analysis on the basis of tumor stage	98

5.17	Logistic regression analysis on the basis of Primary tumor extension	100
5.18	Logistic regression analysis on the basis of lymph node invasion	102
5.19	Logistic regression analysis on the basis of metastasis	104
5.20	Logistic regression analysis on the basis of ECOG	105
5.21	Logistic regression analysis on the basis of KPS	107
5.22	Logistic regression analysis on the basis of clinical response	109
5.23	Functional class of gene polymorphism of <i>MGMT</i> and <i>ERCC1</i> gene	110
5.24	Prediction of deleterious SNP by using SIFT and PROVEAN	111
5.25	Prediction of deleterious SNP by using Polyphen 2	112
5.26	Prediction of protein stability by I-mutant and MuPro	113
5.27	Prediction of deleterious effect of SNP by using PhD-SNP and SNP&GO	114
5.28	Prediction of changes in protein structure in SNP by Elaspic	115
5.29	Prediction of deleterious effect of SNP by Panther and MutPred	122
5.30	Interaction of <i>MGMT</i> gene with other genes	125
5.31	Interaction of <i>ERCC1</i> with other genes	127
5.32	Interaction of <i>MGMT</i> and <i>ERCC1</i> with each other and with other genes	129

LIST OF FIGURES

Figure No.	Title	Page No.
2.1	Cancer statistics	5
2.2	Risk factors of the cancer	6
2.3	Lung cancer statistics	8
2.4	Percentage of lung cancer patients on the basis of stage	8
2.5	Percentage of 5 year survival of lung cancer patients based on stage	9
2.6	DNA damage and Repair system	11
2.7	Removal of alkyl group from O ⁶ methyl guanine by <i>MGMT</i>	12
2.8	Structure of <i>MGMT</i>	13
2.9	Mechanism of action of <i>MGMT</i>	15
2.10	Structure of <i>ERCC1</i> -XPF complex	16
2.11	Mechanism of NER pathway	18
2.12	Chemical structure of cisplatin	20
2.13	Cellular response to platinum exposure	21
4.1	Restriction site of <i>mboII</i> enzyme	34
4.2	Input of SIFT software for <i>MGMT</i> and <i>ERCC1</i>	35
4.3	Input of PROVEAN software for <i>MGMT</i> and <i>ERCC1</i>	36
4.4	Input of Polyphen2 software for <i>MGMT</i> and <i>ERCC1</i>	36
4.5	Input of I-Mutant software for <i>MGMT</i> and <i>ERCC1</i>	37
4.6	Input of MuPro software for <i>MGMT</i> and <i>ERCC1</i>	38
4.7	Input of PhD-SNP software for <i>MGMT</i> and <i>ERCC1</i>	39
4.8	Input of SNP & GO software for <i>MGMT</i> and <i>ERCC1</i>	39
4.9	Input of Elaspic software for <i>MGMT</i> and <i>ERCC1</i>	40
4.10	Input of Panther software for <i>MGMT</i> and <i>ERCC1</i>	41
4.11	Input of MutPred software for <i>MGMT</i> and <i>ERCC1</i>	42
4.12	Input of Mutation 3D software for <i>MGMT</i> and <i>ERCC1</i>	43
4.13	Input of Genemania software for <i>MGMT</i> and <i>ERCC1</i>	43-44
5.1	Isolated genomic DNA	45
5.2	Amplified product of <i>MGMT</i> (rs12917)	46
5.3	Amplified product of <i>ERCC1</i> (rs3212986)	46

5.4	Restriction digested product of <i>MGMT</i> (rs12917)	47
5.5	Restriction digested product of <i>ERCC1</i> (rs3212986)	48
5.6	Frequencies of gender (a) and smoking status (b) in current study population to observe individual effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	51
5.7	Frequencies of SQCC, ADCC and SCLC in current study population to observe individual effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	51
5.8	Frequency of lung cancer stage in present study population to observe individual effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	52
5.9	Frequencies of tumor size (a), metastasis (b), lymph node involvement (c) and clinical response (d) in current study population to observe individual effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	53
5.10	Frequency of regimen in present study population to observe individual effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	54
5.11	Frequencies of KPS (a) and ECOG (b) in current study population to observe individual effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	54
5.12	Frequencies of gender and smoking status in current study population to observe combined effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	56
5.13	Frequency of histology in current study population to observe combined effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	57
5.14	Frequency of lung cancer stage in current study population to observe combined effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	57
5.15	Frequency of tumor size (a), metastasis (b), lymph node involvement (c) and clinical response (d) in current study population to observe combined effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	58
5.16	Frequency of regimen in current study population to observe	59

	combined effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	
5.17	Frequencies of KPS (a) and ECOG (b) in current study population to observe combined effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism	59
5.18	Effect of <i>MGMT</i> (rs12917) polymorphism on overall survival in lung cancer patients (CC vs. CT)	61
5.19	Effect of <i>ERCC1</i> (rs3212986) polymorphism on overall survival in lung cancer patients (CC vs. CA)	62
5.20	Effect of <i>ERCC1</i> (rs3212986) polymorphism on overall survival in lung cancer patients (CC vs. AA)	63
5.21	Effect of <i>ERCC1</i> (rs3212986) polymorphism on overall survival in lung cancer patients (CC vs. CA+AA)	63
5.22	Combinatorial effect of <i>MGMT</i> (rs12917) and <i>ERCC1</i> (rs3212986) polymorphism on overall survival in lung cancer patients (CC vs. CT+CA)	64
5.23	Effect of <i>MGMT</i> gene polymorphism on overall survival in SCLC patients (CC vs. CT)	66
5.24	Effect of <i>ERCC1</i> (rs3212986) gene polymorphism on overall survival in SCLC patients (CC vs. CA)	67
5.25	Effect of <i>ERCC1</i> (rs3212986) gene polymorphism on overall survival in SCLC patients (CC vs. AA)	67
5.26	Effect of <i>ERCC1</i> (rs3212986) gene polymorphism on overall survival in SCLC patients (CC vs. CA+AA)	68
5.27	Combine effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism on overall survival in SCLC patients (CC vs. CT+CA)	69
5.28	Effect of <i>MGMT</i> gene polymorphism on overall survival in ADCC patients (CC vs. CT)	71
5.29	Effect of <i>ERCC1</i> gene polymorphism on overall survival in ADCC patients (CC vs. CA)	71
5.30	Effect of <i>ERCC1</i> gene polymorphism on overall survival in ADCC patients (CC vs. AA)	72
5.31	Effect of <i>ERCC1</i> gene polymorphism on overall survival in ADCC patients (CC vs. CA+AA)	73

5.32	Combinatorial effect of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism on overall survival in ADCC patients (CC vs. CT+CA)	74
5.33	Effect of <i>MGMT</i> gene polymorphism on overall survival in SQCC patients (CC vs. CT)	76
5.34	Effect of <i>ERCC1</i> gene polymorphism on overall survival in SQCC patients (CC vs. CA)	76
5.35	Effect of <i>ERCC1</i> gene polymorphism on overall survival in SQCC patients (CC vs. AA)	77
5.36	Effect of <i>ERCC1</i> gene polymorphism on overall survival in SQCC patients (CC vs. CA+AA)	78
5.37	Combined effects of <i>MGMT</i> and <i>ERCC1</i> gene polymorphism on overall survival in SQCC patients (CC vs. CT+CA)	79
5.38	Mutation position for rs12917	116
5.39	Protein structure of wild and mutant type genotype of <i>MGMT</i> (rs12917)	116
5.40	Mutation position for rs2020893	117
5.41	Protein structure of wild and mutant type genotype of <i>MGMT</i> (rs2020893)	117
5.42	Mutation position for rs228164	117
5.43	Protein structure of wild and mutant type genotype of <i>MGMT</i> (rs228164)	118
5.44	Mutation position for rs2308318	118
5.45	Protein structure of wild and mutant type genotype of <i>MGMT</i> (rs2308318)	118
5.46	Mutation position of rs2308320	119
5.47	Protein structure of wild type and mutant type genotype of <i>MGMT</i> (rs2308320)	119
5.48	Mutation position of rs2308321	120
5.49	Protein structure of wild and mutant type genotype of <i>MGMT</i> (rs2308321)	120
5.50	Mutation position of rs3750824	120
5.51	Protein structure of wild and mutant genotype of <i>MGMT</i> (rs3750824)	121

5.52	Mutation position of rs2308322	121
5.53	Protein structure of wild and mutant type genotype of <i>MGMT</i> (rs2308322)	121
5.54	Mutation position of rs3212977	122
5.55	Protein structure of wild and mutant type genotype of <i>ERCC1</i> (rs3212977)	122
5.56	Position of mutations in protein structure of <i>MGMT</i>	124
5.57	Position of mutations in protein structure of <i>ERCC1</i>	125
5.58	<i>MGMT</i> gene interactions with other genes	127
5.59	<i>ERCC1</i> gene interaction with other genes	129
5.60	Gene interaction of <i>MGMT</i> and <i>ERCC1</i> with each other and with other genes	130

Chapter 1

Introduction

Cancer involves abnormal cell growth and has potential of metastasis (Croce, 2008). It is the second most common disease which leads to death (Hassanpour & Dehghani, 2017). According to international (GLOBOCAN) and national (NCPR-National Cancer Registry Programme authorized by Indian Council of Medical Research) research agencies on cancer, there will be chance of doubling of cancer cases in next 20 year (Mallath *et al.*, 2014). Among different cancer it was reported lung cancer is typical form of cancer in males whereas in females it is the second most common after breast cancer (Thun *et al.*, 2008).

According to American Cancer Society (American Cancer Society, 2017), in the US 234,030 new lung cancer cases were reported in 2017, among which 121,680 were males and rest were females. Approx 83,550 males and 112,350 females dead due to lung cancer (Siegal *et al.*, 2017). In India, among all the new cancer cases reported in 2012, percentage of lung cancer was 6.9% and death rate due to lung cancer was 9.3% (Malik & Raina, 2015).

The most common cause of lung cancer is tobacco, passive and active smoking, asbestos and radiations (Anand *et al.*, 2008). It is seen that long term tobacco-smoking contributes to about 85% of lung cancer cases and 15% of lung cancer cases are contributed by non-smokers (Thun *et al.*, 2008). Lung cancer is categorised into two types – small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Small cell lung cancer shows high malignancy and has short doubling time, hence growth rate is fast and it occurs within lung. NSCLC is the cancer of numerous types of epithelial lung cells and contributes 80-85% of the lung cancer cases and rest is contributed by SCLC (Thun *et al.*, 2008, Perez-Ramirez *et al.*, 2017, Tan *et al.*, 2017, Wang *et al.*, 2016, Olaussen *et al.*, 2006).

It has seen that information present in genomic DNA of normal cells makes proper balance between cell survival and homeostasis. Several factors like endogenous and exogenous (normal cellular metabolism by-products) are instigating for causing DNA damage. Damage in DNA includes single and double strand breaks (DSBs), mis-matches and addition of chemical compounds (Houtgraaf *et al.*, 2006). Whenever there is damage in genomic DNA, it is repaired by DNA repair machinery. Failure of DNA repair leads to accumulation of various kinds of mutation in the gene which interrupts the normal functioning of the gene and transforms this gene into oncogene resulting abnormal cell growth (Kamal *et al.*, 2015).

Hence genomic integrity is necessary for survival and homeostasis of the cell, which is maintained by various DNA repair pathways like Base Excision Repair (BER), Mis-Match Repair (MMR), Double-strand break repair and Nucleotide Excision Repair (NER) including both Non-Homologous End Joining (NHEJ) and Homologous Recombination (HR) (Altieri *et al.*, 2008).

Among several DNA repair pathway gene, it is reported that, O⁶-methylguanine–DNA methyl transferase (*MGMT*) and Excision Repair Cross Complementing Enzyme 1 (*ERCC1*) are involved in DNA damage reversal and NER pathway respectively (Pegg *et al.*, 2007, Felip & Rossel, 2007). Tumor type dictates the chemotherapy regimen. Cisplatin and etoposide are most commonly used in SCLC and in combination with gemcitabine, carboplatin, paclitaxel, topotecan, vinorelbine are also used in some cases. Chemotherapy is used as the first line treatment in NSCLC. Generally a platinum based drug like cisplatin or carboplatin is given. Other drugs like gemcitabine, paclitaxel and vinorelbine are also given in combination. For NSCLC combination of platinum-based drug and vinorelbine is more effective (Yu *et al.*, 2012, Le Chevalier, 2010). As discussed earlier single nucleotide polymorphism in a gene may leads to several kinds of mutation, which may further interrupt the functioning of the gene and structure of the protein. To know about whether a particular SNP disrupts the functioning of a gene and structure of a protein, various computational analysis are there, which can predict the possibility of disease due to mutation.

Computational analysis just predict the possibility of disease to some extent but cannot ensure the disease and its prognostic significance after treatment, so other analysis is required like genotyping, stastical analysis etc. to ensure the disease and to predict the prognostic implication of mutation in gene with chemotherapy in lung cancer patients. Although there were numerous study worldwide about the overall survival rate of lung cancer patients having genetic polymorphism in *MGMT* and *ERCC1* gene after chemotherapy, but there was no any data related to this for north Indian population. So, this work was designed to prognose the significance of gene polymorphism in *MGMT* and *ERCC1* gene among lung cancer patients of north India treated with platinum based doublet chemotherapy. The current study is based on computational analysis, genotyping as well as stastical analysis to find out whether the genetic polymorphism in *MGMT* and *ERCC1* gene contributes to disease or not, which variant is more significant to disease and overall survival rate association.

Chapter 2

Review of literature

2.1 Cancer

2.1.1 Introduction

Cancer is defined as uncontrolled, unwanted, purposeless and damaging growth of cells. Cancer is a Latin word which means crab. Uncontrolled growth of cells is called cancer because it resembles with the crab having claws (spreading into surroundings as like cancerous cells). “Switch on” and “Switch off” are two mechanisms which either promote or stop cell-division. Proto-oncogene and tumor suppressor genes control these two mechanisms. In the presence of growth signals, proto-oncogene promote cell growth and cell proliferation, on the other hand tumor suppressor gene monitor the unregulated cell growth and halt the cell division and proliferation. Cancerous cells are clonal in origin and has irregular, large, darker and sometime multi nucleus (Stephens *et al.*, 2009). Modifications in epigenetic may cause cancer and most common epigenetic modification are methylation of DNA (in cancer cell there is less methylation i.e 5-methyl cytosine), nucleosome position (chromatin modification) and histone modification (lack of mono-acetylated H4K16) (Hassanpur & Dehghani, 2017). Segment of DNA just give us information about the disease or risk of disease and genes are unable to determine the cause of disease. Cause of disease depends upon the interaction of proteins and cells with the external or internal factors of the cell. Internal as well as external factors may cause cancer. Internal factors include immune condition, hormones and genetic mutation, while external factors include radiations, chemicals and diet etc (Anand *et al.*, 2008).

2.1.2 Types of Cancer

Tumors are categorised into basically two types: benign and malignant. Benign tumor grows to a certain size and after that their growth becomes slow and after that stops, they do not show metastasis and can be removed by surgical methods. While malignant tumor is of two types: carcinomas and sarcomas. Tumors of epithelial cell are carcinomas (For e.g. – lung, prostate, colon and pancreas); tumors of connective tissue are sarcomas (For e.g. - fat, nerve, bone and cartilage) and carcinomas are the most common (Stephens *et al.*, 2009). On the basis of origin of tumor cell, cancer is also categorised into following types:

- Lymphoma or leukaemia – cancer of blood cells
- Germ-cell tumor- cancer of germ cells for e.g. -seminoma (testicle) and dys-germinoma (ovary)
- Blastoma- cancer of embryonic cell etc (Ebrahimi *et al.*, 2016).

Most frequent type of cancer among males is prostate cancer, followed by lung and bronchus, colon and rectum, urinary bladder and in females; it is breast cancer, then lung and bronchus, colon and rectum, uterine corpus and thyroid. In children, it is blood cancer, cancer of lymph node and brain (Hassanpour & Dehghani, 2017).

2.1.3 Cancer statistics

GLOBOCAN project is an international research agency on cancer which predicts that in India the cancer cases become doubles in next 20 year. National Cancer Registry Programme NCRP authorised by the Indian Council of Medical Research in 1981 which showed that in that year there was only 28 cancer cases.

Million death studies predict the burden of cancer nationwide as well as region wise. There are other sources which give the important information regarding cancer like publications of Atomic Energy Regulatory board, Medical Council of India and national board of examination websites. Registrar General and Census Commissioner of India provide information regarding cancer and population size, which cover 640930 villages, 7933 towns, 5924 sub-districts, 640 districts and 35 states (Ebrahimi *et al.*, 2016).

Second most common disease which leads to death worldwide is cancer. In US, there were 1,665,540 cases of cancer and 585,720 died in 2014 (Hassanpour & Dehghani, 2017). According to Indian Council of Medical Research (ICMR) cancer report 2016, there are 14.5 lakh cancer cases and it may increase up to 17.3 lakh in next four years. In males mouth cancer was the most common, while in females it was breast cancer. In Aizawl (Mizoram), there are most cancer cases in males and in Papumpare (Arunachal Pradesh), it is in females. Northeast has most cancer cases in males as well as females as shown in figure 2.1 (Weir *et al.*, 2015).

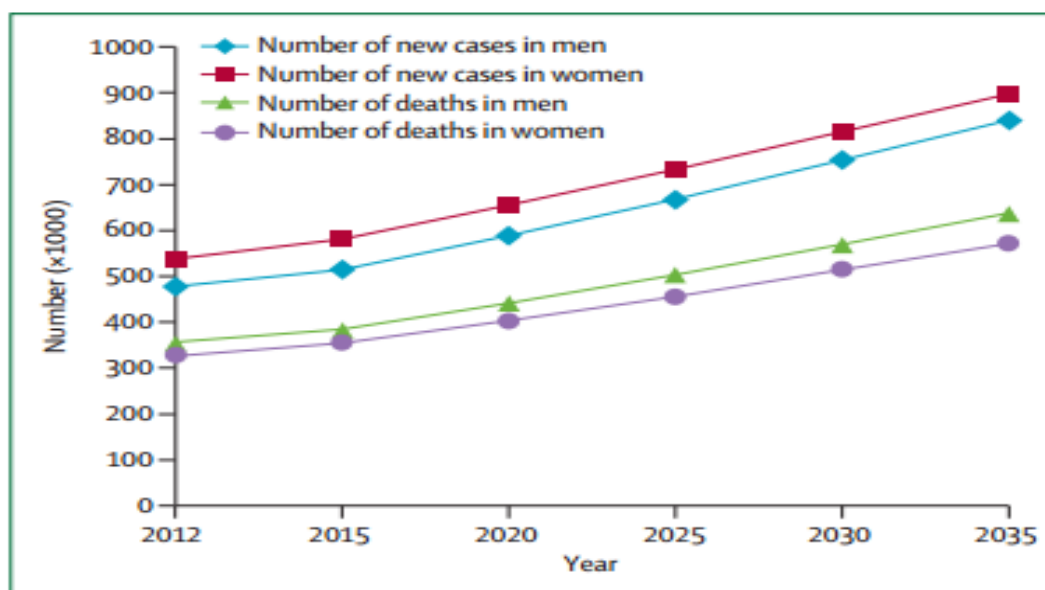


Figure 2.1 Cancer statistics
(Ebrahimi *et al.*, 2016)

2.1.4 Risk factors of cancer

Tobacco: According to US Surgeon Advisory commission report 1964 (<https://profiles.nlm.nih.gov/ps/retrieve/Narrative/NN/p-nid/60>), the primary cause of lung cancer is smoking. When smokers compared with the non-smokers, then it was found that a person who smokes is more prone to lung cancer. Approximately 50 carcinogens are there in tobacco. Among the 50 carcinogens, benzopyrenediol epoxide is directly associated with the lung cancer. Cell-signalling pathways can be altered by the tobacco and mechanism of alteration is still unknown (Anand *et al.*, 2008).

Alcohol: According to epidemiology of alcohol and cancer report 1910 (Lamu, 1910), alcohol is associated with esophageal cancer, breast cancer, liver, colorectal cancer. Mechanism of alcohol leads to carcinogenesis is unknown except mechanism of ethanol (co-carcinogen), acetaldehyde and free-radicals generates on metabolism of ethanol, free radicals binds to DNA and protein, destroys folate which leads to hyper-proliferation (Anand *et al.*, 2008). Alcohol also induces cytochrome P-4502E1, which enhances the production of free-radical, hence leads to hyper-proliferation (Anand *et al.*, 2008).

Diet: Nitrates, pesticides, nitrosamine and dioxins are the most common carcinogens in food (Anand *et al.*, 2008). Red meat can cause bladder, prostate, colorectal, breast, gastric, pancreatic, gastro-intestinal and oral cancers (Anand *et al.*, 2008). But meat consumption

reduces the risk of lung cancer. Carbon containing compounds such as amino acids, heterocyclic amines and pyrolysates are the major carcinogen in red meat (Anand *et al.*, 2008). Phenyl imidazo pyridine is considered to be most carcinogenic which is present in fried beef. Bis-phenol is a carcinogen which migrates in food from plastics. Fatty acids such as saturated and trans, and refined sugar may contribute to carcinogenesis (Anand *et al.*, 2008).

Obesity: It can cause cancer of gall bladder, gastric cardia, prostate, oesophagus, kidneys, pancreas, colon, breast and liver. Due to obesity, signalling pathway such as IGF/insulin/akt affected which leads to carcinogenesis (Anand *et al.*, 2008).

Infectious agents: Viruses such as human papilloma virus, Kaposi's sarcoma-associated herpes virus, Epstein Barr virus, HIV, HBV, HCV and T-lymphotropic virus 1 are associated with several kinds of cancer like liver, cervical cancer, skin cancer, Hodgkin's lymphoma, Kaposi's sarcoma, nasopharyngeal cancer, anogenital cancer, Burkett's lymphoma, adult T-cell leukaemia and B-cell lymphoma (Anand *et al.*, 2008).

Environ-mental pollution: It includes outdoor air pollution (polycyclic aromatic hydrocarbons) and indoor air pollution like tobacco smoke, food additives, and volatile compounds. These associates with the risk of childhood leukaemia and adult brain tumor (Anand *et al.*, 2008).

Radiations: Both ionizing and non-ionizing radiations contribute to lymphoma, thyroid cancers, skin cancers, breast cancer, leukaemia and lung cancer as shown in figure 2.2 (Anand *et al.*, 2008).

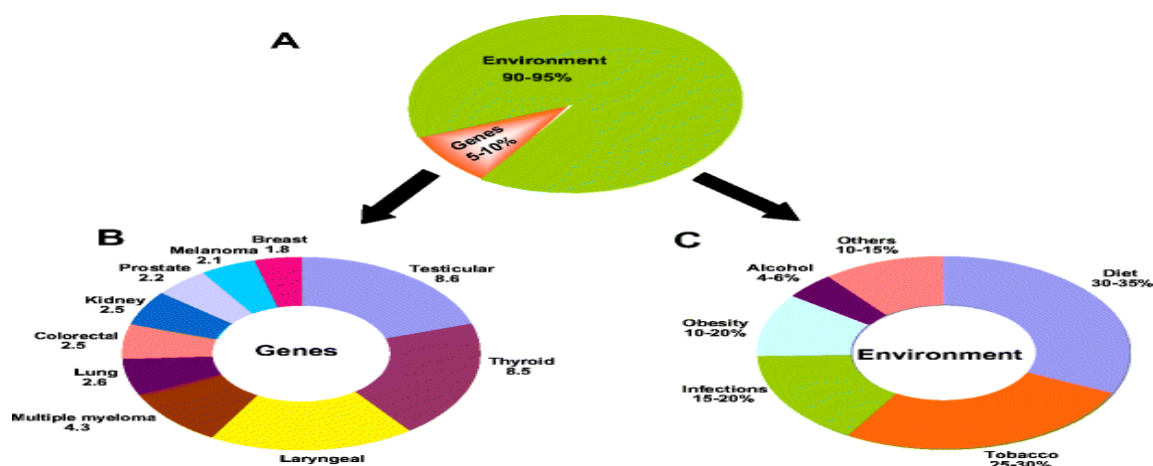


Figure 2.2: The risk factors of cancer
(Anand *et al.*, 2008)

2.2 Lung cancer

2.2.1 Introduction

Lung cancer is defined as abnormal cell growth in the lung tissue. Lung cancer found to be the most common cause of mortality worldwide (Thun *et al.*, 2008). Major symptoms of lung cancer include weight loss, chest discomfort, loss of appetite, coughing, shortness of breath and drastic change in voice (Thun *et al.*, 2008). Lung cancer diagnosis carried out by X-ray and CT and biopsy confirmed it. Lung cancer treatment includes surgery, chemotherapy as well as radiation therapy or sometimes both. Studies shows that 5-year overall survival rate is only 15% but in stage 4 it is less than 1% (Thun *et al.*, 2008).

2.2.2 Classification of lung cancer

Lung cancer is categorised into three types on the basis of arrangement of tumor cell under microscope: small cell lung cancer (SCLC), non-small cell lung cancer (NSCLC) and lung carcinoid tumor. SCLC contributes to only 15-20% of all the lung cancer cases, rest percentage contributed by NSCLC (Thun *et al.*, 2008, Perez-Ramirez *et al.*, 2017, Tan *et al.*, 2017, Wang *et al.*, 2016, Kalikaki *et al.*, 2009). Small cell lung cancer is more aggressive in nature; it is associated with heavy smoking and occur in primary and secondary bronchi of the lungs (Thun *et al.*, 2008).

NSCLC is the most frequent type of lung cancer. Based upon type of cell found in tumor, NSCLC is grouped into three types: adenocarcinoma, squamous cell carcinoma and large cell carcinoma. Among that types of NSCLC, adenocarcinoma is the most common, it is associated with the smokers as well as non-smokers (Thun *et al.*, 2008). It arises in the outer or peripheral areas of the lungs. Whereas squamous cell carcinoma occurs in the centre of chest area in bronchi and large cell tumor arises from the epithelial cells of the lung. Lung carcinoid tumor found to be least frequent and it derives from the central part of bronchi (Thun *et al.*, 2008).

2.2.3 Statistics of Lung cancer

According to Surveillance, Epidemiology and End Result Program (SEER) (www.seer.cancer.gov) of National Cancer Institute U.S, new lung cancer cases in 2018 will be 234,030 and death cases due to lung cancer will be 154,050. In 2015 there were 541,035

lung and bronchus cancer cases and in 2011-2015, 54.6 new cancer cases per 100,000 males and females every year and death due to lung and bronchus cancer was 43.4 per 100,000 men and women per year as shown in figure 2.3.

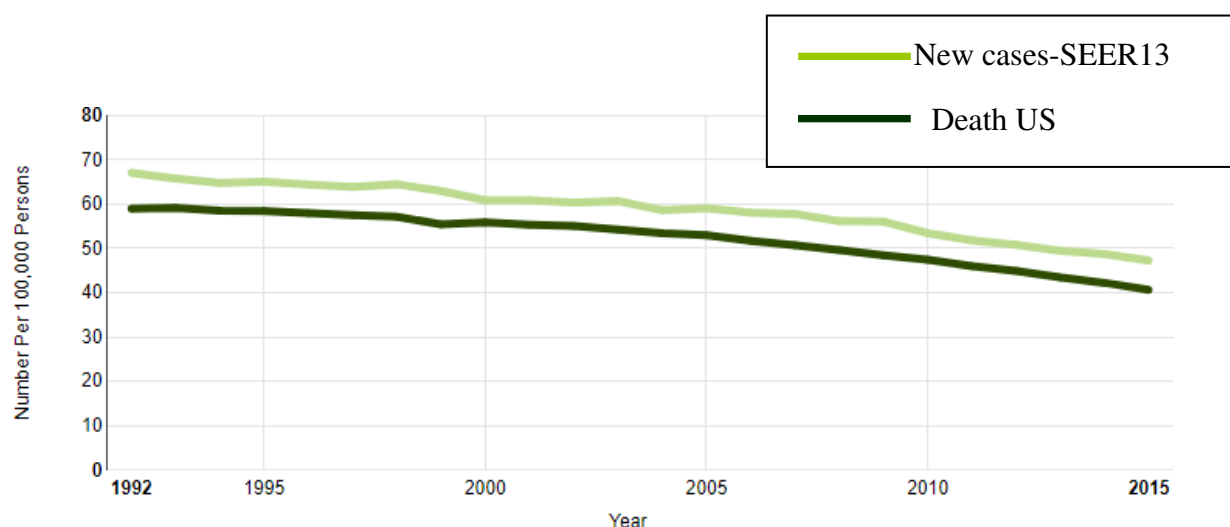


Figure 2.3 Lung and bronchus cancer statistics
(Source: <https://seer.cancer.gov/statfacts/html/lungb.html>)

From 2008-2014, percentage of 5 year relative survival was only 18.6 % as shown in figure 2.4. And on the basis of stage it is described in the figure 2.5, that percentage of 5 year survival is least in that cancer stage which is metastasized (www.seer.cancer.gov).

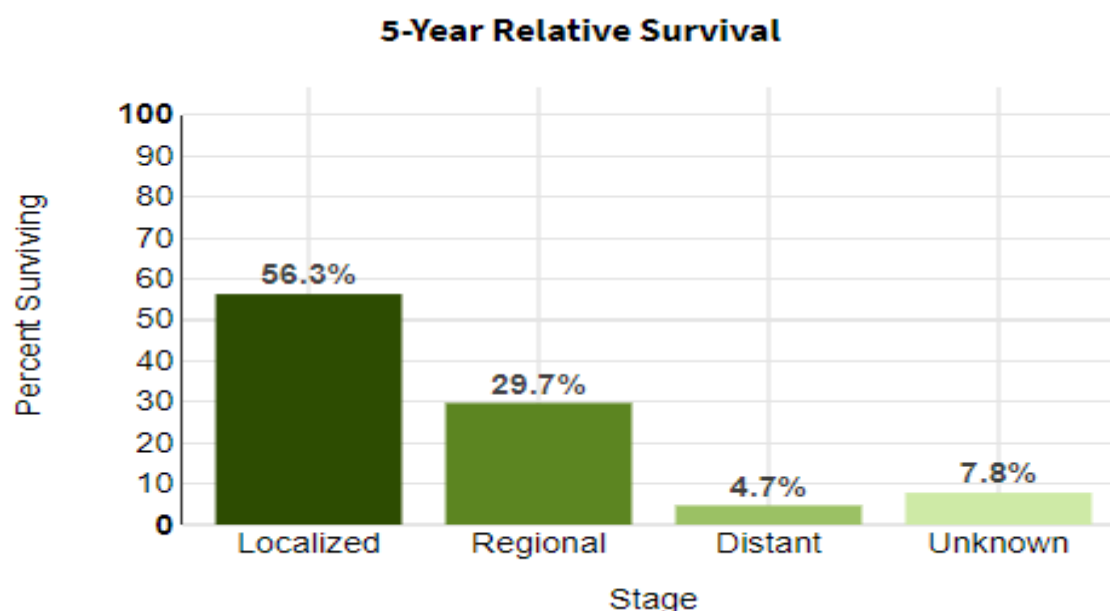


Figure 2.4: Percentage of 5 year relative survival of lung and bronchus cancer patients based on stage (2008-2014)

(Source: <https://seer.cancer.gov/statfacts/html/lungb.html>)

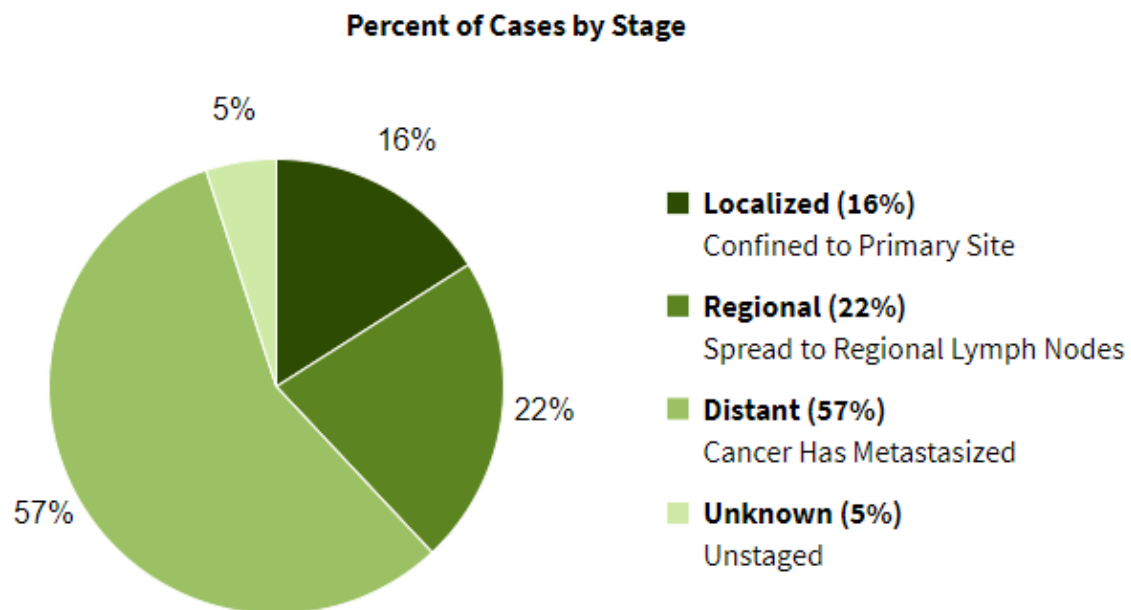


Figure 2.4: Percentage of lung and bronchus cancer patients on the basis of stage
 (Source: <https://seer.cancer.gov/statfacts/html/lungb.html>)

In India, among all the new cancer cases reported in 2012, percentage of lung cancer was 6.9% and death rate due to lung cancer was 9.3%. (Malik *et al.*, 2015).

2.2.4 Lung cancer stages

There are four stages of lung cancer. At stage one, the tumor size is 3 cm and there is no metastasis, at stage two, tumor size is less than 6 cm and metastasis begins, tumor size become more than 6 cm in stage three and metastasis reach in the lymph node, at stage four the tumor migrates to other organs of the body (Travis *et al.*, 2015).

2.2.5 Lung cancer risk factors

Most frequent risk factor of the lung cancer includes smoking (Cigarettes, Bidi, Hooka), second hand smoke, radon gas and asbestos etc. (Travis *et al.*, 2015). Occupational risk of lung cancer are Asbestos, Nickel Refinery workers, Haematite mining, Hard rock mining, Radiations, Chromium, Ethers and Mustard gas, Arsenic, chloromethyl, oils and coke exposures (Travis *et al.*, 2015).

According to Environmental Protection Agency USA and International Agency for Research on Cancer, carcinogen present in tobacco smoke, air pollution and food are polycyclic

aromatic hydrocarbons (PAHs), aromatic amines and tobacco specific nitrosamines (TSNA). (Molina *et al.*, 2013).

2.2.5.1 Lung cancer and smoking in India

From background history of patients having lung cancer, it was found that 87% males and 85% females were active tobacco smokers. Among tobacco related products, percentage of use of bidi, cigarettes, hooka and mixed were 28.4-79%, 9-53.7%, 3.4-77.3% and 7.5-13.6% respectively; and bidi is more carcinogenic in nature than others. Passive smoking that is environmental tobacco smoke is another common lung carcinogen (Anand *et al.*, 2008).

2.3 DNA damage and DNA repair System

DNA is damaged by the some internal as well as external factors (Sato & Itamochi, 2015, Liu & Jiang, 2017, Xu *et al.*, 2013). Any damage occurs in DNA may leads to mutagenesis if not repaired. These mutations trigger the uncontrollable growth of cell, which leads to oncogenesis and sometime apoptosis (Yin *et al.*, 2006). So to repair DNA damage inside the cell, there are various DNA repair systems inside the cell, which works on their own to maintain the integrity of the cell and hence prevents the cell to become cancerous cell. DNA repair mechanisms are DNA direct reversal pathway (no requirement of a template DNA), single strand damage repair (requirement of single strand of DNA) includes base excision repair (BER), mis-match repair (MMR), nucleotide excision repair (NER) and double strand damage repair (requirement of both strands and leads to gene rearrangements) includes double strand break repair (DSBs), non-homologous end joining repair (NHEJ) and micro-homology-mediated end joining (MMEJ) and homologous recombination repair; and interstrand DNA Cross-link repair as shown in figure 2.6 (Sato & Itamochi, 2015, Liu & Jiang, 2017).

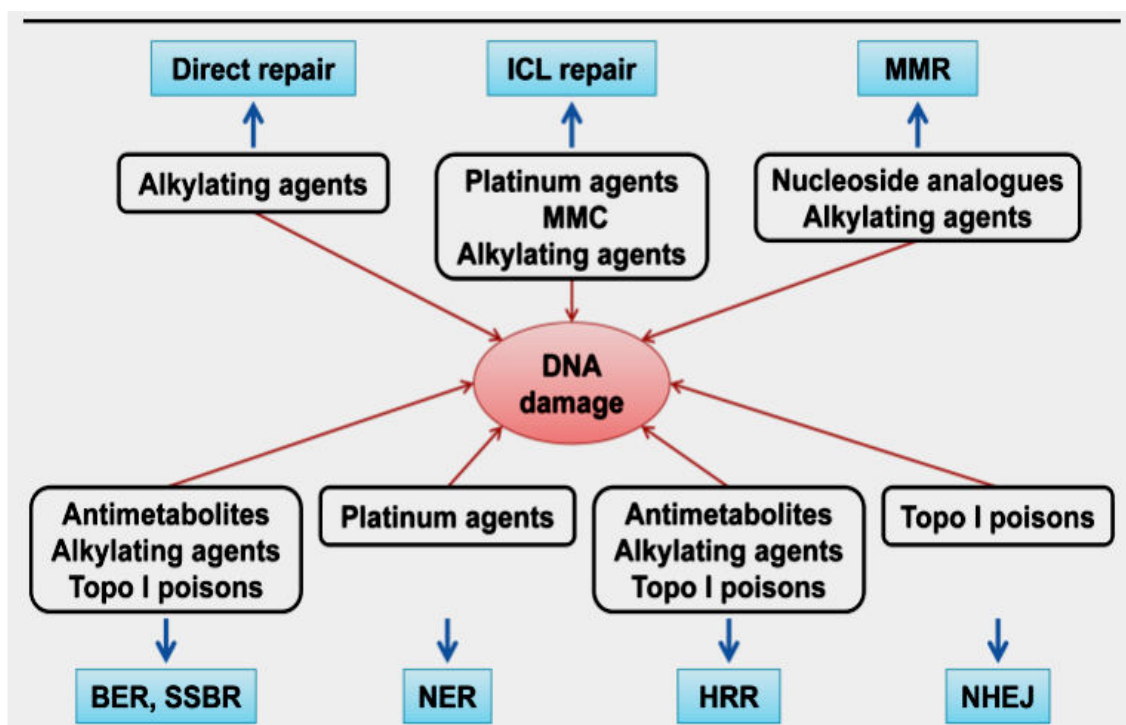


Figure 2.6 DNA damage and repair system

(DNA is damaged by external and internal factors and repaired by their respective DNA damage repair pathways: ICL-Interstrand DNA Cross-link repair, MMR- Mis Match Repair, BER- Base Excision Repair, SSBR- Single Strand Break Repair, NER- Nucleotide Excision Repair, HRR- Homologous Recombination Repair, NHEJ- Non-Homologous End Joining repair)

(Sato & Itamochi, 2015)

2.4 MGMT

2.4.1 Introduction

MGMT gene is also known as ATase, *AGT* and *AGAT* and it encodes MGMT enzyme. This enzyme involves in DNA direct reversal repair (DRR) pathway in humans (Shen *et al.*, 2005). In DNA direct reversal repair pathway, no need of a template because the damage occur at only one of the four nucleotide (A, C, G or T). This is specific to the damage and no breakage of phosphor-diester backbone occurs in this pathway (Liu & Jiang, 2017). Some endogenous and exogenous alkylating agents like cigarette smoke, contaminants from environment and components of diet generates an alkyl group at O⁶ position of guanine in DNA. This O⁶- alkyl guanine is a major carcinogen or premutagen in DNA. *MGMT* removes the alkyl group from O⁶ position in guanine and then this alkyl group transfers to cysteine145 residue of *MGMT*, after this *MGMT* becomes inactive as shown in figure 2.7 (Sato & Itamochi, 2015). *MGMT* shows resistance to chemotherapy among cancer patients

with alkylating agents and it also prevents the cell from mutagenesis and malignant transform (Shen *et al.*, 2005). During DNA replication, chromosomal aberration and sister chromatid exchange, A: T to G: C mutation may be caused by O⁶ methylguanine. *MGMT* gene expression level may vary among cell and tissues and it is less in tumor cells. It is reported that a higher incidence of tumorigenesis induced by nitrosamine is there in *MGMT* knockout mice. These data indicates the susceptibility of cancer due to *MGMT* gene (Li *et al.*, 2005).



Figure 2.7: Removal of alkyl group from O⁶ methyl guanine by *MGMT*
(Sato & Itamochi, 2015)

2.4.2 Protein structure

MGMT gene in humans is located on chromosome 10q26 and carry five exons of genomic DNA of approx 300 kb in size (Shen *et al.*, 2005, Li *et al.*, 2005). *MGMT* protein inclusive of 207 amino acids and has a bound zinc atom, which co-ordinates with four residues: Cys5, Cys24, His29 and His85 in N-terminal domain of the protein. High rate of methylated compound repair is shown by zinc form. When configuration of the protein changes, then

acceptor site Cys145 activates. In C-terminal domain, Cys145, and substrate binding site DNA binding site are located. C-terminal domain is inactive until it mixed with N-terminal domain. N-terminal domain is necessary for activation of C-terminal domain. Structure of *MGMT* protein shown in figure 2.8 (Han *et al.*, 2006).

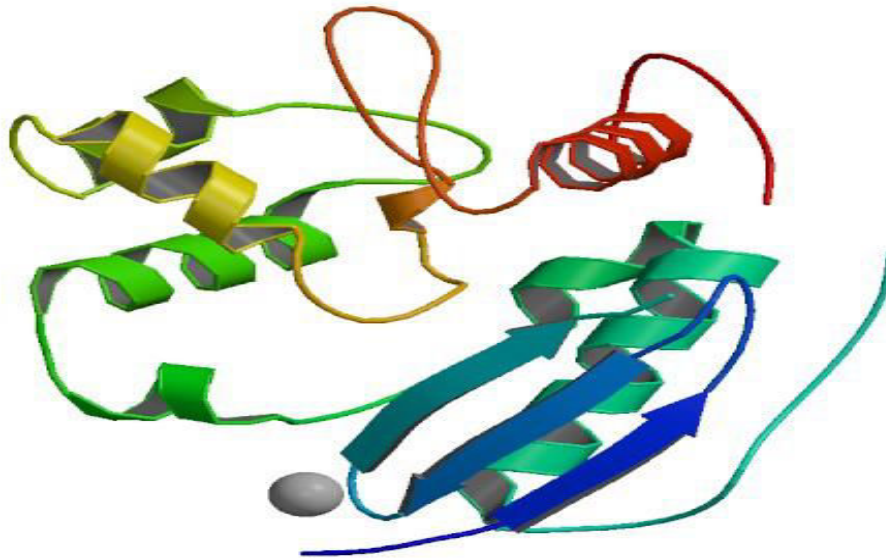


Figure 2.8: Structure of *MGMT* protein

(Source: http://www.quazoo.com/q/O-6-methylguanine-DNA_methyltransferase?alt=O-6-methylguanine-dna_methyltransferase)

2.4.3 Variants of *MGMT*

This gene has 6,846 reported SNPs and minor allelic frequency for rs1803965, rs2282164, rs12917, rs2308318, rs2308321, rs2308327 and rs1625649 are 0.1472, 0.0018, 0.1484, 0.0016, 0.0513, 0.0513 and 0.4032 respectively (Source: <https://www.ncbi.nlm.nih.gov/snp/>). There are approx 7 currently known SNP which affects primary structure of the *MGMT* protein as shown in table 2.1. Among these variants, most frequent are L84F, I143V and K178R. These three variants codes for non-conservative amino acid substitutions within the transcribed region. Very rare variants are G160R and W65C (Shen *et al.*, 2005, Jiao *et al.*, 2006). The polymorphism of Leu84Phe and Ile143Val associated with less potential of repairing of smoking related DNA damage.

Table 2.1: Variants of *MGMT*

Sr. No.	Nucleotide	SNP Id	Exon	Amino Acid Change	Allele frequency
1	C2740123T	rs1803965	3	Leu53Leu	17%
2	G2740159C	rs2282164	3	Trp65Cys (W65C)	1%
3	C2740214T	rs12917	3	Leu84Phe (L84F)	21%
4	G2799046A	rs2308318	5	Gly160Arg (G160R)	<1%
5	A2798995G	rs2308321	5	Ile143Val (I143V)	24%
6	A2799101G	rs2308327	7	Lys178Arg (K178R)	12%

(Shen *et al.*, 2005, Jiao *et al.*, 2006, Du *et al.*, 2013)**2.4.4 Leu84Phe (rs12917)**

Leu84Phe shows germ line polymorphism which causes alteration in repair activity of *MGMT*. Increased prevalence of glioblastoma, breast, prostate cancer, bladder and decreased proportion of endometrial, head and neck cancer linked to Leu84Phe polymorphism. For colon cancer, this polymorphism shows no effect (Fahmideh *et al.*, 2014). L84F is a change of one hydrophobic amino acid side chain which is conservative in nature and away from the active site in the N-domain of the *MGMT*. Leu84 is beside to His85, effect L84F substitution on zinc binding is unknown (Jiao *et al.*, 2006). This variant alters the structure of *MGMT* gene and this recombinant structure provides better zinc binding to *MGMT* and improves the DNA repair rate. Hence the mutant variant that is phe/phe has chance of being selected during evolution (Fahmideh *et al.*, 2014).

2.4.5 *MGMT* and Chemotherapy

Most commonly used chemotherapy in metastatic melanoma is dacarbazine. It acts as methylating cytostatic drug by producing O⁶ methyl guanine, which mis-pairs with the thymine during DNA replication and commence the cycle of repairing mis-match and DNA strand breaks generates, which leads to killing the tumour cells by inducing apoptotic signal transduction (Yin *et al.*, 2006). On the other hand *MGMT* removes the primary cytotoxic O⁶ methyl guanine by transferring the methyl group to an internal cysteine residue at codon 145 in the protein from O⁶ atom in guanine base and hence prevent cytotoxicity and leads to rebellious to the drug as shown in figure 2.9. *MGMT* shows resistance to certain alkylating antitumor agents like temozolomide and methylating drugs (dacarbazine) (Yin *et al.*, 2006).

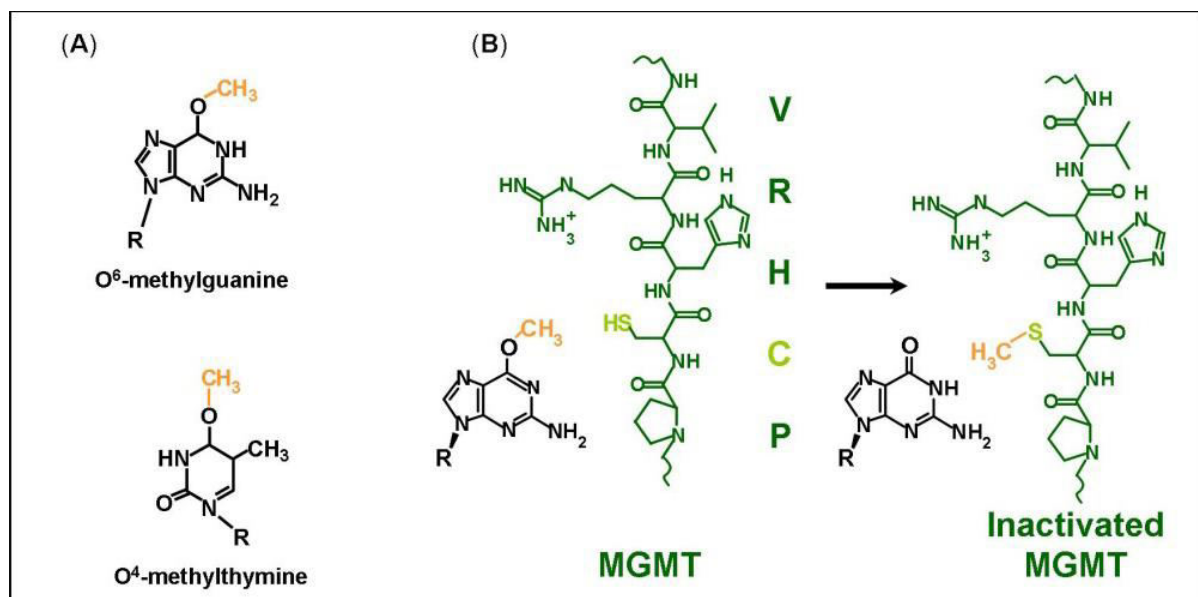


Figure 2.9: Mechanism of action of MGMT

(A: structure of DNA adduct, B: structure of MGMT, MGMT removes the methyl group from the DNA adduct and after removal undergoes inactivated)
(Nay and O'connor, 2013)

Meta-analysis studies show the role of *MGMT* gene polymorphism in lung cancer which remains uncertain and reason of uncertainty is that a single reported work or study may not be enough to find a small effect of the polymorphisms on lung cancer, particularly in relatively small sample sizes. So there is need of multiple studies about leu84phe polymorphism in lung cancer with large sample size (Qiu *et al.*, 2014).

2.4.6 Relation of *MGMT* with Cancer

As from literature no association was observed between *MGMT* (rs12917) and lung cancer patients of Sweden (Egyhazi *et al.*, 2002), South China (Liu *et al.*, 2002), Poland (Krzesiak *et al.*, 2004); patients with bladder cancer in South China (Li *et al.*, 2005); gastrointestinal cancer patients in Asian, Black and Caucasian population (Hu *et al.*, 2014); colorectal cancer patients in Iranian population (Khatami *et al.*, 2009); biliary tract cancer patients in Chinese population (Zhang *et al.*, 2008).

Weak association was found in patients with glioma (Liu & Jiang., 2017) and esophageal cancer (Du *et al.*, 2013) having gene polymorphism in *MGMT* (rs12917) in Caucasian population but not in Asian population; patients with glioma in han Chinese population (Dong *et al.*, 2014); patients with oesophageal squamous cell carcinomas in south African Population (Bye *et al.*, 2011); patients with glioma among white population but not in Asian population (Felini *et al.*, 2007).

2.5 *ERCC1*

2.5.1 Introduction

Excision Repair Cross Complementation group 1 was first gene (NER gene) to be cloned which has an important role in NER pathway (Yin *et al.*, 2006). In NSCLC patients who treated with platinum based chemotherapy, *ERCC1* could be a prognostic bio-marker (Ryu *et al.*, 2014). Function of a protein can be altered by the single nucleotide polymorphism in a gene. When DNA repair machinery gets affected and is unable to repair the DNA Damage, then there is risk of acquiring cancer due to accumulation of DNA damage. On the other hand any defect in DNA damage machinery is related with the high respond of platinum based chemotherapy as it generate an adduct in DNA which disrupt the DNA replication and hence it leads to tumor cell death (Felip & Rossel, 2007).

2.5.2 Protein structure

ERCC1 gene is located on human chromosome 19q13, is 15 kb base pair and codes for 297 amino acids (Felip & Rossel, 2007, Xu *et al.*, 2013). It has 10 exons and 1.1 kb long coding region (Yin *et al.*, 2006). It joins with Xeroderma Pigmentosum complementation group F and forms the *ERCC1*-XPF complex as shown in figure 2.10. This complex cut the damaged portion of 5' DNA strand which is caused by cis-platin chemotherapy (Felip & Rossel, 2007).

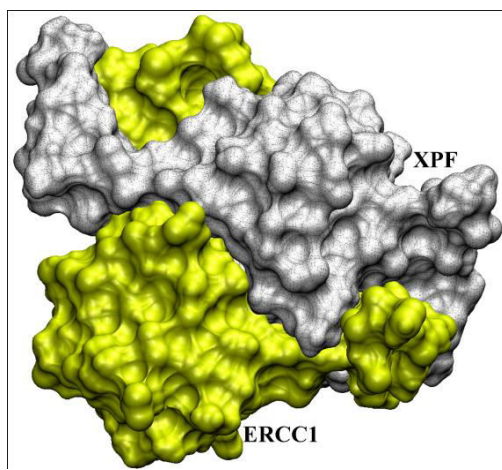


Figure 2.10: Structure of *ERCC1*-XPF complex

(Source: <https://www.intechopen.com/source/html/44543/media/image3.png>)

2.5.3 NER pathway

Nucleotide excision repair pathway repairs the non-specific DNA damage which are caused by chemical inter- or intra- strand additional compound formation, gamma or ultra violet radiation and cross-linking (Felip & Rossel, 2007). NER removes the structurally unrelated DNA damage. Cyclobutane pyrimidine dimer and photoproducts contributes two most common types of damage in DNA which is corrected with the help of NER pathway. There are six genes which encodes for the proteins that has an important role in NER pathway and these genes are *ERCC1*, *ERCC2*, *ERCC3*, *ERCC4* and *ERCC5* (these genes makes complex with Xeroderma Pigmentosum complementation group F, D, B, F, G and A) (Sullivan *et al.*, 2014).

2.5.3.1 Role of *ERCC1* in NER pathway

DNA damage binding factor i.e XPC/hHR23B initiates the NER pathway. After initiation of this pathway TFIIH, XPA and RPA recruited to the site of damage. They open the DNA helix, orient the NER machinery and stabilize the open DNA intermediate. Then single strand DNA binding protein binds on non-damage DNA so that structure specific endonuclease that is XPG and *ERCC1*/XPF orients them on damage DNA. XPG is 3' excision, while *ERCC1*/XPF is 5' incision. These endonuclease remove the damaged part of the DNA and then DNA polymerase δ/ϵ with auxillary factors fill the gap as shown in figure 2.11 (Li *et al.*, 2005, Zhu *et al.*, 2014).

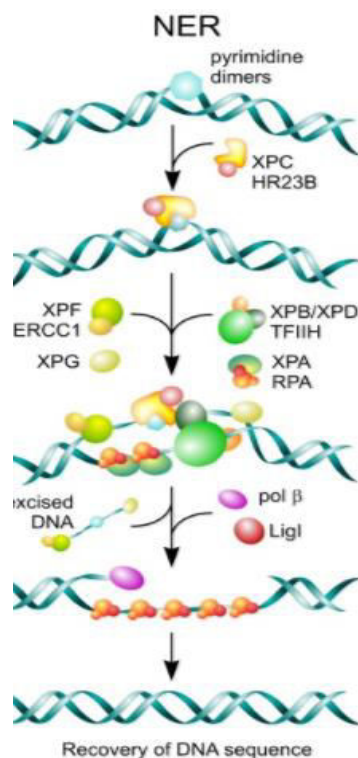


Figure 2.11: Mechanism of NER Pathway

(XPC-Xeroderma Pigmentosum Complementation group C, HR23B- RAD 23 homolog B, TFIIH-Transcription factor II Human, XPA- Xeroderma Pigmentosum, Complementation group A, XPD- Xeroderma Pigmentosum Complementation group D, XPB- Xeroderma Pigmentosum Complementation group B), XPF-Xeroderma Pigmentosum Complementation group F, XPG- Xeroderma Pigmentosum Complementation group G, RPA- Replication protein A, pol β- DNA polymerase beta, Lig I- DNA ligase 1 enzyme. XPC and HR23B binds to DNA damage adduct pyrimidine dimer and initiates NER pathway. Then XPF-ERCC1, XPG, XPB, XPD, TFIIH, XPA, RPA recruit to the site of damage. XPG and ERCC1-XPF cut the DNA damage portion and DNA polymerase and ligase joins the gap)

(Mladenov & Iliakis, 2011)

2.5.4 *ERCC1* variants

Variants of *ERCC1* are rs11615, rs2298881, rs3212948, rs3212961, and rs3212986 (Hui *et al.*, 2014, He *et al.*, 2013, Yu *et al.*, 2015, Zhu *et al.*, 2014). Most common among these are rs11615 and rs3212986 (Hui *et al.*, 2014).

2.5.5 Relation of overall survival with *ERCC1* rs3212986 in cancer patients

No relationship was observed between response/overall survival and *ERCC1* (rs3212986) in NSCLC patients who were treated with platinum based chemotherapy (Sullivan *et al.*, 2014, Liu & Jiang, 2017); in NSCLC patients administered with platinum based chemotherapy in Caucasian and Asian population (Xu *et al.*, 2013, Perez-Ramirez *et al.*, 2017, Yin *et al.*, 2006); in Chinese population (Chen *et al.*, 2012, Shi *et al.*, 2015).

A significant relationship was observed between *ERCC1* rs3212986 and lung cancer patients in response to combination of cisplatin and docetaxel with p value-0.012 (Azuma *et al.*, 2009), p-0.024 (Olaussen *et al.*, 2006), p-0.004 (Takenaka *et al.*, 2010); also in response to patients dispensed with platinum based chemotherapy among Caucasian and Asian population (Tan *et al.*, 2017).

Poor relationship was observed between overall survival and gene polymorphism in *ERCC1* (rs3212986) among NSCLC patients who were treated with platinum based chemotherapy in China (Zhao *et al.*, 2014, Gao *et al.*, 2014, Huang *et al.*, 2014).

No correlation was observed in respond to platinum based doublet chemotherapy and *ERCC1* rs3212986 among osteosarcoma patients in China (Cao *et al.*, 2015, Biason *et al.*, 2012). No correlation was observed between overall survival and gastric cancer patients having *ERCC1* rs3212986 genotype (administered with platinum based chemotherapy) (Yu *et al.*, 2015); also true for Chinese population (Li *et al.*, 2014). Also no correlation was observed between *ERCC1* rs3212986 and overall survival in epithelial ovarian cancer patients dispensed with platinum based chemotherapy (Steffensen *et al.*, 2011). *ERCC1* rs3212986 was not significantly related with therapeutic respond to platinum based doublet chemotherapy in patients with esophageal cancer (Rumiato *et al.*, 2013)

A significant relationship was found between overall survival and *ERCC1* rs3212986 among patients suffering from gastric cancer and administered with chemotherapy in Chinese population (Xue *et al.*, 2015, Mo *et al.*, 2015).

2.6. Platinum based chemotherapy

Platinum is most commonly used to kill tumor cells and it is the first line treatment for NSCLC (Olaussen *et al.*, 2006). It shows cytotoxic activity by forming mono or bifunctional compounds in DNA and it causes inter- or intra- strand cross-linking. These cross links are repaired by the DNA damage repair machinery. NER, BER and DBS pathway is the part of DNA damage repair machinery and all of them are regulated by p53 which can trigger apoptosis or repair of damage in DNA as shown in figure 2.12 (Ryu *et al.*, 2014, Perez-Ramirez *et al.*, 2017).

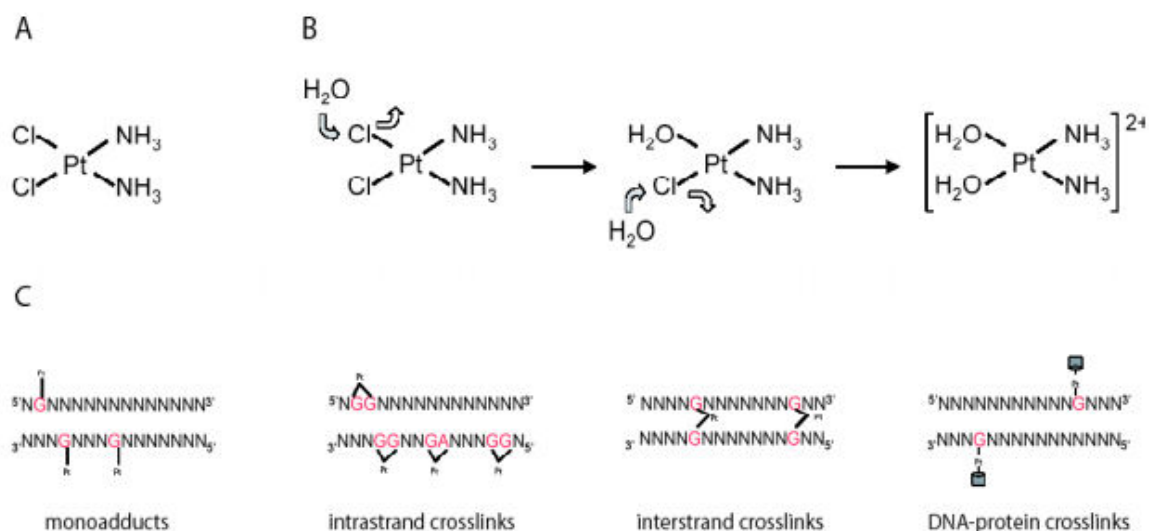


Figure 2.12: A. Chemical structure of cisplatin, B. Changes occur in cisplatin inside the cell and C. Cisplatin induced DNA damage.

(Koberle & Staffa, 2012)

Platinum based chemotherapy in combination with a third generation cytotoxic drug (docetaxel, gemcitabine, paclitaxel and vinorelbine) shows improvement in survival of metastatic patients. To treat NSCLC; cisplatin based chemotherapy regimens is consider as most effective treatment. The mechanism of resistance against platinum based chemotherapy is decreased accumulation inside the cell, increased resilience to DNA damage because of highly potent DNA repair capacity and increased detoxification by conjugation with glutathione as shown in figure 2.13.

There is a hypothesis that when NER does not able to repair the damaged DNA then tumor cells treated with cisplatin induce the tumor cell death by forming inter- or intra-strand DNA additional compounds (Felip & Rossel 2007). It is reported that *ERCC1* gene is related with the tolerance of platinum based drug (Azuma *et al.*, 2009).

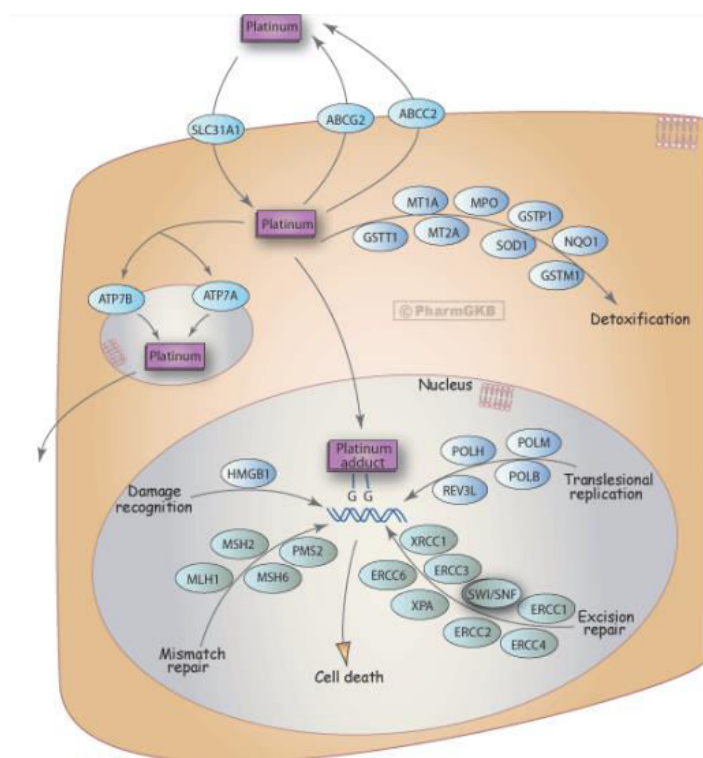


Figure 2.13: Cellular response to platinum exposure

(There are some cellular genes which detoxify or pump out or repair the damage when platinum compound enters inside the cell)

(Van der *et al*, 2011)

Another name of cisplatin is cis-diamminedichloroplatinum (II). Its geometry is square planar and it is a metallic compound. At room temp. it is a crystalline powder of white or deep yellow or yellowish orange color. It is a non-polar compound which is soluble in N,N dimethylformamide and dimethylprimanide and stable at room temp. But with time it can change into trans-form. Molecular weight, density, melting point, log K and water solubility of cisplatin is 301.1 gm/mol, 3.74 g/cm³, 270° C, -2.19 and 2.53 g/L at 25° C respectively. M. Peyrone first synthesized cisplatin in 1844 and its chemical structure was first described by Alfred Werner in 1893. In 1960, it was found by Rosenberg that platinum electrolysis product inhibits the cell division in *E.coli*. This observation is the basis of use of platinum product for treatment of cancers (Van der *et al*, 2011, Ingels *et al.*, 2017).

2.6.1 Cisplatin & lung cancer

Lung cancer is the most frequent cancer now a day's worldwide. The two most common platinum based chemotherapy for NSCLC is cisplatin and carboplatin. Cisplatin is most commonly used treatment as compared to carboplatin because it has an excellent anti-tumor activity. Although it has many side-effects like vomiting, renal toxicity and nausea etc. at stage II and III of non-small cell lung cancer, adjuvant cisplatin-based chemotherapy is more preferable (Ingels *et al.*, 2017). Percentage of patients suffering from lung cancer and who responds to platinum based doublet chemotherapy is less than 30% and 5 year survival rate is only 10%. Clinical stage is the most influential prognostic factor among lung cancer patients (Sullivan *et al.*, 2014, Xu *et al.*, 2014, Zhao *et al.*, 2014).

2.7 Gaps in study

A lot of studies were carried out to predict the significant relationship of *MGMT* (rs12917) gene polymorphism and the risk of lung cancer or other cancer worldwide, but few study was carried out to find overall survival rate among patients who were administered with platinum based doublet chemotherapy. For analysis of chemotherapeutic response among lung cancer patients having *MGMT* (rs12917) gene polymorphism, the North Indian Population was taken as subject in this dissertation for the first time.

Also, a lot of studies were carried out to find overall survival rate among patients with *ERCC1* rs3212986 gene polymorphism treated with platinum based chemotherapy in different population, but nobody studied it for North Indian Population.

Moreover, there was no data available for combined effect of *MGMT* and *ERCC1* gene on therapeutic response. Many parameters were also included in this study like metastasis, TNM staging, lymph node, tumor size and regimen. For the first time insilico analysis was included along with genotyping and statistical analysis to predict the effect of SNP on function and structure of protein.

2.8 Computational analysis of deleterious single nucleotide polymorphism of *MGMT* and *ERCC1* gene

Single nucleotide polymorphism is describe as change in base of the DNA whether adenine, guanine, cytosine or thymine. SNP may lie within coding region, non-coding region or in between the coding and non-coding region. Non-synonymous SNP changes the phenotypic effect of the genes. So to check or predict whether the SNP lies within coding region, non-coding region or among coding and non-coding region, whether SNP causes any phenotypic change in the gene or whether the SNP contributes to benign or malignant tumor, various bioinformatics tools are there. These are SIFT, PROVEAN, I-MUTANT, POLYPHEN2, MuPro, Phd-SNP, SNP & GO, ELASPIC, PANTHER, MutPred, Mutation 3D and Genemania.

SIFT- It stands for **S**eparating **I**ntolerant from the **T**olerant. It is homology-based software which predicts the phenotypic effect of the genes. Its input is a query sequence and it uses that query sequence and on the basis of multiple alignments, it gives a score. This score is the normalised probability of all the substitution, which predicts whether the SNP has deleterious effect or not. If the score is equals to or less than 0.05 then the SNP is consider as deleterious and if score is more than 0.05 then the SNP is tolerated. Prediction of SIFT is based upon conservation of amino acid residues in closely related sequences after performing PSI-BLAST. This can be used for laboratory generated mutations or non-synonymous mutations (Ng and Henikoff, 2003).

PROVEAN- It stands for protein variation effect analyser. Basis of this software is clustering method, in which final score is generated by the average cluster of 30 identical sequences (75% identity). Score predicts whether SNP has functional effect or not (Choi and Chan, 2015).

POLYPHEN2- It stands for polymorphism phenotyping v2. This bioinformatics tool uses comparative and structural evolutionary relationships and predicts whether stability and function of the protein affected or not. It is based on the protein sequence alignment, 3-D structure of protein and several machine learning programmes. Multiple sequence alignment is based on position specific independent counts and calculates the profile matrix (score), which is the likelihood log ratio of one amino acid at a particular position to another amino acid occurring at same position or vice versa (Adzhubei *et al.*, 2013).

I-MUTANT- It is hinged on the support vector machine and predicts the stability of protein when SNP occur. Reliability index (measure of consistency of score) should be $0 < RI < 10$ and calculated from the output of support vector machine (O). Formula to calculate $RI = 20 * \text{abs}(O - 0.5)$ and to calculate Gibbs free energy change (DDG) = Gibbs free energy of mutated amino acid minus Gibbs free energy of wild type amino acids. If the value of DDG is lower than zero then the protein stability decreases and if it is more than zero then protein stability increases (Capriotti *et al.*, 2005).

MuPro-It is based on support vector machine as well as neural network. It is more accurate (84%) as it cross validate results by 20 folds. This predicts the stability of proteins and it does not require tertiary structures to predict stability of proteins. If the value of DDG is lower than zero then the protein stability decreases and if it is more than zero then protein stability increases (Cheng *et al.*, 2006).

PhD-SNP-This tool is also based upon support vector machine. It predicts whether the mutation is disease related or not. Reliability index (measure of consistency of score) should be $0 < RI < 10$ and calculated from the output of support vector machine (O). Formula to calculate $RI = 20 * \text{abs}$ (Capriotti *et al.*, 2017).

SNP & GO- This software is also based upon the support vector machine and also predicts whether the SNP contribute to disease or not. Its accuracy is 82% (Calabrese *et al.*, 2009).

ELASPIC- It is based on the ensemble machine learning programme. It predicts whether protein-protein interaction and protein folding gets affected by the change in nucleotide. This predicts with the help of homology model of domains and domain-domain interactions (Witvliet *et al.*, 2016).

PANTHER-It is based on the likelihood method to predict whether functionality of a protein gets affected by SNP or not. This method aligns the evolutionary related proteins and then calculates the score called substitution position specific evolutionary conservation score (Thomus *et al.*, 2003).

MutPred-It predicts whether structural and features of the protein changes by SNP or not and also predicts the pathogenicity of the SNP. A score less than 0.5 is considered as non-pathogenic while score equals to or more than 0.5 is considered as pathogenic (Pejaver *et al.*, 2017).

Mutation 3D- It is helpful in the visualization and functionality prediction of protein when there is any change in the nucleotide. Visualisation involves the arrangement of amino acid substitution spatially on protein model (Meyer *et al.*, 2016).

Genemania-This predicts the set of data which includes co-localization, protein interaction, genetic interaction, co-expression, pathways and similarity in protein domain etc. related to one gene or set of genes and also predicts genes which are related to input gene (Warde-Farley *et al.*, 2010).

URL for all the software is given in table 2.2 (Abdelraheem *et al.*, 2016, Manickam *et al.*, 2014)

Table 2.2 Bioinformatics tool with URL			
Sr no.	Bioinformatics tool	URL	References
1	SIFT	http://sift.jcvi.org	(Ng and Henikoff, 2003)
2	PROVEAN	http://provean.jcvi.org	(Choi and Chan, 2015)
3	POLYPHEN2	http://genetics.bwh.harvard.edu/pph2/	(Adzhubei <i>et al.</i> , 2013)
4	I-MUTANT	http://folding.biofold.org/cgi-bin/i-mutant2.0.cgi	(Capriotti <i>et al.</i> , 2005)
5	MuPro	http://www.ics.uci.edu/~baldig/mutation.html	(Cheng <i>et al.</i> , 2006)
6	PhD-SNP	http://snps.biofold.org/phd-snp/phd-snp.html	(Capriotti <i>et al.</i> , 2017)
7	SNP & GO	https://snps-and-go.biocomp.unibo.it/snps-and-go/	(Calabrese <i>et al.</i> , 2009)
8	ELASPIC	http://elaspic.kimlab.org/many/	(Witvliet <i>et al.</i> , 2016)
9	PANTHER	http://pantherdb.org/tools/csnpscoreForm.jsp	(Thomus <i>et al.</i> , 2003)
10	MutPred	http://mutdb.org/mutpred	(Pejaver <i>et al.</i> , 2017)
11	Mutation 3D	http://mutation3d.org/index.shtml	(Meyer <i>et al.</i> , 2016)
12	Genemania	http://www.genemania.org/	(Warde-Farley <i>et al.</i> , 2010)

Chapter 3

Objectives

- Computational analysis of deleterious single nucleotide polymorphisms (SNPs) in human *MGMT* and *ERCC1* gene involved in cancer using multiple softwares.
- To predict the clinical outcomes of advanced lung cancer Patients having polymorphism in *ERCC1* and *MGMT* gene treated with platinum based doublet chemotherapy.

Chapter 4

Material and Methods

4.1 Sample Collection

The current study enrolled a total of 214 lung cancer patients who were enlisted from the Department of Pulmonary Medicine, Post Graduate Institute of Medical Education and Research (PGIMER) Chandigarh, India. This study has been revised and approved by the ethics committee of PGIMER, Chandigarh as well as Thapar Institute of Engineering and Technology, Patiala. Informed written agreement was obtained from all enrolled patients or their representatives. All the enlisted patients were histopathologically diagnosed with NSCLC or SCLC. Patients under observation having a previous history of cancer were excluded from the study. There was no age, gender, smoking, histological or TNM stage restrictions. Around, 5 ml of blood was collected in vacutainers from each individual enrolled in present study. Each participant filled up the detailed questionnaire with the help of trained interviewer. The questionnaire compromised information on demographic and smoking characteristics like tobacco habits such as smoking of beedi/cigarette etc. Individuals who smoked regularly were classified as smokers. They were further classified as light and heavy smokers on the basis of pack years (PY) that were calculated by the formula:

$$[(\text{cigarettes or beedis per day}/20)*\text{years smoked}]$$

Pack years (PY) less than 25 were considered as light smokers and pack years (PY) greater than or equals to 25 were considered as heavy smokers. The medical information of cases such as histology, clinical staging, primary tumor size, involvement of lymph node and metastasis, TNM classification were obtained from medical records of the hospital.

4.2 In-vitro analysis

4.2.1 Genomic DNA isolation

Genomic DNA was isolated using standard Protein K digestion, phenol/chloroform extraction and ethanol precipitation method from whole blood samples of Lung cancer cases (Maniatis *et al.*, 1982).

Requirements

- Blood sample
- Washing buffer
- Lysis buffer
- Phenol:Chloroform:Isoamylalcohol (25:24:1)
- Chloroform: Isoamylalcohol (24:1)
- Absolute ethanol
- Ethanol
- Tris EDTA Buffer

Preparation of Buffers

Table 4.1 Preparation of washing buffer			
Sr no.	Chemical reagent	Stock concentration	Working concentration
1	Sucrose	1 M	320 mM
2	Triton X 100	100%	1%
3	MgCl ₂	100 mM	5 mM
4	Tris-Cl	100 mM	10 mM

Table 4.2 Preparation of lysis buffer			
Sr no.	Chemical reagent	Stock concentration	Working concentration
1	Tris-Cl	1 M	400 mM
2	SDS	10%	1%
3	EDTA	0.5 M	60 mM
4	NaCl	5 M	150 mM
5	Proteinase K	10 mg/ml	100 µg/ml

Table 4.3 Preparation of Tris EDTA buffer			
Sr no.	Chemical reagent	Stock concentration	Working concentration
1	Tris-Cl (pH-8)	1 M	10 mM
2	EDTA (pH-8)	0.5 M	1 mM

Procedure of genomic DNA Isolation

- 1 ml of blood was taken and same amount of washing buffer was added to it, and then centrifuged at 3500 rpm for 5 minutes at 4°C.
- After centrifugation, supernatant was discarded and 1 ml washing buffer was added to it and again centrifuged (this step was repeated thrice).
- Then, 1 ml Lysis buffer was added to dissolve the pellet and was incubated at 55°C overnight.
- After incubation, equal volume (1 ml) of Phenol: Chloroform: Isoamylalcohol (25:24:1) was added to it and centrifuged at 8,000 rpm for 10 min (this step was repeated twice).
- Supernatant was taken and equal volume of Chloroform: Isoamylalcohol (24:1) was added to it, after that it was centrifuged at 6500 rpm for 5 min.
- Then aqueous layer was taken and 2.5 times volume of absolute ethanol was added to it, after that it was freeze at -20°C for 1-2 hrs.
- After that it was centrifuged at 12,000 rpm for 10 min. and supernatant was decanted, then 1 ml of ethanol was added to it and centrifuged at 12,000 rpm for 5 min. (this step was repeated twice).
- After washing of pellet, pellet was dried and dissolved in 50 µl TE Buffer (Maniatis *et al.*, 1982).

Quantitative estimation of DNA sample

This was done with the help of UV spectrophotometer by recording absorbance at 260 nm and 280 nm. Ratio of absorbance at 260 nm and 280 nm was used to find out the purity of DNA.

$$\text{Purity of DNA} = \text{Absorbance at 260 nm} / \text{Absorbance at 280 nm}$$

(Ratio of 1.8 indicates pure DNA, 2 indicated RNA and less than 1.8 indicates impurity of proteins and other contaminants)

$$\text{DNA concentration } (\mu\text{g/ml}) = \text{Absorbance at 260 nm} \times 50 \times \text{Dilution Factor}$$

(Where 1 O.D is equals to 50 µg/ml of DNA) (Maniatis *et al.*, 1982)

Qualitative analysis of DNA sample

Requirements

- Horizontal gel electrophoresis apparatus
- Gel casting tray
- Gel combs
- Electrophoresis – grade agarose
- Ethidium bromide solution (10 mg/ml-stock, 10µg/ml-working)
- Electrophoresis buffer (TBE or TAE)
- 6 X loading dye

Preparation of reagents

Table 4.4 Preparation of TBE buffer and 6X loading dye				
	Preparation of TBE buffer		Preparation of 6X loading dye	
Sr no.	Reagent	Amount (for 1000 ml)	Reagent	Amount (for 20 ml)
1	Tris-base	54 g	0.25% Bromophenol blue	0.05 g
2	Boric acid	27.5 g	0.25% xylene cyanol	0.05 g
3	EDTA (0.5 M)	20 ml	40% Sucrose	8 g

Procedure for agarose gel electrophoresis

- 25 ml of 0.8% agarose was prepared with TBE buffer and before pouring, adequate (1-2 µl) amount of ethidium bromide was added to it to visualise the DNA under UV illuminator.
- Gel comb was inserted on the gel casting tray and then agarose was poured, making sure that no bubble formation was takes place and if it so then bubble was removed.
- After solidification of agarose, comb was removed from the tray carefully and then, this was placed in the electrophoretic tank.
- TBE buffer was added to the tank so that the agarose gel sinks and wells were submerged.
- After that DNA sample mixtures (3µl DNA + 2 µl 6X loading dye) was loaded into the wells and then allowed the electrophoresis at 70 V electric supplies.

- After completion of electrophoresis, switched off the supply and gel was visualised under UV illuminator to find out the presence of DNA and then photograph was documented using gel documentation apparatus (Maniatis *et al.*, 1982)

4.2.2 Polymerase chain reaction (PCR)

This technique is used to amplify the specific portion of the DNA i.e desired gene. It is an enzymatic process which occurs at different temperatures. It uses DNA as template and then with the help of primers, dNTPs it amplifies particular gene. Primers are short nucleotide sequences which are complementary to the gene. Amplification of gene involves denaturation in which at high temperature the double stranded DNA denatures to single stranded DNA. After this at a particular temperature which is based on the GC content of the primers, primer anneals to the DNA sequence and after this it extends. This process is repeated to a number of cycles to generate the multiple copies of desired gene.

Requirements

- 10X buffer
- BSA
- Forward primer
- Reverse primer
- dNTPs
- Taq DNA polymerase
- Water
- DNA sample

The purpose of primer is to provide a “free” 3’-OH group to which DNA Taq polymerase can add dNTPs. While designing or choosing primers, some points should be kept in mind that it is unable to anneal with each other to form dimer easily and also does not able to anneal itself to make hairpin structure.

Given below is the list of the primer used in amplification of *MGMT* and *ERCC1*.

For *MGMT* (rs12917)

Forward primer: 5-AAGAGTTCCCCGTGCCGAC-3

Reverse primer: 5-GCCAAACGCTGCCTCTGT-3

Alteration –C-T

Product size- 178 bp (Wang *et al.*, 2010) (Xiang *et al.*, 2015)

For *ERCC1* (rs3212986)

Forward primer: 5-TGCCAGAGACAGTGCCCCAAG-3

Reverse primer: 5-AGCTGCCAAGGAAACCCCCAG-3

Alteration- C-A

Product size-285bp (Chen *et al.*, 2014)

Table 4.5 Reagents used in PCR

Reagents	Stock concentration	Working concentration	Amount used for single 15 µl rxn
Additive BSA	100X	10X	1.5 µl
PCR buffer (Mg concn)	10X 25mM	1X 1.5mM	1.5 µl
Primers f and r	10pmol	0.5pmol	0.75 µl, 0.75 µl
dNTPs	10mM	0.2mM	0.3 µl
Taq polymerase	5U/µl	1U/ µl	0.2 µl
PCR grade water			8 µl
DNA template	100ng/ µl	200ng	2 µl

Table 4.6 Temperature and time conditions for PCR

Steps	Temperature	Time
Initial denaturation	95°C	5 min
Denaturation	94°C	30 sec
Annealing	For <i>MGMT</i> - 62°C For <i>ERCC1</i> -67°C	45sec 45sec
Polymerization	72°C	30 sec
Final extension	72°C	5 min

The above reaction was executed for 29 cycles.

4.2.3 Restriction fragment length polymorphism of *MGMT* and *ERCC1* gene

4.2.3.1 *MGMT* (rs12917)

Two units of *HinfI* enzyme was used to digest the amplified product of *MGMT* gene with the help of NEB buffer. After digestion of *MGMT* gene, *hinfi* enzyme cut the 178 bp product of *MGMT* into 161 bp and 17 bp. When digested mixture was resolved with the help of 6%

polyacrylamide gel by using native polyacrylamide gel electrophoresis then, if there is 161 bp and 17 bp product then it is wild type genotype (CC), if product is 178, 161, 17 bp then it is heterozygous type genotype (CT) and when product is 178bp then it is mutant type genotype (TT) in nature. On the basis of this, samples were classified as wild type genotype, heterozygous type genotype and mutant type genotype.

HinfI isolated from *Haemophilus influenzae* and its recognition sites is GANTC restriction site is GANTC (Source: <https://international.neb.com/products/R0155-HinfI>).

Table 4.7 Components for 6% polyacrylamide gel		
Sr no.	Reagents	Amount for 6 ml
1	30% Acrylamide	1.2 ml
2	Water	3.6 ml
3	5X TBE	1.2 ml
4	10 % APS	100 µl
5	TEMED	5 µl

4.2.3.2 *ERCC1* (rs3212986)

Two units of *mboII* enzyme were used to digest the amplified product of *ERCC1* gene. After digestion of *ERCC1* gene, *mboII* enzyme cut the 285 bp product of *ERCC1* into 181 and 104 bp. When digested mixture was resolved on 2.5 % agarose then, if there is 285 bp product then it is wild type genotype (CC), if there is 285, 181, 104 bp product then it is heterozygous type genotype (CA), if there is 181, 104 bp product then it is mutant type genotype (AA). On the basis of this samples were classified as wild type genotype, heterozygous type and mutant type genotype. Buffer used for this enzyme is Buffer B. Recognition site for this enzyme is GAAGA (N)₈ and restriction site for this enzyme is



Figure 4.1 Restriction site of *mboII* enzyme

(Source: <https://international.neb.com/products/R0148-MboII> #Product %20Information)

4.3 Statistical analysis

Statistical analysis are used to check whether the 2 SNP'S were in Hardy-Weinberg equilibrium or not ($p^2+2pq+q^2=1$); where p is the frequency of the wild-type allele and q is the frequency of the variant allele). χ^2 -test are used to assess the cases and independent t-test has been used for continuous variable like age and smoking. This was used to determine the genotypic frequency in cases using chi-square test. MedCal (version 18.5) software was used to compute genotypic frequency in a population. URL for MedCal software is MedCalc Softwarebvba,Ostend,Belgium; <https://www.medcalc.org>; 2016. Univariate and multivariate analysis are evaluated by using Kaplan-Meier survival analysis and Cox proportional hazardous ratio respectively. Logistic regression analysis were used that gives Odds-ratio (OR), 95% Confidence interval(CI) and p-value. By using logistic regression, cases are taken as independent variable while gender, age, smoking and pack years are taken as dependent variable. A p-value which is less than or equals to 0.05 is considered as highly significant (Chen *et al.*, 2014).

4.4 Bioinformatics methods

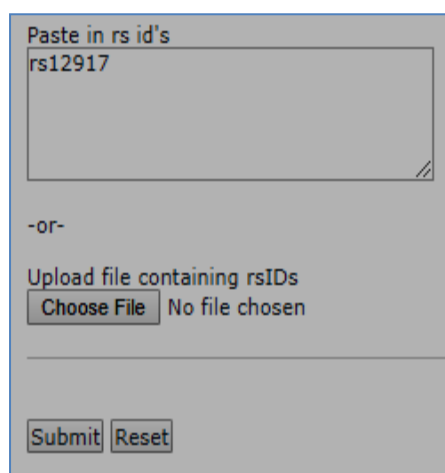
Computational analysis of deleterious single nucleotide polymorphisms (SNPs) in human *MGMT* and *ERCC1* gene involved in cancer using multiple softwares.

4.4.1 Prediction of structural impact of nsSNP on protein

This was performed with the help of SIFT software (Ng and Henikoff, 2003)

a. For *MGMT* rs12917

Input: rsID (rs12917)



Paste in rs id's
rs12917

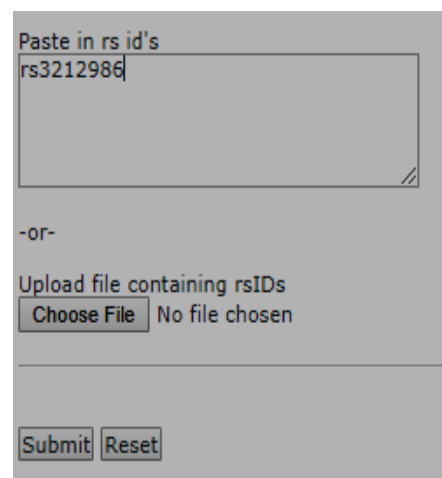
-or-

Upload file containing rsIDs
Choose File No file chosen

Submit Reset

b. For *ERCC1* rs3212986

Input: rsID (rs3212986)



Paste in rs id's
rs3212986

-or-

Upload file containing rsIDs
Choose File No file chosen

Submit Reset

Figure 4.2 Input of SIFT software for *MGMT* (rs12917) and *ERCC1* (rs3212986)
(Source: <http://sift.jcvi.org>)

4.4.2 Prediction of functional effect of amino acid substitution.

For this PROVEAN software was used (Choi and Chan, 2015)

Input: Chromosome no and position, Nucleotide change. These were retrieved from ncbi website.

a. For *MGMT* rs12917

10,129708019,C,T

b. For *ERCC1* rs3212986

19,45409478,G,T

Figure 4.3 Input of PROVEAN software for *MGMT* (rs12917) and *ERCC1* (rs3212986)
(Source: <http://provean.jcvi.org>)

4.4.3 Prediction of deleterious SNP by POLYPHEN2 software (Adzhubei *et al.*, 2013)

a. For *MGMT* rs12917

b. For *ERCC1* rs3212986

Input: rsID (rs12917)

Input: rsID (rs3212986)

Figure 4.4 Input of POLYPHEN 2 software for *MGMT* (rs12917) and *ERCC1* (rs3212986)
(Source: <http://genetics.bwh.harvard.edu/pph2/>)

4.4.4 To predict nsSNP impact on stability of protein with the help of I-Mutant software (Capriotti *et al.*, 2005)

INPUT: Protein sequence in FASTA format (retrieved from Swissprot), Position at which amino acid changes and amino acid changes into which new residue (from SIFT software output), temperature and pH.

Protein sequence of *MGMT* in FASTA format

```
>sp|P16455.1|MGMT_HUMAN RecName: Full=Methylated-DNA--protein-cysteine
methyltransferase; AltName: Full=6-O-methylguanine-DNA methyltransferase;
Short=MGMT; AltName: Full=O-6-methylguanine-DNA-alkyltransferase
MDKDCEMKRTTLDSP LGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGP
EPLMQCTAWLNAYFHQPEAIEEFVVPALHHPVFQQESFTRQVLWKLKVVKFGEVIS
YQQLAALAGNPKAARAVGGAMRGNPVPILIPCHRVCSSGAVGNYSGLLAVKEWL
LAHEGHRLGKPG LGGSSGLAGAWLKGAGATSGSPAGRN
(Source: https://www.ncbi.nlm.nih.gov/protein/)
```

Protein sequence of *ERCC1* in FASTA format

```
>sp|O15446.1|RPA34_HUMAN RecName: Full=DNA-directed RNA polymerase I subunit
RPA34; AltName: Full=A34.5; AltName: Full=Antisense to ERCC-1 protein; Short=ASE-1;
AltName: Full=CD3-epsilon-associated protein; Short=CAST; Short=CD3E-associated
protein; AltName: Full=RNA polymerase I-associated factor PAF49
MEEPQAGDAARFSCPPNFTAKPPASESPRFSLEALTGPDTLWLIQAPADFAPECFNG
RHHVPLSGSQIVKGLAGKRHRVRLSSCPQAGEATLLAPSTEAGGGLTCASAPQGTL
RILEGPQQSLSGSPLQPIPASPPPQIPPGLRPRFCAGGNPPVTGPRSALAPNLLTSGKK
KKEMQVTEAPVTQEAVNGHGALEVDMALGSPENDVRKKKKKKKNQQLKEPEAAGP
VGTEPTVETLEPLGVLPSTTKRKKPKGKETFEPEDKTVKQEQINTEPLEDTVLSPT
KKRKRQKGTEGMEPEEGVTVESQPQVKVEPLEEAIPLPPTKKRKKKEKGQMAMMEPG
TEAMEPVPEPMKPLESPGGTMAPQQPEGAKPQAQAALAAPKKKTKKEKQQDATVE
PETEVVGPELPDDLEPQAAPTSTKKKKKKKERGHTVTEPIQPLEPELPGEGQPEARAT
PGSTKKRKKQSQESRMPETVPQEEMPGLNSESSEGEAPTGRDKKRKQQQQQPV
(Source: https://www.ncbi.nlm.nih.gov/protein/)
```

a. For *MGMT* rs12917

b. For *ERCC1* rs3212986

Figure 4.5 Input of I-Mutant software for *MGMT* (rs12917) and *ERCC1* (rs3212986)

(Source: <http://folding.biofold.org/cgi-bin/i-mutant2.0.cgi>)

4.4.5 Prediction of stability of protein on the basis of change in energy by MuPro software (Cheng *et al.*, 2006)

INPUT: Mutation position, original amino acid, mutated amino acid, protein sequence in FASTA format. Mutation position, original amino acid and mutated amino acid was fetched back from SIFT software output and protein sequence was retrieved from Swissprot.

a. For *MGMT* rs12917

Mutation Name(optional):

Mutation Position: 84

Original Amino Acid: L

Substitute Amino Acid: F

Protein sequence(one plain sequence, no headers):

```
MDKDCMKRTTLDSPGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPEPLM
QCTAHLNAYFHQPEAIEEFVPALHHPVFQESFTRQVLKLLKVVKFGEVISYQQLAAL
AGNPKAARAVGGAMRGNPVPILIPCHRVVCSGAVGNYSGLAVKEWLLAHEGHRLGKPG
LGGSSGLAGAWLKAGATSGSPAGRII
```

Specify the protein structure file if available (optional):

Choose File No file chosen

Predict

b. For *ERCC1* rs3212986

Mutation Name(optional):

Mutation Position: 504

Original Amino Acid: Q

Substitute Amino Acid: K

Protein sequence(one plain sequence, no headers):

```
MEEPQAGDAARFSCPPNFTAKPPASESPRFSLEALTGPDTELWLQAPADFAPECFNGRH
VPLSGSQIVKGLAGKRHYRVLSLSCPQAGEATLLAPSTEAGGGGLTCASAPQGTLRILEG
PQQSLSGSPLOPIPASPPPIPPGLRPRFCAGGNPPVTGPRSALAPNLLTSQKKKKEMQ
VTEAPVTQEAIVNGHGALEVDHMGSPENDVRKKKKKKQQLKEPEAAGPVGTPTVETLE
PLGVLFPSTTKRKKPKGKETFEPEDKTVKQEQINTEPLEDTVLSPTKKRKRQKGTGME
PEEGVTVESQPQVKVEPLEEAIPPTKKRKKKEKGQMMMEPGTEAMEPVPEMKPLESP
GGTMAPQQPEGAKPQAQALAAPKKTKKEKQQDATVEPETEVVGPGLPDDLEPQAAPS
```

Specify the protein structure file if available (optional):

Choose File No file chosen

Predict

Figure 4.6 Input of MuPro software for *MGMT* (rs12917) and *ERCC1* (rs3212986)
(Source: <http://www.ics.uci.edu/~baldig/mutation.html>)

4.4.6 To anticipate whether the mutation is disease related or neutral

This was performed by Phd-SNP software (Capriotti *et al.*, 2017)

INPUT: Protein sequence in FASTA format, Swiss-Prot Code for protein, mutation position, mutated amino acid and basis of prediction. Protein sequence and code for protein was retrieved from Swissprot. Mutation position, mutated amino acid was fetched back from SIFT software output.

a. For *MGMT* rs12917

Protein Sequence: MDKDCMKRTTLDSPGLKLESGCEQGLHEIKLLGKGTSAADAVEVPAPA
AVLGGPEPLM
QCTAWLNAYFHQPEAIEEFPPVPAHHPVFQESFTRQVLWLLKVKVFGE
VTSYQQLAAL
AGNPKAARAVGGMARGNPVPIIPCHRVCSSGAVGNYSGLAVKEWLLA
HEGHRLLGKPG
LGGSSGLAGAWLKAGATSGSPAGRN

Swiss-Prot Code: P16455

Sequence File: Choose File No file chosen

Position: 84

New Residue: F

Prediction: ☐ Sequence-Based
☐ Hybrid Method (old version)
☒ Sequence and Profile-Based

Multi SVM: ☐ 20-fold cross-validation prediction

e-mail:

Submit

b. For *ERCC1* rs3212986

Protein Sequence: MEEPQAGDAARFSCPPNFTAKPPASESPRFSLEALTGPOTELWLIQA
PADFAECFNGRHVPLSGSQIVKGLAGKRRHVRVLSQCPQAGEATL
LAPSTEAGGGLTCASAPQGLRLLEGPOQLSGSPLOTPASPPQT
PPGLRPRFCAGGPNPVTGPRSAAPNLLTSGKKKEQVTEAPVTQ
EAVNGHGALEVDNALGSPHVDKRRKKKKKQQLKEPEAAGPVTEPT
VETLEPLGLVLPSTTKKKKKPKGKETFEPEDKTVKQEQINTEPLEDT
VLSPTKKRRKQKGTGMEPEEGVTVESQPVKVEPLEAIPLPPTKK

Swiss-Prot Code: O15446

Sequence File: Choose File No file chosen

Position: 504

New Residue: K

Prediction: ☐ Sequence-Based
☐ Hybrid Method (old version)
☒ Sequence and Profile-Based

Multi SVM: ☐ 20-fold cross-validation prediction

e-mail:

Submit

Figure 4.7 Input of Phd-SNP software for *MGMT* (rs12917) and *ERCC1* (rs3212986)
(Source: <http://snps.biofold.org/phd-snp/phd-snp.html>)

4.4.7 To predict whether mutation is related to disease or not by SNP & GO software (Calabrese *et al.*, 2009)

INPUT: UNIPROT accession Number for protein, Mutation position, original amino acid, Mutated amino acid. Accession number was fetch from Swissprot website and mutation position, original amino acid, mutated amino acid was retrieved from SIFT software output.

a. For *MGMT* rs12917

UNIPROT Accession Number: P16455

Mutation Position: 84

Wild-type residue: L

Substituting residue: F

Submit Clear

b. For *ERCC1* rs3212986

UNIPROT Accession Number: O15446

Mutation Position: 504

Wild-type residue: Q

Substituting residue: K

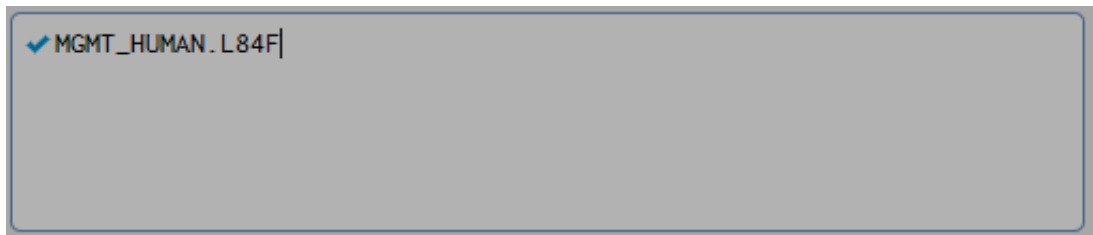
Submit Clear

Figure 4.8 Input of SNP & GO software for *MGMT* (rs12917) and *ERCC1* (rs3212986)
(Source: <https://snps-and-go.biocomp.unibo.it/snps-and-go/>)

4.4.8 Prediction of effects of mutations on folding of protein and protein-protein interactions by ELASPIC Software (Witvliet *et al.*, 2016)

INPUT: name of gene_organism_original amino acid changes at which position into what kind of new amino acid

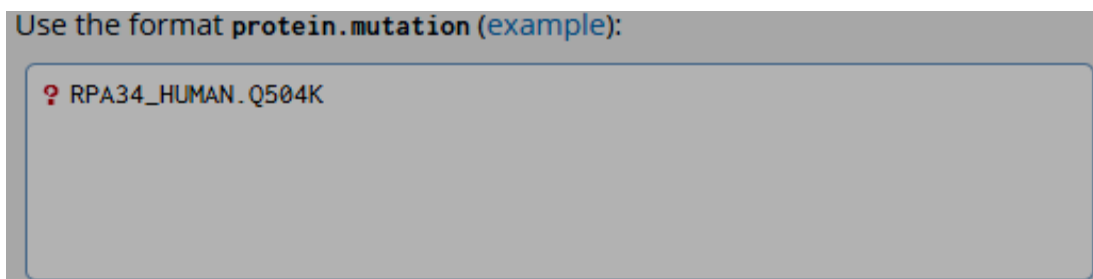
a. For *MGMT* rs12917



A screenshot of a web form input field. The text 'MGMT_HUMAN.L84F' is entered, preceded by a blue checkmark icon. The input field is a light gray rectangle with a thin blue border.

b. For *ERCC1* rs3212986

RPA34 is anti-sense strand to *ERCC1*



A screenshot of a web form input field. Above the input field, the text 'Use the format protein.mutation(example):' is displayed. Inside the input field, the text 'RPA34_HUMAN.Q504K' is entered, preceded by a red question mark icon. The input field is a light gray rectangle with a thin blue border.

Figure 4.9 Input of ELASPIC software for *MGMT* (rs12917) and *ERCC1* (rs3212986)
(Source: <http://elaspic.kimlab.org/many/>)

4.4.9 Prediction of deleterious SNP with the help of PANTHER (Thomus *et al.*, 2003)

INPUT:

Protein sequence: retrieved from Swiss-prot database

Amino acid substitution: |wild type amino acid||position of mutation||mutated amino acid|.

These are fetch from SIFT Software output

Name of organism

For *MGMT* (rs12917)

Enter a protein sequence: ?

```
>sp|P16455|MGMT_HUMAN Methylated-DNA--protein-cysteine
methyltransferase OS=Homo sapiens GN=MGMT PE=1 SV=1
MDKDCEMKRTTLDSPLGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAA
VLGGPEPLM
QCTAWLNAYFHQPEAIEEFVVPALHHPVFQQESFTRQVLWKLLKVVKFGEVI
SYQQLAAL
AGNPKAARAVGGAMRGNPVPILIPCHRVVCSSGAVGNYSGLAVKEWLLA
HEGHR LGKPG
LGGSSGLAGAWLKGAGATSGSPAGRN
```

Enter substitution(s), e.g. A10P ?

L84F

Select single organism

- Homo sapiens
- Mus musculus
- Rattus norvegicus
- Gallus gallus
- Danio rerio

Submit

For *ERCC1* (rs3212986)

Enter a protein sequence: ?

```
PLGVLFPSSTTKRKKPKGKETFEPEDKTVKQEQINTEPLEDTVLSPTKK
RKRQKGTEGME
PEEGVTVESQPQVKVEPLEEAIPLPTTKRKKKEKGQMAMMEPGTEAME
PVEPEMKPLESP
GGTMAPQQPEGAKPQAQAALAAPKKKTKKEKQQDATVEPETEVVGPE
LPDDLEPQAAPTS
TKKKKKKKKERGHTVTEPIQPLEPELPGEGQPEARATPGSTKKRKKQSQ
ESRMPETVPQEE
MPGPPLNSESSEFAPTGRDKKRKQQQQQPV
```

Enter substitution(s), e.g. A10P ?

Q504K

Select single organism

- Homo sapiens
- Mus musculus
- Rattus norvegicus
- Gallus gallus
- Danio rerio

Submit

Fig 4.10 Input of PANTHER software for *MGMT* (rs12917) and *ERCC1* (rs3212986)
(Source: <http://pantherdb.org/tools/csnpscoreForm.jsp>)

4.4.10 Prediction of harmful mutation by MutPred Software (Pejaver et al., 2017)

INPUT:

Email ID

Sequence and Substitution Information: |Swissprot protein code||sequence size||wild amino acid||position of mutation||Mutated amino acid

a. For *MGMT* (rs12917)

E-Mail

anugupta94ng@gmail.com

Sequence and substitution information (FASTA format)

```
>sp|P16455|1-207 L84F
MDKDCCEMKRTTLDSP LGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPEPLM
QCTAWLNAYFHQPEAIEEFVPA LHHPVFQQESFTRQVLWKLLKVVKFGEVISYQQLAAL
AGNPKAARAVGGAMRGNPVPILIPCHRVVCSGAVGNYSGLAVKEWLLAHEGHRLLGKPG
LGGSSGLAGAWLKGAGATSGSPAPGRN
```

OR

Upload FASTA file

Choose File No file chosen

P-value threshold (optional)

0.05

Submit

b. For *ERCC1* (rs3212986)

E-Mail

anugupta94ng@gmail.com

Sequence and substitution information (FASTA format)

```
>sp|Q15446|1-510 Q504K
MEEPQAGDAARFSCPPNFTAKPPASESPRFSLEALTGPDTELWLIQAPADFAPECFNGRH
VPLSGSQIVKGLAGKRHRVRLSSCPQAGEATLLAPSTEAGGGLTCASAPQGTLRILEG
PQQSLSGSPLOPIASPPPQIPPLRPRFCAGGNPPVTGPRALAPNLLTSGKKKKEMQ
VTEAPVTQEA VNGHGALEVD MALGSPMDVRKKKKKKQQLKEPEAAGPVGTEPTVETLE
PLGVLPSTTKRKKPKGKETFEPEDKTVKQEINTEPLEDTVLSPTKKRKRQKGTGEGME
PEEGVTVESQPVKVEPLEEAIPLPTKKRKKKEGQMAMMEPGTEAMEPVEPEMKPLESP
GGTMAPQQPEGAKPQAQAALAAPKKTKKEKQDQATVEPETEVVGPGLPDDLEPQAAPTS
TKKKKKKKERGHVTVEIQPLEPELPGEGQPEARATPGSTKKRKKQSQESRMPETVPQEE
MPGPPLNSES GEEAPTGRDKKRKQQQQQPV
```

OR

Upload FASTA file

Choose File No file chosen

P-value threshold (optional)

0.05

Submit

Fig 4.11 Input of MutPred software for *MGMT* (rs12917) and *ERCC1* (rs3212986)

(Source: <http://mutdb.org/mutpred>)

4.4.11 Prediction of spatial arrangement of amino acid substitution on protein models and structures with the help of Mutation 3D software (Meyer *et al.*, 2016)

INPUT:

Protein ID: retrieved from UNIPROT

Mutation: |wild type amino acid ||position of mutation||mutated amino acid|

a. For *MGMT* (rs12917)

b. For *ERCC1* (rs3212986)

Fig 4.12 Input of Mutation 3D software for *MGMT* (rs12917) and *ERCC1* (rs3212986)

(Source: <http://mutation3d.org/index.shtml>)

4.4.12 Prediction of interaction of gene with other genes with the help of GENEMANIA Software (Warde-Farley *et al.*, 2010)

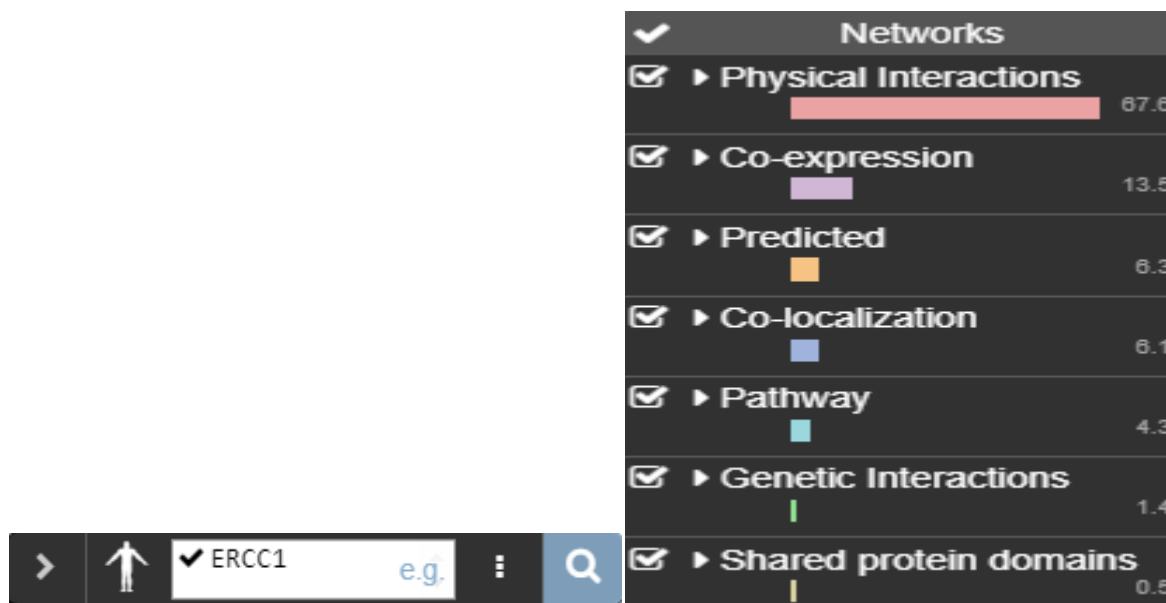
INPUT:

(I) for *MGMT*

a. for all networks

Network Type	Percentage
Physical Interactions	67.64%
Co-expression	13.50%
Predicted	6.35%
Co-localization	6.17%
Pathway	4.35%
Genetic Interactions	1.40%
Shared protein domains	0.59%

For *ERCC1*



For prediction of interaction of *MGMT* and *ERCC1*

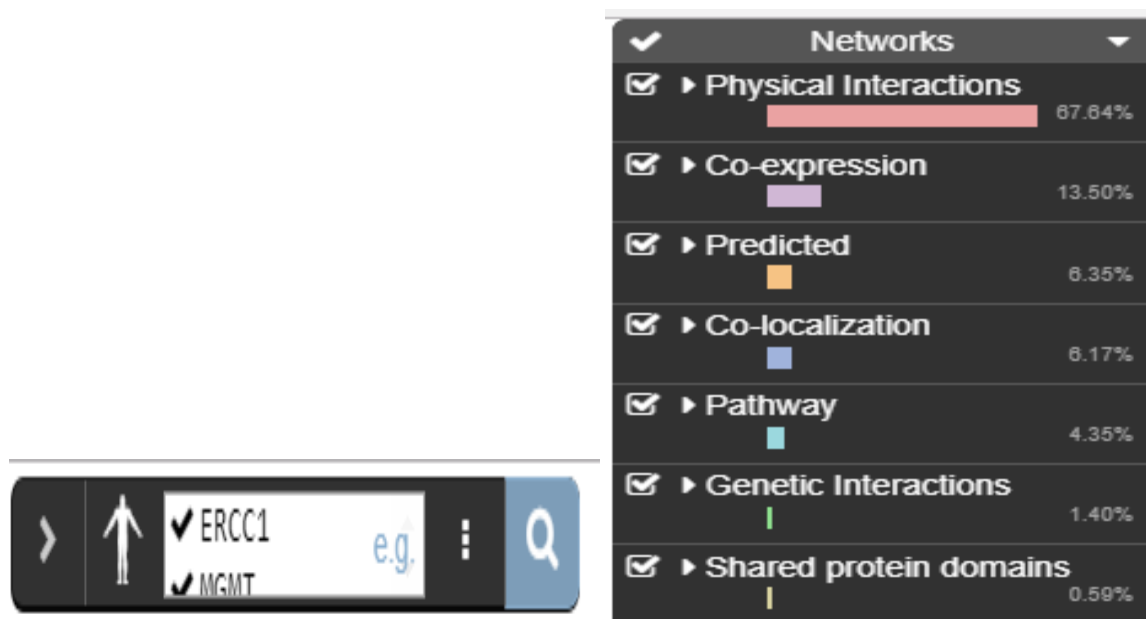


Fig 4.13 Input of GENEMANIA software for *MGMT* (rs12917), *ERCC1* (rs3212986) and interaction of *ERCC1* and *MGMT*

(Source: <http://www.genemania.org/>)

Chapter 5

RESULTS

5.1. Genotyping

5.1.1 DNA Isolation

To isolate the genomic DNA, 1ml of peripheral blood sample was used. Presence of genomic DNA was confirmed with the help of 0.8% agarose gel. Concentrated DNA was further diluted to 100ng/μl with TE buffer. Genomic DNA with concentration of 100ng/μl was used as template to amplify the desire gene with the help of polymerase chain reaction (PCR). The gel of isolated DNA from blood is shown in figure 5.1.

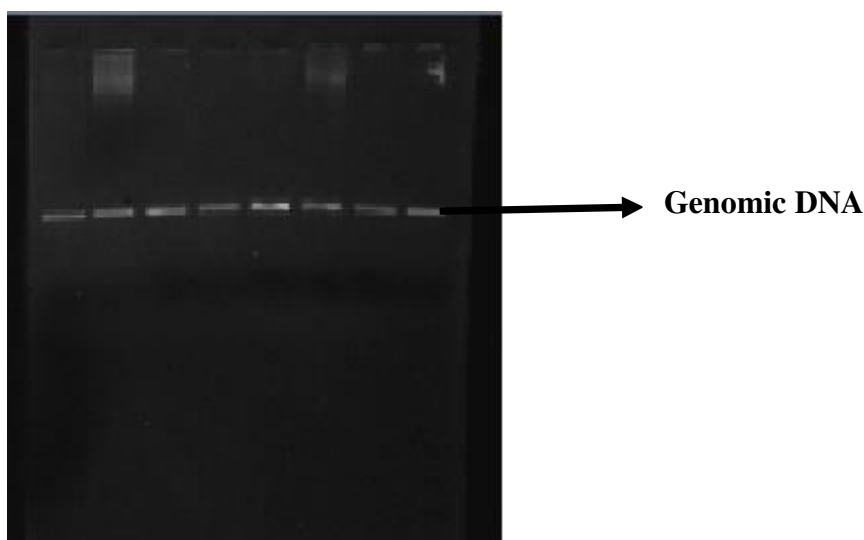


Figure 5.1 Isolated genomic DNA

5.1.2 Amplification of *MGMT* (rs12917) and *ERCC1* (rs3212986) gene

MGMT (rs12917) gene was amplified with the help of its respective primers and reaction mixtures by polymerase chain reaction. Then the amplified product was qualitatively analysed by 2% agarose gel. Amplified product of *MGMT* (rs12917) was 178 bp in size as shown in figure 5.2.

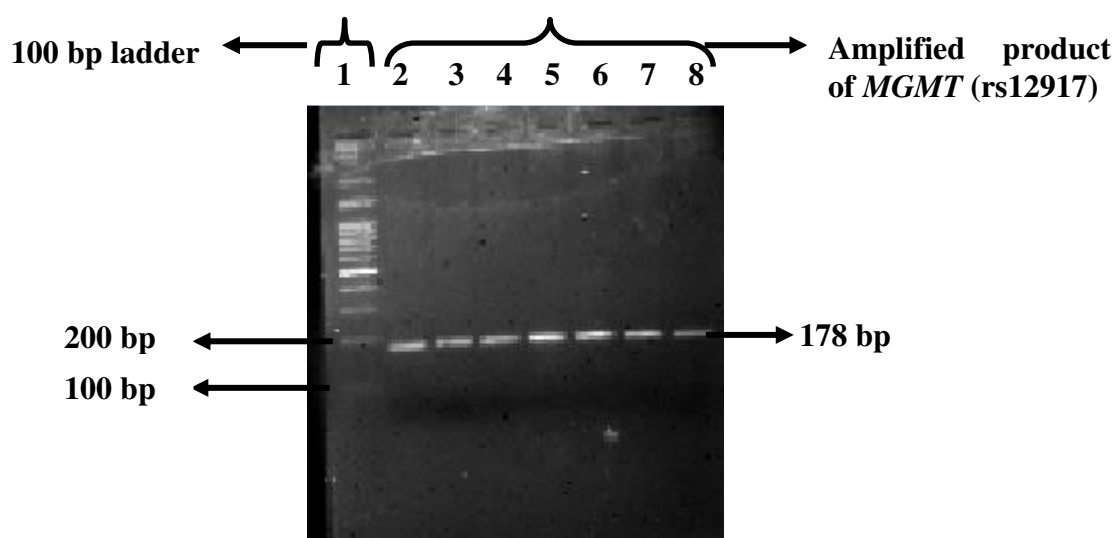


Figure 5.2 Amplified product of *MGMT* (rs12917)

ERCC1 (rs3212986) gene was amplified with the help of its respective primers and reaction mixtures by polymerase chain reaction. 2% agarose gel was used to analyse the amplified product qualitatively. Amplified product of *ERCC1* (rs3212986) was 285 bp in size as shown in figure 5.3.

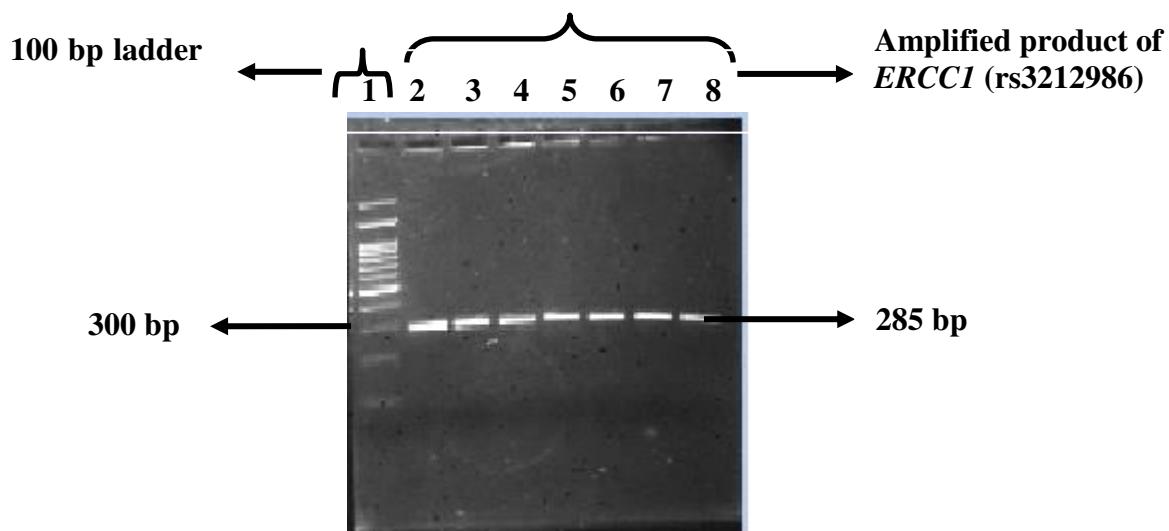


Figure 5.3 Amplified product of *ERCC1* (rs3212986)

5.1.3 Restriction digestion of *MGMT* (rs12917) and *ERCC1* (rs3212986) gene

After restriction digestion of amplified product of *MGMT* (rs12917) gene by *HinfI* enzyme, the digested amplified product was analysed with the help of 6% polyacrylamide gel electrophoresis (PAGE). Different band size on the polyacrylamide gel corresponded to different variants of *MGMT* gene like the band size of 161 and 17 bp corresponded to wild type genotype, band size of 178, 161 and 17 bp corresponded to heterozygous genotype as shown in figure 5.4. There was no band of 178 bp in the current study, no mutant was observed.

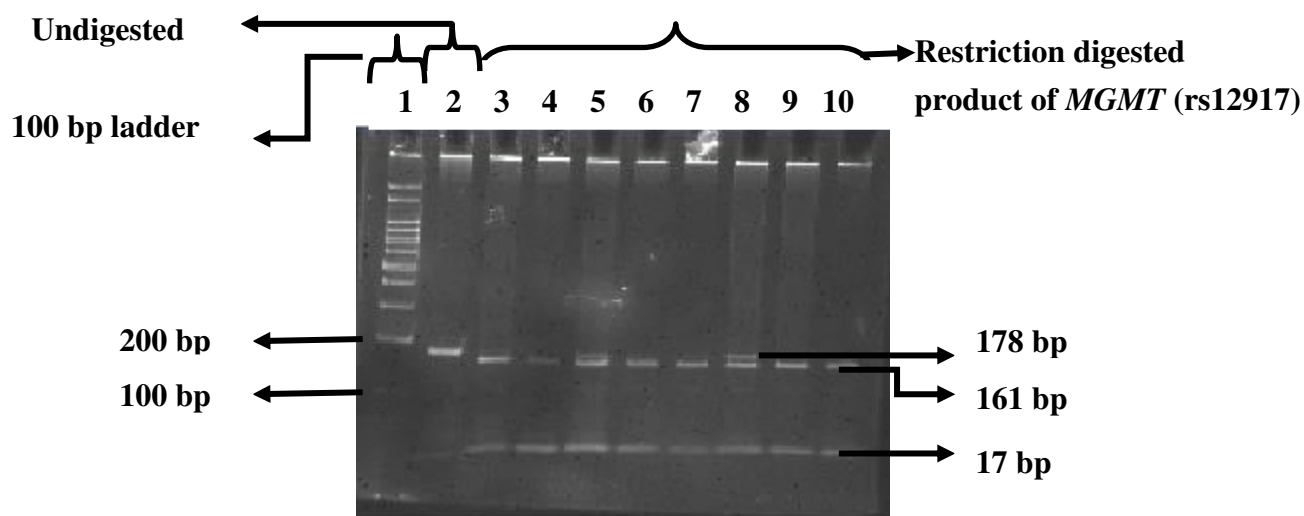


Figure 5.4 Restriction digested product of *MGMT* (rs12917)

(Lane 3, 4, 6, 7, 9, 10-wild genotype; 5, 8-Hetero genotype)

After digestion of amplified product of *ERCC1* (rs3212986) gene with *MboII* enzyme, the digested amplified product was analysed with the help of 2.5% agarose gel electrophoresis. Different band size on the agarose gel corresponded to different variants of *ERCC1* gene like single band of 285 bp size corresponded to wild type genotype, two bands of 181 and 104 bp were of mutant type genotype and three bands of 285, 181 and 104 bp were of heterozygous type genotype as shown in figure 5.5.

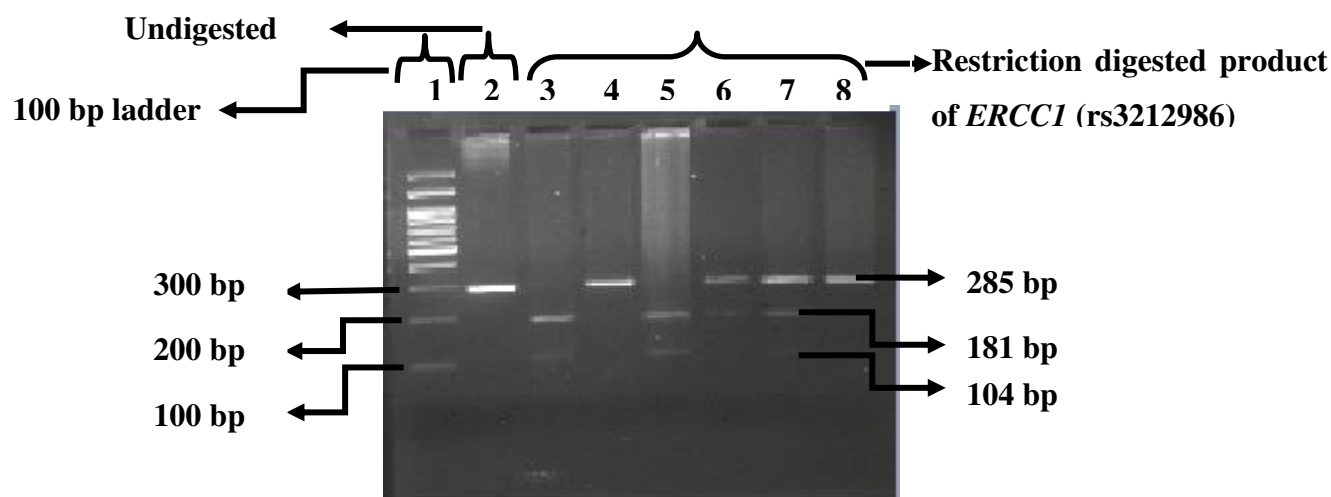


Figure 5.5 Restriction digested product of *ERCC1* (rs3212986)

(Lane 4,8-wild genotype; 6,7-hetero genotype; 3,5-mutant genotype)

5.2 Statistical analysis

5.2.1 Distribution of *MGMT* (rs12917) and *ERCC1* (rs3212986) genotype

The genotypic distribution of *MGMT* (rs12917) included a total sample size of 214 lung cancer patients, among these 167 cases were wild type genotype and 47 were carrying heterozygous type genotype. Genotypic distribution of *MGMT* (rs12917) was found to fall in Hardy-Weinberg equilibrium as there was a non-significant relationship observed between *MGMT* (rs12917) and genotype ($\chi^2 = 3.26$, $P = 0.07$).

Genotypic distribution of *ERCC1* (rs3212986) included a total sample size of 214 lung cancer patients, among these 107 cases were of wild type genotype, 87 cases were harbouring heterozygous type genotype and 20 cases were of mutant type genotype. Genotypic distribution of *ERCC1* (rs3212986) also was found to fall in Hardy-Weinberg equilibrium as there was a non-significant relationship found between *ERCC1* (rs3212986) and genotype ($\chi^2 = 0.14$, $P = 0.70$).

We also evaluated combinatorial effect of *MGMT* (rs12917) and *ERCC1* (rs3212986) in lung cancer patients. Our study showed that out of 214 lung cancer patients, there were 107 patients

who have common genotype for both the SNP. Further it was observed out of 107 cases; 85 cases were of wild type genotype and 22 cases were carrying heterozygous genotype. This genotypic distribution was found to fall in Hardy-Weinberg equilibrium as there was a non-significant relationship found between combinatorial *MGMT* (rs12917), *ERCC1* (rs3212986) and genotype ($\chi^2 = 1.40$, $P = 0.24$).

5.2.2 Demographic distribution of lung cancer patients for *MGMT* (rs12917) and *ERCC1* (rs3212986)

Non-clinical parameters studied for demographic features were age, smoking status, gender and pack years of smoking. Pathological and clinical parameters included tumor stage, histology, tumor size extension, metastasis and lymph node involvement. Clinical response parameters were also studied which included CR (complete response), PR (partial response), SD (stable disease) and PD (progressive disease). Other parameters like regimen, KPOS and ECOG were also studied for lung cancer patients.

Table 5.1 Demographic distribution of lung cancer patients for <i>MGMT</i> and <i>ERCC1</i>			
Variable	Cases, n (%) n=214	Variable	Cases, n (%) n=214
Age (years) Mean $\pm$ SD Range	59.03 $\pm$ 10.24 31-85	Lymph node involvement N0 N1 N2 N3 N4 Others	21 (9.81) 21 (9.81) 89 (41.59) 56 (26.17) 02 (0.93) 25 (11.68)
Gender Male Female	190 (88.78) 24 (11.21)	Metastasis M0 M1 N/A	105 (49.06) 85 (39.72) 24 (11.21)
Smoking status Smokers Non-smokers	182 (85.05) 32 (14.95)	KPS 30-60 70-80 90-100 N/A	30 (14.02) 103 (48.13) 73 (34.11) 8 (3.74)
Pack years Mean $\pm$ SD	28.40 $\pm$ 12.71	ECOG 0 1 2 3	32 (14.95) 69 (32.24) 80 (37.38) 22 (10.28)

		4 N/A	3 (1.40) 8 (3.74)
Histology		Regimen*	
SQCC	80 (37.38)	Docetaxel	49 (22.90)
ADCC	69 (32.24)	Irinotecan	44 (20.56)
SCLC	60 (28.04)	Pemetrexed	45 (21.03)
Others	5 (2.34)	Others	15 (7.01)
		N/A	61 (28.50)
Tumor Size		TMN staging	
T0	6 (2.80)	I	2 (0.93)
T1	8 (3.74)	II	4 (1.87)
T2	26 (12.15)	III	98 (45.79)
T3	47 (21.96)	IV	89 (41.59)
T4	103 (48.13)	N/A	21 (9.81)
N/A	24 (11.21)		
Response			
CR	2 (1.71)		
PR	65 (55.55)		
SD	42 (35.90)		
PD	8 (6.84)		
Abbreviations: n= total number of lung cancer patients, SD- standard deviation, M0- metastasis not involved, M1- metastasis involved, KPS- karnofsky performance scale, ECOG- eastern cooperative oncology group, ADCC- adenocarcinoma, SQCC- squamous cell carcinoma, SCLC- small cell lung carcinoma, PR- partial response, CR- complete response, PD- progressive disease, SD- stable disease, Regimen* - in combination with cisplatin/carboplatin, N/A- data not available.			

Demographic features and clinical data of 214 lung cancer patients are shown in table 5.1. Among 214 Lung cancer cases, the average age with standard deviation of patient was 59.03 ± 10.24 , minimum age was 31 years and maximum age was 85 years. Total number of males in lung cancer cases were 190 (88.78%) and females were 24 (11.21%) as shown in figure 5.6 (a), it showed that lung cancer occur mostly in males. Total number of smokers was high as compared to non-smokers [smokers {182 (85.05%)}, non-smokers {32 (14.95%)}] which indicated that smoking may be a causative agent for lung cancer as shown in figure 5.6 (b). Average pack year for lung cancer patients was 28.40 ± 12.71 .

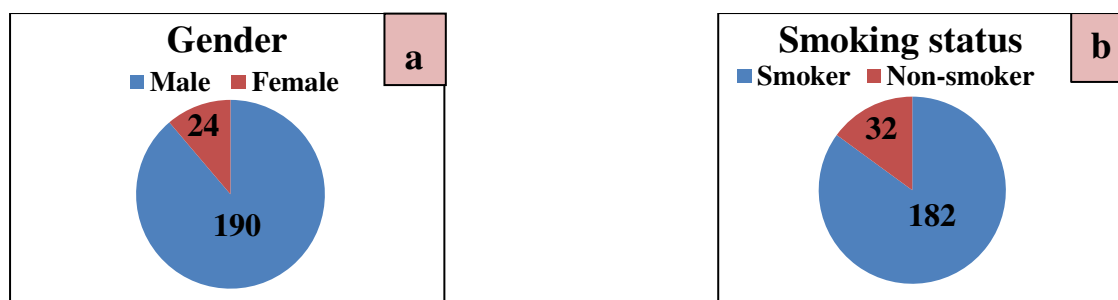
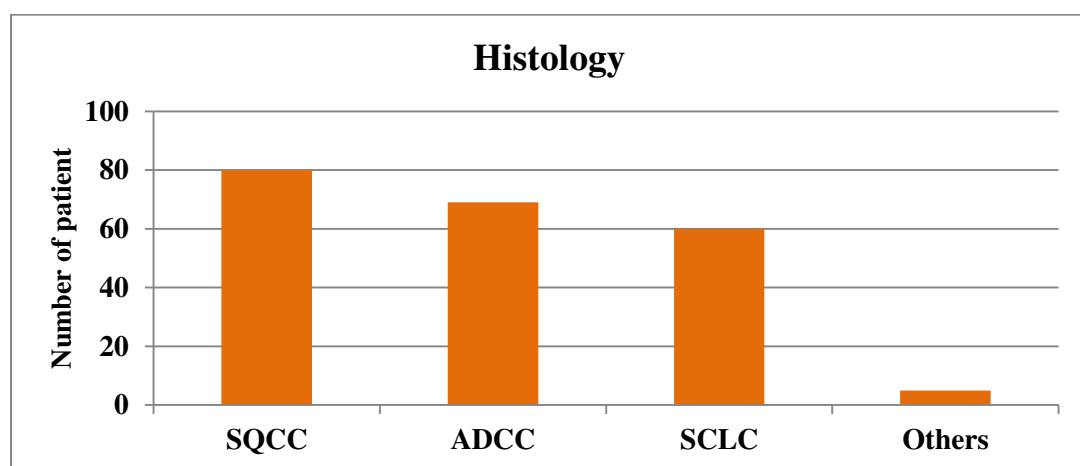


Figure 5.6 Frequencies of gender (a) and smoking status (b) in current study population to observe individual effect of *MGMT* and *ERCC1* gene polymorphism

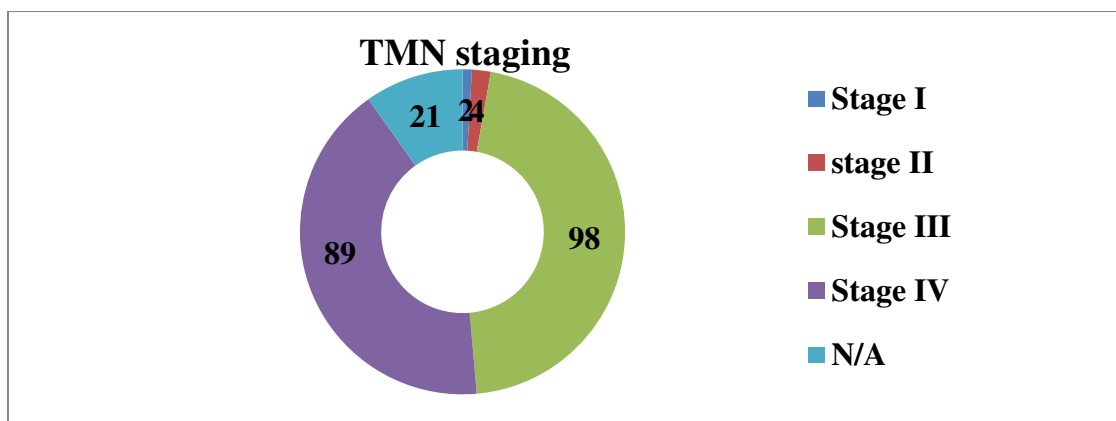
On the basis of histology, the most common type of lung cancer was SQCC [80(37.38%)], then ADCC [69 (32.24%)] and SCLC [60 (28.04%)]. Patients whose histology inconclusive was 5 (2.33%) as shown in figure 5.7.



(Abbreviation: ADCC- adenocarcinoma, SQCC- squamous cell carcinoma, SCLC- small cell lung carcinoma)

Figure 5.7 Frequencies of SQCC, ADCC and SCLC in current study population to observe individual effect of *MGMT* and *ERCC1* gene polymorphism

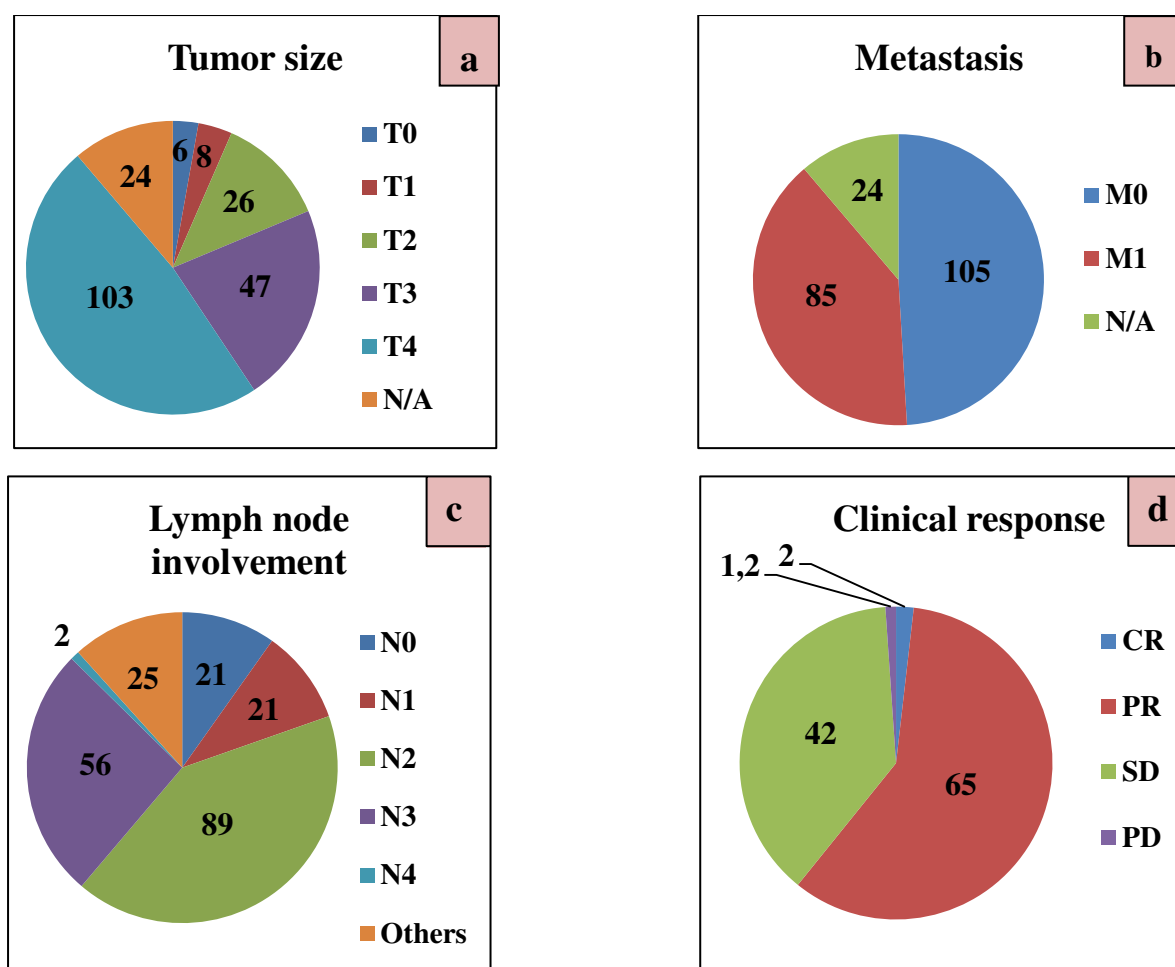
When studied about TNM staging it was found that there were more lung cancer cases of stage III [98 (45.79%)] and stage IV [89 (41.59%)], but number and percentage of cases was less for stage I [2 (0.93%)] and stage 2 [4 (1.87%)]. For 9.81% patients, data was not available. This indicated that stage III and stage IV are more prominent than stage I and stage II cases as shown in figure 5.8.



(Abbreviation: N/A- data not available)

Figure 5.8 Frequency of lung cancer stage in present study population to observe individual effect of *MGMT* and *ERCC1* gene polymorphism

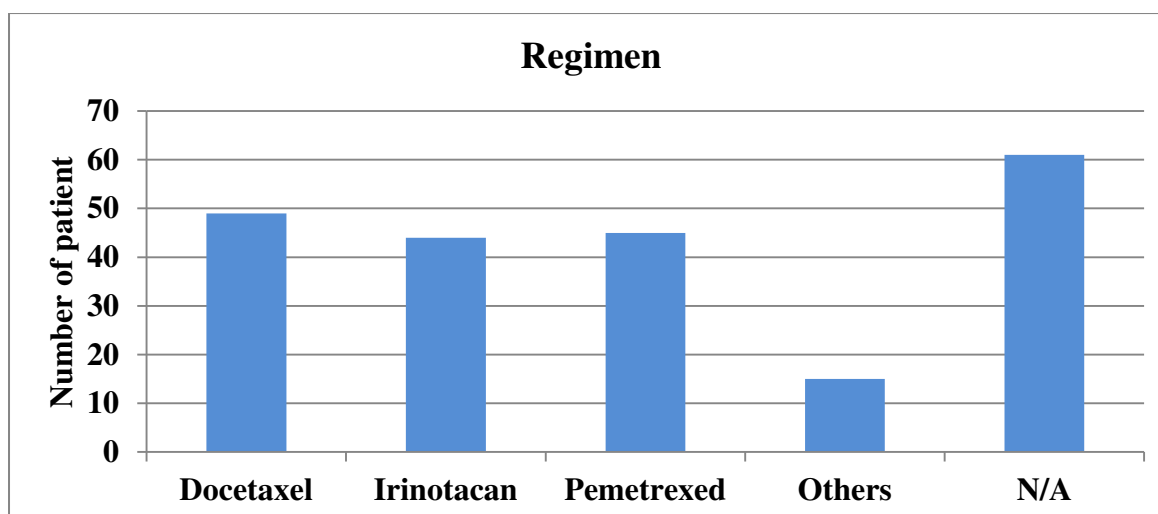
Tumor size among lung cancer patients ranged from 0-4. Highest number and percentage of tumor size in lung cancer cases was of T4 type tumor size; 103 (48.13%), then followed by T3 type; 47 (21.96%), T2 type; 26 (12.15%), T1 type; 8 (3.74%) and T0 type; 6 (2.80%). This study showed that T4 type of tumor size is preponderant over T3, T2, T1 and T0 types respectively. Data for tumor size of twenty four lung cancer patients was not available as shown in figure 5.9 (a). Among lung cancer patients approx 85 (39.72%) cases showed distant metastasis, while 105 (49.06%) cases did not show any metastasis. Data was not available for 24 (11.21) patients. This means that maximum lung cancer cases did not show any metastasis as shown in figure 5.9 (b). When studied about lymph node involvement in lung cancer, it was found that patients showed highest involvement of N2 lymph node [89 (41.59%)] and lowest involvement of N4 lymph node [2 (0.93%)]. 26.17 % patients showed N3 lymph node involvement, while 9.81% patients showed N0 and N1 lymph node involvement. 11.68% patients' data was not reported in current study population as shown in figure 5.9 (c). Among 214 lung cancer cases only 117 cases were computable for clinical response study. A large number of cases showed partial response (PR) (55.55%), while only 1.71% cases showed complete response (CR). 35.90% cancer patients showed stable disease (SD), while 6.84% showed progressive disease (PD) after treated with chemotherapy. This indicated that among cancer cases most common clinical response after chemotherapy is Partial response (PR) and stable disease (SD) as shown in figure 5.9 (d).



(Abbreviation: M0- no metastasis, M1- metastasis, CR- complete response, PR-partial response, SD- stable disease, PD- progressive disease, N/A- Data not available)

Figure 5.9 Frequencies of tumor size (a), metastasis (b), lymph node involvement (c) and clinical response (d) in current study population to observe individual effect of *MGMT* and *ERCC1* gene polymorphism

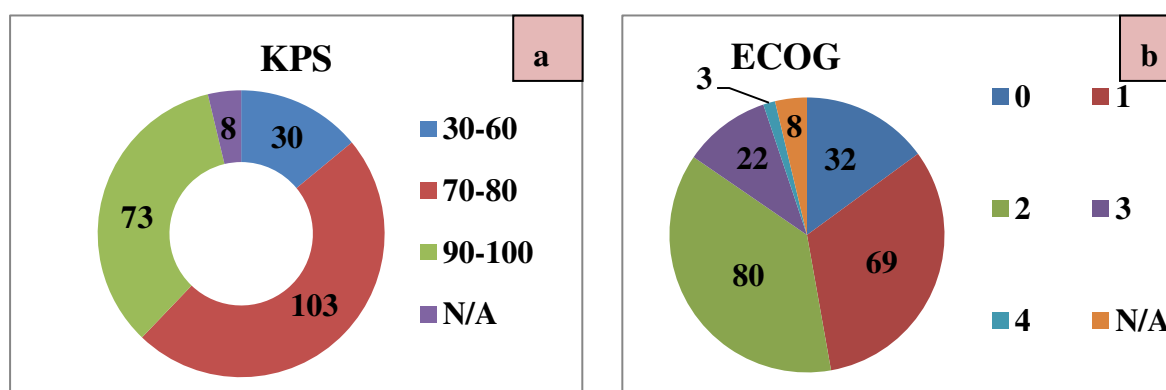
When chemotherapy regimen was studied, it was found that 22.90% patients were administered Docetaxel + cisplatin/ carboplatin treatment, 20.56% were treated with Irinotecan + cisplatin/ carboplatin and 21.03% were administered Pemetrexed + cisplatin/ carboplatin. 7.01% patients were treated with other drugs, while 28.50% patients had no information about regimen as shown in figure 5.10.



(Abbreviation: N/A- data not available)

Figure 5.10 Frequency of regimen in present study population to observe individual effect of *MGMT* and *ERCC1* gene polymorphism

When performance status of lung cancer patients was evaluated, it was observed that 14.02% cases had KPS range between 30-60, 48.13% had KPS range of 70-80 and 34.11 % cases had KPS size of 90-100. For 3.74% cancer cases, data was not available as shown in figure 5.11 (a). On the basis of ECOG, 14.95% cancer patients showed ECOG size of 0, 32.24% cases showed ECOG value of 1, 37.38% cases showed ECOG size of 2, 10.28% cases had ECOG value of 3 and 1.40% showed ECOG size of 4 and no data available for 3.74% cancer cases as shown in figure 5.11 (b).



(Abbreviation: KPOS- karnofsky performance scale, ECOG- eastern cooperative oncology group, N/A- data not available)

Figure 5.11 Frequencies of KPS (a) and ECOG (b) in current study population to observe individual effect of *MGMT* and *ERCC1* gene polymorphism

5.2.3 Demographic distribution of lung cancer patients for combined effect of *MGMT* (rs12917) and *ERCC1* (rs3212986)

Table 5.2 Demographic distribution of lung cancer patients for combined effect of <i>MGMT</i> (rs12917) and <i>ERCC1</i> (rs3212986)			
Variable	Cases, n (%) n-107	Variable	Cases, n (%) n-107
Age (years) Mean $\pm$ SD Range	59 $\pm$ 9.90 35-85	Lymph Node involvement N0 N1 N2 N3 Others	13 (12.15) 12 (11.21) 43 (40.19) 28 (26.17) 11 (10.28)
Gender Male Female	97 (90.65) 10 (9.35)	Metastasis M0 M1 N/A	50 (46.73) 46 (42.99) 11 (10.28)
Smoking status Smokers Non-smokers	94 (87.85) 13 (12.15)	KPS 30-60 70-80 90-100 N/A	17 (15.89) 54 (50.47) 32 (29.91) 4 (3.74)
Pack years Mean $\pm$ SD	28.95 $\pm$ 12.14	ECOG 0 1 2 3 4 N/A	11 (10.28) 37 (34.58) 39 (36.45) 15 (14.02) 1 (0.93) 4 (3.74)
Histology SQCC ADCC SCLC Others	42 (39.25) 34 (31.77) 30 (28.04) 1 (0.93)	Regimen [*] Docetaxel Irinotecan Pemetrexed Others N/A	28 (26.17) 21 (19.63) 25 (23.36) 7 (6.54) 26 (24.30)
Tumor Size T0 T1 T2 T3 T4 N/A	1 (0.93) 3 (2.80) 15 (14.02) 27 (25.23) 50 (46.73) 11 (10.28)	TMN staging I II III IV N/A	2 (1.87) 2 (1.87) 47 (43.92) 48 (44.86) 8 (7.48)
Response PR SD	N=60 37 (61.67) 19 (31.67)		

PD	4 (6.67)		
CR	0 (0)		
Abbreviations: N= total number of lung cancer patients, SD- standard deviation, M0-metastasis not involved, M1-metastasis involved, KPS- karnofsky performance scale, ECOG-eastern cooperative oncology group, ADCC-adenocarcinoma, SQCC-squamous cell carcinoma, SCLC-small cell lung carcinoma, PR-partial response, CR-complete response, PD- progressive disease, SD-stable disease, Regimen*-chemotherapy in combination with cisplatin/carboplatin, N/A- data not available.			

Table 5.2 described the demographical features and clinical data of 107 lung cancer patients. Among 107 lung cancer cases, the average age with standard deviation was 59 ± 9.90 , minimum and maximum age of patient was 35 and 85 years respectively. Number of males in lung cancer cases was high as compared to females *i.e* 97 (90.65%) and 10 (9.35%) respectively, this showed that lung cancer occurs mostly in males as shown in figure 5.12 (a). Total number and percentage of smokers was and non-smokers were 94 (87.85%); 13 (12.15%) respectively; which indicated that smoking may be a causative agent for lung cancer patients. Average pack years for lung cancer patients was observed 28.95 ± 12.14 as shown in figure 5.12 (b).

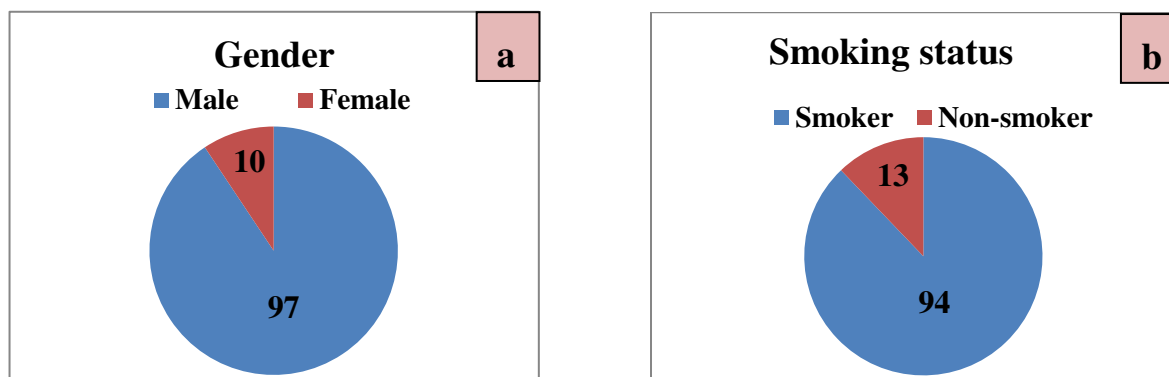
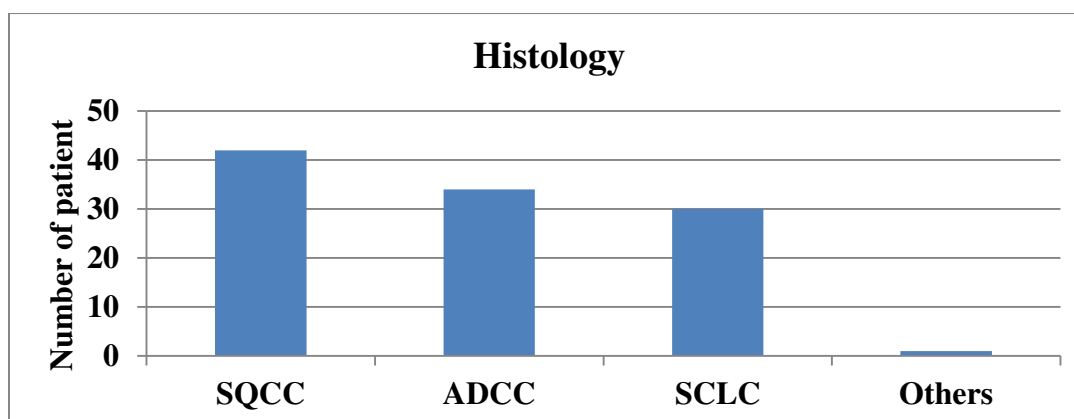


Figure 5.12 Frequencies of gender and smoking status in current study population to observe combined effect of *MGMT* and *ERCC1* gene polymorphism

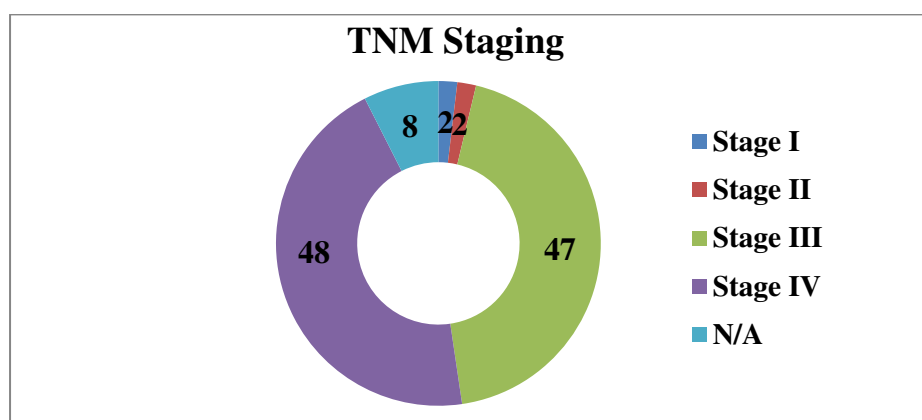
On the basis of histology, the most common type of lung cancer was SQCC; 42 (39.25%), then ADCC; 34 (31.77%) and SCLC; 30 (28.04%). 0.93% patients had vague histology as shown in figure 5.13.



(Abbreviation: ADCC- adenocarcinoma, SQCC- squamous cell carcinoma, SCLC- small cell lung carcinoma)

Figure 5.13 Frequency of histology in current study population to observe combined effect of *MGMT* and *ERCC1* gene polymorphism

There were more patients having stage IV [48 (44.86%)] and stage III [47 (43.92%)] in current study population, but number and percentage of stage I and Stage 2 was less *i.e* 2 (1.87%) and 2 (1.87%) respectively. It indicated that Stage III and Stage IV is more prominent than stage I and stage II as shown in figure 5.14.

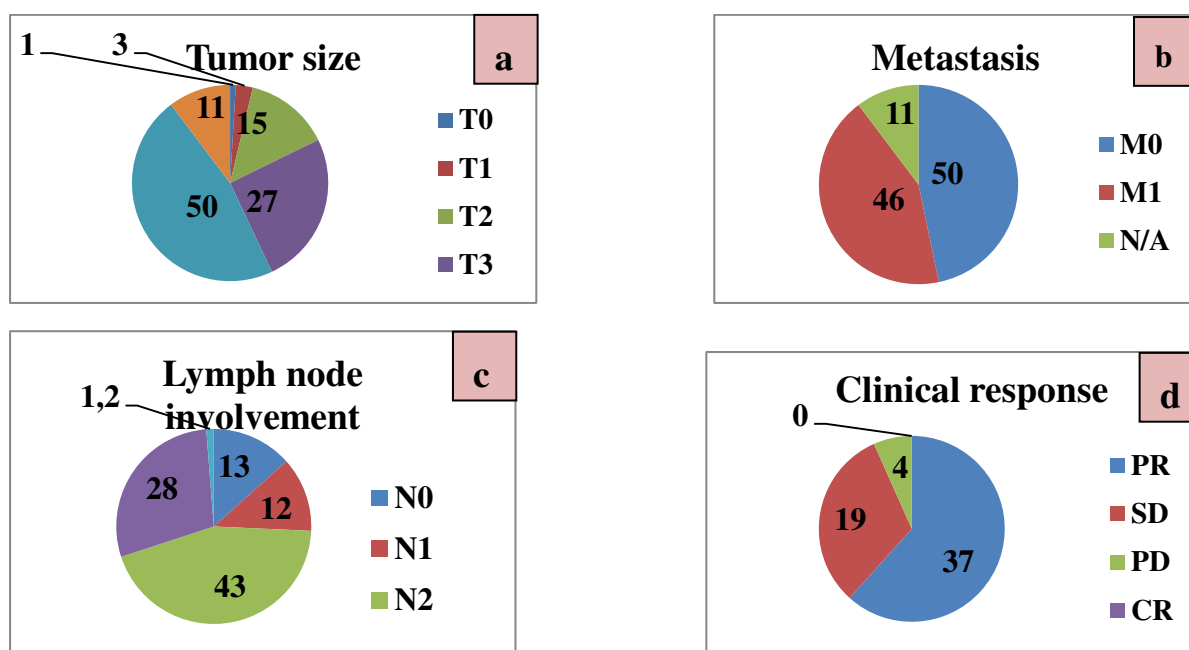


(Abbreviation: N/A-data not available)

Figure 5.14 Frequency of lung cancer stage in current study population to observe combined effect of *MGMT* and *ERCC1* gene polymorphism

In current study population range of tumor size varied from 0 to 4. Highest number and percentage of tumor size in lung cancer cases was of T4 type tumor size [50 (46.73 %)], then followed by T3 type [27 (25.23%)], T2 type [15 (14.02%)], T1 type [3 (2.80%)] and T0 type [1 (0.93%)]. There was no data available for 11 cancer patients. This study showed that T4 type of

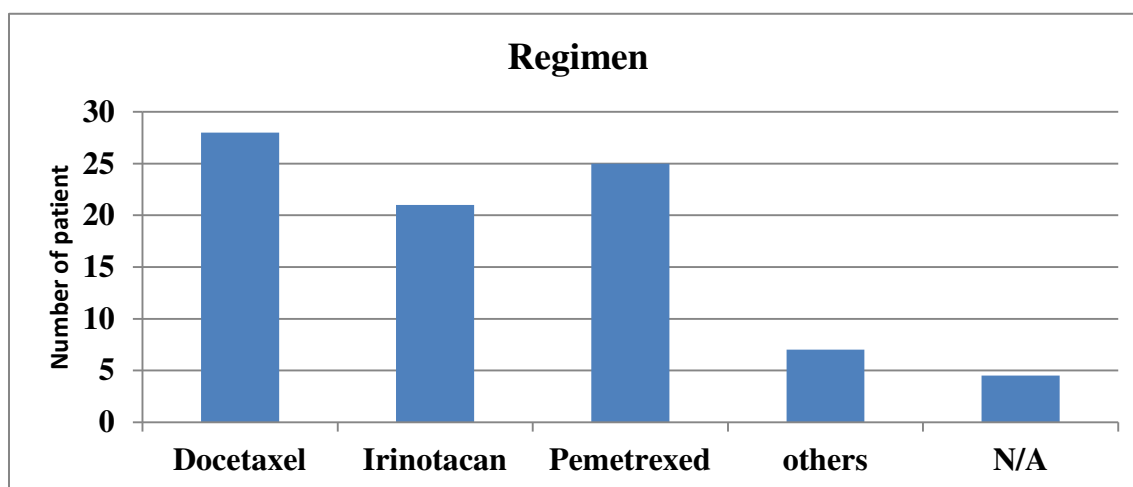
tumor size is dominative over T3, T2, T1 and T0 type as shown in figure 5.15 (a). Among lung cancer patients approx 46 (42.99%) cases showed distant metastasis, while 50 (46.73%) cases did not show any metastasis. Metastasis status for 11 (10.28%) cancer cases was not available in this study as shown in figure 5.15 (b). When lymph node involvement in lung cancer patients was studied, it was observed that patients showed highest involvement of N2 lymph node [43 (40.19%)] and no involvement of N4 lymph node. 26.17% patients showed N3 lymph node involvement, while 11.21% patients showed N0 and 12.15 % patients showed N1 lymph node involvement. 10.28% patients had ambiguous lymph node involvement as shown in figure 5.15 (c). Among 107 Lung cancer cases, data for only 60 cancer cases was available for clinical response study. A large number of cases showed partial response (PR) (61.67%), while none of the patients showed complete response (CR). 31.67% cancer patients had stable disease (SD), while 6.67% had progressive disease (PD) even after treated with chemotherapy as shown in figure 5.15 (d).



(Abbreviation: M0- no metastasis, M1- metastasis, CR- complete response, PR-partial response, SD- stable disease, PD- progressive disease, N/A- data not available)

Figure 5.15 Frequency of tumor size (a), metastasis (b), lymph node involvement (c) and clinical response (d) in current study population to observe combined effect of *MGMT* and *ERCC1* gene polymorphism

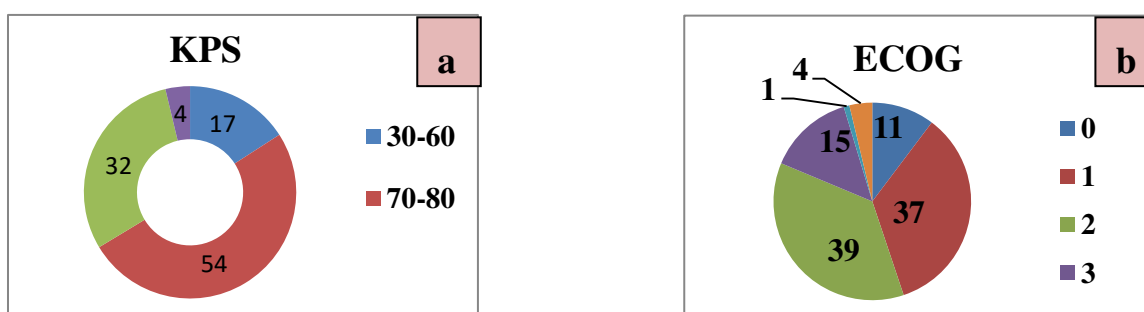
When chemotherapy regimen was observed, 26.17% patients were treated with Docetaxel + cis-platin/carbo-platin, 19.63% were administered Irinotecan + cis-platin/carbo-platin and 23.36% were treated with Pemetrexed + cis-platin/carbo-platin. 6.54% patients were administered other drugs, while regimen data was not available for 24.30% patients as shown in figure 5.16.



(Abbreviation: N/A- data not available)

Figure 5.16 Frequency of regimen in current study population to observe combined effect of *MGMT* and *ERCC1* gene polymorphism

When performance status of lung cancer patients was studied, it was observed that 15.89% cases had KPS range of 30-60, 50.47% had KPS size of 70-80 and 29.91% cases had KPS range between 90-100. KPS data was not available for 3.74% cancer cases as shown in figure 5.17 (a). ECOG value of 0, 1, 2, 3, 4 was shown by 10.28%, 34.58%, 36.45%, 14.02% and 0.93% cancer cases respectively. ECOG data was not available for 3.74% cancer cases as shown in figure 5.17 (b).



(Abbreviation: KPOS- karnofsky performance scale, ECOG- eastern cooperative oncology group, N/A- data not available)

Figure 5.17 Frequencies of KPS (a) and ECOG (b) in current study population to observe combined effect of *MGMT* and *ERCC1* gene polymorphism

5.2.4 Overall survival analysis by Kaplan-Meier and cox regression in lung cancer patients

Univariate Kaplan Maeir analysis was used for evaluation; then adjusted by multivariate cox hazard proportional analysis for various factors such as age, ECOG, gender, histology, KPS, heavy smoker, light smoker, smoking status, stage and considered the genotype as an independent factor. Patients having wild genotype were taken as reference to study overall survival period.

Table 5.3 Overall survival analysis by Kaplan-Meier and cox regression in lung cancer patients

Genotype	Overall Survival						
<i>MGMT</i> (rs12917)	Dead 179 (86.64%)	Alive 35 (16.35%)	Median OS (month)	HR (95% CI)	Log P	Adjusted HR 95% CI	P
CC	137 (76.54)	30 (85.71)	8.27	-	-	-	-
CT	42 (23.46)	5 (14.28)	6.23	1.20 (0.84-1.73)	0.29	1.44 (0.99- 2.09)	0.004
<i>ERCC1</i> (rs3212986)							
CC	92 (51.40)	15 (42.86)	8.30	-	-	-	-
CA	70 (39.11)	17 (48.57)	8.20	0.92 (0.68-1.26)	0.60	1.25 (0.86 -1.82)	0.005
AA	17 (9.50)	3 (8.57)	5.07	1.20 (0.69-2.10)	0.47	1.40 (1.04 -1.87)	0.003
CA+AA	87 (48.61)	20 (57.14)	7.13	0.97 (0.72-1.29)	0.81	1.25 (0.91 -1.73)	0.007
<i>MGMT</i> + <i>ERCC1</i>	91 (85.05)	16 (14.95)					
CC	72 (79.12)	13 (81.25)	9.33	-	-	-	-
CT+CA	19 (20.88)	3 (18.75)	6.33	1.15 (0.68-1.94)	0.60	1.52 (0.86- 2.69)	0.02

Adjusted Hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, gender, genotype, histology, KPS, smoker, smoking status and stage. **Abbreviations:** HR-hazard ratio; CI-confidence interval.

Results for overall survival analysis are summarised in table 5.3. After evaluation of 214 patients from day of diagnosis of disease to last follow-up, it was observed that 179 (86.64%)

patients died and 35 (16.35%) patients were alive. On the basis of genotype, for *MGMT* it was observed that among patients who had expired (179), 76.54% (137) patients had wild type genotype while, 23.46% (42) patients had heterozygous type genotype, and among 35 patients who survived, 85.71% (30) patients had wild type genotype while 14.28% (5) patients had heterozygous type genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype and when univariate analysis was evaluated, a non-significant association was observed between *MGMT* polymorphism and lung cancer patients (MST= 6.23 vs 8.27 months; HR- 1.20; log rank $p=0.29$). When multivariate cox model was analysed a significant relationship was found between *MGMT* genotype and overall survival in lung cancer patients (**adjusted HR=1.44; 95% CI=0.99-2.09; $p=0.004$**) as shown in figure 5.18.

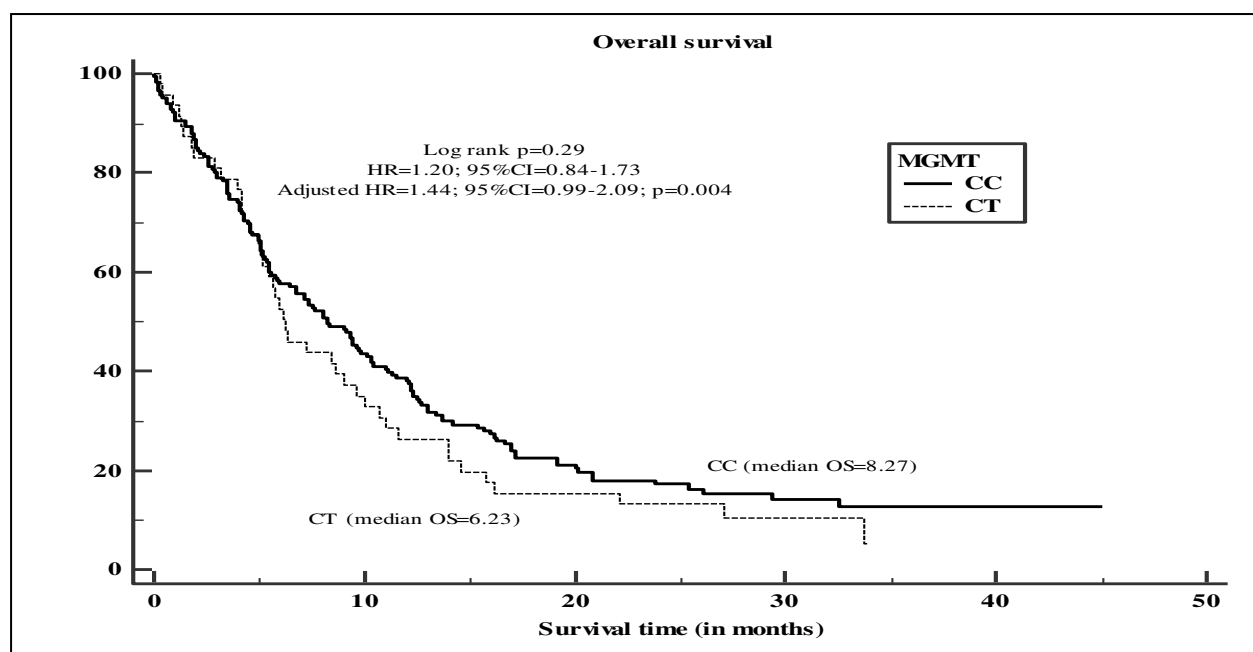


Figure 5.18 Effect of *MGMT* (rs12917) polymorphism on overall survival in lung cancer patients (CC vs. CT)

For *ERCC1* (rs3212986) it was observed that out of 179 patients who were dead, 51.40% (92) patients had wild type genotype, 39.11% (70) patients had heterozygous type genotype, 9.50% (17) patients had mutant type genotype and out of 35 patients who were alive after the study, 42.86% (15) patients had wild type genotype, 48.57% (17) patients had heterozygous type genotype and 8.57% (3) patients had mutant type genotype. Overall survival period of heterozygous type genotype was observed similar to that of wild type genotype (MST= 8.2 vs.

8.3 months; HR=0.92; log rank $p=0.60$). When multivariate cox model was analysed a significant relationship was observed between *MGMT* polymorphism and overall survival in lung cancer patients having heterozygous type genotype (**adjusted HR=1.25; 95% CI=0.86-1.82; $p=0.005$**) as shown in figure 5.19.

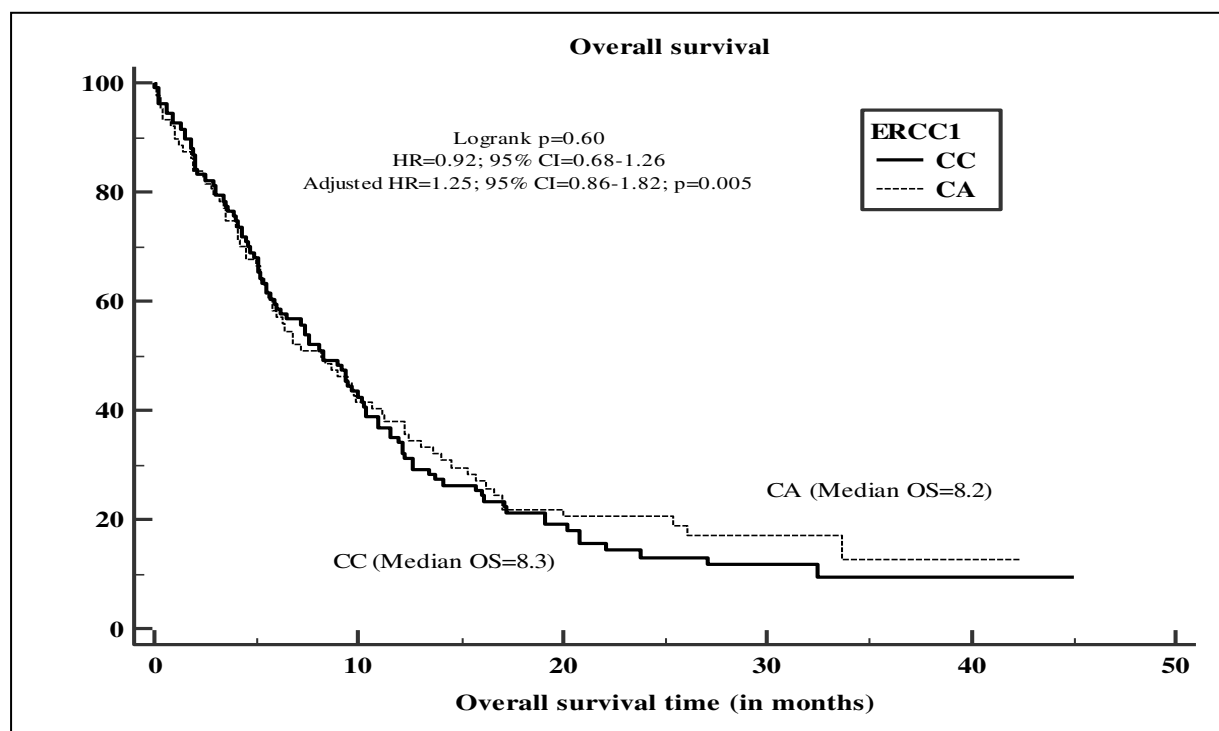


Figure 5.19 Effect of *ERCC1* (rs3212986) polymorphism on overall survival in lung cancer patients (CC vs. CA)

Overall survival period of mutant type genotype was much lower than wild type genotype (MST= 5.07 vs. 8.3 months; HR- 1.20; log rank $p=0.60$) and on applying cox multivariate hazard proportional analysis, a significant relationship was observed in patients having mutant type genotype (**adjusted HR=1.40; 95% CI=1.04-1.87; $p=0.003$**) as shown in figure 5.20. On combining the patients having heterozygous type genotype and patients having mutant type genotype and comparing it with patients having wild type genotype, overall survival period was less as compared to wild type genotype (MST= 7.13 vs. 8.3 months; HR=0.97; log rank $p=0.81$) and a significant association was also observed between *MGMT* polymorphism and lung cancer patients (**adjusted HR=1.25; 95% CI=0.91-1.73; $p=0.007$**) as shown in figure 5.21.

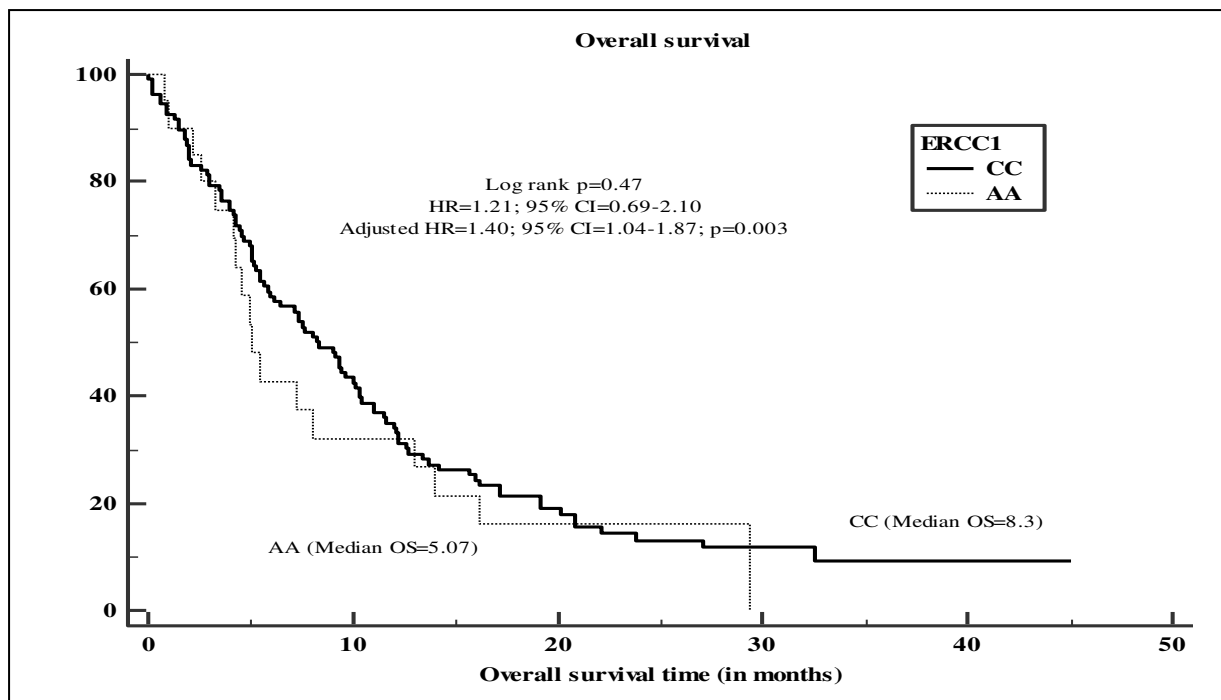


Figure 5.20 Effect of *ERCC1* (rs3212986) polymorphism on overall survival in lung cancer patients (CC vs. AA)

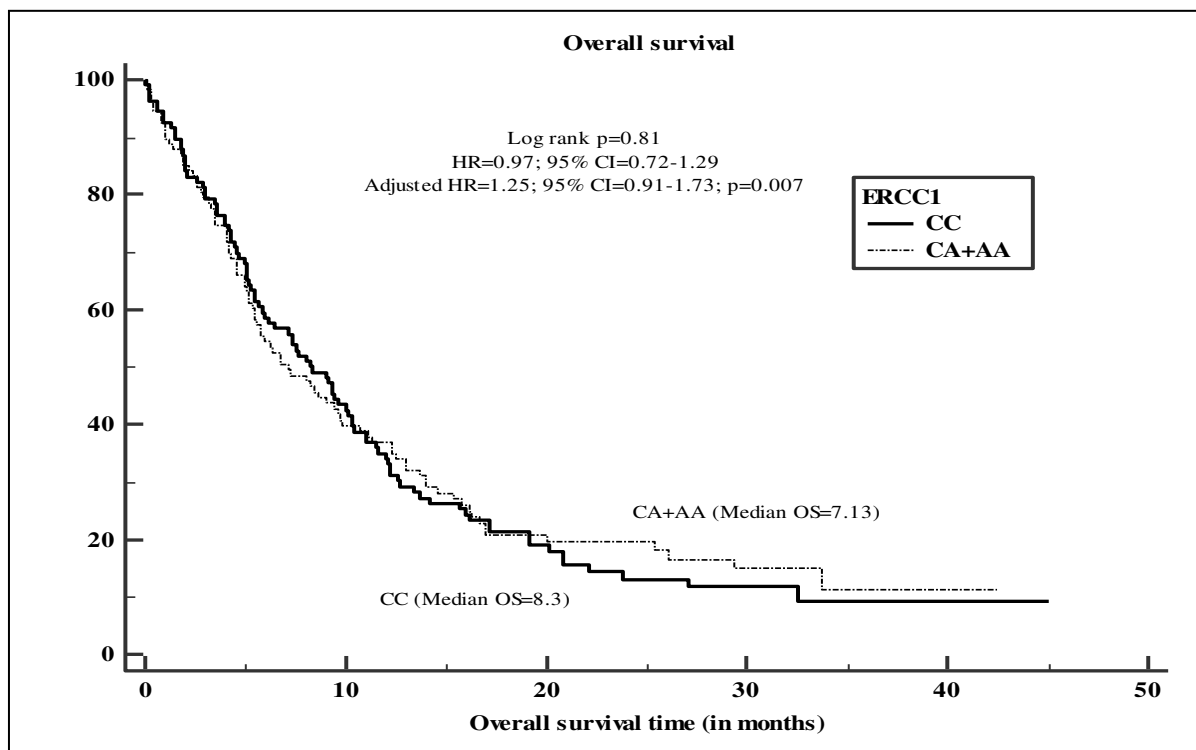


Figure 5.21 Effect of *ERCC1* (rs3212986) polymorphism on overall survival in lung cancer patients (CC vs. CA+AA)

A total of 107 lung cancer cases were analysed to assess the overall survival and the results are summarised in table 5.3. After evaluation of 107 patients from day of diagnosis of disease to last follow-up, it was observed that 91 (85.05%) patients had died and 16 (14.95%) patients were alive. Among patients having common *MGMT* and *ERCC1* genotype, it was observed that 91 subjects were not alive, 79.12% (72) patients had wild type genotype while 20.88% (19) patients had heterozygous type genotype, whereas among alive patients (16), 81.25% (13) patients had wild type genotype while 18.75% (3) patients had heterozygous genotype. Overall survival period of heterozygous type genotype was much lower as compared to wild type genotype (MST= 6.330 vs. 9.330 months; HR- 1.15; p-value- 0.60). It indicated that subjects carrying heterozygous type genotype for both the SNP's; has high risk of death as compared to wild type genotype. When univariate analysis was evaluated, a non-significant association was observed between overall survival and lung cancer patients having common *MGMT* and *ERCC1* genotype whereas multivariate cox model showed a significant association between overall survival and lung cancer patients having common *MGMT* and *ERCC1* genotype (**adjusted HR=1.52; 95% CI=0.86-2.68; p=0.02**) as shown in figure 5.22.

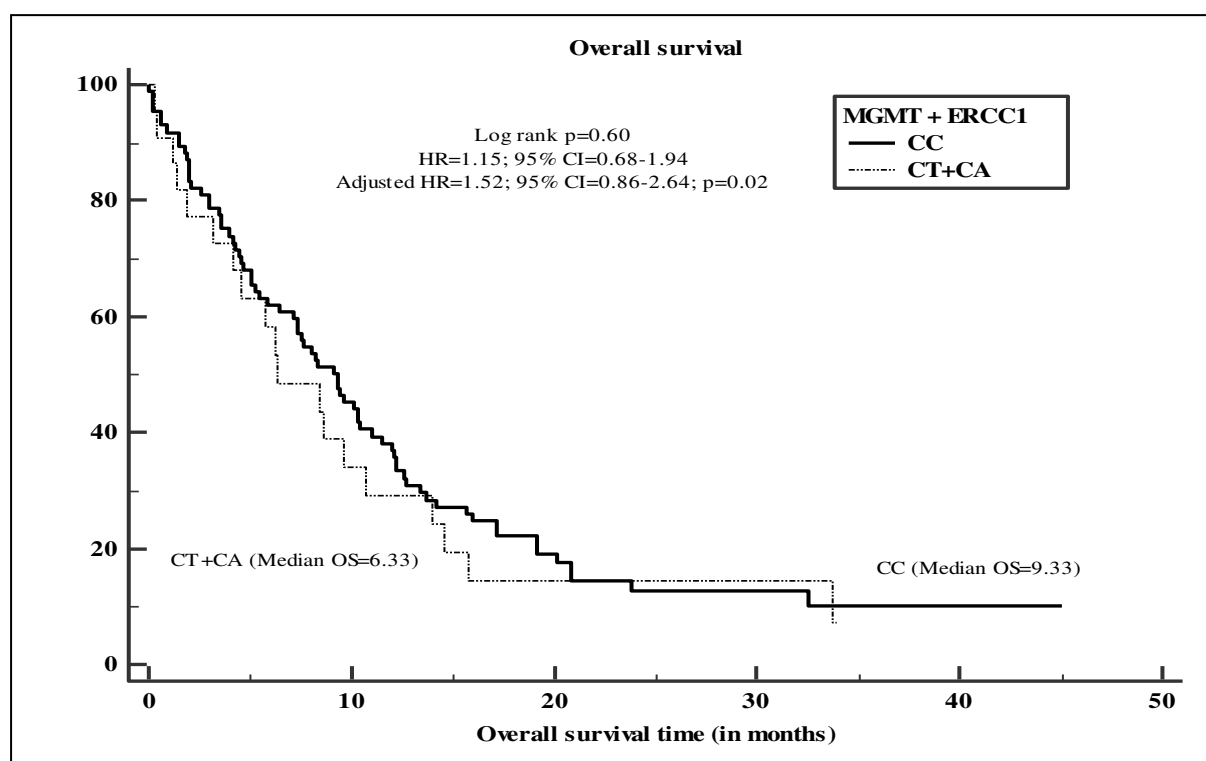


Figure 5.22 Combinatorial effect of *MGMT* (rs12917) and *ERCC1* (rs3212986) polymorphism on overall survival in lung cancer patients (CC vs. CT+CA)

5.2.5 Overall survival analysis by Kaplan Meier and cox regression in lung cancer patients on the basis of histology

5.2.5.1 Overall survival analysis among SCLC cases

Table 5.4 Overall survival analysis on the basis of SCLC							
Genotype	SCLC		Overall survival				
<i>MGMT</i> (rs12917)	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P- value
	55 (91.67)	5 (8.33)					
CC	46 (83.64)	5 (100)	7.33	-	-	-	-
CT	9 (16.36)	0 (0)	7.23	1.05 (0.51-2.18)	0.88	1.81 (0.82 3.99)	0.06
<i>ERCC1</i> (rs3212986)							
CC	26 (53.06)	3 (60)	9.67	-	-	-	-
CA	23 (46.94)	2 (40)	8.40	1.05 (0.60-1.85)	0.86	1.15 (0.61-2.17)	0.18
AA	6 (12.24)	0 (0)	7.23	1.11 (0.44-2.76)	0.82	1.30 (0.71-2.38)	0.39
CA+AA	29 (59.18)	2 (40)	8.40	1.06 (0.63-1.81)	0.82	1.27 (0.68-2.35)	0.09
<i>MGMT</i> + <i>ERCC1</i>	27 (90)	3 (10)					
CC	23 (85.18)	3 (100)	9.67	-	-	-	-
CT+CA	4 (14.81)	0 (0)	5.80	1.47 (0.43-4.99)	0.47	3.44 (0.74-16.1)	0.48
Adjusted Hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, gender, genotype, KPS, smoker and stage. Abbreviations: HR-hazard ratio; CI-confidence interval.							

After evaluation of 60 patients with SCLC (results are summarised in table 5.4) from day of diagnosis of disease to last follow-up, it was found that 55 (91.67%) patients had died and 5 (8.33%) patients were alive. For *MGMT* it was observed that among patients who had died (55), 83.64% (46) patients had wild type genotype whereas 16.36% (9) patients had heterozygous type genotype, and among patients who were alive after the study (5), all patients had wild type genotype. Overall survival period of heterozygous type genotype was slightly lower than that of wild type genotype (MST= 7.23 vs. 7.33 months; HR- 1.05; p-value- 0.88). A non-significant

relationship was observed between *MGMT* polymorphism and overall survival in SCLC patients after evaluated by univariate and multivariate cox regression analysis (adjusted HR=1.81; 95% CI=0.82-3.99; p=0.06) as shown in figure 5.23.

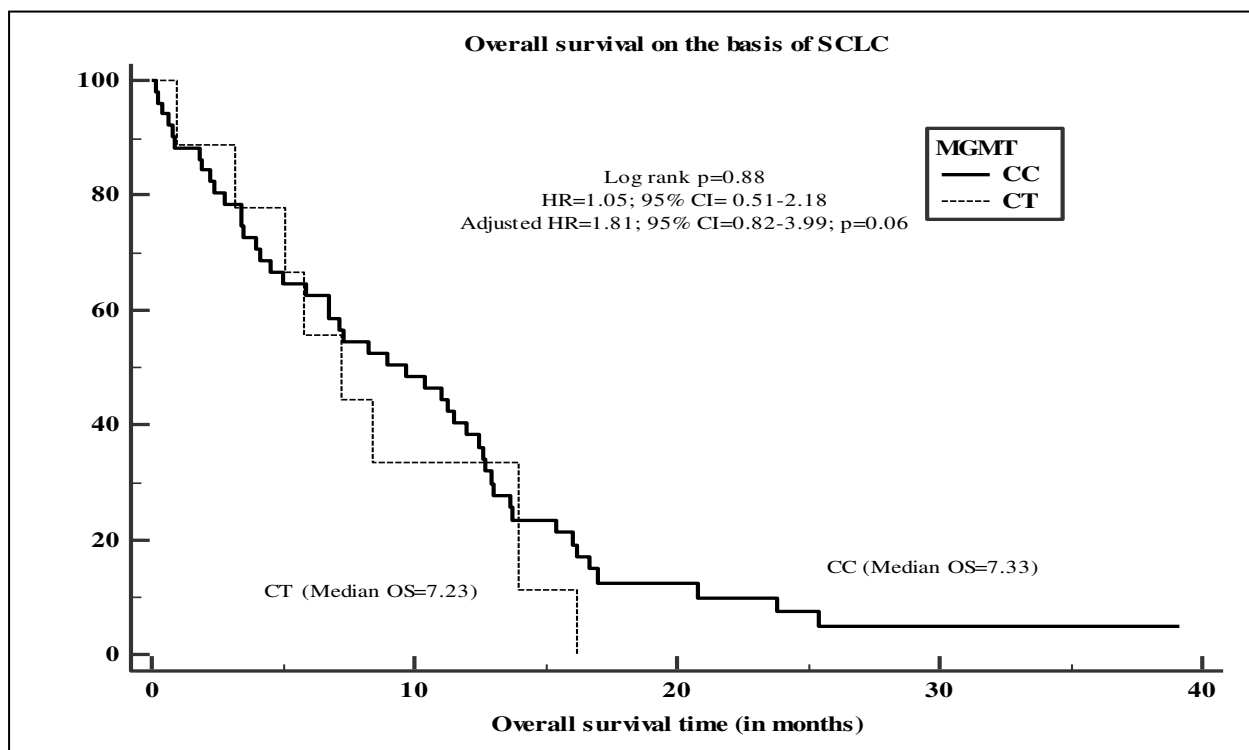


Figure 5.23 Effect of *MGMT* gene polymorphism on overall survival in SCLC patients (CC vs. CT)

For *ERCC1* it was observed that among patients, who had not survived (55), 53.06% (26) patients had wild type genotype, 46.94% (23) patients had heterozygous type genotype and 12.24% (6) patients had mutant type genotype. Among patients who were alive (5), 60% (3) patients had wild type genotype, 40% (2) patients had heterozygous type genotype, none of the patients had mutant genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype (MST= 8.4 vs. 9.67 months; HR=1.05; p=0.85). After evaluated by univariate and multivariate cox model, a non-significant relationship was observed between *ERCC1* polymorphism and overall survival in SCLC patients (adjusted HR=1.15; 95% CI=0.61-2.17; p=0.18) as shown in figure 5.24. Overall survival period of mutant type genotype was also observed much lower as compared to wild type genotype (MST= 7.23 vs. 9.67 months; HR=1.11; p=0.82). After evaluated by univariate and multivariate cox model, a non-significant

relationship was also observed between *ERCC1* polymorphism and overall survival in SCLC patients (adjusted HR=1.30; 95% CI=0.71-2.38; p=0.39) as shown in figure 5.25.

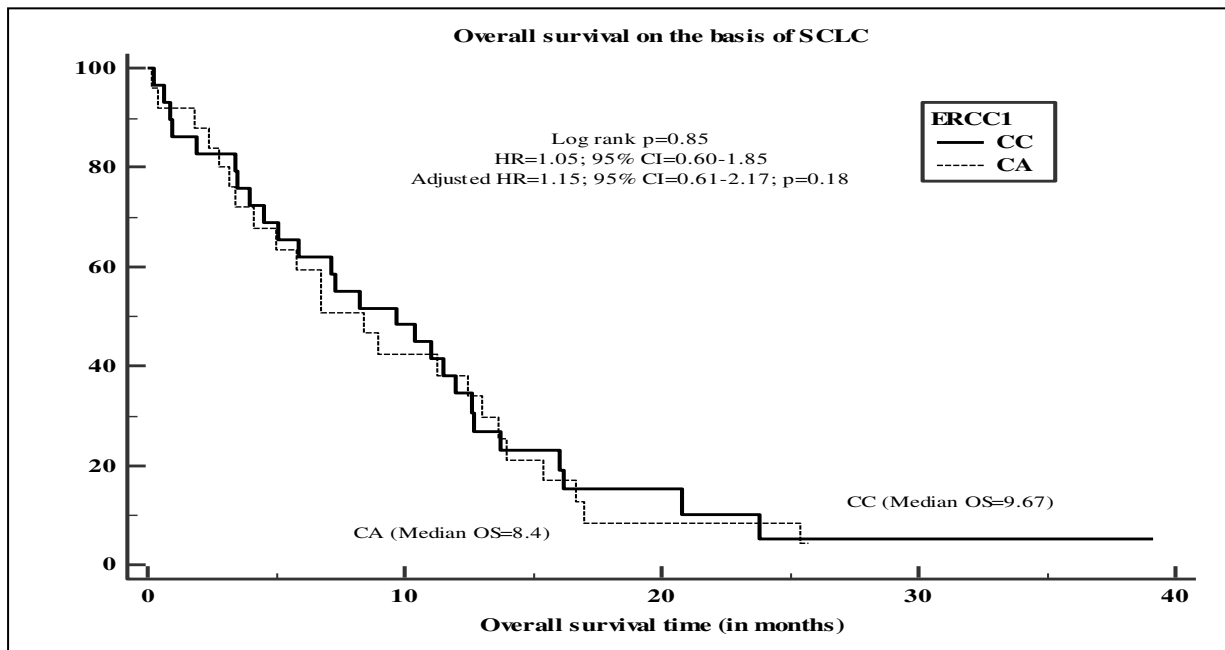


Figure 5.24 Effect of *ERCC1* (rs3212986) gene polymorphism on overall survival in SCLC patients (CC vs. CA)

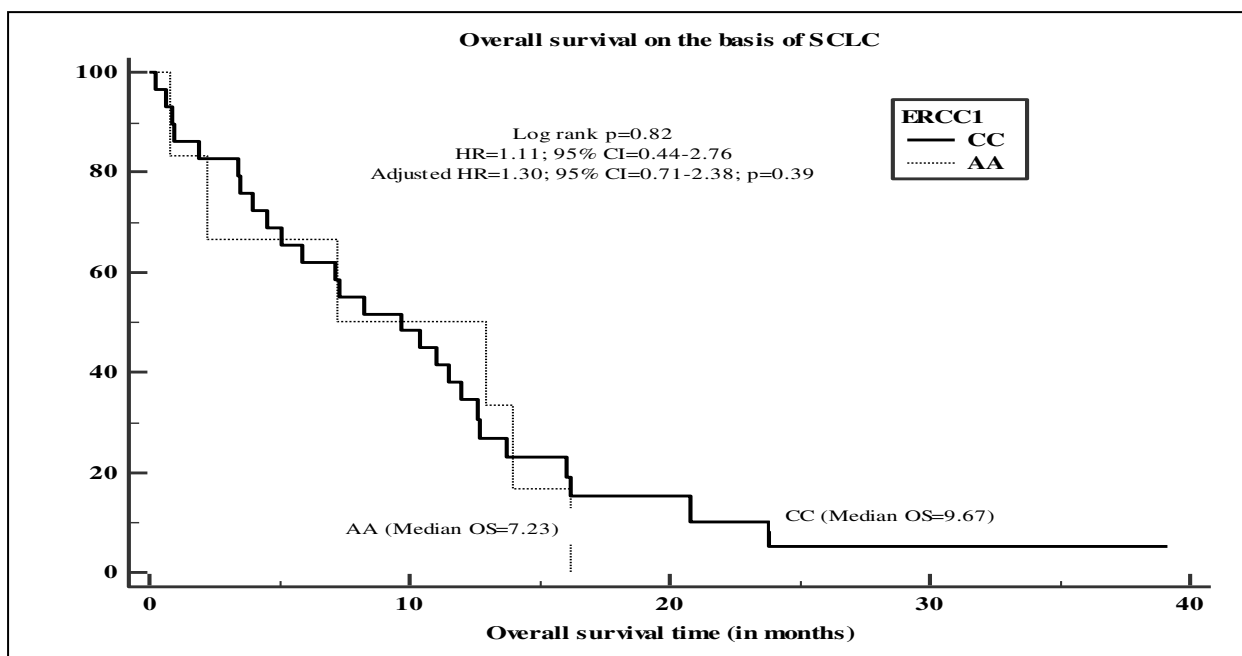


Figure 5.25 Effect of *ERCC1* (rs3212986) gene polymorphism on overall survival in SCLC patients (CC vs. AA)

Since the frequency of mutant type genotype was less, we combined this with heterozygous type genotype and compared against wild type genotype. It was observed that overall survival period was less as compared to wild type genotype (MST= 8.4 vs. 9.67 months; HR=1.06; p=0.82). A non-significant relationship was observed between *ERCC1* polymorphism and overall survival in SCLC patients after evaluated by univariate and multivariate cox regression model (adjusted HR=1.27; 95% CI=0.68-2.35; p=0.09) as shown in figure 5.26.

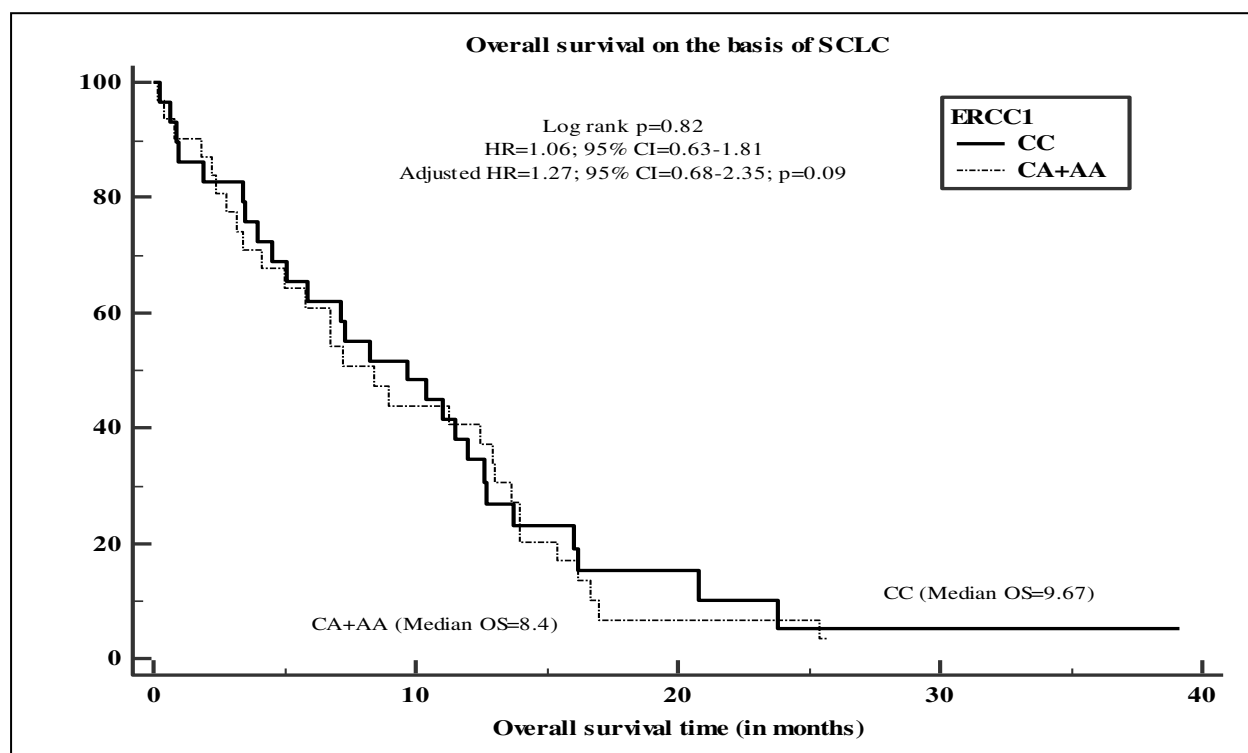


Figure 5.26 Effect of *ERCC1* (rs3212986) gene polymorphism on overall survival in SCLC patients (CC vs. CA+AA)

After evaluation of 30 patients with SCLC having common *MGMT* and *ERCC1* genotype (results are summarised in table 5.4), it was found that 27 (90%) patients did not survive in study period and only 3 (10%) patients were alive. From the current study it was observed that out of 27 patients who had died, 85.18% (23) patients had wild type genotype while 14.81% (4) patients had heterozygous type genotype and among alive patients (3), all patients had wild type genotype. Overall survival period of heterozygous type genotype was much lower than wild type genotype (MST= 5.80 vs. 9.67 months; HR- 1.47; p-value- 0.47). It indicated that death rate was more in heterozygous type genotype. After evaluated by univariate and multivariate cox model, a non-significant relationship was observed between overall survival and patients

with SCLC having common *MGMT* and *ERCC1* genotype (adjusted HR=3.44; 95% CI=0.74-16.1; p=0.48) as shown in figure 5.27.

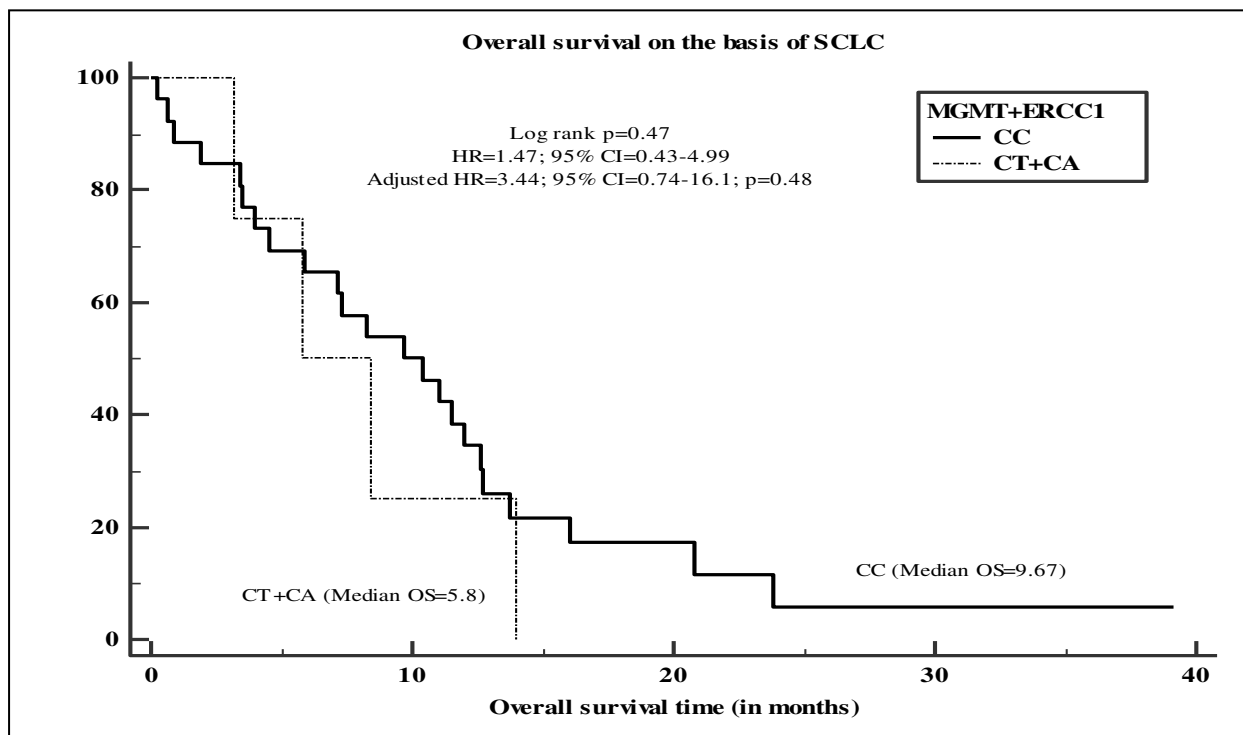


Figure 5.27 Combine effect of *MGMT* and *ERCC1* gene polymorphism on overall survival in SCLC patients (CC vs. CT+CA)

5.2.5.2 Overall survival analysis in ADCC patients

Table 5.5 Overall survival analysis on the basis of ADCC							
Genotype	ADCC		Overall survival				
<i>MGMT</i> (rs12917)	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P- value
CC	57 (82.61)	12 (17.39)	7.60	-	-	-	-
CT	40 (70.17)	9 (75)	5.93	1.13 (0.63-2.02)	0.68	1.56 (0.83-2.94)	0.02
<i>ERCC1</i> (rs3212986)							
CC	34 (64.15)	4 (36.36)	7.57	-	-	-	-
CA	19 (35.85)	7 (63.64)	7.13	0.67 (0.39-1.15)	0.15	0.66 (0.34-1.27)	0.06

AA	4 (7.55)	1 (9.09)	4.97	1.23 (0.40-3.77)	0.70	1.25 (0.65-2.46)	0.01
CA+AA	23 (43.40)	8 (72.73)	6.33	0.74 (0.44-1.24)	0.25	0.78 (0.44-1.40)	0.03
MGMT + ERCC1	31 (91.18)	3 (8.82)					
CC	24 (77.42)	2 (66.67)	7.6				
CT+CA	7 (22.58)	1 (33.33)	4.53	1.00 (0.43-2.33)	1.0	1.51 (0.54-4.28)	0.18
Adjusted Hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, gender, genotype, KPS, smoker and stage. Abbreviations: HR- hazard ratio; CI -confidence interval.							

A total of 69 ADCC patients were followed from day of diagnosis of disease to last follow-up, it was observed that 57 (82.61%) patients did not survive whereas 12 (17.39%) patients were alive as shown in table 5.5. In the case of *MGMT* gene polymorphism, out of 57 ADCC subjects who were dead, 70.17% (40) ADCC patients had wild type genotype while 29.82% (17) patients had heterozygous type genotype, and among 12 ADCC patients who were alive, 75% (9) patients had wild type genotype while 25% (3) patient had heterozygous type genotype. Overall survival period of heterozygous type genotype was observed much lower than wild type genotype (MST= 5.93 vs. 7.6 months; HR- 1.13; p-value- 0.68). It indicated that in heterozygous type genotype has more risk of death rate as compared to wild type genotype. When univariate model was analysed, a non-significant association was observed between *MGMT* polymorphism and overall survival in ADCC patients. But multivariate cox hazard proportional analysis showed a significant relationship between *MGMT* polymorphism and overall survival in ADCC patients (adjusted HR=1.56; 95% CI=0.83-2.94; **p=0.02**) as shown in figure 5.28.

For *ERCC1* gene polymorphism it was observed that among patients who did not survive in study period (57), 64.15% (34) patients had wild type genotype, 35.85% (19) patients had heterozygous type genotype and 7.55% (4) patients had mutant type genotype. Among patients who were alive (12), 36.36% (4) patients had wild type genotype, 63.64% (7) patients had heterozygous type genotype and only 9.09% (1) patient had mutant type genotype as shown in table 5.5. Overall survival period of heterozygous type genotype was observed slightly lower than wild type genotype (MST= 7.13 vs. 7.57 months; HR= 0.67; p=0.15). Analysis of univariate model and multivariate cox model state a non-significant relationship between

ERCC1 gene polymorphism and overall survival in ADCC patients (adjusted HR= 0.066; 95% CI= 0.34-1.27; p= 0.06) as shown in figure 5.29.

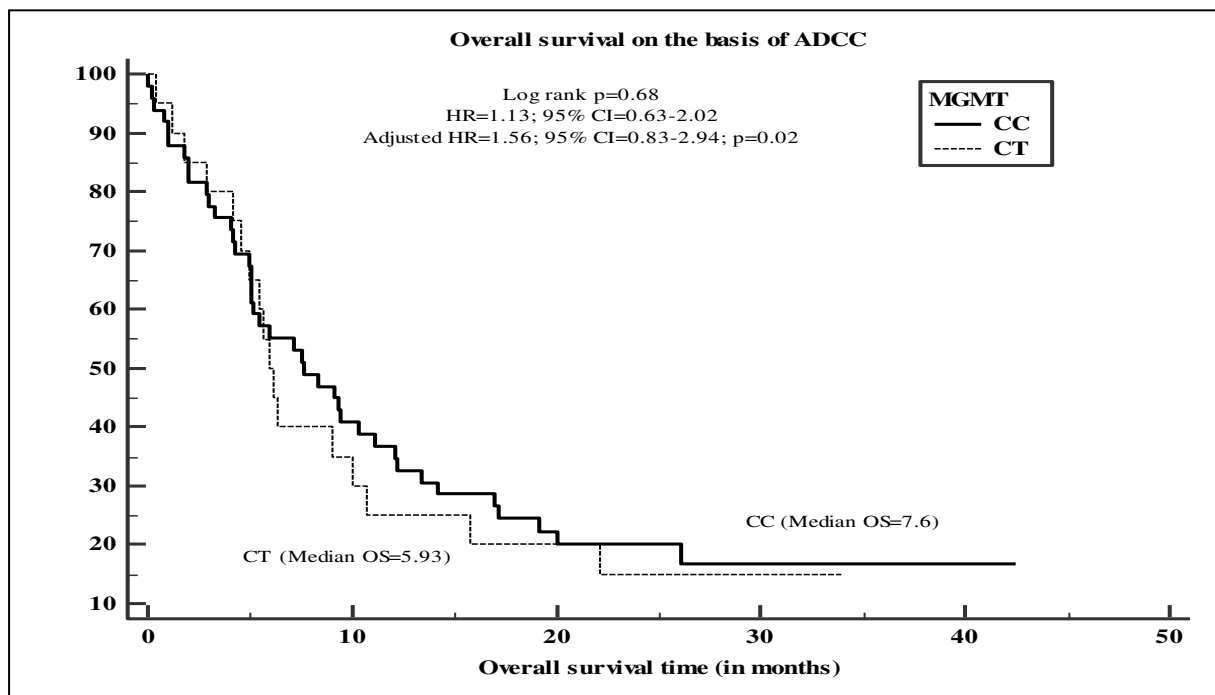


Figure 5.28 Effect of *MGMT* gene polymorphism on overall survival in ADCC patients (CC vs. CT)

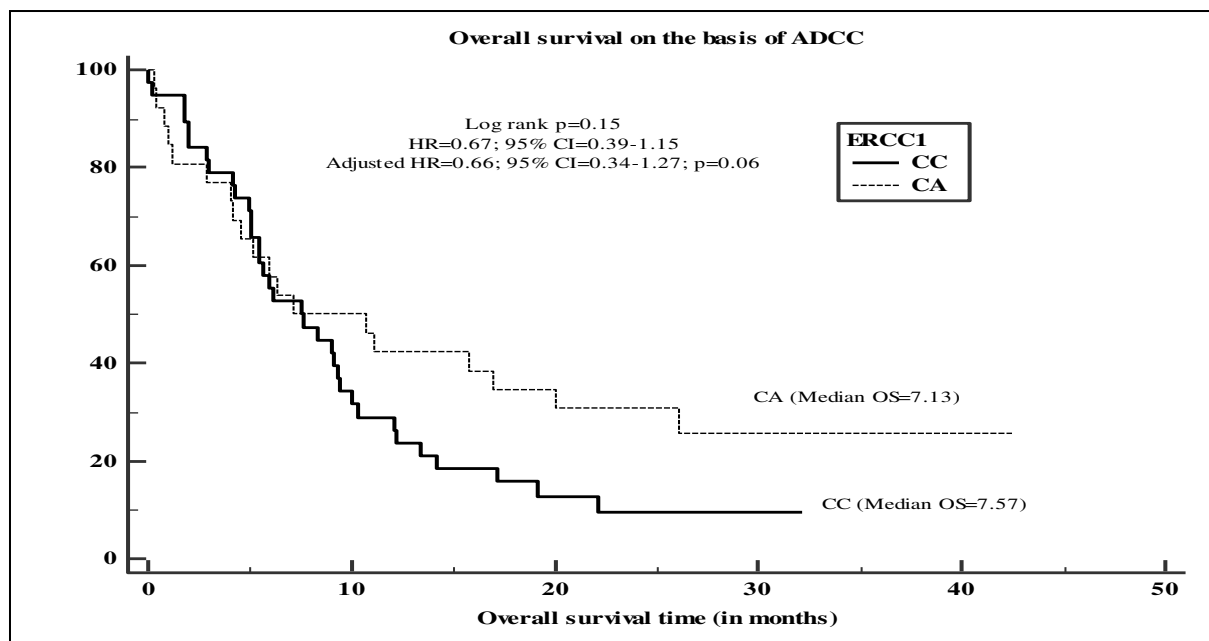


Figure 5.29 Effect of *ERCC1* gene polymorphism on overall survival in ADCC patients (CC vs. CA)

Overall survival period of mutant type genotype was observed much lower than that of wild type genotype (MST= 4.97 vs. 7.57 months; HR=1.23; $p=0.70$). When univariate model was analysed, a non-significant relationship was found between *ERCC1* polymorphism and overall survival in ADCC patients; but multivariate cox regression model showed a significant relationship between *ERCC1* polymorphism and overall survival in ADCC patients (adjusted HR= 1.25; 95% CI=0.65-2.46; $p=0.01$) as shown in figure 5.30.

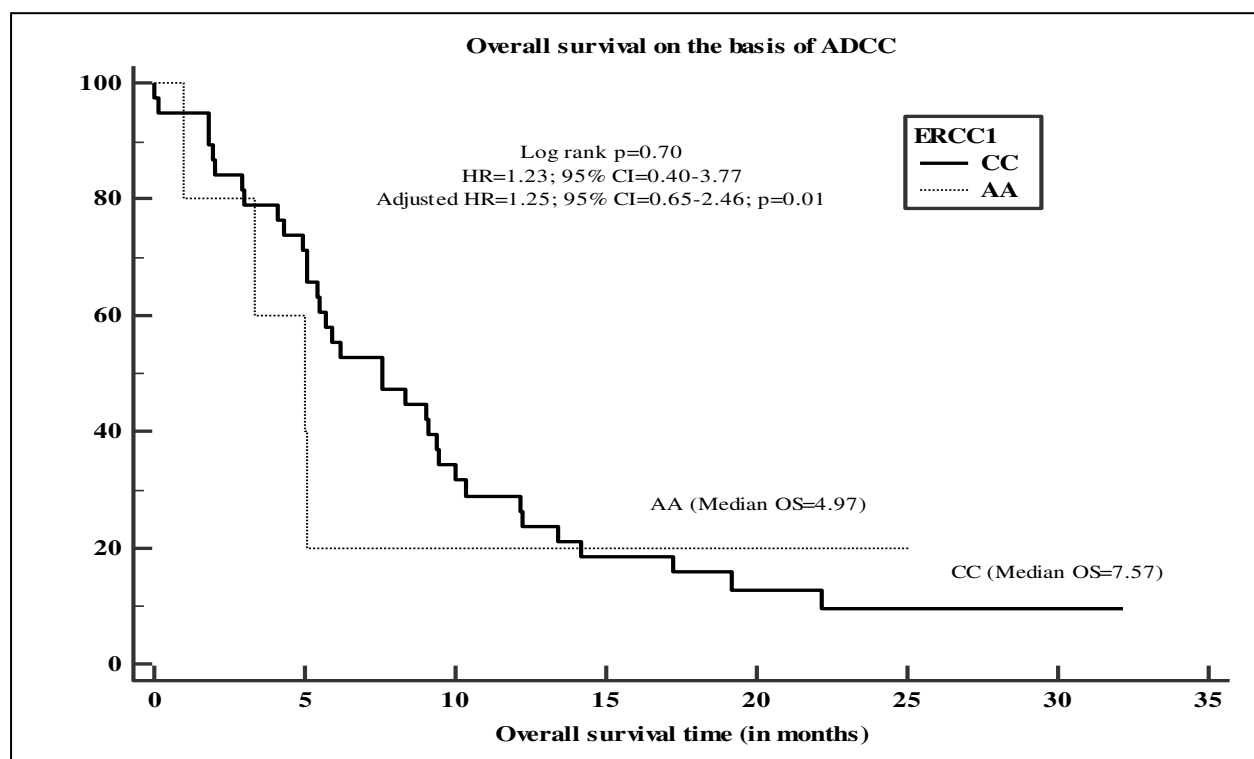


Figure 5.30 Effect of *ERCC1* gene polymorphism on overall survival in ADCC patients (CC vs. AA)

Since the frequency of type mutant genotype was less, so we combined it with the heterozygous type genotype and on comparing this against wild type genotype; overall survival period in this case was observed less as compared to wild type genotype (MST= 6.33 vs. 7.57 months; HR= 0.74; $p= 0.25$). A non-significant relationship was observed between *ERCC1* polymorphism and overall survival in ADCC patients after evaluated by univariate model. But multivariate cox model analysis stated a significant relationship between *ERCC1* polymorphism and overall survival in ADCC patients (adjusted HR= 0.78; 95% CI=0.44-1.40; $p=0.03$) as shown in figure 5.31.

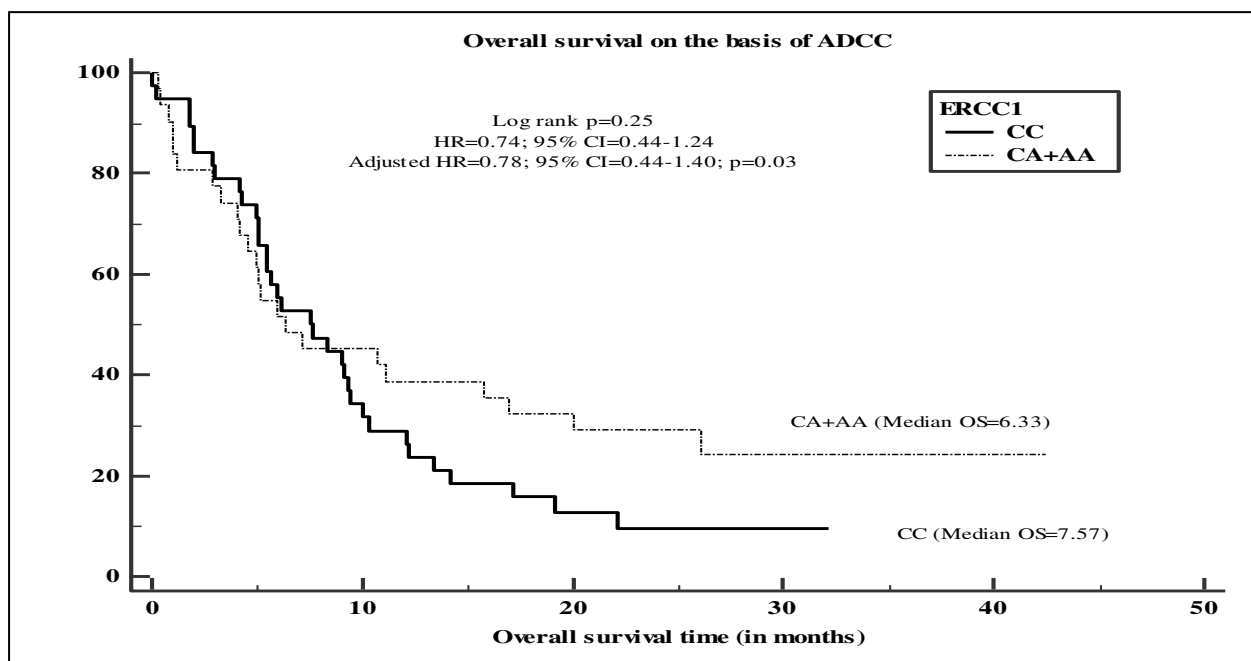


Figure 5.31 Effect of *ERCC1* gene polymorphism on overall survival in ADCC patients (CC vs. CA+AA)

After evaluation of 34 ADCC patients having common *MGMT* and *ERCC1* genotype (results are summarised in table 5.5), it was observed that 31 (91.18%) patients did not survive in study period and only 3 (8.82%) patients were alive. From the current study it was observed that out of 31 patients who had died, 77.42% (24) patients had wild type genotype while 22.58% (7) patients had heterozygous type genotype and among alive patients (3), 66.67% (2) patients had wild type genotype whereas only 33.33% (1) patient had heterozygous type genotype. Overall survival period of heterozygous type genotype was much lower than wild type genotype (MST= 4.53 vs. 7.6 months; HR- 1; p-value- 1). It indicated that death rate was more in heterozygous type genotype. After evaluated by univariate and multivariate cox model, a non-significant relationship was observed between overall survival and ADCC patients having common *MGMT* and *ERCC1* genotype (adjusted $HR=1.51$; 95% $CI=0.54-4.28$; $p=0.18$) as shown in figure 5.32.

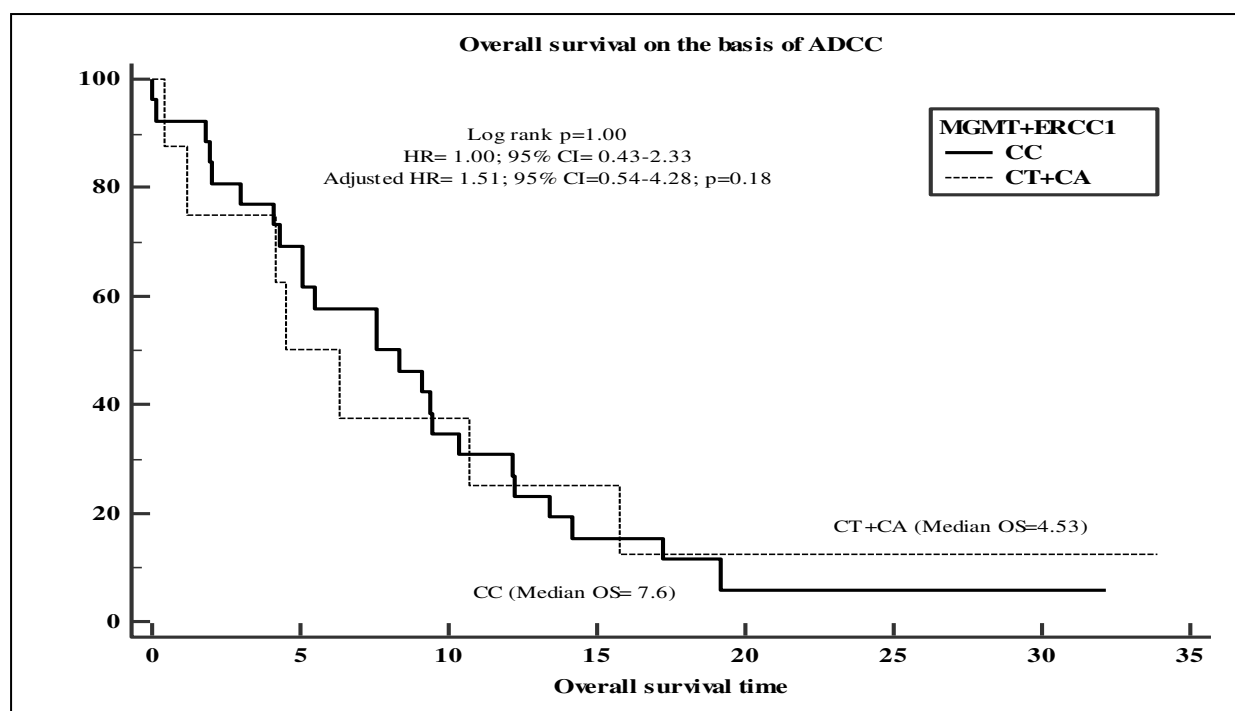


Figure 5.32 Combinatorial effect of *MGMT* and *ERCC1* gene polymorphism on overall survival in ADCC patients (CC vs. CT+CA)

5.2.5.3 Overall survival of analysis on the basis of SQCC

Table 5.6 Overall survival analysis on the basis of SQCC							
Genotype	SQCC		Overall survival				
<i>MGMT</i> (rs12917)	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P- value
CC	64 (80)	16 (20)	8.03	-	-	-	-
CT	49 (76.56)	14 (87.5)	8.63	1.23 (0.67-2.26)	0.49	1.99 (0.98-4.03)	0.06
<i>ERCC1</i> (rs3212986)							
CC	30 (52.63)	8 (57.14)	10.13	-	-	-	-
CA	27 (47.37)	6 (42.86)	6.23	1.30 (0.77-2.21)	0.31	2.55 (1.30-5.02)	0.02
AA	7 (12.28)	2 (14.28)	5.43	1.36 (0.55-3.39)	0.46	1.44 (0.89-2.33)	0.24
CA+AA	34 (59.65)	8 (57.14)	5.77	1.31 (0.80-2.14)	0.28	2.33 (1.26 4.34)	0.01

<i>MGMT</i> + <i>ERCC1</i>	32 (76.19)	10 (23.81)					
CC	24 (75)	8 (80)	10.13	-	-	-	-
CT + CA	8 (25)	2 (20)	8.63	1.27 (0.54-2.97)	0.56	2.69 (0.92-7.87)	0.25
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, gender, genotype, KPS, smoker and stage. Abbreviations: HR -hazard ratio; CI -confidence intervals.							

A total of 80 SQCC patients were followed from day of diagnosis of disease to last follow-up, it was observed that 64 (80%) patients did not survive in study period and 16 (20%) patients were alive as shown in table 5.6. For *MGMT* it was observed that among patients who did not survive (64), 76.56% (49) patients had wild type genotype whereas 23.44% (15) patients had heterozygous type genotype. Among patients who survived during study period (16), 87.5% (14) patients had wild type genotype whereas 12.5% (2) patient had heterozygous type genotype. Overall survival period of heterozygous type genotype was more as compared to wild type genotype (MST= 8.63 vs. 8.03 months; HR=1.23; p-value= 0.49). It indicated that in wild type genotype there was high risk of death as compared to heterozygous type genotype as shown in table 5.6. For multivariate cox hazard proportional analysis, values for HR and P were 1.9909 and 0.0551 respectively. When both univariate (Kaplan) and multivariate cox regression model was analysed a non-significant relationship was observed between *MGMT* genotype and overall survival in SQCC patients (adjusted HR=1.99; 95% CI=0.98-4.03; p=0.06) as shown in figure 5.33.

For *ERCC1* it was observed that among patients who were dead (64), 52.63% (30) patients had wild type genotype, 47.37% (27) patients had heterozygous type genotype and 12.28% (7) patients had mutant type genotype. Among alive patients (20), 57.14% (8) patients had wild type genotype, 42.86% (6) patients had heterozygous type genotype and 14.28% (2) patients had mutant type genotype as shown in table 5.6. Overall survival period of heterozygous type genotype was much lower as compared to wild type genotype (MST= 6.23 vs. 10.13 months; HR=1.30; p=0.31). A non-significant association was observed between *ERCC1* polymorphism and overall survival in SQCC patients after evaluated by univariate model. But when multivariate cox regression was analysed a significant relationship was observed between

ERCC1 polymorphism and overall survival in SQCC patients (adjusted HR=2.55; 95% CI=1.30-5.02; **p=0.02**) as shown in figure 5.34.

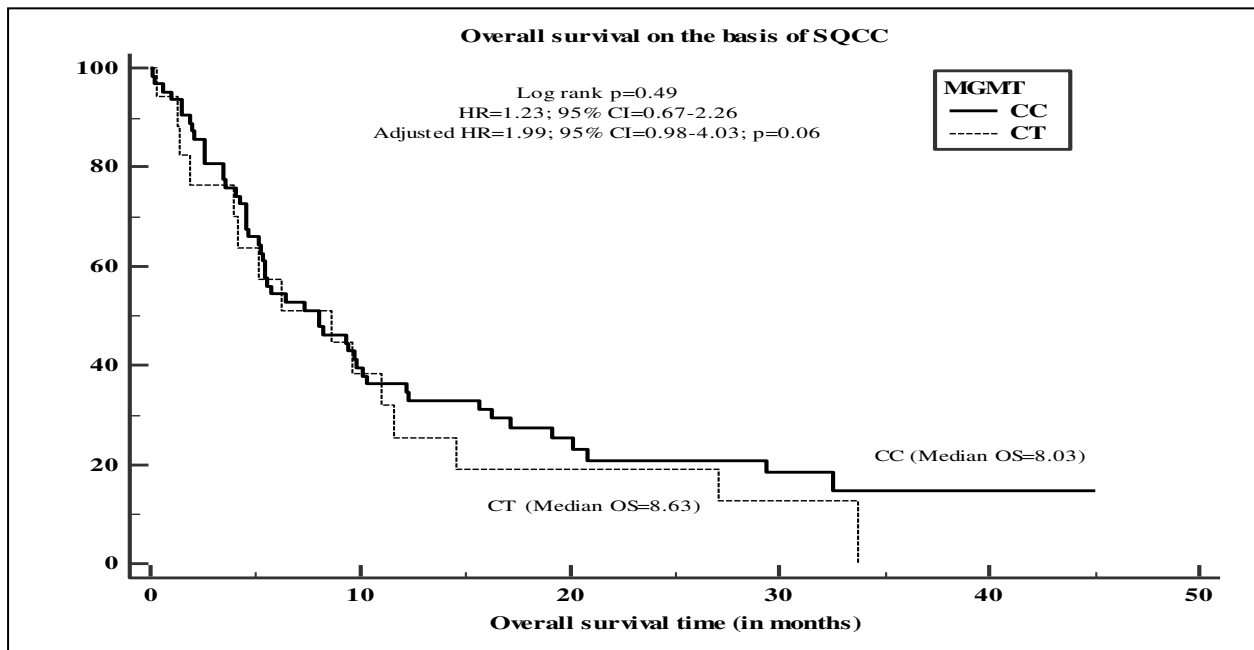


Figure 5.33 Effect of *MGMT* gene polymorphism on overall survival in SQCC patients (CC vs. CT)

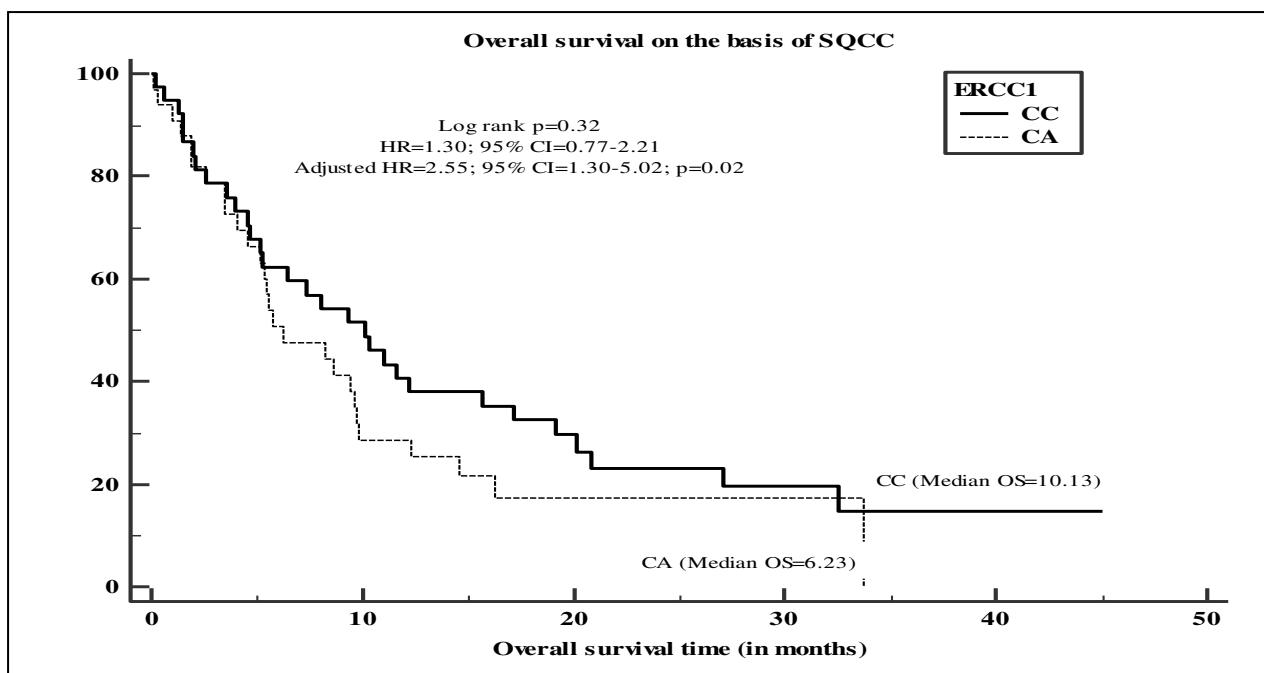


Figure 5.34 Effect of *ERCC1* gene polymorphism on overall survival in SQCC patients (CC vs. CA)

Overall survival period of mutant type genotype was observed much lower than wild type genotype (MST=5.43 vs. 10.13; HR=1.36; $p=0.46$). A non-significant relationship was observed between ERCC1 polymorphism and overall survival in SQCC patients after evaluated by univariate and multivariate model (adjusted HR=1.44; 95% CI= 0.89-2.33; $p=0.24$) as shown in figure 5.35.

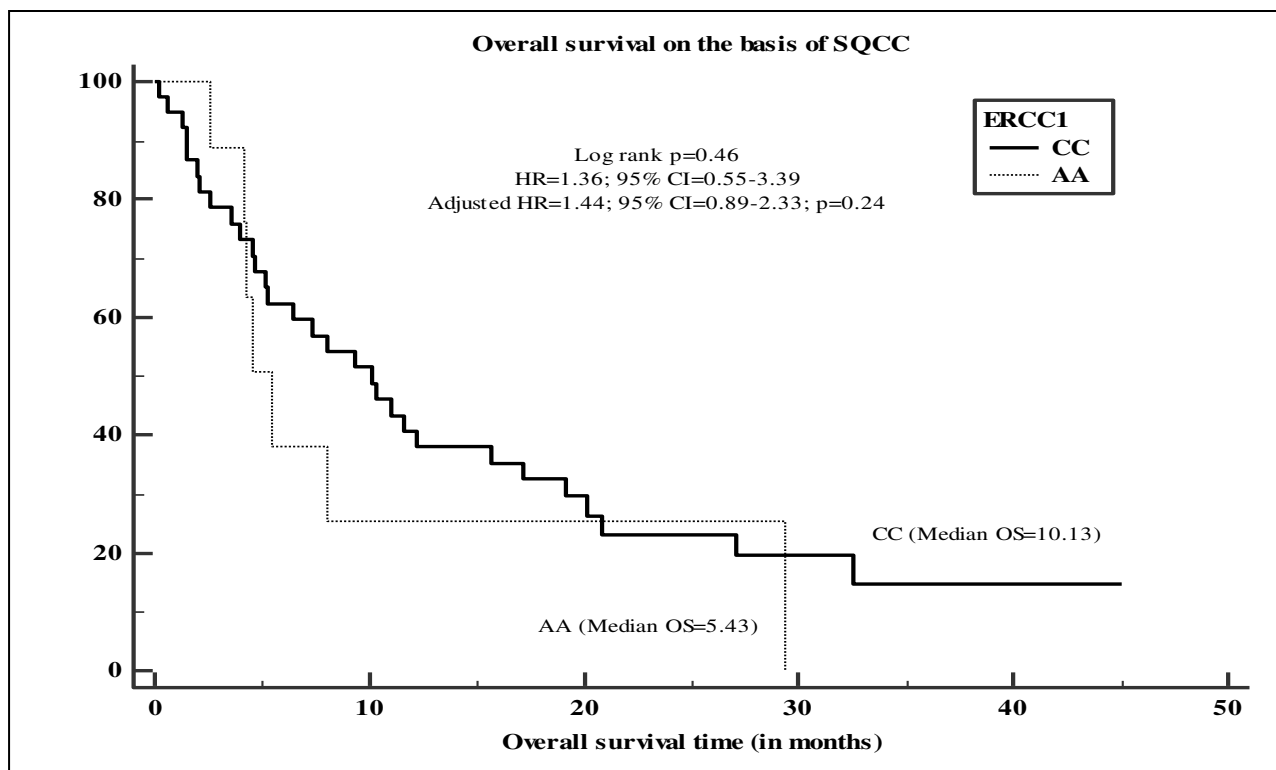


Figure 5.35 Effect of *ERCC1* gene polymorphism on overall survival in SQCC patients (CC vs. AA)

Since the frequency of mutant type genotype was less, so we combined it with heterozygous type genotype and compared against wild type genotype. Overall survival period was much lower than wild type genotype (MST=5.77 vs. 10.13; HR=1.31; $p=0.28$). When univariate model was analyzed; a non-significant relationship was observed between *ERCC1* polymorphism and overall survival in SQCC patients. But a significant relationship was observed between *ERCC1* polymorphism and overall survival in SQCC patients (adjusted HR=2.34; 95% CI=1.26-4.34; $p=0.01$) as shown in figure 5.36.

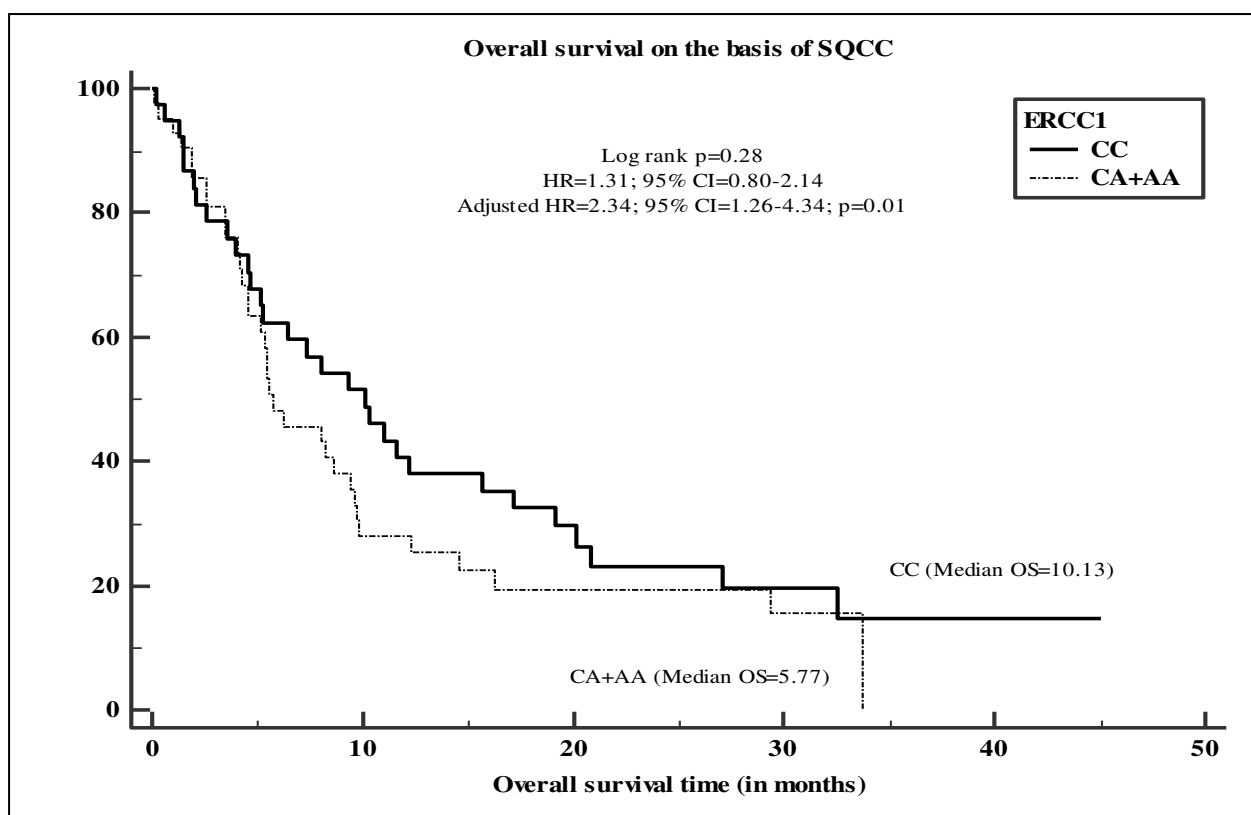


Figure 5.36 Effect of *ERCC1* gene polymorphism on overall survival in SQCC patients (CC vs. CA+AA)

After evaluation of 42 SQCC patients having common *MGMT* and *ERCC1* genotype (results are summarised in table 5.6), it was observed that 32 (76.19%) patients did not survive in study period and 10 (23.81%) patients were alive. From the current study it was observed that out of 32 patients who were dead, 75% (24) patients had wild type genotype while 25% (8) patients had heterozygous type genotype and among alive patients (10), 80% (8) patients had wild type genotype whereas only 20% (2) patient had heterozygous type genotype. Overall survival period of heterozygous type genotype was much lower than wild type genotype (MST= 8.63 vs. 10.13 months; HR- 1.27; p-value- 0.56). It indicated that death rate was more in heterozygous type genotype. After evaluated by univariate and multivariate cox model, a non-significant relationship was observed between overall survival and patients with SQCC having common *MGMT* and *ERCC1* genotype (adjusted HR=2.69; 95% CI=0.92-7.87; p=0.25) as shown in figure 5.37.

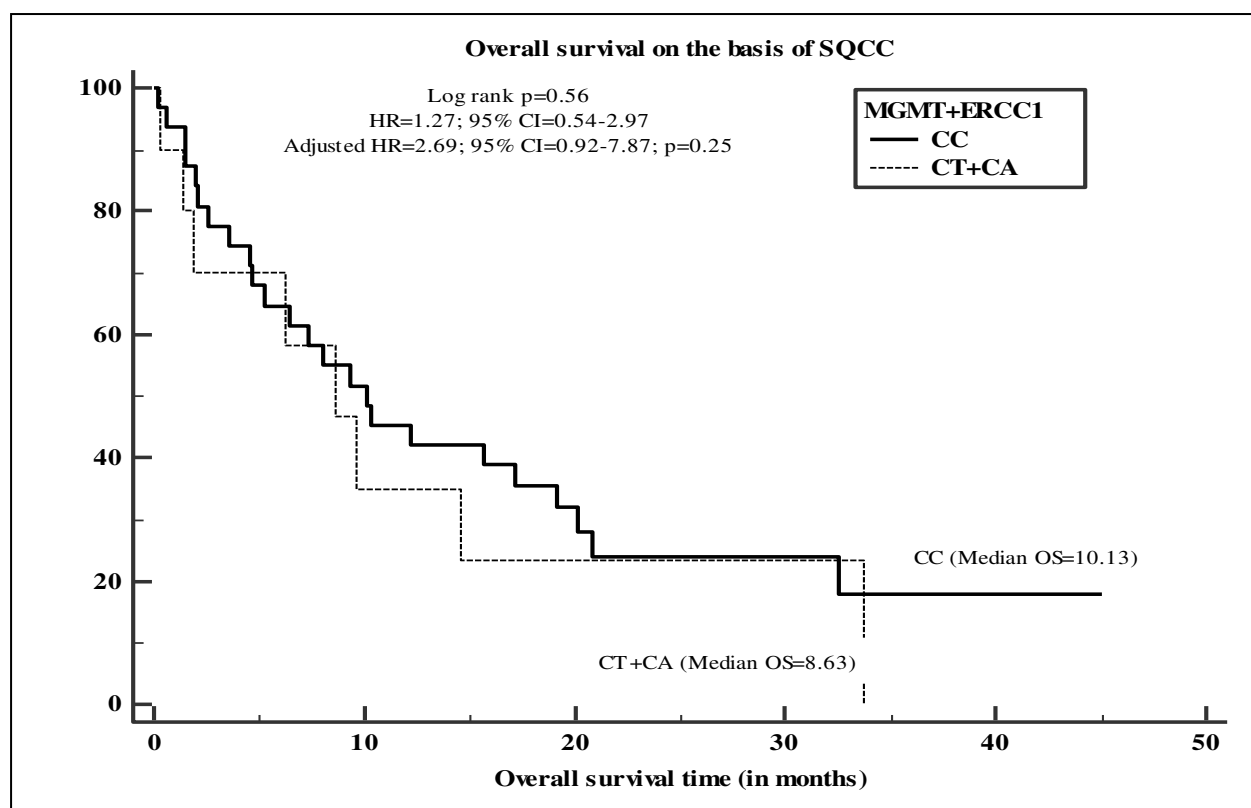


Figure 5.37 Combined effects of *MGMT* and *ERCC1* gene polymorphism on overall survival in SQCC patients (CC vs. CT+CA)

5.2.6 Overall survival analysis by Kaplan Meier and cox regression in lung cancer patients on the basis of smoking habits (light smokers, heavy smokers)

5.2.6.1 Overall survival analysis on the basis of light smokers

Table 5.7 Overall analysis on the basis of Light Smokers							
Genotype	Light Smokers		Overall survival				
<i>MGMT</i> (rs12917)	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P
CC	143 (82.18)	31 (17.82)	9.13	-	-	-	-
CT	34 (23.78)	5 (16.13)	7.23	1.16 (0.78-1.73)	0.46	1.39 (0.92 - 2.12)	0.05
<i>ERCC1</i> (rs3212986)							
CC	72	11	9.13	-	-	-	-

	(50.35)	(35.48)					
CA	57 (39.86)	17 (54.84)	9.43	0.87 (0.61-1.22)	0.42	1.07 (0.73 - 1.57)	0.11
AA	14 (9.79)	3 (9.68)	5.43	1.20 (0.65-2.20)	0.54	1.35 (0.98- 1.87)	0.02
CA+AA	71 (49.65)	20 (64.52)	8.20	0.92 (0.66-1.27)	0.60	1.16 (0.81 - 1.67)	0.09
<i>MGMT</i> + <i>ERCC1</i>	71 (94.67)	4 (5.33)					
CC	56 (78.87)	1 (25)	9.33	-	-	-	-
CT + CA	15 (21.13)	3 (75)	6.23	1.08 (0.60-1.93)	0.80	1.52 (0.78 - 2.98)	0.08
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, gender, genotype, histology, KPS and stage. Abbreviations: HR-hazard ratio; CI-confidence intervals.							

A total of 174 patients who were light smokers were followed from day of diagnosis of disease to last follow-up, it was observed that 143 (82.18%) patients were dead and 31 (17.82%) patients were alive as shown in table 5.7. For *MGMT* it was observed that among patients who were dead (143), 76.22% (109) patients had wild type genotype whereas 23.78% (34) patients had heterozygous type genotype. Among alive patients (31), 83.87% (26) patients had wild type genotype while 16.13% (5) patient had heterozygous type genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype (MST= 7.23 vs. 9.13 months; HR-1.16; p-value- 0.46). It indicated that in heterozygous type genotype; there was high risk of death as compared to wild type genotype. On applying univariate model; a non-significant association was observed between *MGMT* genotype and overall survival in lung cancer patients who were light smokers. But when multivariate cox model was analysed, a non-significant relationship was observed between *MGMT* genotype and overall survival in patients who were light smokers (adjusted HR= 1.39; 95% CI=0.92-2.12; **p=0.05**).

For *ERCC1* genotype it was observed that among patients, who did not survive in study period (143), 50.35% (72) patients had wild type genotype, 39.86% (57) patients had heterozygous type genotype and 9.79% (14) patients had mutant type genotype. Among patients who were alive (31), 35.48% (11) patients had wild type genotype, 54.84% (17) patients had heterozygous type genotype and 9.68% (3) patients had mutant type genotype. On comparing heterozygous type genotype with wild type genotype; overall survival period of wild type genotype was

observed less as compared to heterozygous type genotype (MST= 9.43 vs. 9.13 months; HR=0.87; p-value=0.42). When univariate and multivariate cox regression model was analysed; a non-significant relationship was observed between *ERCC1* and overall survival in light smoker patients (adjusted HR= 1.07; 95% CI= 0.74-1.57; p=0.11). When mutant type genotype was compared with wild type genotype; overall survival period of mutant type genotype was observed much lower than wild type genotype (MST=5.43 vs. 9.13; HR= 1.20; p=0.54). A non-significant association was observed between *ERCC1* genotype and overall survival in patients who were light smokers after evaluated by univariate. But when multivariate cox regression model was analysed; a significant relationship was observed between *ERCC1* genotype and overall survival in patients who were light smokers (adjusted HR= 1.35; 95% CI=0.98-1.87; **p=0.02**). When mutant type genotype was combined with heterozygous type genotype and compared against wild type genotype; overall survival period was observed less as compared to wild type genotype (MST=8.2 vs. 9.13; HR=0.92; p=0.60). On applying both the model *i.e* univariate and multivariate cox regression model; a non-significant relationship was observed between *ERCC1* and overall survival in lung cancer patients who were light smokers (adjusted HR= 1.16; 95% CI=0.81-1.67; p=0.09).

After evaluation of 75 patients who were light smokers throughout the study period having common *MGMT* and *ERCC1* genotype; it was observed that 71 (94.67%) patients were dead and 4 (5.33%) patients were alive. Among patients who were dead (71), 78.87% (56) patients had wild type genotype whereas 21.13% (15) patients had heterozygous type genotype. Among alive patients (4), 25% (1) patients had wild type genotype whereas 75% (3) patients had heterozygous type genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype (MST= 6.23 vs. 9.33; HR- 1.08; p-value- 0.80). A non-significant relationship was observed between overall survival and patients (who were light smokers) having common *MGMT* and *ERCC1* genotype (adjusted HR=1.52; 95% CI= 0.78-2.98; p=0.08).

5.2.6.2 Overall survival of lung cancer patients on the basis of heavy smokers

Table 5.8 Overall survival analysis on the basis of Heavy Smokers							
Genotype	Heavy Smokers		Overall survival				
<i>MGMT</i> (rs12917)	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P
	35 (89.74)	4 (10.26)					
CC	27 (77.14)	4 (100)	7.6	-	-	-	-
CT	8 (22.86)	0 (0)	4.3	1.64 (0.65-4.10)	0.21	1.68 (0.72-3.94)	0.01
<i>ERCC1</i> (rs3212986)							
CC	19 (54.28)	4 (100)	8.0	-	-	-	-
CA	13 (37.14)	0 (0)	5.77	1.46 (0.69-3.07)	0.29	1.73 (0.73 - 4.12)	0.03
AA	3 (8.57)	0 (0)	4.57	1.19 (0.32-4.34)	0.78	2.27 (1.00-5.11)	0.003
CA+AA	16 (45.71)	0 (0)	5.43	1.41 (0.71-2.80)	0.30	2.05 (0.93 -4.50)	0.01
<i>MGMT</i> + <i>ERCC1</i>	19 (82.61)	19 (82.61)					
CC	15 (78.95)	4 (100)	9.37	-	-	-	-
CT + CA	4 (21.05)	0 (0)	6.33	0.60 (0.40-4.46)	0.80	2.83 (0.60-13.33)	0.02
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, gender, genotype, histology, KPS and stage. Abbreviations: HR-hazard ratio; CI-confidence intervals.							

It was observed that out of 39 patients who smoked 25 or more than 25 cigarettes/bidis per day; 35 (89.74%) patients were dead and 4 (10.26%) patients were alive as shown in table 5.8. For *MGMT* genotype it was observed that among dead patients (35), 77.14% (27) patients had wild type genotype while 22.86% (8) patients had heterozygous type genotype. Among alive patients (4), all the patients had wild type genotype. Overall survival period of heterozygous type genotype was much lower than wild type genotype (MST=4.3 vs. 7.6; HR=1.64; p-value- 0.21). A significant association was observed between overall survival and *MGMT* among patients who were heavy smokers after applying cox hazards proportional analysis (adjusted HR=1.68; 95% CI= 0.72-3.94; **p=0.01**). But, this association was not statistically proved by univariate model.

Further, on relating the *ERCC1* genotype with overall survival in heavy smokers, it was observed that among patients who did not survive (35), 54.28% (19) patients had wild type genotype, 37.14% (13) patients had heterozygous type genotype and 8.57% (3) patients had mutant type genotype. Among alive patients (4), all the patients had wild type genotype. On comparing heterozygous type genotype with wild type genotype; overall survival period of heterozygous type genotype was observed less as compared to wild type genotype (MST=5.77 vs. 8.03 months; HR=1.46; p=0.29). When cox hazards proportional model was analysed; a significant association was observed between overall survival and *ERCC1* among patients who were heavy smokers (adjusted HR=1.73; 95% CI= 0.73-4.12; **p=0.03**). But, this association was not statistically proved by univariate model. When mutant type genotype was compared against wild type genotype; overall survival period of mutant type genotype was observed much lower than wild type genotype (MST=4.57 vs. 8.03 months; HR=1.19; p=0.78). A significant association was observed between overall survival and *ERCC1* among patients who were heavy smokers after applying cox hazards proportional analysis (adjusted HR=2.27; 95% CI= 1.01-5.11; **p=0.04**). But, this association was not statistically proved by univariate model. It was observed that frequency of mutant type genotype was less, so we combined it with the heterozygous type genotype and compared against wild type genotype; overall survival period in this case was less as compared to wild type genotype (MST=5.43 vs. 8.03 months; HR=1.41; p=0.30). When cox hazards proportional model was analysed; a significant association was observed between overall survival and *ERCC1* among patients who were heavy smokers (adjusted HR=2.05; 95% CI= 0.93-4.50; **p=0.005**). But, this association was not statistically proved by univariate model.

When combined effect of *MGMT* and *ERCC1* was studied among 23 patients who were heavy smokers; it was observed that 19 (82.61%) patients were dead and 4 (17.39%) patients were alive as shown in table 5.8. Among patients who did not survive in current study period (19), 78.95% (15) patients had wild type genotype while 21.05% (4) patients had heterozygous type genotype. Among alive patients (4), all the patients had wild type genotype. Overall survival period of the heterozygous type genotype was observed much lower than wild type genotype (MST= 6.33 vs. 9.37 months; HR=0.60; p=0.80). A non-significant association was observed between overall survival and combined effect of *MGMT* and *ERCC1* in heavy smokers. But statistically significant relationship was seen between the patients carrying common *MGMT* and

ERCC1 genotype and overall survival after applying multivariate cox regression analysis. (adjusted HR= 2.83; 95% CI=0.60-13.33; **p=0.02**).

5.2.7 Overall survival analysis by Kaplan Meier and cox regression in lung cancer patients the basis of smoking status (smokers, non-smokers)

5.2.7.1 Overall survival analysis among smokers

Table 5.9 Overall survival on the basis of smokers							
Genotype	Smokers		Overall survival				
<i>MGMT</i> (rs12917)	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P
	153 (84.53)	28 (15.47)					
CC	115 (75.16)	24 (85.71)	8.20	-	-	-	-
CT	38 (24.84)	4 (14.29)	6.23	1.25 (0.85-1.84)	0.24	1.34 (0.91-1.99)	0.03
<i>ERCC1</i> (rs3212986)							
CC	82 (53.59)	13 (46.43)	8.27	-	-	-	-
CA	58 (37.91)	13 (46.43)	8.20	0.99 (0.71-1.39)	0.96	1.22 (0.85-1.75)	0.04
AA	14 (9.15)	2 (7.14)	5.07	1.34 (0.71-2.52)	0.31	1.48 (1.08-2.02)	0.003
CA+AA	72 (47.06)	15 (53.57)	7.13	1.04 (0.76-1.43)	0.80	1.31 (0.94-1.84)	0.02
<i>MGMT + ERCC1</i>	79 (84.04)	15 (15.96)					
CC	63 (79.75)	12 (80)	9.33	-	-	-	-
CT + CA	16 (20.25)	3 (20)	8.40	1.19 (0.67-2.12)	0.53	1.34 (0.72-2.47)	0.08
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, gender, genotype, histology, KPS and stage. Abbreviations: HR-hazard ratio; CI-confidence intervals.							

Out of 181 lung cancer patients who were smokers, it was observed that patients (who carried *MGMT* genotype) 153 (84.53%) patients were dead and 28 (15.47%) patients were alive as shown in table 5.9. Among patients who did not survive (153), 75.16% (115) patients had wild

type genotype while 24.84% (38) patients had heterozygous type genotype and among alive patients (28), 85.71% (24) patients had wild type genotype while 14.29% (4) patient had heterozygous type genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype (MST= 6.23 vs. 8.2 months; HR=1.25; p-value= 0.24). On applying univariate analysis a non-significant association was observed between overall survival and MGMT genotype in lung cancer patients who smoked. But a significant relationship was observed after applying multivariate cox regression analysis (adjusted HR= 1.34; 95% CI= 0.91-1.99; **p=0.03**).

For the subjects carrying *ERCC1* genotype it was observed that among dead patients (153), 53.39% (82) patients had wild type genotype, 37.91% (58) patients had heterozygous type genotype and 9.15% (14) patients had mutant type genotype. Out of patients who survived (28), 46.43% (13) patients had wild type genotype, 46.43% (13) patient had heterozygous type genotype and 7.14% (2) patients had mutant type genotype. Overall survival period of heterozygous type genotype was observed slightly lower than wild type genotype (MST= 8.2 vs. 8.27 months; HR= 0.99; p=0.96). A non-significant association was observed between overall survival and lung cancer patients carrying *ERCC1* heterozygous type genotype after applying univariate analysis; but when multivariate cox model was analysed a statistically significant relationship was observed (adjusted HR= 1.22; 95% CI= 0.85-1.75; **p=0.04**).

Overall survival period of mutant type genotype was much lower than compared to wild type genotype (MST=5.07 vs. 8.27; HR= 1.34; p=0.31). A non-significant association was observed between overall survival and lung cancer patients carrying *ERCC1* mutant type genotype after applying univariate analysis; but when multivariate cox model was analysed a statistically significant relationship was observed (adjusted HR= 1.48; 95% CI= 1.08-2.02; **p=0.003**). Since the frequency of mutant type genotype was less, so we combined it with heterozygous type genotype and compared it against wild type genotype. Overall survival period was less as compared to wild type genotype (MST= 7.13 vs. 8.27 months; HR=1.04; p=0.80). A non-significant association was observed between overall survival and lung cancer patients after applying univariate analysis; but when multivariate cox model was analysed a statistically significant relationship was observed (adjusted HR= 1.31; 95% CI= 0.94-1.84; **p=0.003**).

After evaluation of 94 lung cancer patients who harbouring common *MGMT* and *ERCC1* genotype and also who smoked, it was observed that 79 (84.04%) patients did not survive and 15 (15.96%) patients were survived as shown in table 5.9. Among dead patients (79), 79.75% (63) patients had wild type genotype whereas 20.25% (16) patients had heterozygous type genotype. 80% (12) patients who were alive had wild type genotype while 20% (3) patients had heterozygous type genotype. Overall survival period of heterozygous genotype was less as compared to wild genotype (MST=8.40 vs. 9.33; HR=1.19; p=0.53). A non-significant relationship was observed between overall survival and who harbouring common *MGMT* and *ERCC1* genotype and also who smoked after applying both the model univariate as well as multivariate cox regression model (adjusted HR=1.34; 95% CI=0.72-2.47; p=0.08).

5.2.7.2 Overall survival analysis on the basis of non-smokers

Table 5.10 Overall survival analysis on the basis of non-smokers							
Genotype	Non-smokers		Overall survival				
<i>MGMT</i> (rs12917)	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P
	25 (78.13)	7 (21.87)					
CC	21 (84)	6 (85.71)	9.43	-	-	-	-
CT	4 (16)	1 (14.29)	5.93	0.96 (0.33-2.74)	0.93	5.42 (0.93-31.7)	0.32
<i>ERCC1</i> (rs3212986)							
CC	10 (40)	2 (28.57)	9.13	-	-	-	-
CA	12 (48)	4 (57.14)	6.73	0.75 (0.32-1.77)	0.48	0.75 (0.23-2.47)	0.65
AA	3 (12)	1 (14.29)	4.97	0.94 (0.27-3.36)	0.93	1.54 (0.47-5.03)	0.46
CA+AA	15 (60)	5 (71.43)	6.73	0.78 (0.34-1.79)	0.54	0.80 (0.27-2.41)	0.64
<i>MGMT</i> + <i>ERCC1</i>	11 (91.67)	1 (8.33)					
CC	8 (72.73)	1 (100)	9.13	-	-	-	-
CT + CA	3 (27.27)	0 (0)	4.17	0.93 (0.25-3.43)	0.90	96.34 (1.63-568)	0.16

Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, gender, genotype, histology, KPS and stage. **Abbreviations:** HR-hazard ratio; CI-confidence intervals.

A total of 32 patients who did not smoke were followed from day of diagnosis of disease to last follow-up, it was observed that 25 (78.13%) patients were dead and 7 (21.87%) patients were alive as shown in table 5.10. For subjects harbouring *MGMT* genotype it was observed that among dead patients (25), 84% (21) patients had wild type genotype while 16% (4) patients had heterozygous type genotype. Out of 7 alive patients, 85.71% (6) patients had wild type genotype whereas only 14.29% (1) patient had heterozygous genotype. Overall survival period of heterozygous type genotype was much lower than wild type genotype (MST= 5.93 vs. 9.43 months; HR=0.96; p=0.93). After applying both the model univariate as well as multivariate cox regression; a non-significant association was observed between overall survival and *MGMT* in lung cancer patients who did not smoke (adjusted HR=5.42; 95% CI=0.93-31.7; p=0.32).

When patients carrying *ERCC1* genotype was studied; it was observed that among patients who did not survive (25), 40% (10) patients had wild type genotype, 48% (12) patients had heterozygous type genotype and 12% (3) patients had mutant type genotype. Among alive patients (7), 28.57% (2) patients had wild type genotype, 57.14% (4) patient had heterozygous type genotype and only 14.29% (1) patients had mutant type genotype. Overall survival period of heterozygous type genotype and mutant type genotype was less as compared to wild type genotype (MST=6.73 vs. 9.13 months; HR=0.75; p=0.48 and MST=4.97 vs. 9.13; HR=0.94; p=0.93 respectively). When univariate and multivariate model was analysed for heterozygous type genotype and mutant type genotype against wild type genotype; a non-significant relationship was observed between overall survival and *ERCC1* in lung cancer patients who did not smoke (adjusted HR=0.75; 95% CI=0.23-2.47; p=0.65 and adjusted HR=1.54; 95% CI=0.47-5.03; p=0.46 respectively). On combining the mutant type genotype with heterozygous type genotype and compared against wild type genotype; overall survival period was observed less as compared to wild type genotype (MST=6.73 vs. 9.13; HR=0.78; p=0.54). A non-significant association was also observed between overall survival and *ERCC1* genotype in lung cancer patients who were non-smokers after applying univariate and multivariate cox model (adjusted HR=0.80; 95% CI=0.27-2.41; p=0.64).

To study the combined effect of common *MGMT* and *ERCC1* genotype on overall survival among patients who did not smoke a total of 12 patients were evaluated. It was observed; 11 (91.67%) patients were dead and only 1 (8.33%) patient survived as shown in table 5.10. It was observed among dead patients (11), 72.73% (8) patients had wild type genotype while 27.27% (3) patients had heterozygous type genotype. Patient who survived had wild type genotype. Overall survival period of heterozygous type genotype was much lower than wild type genotype (MST=4.17 vs. 9.13; HR=0.93; p=0.90). On applying both the models a non-significant association was observed between overall survival and common *MGMT* and *ERCC1* genotype in lung cancer patients who did not smoke.

5.2.8 Overall survival of lung cancer patients treated with different chemotherapeutic regimens

5.2.8.1 Overall survival of lung cancer patients treated with docetaxel + cisplatin/carboplatin and its association with *MGMT* and *ERCC1* genotype

Table 5.11 Overall survival analysis on the basis of regimen docetaxel + cisplatin/carboplatin							
Genotype	docetaxel+cis-platin/carboplatin		Overall survival				
<i>MGMT</i>	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P- value
CC	42 (85.71)	7 (14.29)	5.77	-	-	-	-
CT	32 (76.19)	7 (100)	9.67	0.87 (0.44-1.73)	0.69	1.00 (0.40-2.49)	0.22
<i>ERCC1</i>							
CC	21 (50)	3 (42.86)	6.47	-	-	-	-
CA	18 (42.86)	2 (28.57)	8.20	1.06 (0.57 -2.0)	0.84	1.86 (0.80-4.35)	0.22
AA	3 (7.14)	2 (28.57)	4.57	0.72 (0.25-2.11)	0.59	1.10 (0.54-2.23)	0.36
CA+AA	21 (50)	4 (57.14)	8.40	0.99 (0.54-1.81)	0.96	1.69 (0.78-3.65)	0.12
<i>MGMT</i> + <i>ERCC1</i>	25 (89.29)	3 (10.71)					

CC	18 (72)	3 (100)	5.27	-	-	-	-
CT + CA	7 (28)	0 (0)	8.63	0.96 (0.40-2.28)	0.92	1.29 (0.35-4.78)	0.50
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, genotype, histology, KPS, smoking habits and stage. Abbreviations: HR-hazard ratio; CI-confidence intervals.							

A total of 49 patients who treated with regimen docetaxel in combination with cisplatin/carboplatin, were followed from day of diagnosis of disease to last follow-up, it was observed that 42 (85.71%) patients were dead and 7 (14.29%) patients were alive as shown in table 5.11. For subjects having *MGMT* genotype it was observed that among dead patients (42), 76.19% (32) patients had wild type genotype whereas 23.81% (10) patients had heterozygous type genotype. All the patients who survived had wild type genotype. Overall survival period of wild type genotype was less as compared to heterozygous type genotype (MST=5.77 vs. 9.67 months; HR=0.87; p=0.69). On applying both the models *i.e* univariate as well as multivariate cox, a non-significant relationship was observed between overall survival and *MGMT* genotype in lung cancer patients after treated with docetaxel in combination with cisplatin/carboplatin (adjusted HR=1; 95% CI= 0.40-2.5; p=0.22).

For *ERCC1* genotype it was observed that among 42 patients who did not survive, 50% (21) patients had wild type genotype, 42.86% (18) patients had heterozygous genotype and 7.14% (3) patients had mutant type genotype. Among alive patients (7), 42.86% (3) patients had wild type genotype, 28.57% (2) patient had heterozygous type genotype and 28.57% (2) patients had mutant type genotype. Overall survival period of heterozygous type genotype was more as compared to wild type genotype (MST=8.2 vs. 6.47 months; HR=1.06; p=0.84). But overall survival period of mutant type genotype was less as compared to wild type genotype (MST=4.57 vs. 6.47 months; HR=0.72; p=0.59). When univariate and multivariate model was analysed for heterozygous type genotype and mutant type genotype against wild type genotype respectively; a non-significant relationship was observed between overall survival and *ERCC1* in lung cancer patients who treated with docetaxel in combination with cisplatin/carboplatin. (adjusted HR=1.86; 95% CI=0.80-4.35; p=0.22 and adjusted HR=1.10; 95% CI=0.54-2.23; p=0.36 respectively). Since the frequency of mutant type genotype was less, we combined it with heterozygous type genotype and compared against wild type genotype; overall survival period was observed more as compared to wild type genotype (MST=8.4 vs. 6.47; HR=0.99;

p=0.96). A non-significant association was also observed between overall survival and *ERCC1* genotype in lung cancer patients who were administered with docetaxel in combination with cisplatin/carboplatin after applying univariate and multivariate cox model (adjusted HR=1.69; 95% CI=0.78-3.65; p=0.12).

After evaluation of 28 patients who were treated with regimen docetaxel in combination with cisplatin/carboplatin and having common *MGMT* and *ERCC1* genotype, it was observed that 25 (89.29%) patients were dead and 3 (10.71%) patients were alive as shown in table 5.11. Among dead patients (25), 72% (18) patients had wild type genotype whereas 28% (7) patients had heterozygous type genotype. All the patients who survived had wild type genotype. Overall survival period of heterozygous type genotype was more as compared to wild type genotype (MST=8.63 vs. 5.27 months; HR= 0.96; p=0.92). When both univariate analysis (Kaplan) and multivariate analysis (cox) were evaluated, a non-significant relationship was found between overall survival in lung cancer patients treated with docetaxel in combination with cisplatin/carboplatin and patients having common *MGMT* and *ERCC1* genotype (adjusted HR= 1.29; 95% CI=0.35-4.78; p=0.50).

5.2.8.2 Overall survival of lung cancer patients treated with irinotecan + cisplatin / carboplatin and its association with *MGMT* and *ERCC1* genotype

Table 5.12 Overall survival analysis on the basis of regimen irinotecan + cisplatin/carboplatin							
Genotype	Irinocis/Irinocarb		Overall survival				
<i>MGMT</i>	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P- value
	36 (81.82)	8 (18.18)					
CC	31 (86.11)	8 (100)	11.27	-	-	-	-
CT	5 (13.89)	0 (0)	5.80	1.81 (0.55-5.96)	0.20	2.58 (0.88-7.54)	0.11
<i>ERCC1</i>							
CC	19 (52.78)	4 (50)	9.67	-	-	-	-
CA	14 (38.89)	4 (50)	11.27	1.00 (0.50-2.00)	1.00	1.22 (0.54-2.76)	0.17
AA	3	0	7.23	1.93	0.28	3.31	0.23

	(8.33)	(0)		(0.40-9.41)		(1.23-8.94)	
CA+AA	17 (47.22)	4 (50)	11.27	1.09 (0.57-2.10)	0.80	1.36 (0.62-2.98)	0.21
MGMT + ERCC1	17 (80.95)	4 (19.05)					
CC	16 (94.12)	4 (100)	9.67	-	-	-	-
CT + CA	1 (5.88)	0 (0)	5.80	2.64 (0.11-62.1)	0.32	1.38 (0.04-44.1)	0.37
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, genotype, histology, KPS, smoking habits and stage. Abbreviations: HR-hazard ratio; CI-confidence intervals.							

A total of 44 patients who administered irinotecan in combination with cisplatin/carboplatin, were studied and observed that 36 (81.82%) patients were dead and 8 (18.18%) patients were alive as shown in table 5.12. For subjects carrying *MGMT* genotype; among dead patients (36), 86.11% (31) patients had wild type genotype and 13.89% (5) patients had heterozygous type genotype. All the survived patients had wild type genotype. Overall survival period of heterozygous type genotype was much lower than wild type genotype (MST=5.8 vs. 11.27 months; HR=1.81; p-value= 0.20). A non-significant relationship was observed between overall survival and *MGMT* genotype in lung cancer patients treated with irinotecan in combination with cisplatin/carboplatin after applying univariate and multivariate model (adjusted HR=2.58; 95% CI=0.88-7.54; p=0.11).

For patients carrying *ERCC1* genotype; among dead patients (36), 52.78% (19) patients had wild type genotype, 38.89% (14) patients had heterozygous type genotype and 8.33% (3) patients had mutant type genotype. 50% (4) patients who survived had wild type genotype and 50% (4) patient had heterozygous type genotype. Overall survival period of heterozygous type genotype was more as compared to wild type genotype (MST=11.27 vs. 9.67 months; HR=1; p=1). But overall survival period of mutant type genotype was less as compared to wild type genotype (MST=7.23 vs. 9.67 months; HR=1.93; p=0.27). When univariate and multivariate model was analysed for heterozygous type genotype and mutant type genotype against wild type genotype respectively; a non-significant relationship was observed between overall survival and *ERCC1* in lung cancer patients who treated with irinotecan in combination with cisplatin/carboplatin (adjusted HR=1.22; 95% CI=0.54-2.76; p=0.17 and adjusted HR=3.31; 95% CI=1.23-8.94; p=0.23 respectively). On combining mutant type genotype with hetero-

zygous type genotype and compared it against wild type genotype; overall survival period was observed more as compared to wild type genotype (MST=11.27 vs. 9.67; HR=1.09; p=0.79). A non-significant association was also observed between overall survival and *ERCC1* genotype in lung cancer patients who were administered with irinotecan in combination with cisplatin/carboplatin after applying univariate and multivariate cox model (adjusted HR=1.36; 95% CI=0.62-2.98; p=0.21).

After evaluation of 28 patients who were treated with regimen irinotecan in combination with cisplatin/carboplatin and having common *MGMT* and *ERCC1* genotype, it was observed that 17 (80.95%) patients were dead and 4 (19.05%) patients were alive as shown in table 5.12. Among dead patients (17), 94.12% (16) patients had wild type genotype whereas only 5.88% (1) patients had heterozygous type genotype. All the patients who survived had wild type genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype (MST=5.8 vs. 9.67 months; HR= 2.64; p=0.32). When both univariate analysis (Kaplan) and multivariate analysis (cox) were evaluated, a non-significant relationship was found between overall survival in lung cancer patients treated with irinotecan in combination with cisplatin/carboplatin and patients having common *MGMT* and *ERCC1* genotype (adjusted HR= 1.38; 95% CI=0.04-44.1; p=0.37).

5.2.8.3 Overall survival of lung cancer patients treated with pemetrexed in combination with cisplatin/carboplatin and its relationship with *MGMT* and *ERCC1* genotype

Table 5.13 Overall analysis on the basis of regimen pemetrexed + cisplatin/carboplatin							
Genotype	Pemecis/pemecarb		Overall survival				
<i>MGMT</i> (rs12917)	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P
	35 (77.78)	10 (22.22)					
CC	23 (65.71)	8 (80)	10.33	-	-	-	-
CT	12 (34.29)	2 (20)	5.70	1.66 (0.76-3.61)	0.15	1.68 (0.77-3.68)	0.10
<i>ERCC1</i> (rs3212986)							
CC	22 (62.86)	3 (30)	9.13	-	-	-	-

CA	11 (31.43)	6 (60)	14	0.55 (0.28-1.09)	0.1	0.53 (0.24- 1.2)	0.06
AA	2 (5.71)	1 (10)	4.97	0.90 (0.23-3.62)	0.89	1.03 (0.45-2.35)	0.62
CA+AA	13 (37.14)	7 (70)	10.67	0.60 (0.31-1.16)	0.13	0.60 (0.28-1.28)	0.09
<i>MGMT</i> + <i>ERCC1</i>	22 (88)	3 (12)					
CC	16 (72.73)	2 (66.67)	9.33	-	-	-	-
CT + CA	6 (27.27)	1 (33.33)	6.33	1.15 (0.44-3.04)	0.77	1.19 (0.36-3.96)	0.46
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, ECOG, genotype, histology, KPS, smoking habits and stage. Abbreviations: HR-hazard ratio; CI-confidence intervals.							

Among 45 patients who were treated with regimen pemetrexed in combination with cisplatin/carboplatin; 35 (77.78%) patients were did not survive and 10 (22.22%) patients were survive as shown in figure 5.13. Among dead patients (35) who had *MGMT* genotype, 65.71% (23) patients had wild type genotype whereas 34.29% (12) patients had heterozygous type genotype. 80% (8) alive patients had wild type genotype and 20% (2) patients had heterozygous type genotype. Overall survival period for heterozygous type genotype was much lower than wild genotype (MST= 5.70 vs. 10.33 months; HR-1.66; p-value- 0.15). A non-significant relationship was observed between *MGMT* genotype and overall survival in lung cancer patients treated with pemetrexed in combination with cisplatin/carboplatin (adjusted HR=1.68; 95% CI=0.77-3.68; p=0.1).

In the cases of subjects having *ERCC1* genotype; it was observed that 62.86% (22) dead patients had wild type genotype, 31.43% (11) dead patients had heterozygous type genotype and 5.71% (2) patients had mutant type genotype. 30% (3) alive patients had wild type genotype, 60% (6) alive patients had heterozygous type genotype and only 10% (1) patient had mutant type genotype. Overall survival period of heterozygous type genotype was more as compared to wild type genotype (MST=14 vs. 9.13 months; HR=0.55; p=0.09). But overall survival period of mutant type genotype was much lower than wild type genotype (MST=4.97 vs. 9.13 months; HR=1.93; p=0.27). When univariate and multivariate model was analysed for heterozygous type genotype and mutant type genotype against wild type genotype respectively; a non-significant relationship was observed between overall survival and *ERCC1* in lung cancer patients who

treated with pemetrexed in combination with cisplatin/carboplatin (adjusted HR=0.53; 95% CI=0.24-1.20; p=0.06 and adjusted HR=1.03; 95% CI=0.45-2.35; p=0.62 respectively). On combining mutant type genotype with hetero-zygous type genotype and compared it against wild type genotype; overall survival period was observed more as compared to wild type genotype (MST=10.67 vs. 9.13; HR=0.60; p=0.13). A non-significant association was also observed between overall survival and *ERCC1* genotype in lung cancer patients who were administered with pemetrexed in combination with cisplatin/carboplatin after applying univariate and multivariate cox model (adjusted HR=0.60; 95% CI=0.28-1.28; p=0.10).

After evaluation of 25 patients who were treated with regimen pemetrexed in combination with cisplatin/carboplatin and having common *MGMT* and *ERCC1* genotype, it was observed that 22 patients were dead and 3 patients were alive as shown in table 5.13. Among dead patients (22), 16 patients had wild type genotype whereas only two patients had heterozygous type genotype. Among alive patients two patients had wild type genotype and only one patient had heterozygous type genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype (MST=6.33 vs. 9.33 months; HR= 1.15; p=0.77). When both univariate analysis (Kaplan) and multivariate analysis (cox) were evaluated, a non-significant relationship was found between overall survival in lung cancer patients treated with pemetrexed in combination with cisplatin/carboplatin and patients having common *MGMT* and *ERCC1* genotype (adjusted HR= 1.19; 95% CI=0.36-3.96; p=0.47).

5.2.9 Overall survival of lung cancer patients and its relationship with *MGMT* and *ERCC1* genotype on the basis of gender

5.2.9.1 Overall survival of female lung cancer patients and its relationship with *MGMT* and *ERCC1* genotype

Table 5.14 Overall survival analysis among female lung cancer patients							
Genotype	Females		Overall survival				
<i>MGMT</i>	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P
	21 (87.5)	3 (12.5)					
CC	17	2	5.13	-	-	-	-

	(80.95)	(66.67)					
CT	4 (19.05)	1 (33.33)	14	0.47 (0.19-1.16)	0.14	0.77 (0.19-3.13)	0.73
<i>ERCC1</i>							
CC	7 (33.33)	3 (100)	5.93	-	-	-	-
CA	12 (57.14)	0 (0)	4.17	1.63 (0.66-4)	0.29	0.91 (0.27-3.03)	0.41
AA	2 (9.52)	0 (0)	5.43	1.16 (0.22-5.97)	0.85	1.87 (0.23-15.5)	0.48
CA+AA	14 (66.67)	0 (0)	5.13	1.58 (0.67-3.73)	0.31	1.03 (0.36-3.01)	0.75
<i>MGMT</i> + <i>ERCC1</i>	8 (80)	2 (20)					
CC	6 (75)	2 (100)	4.50	-	-	-	-
CT + CA	2 (25)	0 (0)	4.17	0.70 (0.16-3.10)	0.60	4.65 (0.1- 10)	0.15
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, genotype, histology, KPS, smoking habits and stage. Abbreviations: HR-hazard ratio; CI-confidence intervals.							

After evaluation of 24 female lung cancer patients, it was observed that 21 (87.5%) patients were dead and 3 (12.5%) patients were alive as shown in table 5.14. For subjects having *MGMT* genotype; 80.95% (17) dead patients had wild type genotype and 19.05% (4) dead patients had heterozygous type genotype. 66.67% (2) alive patients had wild type genotype and only 33.33% (1) patient had heterozygous type genotype. Overall survival period for heterozygous genotype was much higher than wild type genotype (MST=14 vs. 5.13 months; HR=0.47; p=0.14). When both univariate analysis (Kaplan) and multivariate analysis (cox) was evaluated, a non-significant relationship was observed between *MGMT* genotype and overall survival in female lung cancer patients (adjusted HR= 0.77; 95% CI=0.19-3.13; p=0.73).

In the cases of subjects having *ERCC1* genotype; it was observed that 7 dead patients had wild type genotype, 12 dead patients had heterozygous type genotype and 2 patients had mutant type genotype. All the alive patients had wild type genotype. Overall survival period of heterozygous type genotype and mutant type genotype was less as compared to wild type genotype (MST=4.17 vs. 5.93 months; HR=1.63; p=0.29 and MST=5.43 vs. 5.93 months; HR=1.16; p=0.85 respectively). When univariate and multivariate model was analysed for heterozygous

type genotype and mutant type genotype against wild type genotype respectively; a non-significant relationship was observed between overall survival and *ERCC1* in female lung cancer patients (adjusted HR=0.91; 95% CI=0.27-3.03; p=0.41 and adjusted HR=1.87; 95% CI=0.23-15.48; p=0.47 respectively). On combining mutant type genotype with hetero-zygous type genotype and compared it against wild type genotype; overall survival period was observed less as compared to wild type genotype (MST=5.13 vs. 5.93; HR=1.58; p=0.31). A non-significant association was also observed between overall survival and *ERCC1* genotype in female lung cancer patients after applying univariate and multivariate cox model (adjusted HR=1.03; 95% CI=0.36-3.01; p=0.75).

After evaluation of 10 female patients having common *MGMT* and *ERCC1* genotype, it was observed that 8 patients were dead and 2 patients were alive as shown in table 5.14. Among dead patients (8), 6 patients had wild type genotype whereas only two patients had heterozygous type genotype. All the alive patients had wild type genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype (MST=4.17 vs. 4.5 months; HR= 0.70; p=0.60). When both univariate analysis (Kaplan) and multivariate analysis (cox) were evaluated, a non-significant relationship was found between overall survival in female lung cancer patients having common *MGMT* and *ERCC1* genotype (adjusted HR= 4.65; 95% CI=0.1-10; p=0.15).

5.2.9.2 Overall survival of male lung cancer patients having *MGMT* and *ERCC1* genotype

Table 5.15 Overall survival analysis among male lung cancer patients							
Genotype	Males		Overall survival				
<i>MGMT</i>	Dead N (%)	Alive N (%)	Median OS (Months)	HR 95% CI	Log P	HR 95% CI	P
	158 (83.16)	32 (16.84)					
CC	120 (75.95)	28 (87.5)	9.13	-	-	-	-
CT	38 (24.05)	4 (12.5)	6.17	1.36 (0.92-2.03)	0.09	1.34 (0.91-1.97)	0.01
<i>ERCC1</i>							
CC	85 (53.80)	12 (37.5)	8.30	-	-	-	-
CA	58	17	9	0.83	0.28	1.07	0.03

	(36.71)	(53.13)		(0.60-1.16)		(0.75-1.53)	
AA	15 (9.49)	3 (9.37)	4.97	1.19 (0.66-2.12)	0.54	1.44 (1.05-1.98)	0.003
CA+AA	73 (46.20)	20 (62.5)	8.20	0.89 (0.65-1.21)	0.45	1.19 (0.85-1.66)	0.01
MGMT + ERCC1	83 (85.57)	14 (14.43)					
CC	66 (79.52)	11 (78.57)	9.33	-	-	-	-
CT + CA	17 (20.48)	3 (21.43)	6.33	1.22 (0.69-2.14)	0.47	1.39 (0.77-2.49)	0.11
Adjusted hazard ratios, 95% confidence intervals and their respective p-values were calculated, after adjusting for age, genotype, histology, KPS, smoking habits and stage. Abbreviations: HR-hazard ratio; CI-confidence intervals.							

After evaluation of 190 male lung cancer patients, it was observed that 158 patients were dead and 32 patients were alive as shown in table 5.15. For subjects having *MGMT* genotype; 120 dead patients had wild type genotype and 38 dead patients had heterozygous type genotype. 28 alive patients had wild type genotype and only 4 patients had heterozygous type genotype. Overall survival period for heterozygous genotype was much lower than wild type genotype (MST=6.17 vs. 9.13 months; HR=1.36; p=0.09). When univariate model was analysed, a non-significant relationship was observed between *MGMT* genotype and overall survival in male lung cancer patients; but multivariate cox regression model showed a significant relationship between *MGMT* genotype and overall survival in male lung cancer patients (adjusted HR= 1.34; 95% CI=0.91-1.97; **p=0.01**).

In the cases of subjects having *ERCC1* genotype; it was observed that 85 dead patients had wild type genotype, 58 dead patients had heterozygous type genotype and 15 patients had mutant type genotype. 12 alive patients had wild type genotype, 17 alive patients had heterozygous type genotype and only 3 patients had mutant type genotype. Overall survival period of heterozygous type genotype was more as compared to wild type genotype (MST=9 vs. 8.3 months; HR=0.83; p=0.28); but overall survival period of mutant type genotype was much lower than wild type genotype (MST=4.97 vs. 8.3 months; HR=1.19; p=0.54). When univariate model was analysed for heterozygous type genotype and mutant type genotype against wild type genotype respectively; a non-significant relationship was observed between overall survival and *ERCC1* in male lung cancer patients. But on applying multivariate cox model for heterozygous type genotype and mutant type genotype against wild type genotype respectively; a significant

relationship was observed between overall survival and *ERCC1* in male lung cancer patients (adjusted HR=1.07; 95% CI=0.75-1.53; **p=0.03** and adjusted HR=1.44; 95% CI=1.05-1.98; **p=0.003** respectively). On combining mutant type genotype with hetero-zygous type genotype and compared it against wild type genotype; overall survival period was observed less as compared to wild type genotype (MST=8.2 vs. 8.3 months; HR=0.89; p=0.45). A non-significant association was also observed between overall survival and *ERCC1* genotype in male lung cancer patients after applying univariate model, but on applying multivariate cox model; relationship observed was significant (adjusted HR=1.19; 95% CI=0.85-1.66; p=0.02).

After evaluation of 97 male patients having common *MGMT* and *ERCC1* genotype, it was observed that 83 patients were dead and 14 patients were alive as shown in table 5.15. Among dead patients (83), 66 patients had wild type genotype whereas only 17 patients had heterozygous type genotype. 11 alive patients had wild type genotype and 3 patients had heterozygous type genotype. Overall survival period of heterozygous type genotype was less as compared to wild type genotype (MST=6.33 vs. 9.33 months; HR= 1.22; p=0.47). When both univariate analysis (Kaplan) and multivariate analysis (cox) were evaluated, a non-significant relationship was found between overall survival in male lung cancer patients having common *MGMT* and *ERCC1* genotype (adjusted HR= 1.39; 95% CI=0.77-2.49; p=0.11).

5.2.10 Logistic regression analysis of lung cancer patients on the basis of tumor stage and its relationship with *MGMT* and *ERCC1* genotype

Patients having wild type genotype were taken as reference to study odds ratio.

Table 5.16 Logistic regression analysis on the basis of tumor stage					
Genotype	Tumor stage		Logistic regression analysis		
<i>MGMT</i>	Stage 3 N (%) 98 (52.41)	Stage 4 N (%) 89 (47.59)	Adjusted Odds ratio	95% CI (Confidence Intervals)	P value
CC	73 (74.49)	69 (77.53)	-	-	-
CT	25 (25.51)	20 (22.47)	0.90	(0.45-1.79)	0.76

<i>ERCC1</i>					
CC	46 (46.94)	49 (55.06)	-	-	-
CA	43 (43.88)	32 (35.95)	0.62	(0.33-1.17)	0.14
AA	9 (9.18)	8 (8.99)	0.86	(0.50-1.49)	0.60
CA+AA	52 (53.06)	40 (44.94)	0.63	(0.34-1.15)	0.13
<i>MGMT + ERCC1</i>	47 (49.47)	48 (50.53)			
CC	35 (74.47)	39 (81.25)	-	-	-
CA+CT	12 (25.53)	9 (18.75)	0.58	(0.20-1.70)	0.32
Unconditional logistic analysis was used to calculate adjusted Odds ratios, 95% confidence intervals and their respective p-values after adjusting for Age, Gender, Smoking, Histology and Genotype					

Regression analysis for association of tumor stage with *MGMT* and *ERCC1* genotype is summarised in table 5.16. Among 187 Subjects having *MGMT* genotype; 98 (52.41%) cases were in stage 3 and 89 (47.59%) cases were in stage 4. 74.49% (73) cases with stage 3 were observed carrying wild type genotype and 25.51% (25) cases of stage 3 were carrying heterozygous type genotype whereas among patients having stage 4 (89), 77.53% (69) cases were harbouring wild type genotype and 22.47% (20) cases were carrying heterozygous type genotype. When regression model was evaluated, non-significant relationship was observed between *MGMT* genotype and tumor stage among lung cancer patients (AOR=0.90; 95% CI=0.45-1.79; p=0.76).

Among patients having *ERCC1* genotype with stage 3, 46.94% (46) cases were carrying wild type genotype, 43.88% (43) cases had heterozygous type genotype and 9.18% (9) cases had mutant type genotype. Among 89 cases having tumor stage 4, 55.06% (49) cases had wild type genotype, 35.95% (32) cases had heterozygous type genotype, and 8.99% (8) cases had mutant type genotype. On evaluating heterozygous type genotype and mutant type genotype against wild type genotype respectively with the help of logistic regression model, a non-significant relationship was observed between *ERCC1* and tumor stage among lung cancer patients (AOR=0.62; 95% CI=0.33-1.17; p=0.14 and AOR=0.86; 95% CI=0.5-1.49; p=0.60

respectively). Since the frequency of mutant type genotype was less, we combined it with heterozygous type genotype and compared it against wild type genotype. On applying logistic regression model; a non-significant relationship was observed between tumor stage and *ERCC1* genotype in lung cancer patients (AOR=0.63; 95% CI=0.34-1.15; p=0.13).

Regression analysis for association of tumor stage for patients having common *MGMT* and *ERCC1* genotype is summarise in table 5.16. Among 95 lung cancer cases, 47 (49.47%) cases were in diagnosed with stage 3 and 48 (50.53%) cases were in stage 4. 74.47% (35) cases diagnosed with stage 3 had wild type genotype and 25.53% (12) cases had heterozygous type genotype whereas among patients having stage 4 (48), 81.25% (39) cases had wild type genotype and 18.75% (9) cases had heterozygous type genotype. When regression model was analysed a non-significant association was observed between patients having common *MGMT* and *ERCC1* genotype and tumor stage (AOR=0.58; 95% CI=0.20-1.70; p=0.32).

5.2.11 Logistic regression analysis of lung cancer patients on the basis of primary tumor extension and its relationship with *MGMT* and *ERCC1* genotype

Table 5.17 Logistic regression analysis on the basis of Primary tumor extension					
Genotype	Primary tumor extension		Logistic regression analysis		
<i>MGMT</i>	Tx+T1+T2	T3+T4	Odds ratio	95% CI	P value
	N	N			
	(%)	(%)			
	40	150			
	(21.05)	(78.95)			
CC	29	115	-	-	-
	(72.5)	(76.67)			
CT	11	35	0.70	(0.31 - 1.57)	0.38
	(27.5)	(23.33)			
<i>ERCC1</i>					
CC	22	75	-	-	-
	(55)	(50)			
CA	16	59	1.08	(0.51 - 2.27)	0.84
	(40)	(39.33)			
AA	2	16	1.45	(0.65 - 3.24)	0.37
	(5)	(10.67)			
CA+AA	18	75	1.22	(0.59 - 2.50)	0.60
	(45)	(50)			
MGMT+	19	77			

ERCCI	(19.79)	(80.21)			
CC	15 (78.95)	60 (77.92)	-	-	-
CT+CA	4 (21.05)	17 (22.08)	1.13	(0.27- 4.70)	0.87
Unconditional logistic analysis was used to calculate adjusted Odds ratios, 95% confidence intervals and their respective p-values after adjusting for age, gender, smoking, histology and genotype					

Regression analysis for association of primary tumor extension with *MGMT* and *ERCCI* genotype was summarised in table 5.17. A total of 190 subjects having *MGMT* genotype were studied for evaluating odds ratio. Since the frequencies of patients having tumor type Tx, T1 and T2 was less so we combined these three tumor type and it included a total of 40 cases; among them 29 cases were wild type genotype and 11 cases were heterozygous type genotype. Similarly we combined all the cases having T3 and T4 tumor type and among these cases (150) 115 cases were wild type genotype and 35 cases were heterozygous type genotype. A non-significant relationship was observed between primary tumor extension and lung cancer patients having *MGMT* genotype (AOR= 0.70; 95% CI=0.31-1.57; p=0.38).

On combining all the cases having Tx, T1 and T2 tumor type; we observed 22 patients had wild type genotype, 16 patients carrying heterozygous type genotype and only 2 patients had mutant type genotype. We also combined all the cases having T3 and T4 tumor type and observed 75 cases had wild type genotype, 59 cases had heterozygous type genotype and 16 patients carrying mutant type genotype. A non-significant relationship was observed between *ERCCI* and primary tumor extension among lung cancer patients carrying heterozygous type genotype and mutant type genotype respectively (AOR=1.08; 95% CI=0.51-2.27; p=0.84 and AOR=1.45; 95% CI=0.65-3.24; p=0.37 respectively). Since the frequency of mutant type genotype was less, we combined it with heterozygous type genotype and compared it against wild type genotype. On applying logistic regression model; a non-significant relationship was observed between primary tumor extension and *ERCCI* genotype in lung cancer patients (AOR=1.22; 95% CI=0.59-2.50; p=0.60).

Regression analysis for association of primary tumor extension for patients having common *MGMT* and *ERCCI* genotype is summarise in table 5.17. Among 96 lung cancer cases, 19 cases were combination of tumor type Tx, T1 and T2 whereas 77 cases were combination of tumor type T3 and T4. Among combination of Tx, T1 and T2 tumor type cases, 15 cases were carrying

wild type genotype and only 4 cases had heterozygous type genotype. Out of combination of cases having T3 and T4 tumor type; 60 cases had wild type genotype and 17 cases had heterozygous type genotype. When regression model was analysed a non-significant association was observed between patients having common *MGMT* and *ERCC1* genotype and primary tumor extension (AOR=1.13; 95% CI=0.27-4.70; p=0.87).

5.2.12 Logistic regression analysis of lung cancer patients on the basis of lymph node invasion and its relationship with *MGMT* and *ERCC1* genotype

Table 5.18 Logistic regression analysis on the basis of lymph node invasion					
Genotype	Lymph node invasion		Overall survival		
<i>MGMT</i>	N0+N1 N (%)	N2+N3+N4 N (%)	Odds ratio	95% CI	P value
	41 (21.58)	149 (78.42)			
CC	30 (73.17)	114 (76.51)	-	-	-
CT	11 (26.83)	35 (23.49)	0.81	(0.36 - 1.79)	0.59
<i>ERCC1</i>					
CC	26 (63.41)	71 (47.65)	-	-	-
CA	11 (26.83)	64 (42.95)	2.17	(0.98 - 4.79)	0.06
AA	4 (9.76)	14 (9.40)	1.17	(0.63 - 2.17)	0.61
CA+AA	15 (36.59)	78 (52.35)	1.92	(0.93 - 3.96)	0.08
<i>MGMT + ERCC1</i>	31 (30.39)	71 (69.61)			
CC	25 (80.65)	55 (77.46)	-	-	-
CT, CA	6 (19.35)	16 (22.54)	1.52	(0.48 - 4.76)	0.48
Unconditional logistic analysis was used to calculate adjusted Odds ratios, 95% confidence intervals and their respective p-values after adjusting for age, gender, smoking, histology and genotype					

Regression analysis for association of lymph node with *MGMT* and *ERCC1* is summarised in table 5.18. A total of 190 subjects having *MGMT* genotype were studied for evaluating odds ratio. Since the frequencies of patients having involvement of N0 and N1 lymph node was less

so we combined these two types of lymph node involvement and it included a total of 41 cases; among them 30 cases had wild type genotype and 11 cases had heterozygous type genotype. Similarly we combined all the cases having involvement of N2, N3 and N4 lymph node and among these cases (149) 114 cases had wild type genotype and 35 cases had heterozygous type genotype. A non-significant relationship was observed between lymph node invasion and lung cancer patients having *MGMT* genotype (AOR= 0.81; 95% CI=0.36-1.79; p=0.59).

On combining all the cases having involvement of N0 and N1 lymph node; we observed 26 patients had wild type genotype, 11 patients carrying heterozygous type genotype and only 4 patients had mutant type genotype. We also combined all the cases having involvement of N2, N3 and N4 lymph node and observed 71 cases had wild type genotype, 64 cases had heterozygous type genotype and 14 patients carrying mutant type genotype. A non-significant relationship was observed between *ERCC1* and lymph node invasion among lung cancer patients carrying heterozygous type genotype and mutant type genotype respectively (AOR=2.17; 95% CI= 0.98-4.80; p=0.06 and AOR=1.17; 95% CI=0.63-2.17; p=0.61 respectively). Since the frequency of mutant type genotype was less, we combined it with heterozygous type genotype and compared it against wild type genotype. On applying logistic regression model; a non-significant relationship was observed between lymph node invasion and *ERCC1* genotype in lung cancer patients (AOR=1.92; 95% CI=0.93-3.96; p=0.08).

Regression analysis for association of lymph node invasion for patients having common *MGMT* and *ERCC1* genotype is summarise in table 5.18. Among 102 lung cancer cases, 31 cases were combination of involvement of N0 and N1 lymph node whereas 71 cases were combination of involvement of N2, N3 and N4 lymph node. Among combination of involvement of N0 and N1 lymph node cases, 25 cases were carrying wild type genotype and only 6 cases had heterozygous type genotype. Out of combination of cases having involvement of N2, N3 and N4 lymph node; 55 cases had wild type genotype and 16 cases had heterozygous type genotype. When regression model was analysed a non-significant association was observed between patients having common *MGMT* and *ERCC1* genotype and lymph node invasion (AOR=1.52; 95% CI=0.48-4.76; p=0.48).

5.2.13 Logistic regression analysis of lung cancer patients on the basis of metastasis and its relationship with *MGMT* and *ERCC1* genotype

Table 5.19 Logistic regression analysis on the basis of metastasis					
Genotype	Metastasis		Logistic regression		
<i>MGMT</i>	M0 N (%) 105 (55.26)	M1 N (%) 85 (44.74)	Odds ratio	95% CI	P value
CC	78 (74.29)	66 (77.65)	-	-	-
CT	27 (25.71)	19 (22.35)	0.90	(0.44 - 1.82)	0.76
<i>ERCC1</i>					
CC	49 (46.67)	48 (56.47)	-	-	-
CA	45 (42.86)	30 (35.29)	0.54	(0.28 - 1.05)	0.07
AA	11 (10.48)	7 (8.24)	0.69	(0.39 - 1.24)	0.21
CA+AA	56 (53.33)	37 (43.53)	0.53	(0.28 - 0.98)	0.04
<i>MGMT + ERCC1</i>	50 (52.08)	46 (47.92)			
CC	37 (74)	38 (82.61)	-	-	-
CT+CA	13 (26)	8 (17.39)	0.47	(0.15 - 1.46)	0.19
Unconditional logistic analysis was used to calculate adjusted Odds ratios, 95% confidence intervals and their respective p-values after adjusting for age, gender, smoking status, histology and genotype					

Regression analysis for association of metastasis with *MGMT* and *ERCC1* is summarised in table 5.19. A total of 190 subjects having *MGMT* genotype were studied for evaluating odds ratio. Among patients who did not show metastasis (105), 78 patients had wild type genotype and 27 patients had heterozygous type genotype. Out of 85 patients who showed distant metastasis, 66 carrying wild type genotype and 19 cases had heterozygous type genotype. A non-significant relationship was observed between metastasis and lung cancer patients having *MGMT* genotype (AOR= 0.90; 95% CI=0.44-1.82; p=0.76).

Out of 105 patients who did not show metastasis and having *ERCC1* genotype, 49 patients were carrying wild type genotype, 45 patients had heterozygous type genotype and 11 patients had mutant type genotype. Among patients who showed metastasis, 48 patients had wild type genotype, 30 patients had heterozygous type genotype and only 7 patients had mutant type genotype. A non-significant relationship was observed between *ERCC1* and metastasis among lung cancer patients carrying heterozygous type genotype and mutant type genotype respectively (AOR=0.54; 95% CI= 0.28-1.05; p=0.07 and AOR=0.69; 95% CI=0.39-1.24; p=0.21 respectively). Since the frequency of mutant type genotype was less, we combined it with heterozygous type genotype and compared it against wild type genotype. On applying logistic regression model; a non-significant relationship was observed between metastasis and *ERCC1* genotype in lung cancer patients (AOR=0.53; 95% CI=0.28-0.98; p=0.04).

Regression analysis for association of metastasis for patients having common *MGMT* and *ERCC1* genotype is summarise in table 5.19. Among 96 lung cancer cases, 50 cases did not show metastasis and 46 cases showed metastasis. Out of cases which did not show metastasis, 37 patients had wild type genotype and 13 patients had heterozygous type genotype. Among cases which showed metastasis, 38 cases had wild type genotype and only 8 patients had heterozygous type genotype. When regression model was analysed a non-significant association was observed between patients having common *MGMT* and *ERCC1* genotype and metastasis (AOR=0.47; 95% CI=0.15-1.46; p=0.19).

5.2.14 Logistic regression analysis of lung cancer patients on the basis of ECOG and its relationship with *MGMT* and *ERCC1* genotype

Table 5.20 Logistic regression analysis on the basis of ECOG					
Genotype	ECOG		Logistic regression analysis		
<i>MGMT</i>	0,1 N (%) 101 (49.03)	2,3,4 N (%) 105 (50.97)	Odds ratio	95% CI	P value
CC	74 (73.27)	86 (81.90)	-	-	-
CT	27 (26.73)	19 (18.10)	0.61	(0.31 - 1.18)	0.14

<i>ERCC1</i>					
CC	49 (48.51)	53 (50.48)	-	-	-
CA	42 (41.58)	42 (40)	0.89	(0.50 - 1.61)	0.71
AA	10 (9.90)	10 (9.52)	1.02	(0.62 - 1.68)	0.94
CA+AA	52 (51.49)	52 (49.52)	0.93	(0.53 - 1.62)	0.79
<i>MGMT + ERCC1</i>	48 (46.60)	55 (53.40)			
CC	36 (75)	45 (81.82)	-	-	-
CT+CA	12 (25)	10 (18.18)	0.58	(0.21 - 1.58)	0.28
Unconditional logistic analysis was used to calculate adjusted Odds ratios, 95% confidence intervals and their respective p-values after adjusting for age, gender, smoking status, histology and genotype					

Regression analysis for association of ECOG with *MGMT* and *ERCC1* is summarised in table 5.20. A total of 206 subjects having *MGMT* genotype were studied for evaluating odds ratio. Since the frequencies of patients having ECOG value of 0 and 1 was less so we combined these two types of ECOG values and it included a total of 101 cases; among them 74 cases had wild type genotype and 27 cases had heterozygous type genotype. Similarly we combined all the cases having ECOG size of 2, 3 and 4; and among these cases (105) 86 cases had wild type genotype and 19 cases had heterozygous type genotype. A non-significant relationship was observed between ECOG and lung cancer patients having *MGMT* genotype (AOR= 0.61; 95% CI=0.31-1.18; p=0.14).

On combining all the cases having ECOG size of 0 and 1; we observed 49 patients had wild type genotype, 42 patients carrying heterozygous type genotype and only 10 patients had mutant type genotype. We also combined all the cases having ECOG range from 2-4 and observed 53 cases had wild type genotype, 42 cases had heterozygous type genotype and 10 patients carrying mutant type genotype. A non-significant relationship was observed between *ERCC1* and ECOG among lung cancer patients carrying heterozygous type genotype and mutant type genotype respectively (AOR=0.89; 95% CI= 0.5-1.6; p=0.71 and AOR=1.02; 95% CI=0.62-1.68; p=0.94 respectively). Since the frequency of mutant type genotype was less, we combined it with heterozygous type genotype and compared it against wild type genotype. On applying logistic

regression model; a non-significant relationship was observed between ECOG and *ERCC1* genotype in lung cancer patients (AOR=0.93; 95% CI=0.53-1.62; p=0.79).

Regression analysis for association of ECOG for patients having common *MGMT* and *ERCC1* genotype is summarised in table 5.20. Among 103 lung cancer cases, 48 cases were combination of ECOG value of 0 and 1 whereas 55 cases were combination of ECOG size of 2-4. Among combination of cases having ECOG value of 0 and 1, 36 cases were carrying wild type genotype and only 12 cases had heterozygous type genotype. Out of combination of cases having ECOG size of 2-4; 45 cases had wild type genotype and 10 cases had heterozygous type genotype. When regression model was analysed a non-significant association was observed between patients having common *MGMT* and *ERCC1* genotype and ECOG (AOR=0.58; 95% CI=0.21-1.58; p=0.28).

5.2.15 Logistic regression analysis of lung cancer patients on the basis of KPS and its relationship with *MGMT* and *ERCC1* genotype

Table 5.21 Logistic regression analysis on the basis of KPS					
Genotype	KPS		Logistic regression analysis		
<i>MGMT</i>	80-100 N (%) 130 (63.11)	Below 80 N (%) 76 (36.89)	Odds ratio	95% CI	P value
CC	93 (71.54)	67 (88.16)	-	-	-
CT	37 (28.46)	9 (11.84)	0.33	(0.15 - 0.74)	0.007
<i>ERCC1</i>					
CC	64 (49.23)	38 (50)	-	-	-
CA	54 (41.54)	30 (39.47)	0.97	(0.53 - 1.79)	0.93
AA	12 (9.23)	8 (10.53)	1.09	(0.66 - 1.80)	0.74
CA+AA	66 (50.77)	38 (50)	1.05	(0.59 - 1.87)	0.87
<i>MGMT</i> + <i>ERCC1</i>	65 (63.11)	38 (36.89)			
CC	48 (73.85)	33 (86.84)	-	-	-

CT+CA	17 (26.15)	5 (13.16)	0.45	(0.15 -1.36)	0.16
Unconditional logistic analysis was used to calculate adjusted Odds ratios, 95% confidence intervals and their respective p-values after adjusting for age, gender, smoking status, histology and genotype					

A total of 206 subjects having *MGMT* genotype were studied for evaluating odds ratio. Among patients who had KPS range of 80-100 (130), 93 patients had wild type genotype and 37 patients had heterozygous type genotype. Out of 76 patients who had KPS value of below 80, 67 carrying wild type genotype and only 9 cases had heterozygous type genotype. A significant relationship was observed between KPS and lung cancer patients having *MGMT* genotype (AOR= 0.33; 95% CI=0.15-0.74; **p=0.007**).

Out of 130 patients who had KPS size of 80-100 and having *ERCC1* genotype, 64 patients were carrying wild type genotype, 54 patients had heterozygous type genotype and 12 patients had mutant type genotype. Among patients who had KPS value below 80, 38 patients had wild type genotype, 30 patients had heterozygous type genotype and only 8 patients had mutant type genotype. A non-significant relationship was observed between *ERCC1* and KPS among lung cancer patients carrying heterozygous type genotype and mutant type genotype respectively (AOR=0.97; 95% CI= 0.53-1.79; p=0.93 and AOR=1.09; 95% CI=0.66-1.80; p=0.74 respectively). Since the frequency of mutant type genotype was less, we combined it with heterozygous type genotype and compared it against wild type genotype. On applying logistic regression model; a non-significant relationship was observed between KPS and *ERCC1* genotype in lung cancer patients (AOR=1.05; 95% CI=0.59-1.87; p=0.87).

Among 103 lung cancer cases having common *MGMT* and *ERCC1* genotype, 65 cases had KPS value of 80-100 and 38 cases had KPS range below 80. Out of cases which had KPS range of 80-100, 48 patients had wild type genotype and 17 patients had heterozygous type genotype. Among cases which had KPS value below 80, 33 cases had wild type genotype and only 5 patients had heterozygous type genotype. When regression model was analysed a non-significant association was observed between patients having common *MGMT* and *ERCC1* genotype and KPS (AOR=0.45; 95% CI=0.15-1.36; p=0.16).

5.2.16 Logistic regression analysis of lung cancer patients on the basis of clinical response and its relationship with *MGMT* and *ERCC1* genotype

Table 5.22 Logistic regression analysis on the basis of clinical response					
Genotype	Clinical response		Logistic regression analysis		
<i>MGMT</i>	CR/PR N (%)	SD/PD N (%)	Odds ratio	95% CI	P value
	67 (57.26)	50 (42.74)			
CC	54 (80.60)	37 (74)	-	-	-
CT	13 (19.40)	13 (26)	1.57	(0.65 - 3.82)	0.32
<i>ERCC1</i>					
CC	35 (52.24)	23 (46)	-	-	-
CA	26 (38.81)	22 (44)	1.19	(0.53 - 2.64)	0.67
AA	6 (8.96)	5 (10)	0.97	(0.46 - 2.03)	0.93
CA+AA	32 (47.76)	27 (54)	1.17	(0.54 - 2.52)	0.69
<i>MGMT</i> + <i>ERCC1</i>	36 (60)	24 (40)			
CC	29 (80.56)	18 (75)	-	-	-
CT,CA	7 (19.44)	6 (25)	1.1623	(0.3253 - 4.1531)	0.8170
Unconditional logistic analysis was used to calculate adjusted Odds ratios, 95% confidence intervals and their respective p-values after adjusting for age, gender, smoking status, histology and genotype.					

A total of 117 subjects having *MGMT* genotype were studied for evaluating odds ratio. Among patients who did response to the administered drug (67), 54 patients had wild type genotype and only 3 patients had heterozygous type genotype. Out of 50 patients who did not response to administered drug; 37 carrying wild type genotype and only 13 cases had heterozygous type genotype. A non-significant relationship was observed between clinical response and lung cancer patients having *MGMT* genotype (AOR= 1.57; 95% CI=0.65-3.82; p=0.32).

Out of 67 patients who had partial or clear response to the administer drug and having *ERCC1* genotype, 35 patients were carrying wild type genotype, 26 patients had heterozygous type

genotype and 6 patients had mutant type genotype. Among patients who did not response to drug, 23 patients had wild type genotype, 22 patients had heterozygous type genotype and only 5 patients had mutant type genotype. A non-significant relationship was observed between *ERCC1* and clinical response among lung cancer patients carrying heterozygous type genotype and mutant type genotype respectively (AOR=1.19; 95% CI= 0.53-2.64; p=0.67 and AOR=0.97; 95% CI=0.46-2.03; p=0.93 respectively). Since the frequency of mutant type genotype was less, we combined it with heterozygous type genotype and compared it against wild type genotype. On applying logistic regression model; a non-significant relationship was observed between clinical response and *ERCC1* genotype in lung cancer patients (AOR=1.17; 95% CI=0.54-2.52; p=0.69).

Among 60 lung cancer cases having common *MGMT* and *ERCC1* genotype, 36 cases did response to drug and 24 cases did not response to drug. Out of cases which did response to drug, 29 patients had wild type genotype and 7 patients had heterozygous type genotype. Among cases which did not response to drug, 18 cases had wild type genotype and only 6 patients had heterozygous type genotype. When regression model was analysed a non-significant association was observed between patients having common *MGMT* and *ERCC1* genotype and clinical response (AOR=1.16; 95% CI=0.33-4.15; p=0.82).

5.3 Computational analysis of gene polymorphism in *MGMT* and *ERCC1* gene by using multiple bioinformatics tools

Total number of SNP for *MGMT* and *ERCC1* was 68838, 16635 respectively (<https://www.ncbi.nlm.nih.gov/snp>) and their classification on the basis of functional class was described in table 5.43.

Table 5.23 Functional class of gene polymorphism of MGMT and ERCC1 gene			
Sr. No.	Functional class	SNP of <i>MGMT</i> gene	SNP of <i>ERCC1</i> gene
1	3' splice site	2	5
2	3'Utr	995	1525
3	5'splice site	6	13
4	5'Utr	105	632
5	Coding synonymous	98	411

6	Intron	66872	15256
7	Mis-sense	199	792
8	Non-sense	13	35
9	stop gained	13	28
Note: One <i>SNP</i> may have more than one functional class and some <i>SNP</i> may have unknown functional class.			

SIFT Software

Among 68838 SNP of *MGMT* there were only 9 SNP whose data was present on SIFT software and for rest SNP no data was available. Among 9 SNPs, it was observed only three SNP were deleterious and six were tolerated. Among 16635 SNP of *ERCC1* there were only 11 SNP whose data was present on SIFT software and for rest SNP no data was available. Among 11 SNPs, only three SNP were deleterious and eight were tolerated as shown in table 5.24.

Table 5.24 Prediction of deleterious SNP by using SIFT and PROVEAN						
Sr. No.	rs ID	Gene	Amino acid change	Score	SIFT Prediction	PROVEAN Prediction
1	rs3212986*	<i>ERCC1</i>	Q504K	1	Tolerated	Non-Protein Coding
2	rs3212989	<i>ERCC1</i>	T282A	0.73	Tolerated	Non-Protein Coding
3	rs735482	<i>ERCC1</i>	K259T	0.59	Tolerated	Non-Protein Coding
4	rs762562	<i>ERCC1</i>	K373E	1	Tolerated	Protein Coding
5	rs3188420	<i>ERCC1</i>	P77H	0.05	Damaging	Non-Protein Coding
6	rs3212977	<i>ERCC1</i>	V242T	0.02	Damaging	Non-Protein Coding
7	rs35448603	<i>ERCC1</i>	P484L	0.2	Tolerated	Non-Protein Coding
8	rs112438131	<i>ERCC1</i>	A372V	0.42	Tolerated	Protein Coding
9	rs112625502	<i>ERCC1</i>	E277T	0.55	Tolerated	Non-Protein Coding
10	rs113047593	<i>ERCC1</i>	E225K	0.55	Tolerated	Non-Protein Coding
11	rs117289933	<i>ERCC1</i>	F12L	0.01	Damaging	Non-Protein Coding
12	rs12917*	<i>MGMT</i>	L84F	0.62	Tolerated	Non-Protein Coding
13	rs2020893	<i>MGMT</i>	Q61K	0.85	Tolerated	Non-Protein Coding
14	rs228164	<i>MGMT</i>	R96C	0.01	Damaging	Non-Protein Coding
15	rs2308318	<i>MGMT</i>	G160R	0.01	Damaging	Non-Protein Coding
16	rs2308320	<i>MGMT</i>	E166D	0.11	Tolerated	Non-Protein Coding
17	rs2308321	<i>MGMT</i>	I143V	0.86	Tolerated	Non-Protein Coding

18	rs2308327	<i>MGMT</i>	K178R	0.25	Tolerated	Non-Protein Coding
19	rs3750824	<i>MGMT</i>	R128Q	0.05	Damaging	Non-Protein Coding
20	rs2308322	<i>MGMT</i>	F89S	0.37	Tolerated	Non-Protein Coding
Note: rs3212986 (<i>ERCC1</i>) and rs12917 (<i>MGMT</i>) both were tolerated and lies in non-protein coding region according to prediction of SIFT and PROVEAN respectively						

PROVEAN

All the SNP of the *MGMT* were observed outside of non-protein coding region. Among 11SNP of *ERCC1*, only two SNP were observed inside the protein coding region and rest were observed in non-protein coding as shown in table 5.24.

Polyphen 2

Among 68838 SNP of *MGMT* there were only 8 SNP whose data was present on Polyphen 2 software and for rest SNP no data was available. Among 8 SNPs, only three SNPs were possibly damaging, one was probably damaging and four were Benign (tolerated). Among 16635 SNP of *ERCC1* there were only 5 SNP whose data was present on Polyphen 2 software and for rest SNP no data was available. Among 5 SNPs, only one SNP was probably damaging and four were Benign (tolerated) as shown in table 5.25.

Table 5.25 Prediction of deleterious SNP by using Polyphen 2

Sr No	rs ID	Gene	Score	Sensitivity	Specificity	Prediction
1	rs3212986	<i>ERCC1</i>	0.199	0.88	0.73	Benign
2	rs3212989	<i>ERCC1</i>	0.014	0.95	0.53	Benign
3	rs735482	<i>ERCC1</i>	0	1	0	Benign
4	rs762562	<i>ERCC1</i>	0	1	0	Benign
5	rs3212977	<i>ERCC1</i>	0.992	0.49	0.95	<u>Probably Damaging</u>
6	rs12917	<i>MGMT</i>	0.456	0.83	0.8	Possibly Damaging
7	rs2020893	<i>MGMT</i>	0.007	0.97	0.46	Benign
8	rs228164	<i>MGMT</i>	0	1	0	Possibly damaging
9	rs2308318	<i>MGMT</i>	0.931	0.67	0.91	<u>Probably damaging</u>
10	rs2308320	<i>MGMT</i>	0.335	0.85	0.77	Benign
11	rs2308321	<i>MGMT</i>	0.007	0.97	0.46	Benign
12	rs2308327	<i>MGMT</i>	0	1	0	Benign
13	rs2308322	<i>MGMT</i>	0.451	0.83	0.8	Possibly damaging
Note: According to prediction of Polyphen 2, rs3212986 (<i>ERCC1</i>) was benign (tolerated) in nature while rs12917 (<i>MGMT</i>) was damaging in nature.						

I-mutant vs 2.0

Among 9 SNP of *MGMT* gene, all the SNPs were observed within reliability index and had decreased stability of proteins. Among 11 SNPs of *ERCC1*, 10 SNPs were observed within reliability index and 7 SNPs had decreased protein stability whereas 4 SNPs had increased protein stability as shown in table 5.26.

Table 5.26 Prediction of protein stability by I-mutant and MuPro							
			I-Mutant Prediction			MuPro Prediction	
Sr no	rs ID	Gene	RI	DDG	Stability	DDG	Stability
1	rs3212986*	<i>ERCC1</i>	5	-0.29	Decrease	-0.76	Decrease
2	rs3212989	<i>ERCC1</i>	9	-1.91	Decrease	-1.46	Decrease
3	rs735482	<i>ERCC1</i>	6	0.84	Increase	-0.56	Decrease
4	rs762562	<i>ERCC1</i>	1	0.4	Increase	-0.1	Decrease
5	rs3188420	<i>ERCC1</i>	6	-0.54	Decrease	-0.9	Decrease
6	rs3212977	<i>ERCC1</i>	9	-1.41	Decrease	-2.11	Decrease
7	rs35448603	<i>ERCC1</i>	1	1.26	Increase	0.07	Increase
8	rs112438131	<i>ERCC1</i>	0	0.36	Increase	-0.28	Decrease
9	rs112625502	<i>ERCC1</i>	7	-0.03	Decrease	-1.54	Decrease
10	rs113047593	<i>ERCC1</i>	7	-0.81	Decrease	-0.33	Decrease
11	rs117289933	<i>ERCC1</i>	2	-0.26	Decrease	-0.51	Decrease
12	rs12917*	<i>MGMT</i>	6	-0.53	Decrease	-1.32	Decrease
13	rs2020893	<i>MGMT</i>	1	-0.06	Decrease	-0.78	Decrease
14	rs228164	<i>MGMT</i>	5	-1.09	Decrease	-0.69	Decrease
15	rs2308318	<i>MGMT</i>	7	-1.17	Decrease	-0.94	Decrease
16	rs2308320	<i>MGMT</i>	6	-0.27	Decrease	-0.91	Decrease
17	rs2308321	<i>MGMT</i>	7	-0.58	Decrease	-0.77	Decrease
18	rs2308327	<i>MGMT</i>	4	-0.56	Decrease	-0.07	Decrease
19	rs3750824	<i>MGMT</i>	3	-0.36	Decrease	-1.03	Decrease
20	rs2308322	<i>MGMT</i>	8	-2.31	Decrease	-1.38	Decrease
Note: rs3212986 (<i>ERCC1</i>) and rs12917 (<i>MGMT</i>) both were observed in reliability index and had decreased protein stability according to both the softwares. Abbreviation: RI- reliability index; DDG: Difference in delta G of mutant genotype and wild genotype.							

MuPro

Among 9 SNP of *MGMT* gene, all the SNPs had decreased stability of proteins. Among 11 SNPs of *ERCC1*, 10 SNPs had decreased protein stability but only one SNP had increased protein stability as shown in table 5.26.

PhD-SNP

Among 9 SNP of *MGMT* gene, all the SNPs were observed within reliability index and 5 SNPs had deleterious effect and 4 SNPs were neutral. Among 11 SNPs of *ERCC1*, 10 SNPs were observed within reliability index and all the SNP had neutral effect as shown in table 5.27.

Table: 5.27 Prediction of deleterious effect of SNP by using PhD-SNP and SNP&GO

			PhD-SNP Prediction		SNP&GO Prediction	
Sr no	rs ID	Gene	RI	Prediction	RI	Prediction
1	rs3212986*	<i>ERCC1</i>	10	Neutral	10	Neutral
2	rs3212989	<i>ERCC1</i>	8	Neutral	9	Neutral
3	rs735482	<i>ERCC1</i>	8	Neutral	10	Neutral
4	rs762562	<i>ERCC1</i>	6	Neutral	9	Neutral
5	rs3188420	<i>ERCC1</i>	7	Neutral	8	Neutral
6	rs3212977	<i>ERCC1</i>	1	Neutral	4	Neutral
7	rs35448603	<i>ERCC1</i>	9	Neutral	10	Neutral
8	rs112438131	<i>ERCC1</i>	9	Neutral	10	Neutral
9	rs112625502	<i>ERCC1</i>	6	Neutral	8	Neutral
10	rs113047593	<i>ERCC1</i>	6	Neutral	9	Neutral
11	rs117289933	<i>ERCC1</i>	5	Neutral	8	Neutral
12	rs12917*	<i>MGMT</i>	5	Neutral	9	Neutral
13	rs2020893	<i>MGMT</i>	3	Disease	7	Neutral
14	rs228164	<i>MGMT</i>	7	Disease	4	Neutral
15	rs2308318	<i>MGMT</i>	7	Disease	4	Disease
16	rs2308320	<i>MGMT</i>	4	Neutral	0	Neutral
17	rs2308321	<i>MGMT</i>	8	Neutral	9	Neutral
18	rs2308327	<i>MGMT</i>	6	Neutral	9	Neutral
19	rs3750824	<i>MGMT</i>	2	Disease	2	Disease
20	rs2308322	<i>MGMT</i>	8	Disease	4	Neutral

Note: rs3212986 (*ERCC1*) and rs12917 (*MGMT*), both were observed in reliability index and had negligible deleterious effect according to PhD-SNP and SNP & GO software.

Abbreviation: RI- reliability index.

SNP & GO

Among 9 SNP of *MGMT* gene, 8 SNPs were observed within reliability index and 2 SNPs had deleterious effect and 7 SNPs were neutral. Among 11 SNPs of *ERCC1*, 7 SNPs were observed within reliability index and all the SNP had neutral deleterious effect as shown in table 5.27.

Elastic

Among 9 SNP of *MGMT* gene, 8 SNPs were observed within protein domain and only one SNP was observed outside of protein domain. Among 11 SNPs of *ERCC1*, 10 SNPs were observed outside of protein domain and only one SNP was observed within protein domain as shown in table 5.28.

Table 5.28 Prediction of changes in protein structure in SNP by Elastic

Sr no	rs ID	Gene	DDG	Prediction
1	rs3212986*	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
2	rs3212989	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
3	rs735482	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
4	rs762562	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
5	rs3188420	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
6	rs3212977	<i>ERCC1</i>	0.915	Mutation falls Inside of protein domain
7	rs35448603	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
8	rs112438131	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
9	rs112625502	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
10	rs113047593	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
11	rs117289933	<i>ERCC1</i>	-	Mutation falls Outside of protein domain
12	rs12917*	<i>MGMT</i>	-0.097	Mutation falls Inside of protein domain
13	rs2020893	<i>MGMT</i>	0.808	Mutation falls Inside of protein domain
14	rs228164	<i>MGMT</i>	2.044	Mutation falls Inside of protein domain
15	rs2308318	<i>MGMT</i>	-0.627	Mutation falls Inside of protein domain
16	rs2308320	<i>MGMT</i>	0.664	Mutation falls Inside of protein domain
17	rs2308321	<i>MGMT</i>	0.376	Mutation falls Inside of protein domain
18	rs2308327	<i>MGMT</i>	-	Mutation falls Outside of protein domain
19	rs3750824	<i>MGMT</i>	0.363	Mutation falls Inside of protein domain
20	rs2308322	<i>MGMT</i>	2.76	Mutation falls Inside of protein domain

Note: rs3212986 (*ERCC1*), mutation was observed outside of the protein domain while in rs12917 (*MGMT*) mutation was observed inside of the protein and the protein became unstable

For rs12917 (*MGMT*)

Mutation was observed within protein domain of *MGMT* at position 84 from leucine to phenylalanine as shown in figure 5.38 and the sky blue colored protein structure was of wild type genotype while grey colored protein structure was of mutant type genotype as shown in figure 5.39.

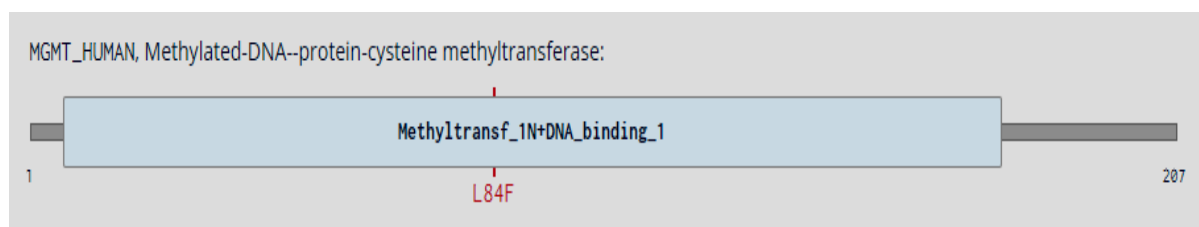


Figure 5.38 Mutation position for rs12917

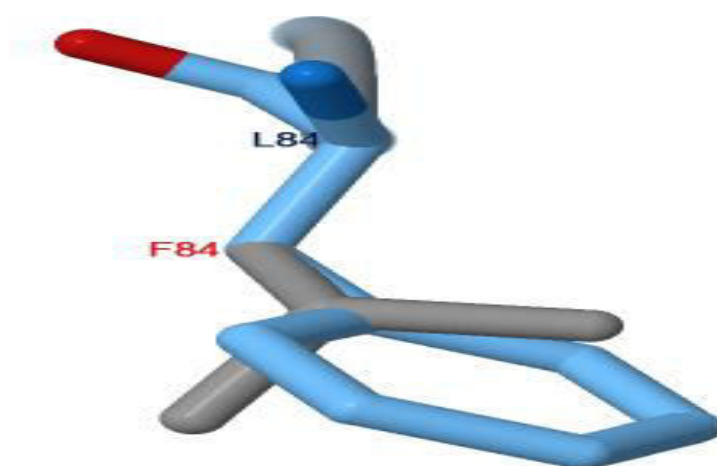


Figure 5.39 Protein structure of wild and mutant type genotype of *MGMT* (rs12917)

For rs2020893 (*MGMT*)

Mutation was observed within protein domain of *MGMT* at position 61 from glutamine to lysine as shown in figure 5.40 and the sky blue colored protein structure was of wild type genotype while grey colored protein structure was of mutant type genotype as shown in figure 5.41.

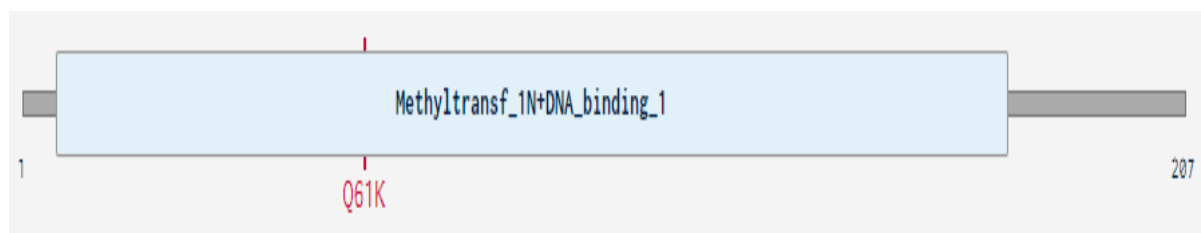


Figure 5.40 Mutation position for rs2020893

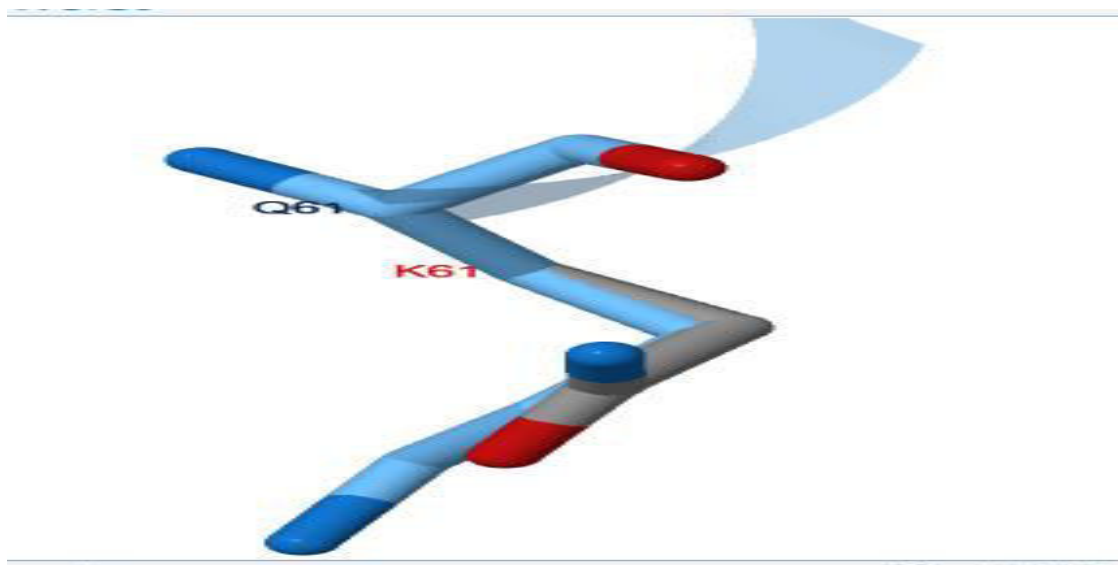


Figure 5.41 Protein structure of wild and mutant type genotype of *MGMT* (rs2020893)

For rs228164 (*MGMT*)

Mutation was observed within protein domain of *MGMT* at position 96 from arginine to cysteine as shown in figure 5.42 and the sky blue colored protein structure was of wild type genotype while grey colored protein structure was of mutant type genotype as shown in figure 5.43.

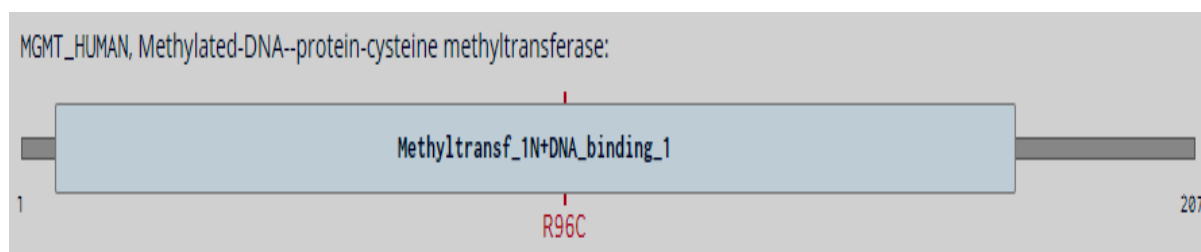


Figure 5.42 Mutation position for rs228164

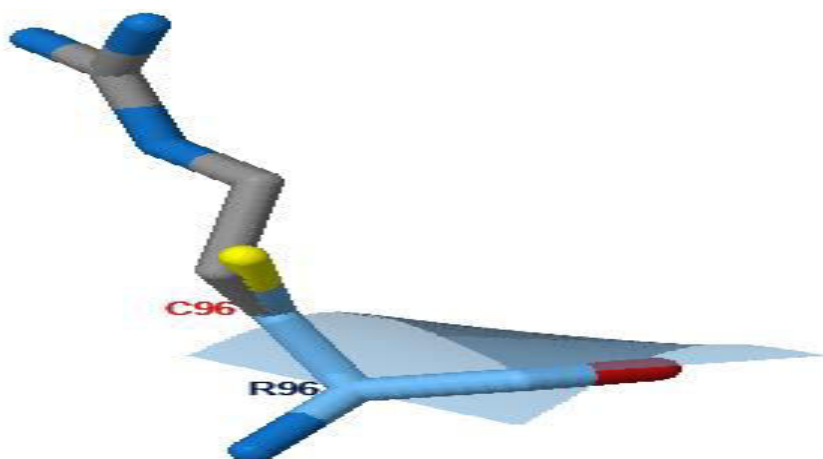


Figure 5.43 Protein structure of wild and mutant type genotype of *MGMT* (rs228164)

For rs2308318 (*MGMT*)

Mutation was observed within protein domain of *MGMT* at position 160 from glycine to arginine as shown in figure 5.44 and the sky blue colored protein structure was of wild type genotype while grey colored protein structure was of mutant type genotype as shown in figure 5.45.

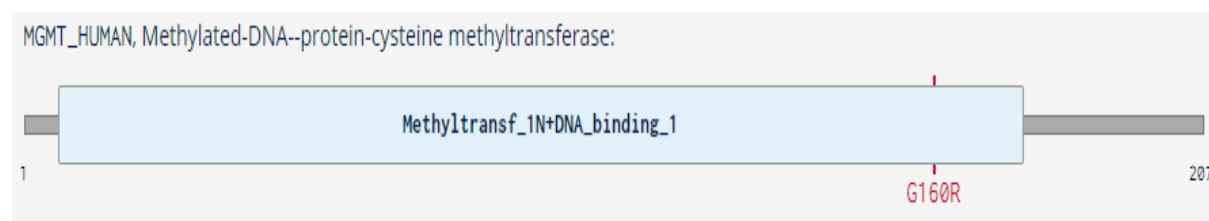


Figure 5.44 Mutation position for rs2308318

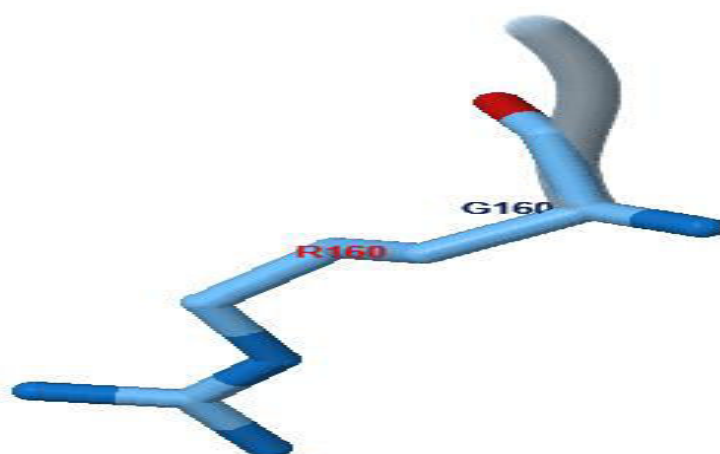


Figure 5.45 Protein structure of wild and mutant type genotype of *MGMT* (rs2308318)

For rs2308320 (*MGMT*)

Mutation was lies within protein domain of *MGMT* at position 166 from glutamic acid to aspartic acid as shown in figure 5.46 and the sky blue colored protein structure was of wild type genotype while grey colored protein structure was of mutant type genotype as shown in figure 5.47.

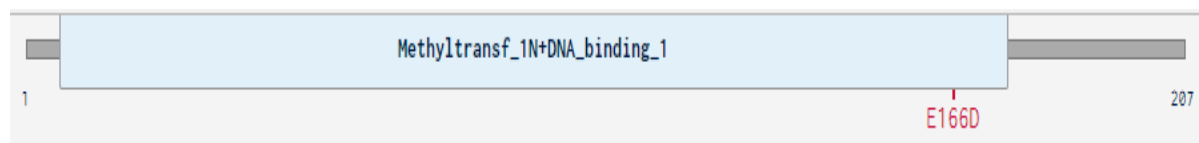


Figure 5.46 Mutation position of rs2308320

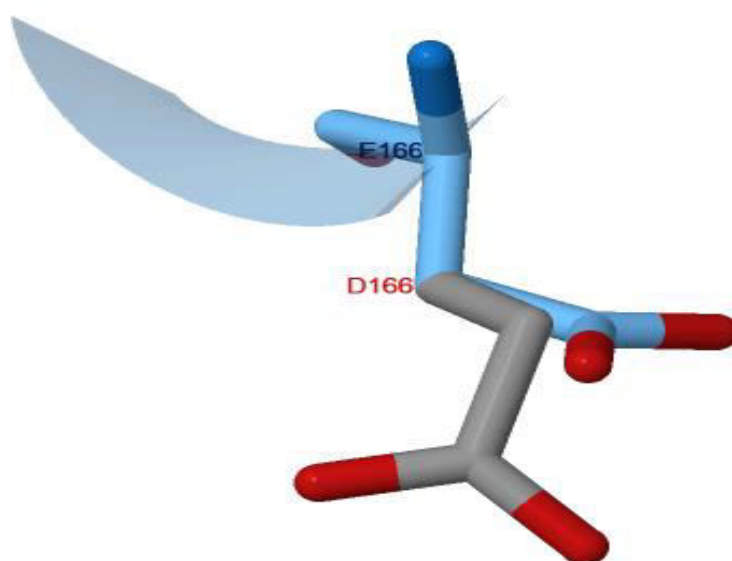


Figure 5.47 Protein structure of wild type and mutant type genotype of *MGMT* (rs2308320)

For rs2308321 (*MGMT*)

Mutation was observed within protein domain of *MGMT* at position 143 from isoleucine to valine as shown in figure 5.48 and the sky blue colored protein structure was of wild type genotype while grey colored protein structure was of mutant type genotype as shown in figure 5.49.

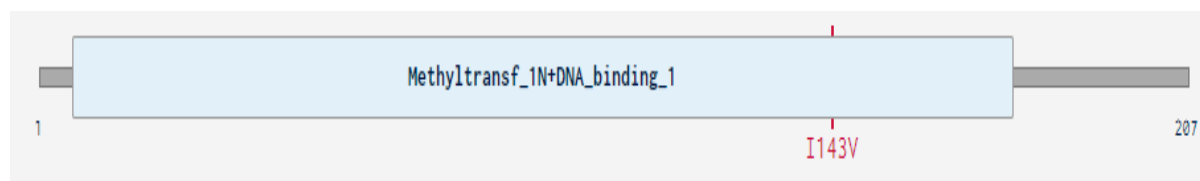


Figure 5.48 Mutation position of rs2308321

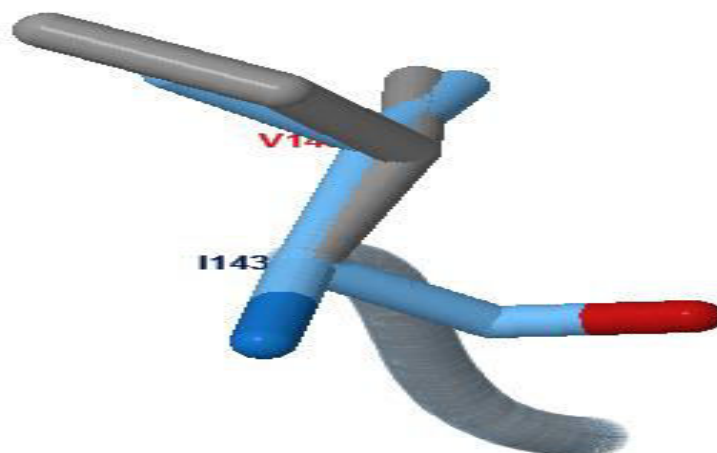


Figure 5.49 Protein structure of wild and mutant type genotype of *MGMT* (rs2308321)

For rs3750824 (*MGMT*)

Mutation was observed within protein domain of *MGMT* at position 128 from arginine to glutamine as shown in figure 5.50 and the sky blue colored protein structure was of wild type genotype while grey colored protein structure was of mutant type genotype as shown in figure 5.51.

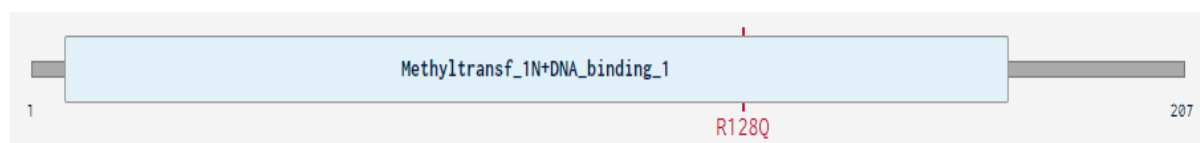


Figure 5.50 Mutation position of rs3750824



Figure 5.51 Protein structure of wild and mutant genotype of *MGMT* (rs3750824)

For rs2308322 (*MGMT*)

Mutation was observed within protein domain of *MGMT* at position 89 from phenylalanine to serine as shown in figure 5.52 and the sky blue colored protein structure was of wild type genotype while grey colored protein structure was of mutant type genotype as shown in figure 5.53.

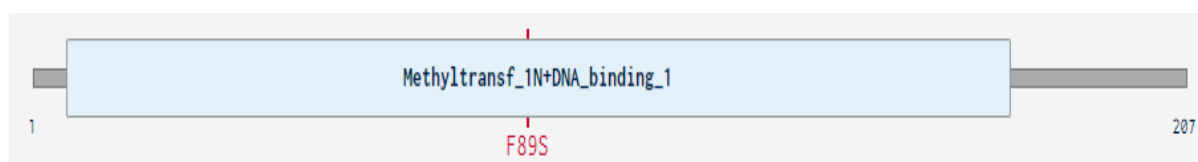


Figure 5.52 Mutation position of rs2308322

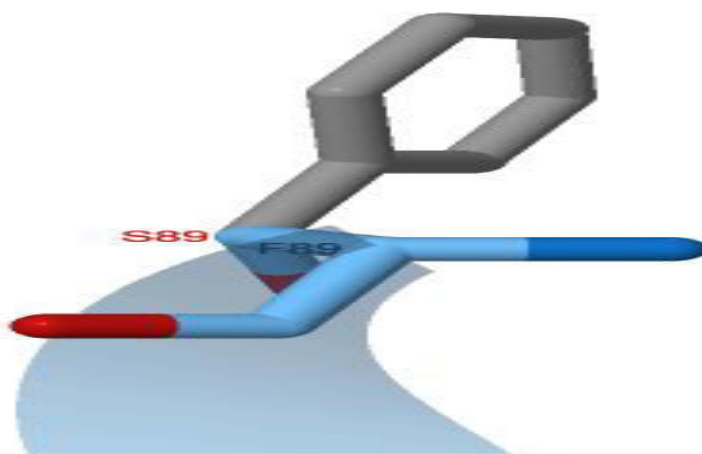


Figure 5.53 Protein structure of wild and mutant type genotype of *MGMT* (rs2308322)

For rs3212977 (*ERCC1*)

Mutation was observed within protein domain of *ERCC1* at position 242 from valine to threonine as shown in figure 5.54 and the sky blue colored protein structure was of wild type genotype while rough portioned protein structure was of mutant type genotype as shown in figure 5.55.

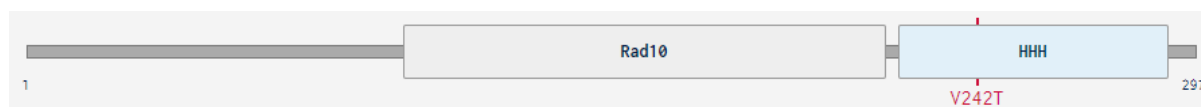


Figure 5.54 Mutation position of rs3212977

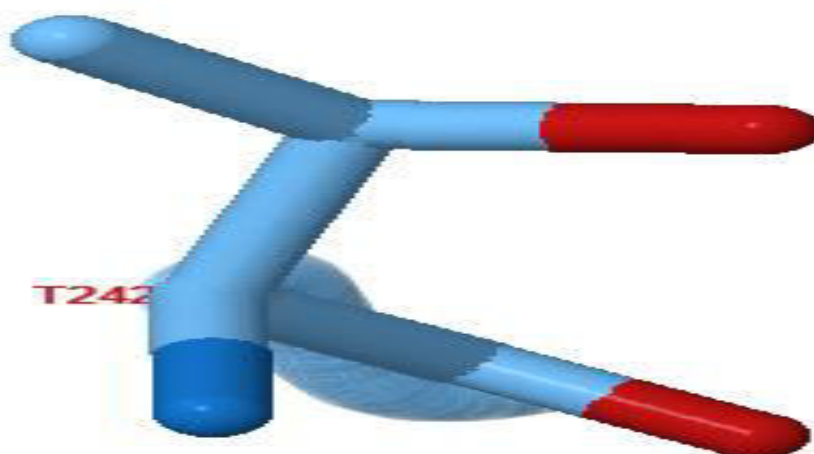


Figure 5.55 Protein structure of wild and mutant type genotype of *ERCC1* (rs3212977)

PANTHER

Among 9 SNP of *MGMT* gene, 5 SNPs were observed possibly damaging, three were probably damaging and only one SNP was probably benign. Among 11 SNPs of *ERCC1*, 3 SNPs were probably damaging and 8 SNP were probably benign as shown in table 5.25.

Table 5.29 Prediction of deleterious effect of SNP by Panther and MutPred					
			Panther	MutPred prediction	
Sr no	rs ID	Gene	Prediction	Score	Prediction
1	rs3212986*	<i>ERCC1</i>	Probably benign	0.03	non-pathogenic
2	rs3212989	<i>ERCC1</i>	Probably benign	0.027	non-pathogenic

3	rs735482	<i>ERCC1</i>	Probably benign	0.037	non-pathogenic
4	rs762562	<i>ERCC1</i>	Probably benign	0.026	non-pathogenic
5	rs3188420	<i>ERCC1</i>	Probably damaging	0.185	non-pathogenic
6	rs3212977	<i>ERCC1</i>	Probably damaging	0.576	pathogenic
7	rs35448603	<i>ERCC1</i>	Probably benign	0.034	non-pathogenic
8	rs112438131	<i>ERCC1</i>	Probably benign	0.02	non-pathogenic
9	rs112625502	<i>ERCC1</i>	Probably benign	0.08	non-pathogenic
10	rs113047593	<i>ERCC1</i>	Probably benign	0.105	non-pathogenic
11	rs117289933	<i>ERCC1</i>	Probably damaging	0.322	non-pathogenic
12	rs12917*	<i>MGMT</i>	<u>Possibly damaging</u>	0.097	non-pathogenic
13	rs2020893	<i>MGMT</i>	<u>Possibly damaging</u>	0.294	non-pathogenic
14	rs228164	<i>MGMT</i>	<u>Possibly damaging</u>	0.469	non-pathogenic
15	rs2308318	<i>MGMT</i>	Probably damaging	0.921	pathogenic
16	rs2308320	<i>MGMT</i>	<u>Possibly damaging</u>	0.496	non-pathogenic
17	rs2308321	<i>MGMT</i>	<u>Possibly damaging</u>	0.115	non-pathogenic
18	rs2308327	<i>MGMT</i>	Probably benign	0.132	non-pathogenic
19	rs3750824	<i>MGMT</i>	Probably damaging	0.893	pathogenic
20	rs2308322	<i>MGMT</i>	Probably damaging	0.232	non-pathogenic
Note: rs3212986 (<i>ERCC1</i>), this SNP was probably benign and rs12917 (<i>MGMT</i>), this SNP was possibly damaging according to PANTHER and both were non-pathogenic according to MutPred.					

MutPred

Among 9 SNP of *MGMT* gene, 2 SNPs were pathogenic, seven were non-pathogenic in nature. Among 11 SNPs of *ERCC1*, only one SNPs was pathogenic and 10 SNP were pathogenic as shown in table 5.25.

Mutation 3-D

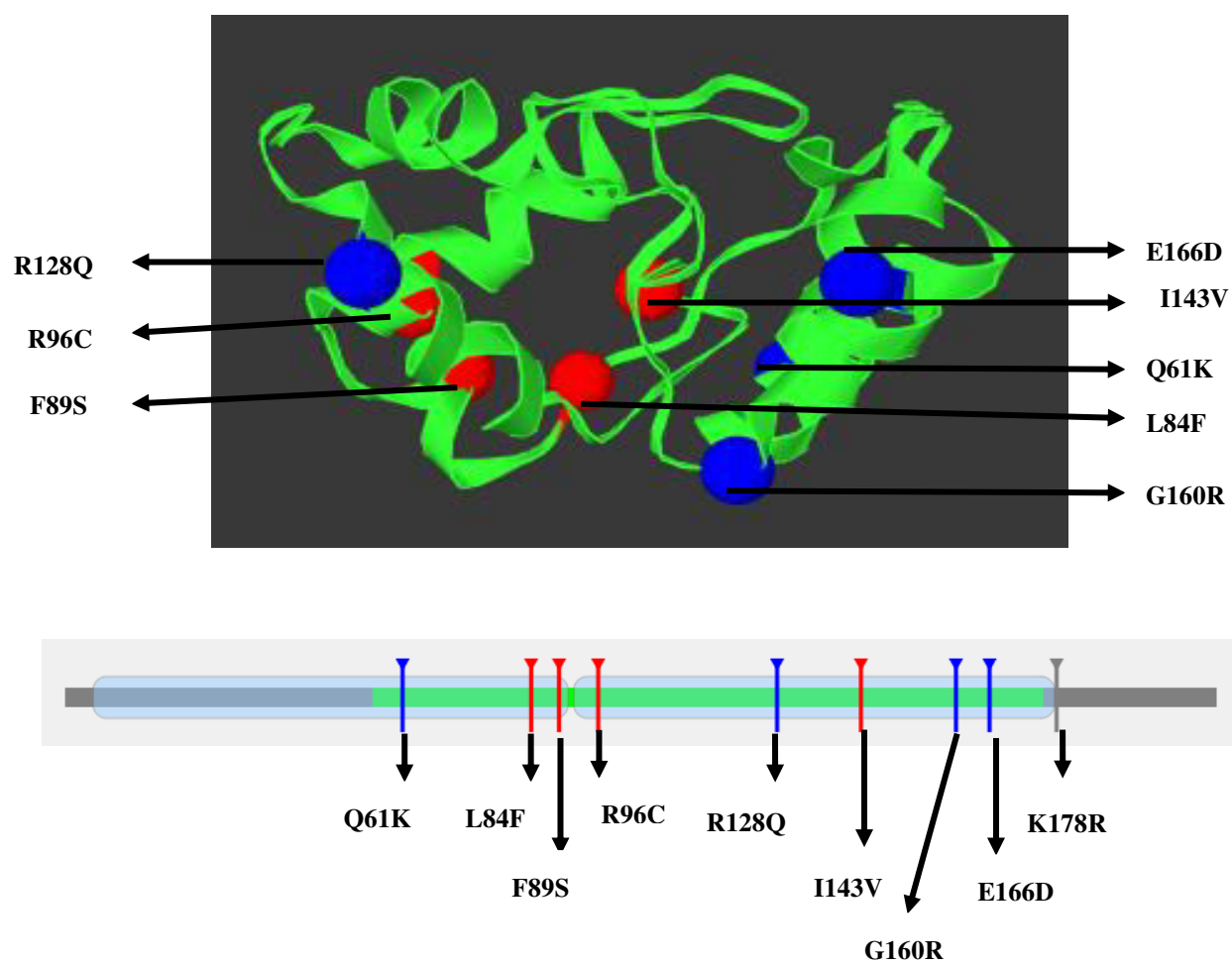


Figure 5.56 Position of mutations in protein structure of *MGMT*

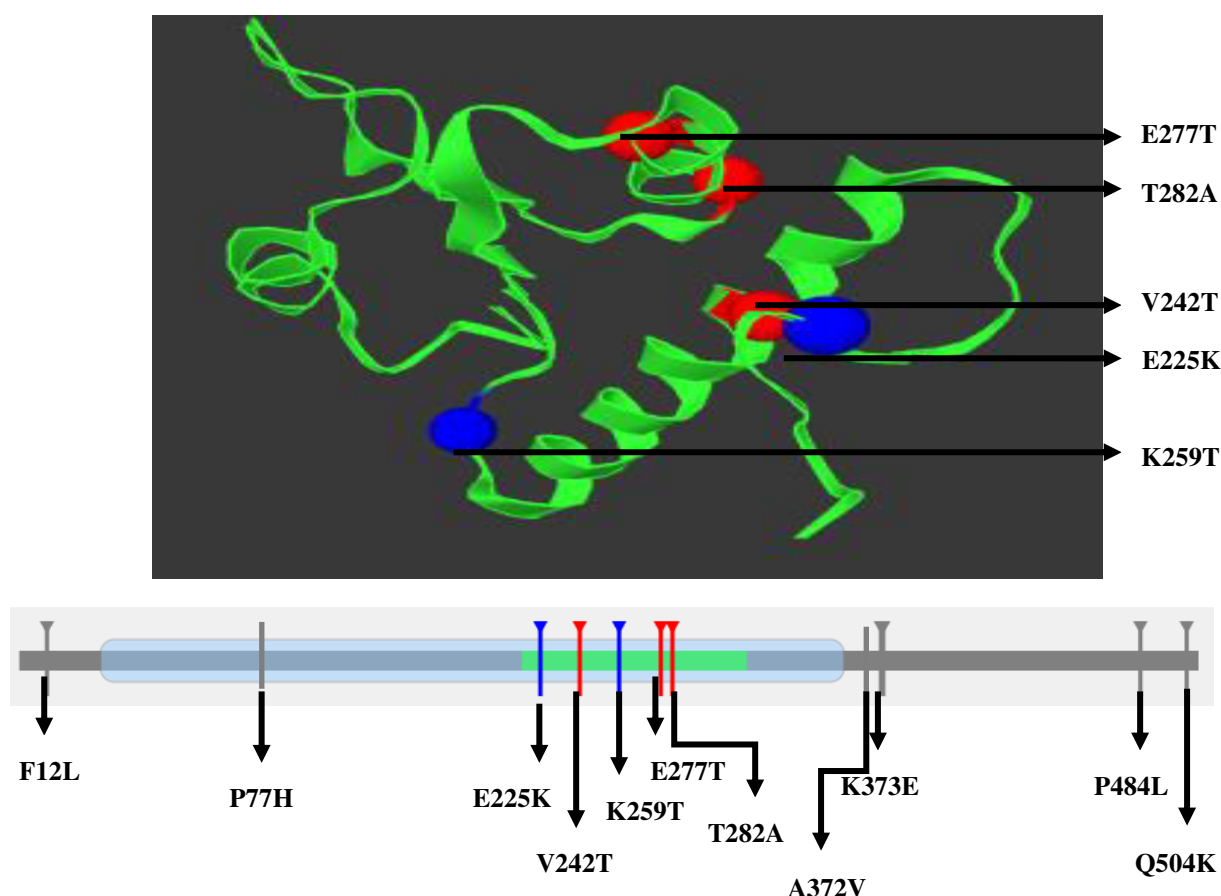


Figure 5.57 Position of mutations in protein structure of *ERCCI*

Out of 9 SNP of *MGMT* only one SNP (K178R) was observed outside of protein domain as shown in figure 5.56. Out of 11 SNP of *ERCCI* only five SNPs were observed within the protein domain and rest were observed outside of protein domain as shown in figure 5.57.

Genemania

Interaction of *MGMT* gene with other genes was summarised in table 5.56 and shown in figure 5.58.

Table 5.30 Interaction of <i>MGMT</i> gene with other genes						
Sr. No.	Network	Physical Interaction	Co-expression	Predicted	Co-localization	Shared Protein Domain
1	<i>MGMT</i>	<i>MGMT</i>	<i>MGMT</i>	<i>MGMT</i>	<i>MGMT</i>	-
2	<i>PRSS18</i>	-	-	<i>PRSS18</i>	-	<i>PRSS18</i>

3	<i>SIGIRR</i>	-	<i>SIGIRR</i>	-	-	-
4	<i>TECR</i>	-	<i>TECR</i>	-	-	-
5	<i>HIST1 H2AM</i>	<i>HIST1H2AM</i>	-	-	-	<i>HIST1H2AM</i>
6	<i>ABHD14B</i>	-	<i>ABHD14 B</i>	<i>ABHD14B</i>	<i>ABHD14B</i>	<i>ABHD14B</i>
7	<i>TRAPP8A</i>	-	<i>TRAPPC8 A</i>	-	-	-
8	<i>PABPC3</i>	<i>PABPC3</i>	<i>PABPC3</i>	-	-	-
9	<i>ECHS1</i>	-	<i>ECHS1</i>	<i>ECHS1</i>	<i>ECHS1</i>	-
10	<i>H2BFM</i>	<i>H2BFM</i>	-	-	-	<i>H2BFM</i>
11	<i>TTPA</i>	<i>TTPA</i>	<i>TTPA</i>	-	<i>TTPA</i>	-
12	<i>GALT</i>	-	<i>GALT</i>	<i>GALT</i>	<i>GALT</i>	-
13	<i>OSGEPL1</i>	-	-	<i>OSGEPL1</i>	-	-
14	<i>HAGH</i>	-	<i>HAGH</i>	<i>HAGH</i>	<i>HAGH</i>	-
15	<i>CDKL5</i>	<i>CDKL5</i>	<i>CDKL5</i>	-	-	<i>CDKL5</i>
16	<i>CDK12</i>	<i>CDK12</i>	<i>CDK12</i>	<i>CDK12</i>	-	<i>CDK12</i>
17	<i>TSTA3</i>	-	<i>TSTA3</i>	-	-	-
18	<i>GPN2</i>	-	<i>GPN2</i>	<i>GPN2</i>	-	-
19	<i>PEX13</i>	-	-	<i>PEX13</i>	-	-
20	<i>EP300</i>	<i>EP300</i>	<i>EP300</i>	<i>EP300</i>	-	-
21	<i>WNT9B</i>	<i>WNT9B</i>	-	-	-	-

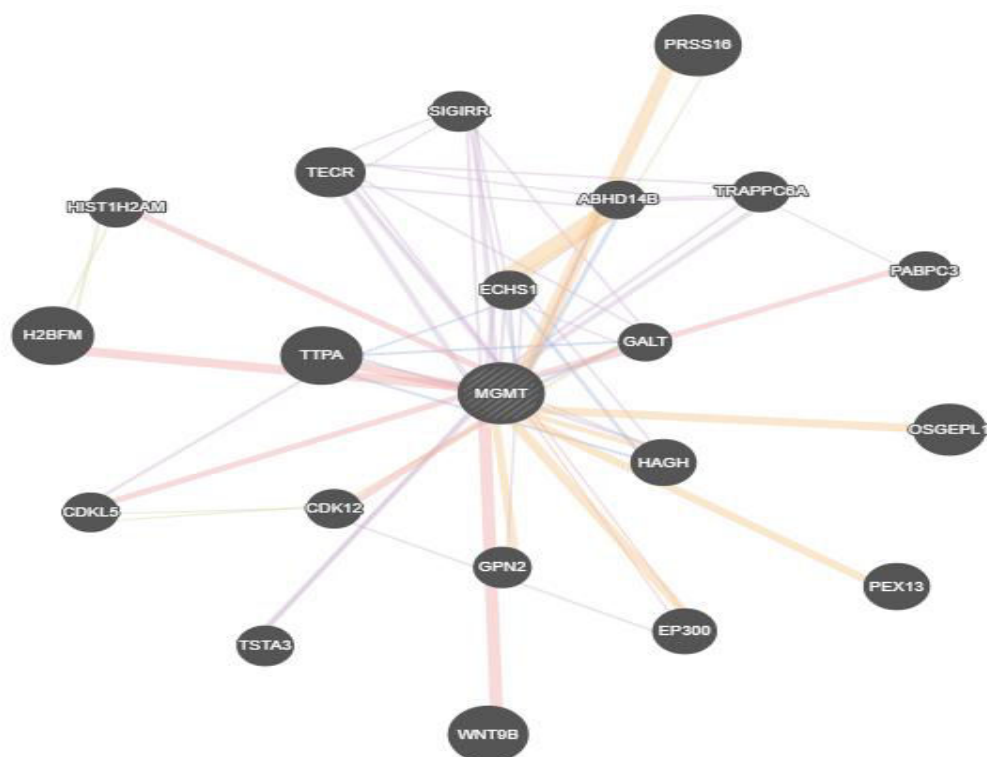


Figure 5.58 *MGMT* gene interactions with other genes

Interaction of *ERCC1* gene with other genes was summarised in table 5.27 and shown in figure 5.59.

Table 5.31 Interaction of <i>ERCC1</i> with other genes							
Sr. No.	Network	Physical Interaction	Co-expression	Predicted	Co-localization	Pathway	Shared Protein Domain
1	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>
2	<i>TWF2</i>	<i>TWF2</i>	<i>TWF2</i>	<i>TWF2</i>	-	-	-
3	<i>C15orf40</i>	<i>C15orf40</i>	<i>C15orf40</i>	-	-	-	-
4	<i>MUTYH</i>	<i>MUTYH</i>	<i>MUTYH</i>	-	-	-	<i>MUTYH</i>
5	<i>RAD52</i>	<i>RAD52</i>	<i>RAD52</i>	<i>RAD52</i>	-	-	-
6	<i>MGME1</i>	-	<i>MGME1</i>	-	<i>MGME1</i>	-	<i>MGME1</i>
7	<i>TOPBP1</i>	<i>TOPBP1</i>	<i>TOPBP1</i>	-	-	-	-
8	<i>NTHL1</i>	<i>NTHL1</i>	<i>NTHL1</i>	<i>NTHL1</i>	-	-	<i>NTHL1</i>
9	<i>ERCC4</i>	<i>ERCC4</i>	-	<i>ERCC4</i>	-	<i>ERCC4</i>	<i>ERCC4</i>

10	<i>ERCC5</i>	<i>ERCC5</i>	<i>ERCC5</i>	<i>ERCC5</i>	-	<i>ERCC5</i>	-
11	<i>XPA</i>	<i>XPA</i>	<i>XPA</i>	<i>XPA</i>	<i>XPA</i>	<i>XPA</i>	-
12	<i>XPC</i>	<i>XPC</i>	<i>XPC</i>	<i>XPC</i>	<i>XPC</i>	<i>XPC</i>	-
13	<i>DDB2</i>	<i>DDB2</i>	<i>DDB2</i>	-	<i>DDB2</i>	<i>DDB2</i>	<i>DDB2</i>
14	<i>DDB1</i>	<i>DDB1</i>	<i>DDB1</i>	<i>DDB1</i>	<i>DDB1</i>	<i>DDB1</i>	-
15	<i>RAD23B</i>	<i>RAD23B</i>	<i>RAD23B</i>	<i>RAD23B</i>	-	<i>RAD23B</i>	-
16	<i>TCEA1</i>	-	<i>TCEA1</i>	-	-	<i>TCEA1</i>	-
17	<i>XAB2</i>	<i>XAB2</i>	<i>XAB2</i>	<i>XAB2</i>	-	<i>XAB2</i>	-
18	<i>ERCC8</i>	<i>ERCC8</i>	<i>ERCC8</i>	<i>ERCC8</i>	-	<i>ERCC8</i>	<i>ERCC8</i>
19	<i>ERCC6</i>	<i>ERCC6</i>	-	<i>ERCC6</i>	<i>ERCC6</i>	<i>ERCC6</i>	-
20	<i>SLX4</i>	<i>SLX4</i>	-	-	-	-	-
21	<i>BCKDHA</i>	-	<i>BCKDHA</i>	-	<i>BCKDHA</i>	-	-

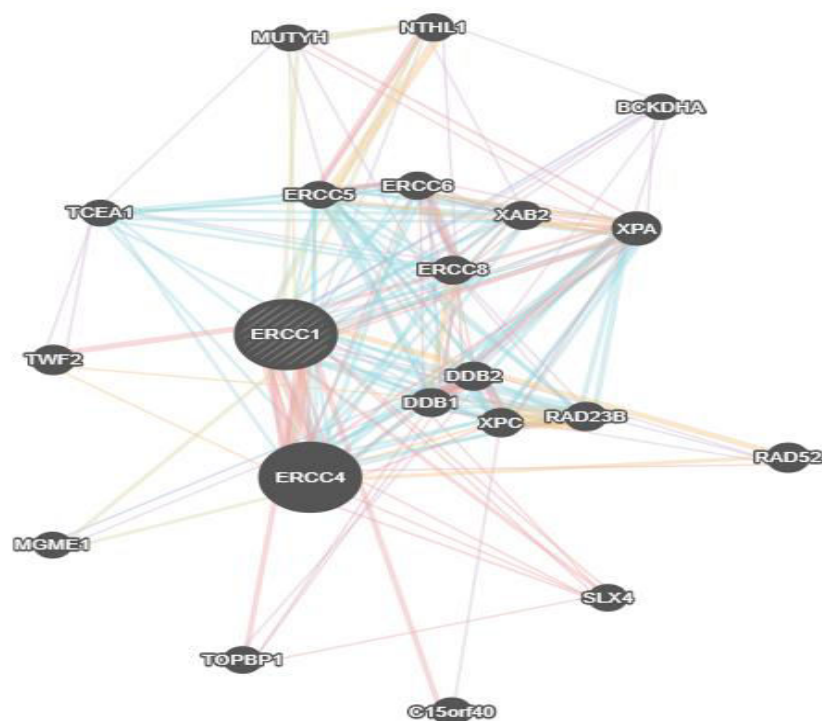


Figure 5.59 *ERCC1* gene interaction with other genes

Interaction of *MGMT* and *ERCC1* with each other and with other genes was summarised in table 5.28 and shown in figure 5.60

Table 5.32 Interaction of <i>MGMT</i> and <i>ERCC1</i> with each other and with other genes							
Sr. No.	Network	Physical Interaction	Co-expression	Predicted	Co-localization	Pathway	Shared Protein Domain
1	<i>MGMT</i>	-	<i>MGMT</i>	-	-	-	-
2	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>	<i>ERCC1</i>
3	<i>TWF2</i>	<i>TWF2</i>	<i>TWF2</i>	<i>TWF2</i>	-	-	-
4	<i>C15orf40</i>	<i>C15orf40</i>	<i>C15orf40</i>	-	-	-	-
5	<i>MUTYH</i>	<i>MUTYH</i>	<i>MUTYH</i>	-	-	-	<i>MUTYH</i>
6	<i>RAD52</i>	<i>RAD52</i>	<i>RAD52</i>	<i>RAD52</i>	-	-	-
7	<i>MGME1</i>	-	<i>MGME1</i>	-	<i>MGME1</i>	-	<i>MGME1</i>
8	<i>TOPBP1</i>	<i>TOPBP1</i>	<i>TOPBP1</i>	-	-	-	-
9	<i>NTHL1</i>	<i>NTHL1</i>	<i>NTHL1</i>	<i>NTHL1</i>	-	-	<i>NTHL1</i>

10	<i>ERCC4</i>	<i>ERCC4</i>	-	<i>ERCC4</i>	-	<i>ERCC4</i>	<i>ERCC4</i>
11	<i>ERCC5</i>	<i>ERCC5</i>	<i>ERCC5</i>	<i>ERCC5</i>	-	<i>ERCC5</i>	-
12	<i>XPA</i>	<i>XPA</i>	<i>XPA</i>	<i>XPA</i>	<i>XPA</i>	<i>XPA</i>	-
13	<i>XPC</i>	<i>XPC</i>	<i>XPC</i>	<i>XPC</i>	<i>XPC</i>	<i>XPC</i>	-
14	<i>DDB2</i>	<i>DDB2</i>	<i>DDB2</i>	-	<i>DDB2</i>	<i>DDB2</i>	<i>DDB2</i>
15	<i>DDB1</i>	<i>DDB1</i>	<i>DDB1</i>	<i>DDB1</i>	<i>DDB1</i>	<i>DDB1</i>	-
16	<i>RAD23B</i>	<i>RAD23B</i>	<i>RAD23B</i>	<i>RAD23B</i>	-	<i>RAD23B</i>	-
17	<i>TCEA1</i>	-	<i>TCEA1</i>	-	-	<i>TCEA1</i>	-
18	<i>XAB2</i>	<i>XAB2</i>	<i>XAB2</i>	<i>XAB2</i>	-	<i>XAB2</i>	-
19	<i>ERCC8</i>	<i>ERCC8</i>	<i>ERCC8</i>	<i>ERCC8</i>	-	<i>ERCC8</i>	<i>ERCC8</i>
20	<i>ERCC6</i>	<i>ERCC6</i>	-	<i>ERCC6</i>	<i>ERCC6</i>	<i>ERCC6</i>	-
21	<i>SLX4</i>	<i>SLX4</i>	-	-	-	-	-
	<i>BCKDHA</i>	-	<i>BCKDHA</i>	-	<i>BCKDHA</i>	-	-

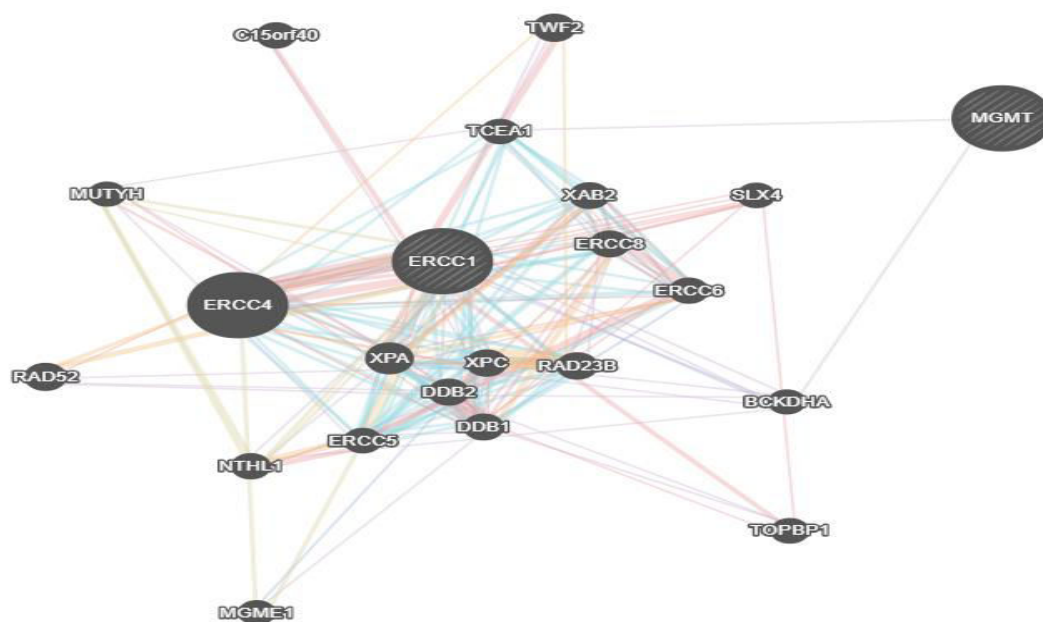


Figure 5.60 Gene interaction of *MGMT* and *ERCC1* with each other and with other genes

Chapter 6

Discussion

Gene polymorphism in *MGMT* and *ERCC1* with rsID as rs12917 and rs3212986 respectively were studied individually and in combination; for prognostic significance of a given platinum based doublet chemotherapy against the 214 lung cancer patients in north Indian population for the first time. For this work various parameters were considered like non-clinical parameters (age, smoking status, gender and pack years of smoking), pathological and clinical parameters (tumor stage, histology, tumor size extension, metastasis and lymph node involvement), clinical response parameters (complete response, partial response, stable disease and partial disease) and other parameters (Regimen, KPOS and ECOG).

A large number of data has been published about association of risk of cancer with the gene polymorphism in *MGMT* gene (rs12917). According to Li *et al.*, (2005) risk of bladder cancer in southern Chinese population due to gene polymorphism in *MGMT* (rs12917) was high in mutant type genotype as compared to wild type genotype among aged and non-drinkers when comparing along with other SNPs among 167 bladder cancer patients and 204 normal people. Parameters included in this study were age, gender, smoking status and use of alcohol. But when rs12917 individually was studied then, no risk of bladder cancer was observed. In the study of Molina *et al.*, (2013), comparison of DNA adduct (caused by hydrocarbons) with rs12917 (*MGMT*) was studied in 93 people of Mexican city. High level of DNA adduct was detected in mutant type genotype as compared to wild type genotype and logistic regression analysis also supported this observation.

Another study by Loh *et al.*, (2011) included 192 cancer cases and 1155 controls; indicated non-significant association of rs12917 (*MGMT*) with any cancer incidence. Meta-analysis of 33 studies by Liu *et al.*, (2017) showed high risk of glioma was associated with rs12917 (*MGMT*) and same was also reported in study of Felini *et al.*, (2007). Further meta analysis of 23 studies conducted by Hu *et al.*, (2014); T allele of heterozygous type genotype of rs12917 (*MGMT*) was native to reduced risk of colorectal cancer in Asian population.

Bye *et al.*, (2011) reported association of risk of oesophageal squamous cell carcinoma in two South African populations (mixed and black). For black population no cancer risk was observed for rs12917 and for mixed ancestry population a significant relationship was observed between cancer and rs12917 (*MGMT*). Meta-analysis of 98 study subjects by Du *et al.*, (2013) showed that mutant type genotype of rs12917 (*MGMT*) was linked with high risk

of cancer as compared to wild type genotype by 31 % (when studied in subgroups, it is highly related with risk of esophageal cancer among Caucasian population). This study also mentioned that rs12917 (*MGMT*) was a compelling risk factor for cancer.

According to meta-analysis of 14 studies conducted by Hung *et al.*, (2008) a non-significant association was observed between rs12917 (*MGMT*) and non-small cell lung cancer. Study of Zhang *et al.*, (2008); revealed rs12917 (*MGMT*) was linked with biliary tract cancer in Chinese population. Further according to Jiao *et al.*, (2006); it was also associated with risk of B-cell lymphoma and follicular lymphoma.

No data has been published regarding prognostic significance of gene polymorphism in *MGMT* in response to platinum based doublet chemotherapy but a few studies have been reported about prognostic significance of gene polymorphism in *MGMT* in response to chemotherapy other than platinum based doublet chemotherapy. In the study of Cao *et al.*, (2015), prognostic significance of *MGMT* gene polymorphism in 445 gastric cancer patients in china treated with fluorouracil-based adjuvant chemotherapy was evaluated. Mutant type genotype at stage II did response to the treatment as compared to wild type genotype and patients with Stage III.

In the current study, for the first time 214 lung cancer patients of North India treated with platinum based doublet chemotherapy were taken as study subjects to evaluate the prognostic implication of *MGMT* gene polymorphism. Our study did not report any prognostic compelling relationship of overall survival with lung cancer patients having *MGMT* genotype on applying univariate and logistic regression models. But when multivariate cox model was analysed then a significant relationship was observed between overall survival and male lung cancer patients who smoked and carrying *MGMT* genotype. Also a non-significant relationship was observed between clinical response and patients having *MGMT* genotype who were administered with various chemotherapeutic regimen like docetaxel + cisplatin/ carboplatin, irinotecan + cisplatin/carboplatin or pemetrexed + cisplatin/carboplatin.

Similarly a lot of data has been published on prognostic significance of *ERCC1* gene polymorphism (rs3212986) in response to platinum based doublet chemotherapy in lung cancer patients, but outcomes of the studies were not consistent. According to Sullivan *et al.*, (2014) and Li *et al.*, (2010) chemotherapeutic response was associated with the gene polymorphism in *ERCC1* (rs3212986) in the lung cancer patients who diagnosed with stage

III. A momentous association was also observed in the study of Azuma *et al.*, (2009) between chemotherapeutic response of docetaxel in combination with cisplatin/carboplatin and lung cancer patients. It was also reported in the study of Lv *et al.*, (2014); that chemotherapeutic response was associated with lung cancer after analysed by univariate, multivariate and logistic regression model. Li *et al.*, (2010) also reported in his work that there was a significant relationship between chemotherapeutic response and *ERCC1* (rs3212986) in lung cancer patients.

According to Zhao *et al.*, (2014) a poor relationship was observed between chemotherapeutic response and *ERCC1* (rs3212986) in lung cancer patients who did show metastasis with stage III/IV. Poor association between chemotherapeutic response and *ERCC1* (rs3212986) was also observed in the published work of Takenaka *et al.*, (2010).

Gao *et al.*, (2014) reported a poor association was between outcomes of chemotherapy and *ERCC1* (rs3212986) in lung cancer patients with stage III and IV. The percentage of CR/PR and SD/PD to treatment was 52.76% and 55.83% respectively in Chinese population according to their study, whereas in this dissertation work it was 57.26% and 42.74% respectively in north Indian population. Median survival month varies among different population like according to Gao *et al.*, (2014) it was 19.3 months in Chinese population whereas in this dissertation work median survival month was 59 month in north Indian population.

Our study did not report any prognostic compelling relationship between overall survival and lung cancer patients having *ERCC1* genotype on applying univariate and logistic regression models and this study was supported by reported data of Shi *et al.*, (2015), Koc *et al.*, (2011) and also by various meta-analysis studies like meta-analysis of 39 published work which included 4606 lung cancer patients with stage III or IV, meta-analysis study of Perer-Ramirez *et al.*, (2016), meta-analysis study of 17 published work conducted by Yin *et al.*, (2011).

But when multivariate cox model was analysed then a significant relationship was observed between overall survival and male lung cancer patients who smoked and carrying *ERCC1* genotype. Also a non-significant relationship was observed between clinical response and patients having *ERCC1* genotype who were administered with various chemotherapeutic regimens like docetaxel + cisplatin/carboplatin, irinotecan + cisplatin/carboplatin or pemetrexed + cisplatin/carboplatin.

In the present study we also studied the combined effect of *MGMT* and *ERCC1* genotype on overall survival rate among lung cancer patients in north Indian population which was not studied earlier. A non-significant relationship was observed between overall survival and lung cancer patients having common *MGMT* and *ERCC1* genotype on applying univariate and logistic regression model. But a significant relationship was observed between overall survival and male lung cancer patients who smoked and carrying common *MGMT* and *ERCC1* genotype after analysed by univariate and logistic regression model. Also a non-significant relationship was observed between clinical response and patients having common *MGMT* and *ERCC1* genotype who were administered with various chemotherapeutic regimens like docetaxel + cisplatin/carboplatin, irinotecan + cisplatin/carboplatin or pemetrexed + cisplatin/carboplatin.

Computational analysis was performed for the first time in this study to investigate whether the gene polymorphism in *MGMT* and *ERCC1* contributes to any change in structure and function of the protein and also for the analysis of interaction of these genes with each other and with another genes. For this analysis various softwares were operated like SIFT, Provean, Polyphen 2, I-mutant, PhD-SNP, SNP & GO, MuPro, MutPred, Mutation 3-D, Elaspic, Panther and Genemania.

Sometime output from these softwares were inconsistent to each other viz. results of one software regarding one gene may be contradictory to the result of another software. This was explained by these findings: genetic polymorphism in *MGMT* gene (rs12917) contributed to deleterious SNP by Polyphen 2 and Panther software whereas it was tolerated according to SIFT, PhD-SNP, SNP & GO and MutPred softwares. Further according to Provean software this mutation was observed in non-protein coding region while according to Elaspic software this mutation was within protein coding region (protein domain). But in case of genetic polymorphism in *ERCC1* (rs3212986), the result of one software was consistent with the another software: according to SIFT, PhD-SNP, SNP & GO, MutPred, Panther and Polyphen 2 the mutation was tolerated and; also according to Provean and Elaspic software this mutation observed in non-protein coding region (outside of protein domain).

If any mutation observed outside of protein domain then this does not able to affect the structure and functioning of the protein. Findings reported in case of *ERCC1* (rs3212986) were supported by this hypothesis as this mutation was lie outside of protein coding region and hence its affect was non-deleterious in nature. But the outcomes of *MGMT* (rs12917)

were not fully supported by this hypothesis because the output of software was somehow different with each other.

Computational analysis just predicts the effect of mutation on protein coding region and that is why this analysis is not sufficient to study the overall effect of mutation on protein coding region, there is need of other analysis too for investigate the overall effect of mutation on protein coding region like in-vitro analysis, in-vivo analysis and statistical analysis.

Chapter 7

Conclusion

Total number of 214 lung cancer patients were selected as subject to study the relationship between response to platinum based doublet chemotherapy and genetic polymorphism in *MGMT* (rs12917) and *ERCC1* (rs3212986) gene individually and in combination, among lung cancer patients treated in Post-Graduate Institute of Medical Education and Research, Chandigarh on the basis of various parameters and with the help of univariate, multivariate and logistic regression analysis.

- Genotypic distribution of *MGMT* and *ERCC1* gene and patients having common *MGMT* and *ERCC1* genotype falls in the Hardy-Weinberg equilibrium.
- From the present study it was concluded there was no association between overall survival and genetic polymorphism in *MGMT* (rs12917) and *ERCC1* (rs3212986) (when analysed individually and in combination) among lung cancer patients on the basis of parameters like, smoking status, pack per years, gender, regimen and histology (analyzed by Univariate statistical method).
- Also when analysed by logistic regression method, a non-significant relationship of genetic polymorphism in *MGMT* (rs12917) and *ERCC1* (rs3212986) (when analysed individually and in combination) with the tumor stage, primary tumor extension, metastasis, lymph node involvement, ECOG and clinical response was observed among lung cancer patients. In contrast a significant association was observed among lung cancer patients having only *MGMT* genotype with the KPS values (**p=0.007**).
- Furthermore according to multivariate analysis, a non-significant relationship of both the SNPs (when studied individually and in combination) was observed with overall survival among SCLC patients, patients who did not smoke, patients treated with docetaxel, irinotecan and pemetrexed in combination with cisplatin/carboplatin and female patients. But a significant relationship with overall survival was observed among patients who smoked 25 or more than 25 cigarettes/bidis per day.
- A significant association was observed with the help of multivariate cox model; between overall survival and among ADCC patients (**p=0.02**), patients who were light smokers (**p=0.05**), patients who smoked (**p=0.03**) as well as among male patients (**p=0.01**) having *MGMT* genotype. But a non-significant relationship was observed between overall survival and SQCC patients harbouring *MGMT* genotype.

- When combined effect of *MGMT* and *ERCC1* was studied with the help of multivariate cox model; a non-significant association was observed between overall survival and among ADCC patients, SQCC patients, patients who were light smokers, patients who smoked as well as among male patients.
- On applying multivariate cox model on the basis of ADCC patients having *ERCC1* genotype; a significant relationship was observed among patients carrying mutant type genotype (**p=0.01**) with the overall survival. But this is not true among SQCC patients as the relationship observed in this case was non-significant.
- A significant relationship was observed between overall survival and male patients who smoked and carrying *ERCC1* genotype after analysed by multivariate cox model.
- In-silico analysis of *MGMT* (rs12917) concluded that according to some software it was deleterious (Polyphen 2 and Panther) and according to other it was tolerated (SIFT, PhD-SNP, SNP & GO and MutPred). And also this mutation was observed within protein domain according to Elaspic and according to Provean it was away from protein domain. This is due to different algorithm and machine learning programs of different softwares. But in case of *ERCC1* (rs3212986), it was concluded from the output of all the softwares that this was tolerated in nature and also present outside of protein domain.

References

1. Abdelraheem, N. E., Osman, M. M., Elgemaabi, O. M., Alla, A. A. F., Ismail, M. M., Osman, S. A., Ibrahim, A. I., Babekir, N. E., Mekki, S. O., & Hassan, M. A. (2016). Computational Analysis of Deleterious Single Nucleotide Polymorphisms (SNPs) in Human MutS Homolog6 (MSH6) Gene. *American Journal of Bioinformatics Research*, 6(2), 56-97.
2. Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 76(1), 7-20.
3. Altieri, F., Grillo, C., Maceroni, M., & Chichiarelli, S. (2008). DNA damage and repair: from molecular mechanisms to health implications. *Antioxidants & redox signaling*, 10(5), 891-938.
4. Anand, P., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., Sung, B., & Aggarwal, B. B. (2008). Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical research*, 25(9), 2097-2116.
5. Azuma, K., Sasada, T., Kawahara, A., Takamori, S., Hattori, S., Ikeda, J., Itoh, K., Yamada, A., Kage, M., Kuwano, M., & Aizawa, H. (2009). Expression of *ERCC1* and class III β -tubulin in non-small cell lung cancer patients treated with carboplatin and paclitaxel. *Lung Cancer*, 64(3), 326-333.
6. Biason, P., Hattinger, C. M., Innocenti, F., Talamini, R., Alberghini, M., Scotlandi, K., Zanusso, C., Serra, M., & Toffoli, G. (2012). Nucleotide excision repair gene variants and association with survival in osteosarcoma patients treated with neoadjuvant chemotherapy. *The pharmacogenomics journal*, 12(6), 476.
7. Bye, H., Prescott, N. J., Matejcic, M., Rose, E., Lewis, C. M., Parker, M. I., & Mathew, C. G. (2011). Population-specific genetic associations with oesophageal squamous cell carcinoma in South Africa. *Carcinogenesis*, 32(12), 1855-1861.
8. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human Mutation*. 30(8), 1237-1244
9. Cao, Z. H., Yin, H. P., Jiang, N., & Yu, B. (2015). Association between *ERCC1* and *ERCC2* gene polymorphisms and chemotherapy response and overall survival in osteosarcoma. *Genet Mol Res*, 14(3), 10145-51.
10. Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acid research*, 33 (suppl_2), W306-W310.
11. Capriotti, E., & Fariselli, P. (2017). PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic acids research*, 45(W1), W247-W252.
12. Chen, D. Q., Yao, D. X., Zhao, H. Y., & Yang, S. J. (2012). DNA repair gene *ERCC1* and *XPD* polymorphisms predict glioma susceptibility and prognosis. *Asian Pacific Journal of Cancer Prevention*, 13(6), 2791-2794.
13. Chen, W. H., Xin, P. L., Pan, Q. X., Chen, Y. Y., Wang, C. R., Zhang, Z. S., Chen, Y. F., Zhang, C. Y., & Cai, W. J. (2014). *ERCC1* single nucleotide polymorphism

- C8092A, but not its expression is associated with survival of esophageal squamous cell carcinoma patients from Fujian province, China. *PloS one*, 9(9), e106600.
14. Cheng, J., Randall, A., & Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines, *Proteins: Structure, Function and Bioinformatics*, 62(4), 1125-1132.
 15. Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16), 2145-2147.
 16. Croce, C. M. (2008). Oncogenes and cancer. *New England Journal of Medicine*, 358(5), 502-511.
 17. Dong, Y. S., Hou, W. G., Li, X. L., Jin, T. B., Li, Y., Feng, D. Y., Liu, D. B., Gao, G. D., Yin, Z. M., & Qin, H. Z. (2014), Genetic association ofCHEK2, GSTP1, and ERCC1 with glioblastoma in the Han Chinese population, *Tumor Biol.*, 35(5), 4937–4941.
 18. Du, L., Wang, H., Xiong, T., Ma, Y., Yang, J., Huang, J., Zeng, D., Wang, X., Huang, H., & Huang, J. (2013). The polymorphisms in the MGMT gene and the risk of cancer: a meta-analysis. *Tumor Biology*, 34(5), 3227-3237.
 19. Ebrahimi, E., Khandaghi, A. A., Valipour, F., Babaie, S., Asghari, F., Motaali, S., Abbasi, E., Akbarzadeh, A., & Davaran, S. (2016). In vitro study and characterization of doxorubicin-loaded magnetic nanoparticles modified with biodegradable copolymers. *Artificial cells, nanomedicine, and biotechnology*, 44(2), 550-558.
 20. Egyhazi, S., Ma, S., Smoczynski, K., Hansson, J., Platz, A., & Ringborg, U. (2002). Novel O6-methylguanine-DNA methyltransferase SNPs: A frequency comparison of patients with familial melanoma and healthy individuals in Sweden. *Human mutation*, 20(5), 408-409.
 21. Fahmideh, M. A., Schwartzbaum, J., Frumento, P., & Feychting, M. (2014). Association between DNA repair gene polymorphisms and risk of glioma: A systematic review and meta-analysis *Neuro-Oncology*, 16(6), 807–814.
 22. Felini, M. J., Olshan, A. F., Schroeder, J. C., North, K. E., Carozza, S. E., Kelsey, K. T., Liu, M., Rice, T., Wiencke, J. K., & Wrensch, M. R. (2007). DNA repair polymorphisms XRCC1 and MGMT and risk of adult gliomas. *Neuroepidemiology*, 29(1-2), 55-58.
 23. Felip, E., & Rosell, R. (2007). Testing for excision repair cross-complementing 1 in patients with non-small-cell lung cancer for chemotherapy response. *Expert review of molecular diagnostics*, 7(3), 261-268.
 24. Gao, H., Ge, R. C., Liu, H. Y., Wang, Y., & Yan, S. (2014). Effect of ERCC1 polymorphism on the response to chemotherapy and clinical outcome of non-small cell lung cancer. *Genet Mol Res*, 13(4), 8997-9004.
 25. Han, J., Hankinson, E.S., & Vivo, D.I., (2006). Polymorphisms in O⁶-methylguanine DNA methyltransferase and endometrial cancer risk. *Carcinogenesis*, 27(11), 2281-2285.
 26. Hassanpour, S. H., & Dehghani, M. (2017). Review of cancer from perspective of molecular. *Journal of Cancer Research and Practice*, 4(4), 127-129.

27. He, J., Wang, M. Y., Qiu, L. X., Zhu, M. L., Shi, T. Y., Zhou, X. Y., Sun, M. H., Yang, Y. J., Wang, Y. N., Li, J., Yu, H. P., & Wei, Q. Y. (2013). Genetic variations of mTORC1 genes and risk of gastric cancer in an Eastern Chinese population. *Molecular carcinogenesis*, 52(S1), 70-79.
28. Houtgraaf, J. H., Versmissen, J., & 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.
29. <http://elaspic.kimlab.org/many/> (visited on : 1-09-2018)
30. <http://folding.biofold.org/cgi-bin/i-mutant2.0.cgi> (visited on : 1-09-2018)
31. <http://genetics.bwh.harvard.edu/pph2/> (visited on : 1-09-2018)
32. <http://mutation3d.org/index.shtml> (visited on : 1-09-2018)
33. <http://mutdb.org/mutpred> (visited on : 1-09-2018)
34. <http://pantherdb.org/tools/csnpscoreForm.jsp> (visited on : 1-09-2018)
35. <http://profiles.nlm.nih.gov/ps/retrieve/Narrative/NN/p-nid> (visited on : 1-09-2018)
36. <http://provean.jcvi.org> (visited on : 1-09-2018)
37. <http://sift.jcvi.org> (visited on : 1-09-2018)
38. <http://snps.biofold.org/phd-snp/phd-snp.html> (visited on : 1-09-2018)
39. http://www.apiindia.org/pdf/medicine_update_2012/pulmonology_09.pdf (visited on : 1-09-2018)
40. <http://www.genemania.org/> (visited on : 1-09-2018)
41. <http://www.ics.uci.edu/~baldig/mutation.html> (visited on : 1-09-2018)
42. http://www.quazoo.com/q/O-6-methylguanine-DNA_methyltransferase?alt=O-6-methylguanine-dna_methyltransferase (visited on : 1-09-2018)
43. <https://international.neb.com/products/R0148-MboII#Product%20Information> (visited on : 1-09-2018)
44. <https://international.neb.com/products/R0155-HinfI> (visited on : 1-09-2018)
45. <https://snps-and-go.biocomp.unibo.it/snps-and-go/> (visited on : 1-09-2018)
46. <https://www.intechopen.com/source/html/44543/media/image3.png> (visited on : 1-09-2018)
47. <https://www.intechopen.com/books/bladder-cancer-from-basic-science-to-robotic-surgery/the-molecularbasis-of-cisplatin-resistance-in-bladder-cancer-cells> (visited on : 1-09-2018)
48. <https://www.ncbi.nlm.nih.gov/protein/> (visited on : 1-09-2018)
49. <https://www.ncbi.nlm.nih.gov/snp/> (visited on : 1-09-2018)
50. Hu, Y., Zhou, M., Li, K., Zhang, K., Kong, X., Zheng, Y., Li, J. & Liu, L. (2014). Two DNA repair gene polymorphisms on the risk of gastrointestinal cancers: a meta-analysis. *Tumor Biology*, 35(3), 1715-1725.
51. Huang, S. J., Wang, Y. F., Jin, Z. Y., Sun, J. Y., & Guo, Z. L. (2014). Role of *ERCC1* variants in response to chemotherapy and clinical outcome of advanced non-small cell lung cancer. *Tumor Biology*, 35(5), 4023-4029.

52. Hui, L., Yue, S., Gao, G., Chang, S., & Li, X. (2014), Association of single-nucleotide polymorphisms in *ERCC1* and *ERCC2* with glioma risk, *Tumor Biol.* 35(5), 7451–7457.
53. Hung, R. J., Christiani, D. C., Risch, A., Popanda, O., Haugen, A., Zienolddiny, S., Benhamou, S., Bouchardy, C., Lan, Q., Spitz, M. R., Wichmann, H. E., LeMarchand, L., Vinesis, P., Matullo, G., Kiyohara, C., Zhang, Z. F., Pezeshki, B., Harris, C., Mechanic, L., Seow, A., Ng, D. P., Dabrowska, N. S., Zaridze, D., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Foretova, L., Janout, V., Bencko, V., Caporaso, N., Chen, C., Duell, E. J., Goodman, G., Field, J. K., Houlston, R. S., Hong, Y. C., Landi, M. T., Lazarus, P., Muscat, J., McLaughlin, J., Schwartz, A. G., Shen, H., Stucker, I., Tajima, K., Matsuo, K., Thun, M., Yang, P., Wiencke, J., Andrew, A. S., Monnier, S., Boffeta, P., Brennan, P. (2008). International Lung Cancer Consortium: pooled analysis of sequence variants in DNA repair and cell cycle pathways. *Cancer Epidemiology and Prevention Biomarkers*, 17(11), 3081-3089.
54. Ingels, A., Dinhof, C., Garg, A. D., Maddau, L., Masi, M., Evidente, A., Berger, W., Dejaegher, B., & Mathieu, V. (2017). Computed determination of the in vitro optimal chemocombinations of sphaeropsidin A with chemotherapeutic agents to combat melanomas. *Cancer chemotherapy and pharmacology*, 79(5), 971-983.
55. Jiao, L., Bondy, M. L., Hassan, M. M., Wolff, R. A., Evans, D. B., Abbruzzese, J. L., & Li, D. (2006). Selected polymorphisms of DNA repair genes and risk of pancreatic cancer. *Cancer detection and prevention*, 30(3), 284-291.
56. 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. *Clinical lung cancer*, 10(2), 118-123.
57. Kamal, A., Shaik, T. B., & Malik, M. S. (2015). Embracing synthetic lethality of novel anticancer therapies. *Expert opinion on drug discovery*, 10(10), 1119-1132.
58. Khatami, F., Mohebi, S. R., Ghiasi, S., Haghighi, M. M., Safaei, A., Hashemi, M., & Zali, M. R. (2009). Amino acid substitution polymorphisms of two DNA methyltransferases and susceptibility to sporadic colorectal cancer. *Gastroenterology and Hepatology from bed to bench*, 1(3) 105-112.
59. Köberle, B., & Píe-Staffa, A. (2012). The Molecular Basis of Cisplatin Resistance in Bladder Cancer Cells. In *Bladder Cancer-From Basic Science to Robotic Surgery. InTech. Chap-13* 265-290.
60. Koç, E., Caner, V., Büyükpınarbaşılı, N., Tepeli, E., Türk, N. Ş., Cetin, G. O., & Bağcı, G. (2012). The determination of relationship between “excision repair cross-complementing group 1”(ERCC1) gene T19007C and C8092A single nucleotide polymorphisms and clinicopathological parameters in non-small cell lung cancer. *Molecular biology reports*, 39(1), 375-380.
61. Krześniak, M., Butkiewicz, D., Samojedny, A., Chorży, M., & Rusin, M. (2004). Polymorphisms in TDG and MGMT genes—epidemiological and functional study in lung cancer patients from Poland. *Annals of human genetics*, 68(4), 300-312.
62. Lamu, L., (1910) Etude de statistique clinique de 131 cas de cancer de l’oesophagi et du cordia, *Archives des Maladies Digestifs et de Malnutrition* 4, 451-456.

63. Le Chevalier, T. (2010). Adjuvant chemotherapy for resectable non-small-cell lung cancer: where is it going?. *Annals of Oncology*, 21(suppl_7), vii196-vii198.
64. Li, C., Liu, J., Li, A., Qian, L., Wang, X., Wei, Q., Zhou, J., & Zhang, Z. (2005). Exon 3 polymorphisms and haplotypes of O6-methylguanine-DNA methyltransferase and risk of bladder cancer in southern China: a case-control analysis. *Cancer letters*, 227(1), 49-57.
65. Li, F., Sun, X., Sun, N., Qin, S., Cheng, H., Feng, J., Chen, B., Cheng, L., Lu, Z., Ji, J., & Zhou, Y. (2010). Association Between polymorphisms of ERCC1 and XPD and clinical response to platinum-based chemotherapy in advanced non-small cell lung cancer. *American journal of clinical oncology*, 33(5), 489-494.
66. Li, J., Zuo, X., Lv, X. Kong, F., Xu, W., & Yang, S. (2014). Association of DNA repair gene polymorphisms with response to chemotherapy and prognosis of gastric cancer in a Chinese population. *Tumor Biol* 35(8), 7569–7574.
67. Liu, K., & Jiang, Y. (2017). Polymorphisms in DNA repair gene and susceptibility to glioma: a systematic review and meta-analysis based on 33 studies with 15 SNPs in 9 Genes. *Cellular and molecular neurobiology*, 37(2), 263-274.
68. Liu, R., Zhuang, Z., He, C., & He, Y. (2002). Relationship between genetic polymorphisms of human O⁶-methylguanine-DNA methyltransferase (MGMT) gene and susceptibility to tumors. *Carcinogenesis, Teratogenesis and Mutagenesis*, 14(2), 101-106.
69. Loh, Y. H., Mitrou, P. N., Wood, A., Luben, R. N., McTaggart, A., Khaw, K. T., & Rodwell, S. A. (2011). SMAD7 and MGMT genotype variants and cancer incidence in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk Study. *Cancer epidemiology*, 35(4), 369-374.
70. Lv, H., Han, T., Shi, X., Yao, Y., Yao, Y., Qiu, W., Yue, L., & Liang, J. (2014). Genetic polymorphism of GSTP1 and ERCC1 correlated with response to platinum-based chemotherapy in non-small cell lung cancer. *Medical oncology*, 31(8), 86.
71. Malik, P. S., & Raina, V. (2015). Lung cancer: Prevalent trends & emerging concepts. *The Indian journal of medical research*, 141(1), 5.
72. Mallath, M. K., Taylor, D. G., Badwe, R. A., Rath, G. K., Shanta, V., Pramesh, C. S., Digumarti, R., Sebastian, P., Borthakur, B. B., Kalwar, A., Kapoor, S., Kumar, S., Gill, J. L., Kuriakose, M. A., Malhotra, H., Sharma, S. C., Shukla, S., Viswanath, L., Chacko, R. T., Pautu, J. L., Reddy, K. S., Sharma, K. S., Purushotham, A. D., & Sullivan, R. (2014). The growing burden of cancer in India: epidemiology and social context. *The lancet oncology*, 15(6), 205-212.
73. Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982). *Molecular cloning: a laboratory manual* 545. Cold Spring Harbor, NY: Cold Spring harbor laboratory.
74. Manickam, M., Ramanan, P., Singh, P., & Talwar, P. (2014). In silico identification of genetic variants in glucocerebrosidase (GBA) gene involved in Gaucher's disease using multiple software tools. *Frontiers in genetics*, 5, 148.
75. Meyer, M. J., Lapcevic, R., Romero, A. E., Yoon, M., Das, J., Beltrán, J. F., Mort, M., Stenson, P. D., Cooper, D. N., Paccanaro, A., & Yu, H. (2016). mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome. *Human mutation*, 37(5), 447-456.

76. Mladenov, E., & Iliakis, G. (2011). The pathways of double-strand break repair. In *DNA Repair-On the Pathways to Fixing DNA Damage and Errors. InTech. Chap-8*, 143-168.
77. Mo, J., Luo, M., Cui, J., & Zhou, S. (2015). Prognostic value of ERCC1 and ERCC2 gene polymorphisms in patients with gastric cancer receiving platinum-based chemotherapy. *International journal of clinical and experimental pathology*, 8(11), 15065.
78. Molina, E., Pérez-Morales, R., Rubio, J., Petrosyan, P., Cadena, L. H., Arlt, V. M., Phillips, D. H., & Gonshebbat, M. E. (2013). The GSTM1null (deletion) and MGMT84 rs12917 (Phe/Phe) haplotype are associated with bulky DNA adduct levels in human leukocytes. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 758(1), 62-68.
79. Nay, L. S., & O'connor, R. T. (2013), Direct Repair in Mammalian Cells, *New Research directions in DNA repair*, Chp-5, 124-162.
80. Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids Research*, 31(13), 3812-3814.
81. Olaussen, K. A., Dunant, A., Fouret, P., Brambilla, E., André, F., Haddad, V., Taranchon, E., Filipits, M., Pirker, R., Popper, H. H., Stahel, R., Sabatier, L., Pignon, J. P., Tursz, T., Chevalier, T. L., & Soria, J. C. (2006). DNA repair by *ERCC1* in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *New England Journal of Medicine*, 355(10), 983-991.
82. Pegg, A. E., Fang, Q., & Loktionova, N. A. (2007). Human variants of O 6-alkylguanine-DNA alkyltransferase. *DNA repair*, 6(8), 1071-1078.
83. Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K. A., Lin, G. N., Nam, H. J., Mort, M., Cooper, D. N., Sebat, J., Lakoucheva, L. M., & Mooney, S. D. (2017). MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *BioRxiv*, 134981.
84. Pérez-Ramírez, C., Cañadas-Garre, M., Molina, M. Á., Robles, A. I., Faus-Dáder, M. J., & Calleja-Hernández, M. Á. (2017). Contribution of genetic factors to platinum-based chemotherapy sensitivity and prognosis of non-small cell lung cancer. *Mutation Research/Reviews in Mutation Research*, 5(5), 517-524.
85. Qiu, Z. X., Xue, F., Shi, X. F., He, X., Ma, H. N., Chen, L., & Chen, P. Z. (2014). MGMT Leu84Phe gene polymorphism and lung cancer risk: a meta-analysis. *Tumor Biology*, 35(5), 4381-4387.
86. Rumiato, E., Cavallin, F., Boldrin, E., Cagol, M., Alfieri, R., Basso, D., Castoro, C., Ermanno, A., Alberto, A., Alberto, R., & Saggioro, D. (2013). *ERCC1* C8092A (rs3212986) polymorphism as a predictive marker in esophageal cancer patients treated with cisplatin/5-FU-based neoadjuvant therapy. *Pharmacogenetics and genomics*, 23(11), 597-604.
87. Ryu, J. S., Memon, A., & Lee, S. K. (2014). *ERCC1* and personalized medicine in lung cancer. *Annals of translational medicine*, 2(4) 32-33.
88. Sato, S., & Itamochi, H. (2015). DNA Repair and Chemotherapy. In *Advances in DNA Repair. InTech. Chap-12*, 359-380.

89. Siegel, R. L., Miller, K. D., & Jemal, A. (2017). Cancer statistics, 2017. *CA: a cancer journal for clinicians*, 67(1), 7-30.
90. Shen, J., Terry, M. B., Gammon, M. D., Gaudet, M. M., Teitelbaum, S. L., Eng, S. M., Sagiv, K. S., Neugut, A. I., & Santella, R. M. (2005). MGMT genotype modulates the associations between cigarette smoking, dietary antioxidants and breast cancer risk. *Carcinogenesis*, 26(12), 2131-2137.
91. Shi, Z. H., Shi, G. Y., & Liu, L. G. (2015). Polymorphisms in *ERCC1* and *XPF* gene and response to chemotherapy and overall survival of non-small cell lung cancer. *International journal of clinical and experimental pathology*, 8(3), 3132.
92. Steffensen, K. D., Waldstrøm, M., & Jakobsen, A. (2011). DNA-repair *ERCC1* gene polymorphisms in epithelial ovarian cancer and relation to platinum resistance and survival. *Journal of Cancer Therapy*, 2(2), 140.
93. Stephens, F., Aigner, K., & Allen-Mersh, T. G. (2009). *Basics of oncology*. Dordrecht: Springer 1(2).
94. 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 letters*, 353(2), 160-166.
95. Surveillance, Epidemiology, and End Results (SEER) program (www.seer.cancer.gov) Research Data (1973-2015), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission.
96. Takenaka, T., Yano, T., Kiyohara, C., Miura, N., Kouso, H., Ohba, T., Kometani, T., Shoji, F., Yoshino, I., & Maehara, Y. (2010). Effects of excision repair cross-complementation group 1 (*ERCC1*) single nucleotide polymorphisms on the prognosis of non-small cell lung cancer patients. *Lung Cancer*, 67(1), 101-107.
97. Tan, L. M., Qiu, C. F., Zhu, T., Jin, Y. X., Li, X., Yin, J. Y., Zhang, W., Zhou, H. H., & Liu, Z. Q. (2017). Genetic polymorphisms and platinum-based chemotherapy treatment outcomes in patients with non-small cell lung cancer: a genetic epidemiology study based meta-analysis. *Scientific reports*, 7(1), 5593.
98. Thun, M. J., Hannan, L. M., Adams-Campbell, L. L., Boffetta, P., Buring, J. E., Feskanich, D., Flanders, W. D., Jee, S. H., Katanoda, K., Kolonel, L. N., Lee, I., Marugame, T., Palmer, J. R., Riboli, E., Sobue, T., Tang, E. A., Wilkens L. R. & Samet J. M. (2008). Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS medicine*, 5(9), e185.
99. Travis, W. D., Brambilla, E., Burke, A. P., Marx, A., & Nicholson, A. G. (2015). Introduction to the 2015 World Health Organisation Classification of tumors of the lung, pleura, thymus, & heart. *Journal of Thoracic Oncology*, 10(9), 1240-1242.
100. van der Straaten, T., Mullenders, L., & Guchelaar, H. J. (2011). Pharmacogenetics of cancer and DNA repair enzymes. In *DNA Repair and Human Health*. 51(7) 467-480.
101. Wang, Q., Wang, A. H., Tan, H. S., Feng, N. N., Ye, Y. J., Feng, X. Q., Liu, G., Zheng, Y. X., & Xia, Z. L. (2010). Genetic polymorphisms of DNA repair genes and chromosomal damage in workers exposed to 1, 3-butadiene. *Carcinogenesis*, 31(5), 858-863.

102. Wang, S., Wang, J., Bai, Y., Wang, Q., Liu, L., Zhang, K., Hong, X., Deng, Q., Zhang, X., He, M., Wu, T., Xu, P., & Guo, H. (2016). The genetic variations in DNA repair genes ERCC2 and XRCC1 were associated with the overall survival of advanced non-small-cell lung cancer patients. *Cancer medicine*, 5(9), 2332-2342.
103. Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz M., Grouios C., Kazi, F., Lopes, T. C., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., & Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*, 38(suppl_2), W214-W220
104. Weir, H. K., Thompson, T. D., Soman, A., Møller, B., & Leadbetter, S. (2015). The past, present, and future of cancer incidence in the United States: 1975 through 2020. *Cancer*, 121(11), 1827-1837.
105. Witvliet, D. K., Strokach, A., Giraldo-Forero, A. F., Teyra, J., Colak, R., & Kim, P. M. (2016). ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity. *Bioinformatics*, 32(10), 1589-1591.
106. Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., & Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome research*, 13(9), 2129-2141.
107. www.beaconhospital.com.my/lung-cancer (visited on: 1-09-2018).
108. Xiang, M., Sun, L., Dong, X., Yang, H., Liu, W. B., Zhou, N., Han, X., Zhou, Z., Cui, Z., Liu, J. Y., Cao, J., & Ao, L. (2015). Association between genetic polymorphisms of DNA repair genes and chromosomal damage for 1, 3-butadiene-exposed workers in a matched study in China. *BioMed research international*, 2015.
109. Xu, T. P., Shen, H., Liu, L. X., & Shu, Y. Q. (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(2), 265-274.
110. Xu, Z., Ma, W., Gao, L., & Xing, B. (2014). Association between ERCC1 C8092A and ERCC2 K751Q polymorphisms and risk of adult glioma: a meta-analysis. *Tumor Biology*, 35(4), 3211-3221.
111. Xue, M. H., Li, G. Y., Wu, X. J., Zhang, C. X., Zhang, C. F., & Zhu, K. X. (2015). Genetic variability of genes in NER pathway influences the treatment outcome of gastric cancer. *International journal of clinical and experimental pathology*, 8(5), 5563.
112. Yin, J., Vogel, U., Guo, L., Ma, Y., & Wang, H. (2006). Lack of association between DNA repair gene ERCC1 polymorphism and risk of lung cancer in a Chinese population. *Cancer genetics and cytogenetics*, 164(1), 66-70.
113. Yin, M., Yan, J., Voutsina, A., Tibaldi, C., Christiani, D. C., Heist, R. S., Rosell, R., Booton, R. & Wei, Q. (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(3), 370-377.

114. Yu, D., Shi, J., Sun, T., Du, X., Liu, L., Zhang, X., Lu, C., Tang, X., Li, M., Xiao, L., Zhang, Z., Yuan, Q., & Yang, M. (2012). Pharmacogenetic role of *ERCC1* genetic variants in treatment response of platinum-based chemotherapy among advanced non-small cell lung cancer patients. *Tumor Biology*, 33(3), 877-884.
115. Yu, W. H., Wang, Y. X., Guo, J. Q., Wang, Y. L., Zheng, J. S., & Zhu, K. X. (2015). Genetic variability of *ERCC1* and *ERCC2* influences treatment outcomes in gastric cancer. *Genet Mol Res*, 14(4), 17529-17535.
116. Zhang, M., Huang, W. Y., Andreotti, G., Gao, Y. T., Rashid, A., Chen, J., Sakoda, L. C., Shen, M. C., Wang, B. S., Chanock, S. & Hsing, A. W. (2008). Variants of DNA repair genes and the risk of biliary tract cancers and stones: a population-based study in China. *Cancer Epidemiology and Prevention Biomarkers*, 17(8), 2123-2127.
117. Zhao, X., Zhang, Z., Yuan, Y., & Yuan, X. (2014). Polymorphisms in *ERCC1* gene could predict clinical outcome of platinum-based chemotherapy for non-small cell lung cancer patients. *Tumor Biology*, 35(8), 8335-8341.
118. Zhu, J., Hua, R. X., Jiang, J., Zhao, L. Q., Sun, X., Luan, J., Lang, Y., Sun, Y., Shang, K., Peng, S. & Ma, J. (2014). Association studies of *ERCC1* polymorphisms with lung cancer susceptibility: a systematic review and meta-analysis. *PloS one*, 9(5), e97616.

Prognostic significance of *MGMT* & *ERCC1* gene polymorphism in lung cancer patients treated with platinum based doublet chemotherapy

ORIGINALITY REPORT

14%

SIMILARITY INDEX

8%

INTERNET SOURCES

13%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1

&NA;, . "15th World Conference on Lung Cancer :", Journal of Thoracic Oncology, 2013.

Publication

2%

2

Shweta Lawania, Navneet Singh, Digambar Behera, Siddharth Sharma. "Association of XPA Polymorphisms Towards Lung Cancer Susceptibility and its Predictive Role in Overall Survival of North Indians", Biochemical Genetics, 2018

Publication

1%

3

"Abstracts", Journal of Thoracic Oncology, 2009

Publication

1%

4

Ji Eun, Kyung Im Kim, Sung Soo Yoon, Bong Jin Hahm, Sang Min Lee, Jeong Hyun Yoon, Wan Gyoon Shin, Hye Suk Lee, and Jung Mi Oh. "Psychological Distress

<1%