# Genetic and Epigenetic Alterations in Lung Cancer Associated with Tobacco Exposure

## THESIS

Submitted in partial fulfillment

of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

by

**Rakhshan Ihsan**

Under the Supervision of

**Dr. Sunita Saxena**

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE**

**PILANI (RAJASTHAN) INDIA**

**2013**

राष्ट्रीय विकृति विज्ञान संस्थान
(भारतीय आयुर्विज्ञान अनुसंधान परिषद्)
"श्रीरामाचारी भवन"
सफदरजंग अस्पताल परिसर,
पोस्ट बाक्स नं॰ 4909, नई दिल्ली-110029 (भारत)

NATIONAL INSTITUTE OF PATHOLOGY
(Indian Council of Medical Research)
"SRIRAMACHARI BHAWAN"
Safdarjang Hospital Campus,
Post Box No. 4909, New Delhi-110029 (India)

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

# CERTIFICATE

This is to certify that the thesis entitled **"Genetic and Epigenetic Alterations in Lung Cancer Associated with Tobacco Exposure"** and submitted by **Rakhshan Ihsan,** ID No **2007PHXF439P** for award of Ph.D Degree of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor:

DR. SUNITA SAXENA
डॉ॰ सुनीता सक्सेना
निदेशक / Director
राष्ट्रीय विकृति विज्ञान संस्थान, भा.आ.अं.प.
National Institute of Pathology - ICMR
सफदरजंग अस्पताल परिसर
Safdarjung Hospital Campus
नई दिल्ली-110029 / New Delhi-110029

Name in capital letters: **DR. SUNITA SAXENA**

Designation: **DIRECTOR**
**NATIONAL INSTITUTE OF PATHOLOGY (ICMR)**

Date: **24th September 2013**

Telephone/दूरभाष +91-11-26198402-6, 26716473-80 Fax/फैक्स +91-11-26198401, 26160135
e-mail/ईमेल info@instpath.gov.in Visit us: www.instpath.gov.in Tel.Fax/टैलीफैक्स +91-11-26170499

*The quest for certainty blocks the search for meaning. Uncertainty is the very condition to impel man to unfold his powers.*

-*Erich Fromm*

## *Acknowledgement*

*It would not have been possible to write this doctoral thesis without the help and support of the kind people around me. I owe my gratitude to all those who have made this dissertation possible and to only some of whom it is possible to give particular mention here.*

*This doctoral thesis would not have been possible without the able guidance of my supervisor* ***Dr. Sunita Saxena****. I express my sincere gratitude and heartfelt thanks to her for the continuous support throughout my Ph.D study and research. Her professional knowledge, wise advices and kind personality have been leading me through every step of the study. She patiently provided the vision and encouragement in times of new ideas and difficulties. I appreciate her freedom to develop and apply my intellectual abilities and creativity throughout my work. I also remain indebted for her understanding and support during the times of my personal problems. You have been more than a mentor to me; it has been a real pleasure to work with you.*

*I am very grateful to my Co supervisor,* ***Dr Sujala Kapur*** *for insightful comments, support and for many motivating discussions. Maam, thank you for your helpful, wise guidance and incessant support throughout the study.*

*I express my warmest gratitude to* ***Prof. Dr. Bijendra Nath Jain****, Vice Chancellor, B.I.T.S Pilani for providing me opportunity to pursue the off campus Ph.D of the Institute. I express my gratitude to* ***Prof. S. K. Verma****, Dean, Research and Consultancy Division (RCD), BITS, Pilani for his constant official support and encouragement. In addition, I would like to thank my Doctoral Advisory Committee (DAC) members,* ***Dr. Lalita Gupta*** *and* ***Dr. Vishal Saxena.*** *I would also like to thank* ***Dr. Hemanth Jadav*** *and* ***Dr. Sharad Shrivastava*** *nucleus members of RCD and all office staff of RCD B.I.T.S, Pilani, for their cooperation and guidance in all official procedures associated with the doctoral program.*

*I am most grateful to all my collaborators,* ***Dr. Amal Kataki****,* ***Dr Jagannath Sarma****,* ***Dr. Eric Zomawia*** *and* ***Dr Yogesh Verma*** *for their valuable contribution in sample collection and for their assistance with the clinical information. I would also like to thank*

*all participants in this study, who by giving samples and personal information, have made it possible to perform the analysis and conclusion in this work.*

*I wish to express my gratitude to* ***Dr. L.C Singh*** *for providing me the support and immense help in my Real-Time PCR and Sequencing experiments. I am also very much thankful to* ***Dr A.K Mishra****, for his immense help in statistical analysis and insightful comments during manuscripts preparations. I enjoyed our discussions and talks very much and I will never forget that you always had a few minutes for me even if you were very busy. I sincerely thank* ***Dr Usha Agarwal*** *and* ***Dr. Raju*** *for their kind suggestions, encouragement and everyday support.*

*Very special thanks to my long time friends and colleagues* ***Mishi*** *and* ***Pradeep.*** *You both kept things light and me smiling. With Mishi, I have laughed, cried and among other things, been 'scolded' together! With Pradeep, I had invaluable support day in and day out, during all these years. I really appreciate your very patient nature; you must have nerves of steel to work alongside me! Thank you for providing the camaraderie, fun and many joyful moments. Pradeep 'sir', this thesis would not have happened without your help. Much as happened and changed in the time but in you, I have a life-long friend and colleague.*

*I am indebted to my student colleagues for providing a stimulating and fun filled environment. I would also like to thank* ***Thoudam Regina Devi****,* ***Dhirendra Singh Yadav*** *and* ***Abha Soni,*** *for their encouragement and instructive suggestions. It was a pleasure and privilege working with them. I also warmly thank my colleagues* ***Shanti, Asheema, Nitu*** *and* ***Ashish,*** *who created a pleasant and congenial working atmosphere that I truly enjoyed.*

*I would also like to thank* ***Mrs Sartha*** *(Administrative Officer) for her help in various matters. I am also thankful to other staff of office, store, library, computer and establishment section for their kind help and support.*

*I wish to thank all the students and staff that have passed through the Institute and friends from the other walks of life- thanks for the laughs and keeping me relatively "sane". I would also like I extend my heartfelt thanks to my "Hamdard" friends-* ***Dr. Shams Tabrez*** *and* ***Dr Abhay Singh.*** *Thankyou Shams, for the constant encouragement*

*and belief. You are an excellent example of faith, hard work and perseverance. You have always been a source of positive energies.*

*This list would be incomplete without thanking my lovely friend and roommate* ***Mridu,*** *without whom life would be bleak. Thanks a lot to be such a warm person and for all the fun we had. Thank you, for providing solace from the maddening crowd and making my stay in Delhi most fun filled. I will always remain indebted for all the help especially in the last part of my work.*

*Most importantly, none of this would have been possible without the love and patience of my family. My parents (****Mamma and Papa****) were the first to believe that I can achieve my dreams. They have always inspired, encouraged and provided infallible support in everything that I have done till now. My deep love and appreciation for the personal sacrifices they made, while encouraging me to reach new levels of excellence. I hope I have made them proud, as I am proud of them. A huge hug to* ***Samar****, for the love, laughs and sisterhood. I owe so much to them for lovingly and unselfishly caring for Daniya, allowing me to focus on my work. Special thanks to my brother* ***Ajmal*** *and my sister-in-law* ***Rabiya*** *for their love, constant encouragement and strength throughout. Love goes to my ever so cute and loving niece* ***Naba*** *who refreshes my life by just being around. My immense gratitude to my* ***in-laws*** *for the liberty and encouragement to do my research work.*

*And most of all for my husband* ***Ragbat****, whose unfailing love, patient support, encouragement and understanding during all these years is so appreciated. I can never ever thankyou enough for sticking by my side and unconditionally loving me during my good and bad times. Finally, my love and thanks to my daughter* ***Daniya****, who has been the most valuable gift in my life. Your smile has reminded me of more important things in life than work. I am deeply sorry for the time we spent apart while completing this dissertation. I love you both so much, my darlings!*

*Finally but most importantly, I must acknowledge and thank* ***The Almighty*** *for blessing, protecting and guiding me throughout this period. It is under His grace that we live, learn and flourish. I will keep on trusting You for my future.*

***Rakhshan Ihsan***

# *Abstract*

## *Abstract*

Lung cancer (LC) incidences are on the rise globally and this has been attributed to both genetic and environmental risk factors. In India it constitutes 6.2% of all cancers with approximately 58,000 incident cases reported in 2008 and is the most frequent cancer in males. North eastern (NE) part of India is showing a steady rise in cancer incidences and lung cancer is among the ten leading sites. This region, due to its geographic location and the presence of diverse populations is a hotspot for genetic studies. The absolute risk of lung cancer is not known but smoking is considered as a relative risk factor. The area also reports tobacco use in variety of ways of chewing and smoking that are different from the rest of India. Unlike the Western nations or other parts of India, use of tobacco or alcohol in crude form is more prevalent in this region. Therefore, the current challenges in the management of lung cancer in this region are to obtain a better understanding of the underlying molecular alterations and environmental risk factors to provide effective early detection, prognostic and predictive marker.

The thesis explores the genetic and epigenetic variations in LC and their complex interplay with environmental exposures to present a comprehensive molecular portrait of the disease etiology in the studied population. Since many polymorphic genetic variations produce proteins with increased, decreased or a complete loss of enzymatic activity, they are relevant factors in the gene-environment interplay. The role of these genetic alterations and their interactions with environmental risk factors in lung carcinogenesis may determine interindividual susceptibility to cancer. Variations in xenobiotic metabolizing genes, involved in tobacco-smoke carcinogen metabolism, can result in variable amounts of harmful DNA adducts that can ultimately lead to cancer. In association study which included xenobiotic metabolizing gene and environmental risk factor, we first showed that genetic variants of CYP1A1 and EPHX1 genes show independent contrasting main effect in LC. We then demonstrated that there are distinct gene-gene and gene environment interactions that were associated with smoking in LC

risk by applying novel data mining approaches such as CART and MDR. Combinations of EPHX1 Tyr113His and SULT1A1 Arg213His were identified risk signatures in smokers. Moreover the association identified in the study remained true even at low prior probabilities of FPRP testing, thus minimizing the concerns raised by the multiple hypotheses testing for false discovery rate. This approach was important to unravel the gene-environment interactions, especially relevant in the initiation of lung cancer. In addition to this, interactions of p53 genotypes and betel quid chewing conferred significant increase risk to LC in the study. Further, gene dosage effects of *GSTT1* and *GSTM1* copy number revealed reduced LC risk associated with the interaction of smoking with hemizygous and null genotypes of GSTT1 gene.

Gene expression profiles identify unique gene signatures that provide novel insights into fundamental cancer biology at the molecular level. Data mining and computation analysis was also done to explore possible networks and pathways. A total of 734 genes were differentially expressed which enriched to epidermal growth factor, homeobox related transcription activity terms and MAPK signaling. A signature of 24 differentially expressed zinc finger proteins including a zinc transporter *SLC30A1* was identified. We also identified differential expression of *TMSB10, RPS*, *PPFIA1*, *TNS3, NGFR, CLK3* and *PFDN6* genes for first time in NSCLC.

Finally, we explored aberrant promoter hypermethylation, an epigenetic change that occurs early in lung tumorigenesis resulting in silencing and inactivation of genes. Variations in methylation pattern have been reported to vary with ethnicity probably owing to complex epistasis or gene environment interactions. Methylation analysis showed promoter of p16 gene was the most frequently methylated followed by RASSF1A, DAPK and GSTP1 and also reports the higher frequency of GSTP1 promoter methylation in the population as compared to earlier studies. Further, results suggest an association of p16 and RASSF1A promoter methylation with smoking and betel quid chewing with increased risk of lung cancer.

Overall, our findings will help in understanding the etiology of lung cancer in the studied population and others where similar risk habits are highly prevalent.

## *Table of Contents*

## *List of Tables*

## *List of Figures*

| | | |
|---|---|---|
| α | = | Alpha |
| β | = | Beta |
| γ | = | Gamma |
| µg | = | Microgram |
| µl | = | Microlitre |
| AAR | = | Age Adjusted Rate |
| CI | = | Confidence Interval |
| CART | = | Classification and Regression Tree |
| cDNA | = | Complementary Deoxyribonucleic Acid |
| CYP | = | cytochrome P450 |
| D' | = | Linkage Disequilibrium |
| DAVID | = | Database for Annotation, Visualization and Integrated Discovery |
| DEPC | = | Diethyl pyrocarbonate |
| DNA | = | Deoxyribonucleic acid |
| dNTP | = | Deoyribose Nucleotide Triphosphate |
| EDTA | = | Ethylenediaminetetraacetic acid |
| EtBr | = | Ethidium Bromide |
| FPRP | = | False Positive Report Probability |
| FDR | = | False Discovery Rate |
| HWE | = | Hardy–Weinberg Equilibrium |
| IPA | = | Ingenuity Pathways Analysis |
| LC | = | Lung Cancer |
| LR | = | Logistic Regression |
| mg | = | milligram |
| mRNA | = | messenger Ribonucleic Acid |

| | | |
|---|---|---|
| MDR | = | Multifactor Dimensionality Reduction |
| MSP | = | Methylation Specific PCR |
| NE | = | North East |
| NNK | = | N-nitrosamines, 4- methylnitrosamino-1-3-pyridyl-1-butanone |
| NSCLC | = | Non Small Cell Lung Cancer |
| OD | = | Optical Density |
| OR | = | Odd Ratio |
| PBS | = | Phosphate Buffered Saline |
| PAH | = | Polycyclic Aromatic Hydrocarbon |
| PBCR | = | Population Based Cancer Rregistry |
| PCR | = | Polymerase Chain Reaction |
| RFLP | = | Restriction Fragment Length Polymorphism |
| RIN | = | RNA Integrity Number |
| RNA | = | Ribonucleic acid |
| RT | = | Room Temperature |
| RT-PCR | = | Reverse Transcriptase- Polymerase Chain Reaction |
| TBA | = | Testing Balance Accuracy |
| Tris | = | Tris (hydroxymethyl) amino acid |
| TSG | = | Tumour Suppressor Gene |
| SNP | = | Single Nucleotide Polymorphism |
| SCLC | = | Small Cell Lung Carcinoma |
| TAE | = | Tris Acetate Ethylene Diamine Tetra Acetate |
| TBE | = | Tris Boric Ethylene Diamine Eetra Acetate |
| UTR | = | Untranslated region |
| UV | = | Ultra Violet |
| WHO | = | World Health Organization |

# *Chapter 1*

# *Introduction*

# *Introduction*

Lung cancer (LC) is one of the most prevalent and the leading cause of cancer-related deaths in men and second leading cause of cancer deaths in women around the world [Jemal et al., 2011]. The incidence of the disease is increasing, particularly in developing countries. The rates are 2 to 5 times higher in developed countries compared with the developing ones, a result of variations in a disparate set of risk factors and diagnostic practices [Jemal et al., 2011]. In males, the highest LC incidence rates are in Eastern and Southern Europe followed by North America and Micronesia while rates are low in sub-Saharan Africa. In females, the highest LC incidence rates are found in North America followed by Northern Europe, and Australia/ New Zealand. In India, lung cancer constitutes 6.2% of all cancers with approximately 58,000 incident cases reported in 2008 [Ferlay et. al., 2010]. It is the third largest cause of cancer mortality in India accounting for nearly 8.3% of all cancer related deaths in the country [Ferlay et. al., 2010]. In North eastern (NE) part of India, Lung cancer is among the ten leading sites, with the highest age-adjusted incidence rate (AAR) in Aizwal district (36.0 in males and 38.7 in females) followed by Mizoram state (24.5 in males and 26.3 in females) and Imphal district (25.1 in males and 19.8 in females ) [Zomawia 2010].

Decades of research have contributed to our understanding that lung cancer is a multi-step process involving genetic and epigenetic alterations where resulting DNA damage transforms normal lung epithelial cells into lung cancer [Wistuba et. al., 2006]. Risk factors consistently associated with lung cancer include smoking and tobacco smoke carcinogen. Tobacco smoking is the most important cause of lung cancer, accounting for about 85% of cases. The risk of cancer differs by age, smoking intensity, and smoking duration; the risk of cancer declines after smoking cessation, but it never returns to baseline. 15-25% lung cancer cases occur in patients who have never smoked (less than 100 cigarettes in a lifetime) [Larsen and Minna 2011]. These etiological differences are

associated with differences in tumor acquired molecular changes. Family history is said to impart a twice higher risk in the offsprings of lung cancer patients than risk in the general population [Lorenzo Bermejo et. al., 2005]. Other possible risk factors include exposure to secondhand smoke (Passive smoking) and exposure to industrial toxins, such as asbestos, radiation, arsenic, chromates, nickel, chloromethyl ethers, mustard gas, or coke-oven emissions, encountered or breathed in at work.

The chronic exposure to tobacco smoke carcinogens induces genetic and epigenetic changes in lung epithelial cells which transforms them to a malignant stage. Research groups have used association studies to assess various candidate genes including those encoding enzymes that either activate or inactivate carcinogens found in tobacco smoke. The evidence is strong for xenobiotic metabolizing genes. Xenobiotic-metabolizing phase I and phase II enzymes present in the human lung often play a dominant role in disposition of the carcinogenic constituents of tobacco smoke and also their pharmacological and toxicological effects. The phase I xenobiotic metabolizing enzymes such as cytochrome P-450s (CYPs), alcohol dehydrogenase (ALDH) and epoxide hydroxylase (EPHX) usually activate the procarcinogens through oxidation and dehydrogenation thereby converting them into reactive metabolites. Phase II metabolic enzymes such as glutathione-S-transferases (GST), sulfotransferase (SULT) and N-acetyltransferase (NAT) generally result in inactivation or detoxification of these reactive metabolites. Equilibrium between expression and activity levels of these xenobiotic-metabolizing enzymes of both phase I and II determine the relative level of detoxification of carcinogens. Since multiple gene and gene-related alterations and their interaction with environment contribute to lung cancer development and progression, multigenic approaches have to be used in association studies to assess the combined effects of genes that interact and function in the same pathway. Further, high order interactions in multigenic approach allow more precise delineation of risk groups than single gene locus analysis.

It is reported that exposure to environmental carcinogens, such as tobacco smoke induces lung cancer in mice via both genetic and epigenetic events [Hutt et. al., 2005]. Loss of tumor suppressor gene (TSG) function is an important step in lung

carcinogenesis and usually results from inactivation of both alleles, with LOH inactivating one allele through chromosomal deletion and point mutation, while epigenetic or transcriptional silencing inactivating the second allele [Breuer et. al., 2005]. Epigenetic events can lead to changes in gene expression without any changes in DNA sequence and therefore are potentially reversible [Bird et. al., 2002]. Aberrant promoter hypermethylation occurs early in lung tumorigenesis resulting in silencing of gene transcription and therefore a common method for inactivation of TSGs in lung cancer. This includes genes involved in tissue invasion, DNA repair, detoxification of tobacco carcinogens, and differentiation. Hypermethylation of several TSGs have been reported in lung cancer, among them p16, GSTP1, RASSF1A are the most frequent [Schwartz et. al., 2007, Risch et. al., 2008]. Moreover, results from previous studies have reported that promoter methylation of some genes occurred more frequently in lung tumors from smokers, compared with non-smokers [Kim et. al., 2001, Divine et. al., 2005, Toyooka et. al., 2006].

Progress of phenotypes from normal to advanced carcinoma is controlled by a transcriptional hierarchy that coordinates the action of hundreds of genes. Genetic and epigenetic mechanisms underlying lung cancer development and progression continue to emerge, spearheaded by the development of technologies allowing genome-wide analysis of DNA copy-number, mutations and gene expression. Profiling the lung cancer transcriptome has imparted biologically- and clinically-relevant information such as novel dysregulated genes and pathways and gene signatures that can predict patient prognosis and response to treatment [Anguiano et. al., 2008]. Microarray expression profiling has provided important information regarding lung carcinogenesis further it has also been applied to divide lung cancers into sub-types [Garber et. al., 2001]. In an analysis of 32 NSCLC specimens and 7 normal specimens, unsupervised hierarchical analysis segregated tumors on the basis of histologic type and differentiation [Borczuk et. al., 2003]. Another study classified adenocarcinoma into subtypes, identifying a group with significantly poor outcome [Bhattacharjee et. al., 2001]. Also, gene expression profiling has been used as a discovery tool for early detection biomarkers and for genes induced by cigarette smoking [Sugita et al., 2002].

## Gap in Literature

India is a developing country with one of the most diverse populations in the world with North Eastern part leading in the incidences of lung cancer among top ten sites of cancer. In north east India, tobacco smoking habit is much more rampant than any other part of India. Another risk factor, is the widespread use of fermented, raw and wet variety of BN, locally called ‘kwai’ or ‘tambul’ which is primarily consumed with betel leaf and slaked lime. Recent researches have has also generated sufficient evidences to implicate betel nut as well as betel quid, with or without tobacco, as a suspected carcinogen to humans. The equilibrium between expression and activity levels of xenobiotic-metabolizing enzymes of both phase I and II will determine the relative level of detoxification of carcinogens. Given the importance of the xenobiotic metabolizing genes and environmental risk, this area deserves increased attention particularly in high risk population of NE India where there is no literature evidence in LC. A high order gene-gene interaction in multigenic approach may help in more precise delineation of the risk groups of the disease. According to our current knowledge, the most commonly mutated gene, in lung cancer is the p53 tumor suppressor gene. Since there is no literature available in this population about the role of p53 codon 72 polymorphism and its interaction with environmental risk factor in lung carcinogenesis, we investigated this relationship in present study.

Besides genetic susceptibility, there is also evidence to suggest that lung cancer is driven by epigenetic changes like DNA methylation. Many studies from western countries have investigated the role of methylation status of candidate tumor suppressor genes such as p16, RASSF1A, DAPK and GSTP1 in LC for risk assessment, early detection, disease progression and prognosis. However, there is dearth of literature in India on methylation in lung cancer. This impelled us to investigate the role of promoter methylation of these genes in LC patients from north east India and their interaction with risk habits.

Gene expression arrays are employed to discover changes in the DNA expression that occur in neoplastic transformation. In contrast to candidate approaches the microarray studies have aimed at developing exploratory gene profiles of cancer cells to identify genes related to tumorigenesis, delineate molecular phenotypes and identify functional gene clusters as potential markers of biological behavior. The gene expression

profile of LC has so far not been investigated in this part of India. Therefore, cDNA microarray gene expression analysis was done to obtain the molecular signature of patients with NSCLC.

Most of the studies on lung cancer have been reported from the Western population, where etiology and genetic factors differ considerably from Asian populations. To our knowledge, there are no reports on genetic and epigenetic aspects of lung cancer from Indian population. The ethnic NE population of India, due to its unique, strategic geographic location and the presence of linguistically, culturally and demographically diverse populations is a hotspot for both genetic and epigenetic studies. However, studies exploring genetic aspects of the disease are lacking from this population. Identification of genes and pathways involved will not only enhance our understanding of the disease biology in the population, it will also provide new targets for early diagnosis and facilitate treatment.

# *Chapter 2*

# *Review of Literature*

# *Review of Literature*

Lung cancer (LC) is the most common cancer worldwide, with an estimated 1,600,000 new cases and 1,380,000 deaths in 2008 [Jemal et al., 2011]. In the United States, there will be an estimated 228,000 new cases of lung cancer and 159,500 deaths in 2012 [Siegel et al., 2013]. Lung cancer is a disease characterized by uncontrolled reproduction of cell in tissues of the lung. The accumulation of these cells is called a tumor. Lung tumors derive from pluripotential cells, i.e. cells that have the ability to mature or differentiate into any of the cells in the lung, which line the tracheobronchial tree or alveoli. Lung cancers represent a heterogeneous collection of tumors that are characterized by a large number of abnormalities of both chromosome number and structure. The genetic alterations displayed by a given tumor are the result of a combination of changes that are directly or indirectly caused by inducing factors, such as tobacco carcinogens, and those that rise up secondarily as a consequence of defects in genes that maintain genomic stability. Although a large number of genes which recurrently aberrated in LCs have been identified, however numerous genes contributing to lung carcinogenesis are yet to be revealed.

## HISTOLOGICAL SUBTYPES OF LUNG CANCER

Based on the morphological criteria, lung cancer is broadly divided into two classes — non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) (Figure 2.1).

**Non-small cell lung cancer (NSCLC)** is a collection of several tumor histologies including: adenocarcinoma, squamous-cell carcinoma and large cell carcinoma. It accounts for approximately 75%-80% of all lung cancers and is characterized by slower growth and spread than SCLC [Travis et. al., 1995]. NSCLC can be surgically resected and are characterized by better prognosis, which is reflected in longer overall patient survival (Figure 2.2)

Adenocarcinomas arise more peripherally in the lung from mucous glands and the cells retain some of the tubular, acinar or papillary differentiation and mucus production. They commonly invade pleura and mediastinal lymph nodes and often metastasise to the brain and bones. They bear similarity to secondary tumors and must be distinguished by CT scans and other investigations to check for presence of a primary. Adenocarcinoma commonly arises around scar tissue and is also associated with asbestos exposure. Adenocarcinomas are proportionally less common in non-smokers.

Squamous cell carcinoma (SCC) (epidermoid carcinoma) accounts for about 30 or 40% of primary lung tumors. They grow most commonly in the central areas in or around major bronchi. They grow in a stratified or pseudoductal arrangement, the cells have an epithelial pearl formation with individual cell keratinization. Smoking increases the incidence of all the major histological subtypes but SCLC and SCC seem to be most strongly associated with smoking [Vainio et. al., 1994].

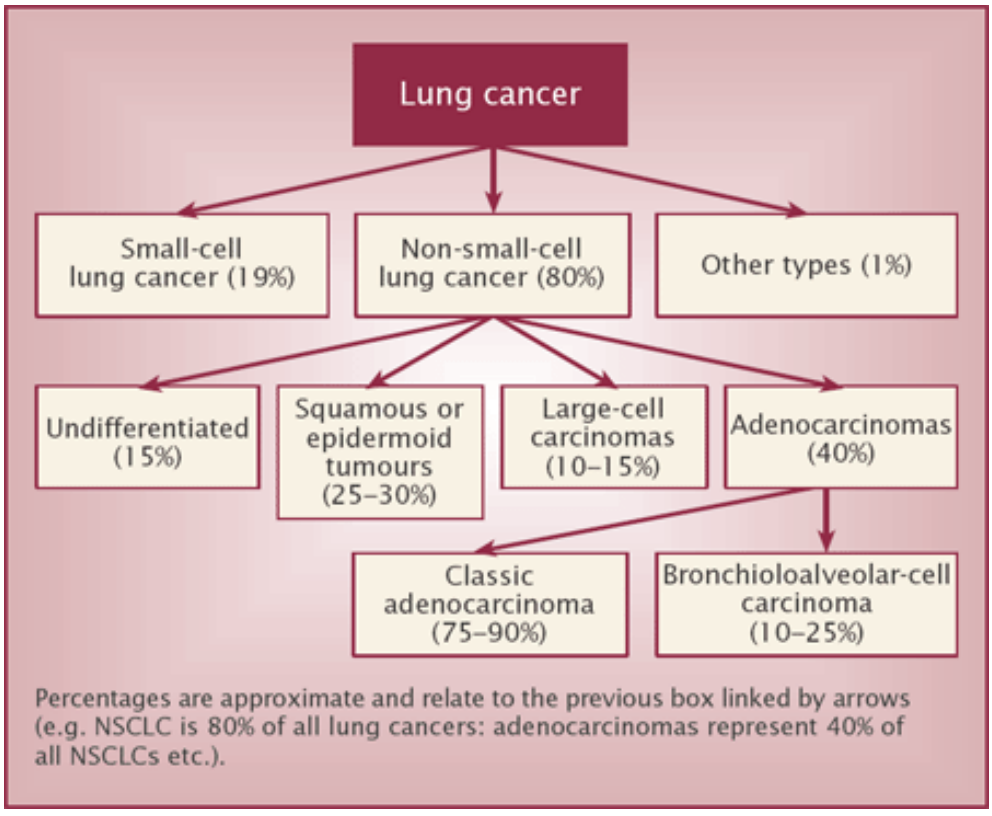



**Figure 2.1: Histological classification of lung cancer**

SCC is the commonest histological type in India whereas adenocarcinoma is gradually becoming the predominant subtype in the Western world [Thippanna et. al., 1999]. The clinical profile of lung cancer in India differs from the West, in that Indian patients present almost 15-20 years earlier, in the 5th or 6$^{th}$ decades of life [Jindal et. al., 1990].

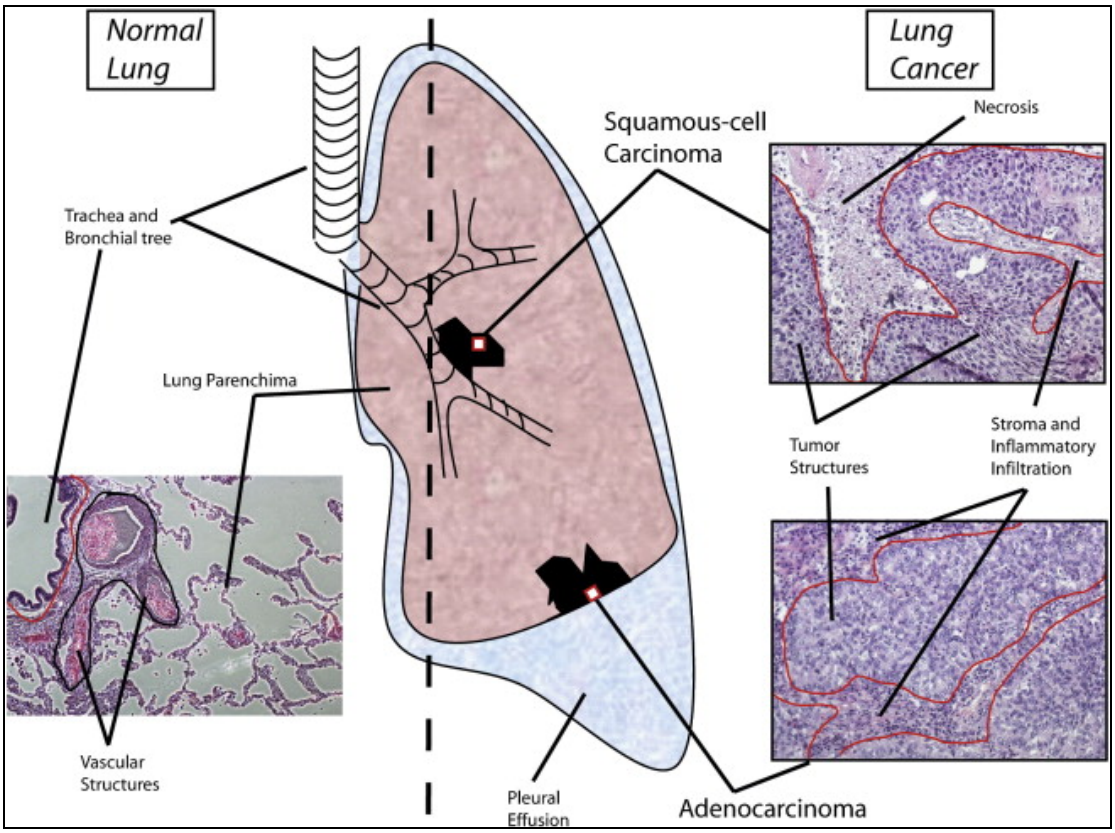



**Figure 2.2: Schematic representation of the anatomical and histological structures of normal lung and lung cancer (Adapted from Lehtiö et al., 2010).**

**Small cell lung cancers (SCLCs)** account for 20%-25% of lung cancers, tend to grow quickly and are classified simply as either limited or extensive stage. Small cell lung cancers have the poorest prognosis, are inoperable and therefore are generally treated through chemotherapy and radiation therapy. Under the microscope they form sheets of darkly staining cells with prominent nuclei and little cytoplasm. Their secretory activity can be seen as the presence of neurosecretory granules in the cytoplasm seen by electron microscopy (Figure 2.3). This form is very strongly linked to smoking as a causative factor.

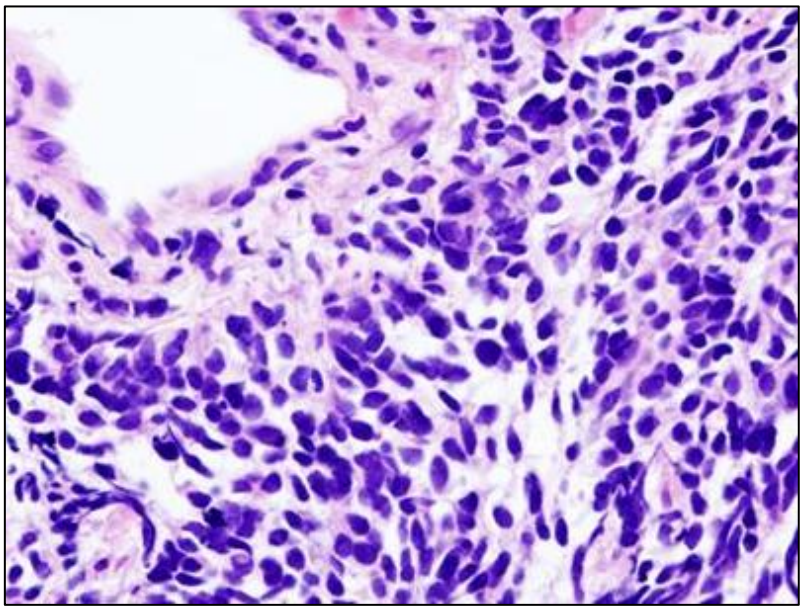


**Figure 2.3: Small cell lung carcinoma (microscopic view of a core needle biopsy)**

## STAGING IN LUNG CANCER

NSCLC is staged using the traditional TNM solid tumour staging system which is based on tumour size, nodal status and presence or absence of metastases [Sobin et. al., 2002]. TNM staging is used to group NSCLCs more broadly into 4 stage categories: I, II, III and IV, the first three of which can be subdivided into A and B subtypes (Figure 2.4). Stage I cancers are confined to the lung and are no larger than 5cm while stage II cancers may have some limited spread beyond the primary tumour and are no larger than 7cm [Sobin et. al., 2002].. Stage IIIA cancers are characterized by greater spread within the lung itself or connected organs (excluding the opposite lung) or lymph nodes on the same side of the chest [Sobin et. al., 2002]. Stage IIIB cancers are those which have greater spread into connecting organs and/or lymph nodes above the collar bone or on the contralateral side of the body [Sobin et. al., 2002]. Stage IV consists of disease that has metastasized either to the opposite lung, the fluid surrounding the lungs or heart or to other more distant parts of the body including the brain, liver, and bones [Sobin et. al., 2002]. Stages III (usually restricted to IIIB) and stage IV are frequently collectively referred to as ‘advanced’ stage disease. Approximately 34% of patients are diagnosed with stage I or II, 27% diagnosed with stage III, and 39% diagnosed with stage IV [Morgensztern et. al., 2010].

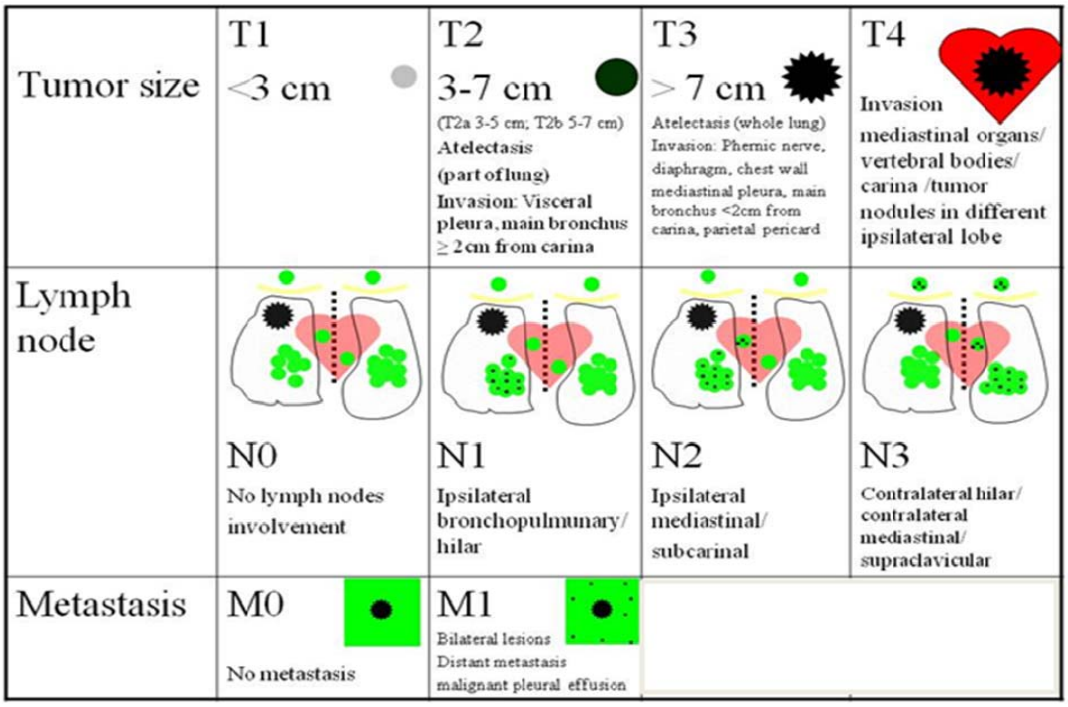



**Figure 2.4: TNM staging in non small cell lung cancer.**
(Adapted from: http://www.thebestoncologist.com)

## EPIDEMIOLOGY OF LUNG CANCER

### Worldwide

Lung cancer is the leading cause of cancer-related mortality worldwide, with nearly 1.4 million deaths each year [Jemal et. al., 2010]. It accounted for 13% (1.6 million) of the total cases and 18% (1.4 million) of the deaths in 2008 [Jemal et al., 2011]. Lung cancer is the leading cancer site in males, comprising 17% of the total new cancer cases and 23% of the total cancer deaths [Jemal et al., 2011]. The incidence is 2 to 5 times higher in developed countries compared with developing countries due variations in a disparate set of risk factors, cancer awareness and diagnostic practices available [Jemal et al., 2011]. In males, the highest lung cancer incidence rates are in Eastern and Southern Europe, North America, Micronesia and Polynesia, and Eastern Asia, while rates are low in sub-Saharan Africa (Figure 2.5). In females, the highest lung cancer incidence rates are found in North

America, Northern Europe, and Australia/ New Zealand. Several countries or registries in Asia have higher shown lung cancer rates than U.S. Asians for both men and women [Jemal et. al., 2010]. International variations in lung cancer rates and trends largely reflect differences in the stage and degree of the tobacco epidemic. Environmental exposures other than smoking such as, radon and asbestos, certain metals (chromium, cadmium, and arsenic), some organic chemicals, radiation, air pollution, coal smoke, and indoor emissions from burning other fuels also contribute to regional variation in lung cancer rate.

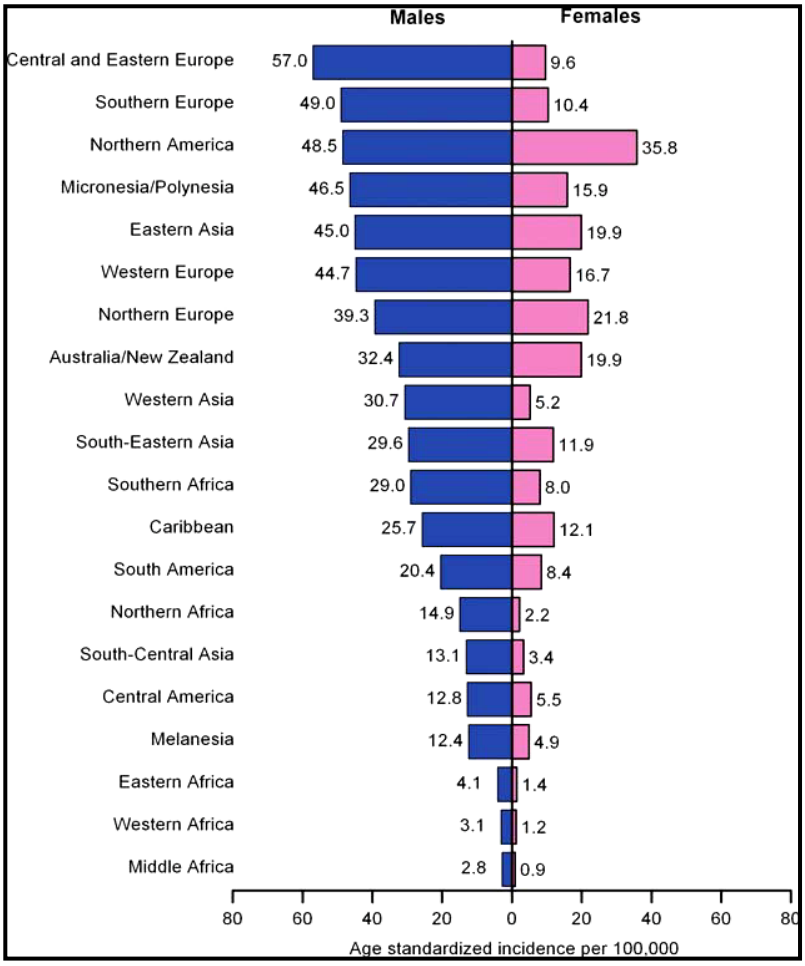



**Figure 2.5: Age-Standardized Lung Cancer Incidence Rates by Sex and World Area. Source: GLOBOCAN 2008.**

## India

In India, lung cancer constitutes 6.2% of all cancers with approximately 58,000 incident cases reported in 2008 [Ferlay et. al., 2011]. It is the most common cancer among males in Chennai, Delhi, Mumbai and Bhopal and among ten most common cancers in other PBCRs viz. Bangalore, Ahmedabad and Barshi including both urban and rural population [NCRP 2008]. It is the third largest cause of cancer mortality in India accounting for nearly 8.3% of all cancer related deaths in the country [GLOBOCAN 2008]. Among males, it is the leading cause of cancer mortality, accounting for 13% of all cancer deaths [GLOBOCAN 2008]. Data from all urban and rural population-based cancer registries in India suggest that a steady rise in cancer incidences in North Eastern (NE) part of India. Lung cancer is among the ten leading sites, with the highest age-adjusted incidence rate (AAR) in Mizoram state (24.5 in males and 26.3 in females). Aizwal district alone shows an AAR of 36.0 in males and 38.7 in females which is almost three to ten times higher than Delhi [Zomawia et. al., 2010]. Incidence of lung cancer is also high among males in Silchar and Imphal districts (Figure 2.6-2.7). Among the metropolitan states incidence is highest in Kolkata among males and in Chennai among females. The northeastern region may be called as “hot-spot” for lung cancer and particularly in women of this region, which is akin to that seen in women in the western world. Further, the assessment of risk factors in Indian patients revealed that bidi smoking is the major risk factor for lung cancer in India in contrast to the cigarette or cigar smoking in USA [Notani et. al., 1974]. Further, in nonsmoking Indian women indoor air pollution due to domestic cooking fuels particularly the biomass fuel is a significant risk factor [Behera et. al., 2005].

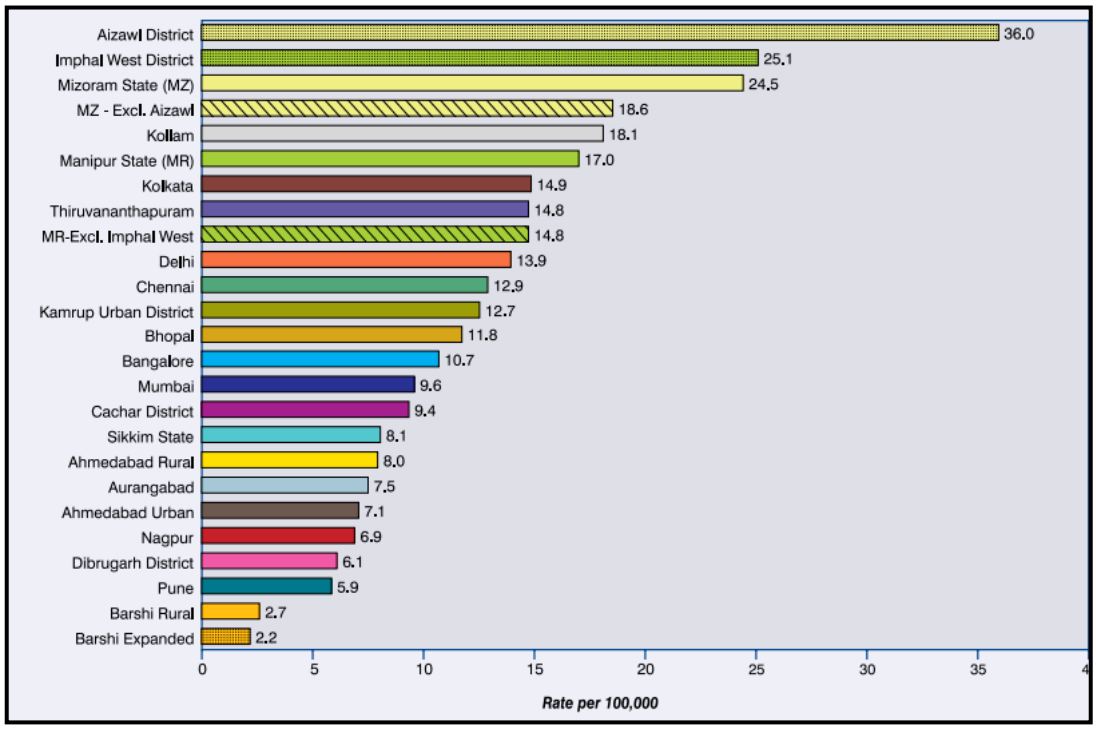



**Figure 2.6: Age adjusted incidence rates of all PBCRs for lung cancer in males (NCRP 2008)**

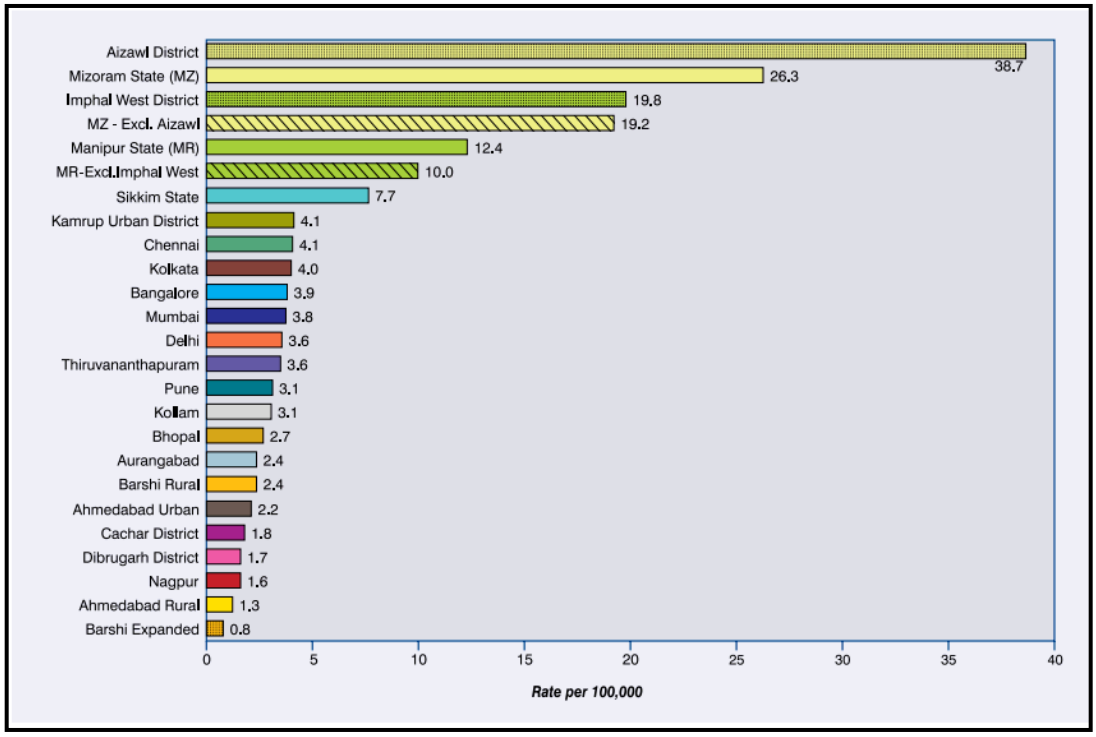



**Figure 2.7: Age adjusted incidence rates of all PBCRs for lung cancer in females (NCRP 2008)**

## ETIOLOGY OF LUNG CANCER

Like in many other solid malignancies, there is great individual variation in the susceptibility to develop lung cancer on exposure to respiratory carcinogens. The ultimate development of lung cancer is multifactorial, depending on the interrelationship between exposure to environmental agents (either etiologic or protective) and individual susceptibility to these agents. The environmental factors also refer to a broader set of interactions like socio-economic status, which is a constellation of different determinants for lung cancer risk. Some of the risk factors for lung cancer can interact synergistically and there are substantial gene-environment interactions. The predominant etiological factor for development of lung cancer remains to be tobacco smoking, while there is increasing knowledge about the other factors leading to the development of lung cancer in never smokers [Subramanian et. al., 2007]. A 'never smoker' is commonly defined as an individual who has smoked less than 100 cigarettes over his or her lifetime [WHO 1998]. The World Health Organisation estimates that worldwide 25% of lung cancer occurs in never smokers and it is the 7$^{th}$ largest cause of cancer-related mortality in the world [Parkin et. al., 2005]. Environmental tobacco smoke (ETS), household fumes, air pollution and occupational exposure to ionizing radiation, radon gas and asbestos are said causes of lung cancer in never smokers. Besides these, pre-existing lung diseases and inherited genetic susceptibility can also account for cancer risks among the group.

Conceptually, subjects with different risk factor patterns (at individual and societal levels) are exposed to different degrees of potential respiratory carcinogens. Upon exposure, there are various mechanisms in handling the carcinogens, in which the efficacy can be determined genetically. Some of the mechanisms are essential for activation of pro-carcinogens, while others are protective by inactivating the effects of carcinogens. The carcinogens or their metabolites, once escaped from the natural body defense mechanisms, may cause nuclear damage to human cells. These damages can be lethal or sublethal, depending on the particular agents and their levels. Under normal circumstances, some of these damages can be rectified by DNA repair mechanisms, which can also be determined by genetic factors. The accumulation of DNA damages, as in the multi-stage theory for carcinogenesis, can ultimately lead to malignant transformation and subsequent progression.

## Smoking

The discovery of cause and effect relationship between smoking and lung cancer development was one of the major landmarks in epidemiology over the past century. Smoking is now known to be the major risk factor for lung cancer and accounts for 90% of all cases [IARC 2004]. Risk to smoking is related to duration, intensity and age of initiation [IARC 2004]. Smoking accounts for about 80% of global lung cancer deaths in men and 50% of the deaths in women [Ezzati et. al., 2003 and 2005]. Epidemiology of lung cancer is a direct reflective of the trends of smoking around the world. Male lung cancer death rates are decreasing in most Western countries, including many European countries, North America and Australia where the tobacco epidemic peaked by the middle of the last century [Peto et. al., 2006, Jemal et. al., 2008, Bray et. al., 2010]. In contrast, lung cancer rates are increasing in countries such as China and several other Asian and African countries where the epidemic has been established more recently and smoking prevalence continues to either increase or show signs of stability [Lam et. al., 2004, Youlden et. al., 2008, Jemal et. al., 2010]. Lung cancer trends among females lag behind males because females started smoking in large numbers several decades later than males [Harris et. al., 1983].

The risks of lung cancer among tobacco smokers have been quantified, which vary according to duration of smoking and number of cigarettes smoked per day. There are various risk models to determine the contributory risks from the number of cigarettes smoked, the duration of smoking, and age. Among them, a quantitative model for lung cancer risk based on the data obtained from the cohort study of British physicians was proposed [Doll et al, 1978], which suggested a stronger effect of duration of smoking than amount smoked per day. Therefore, the exponential effect of the duration of smoking on lung cancer risk increases dramatically the lifetime risk for regular smokers since childhood, thus leading to development of lung cancer at younger ages. On the other hand, the lung cancer risks after smoking cessation also help to illustrate clearly the causal relationship between tobacco smoke and lung cancer. In fact, risk of lung is lower for smokers who abstain from smoking at any age and the cancer risk decreases progressively according to the duration of abstinence from smoking and the previous duration of smoking.

In India, smoking of tobacco is predominantly in the form of bidi, followed by cigarette, hukah, chillum and chutta [Chaudhry 2010]. According to National Family Health Survey (NFHS)-3 carried out during 2005-06, prevalence of tobacco use was 57% in men and 10.8% in women [IIPS 2007]. One third of men (33.4%) and 1.4% of women were cigarette/ bidi smokers. The number of adult current daily smokers is reported to be higher in the rural areas (31.3%) as compared to urban areas (21.5%) [IIPS 2006]. The relative risk of developing lung cancer is 2.64 and 2.23 for bidi and cigarette smokers respectively with 2.45 as the overall relative risk [Notani et. al., 1974]. This could be due to the fact that filtered cigarettes and lower tar yields slightly reduce the risk of lung cancer associated with cigarette smoking compared to the unfiltered bidis. However, smoking starts at somewhat older ages in India than it does in Europe and North America and the average daily consumption per smoker is lower [Gupta et. al., 1996, Gajalakshmi 2003].

Tobacco problem in the North-East is more complex than probably that any other state in India, with a large burden of tobacco related diseases and death. In the North Eastern states tobacco is smoked – in cigarettes, bidis and pipes. Tobacco smoking is relatively more common in the Northeastern states compared to the Northern and Southern states of India. Prevalence of tobacco smoking is high in Mizoram (59.4 %), Meghalaya (54.2%) followed by Tripura (48.5%) among men [Chaturvedi et al., 1998]. Further, a recent study reported prevalence of cigarette smoking of about 36.7 per cent among male school personnel and 10.0 per cent among females in North East India quite higher than other parts of India where it ranges between 13%-19% in males and 1%-4% in females [Sinha et al., 2007].Moreover, in North East, cigarette smoking is more prevalent than bidi smoking [Sinha et al., 2007].

**Environmental tobacco smoke**

Environmental tobacco smoke (ETS) can be derived from two different sources: side-stream smoke coming directly from a burning cigarette and mainstream smoke from the exhaled breath of the smoker. The presence of higher concentrations of some carcinogens like benzo(a)pyrene, nitrosamine and polonium has been found in side-stream compared with mainstream smoke [Lam et. al., 1988, Chan-Yeung et. al., 2003]. Therefore, it would be anticipated that ETS imposed an increased lung cancer risk. Passive smoking, the

involuntary inhalation of tobacco smoke by nonsmokers, also has been found to cause lung cancer. Passive smokers inhale a complex mixture of smoke that is widely referred to as environmental tobacco smoke (ETS). Early evidences of passive smoking were reported from Japan where one study showed that among nonsmoking women, those whose husbands smoked cigarettes were at higher risk for lung cancer than those whose husbands were nonsmokers [Hirayama et. al., 1981]. Since then there have been several studies on ETS and risk of lung cancer and meta-analyses studies have shown 20% increased risk for non-smoking women and 30% increased risk for non-smoking men ever lived with smoking spouses [Boffetta et. al., 2002, Brennan et. al., 2004].

**Tobacco smoke carcinogens**

Apart from the epidemiological evidence of smoking associated with lung cancer development, there are also biological reasons to explain the relationship. Tobacco smoke has been known to contain over 4000 chemicals of which more than 60 known carcinogens have been detected in cigarette smoke based upon an evaluation by the International Agency for Research on Cancer (IARC 1987). Important carcinogens include polycyclic aromatic hydrocarbons, nitrosamines, aromatic amines, chromium and polonium. Figure 2.8 lists several carcinogens detected in tobacco smoke. The total amount of carcinogens in cigarette smoke adds up to 1–3 mg per cigarette. Cigarette smoke contains a mixture of carcinogens, including a small dose of polycyclic aromatic hydrocarbons (PAHs) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) among other lung carcinogens, tumor promoters, and co-carcinogen. Considerable evidence favors PAHs and NNK as major aetiological factors in lung cancer.

| *Chemical class* | *Number of compounds‡* | *Representative carcinogens and typical amounts in mainstream smoke (ng per cigarette)§* | |
|---|---|---|---|
| PAH‖ | 14 | BaP | 9 |
| | | Dibenz[*a*,*h*]anthracene | 4 |
| Nitrosamines | 8 | NNK | 123 |
| | | NNN | 179 |
| Aromatic amines‖ | 12 | 4-Aminobiphenyl | 1.4 |
| | | 2-Naphthylamine | 10 |
| Aldehydes | 2 | Formaldehyde | 16,000 |
| | | Acetaldehyde | 819,000 |
| Phenols | 2 | Catechol | 68,000 |
| Volatile hydrocarbons | 3 | Benzene<br>1,3-Butadiene | 59,000<br>52,000 |
| Nitro compounds | 3 | Nitromethane | 500 |
| Other organic compounds | 8 | Ethylene oxide<br>Acrylonitrile | 7,000<br>10,000 |
| Inorganic compounds | 9 | Cadmium | 132 |
| **Total** | **61** | | |

**Figure 2.8: Types of carcinogen in tobacco smoke (Adapted from Hecht et. al., 2003)**

PAH are a group of more than 100 chemicals which result from incomplete combustion of tobacco and other organic products, many of which are known carcinogens. Among the PAHs, benzo(a)pyrene (BaP) is the most extensively studied compound, and its ability to induce lung tumors upon local administration or inhalation is well documented [IARC 1972 and 1983]. PAH undergo metabolic activation and, as a first step in the carcinogenic process, can form DNA and protein adducts (Figure 2.9). Polycyclic aromatic hydrocarbons (PAH) have been found to induce mutations in the p53 gene, which is crucial for cell cycle dysregulation and carcinogenesis. G to T transversion in the p53 gene has been considered as a molecular signature of tobacco mutagens in smoking-related lung cancers due to the following reasons [Hainaut et. al., 2001, Vineis et. al., 2004]: (1) PAHs are the main carcinogens in tobacco smoke causing G to T transversions; (2) PAH adducts are present in DNA extracted from human tissues exposed to tobacco smoke; (3) G to T transversions are more frequently found in lung cancers from smokers compared to non-

smokers; (4) a non-transcribed strand bias of G to T transversions can be attributed to the preferential repair of adducts on the transcribed strand. NNK is a tobacco-specific carcinogen and levels of its metabolites directly represent markers of tobacco effects [Carmella et. al., 1995, Anderson et. al., 2003]. Apart from adduct formation, the smoke components also induce a variety of genetic and epigenetic changes involved in transforming a normal cell into tumor cell. They induce sister chromatid exchange, oxidative damage as well as mutations in tumor suppressor genes and oncogenes. Besides the carcinogens cigarette smoke is also contain tumor promoters. Substantial levels of catechols, acrolein and other agents such as nitrogen oxides, acetaldehyde and formaldehyde contribute indirectly to pulmonary carcinogenicity by increasing the carcinogenic activity of carcinogens.

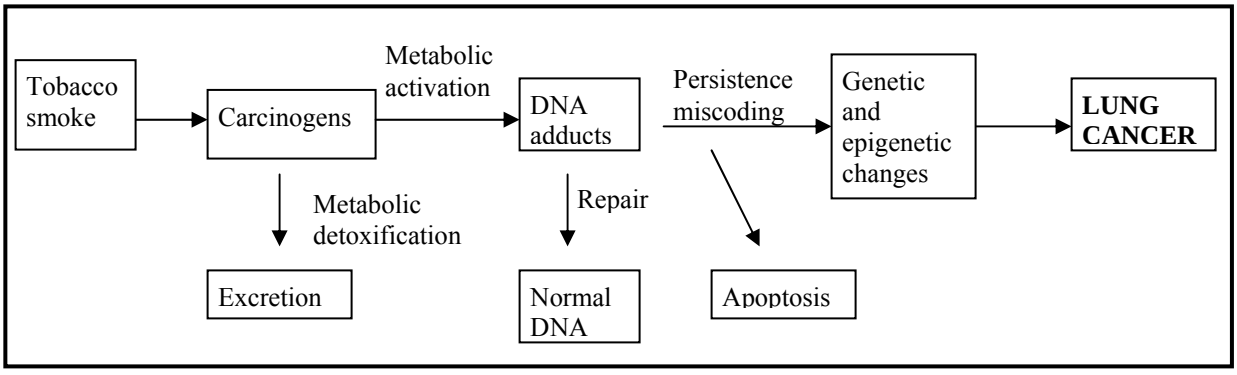



**Figure 2.9: Schematic representation linking tobacco smoke and lung cancer through tobacco smoke carcinogens**

**Occupational risk**

Lung cancer is the most common form of cancers associated with occupational exposures to potentially carcinogenic chemicals [Doll et. al., 1981). Environmental exposures in certain occupations have shown to be related to lung cancer [Jockel et. al., 1992]. The International Agency for Research on Cancer has identified 12 occupational exposure factors as being carcinogenic to the human lung [aluminum production, arsenic, asbestos, bis-chloromethyl ether, beryllium, cadmium, hexavalent chromium, coke and coal gasification fumes, crystalline silica, nickel, radon, and soot). The most notable

occupational carcinogens include tar and soot [Doll et. al., 1965, Lawther et. al., 1965, Lloyd 1971], and heavy metals like arsenic [Lee et. al., 1969, Ott et. al., 1974], chromium [Bidstrup et. al., 1956, Alderson et. al., 1981] and nickel [Doll 1958, Kreyberg, 1978]. Combustion of fossil fuels is the main source of carcinogens such as polycyclic aromatic hydrocarbons and heavy metals like arsenic, nickel and chromium [Friberg et. al., 1978]. Also, augmented levels of arsenic in drinking water have been associated with an increase in the incidence of lung cancer. Asbestos are mineral fibers found naturally in rocks and widely used by industry. Exposure to asbestos fibers, such as chrysotile, amosite, anthophyllite and mixed fibers containing crocidolite, has resulted in a high incidence of lung cancer [IARC 1987]. Radon was established as a risk factor for lung cancer through studies on exposed individuals, for example uranium miners. Radon in indoor environments is considered to be a significant cause of lung cancer. Inhaled particles of radon generate alpha-emissions that cause DNA damage through double-strand breaks, large chromosomal aberrations, mainly deletions and also point mutations [Prise et. al., 2001, McDonald et. al., 1995]. In the absence of other causes of death, the risk of lung cancer for never smokers exposed to concentrations of 0, 100 and 400 Bq/m3 of radon are respectively 0.4%, 0.5% and 0.7%. In ever smokers, these risks are approximately 25 times greater [Darby et. al., 2005]. Individuals employed in the stone masonry, quarrying or ceramics industries are constantly exposed to crystalline silica dust. A multicentric study conducted in Europe among 6,000 individuals suggested that occupational exposure to crystalline silica is carcinogenic [Cassidy 2007]. However, other studies have suggested silicosis to be associated with higher risk of lung cancer suggesting an indirect role of crystalline silica [Kurihara et. al., 2004]. A study in Mumbai showed a significantly elevated risk (adjusted for smoking) for textile workers, ship and dockyard workers and hand wood workers [Notani et. al., 1993]. Higher lung cancer rates in Chinese women with low smoking prevalence reflect indoor air pollution from unventilated coal-fueled stoves and from cooking fumes [Boffetta et. al., 2003, Thun et. al., 2008]. Outdoor air pollution, which includes combustion generated carcinogens, is also considered to contribute to the lung cancer burden in urban population.

Compared to tobacco smoking, the risk from occupational exposures remains small, but relatively large compared to most other exposure classes. Also, tobacco

smoking has been shown to enhance the effect of known occupational carcinogens for lung cancer [Saracci et. al., 1994]. Hammond et al. reported asbestos and cigarette smoking acting synergistically in causing lung cancer with a multiplicative effect [Hammond et. al., 1979], as smoking may enhance the retention of asbestos fibers in the lungs [Churg et. al., 1995].

**Family history**

Tokuhata and Lilienfeld provided the first epidemiological evidence of familial aggregation of lung cancer, suggesting the interaction of genes, shared environment and common lifestyle factors in the aetiology of the disease [Tokuhata et. al., 1963]. Since then several studies have found evidence of a close association between the development of lung cancer and hereditary factors. One study on association between lung cancer incidence and family history of lung cancer reports that family history of lung cancer in a first-degree relative was associated with a significantly increased risk of lung cancer [Nitadori et. al., 2006]. The association was stronger in women than in men and in never-smokers than in current smokers. These findings support the hypothesis that genetic susceptibility to lung cancer might act as both an independent risk factor and an effect modifier of environmental risk factors.

**Diet and lung cancer**

The main dietary factors in relation to lung cancer risk include fruits, vegetables and specific antioxidant micronutrients, which are mostly built on the hypothesis that antioxidants may protect against oxidative DNA damage leading to prevention of cancer. There is increasing evidence that some dietary factors may increase the risk of lung cancer. β-carotene is hypothesized to have a protective role. Reports from Western population show that persons with the lowest intake of foods rich in beta-carotene had the highest risk for lung cancer [Sikora et. al., 1990]. Animal experiments suggested β-carotene might exert prooxidant effect under an oxygen pressure of 100%, with potential cooperative interaction with α-tocopherol [Lee et. al., 2003]. In smokers' lungs, there is a rich source of free radicals which may facilitate alteration of β-carotene and formation of oxidated metabolites, thus further facilitating carcinogenesis [Lee et. al., 2003]. Smoking with

deficiency of Vit A increases the chance of developing squamous cell carcinoma. Deficiency of retinoids leads to squamous cell transformation and there is increased B(a)-P DNA adduct formation. This is reversed by addition of retinoids. However none of the three large clinical trials showed a chemoprotective effect from betacarotene or vitamins A or E dietary supplements [Shekelle et. al., 1981]. Plant carotenoids alpha-carotene (found in carrots and tomatoes) and lycopene (found in tomatoes) are associated with 20-25% lower risk of lung cancer [Behera et. al., 1998].

## GENETICS OF LUNG CANCER

Although smoking is the primary risk factor for most lung cancers, genetic predisposition may play an important role. Familial aggregation studies suggest a greater genetic component in the risk for younger individuals developing lung cancer, for lifetime nonsmokers, and possibly for women. Molecular genetic studies have also shown that multiple genetic loci contribute to sporadic lung cancers. The molecular abnormalities are found in both growth-promoting oncogenes and growth-suppressing tumor suppressor genes. Tumor suppressor genes reported to be involved in lung cancer include p53, p16 and Rb. Cytogenetic studies have identified many chromosomal changes in lung cancer with numerical abnormalities, and structural aberrations including deletions and transolocations. These mutations include activation of the dominant cellular protooncogenes (which promote oncogenesis) of the ras and myc family and inactivation of the recessive or tumor suppressor genes (these genes help suppression of tumor development). Small cell Lung cancer is associated with oncogenes like c-myc, L-myc, N-myc, c-raf and tumor suppressor genes like p53 and Rb. Non small cell lung cancer is associated with K-ras, N-ras, H-ras, c-myc, c-raf and tumor suppressor genes like p16 and Rb. Along with these, abnormal transcription of the FHIT gene was reported in 40% of NSCLC [Sozzi et. al., 1997] and its function was related to proapoptotic signaling. DLC1 (deleted in lung cancer 1) was cloned through large-scale sequencing of 3p21.3 and was found to show aberrant or no transcription in primary NSCLC [Wang et. al., 2007]. Recently, p34 and CYGB, genes previously implicated only in sporadic head and neck cancers, were also added to the list of candidate tumor suppressor genes involved in the pathogenesis of lung.

Genetic polymorphisms are common variations in the genetic code, typically defined as comprising at least 1% of the population or sample of interest. A single nucleotide polymorphism (SNP) is a commonly found alteration in the DNA sequence at a single nucleotide locus. Candidate gene approaches of typing SNPs have frequently employed the following criteria for investigation (i) biological plausibility of risk modification by the enzyme; (ii) known or suspected phenotypic relevance of the genetic polymorphism; as well as (iii) medium to high frequency of polymorphism in Caucasian population (to ensure public health relevance of results). Polymorphism in low-penetrance, high-prevalence genes may explain a greater part of genetic predisposition. Genes with mechanistically plausible variants coding for enzymes involved in the activation, detoxification and repair of damage caused by tobacco smoke and inflammatory pathways, have been extensively studied. These genes influence lung cancer risk as a result of gene-environment interactions. Alterations in these pathways are hypothesized to affect an individual's processing of tobacco carcinogens and therefore risk of developing lung cancer. Genetic polymorphisms can affect lung cancer risk includes effects on smoking behavior and nicotine addiction, carcinogen metabolism, carcinogen detoxification, DNA repair, cell cycle control, apoptosis, signal transduction, and virtually every other part of cellular function and response.

In an alternative approach known as haplotype analysis, genes are again selected based on their known or putative involvement in important lung carcinogenic pathways [Yagil et. al., 2004]. A series of SNPs located in the same region of a chromosome are analyzed together as a unit known as haplotype blocks which remain conserved from one generation to another. SNPs are highly linked (or correlated) to other SNPs due to linkage disequilibrium. These SNPs discriminate between haplotypes and are known as haplotype-tagging SNPs (htSNPs). The htSNPs are used to construct common haplotypes, which are then evaluated for lung cancer risk susceptibility assuming that one or more of the SNPs within the haplotype block are functional. Risk associations observed with the haplotype will then serve as surrogates of the association with the unidentified functional SNP(s).

## GENETIC POLYMORPHISMS IN XENOBIOTIC-METABOLIZING GENES

Individual variability in xenobiotic or specifically tobacco carcinogen metabolism may partly explain differential susceptibility to lung cancer. Metabolism of PAHs, tobacco-specific nitrosamines and aromatic amines in cigarette smoke occurs via two classes of enzymes: phase I enzymes that carry out oxidation/reduction/hydrolysis reactions to metabolically activate procarcinogens to genotoxic electrophilic intermediates and phase II enzymes that conjugate the intermediates to water-soluble derivatives, thus completing the detoxification cycle (Figure 2.10). Many of the xenobiotic metabolizing enzymes show polymorphisms which have been well characterized and are known to affect the enzyme activity.

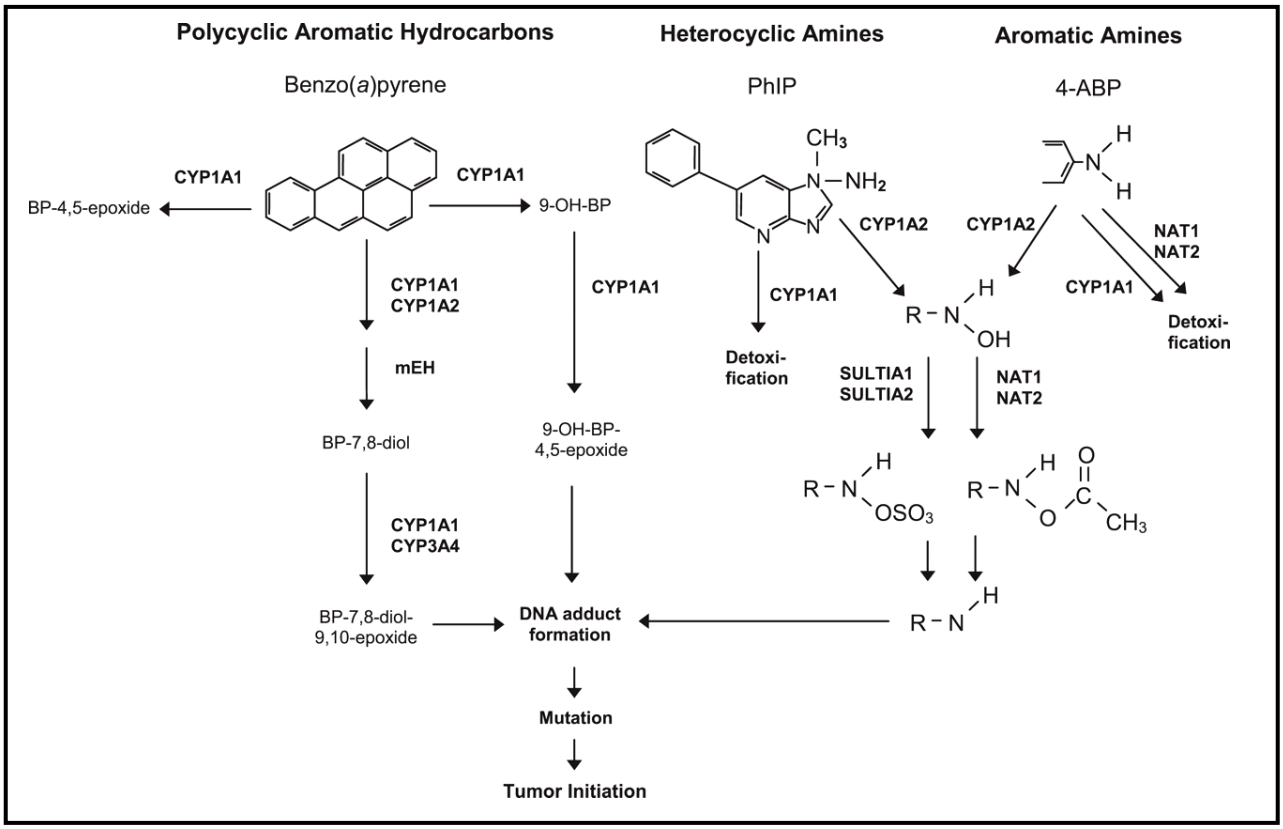



**Figure 2.10: Schematic representation of Xenobiotic metabolization of tobacco smoke carcinogens by Phase I and Phase II enzymes. [Adapted from Goode et. al., 2007]**

**Gene Name:** CYP1A1: Cytochrome P450 1A1; SULT1A1: Sulfotransferase 1A1; mEH: Microsomal epoxide hydrolase; NAT1: N-acetyltransferase 1

## CYP1A1

Phase I enzymes activate PAH and arylamines that ultimately produce carcinogen-DNA adducts. Benzo[a]pyrene, the prototypical PAH, is first oxidized by P450 enzyme CYP1A1 into an arene oxide. Several CYP1A1 genetic polymorphisms have been identified. Evidence for an association between CYP1A1 polymorphisms and risk of lung cancer originally came from studies in Japanese populations where variant alleles occur at higher frequencies than in Caucasians, with reports of 0.2-fold increased risk [Kawajiri et. al., 1990]. There are two known functional polymorphisms in the CYP1A1 gene. The m1 polymorphism is a 6235T>C substitution (CYP1A1*2A, rs 4646903) creating an MspI restriction site in the 3'-flanking region, which has also been associated experimentally with increased catalytic activity [Landi et. al., 1994]. The m2 polymorphism in CYP1A1 is a Ile-Val mutation in the exon 7 (CYP1A1*2C, rs 1048943), a heme-binding region, resulting in a 2-fold increase in microsomal enzyme activity and is in complete linkage disequilibrium in Caucasians with the CYP1A1 MspI (m1) mutation. Although the Ile-Val mutation in the CYP1A1 allele did not increase activity in vitro [Zhang et. al., 1996], it might be linked to other functional polymorphisms, for example in the regulatory region important for CYP1A1 inducibility.

Co-relational studies are available between CYP1A1 variants and alone or in combination with GSTM1, and the formation of bulky (PAH)-DNA adducts in human tissues and leukocytes. However the findings are controversial. Smokers with the exon 7 Ile-Val mutation were found to have more PAH-DNA adducts in their white blood cells than smokers without the variant [Mooney et. al., 1997]. Several studies have showed a weak or no effect of m1 and m2 on adduct levels [Ichiba et. al., 1994, Schoket et. al., 1998]. Other studies have shown that lung and leukocytes of Caucasian smokers with the CYP1A1 m1/m1-GSTM1 0/0 combination clearly contained more BPDE-DNA adducts [Rojas et. al., 1998 and 2000]. Associations of CYP1A1 polymorphism with lung cancer varies with ethnicity. A pooled analysis suggested that genetic polymorphisms in CYP1A1 are associated with lung cancer risk among Asian populations [Lee et. al., 2008]. In a pooled analysis using data from 22 studies, a significant 2.4-fold increased in risk was observed in individuals carrying the MspI variant [Vineis et. al., 2003]. Substantial studies

have reported no association of CYP1A1*2A allele with lung cancer in Caucasian population [Shields et. al., 1993, D'Errico et. al., 1999]. The CYP1A1*2C polymorphism has consistently been associated with lung cancer risk in Asian subjects [Nakachi et. al., 1993, Kihara et. al., 1995] whereas reports in Caucasian population have been more variable [Le Marchand et. al., 2003, Taioli et. al., 2003, Vineis et. al., 2007]. Much of this variability can be attributed to smoking, since the mutation is believed to be important among light- and non-smokers but not among heavy-smokers [Le Marchand et. al., 2003, Taioli et. al., 2003, Hung et. al., 2003].

**EPHX1**

Microsomal epoxide hydrolase (EPHX1) another important Phase I biotransformation enzyme is involved in the first-pass metabolism of highly reactive epoxide intermediates and oxygen radicals. It catalyzes the hydrolysis of various epoxides and reactive epoxide intermediates into less reactive and more water soluble dihydrodiols. EPHX1 plays dual role in carcinogenesis depending on the exposure to type of environmental substrates. Besides providing protection against the toxicity of reactive epoxides intermediate, EPHX1 along with CYP enzymes play a key role in the metabolic activation of procarcinogens such as benzo(a)pyrene (BP) present in tobacco smoke leading to highly reactive carcinogenic diol-epoxides [Miyata et. al., 1999]. Two relatively common genetic polymorphisms, one in exon 3 (T>C, Tyr113His) and other in exon 4 (A>G, His139Arg) produce two protein variants that have been shown to influence the enzyme activity. The exon 3, 113His allele shows reduced enzyme activity by at least 50% (slow allele) whereas exon 4, Arg139 allele has increased activity by at least 25% (fast allele) [Hassett et. al., 1994a and 1994b]. Some previous reports have studied the relation between lung cancer risk and the mEH predicted activity. Smith and Harrison [Smith et. al., 1997] studied a group of lung cancer patients of Caucasian origin and found an association between predicted 'low-activity' mEH and susceptibility to emphysema but they did not confirm a similar association with lung cancer risk. Conversely, Benhamou et al. [1998] reported an association between predicted 'high-activity' mEH and lung cancer risk among French Caucasians. The study of London et al. [2000] in Los Angeles County found an association between susceptibility to lung cancer and predicted 'high' activity among African-

Americans although the same study reported that a similar association was not found among white Caucasians. Lin et al. [2000] also found a significant risk excess of squamous cell carcinoma (but not other histological types of lung cancer) with high mEH activity.

**GST**

Glutathione S-transerases (GSTs) are phase II biotransformation enzyme that catalyzes the transfer of glutathione to reactive electrophiles, thus protecting cellular macromolecules from interacting with electrophiles containing electrophilic heteroatoms (-O, -N, and -S) thereby protecting the cellular environment from damage. The co-substrate in the reaction is the tripeptide glutathione, which is synthesized from γ-glutamic acid, cysteine, and glycine. Glutathione exists in the cell as oxidized (GSSG) or reduced (GSH), and the ratio of GSH: GSSG is critical in maintaining a cellular environment in the reduced state. The addition of GSH to the xenobiotic compound gives it a molecular 'flag' which allows the xenobiotic-conjugate to be removed from the cell during phase III of drug metabolism, a process which requires the participation of drug transporters such as multi-drug resistance associated protein [Hayes et. al., 1999]. GSTs have evolved with GSH, and are abundant throughout most life forms.

GSTs are divided into two distinct super-family members: the membrane bound microsomal and cytosolic family members. Cytosolic GSTs are subject to significant genetic polymorphisms in human populations and are divided into six classes [alpha (GSTA2, mu (GSTM1), pi (GSTP1), theta (GSTT1), zeta (GSTZ1) and omega (GSTO1)], which share ~30% sequence identity. GSTT1 is considered as one of the most ancient enzymes among GSTs and it exhibits a different catalytic activity compared to other GSTs. The GSTM1 and GSTT1 both exhibit deletion polymorphism that results in lack of enzyme activity. GSTP1, located on chromosome 11 (11q13), encodes the major enzyme involved in the inactivation of tobacco-related procarcinogens. The GSTP1 Ile105Val polymorphism is associated with reduced catalytic activity which may result in an increased susceptibility to cancer [Ali-Osman et. al., 1997]. Some studies have suggested that 105Val has different enzyme heat stability and affinity, and lower 1-chloro-2,4 dinitrobenzene (CDNB) conjugating activity, whereas other in-vitro studies have

shown that GSTP1 105Val was more active in conjugation reactions toward carcinogenic diol epoxides of PAHs.

The prevalence of GSTM1 null is around 50% and GSTT1 null genotype ranged from 15–25% in Caucasians to 60–80% in Asians. One of the first meta-analyses conducted of GSTM1-null variants [Houlston et. al., 1999] showed a modest increase in lung cancer among carriers of the GSTM1-null genotype (OR 5 1.13, 95% CI 1.04–1.25). A larger meta-analysis reported that lung cancer risk increased by 17% in those who were GSTM1 null (95% CI 1.07–1.27) (Benhamou et. al., 2002]. In another meta analysis of 130 studies an 18% increased risk of lung cancer was observed among individuals with the GSTM1-null genotype (95% CI 1.14–1.23), but when analyzing data only from the larger studies there was no association [Ye et. al., 2006]. Studies on GSTT1 and lung cancer report both no association and increased risk in carriers of the null allele. In a studies reporting their results stratified by smoking, the GSTT1-null genotype was found to increase the risk of lung cancer in smokers in some [Hou et. al., 2001, Nazar-Stewart et. al., 2003] but not all [Malats et. al., 2000, Taioli et. al., 2003]. A meta-analysis summarizing the effect of GSTP1 on the risk of lung cancer, found an OR of 1.3 [1.1—1.6) for the GSTP1 variant allele [Stucker et. al., 2002]. Miller et. al., [2003] reported ORs stratified by smoking exposure and found an increased risk for lung cancer in carriers of the variant allele at all levels of smoking exposure whereas Perera et. al., [2002] reported increased risk in current and former smokers.

**SULT1A1**

Sulfotransferases (SULTs), the enzymes of phase II metabolism catalyzes the sulfation of a variety of phenolic and estrogenic compounds including endogenous and environmental estrogens. SULT1A1, a member of phenol SULT1 family, is important due to its extensive tissue distribution and abundance [Glatt et. al., 2000]. The Arg213His polymorphism identified in the SULT1A1 gene has functional consequences for the translated protein in the variant allele (His213, SULT1A1*2) [Nowell et. al., 2000]. Association of this polymorphism and risk of cancer are inconsistent, from null association with risk of colorectal cancer [Wong et. al., 2002] and prostate cancer [Steiner et. al., 2000] to increase in risk of breast cancer associated with His213 allele [Zheng et. al., 2001]. Another study

on colorectal cancer showed a significantly reduced risk for individuals carrying His213 allele [Bamber et. al., 2001]. Study [Wang et. al., 2002] on lung cancer in Caucasian population reported that the variant A allele of SULT1A1 was associated with an increased risk of the cancer (OR, 1.41; 95% CI, 1.04--1.91).

Lung cancer is a multifactorial disease involving complex interactions between multiple genes and environmental risk factors. The low-penetrance feature of individual genes may be responsible for the lack of consistency in cancer association studies of common SNPs. A pathway-based multigenic approach to assess the combined effects of a panel of polymorphisms that interact and function in the same pathway should be applied to cancer association studies to reveal complex gene–gene and gene–smoking interactions in modulating cancer risk. Data mining approaches, in analysis, further explore high-order gene–gene and gene–environment interactions in cancer susceptibility.

## COPY NUMBER POLYMORPHISM OF GSTM1 AND GSTT1 IN LUNG CANCER

The majority of polymorphisms affecting genes involved in carcinogen metabolism are single nucleotide polymorphisms. SNP account for >90% of the variations while deletions are less common and the complete absence of a gene in form of a null allele is rare. Null polymorphism is quite evident in GST genes. GST enzymes are encoded by eight distinct loci including alpha, kappa, mu, omega, pi, sigma, theta and zeta. Among these GST isoforms, GSTM1 and GSTT1 is of particular interest because both genes possesses a null polymorphism, referred to as GSTM1-null and GSTT1 null which results in a complete absence of GSTM1 enzyme activity. The GST null allele generally arises from homologous unequal crossing over between two highly identical sequences flanking the GST gene. A GSTM1 null allele is thought to result from homologous unequal crossing over between two highly identical 4.2 kb repeated sequences flanking the GSTM1 gene, resulting in a 15 kb deletion including the entire GSTM1 gene [Sprenger et. al., 2000]. A similar mechanism results in the GSTT1 null allele [Xu et. al., 1998].

In view of the importance of glutathione S-transferases in cellular detoxification, the enzyme deficiency associated with the null genotypes has attracted considerable attention with regard to cancer epidemiology. For this reason, more than 700

studies of GSTM1 and GSTT1 genotypes in relation to lung, breast, colon, brain, and with various other types of cancer has been published. Complete gene deletions (loss of functional gene copy) of GSTM1 and GSTT1 are relatively common and their frequencies vary between populations. In Caucasians, the frequencies of homozygous deletions are approximately 50% for GSTM1 and 30-30% for GSTT1 respectively, compared with >22% and >14% in Asians, and >27% and >37% in Africans [Zhang et. al., 1999; Garte et. al., 2001; Neri et. al., 2006; Piaccentini et. al., 2011].

However, most of these studies investigating the effect of genetic polymorphisms in GSTM1 and GSTT1 do not distinguish between individuals with one or two copies of the genes which resulted in inconsistent or contradictory publications on the association of the genotypes with various malignancies. True genotyping is important because of the gene dosage effect associated with having two, one, or no alleles. In these genes, a trimodal phenotype pattern exists in which individuals with two, one, or no functional enzymes are fast, intermediate, and slow conjugators respectively [Sprenger et. al., 2000, Covault et. al., 2003]. Only few studies have been able to distinguish between genotypes with one or two copies of GSTM1 or GSTT1.

A limitation of earlier studies was the PCR assay, which did not truly genotype GSTM1 but only identified -/- homozygosity without being able to separate the +/- and -/- genotypes. New methods using either long-range PCR or real-time PCR are available to definitively identify +/+, +/−, and −/− genotypes. An example of the scientific advantage of this unambiguous genotyping is provided by Moore et al., [2005], which showed that associations between colorectal adenomas and GSTM1 wild-type and GSTT1 null alleles only became apparent with a real-time PCR assay that distinguished heterozygous from wild-type individuals. Similarly another study of GSTM1 in breast cancer, which used real-time PCR to reveal a more complicated association between GSTM1 allele numbers and risk than would be apparent by using older methods [Yu et. al., 2009].

In lung cancer, meta-analyses have indicated that carriers of GSTM1 0/0 or GSTT1 0/0 have a slightly higher risk of lung cancer as compared with carriers of at least one functional allele [Raimondi et. al., 2006]. Though this review mainly comprised of

studies based on null and non-null approach. To our knowledge, only two studies that classified hemizygous deletion separately from wild-type have report on the associations between GSTM1 and GSTT1 genotypes and lung cancer risk observed no difference between the hemizygous and homozygous null genotypes of GSTM1 and GSTT1 when compared to the referent wild-type genotype [Sorensen et. al., 2007; Lam et. al., 2009]. However, the strength of association was attenuated when data was analysed using traditional genotyping classification. Thus, there is need to clarify the role of GSTT1 and GSTM1 genes based on the number of functional alleles of the genes with lung cancer risk.

## GENETIC POLYMORPHISM IN p53 GENE

The p53 gene is one of the most mutated genes in human tumors and has been referred to as the ‘‘emergency brake’’ because of its tumor-preventing apoptotic and cell-cycle-checkpoint functions in physiologically stressful situations. This gene is an important component in the response to DNA damage, participating in the DNA-repair process and preventing mutations and aneuploidy that result from cellular replication. Therefore, the wild-type p53 gene suppresses cellular transformation by activated oncogenes, thus inhibiting the growth of malignant cells. Mutations in p53 are found in over 50% of all human cancers [Hollstein et. al., 1996], comprising more than 50 different cells and tissue types, indicating that there is a powerful selection for loss of p53 activity during tumor development.

Studies have reported a relationship between tobacco smoke exposures, carcinogen–DNA adduct formation, tumor-specific mutation of TP53 gene leading probably to high cancer risk. p53 gene plays a significant role in the regulation of cellular response to benzo[a]pyrene, one of the most important polycyclic aromatic hydrocarbon compounds of tobacco smoke [Xiao et. al., 2007]. Recent studies have also documented that there is a strong coincidence in mutational hotspots and sites of preferential formation of polycyclic aromatic hydrocarbon adducts along the p53 gene in lung. Apart from mutation, genetic polymorphisms in p53 gene that alter the enzyme activity are also implicated in tobacco related cancers. Several polymorphisms have been identified in the TP53 gene [Olivier et. al., 2002]. However, most of these are single-nucleotide polymorphisms affecting a single base. Many of these natural variants are localized in non-

coding regions of the gene (introns). Among the polymorphisms found in the coding regions [exons) of TP53, only two alter the amino-acid sequence of its product, proline (P) to serine (S) at residue 47 and arginine (R) to proline (P) at residue 72.

The polymorphic variant at codon 72 which is a guanine to cytosine change (G>C) at codon 72 results in a arginine to proline (Arg>Pro) amino acid substitution. The polymorphism results in a structural change of the protein giving rise to variants of distinct electrophoretic mobility [Harris et. al., 1986]. This polymorphism occurs in a proline-rich region of p53, which is known to be important for the growth suppression and apoptotic functions of this protein [Sakamuro et. al., 1997]. The functional impact of this polymorphism has been reported and the Arg/Arg genotype seems to induce apoptosis with faster kinetics and to suppress transformation more efficiently than the Pro/Pro genotype [Thomas et. al., 1999]. The Pro allele induces higher levels of cell cycle arrest [Pim et. al., 2004], more efficient in activating p53-dependent DNA repair [Siddique et. al., 2006], less efficient in inducing apoptosis [Dumont et. al., 2003] and associated with an increased frequency of TP53 mutations in non-small cell lung cancer [Mechanic et. al., 2005, Szymanowska et. al., 2006]. The variant allele Pro could modify the response toward cell cycle arrest or apoptosis developing more mutations that increase the risk of developing cancer.

Beckman et. al. [Beckman et. al., 1994] first demonstrated a significant difference in the allelic distribution of the Arg72 and Pro72 variants. They first noted a significant difference in the Pro72 allele frequency between a Nigerian population (African Black) and a Swedish population (Western Europe), which were 17 and 63%, respectively; in contrast, they did not note any differences between populations living on the same geographical latitude. Furthermore, the frequency of the Pro72 allele differs with latitude, increasing in a linear manner as populations near the equator [Sjalander et. al., 1995]. This suggests that differences in activity of the variants might be subject to selection in areas of high ultraviolet light exposure. Ethnicity thus can be a confounder of the OR in genotype-disease association in case of this polymorphism. Likewise several studies have reported association of the p53 codon 72 polymorphism with lung cancer susceptibility whereas other studies have reported no significant differences between lung cancer and healthy controls in relation to genotypic polymorphic frequencies.

## GENE EXPRESSION PROFILING IN LUNG CANCER

Lung cancer formation and progression involves alterations in multiple genes. These alterations involve somatic gene mutations and regulation disturbance, causing dramatic gene expression changes in the tumor cells. The traditional approach to cancer research involves focusing on candidate gene or a discrete set of genes in any particular biological context. However this leads to many important biological changes either missed or uncovered in a serendipitous manner. Also, the candidate gene approach is largely limited by its reliance on the priori knowledge about the physiological, biochemical or functional aspects of possible candidates which in some cases may produce based results. A genome-wide scanning through microarray enables simultaneous measurement of the expression levels of thousands of genes in cells of a given biological sample. It usually proceeds without any presuppositions regarding the importance of specific functional features of the studied genes. Generation of vast amount of information, coupled with advances in technologies developed for the experimental use of such information allow for a description of biological processes from a view of global genetic perspective.

The identification of the gene expression profiles through microarray may help to better characterize human cancer. High-throughput expression profiling can be used to compare the level of gene transcription in clinical conditions in order to: 1) identify diagnostic or prognostic biomarkers; by seeking those genes that are induced or repressed which are then prioritized for further study. This typically involves going to the literature to acquire knowledge about known genes and often define potential new roles for known genes 2) classify diseases [eg, tumors with different prognosis that are indistinguishable by microscopic examination); This can be done by analyzing large-scale patterns of gene expression and deducing similarities within and among patient populations thereby classifying them more accurately. 3) monitor the response to therapy; and 4) understand the mechanisms involved in the genesis of disease processes. Biological pathway discovery remains the most tantalizing use of whole genome analysis with regard to understanding the complexities of large networks of interacting genes and their encoded proteins (Figure 2.11).

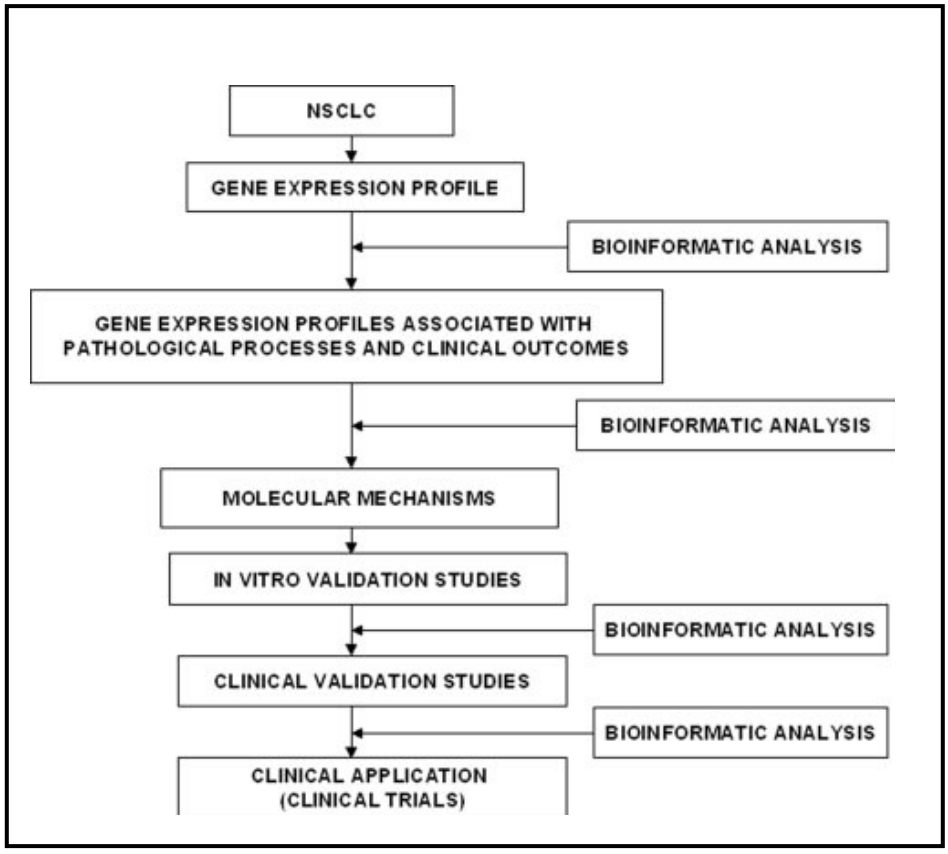



**Figure 2.11: Scheme for the use of gene expression profiling to elucidate the molecular mechanisms underlying important pathological processes and clinical outcomes with resulting clinical application and outcome in lung cancer. (Adapted from: Petty et. al., 2004)**

**Microarray: technique and data analysis**

Microarray technology provides a relatively rapid, reliable, reproducible, and quantitative approach for simultaneously monitoring expression levels of thousands of genes. Basically, the approach is to create a spotted array of thousands of different DNA molecules or probes corresponding to thousands of different genes. The probes are typically either oligonucleotide [Carvalho et. al., 2004], cDNA [Pollack et. al., 1999] or bacterial artificial chromosome (BAC) based sequences [Solinas-Toldo et. al., 1997]. Starting with an RNA sample, a series of biochemical reactions generates a fluorescently labeled cRNA or ss-cDNA probe which is hybridized to the microarray and scanned with a laser scanner. The expression levels are measured by the fluorescence intensity of bound

probe to each spot. A digital image is formed where the intensities reflect the expression of the genes. Image analysis techniques are applied to locate, segment, and quantify the spot intensities. At the time of segmentation and intensity quantification, the quality of the measurements can be assessed using the spatial features of the spot, e.g. its shape or size, or by comparing the spot foreground and background intensity distributions [Wang et. al., 2001, Li et. al., 2005]. Prior to further analysis, data preprocessing or normalization is usually performed. Normalization methods generally calculate a scaling factor or function to correct for non-biological effects in the data. Locally weighted scatter plot smoothing [LOWESS] [Cleveland 1979] has become a commonly used option in cases where there is, for example, a need to adjust for the dye bias.

Set norms for standardization of microarray experiments have been addressed and the Minimum Information about a Microarray Experiment (MIAME) has been established. It describes the information needed to enable the results of an experiment to be interpreted unambiguously allowing reproduction of the experiments [Brazma et. al., 2001, Barrett et. al., 2005]. Microarray data are commonly presented as a logarithmic ratio of the measurements between the sample and the reference. Various statistical approaches are available that can be used to identify changes associated with a specific outcome. The earliest and simplest methods use thresholds to detect changes in gene expression or apply statistical hypothesis testing such as t-test to detect differences in the means between the two given conditions. Another approach for analyzing the array data is the Receiver Operating Characteristic (ROC) curve [Swets 1988], which provides a non-parametric approach for analyzing the diagnostic value of the probes in a two-group classification setting. The curve displays the relationship between the proportion of true and false positive classifications, thus the area under the curve yields an estimate of a correct diagnosis when the probe is used to classify the groups.

Pathway analysis and gene ontology are data integration approach used to group genes into biologically meaningful categories and test the categories for deregulation or enrichment of differentially expressed genes. Investigations of groups of genes are justified as a change in the expression of one gene or protein influences the expression of other genes or proteins through the signaling cascades. Pathway analyses are frequently

performed on gene expression data as moderate changes in a number of components of one pathway may be enough to indicate differential regulation of the whole pathway. Furthermore, it has been shown that the consistency of analyses across independent data of a similar kind are remarkably improved by analyzing a defined set of genes that share a biological function, chromosomal location or regulation instead of single genes [Subramanian et. al., 2005].

The Gene Ontology (GO) project is a collaborative effort that aims at consistent descriptions of gene products in different databases [Ashburner et. al., 2000; Gene Ontology Consortium 2006]. The project has developed vocabulary terms for describing the biological processes, molecular functions, cellular pathways of genes and gene products in a species independent manner. Information about the deregulated pathways can be obtained by testing GO categories for the enrichment of differentially expressed genes. The enrichment calculations are performed on a predefined list of putative, differentially expressed genes or on lists of differentially expressed genes from the data [Breitling et. al., 2004].

In addition to analyzing global gene expression through microarrays, Real-time RT-PCR has also gained prominence as a reliable and specific approach for gene expression quantification and validation. The primary advantage of real time quantitation is the relative simplicity of experiments and high degree of analytical precision. Microarray studies can reveal changes in expression of a smaller number of genes that are used for subsequent hypothesis generation and testing. Once identified, the smaller gene set can be analyzed by real-time RT-PCR, which is better suited to analyzing multiple samples.

**Microarrays in lung cancer**

Microarray expression profiling has provided important information regarding lung carcinogenesis. Gene expression signatures have distinguished lung tumors from normal lung. Expression profiling has divided lung cancers into sub-types [Garber et. al., 2001]. In an analysis of 32 NSCLC specimens and 7 normal specimens, unsupervised hierarchical analysis segregated tumors on the basis of histologic type and differentiation [Borczuk et. al., 2003]. Another study classified adenocarcinoma into subtypes, identifying

a group with significantly poor outcome [Bhattacharjee et. al., 2001]. Another study identified a 50-gene risk index could separate two groups by survival [Beer et. al., 2002]. Genes associated with poor survival included ERBB2, REG1A, VEGF and CRK. Expression signatures have also partitioned patients into prognostic groups [Bhattacharjee et. al., 2001; Garber et. al., 2001; Beer et. al., 2002]. Global expression studies have been used to predict response to treatment. A clinical study of microarray as a predictor of benefit from chemotherapy in NSCLC identified a 15-gene signature that correlated with survival [Winton et. al., 2005]. Altorki et al. [2010] examined safety and efficacy of short-term, preoperative pazopanib (an oral angiogenesis inhibitor targeting VEGFR, platelet-derived growth factor receptor, and c-kit) monotherapy in patients with NSCLC, found several target genes dysregulated validating target-specific response. Another study identified 92 gene, related to inflammation, cell adhesion, migration and metastasis, being differentially expressed between adenocarcinomas and squamous cell carcinomas providing avenues for further research for novel signatures [McDoniels-Silvers et. al., 2002].

Translation of the findings into clinical practice has, however, been problematic owing to the low concordance of the results. Comparision of various prognostic gene expression signatures that had been presented previously for the classification of NSCLC patients have shown that the prognostic gene lists presented in different reports had minimal overlap with only one common gene that had been identified in four separate studies. Recent results are, however, more promising and show that microarrays are of value in LC research.

## DNA METHYLATION AND CANCER

Genes involved in cancer pathogenesis require inactivation of both alleles. One allele is frequently inactivated by allelic loss, while the other one is inactivated by multiple mechanisms, including point mutations and homozygous deletions, or by a process known as aberrant methylation, a process that is limited to certain cytosine nucleotides. In vertebrates, methylation is limited to the dinucleotide CpG. The CpG dinucleotide, which is usually underrepresented in the genome, is clustered in the promoter regions of some genes. These promoter regions have been termed CpG islands. CpG islands are protected

from methylation in normal cells, with the exception of genes on the inactive X chromosome and imprinted genes. Hypermethylation of the CpG islands of gene promoter is one of the earliest and most frequent alterations leading to cancer [Baylin et. al., 2000, Jones et. al., 2002]. Cytosine methylation is a post-replicative epigenetic modification of DNA that plays a crucial role in physiology and carcinogenesis [Jones et. al., 2002]. The following three different alterations in DNA methylation are common in human cancer: (1) global hypomethylation, often seen within the body of genes; (2) dysregulation of DNA methyltransferase I, the enzyme involved in maintaining methylation patterns, and potentially other methyltransferases; and (3) regional hypermethylation in normally ummethylated CpG islands.

The distribution and methylation status of CpG sites are nonrandom. CpG sites occur relatively infrequently in much of the human genome except for discreet CpG-rich regions known as CpG islands. These islands are ~200-1,000 bp in length and often coincide with the 5'ends of genes. There are approximately 29,000 CpG islands in the human genome, although estimates vary widely, depending on the stringency of the definition [Antequera et. al., 1993]. Approximately 80% of all CpG sites are methylated and located primarily in repetitive sequences and the centromeric repeat regions of chromosomes [Herman et. al., 2003]. The remaining 20% is unmethylated and preferentially found in short sequence stretches which range from 0.5 to 5 kb that occur at average intervals of 100 kb [Colot et. al., 1999]. These stretches, or CpG islands, are often methylation-free in somatic tissues and, to a large extent, have been maintained through evolution. Current estimates indicate that 50% to 60% of human genes are associated with a CpG island [Larsen et. al., 1992, Takai et. al., 2002].

The DNA methylation field has advanced significantly over the past 2 decades and it is now well-accepted that the establishment and maintenance of DNA methylation patterns is essential for normal development [Benvenisty et. al., 1985, Sanford et. al., 1987, Monk et. al., 1987], initiation and preservation of genomic imprinting [Reik et. al., 1987, Chaillet et. al., 1991], X-chromosome inactivation [Mohandas et. al., 1981, Ariel et. al., 1995] overall genomic stability [Chen et. al., 1998, Tuck-Muller et. al., 2000, Sciandrello et. al., 2004] and regulation of tissue-specific gene expression [Gyory et al.,

2005, Kitamura et. al., 2006]. Furthermore, it is now recognized that DNA methylation is commonly altered in neoplastic transformations [Feinberg et. al., 1983, Baylin et. al., 1998, Jones et. al., 2002].

Hypermethylation of the CpG island of tumor-related genes can result in transcriptional silencing of the gene with subsequent loss of protein expression (Figure 2.12). Many cellular pathways are inactivated by this epigenetic event, including DNA repair, cell cycle, apoptosis, cell adherence, and detoxification [Esteller et. al., 2002]. Aberrant promoter methylation has been reported for several genes in a number of malignancies, and the variety of genes involved suggests that specific tumors may have their own distinct pattern of methylation [Esteller et. al., 2001].

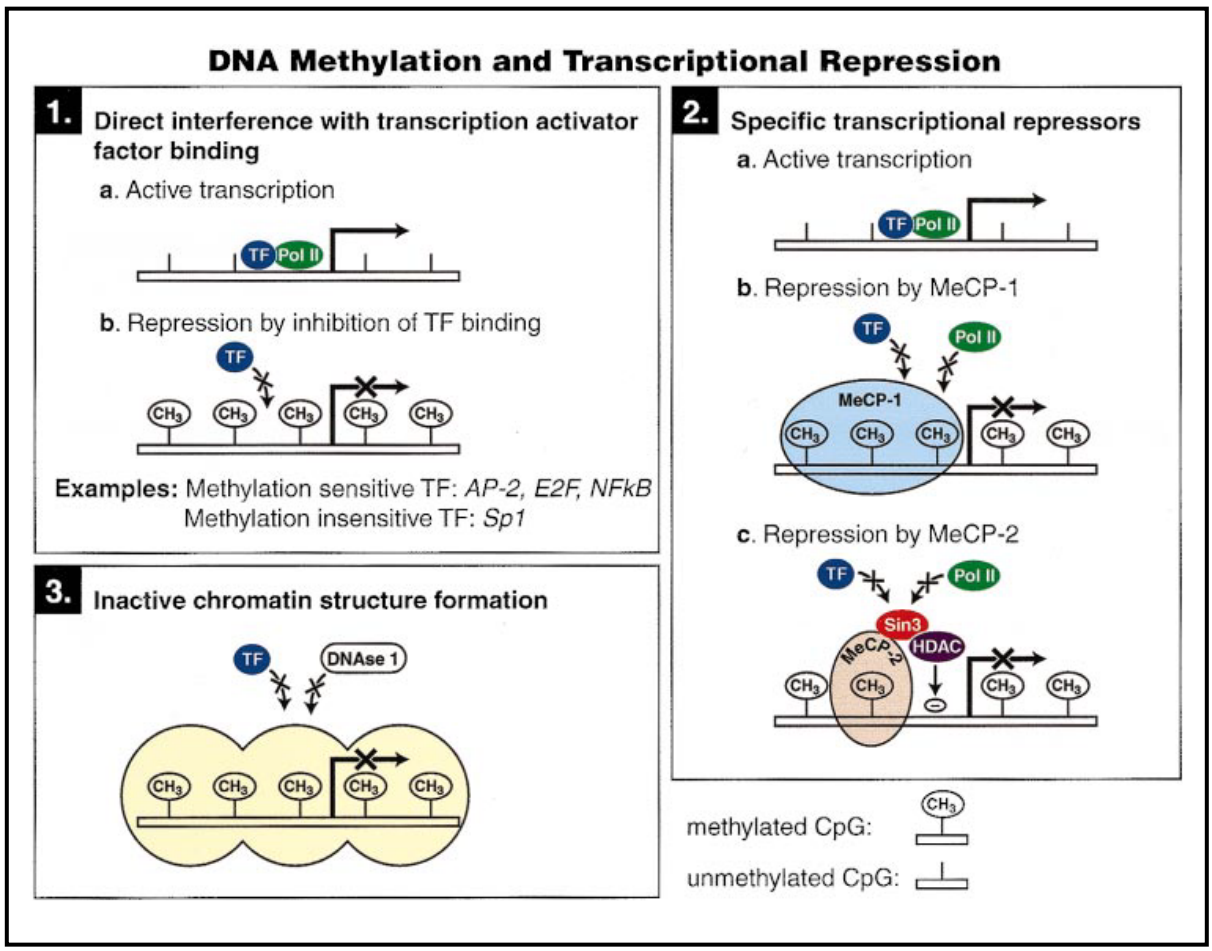



**Figure 2.12. Proposed mechanisms of transcriptional repression mediated by cytosine methylation. [Adapted from Singal et. al., 1999)**

## Detection of DNA methylation

Various methods have been developed to analyze DNA methylation. Amongst these methods, methylation specific PCR (MSP), or quantitative, such as quantitative MSP and pyrosequencing are relatively inexpensive and highly sensitive. These are widely used in retrospective studies and have potential in a diagnostic settings. Recently, genome-wide technologies such as expression and DNA microarrays have been adapted to analyze patterns of DNA methylation and screen for novel disease markers. Basically, these techniques use restriction enzyme- and sodium bisulfite based approaches which directly detect methylation at the level of a single gene or the whole genome.

In 1996, Herman and colleagues [Herman et. al.] introduced methylation specific PCR (MSP). MSP is based on the use of two distinct methylation specific primer sets for the sequence of interest (Figure 2.13). The unmethylated (U) primer will only amplify sodium bisulfite converted DNA in unmethylated condition, while the methylated (M) primer is specific for sodium bisulfite converted methylated DNA. MSP provides a positive, sensitive (detection of 1 methylated allele in a background of 1000 unmethylated alleles), quick and cost-effective test to analyze the methylation status of CpG dinucleotides in a CpG-island. Up to 1 μg single stranded DNA is treated with sodium bisulfite [pH 5.0, final concentration 2.5–3 M) for 16 hours at 50◦C for optimal cytosine to uracil conversion. After sodium bisulfite conversion, DNA is single stranded. The primer should be at least 23-24 bp in length to achieve gene specific primer annealing and should contain one to three CpG dinucleotides in the 3’ region for optimal discrimination between methylated and unmethylated DNA and increased specificity. Specific annealing temperature and cycles of amplification are crucial for proper and specific amplification of products. To interpret the PCR results properly, it is important to include positive, negative and H2O controls. Positive controls are specific for unmethylated and methylated sequences and give an impression of the specificity of the respective reactions. After amplification, the products of the reactions can be visualized by 6–8% non-denaturing polyacrylamide gel electrophoresis or high percentage horizontal agarose gels.

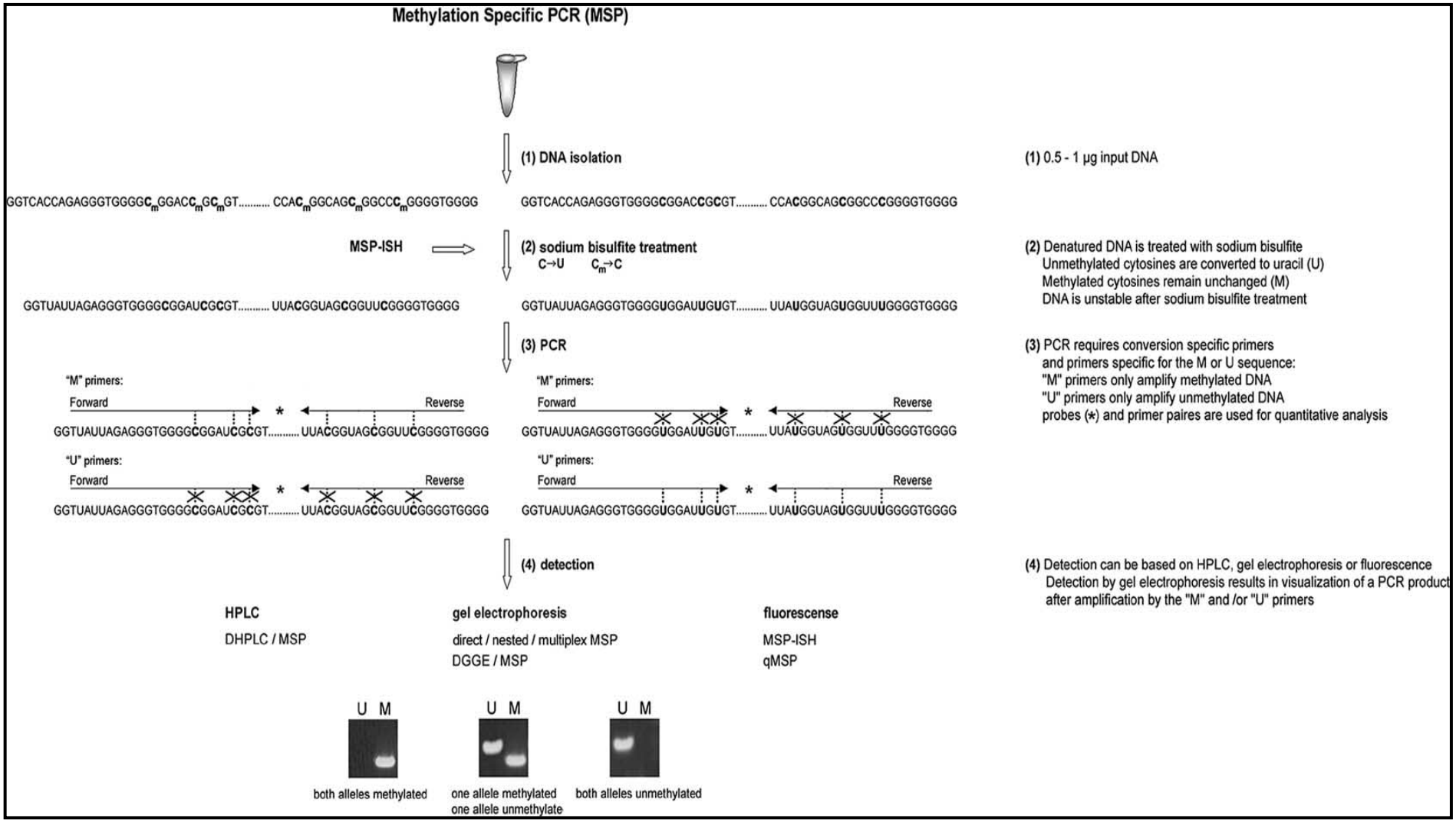



**Figure 2.13: Methylation-specific PCR [Adapted from Derks et. al., 2004)**

**DNA methylation in lung cancer**

There is ample evidence that DNA methylation patterns are profoundly altered in lung cancer. Shiraishi et. al., [1989] reported high levels of DNA methylation in chromosomes 3p and 13q. It is now evident with several findings that that DNA methylation could provide 1 or 2 of the hits necessarily to inactivate tumor suppressor genes, as postulated by Knudson's 2 hit hypothesis for oncogenic transformation [Knudson et. al., 1971]. In the early 1990s, Vertino et. al., [1993] demonstrated for the first time that de novo methylation of CpG islands and demethylation of non-CpG island sequences occur at different stages of immortalization and oncogenic transformation of bronchial epithelial cells. Since then, many genes, primarily tumor suppressor genes such as RASSF1A [Agathanggelou et. al., 2001], p16 [Grote et. al., 2005], MLH1 [Hsu et. al., 2005], MGMT [Zochbauer-Muller et. al., 2001] and DAPK [Toyooka et. al., 2003] among others, have been described as aberrantly methylated and silenced in human lung cancer [Otterson et. al., 1995, Zochbauer-Muller et. al., 2001, Tsou et. al., 2005, Bowman et. al., 2006]

**p16**

p16 is the most commonly altered gene in human malignancies [Hirama et. al., 1995]. The p16INK4a is a tumor suppressor gene on chromosome band 9p21 that encodes the p16 protein and is an inhibitor of cyclin-dependent kinase 4 (Cdk4) and functions as a cell cycle regulator. It phosphorylates the serine/threonine residues of the retinoblastoma protein, for cell cycle progression from G1 checkpoint into the S phase [Serrano et. al., 1993, Kamb et. al., 1994]. It is frequently inactivated in different types of malignancies, including lung cancer predominantly through homozygous deletion [Kamb et. al., 1994] or in association with aberrant promoter region hypermethylation [Merlo et. al., 1995]. Promoter hypermethylation of the gene has also been observed in cancer free individuals exposed to tobacco carcinogens [Toyooka et. al., 2001, Soria et. al., 2002]. Methylation of the p16 gene is reported to be associated with loss of gene transcription [Merlo et. al., 1995, Swafford et. al., 1997]. In NSCLCs, inactivation of the p16 gene has been detected in more than 70% of cell lines [Kamb et. al., 1994] and 50% of primary NSCLCs

[Cespedes et. al., 1999]. Methylation [15–35%) at the 5' CpG islands of the p16 gene has been identified as an alternative to mutation or deletion as a mechanism of p16 inactivation in NSCLCs [Merlo et. al., 1995, Kashiwabara et. al., 1998]. Studies have suggested that tobacco smoke can affect the methylation of p16 in NSCLC.

**RASSF1A**

Allelic loss of human chromosome 3p is an early and frequent event in the development of several cancers, including lung cancer [Wistuba et. al., 2000, Kok et. al., 1997]. Nine genes are located in or on the border of the breast cancer-defined sub region. One of these genes, which spans 7.6 kb of genomic DNA, has a predicted Ras-association domain and homology to the Ras-effector Nore1, it has, therefore, been termed "RASSF1" [Vavvas et. al., 1998]. RASSF1A is a potential tumor suppressor that interacts with Cdc20, activator of the anaphase-promoting complex, to inhibit complex activity and prevent mitotic progression. The protein is also shown to inhibit the accumulation of cyclin D1 and thus induce cell cycle arrest. The most common inactivation mechanism of RASSF1A is promoter hypermethylation [Dammann et. al., 2000, Burbee et. al., 2001]. RASSF1A has been studied in many tumors in which methylation correlates with reduced expression [Shivakumar et. al., 2002]. Recent studies on resected tumors, cell lines, sputum and bronchial aspirates have reported hypermethylation of the RASSF1A promoter in up to 60% of NSCLC [Dammann et. al., 2000, Burbee et. al., 2001, Grote et. al., 2006]. Furthermore, RASSF1A promoter hypermethylation was reported as a prognostic indicator in NSCLC [Wang et. al., 2011]. Another study on 107 resected NSCLC reports that RASSF1A methylation was associated with impaired patient survival [Burbee et. al., 2001]. These findings give rise to the hypothesis that RASSF1A hypermethylation may be a promising molecular biomarker for lung cancer diagnosis.

**DAPK**

The Death-associated protein kinase [DAPK) gene is located on chromosome 9p34.1 and encodes an actin associated Ca+/calmodulin-regulated serine/threonine kinase involved in apoptosis [Shohat et. al., 2002]. DAPK is involved in tumor necrosis factor-a and Fas-induced apoptosis, and has been demonstrated to be an essential mediator in IFN-g–

induced programmed cell death [Cohen et. al., 1999]. DAPK suppresses tumor growth and metastasis by increasing the occurrence of apoptosis in vivo [Inbal et. al., 1997]. DAPK has been shown to be methylated in certain malignant tumors including laryngeal squamous cell carcinoma, leukemia, lung carcinoma, prostate carcinoma, myeloma, and gastric carcinoma. Promoter hypermethylation leads to inactivation of DAPK and is found to be associated with aggressive and metastatic phenotype [Toyooka et. al., 2003]. Promoter methylation of the DAPK gene has been found in 20% to 40% of NSCLC [Tang et. al., 2000, Kim et. al., 2001]. Another study associated loss of DAPK expression with poor overall survival rates of NSCLC patients [Deiss et. al., 1995, Inbal et. al., 1997].

**GSTP1**

GSTP1, located at 11q13, belongs to a supergene family of enzymes, the GSTs, involved in the detoxification of electrophilic compounds, such as carcinogens and cytotoxic drugs, by glutathione conjugation [Henderson et. al., 1998]. In addition, these enzymes are believed to play a role in the protection of DNA from oxidative damage [Ryberg et. al., 1997]. Besides xenobiotic metabolization GSTP1 plays a role in regulating the Map kinase pathway via protein-protein interactions as it is an inhibitor of c-Jun NH2-terminal kinase 1, a kinase involved in stress response, apoptosis, and cellular proliferation [Adler et. al., 1999, Lee et. al., 1996]. Elevated expression of the GSTP1 gene has been reported to correlate with drug resistance in human cancers [Lee et. al., 1996], and high levels have been associated with poor prognosis in breast and colon cancer [Gilbert et. al., 1993, Mulder et. al., 1995]. Over the last few years, several studies have revealed that GSTP1 was somatically inactivated by hypermethylation of the promoter region. Inactivation of the GSTP1 expression was found to be associated with methylation of the CpG dinucleotides in the promoter region of the gene [Nakayama et. al., 2003]. Silencing of this gene by promoter hypermethylation leads to DNA damage and the initiation of cancer. Furthermore, GSTP1 inactivation may lead to an increased cell vulnerability to oxidative DNA damage and to the accumulation of DNA base adducts, which can make a

tumor cell move to acquire other relevant genetic alterations in prostatic carcinogenesis [Nelson et. al., 2001, Berhane et. al., 1994]. Hypermethylation of GSTP1 has been found in 7–9% of NSCLCs [Zochbauer-Muller et. al., 2001, Esteller et. al., 1998].

Thus, aberrant methylation of genes has formed basis in development of biomarkers for early detection of lung cancer. The use of genomic DNA in detection of methylation has advantages over mRNA, miRNA, and certain proteins. Genomic DNA is highly stable, easy to extract, and can survive harsh conditions. It has potential application as a noninvasive, rapid, and sensitive tool which can lead to the development of clinically relevant biomarker for early detection of susceptibility to cancer, prediction of a likely treatment effect, and assessment of tumor response to therapy.

# *Chapter 3*

# *Aim and Objectives*

# *Aim and Objectives*

## Aim

To investigate the role of genetic and epigenetic alterations and their interaction with environmental risk factors in susceptibility to lung cancer in a high risk population from North East India.

## Objectives

1. **To study high order gene-gene and gene-environment interaction in lung cancer with reference to polymorphisms in xenobiotic metabolizing genes using multigenic approaches**

   A case-control study was designed to identify the association of eight polymorphisms in six xenobiotic metabolizing genes (GSTM1, GSTT1, GSTP1, CYP1A1, EPHX1, and SULT1A1) and their interaction with environment in risk assessment of lung cancer. Classification and regression trees (CART) and multifactor dimensionality reduction (MDR) analysis method were used to explore high order gene-gene and gene-environment interactions.

2. **Association of copy number polymorphism of GSTM1 and GSTT1 in susceptibility to lung cancer**

   The aim of this objective is to examine the relationship between GSTM1 and GSTT1 gene and lung cancer risk by assessing potential gene dosage effects and gene-environment interactions. Quantitative real-time TaqMan PCR for GSTM1 and GSTT1 gene was used to determine the copy number of the gene.

3. **Association of p53 codon 72 polymorphism and its interaction with tobacco smoke, betel quid chewing and alcohol consumption with risk to lung cancer**

   In this objective, we investigated the influence of p53 codon 72 polymorphism and its interaction with tobacco, betel quid and alcohol use in lung cancer using case-control design. Estimates of risk to lung cancer, imparted by p53 genotypes and other covariates as tobacco smoking, chewing, betel quid chewing and alcohol was determined using univariate and multivariable conditional logistic regression models.

4. **Gene expression profile of lung cancer (Non-Small Cell Lung Cancer) using microarray technology**

   Microarray was used to determine the expression profile of lung cancer particularly of non-small cell lung cancer patients of north east India. Further, validation of microarray data was done by quantitative real-time reverse transcriptase–polymerase chain reaction (RT-PCR). The eventual goal is to identify the new markers for therapy and to customize therapy based on an individual tumor genetic composition.

5. **To study the role of promoter hypermethylation of tumor suppressor genes or tumor-related genes in lung cancer**

   The aim is to investigate the role of epigenetic alterations, particularly inactivation of tumor suppressor genes or tumor-related genes through promoter hypermethylation, in lung carcinogenesis. Therefore, we determine the frequency of promoter hypermethylation in a panel of tumor suppressor genes (p16, RASSF1A, DAPK) and xenobiotic gene (GSTP1) in lung cancer patients using methylation spedific PCR. Further, the interaction between methylation of these interrogated genes and their relation with clinicopathologic parameters and risk habits in these patients was also studied.

# *Chapter 4*

# *High order gene-gene and gene-environment interaction in lung cancer with reference to polymorphisms in xenobiotic metabolizing genes using multigenic approaches*

# *Introduction*

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer death globally [Jemal et. al., 2011]. The occurrence of lung cancer is attributed to a complex interplay of genetic factors and environmental exposures predominantly tobacco smoking. Of the many carcinogenic components of tobacco smoke, the polycyclic aromatic hydrocarbons (PAH) and nitrosamines are among the most potent. Xenobiotic metabolizing enzymes convert these carcinogens to more polar and water soluble metabolites for a facilitated removal from the body. Individuals possessing modified ability to metabolize these carcinogens are at increased risk of developing cancer. Thus genetic variants in xenobiotic metabolizing genes can influence their clearance from circulation and determine response to such carcinogens. The phase I xenobiotic metabolizing enzymes like cytochrome P-450s (CYPs), alcohol dehydrogenase (ALDH) and epoxide hydroxylase (EPHX) usually activate the procarcinogens through oxidation and dehydrogenation thereby converting them into reactive metabolites. Phase II metabolic enzymes such as glutathione S-transferases (GST), sulfotransferase (SULT) and N-acetyltransferase (NAT) generally result in inactivation or detoxification of these reactive metabolites. Equilibrium between expression and activity levels of these xenobiotic-metabolizing enzymes of both phase I and II determine the relative level of detoxification of carcinogens. However, these pathways are also known to activate toxic and carcinogenic chemicals to electrophilic forms that react irreversibly with macromolecules such as proteins and nucleic acids leading to carcinogenesis.

Single nucleotide polymorphisms (SNPs) in xenobiotic metabolizing genes have been studied extensively with risk of lung cancer. *CYP1A1* is a phase I, predominantly extrahepatic, microsomal enzyme. It contributes to aryl hydrocarbon hydroxylase activity, catalyzing the first step in the metabolism of a number of PAHs, such as the tobacco carcinogen benzo[a]pyrene, to their ultimate DNA-binding forms (Hu et. al., 1997). Two functional polymorphisms are known in the *CYP1A1* gene, a T to C transition, 1194 bp downstream of exon 7, generating a new MspI cleavage site and the closely linked exon 7, A to G transition (isoleucine-valine, ile462val) polymorphis. Both the polymorphism are

associated with increase in *CYP1A1* enzymatic activity towards benzo[a]pyrene and higher inducibility or enchanced catalytic activity of the valine-type *CYP1A1* enzyme. Functional consequences of the 3801T/C polymorphism located in the noncoding region, were previously thought to be due to linkage with another polymorphism in, for example, the coding region or the aryl hydrocarbon receptor. However, polymorphisms in noncoding sequences may influence gene function by altering the level, location, or timing of gene expression or messenger RNA stability (Tabor et. al., 2002). Microsomal epoxide hydrolase (*EPHX1*) is another phase I enzyme that catalyses hydrolysis of arene and alkene oxides to water soluble transdihydrodiols. In contrast to highly reactive epoxides, dihydrodiols are mostly inert and can be excreted after conjugating to glutathione. *EPHX1* converts the tobacco combustion product benzo(a)pyrene-derived benzo(a)pyrene 7,8-epoxide to the less toxic transdihydrodiol derivative, benzo(a)pyrene 7,8 diol. Although the enzyme activity of *EPHX1* is detoxifying with respect to the epoxide, the diol subsequently serves as the primary substrate for cytochrome P450 conversion to the highly reactive benzo(a)pyrene 7,8 dihydrodiol 9,10-epoxide (BPDE) (Cortessis et. al., 2001). BPDE forms adducts in DNA hotspots and is considered a major tobacco-derived benzo(a) pyrene carcinogen (Sims et. al., 1974). Two distinct *EPHX1* polymorphisms, one in exon 3 (T>C, Tyr113His) and other in exon 4 (A>G, His139Arg) have been shown to influence the enzyme activity. In exon 3, 113His allele shows reduced enzyme activity by at least 50% (slow allele) and exon 4, Arg139 allele has increased activity by at least 25% (fast allele).

Glutathione-S-transferases (GSTs) belong to a complex multigenic family of phase II metabolising enzymes. They have been found to be responsible for detoxification of a large number of electrophiles by conjugation reaction. *GSTM1* and *GSTP1* metabolise large hydrophobic electrophiles, such as polycyclic aromatic hydrocarbons derived epoxides (Hayes et. al., 1995), whereas *GSTT1* is involved in the metabolism of smaller compounds, such as monohalomethane and ethylene oxide (Landi 2000). GSTs also metabolise compounds formed during oxidative stress, such as hydroperoxides and oxidized lipids, and they are transcriptionally activated during oxidative stress (Hayes 2005). A homozygous gene deletion that results in total lack of the enzyme (null genotype) is a common polymorphism in both the *GSTM1* and the *GSTT1* gene. The *GSTP1* variant investigated in the current study

consists of an A-to-G substitution at base pair 313 at codon 105 resulting in an amino acid difference, from isoleucine to valine (Kellen et. al., 2007). This codon is located in the substrate-binding site of *GSTP1*, and the corresponding allozymes exhibited differential catalytic activities toward diverse substrates (Hu et. al., 1997). Studies have shown that the activity of the isoleucine 105 variant toward several carcinogenic diol epoxides is lower compared with that of the valine 105 form (Hu et. al., 1997a and b). Sulfotransferases (SULTs), a superfamily of multifunctional enzymes, catalyze sulfonate conjugation that is an important pathway in the phase II metabolism of numerous endogenous and exogenous compounds. Although sulfonation is, in general, considered a detoxification reaction, several SULTs, particularly *SULT1A1*, are involved in the bioactivation of certain procarcinogens, including heterocyclic amines and PAHs. Functional polymorphism located in the coding region 638G>A, resulting in a substitution of histidine for arginine (Arg213His), leads to decreased enzyme activity. The substitution in amino acid sequence is associated with both a decreased substrate affinity and a lower level of protein (Jones et. al., 1993).

A majority of the molecular epidemiological studies consider only the main effects of these SNPs and their observed strength of associations could be challenged by penetrance of the genetic variant. Furthermore, a single locus cannot account for genetic susceptibility in a complex disease such as cancer which involves multiple genetic variations and gene-environment interactions. Current evidences suggest that high order interactions in multigenic approach allow more precise delineation of risk groups [Ritchie et. al., 2001, Liu et. al., 2011].

In the present objective, two data mining approaches, classification and regression trees (CART) and multifactor dimensionality reduction (MDR) was applied along with logistic regression (LR) to detect high order gene-gene and gene environment interactions. Both CART and MDR assume model free and non-parametric methods of estimating non-linear interactions with low false-positives even on relatively small sample sizes. Model validation through permutation testing and false positive report probabilities were also done to overcome inaccurate estimation. Interaction entropy graphs were constructed to interpret combination effects identified by MDR. To further analyze possible effects of the EPHX1 and CYP1A1 SNPs, we estimated their haplotype frequencies and risk imparted towards lung cancer.

## EXPERIMENTAL METHODS

### Materials

Agarose, Tris base, EDTA, NaCl, SDS, Triton X-100 and other fine chemicals were purchased from Sigma Chemicals, USA. Taq polymerase, dNTPs, MgCl2, was obtained from Invitrogen and MBI fermentas USA. Oligos were synthesized by Microsynth, Switzerland. RNA later, DNA and RNA extraction kit were purchased from Qiagen Sciences, USA and Himedia, India.

### Chemicals used

LYSIS BUFFER I: 30mm Tris-Hcl (Ph-8), 5mm EDTA, 50 Mm Nacl; LYSIS BUFFER II: 75mm Nacl, 2mm EDTA (Ph-8); SDS STOCK: 20 gm of SDS dissolved in 80 ml of TDW at $65^0$C. Make up volume up to 100 ml ; PROTEINASE K: 10 mg dissolved in 1 ml of TDW:1% ; AGAROSE: 1gm of agarose dissolved 1% TAE buffer.

### Patient recruitment and sample collection

The study was conducted in 188 histopathologically diagnosed lung cancer cases registered at Dr. Bhubaneswar Borooah Cancer Institute, Guwahati, Civil Hospital, Aizawl, and Sir Thutob Namgyal Memorial Hospital, Gangtok, the collaborating centers in north east India.

### Inclusion criteria

Incident cases during the period of December 2006 to 2009 and willing to participate in the study were included. 290 voluntary, age (±5 years) and sex matched individuals were selected from the unrelated attendants who accompanied cancer patients. This provided a readily available and cooperative source of controls from the same socio-economic background as the cases reducing confounding biases. Patients with only lung as their primary site of cancer were included. Final selected controls were included on the basis of no history of any obvious disease and those not taking any medication at the time of recruitment

**Exclusion criteria**

Any subject with history of familial malignancy or pulmonary infectious disease was excluded both from case and control. Patients unwilling or too ill to participate in the study were excluded. Patients who had taken any form of treatment earlier (Secondary cases) were also excluded from the study**.**

**Patient details**

All subjects provided written informed consent for participation in this research which was done under a protocol approved by the institutional ethics committee of Regional Medical Research Centre, North East Region (Indian Council of Medical Research). Smokers, chewers and drinkers were classified into two categories ever and never. For smoking, an individual who had never smoked or smoked less than 100 cigarettes in their lifetime and were not smoking at the time of reporting was considered never smoker or non-smokers. Ever smokers or smokers category included current smokers, and those who had quit within <1 year of reporting [WHO 1998]. As our collaborating centers were public hospitals a large majority of subjects belonged to lower to middle socio-economic background. Demographic data and characteristics such as age, sex, smoking habit, usage of tobacco, betel quid and alcohol, were obtained from subjects in a standard questionnaire used for all the centers, in an in-person interview by a trained data collector. A majority of cases and controls were literate with full primary schooling and some upto the college level. The occupational history of the study participants revealed that most of them were farm laborers or engaged in petty jobs and the nature of such jobs did not exposed them to any occupational hazards. Any history of past or present illness was enquired or if undergoing any medication at the time of enrolment.

**Collection of blood samples**

Peripheral blood samples (4-5 ml) were obtained from all patients and controls in EDTA coated vials and stored in $-20^0$C until transported to the laboratory where the study was performed.

**Extraction of Genomic DNA**

Genomic DNA was isolated using Qiagen Blood DNA Isolation kit (Qiagen GmbH, Germany) and stored at -30$^0$C till further analysis.

**Quantification of Genomic DNA**

For the quantification of DNA, readings were taken in ND-1000 spectrophotometer (Nanodrop Technologies Inc USA). Precisely 1.5µl of the sample was loaded on the pedestal of the instrument. Readings were taken in specific module for DNA after taking measurement for blank. The ratio of absorbance at 260 and 280 nm is used to assess the purity of DNA. A ratio of ~1.8 is accepted as "pure" for DNA.

**Agarose Gel Electrophoresis of Extracted DNA Samples**

In order to check the quality of the extracted DNA from blood samples, agarose gel electrophoresis was carried out in a 0.8% agarose gel in TAE buffer. 0.8 gram of agarose was dissolved in 100 ml of 1x TAE buffer and boiled. The solution was cooled to 45-50$^0$ C and 5-6 ul of EtBr was added in to the solution. Then solution was poured in to the casting tray with a comb. After solidification, gel was placed in electrophoresis tank containing 1% TAE buffer. The DNA samples (5 µl~250-500ng) were mixed with 6X loading dye (1 µl) and loaded into the slot/well of submerged gel. Applying a constant current of 100 mA the gel was run for up to 30 minutes. Gels were visualized under the gel documentation system and images acquired (Figure 4.1).

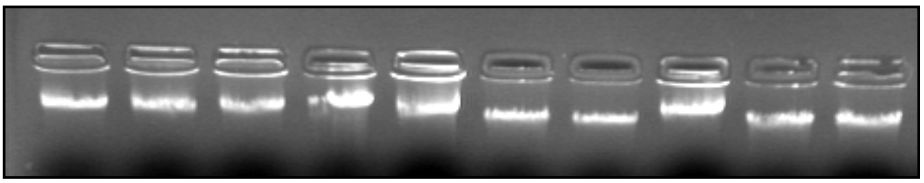


**Figure 4.1: Agarose gel picture showing quality of genomic DNA isolated from the subjects**

**Genotyping protocol:**

**Genotyping of GSTM1 and GSTT1 by Multiplex PCR**

A multiplex PCR method was used to detect the presence or absence of the GSTM1 and GSTT1 genes in the genomic DNA samples (Table 4.1). This method had both GST primers sets in the same PCR reaction and included a third primer set for β-globin as internal control to ensure proper functioning of PCR. The PCR was carried out for an initial activation step at $94^0$C for 4 min, 20 cycles of denaturation at $93^0$C for 1 min; annealing at $60^0$C for 1 min; $72^0$C for 1 min and in addition with these there were 15 cycles of denaturation at $93^0$C for 1 min; annealing at $50^0$C for 1 min; $72^0$C for 1 min and a final extension at $72^0$C for 10 min. The PCR products were electrophoresed in 2.5% agarose gels containing ethidium bromide, prepared and run in 0.5X TBE buffer. The absence of 459 bp band indicates GSTT1 null and the absence of 219bp indicates GSTM1 null genotypes All 188 cases and 290 controls were subjected to agarose gel electrophoresis after multiplex PCR. Representative figures are shown from Figure 4.2.

**Table 4.1: Multiplex PCR for genotyping of GSTM1 and GSTT1 polymorphism**

| Components | Master Stock | Working Stock | Reaction l (25µl) |
|---|---|---|---|
| Nuclease-free water ( 25 µl) | X | | 15.3 |
| PCR buffer | 10X | 1X | 2.5 |
| $MgCl_2$ | 25 mM | 1.0 mM | 1.0 |
| dNTP mix (2.5 mM each) | 25 mM | 0.2 mM | 0.2 |
| GSTM1 (Forward primer) | 10 µM | 0.25 µM | 0.625 |
| GSTM1 (Reverse primer) | 10 µM | 0.25 µm | 0.625 |
| GSTT1 (Forward primer) | 10 µM | 0.25 µM | 0.625 |
| GSTT1 (Reverse primer) | 10 µM | 0.25 µm | 0.625 |
| β-Globin (Forward primer) | 10 µM | 0.25 µM | 0.625 |
| β-Globin (Reverse primer) | 10 µM | 0.25 µm | 0.625 |
| Taq DNA polymerase | 5 U/µl | 1.25U | 0.25 |
| Template (DNA) | Y | 100 to 300 ng | 2.0 |

**Genotyping of GSTP1, EPHX1, CYP1A1 and SULT1A1**

Polymorphism in GSTP1, EPHX1 Exon3 and Exon4, CYP1A12A, CYP1A12C and SULT1A1 were genotyped using PCR-RFLP (polymerase chain reaction-restriction fragment length polymorphism) method. Standard PCR were performed on PTC-200 (MJ Research, USA). The PCR reaction were performed in a volume of 25µl with a final concentration of 1X PCR Buffer (MBI Fermentas), 1.5mM $Mgcl_2$, 200 µM dNTPs, 2.0 mM , 1 Unit of Taq DNA polymerase and 100-300 ng of DNA Template. Negative controls were included in all PCR-runs to prevent misjudging following contamination of samples. PCR amplification consist of 35 cycles of denaturation at $94^0$C for 45s; annealing at depend on gene for 45s; $72^0$C for 45s followed by a final extension at $72^0$C for 10 min. PCR products were loaded on 2.5% agarose gel and subjected to gel electrophoresis in 0.5X TBE buffer, stained with ethidium bromide and visualized under UV. Detail of single-nucleotide polymorphisms (SNPs) selected for the study is summarized in Table 4.2. Sequence of the primer and their annealing temperatures are given in the Table 4.3.

**Table 4.2: Detail of the single-nucleotide polymorphisms (SNPs) selected for the study**

| Gene | Chr[a] | SNP | Loc | Polymorphism | |
|---|---|---|---|---|---|
| | | | | Nucleotide | Codon |
| GSTT1 | 22q11.23 | GSTT1 | Gene | Presence>Null | Deletion |
| GSTM1 | 1p13.3 | GSTM1 | Gene | Presence>Null | Deletion |
| GSTP1 | 11q13 | rs1695 | Exon 5 | 313A>G | Ile105Val |
| EPHX1 | 1q42.1 | rs2234922 | Exon 4 | 418A>G | His139Arg |
| EPHX1 | 1q42.1 | rs1051740 | Exon 3 | 339T>C | Tyr113His |
| CYP1A1 | 15q24.1 | rs4646903 | 3'UTR | 6235T>C | NA[b] |
| CYP1A1 | 15q24.1 | rs1048943 | Exon 7 | 2454A>G | Ile462Val |
| SULT1A1 | 16p12.1 | rs9282861 | Exon 7 | 638G>A | Arg213His |

[a]Chromosomal position is based on NCBI Build.

[b]Not Applicable

**Table 4.3: Sequence of primers used in the study**

| Gene | Primer sequence | $T^0C$ | PCR (bp) |
|---|---|---|---|
| GSTT1 | 5'-TTCCTTACTGGTCCTCACATCTC-3'<br>5'-TCACCGGATCATGGCCAGCA-3' | - | 459 |
| GSTM1 | 5'-GAACTCCCTGAAAAGCTAAAGC-3'<br>5'- GTTGGGCTCAAATATACGGTGG-3' | - | 219 |
| GSTP1 | 5'-CCAGTGACTGTGTGTTGATC-3'<br>5'-CAACCCTGGTGCAGATGCTC-3' | 62 | 189 |
| EPHX1 Exon 4 | 5'-ACATCCACTTCATCCACGT-3'<br>5'-ATGCCTCTGAGAAGCCAT-3' | 56 | 210 |
| EPHX1 Exon 3 | 5'-GATCGATAAGTTCCGTTTCACC-3'<br>5'-ATCCTTAGTCTTGAAGTGAGGAT-3' | 52 | 162 |
| CYP1A12A | 5'-TAGGAGTCTTGTCTCATGCCT-3'<br>5'-CAGTGAAGAGGTGTAGCCGCT-3' | 61 | 340 |
| CYP1A12C | 5'- GAAAGGCTGGGTCCACCCTCT -3'<br>5'-CCAGGAAGAGAAAGACCTCCCAGCGGGCCA-3' | 63 | 333 |
| SULT1A1 | 5'-AGTTGGCTCTGCAGGGTTTCT-3'<br>5'-ACCACGAAGTCCACGGTCTC-3' | 59 | 200 |

**RFLP analysis of GSTP1, EPHX1 Exon3, EPHX1 Exon4, CYP1A12A, CYP1A12C and SULT1A1 polymorphism**

Restriction digestion of the amplified fragments was carried out for the above polymorphism in a water bath (Table 4.4). Heat inactivation of enzyme was done at $80^0$C for 20 minutes after completion of incubation with enzyme. Restriction enzymes that cleave the DNA specifically for different alleles were used (Table 4.5) and the alleles of each specific sample can be observed as a specific band pattern on the gel. All 188 cases and 290 controls were subjected to agarose gel electrophoresis after PCR and RFLP. Representative figures are shown from Figure 4.3 to Figure 4.8. The genotyping results were confirmed by repeated analysis of approximately 10% of all samples randomly chosen

**Table 4.4: Standard protocol used for the RFLP experiment**

| Components | Stock conc. | Working conc. | 1 reaction (µl) |
|---|---|---|---|
| Water | | | 3 |
| Buffer | 10X | 1 | 1.5 |
| Enzyme* | 10Units/µl | 5Units | 0.5 |
| PCR product | | | 10.0 |
| * Enzymes are specific for each polymorphism given in the table 4.5 | | | |

**Table 4.5: Detail of the RFLP enzymes used for each polymorphism**

| Gene | Enzyme | Site | Incubation Condition | PCR | RFLP product (bp) | |
|---|---|---|---|---|---|---|
| | | | | | Homo wild | Homo variant |
| GSTP1 | BsmA1 | 5'-GTCTC^-3'<br>3'-CAGAG^-5' | $55^0$C for 8 hrs | 189 | 189 | 148+41 |
| P53 | BstUI | 5'-CG^CG-3'<br>3'-GC^GC-5' | $37^0$C overnight | 199 | 113+86 | 199 |
| EPHX1 ex4 | RsaI | 5'-GT^AC-3'<br>3'CA^TG-5' | $37^0$C for 4hrs | 210 | 210 | 164 + 46 |
| EPHX1 ex3 | EcoRV | 5'-GAT^ATC-3'<br>3'CTA^TAG-5' | $37^0$C overnight | 162 | 140+22 | 162 |
| CYP1A12A | MspI | 5'-CC^GG-3'<br>3'GG^CC-5' | $37^0$C overnight | 340 | 340 | 220+140 |
| CYP1A12C | NcoI | 5'-C^CATGG-3'<br>3'GGTAC^C-5' | $37^0$C overnight | 333 | 69+32+232 | 69 + 264 |
| SULT1A1 | HhaI | 5'-GCG^C-3'<br>3'C^GCG-5' | $37^0$C for 3hrs | 200 | 160+40 | 200 |

**Statistical Analysis:**

Cases were individually matched with controls on the basis of age (±5 years), sex and ethnicity, in a ratio of approximately 1:1.5. Difference in the distribution of demographic characteristics and genotype frequencies between cases and controls were evaluated using the Chi Square ($\chi^2$) and Fisher's Exact test wherever appropriate. Hardy–Weinberg equilibrium (HWE) was assessed by using the $\chi^2$-test. Estimates of risk to cancer, imparted by genotypes and other covariates as tobacco smoking, tobacco chewing, betel quid chewing and alcohol consumption were determined by deriving the odds ratio (OR) and its corresponding 95% confidence interval (95% CIs) using multivariable conditional logistic regression. For all the tests a two sided $p<0.05$ was considered statistically significant. The data analysis was performed on the Intercooled Stata 8.0 statistical software package (Stata Co., College Station, TX).

**Haplotype Analysis**

Haplotypes were constructed from the unphased diploid genotype data using the Expectation Maximization-based algorithm. Individual haplotypes and their estimated population frequencies were inferred and estimates of linkage disequilibrium (D') between SNPs were calculated using Haploview software ver.4.1.

**Identification of high order Interactions**

High order interactions were determined using CART, MDR and interaction entropy graphs.

**CART**

A binary recursive partitioning method was used to produce a decision tree that identified specific combinations of contributing factors associated with lung cancer risk using the commercially available CART software (version 6.6, Salford Systems) [Steinberg et. al., 1997]. CART is a binary recursive partitioning method that creates a decision tree which describes how well each genotype or environmental factor variable predicts class (eg. Lung cancer case-control status). The CART model selects the variable used to split each branch and the split point. Splitting rules are used to stratify data into subsets of individuals, which are represented in the CART decision tree as nodes. The tree building process continues until

the terminal nodes have no subsequent statistically significant splits or they reach a pre-specified minimum size. Optimal tree was determined by reducing overfitting trees using one standard error (1-SE) rule and repeating 10 fold cross validation. In this study, tree splitting was done till terminal nodes reached a pre- specified minimum size of 10 subjects. Optimal tree was selected using one standard error (1-SE) rule and 10 fold cross validation. Subgroups of individuals with differential risk patterns were identified in the different order of nodes, indicating the presence of gene-gene and gene-environment interactions. Fischer's Exact test was used to calculate relative risk in each terminal node of the tree.

**MDR**

The MDR software was developed by Ritchie et. al., 2001 and reviewed by Hahn et. al, 2003. Genotype and environmental factors were pooled into high and low risk group, effectively reducing the multifactor prediction from n dimension to one dimension using MDR software (version 2.0 beta) (http://www.epistasis.org). This new one-dimensional variable can be evaluated for its ability to classify and predict disease status using cross-validation and permutation testing that minimizes false positive results by multiple examinations of the data. We applied Tuned ReliefF (TuRF) filter algorithm to remove noisy SNPs and avoid overfitting of data. Best models for each locus were selected by repeating the analysis for up to 10 seeds and applying 10 folds cross validation each time. Statistical significance of the best models selected for each locus was determined using 1000 fold permutation testing. The fitness of an MDR model was assessed by estimating the maximum values of cross validation consistency (CVC) and testing accuracy (TA). p-values hence obtained for TBA and cross validation consistency (CVC), were considered statistically significant at 0.05 levels.

**False Positive Report Probability (FPRP)**

Reports of gene-environment interaction studies are often challenged by false positive discoveries especially when results are generated by multiple comparisons. To estimate the FPRP and to evaluate robustness of the findings from MDR analysis we used the Bayesian approach described by Wacholder et. al., 2004. The method requires prior probabilities that the genetic variant and disease association is real. As prior probability can be a subjective measure and can be influenced by several factors, usually a wide range is reported by studies.

Considering poor epidemiological data from the study population and inconsistent association of the SNPs with lung cancer risk we set a fairly wider range of prior probabilities ($10^{-6}$ to $10^{-1}$) with an estimated statistical power to detect an OR of 1.5 and 2.0 and α level equal to the observed p-value. The FPRP cutoff point was stringently kept to 0.2.

**Interaction entropy graphs**

Interaction graphs were built to visualize and interpret the results obtained from MDR using Orange machine learning software package [Demsar and Jupan 2004]. Interaction graphs use entropy estimates as described by Jakulin et. al., 2003 for determining the gain in information about a class variable (e.g. case–control status) from merging two variables together over that provided by the variables independently**.** This measure of entropy is useful for building interaction graphs that facilitate the interpretation of the relationship between variables. Interaction graphs are comprised of a node for each variable with pairwise connections between them. The percentage of entropy removed (i.e. information gain) by each variable is visualized for each node. The percentage of entropy removed for each pairwise Cartesian product of variables was visualized for each connection. Thus, the independent main effects of each SNP can be compared to the interaction effect. Positive entropy (plotted in green) indicates non-linear interaction while negative entropy (plotted in red) indicates redundancy. Entropy value equal to zero indicates independence or a mixture of synergy and redundancy.

**RESULTS**

**Characteristics of study subjects**

The distribution of demographic characteristics and main effects of genetic and environmental factors is summarized in Table 4.6. The frequency distribution of males and females were 77.1% and 22.9% in cases and 76.2% and 23.85 in controls respectively. Mean age of cases and controls was 60.41±10.58 (range 30-82yrs) and 57.19±10.75 (range 32-85yrs) respectively. The distribution of all SNPs in control was in agreement with HWE ($p>0.05$), however alleles of EPHX1 Tyr113His and SULT1A1 Arg213His polymorphisms in cases did not follow HWE ($p<0.001$ and $p=0.004$ respectively). Risk habits such as smoking, tobacco chewing and betel quid chewing were predominant among cases. However only smoking and

betel quid chewing were significantly associated with increased risk for lung cancer (OR=3.06;95%CI=1.94-4.83;p<0.001 and OR=1.86; 95%CI=1.21-2.84;p=0.004 respectively).

**Table 4.6: Demographic data of lung cancer cases and controls**

| Variables | Categories | Cases n (%) | Controls n (%) | OR (95% C.I.) | p value |
|---|---|---|---|---|---|
| **Sex** | Male | 145 (77.1) | 221 (76.2) | | |
| | Female | 43 (22.9) | 69 (23.85) | | |
| **Age Mean (Range)** | | 60.41±10.58 (30-82) | 57.19±10.75 (32-85) | | |
| **Smoking status**[a] | Non smokers | 56 (29.8) | 151 (52.1) | 1.0 | |
| | Smokers | 132 (70.2) | 139 (47.9) | **3.06 (1.94-4.83)** | **<0.001** |
| **Tobacco chewing**[a] | Non chewers | 92 (48.9) | 172 (59.3) | 1.0 | |
| | Chewers | 96 (51.1) | 118 (40.7) | 1.24 (0.82-1.85) | 0.293 |
| **Betel quid chewing**[a] | Non chewers | 52 (27.7) | 131 (45.2) | 1.0 | |
| | Chewers | 136 (72.3) | 159 (54.8) | **1.86 (1.21-2.84)** | **0.004** |
| **Alcohol consumption** | Non alcoholic | 135 (71.8) | 207 (71.4) | 1.0 | |
| | Alcoholic | 53 (28.2) | 83 (28.6) | 0.87 (0.56- 1.37) | 0.57 |

[a] $\chi^2$ significant; p<0.05

Bold number indicate significant p value < 0.05

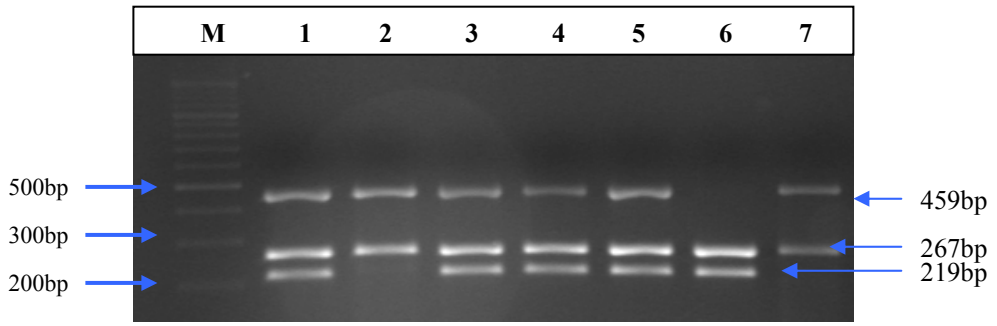



**Figure 4.2: EtBr stained agarose gel electrophoresis showing multiplex PCR for GSTM1 and GSTT1 polymorphisms**. Lane 1,2,3,4,5 and 7-samples with 459 bp represent wild type GSTT1; Lane 1,3,4,5 and 6 samples with 219bp represent GSTM1 gene; The presence of 267bp in all lanes represent β-globin gene used as internal control. M-100bp ladder

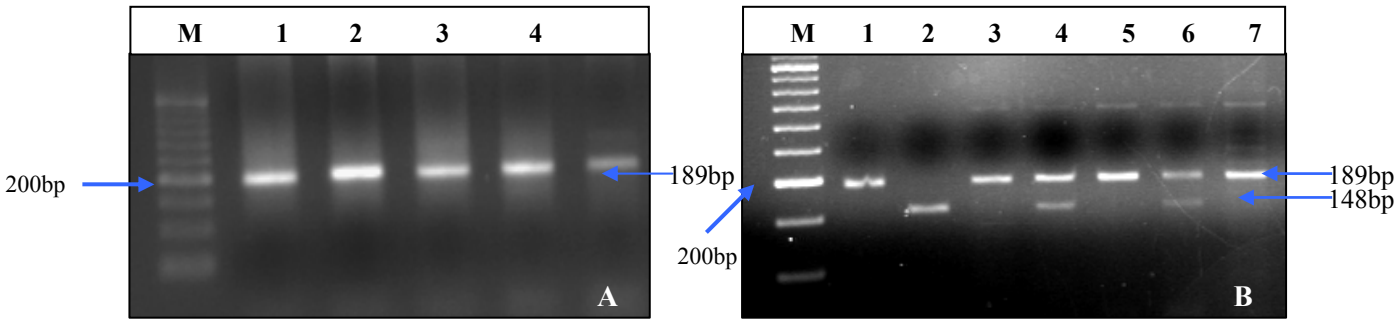



**Figure 4.3**: **EtBr stained agarose gel electrophoresis showing GSTP1 polymorphism**. A- showing PCR amplification of GSTP1 gene (189bp). B-RFLP of GSTP1 PCR product; Lane 1,3,5 and 7 samples with 189 bp represent wild type Ile/Ile allele (AA genotype); Lane 4 and 6-sample with 189 bp and 148 bp represent heterozygous Ile/Val allele (AG genotype); Lane 2 - sample with 148bp and 41 bp not visible represent homozygous Val/Val allele (GG genotype). M-50bp ladder

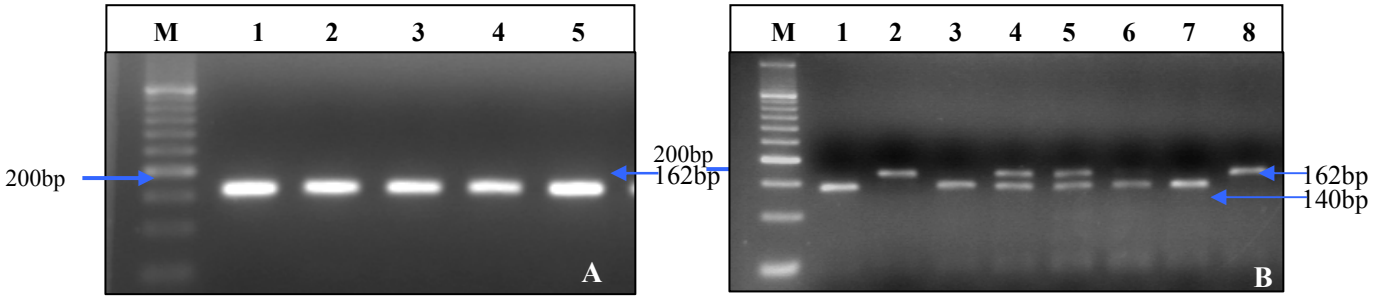



**Figure 4.4: EtBr stained agarose gel electrophoresis for RFLP analysis of EPHX1 exon3 polymorphism:** A- showing PCR amplification of exon3 of EPHX1 gene. B-RFLP of EPHX1 exon 3 PCR product; Lane 1,3,6,7-samples with two band (140bp and 22bp not visible) represent homozygous Tyr/Tyr allele (TT genotype); Lane 4,5-samples with all three bands (162bp, 140bp and 22bp not visible) represent heterozygous Tyr/His allele (TC genotype); Lane 2,8- samples with only one band (162bp) represent homozygous His/His allele (CC genotype). M-50bp ladder.

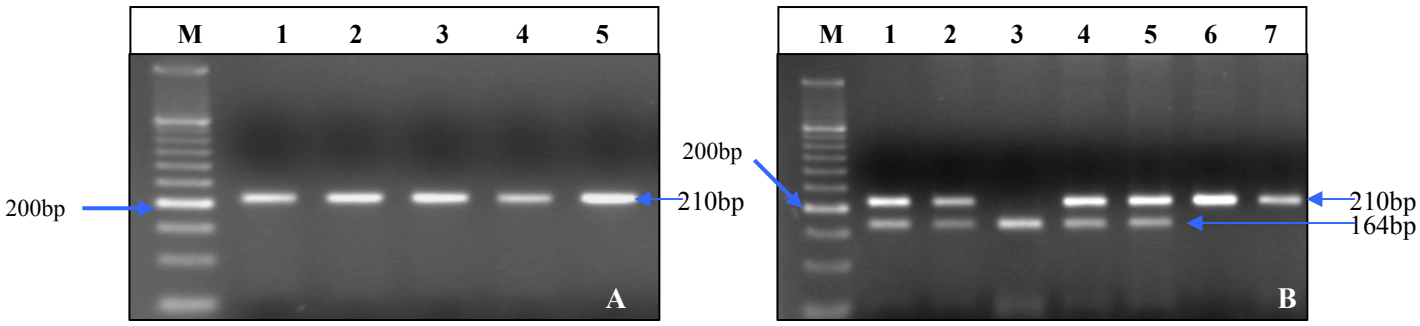



**Figure 4.5: EtBr stained agarose gel electrophoresis for RFLP analysis of EPHX1 exon 4 polymorphism.** A-showing PCR amplification of exon4 of EPHX1 gene. B-RFLP of EPHX1 exon 4 PCR product. Lanes 6 and 7 show the undigested product (210bp) representing His/His allele. Lane 3 show the digested products (164bp + 46bp, not visible) representing the homozygous Arg/Arg allele. Lanes 1, 2, 4 and 5 show the presence of two bands representing the heterozygous condition of His/Arg allele. M-50bp ladder.

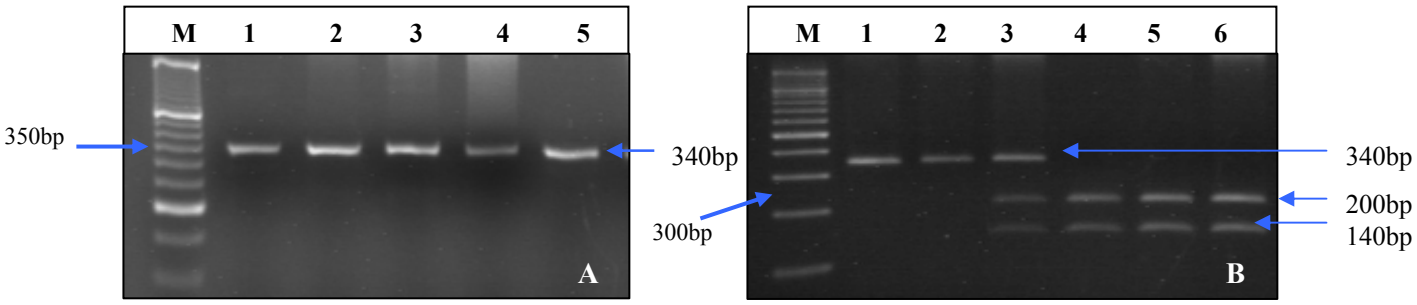



**Figure 4.6: EtBr stained agarose gel electrophoresis for RFLP analysis of CYP1A12A polymorphism: A**- showing PCR amplification of CYP1A12A. **B**-RFLP of CYP1A12A PCR; Lane 1,2-samples with single band (340bp) represent wild type TT genotype; Lane 3-samples with all three band (340bp, 200bp ,140bp) represent heterozygous TC genotype; Lane 4,5,6- samples with two band (200bp and 140bp) represent homozygous CC genotype.

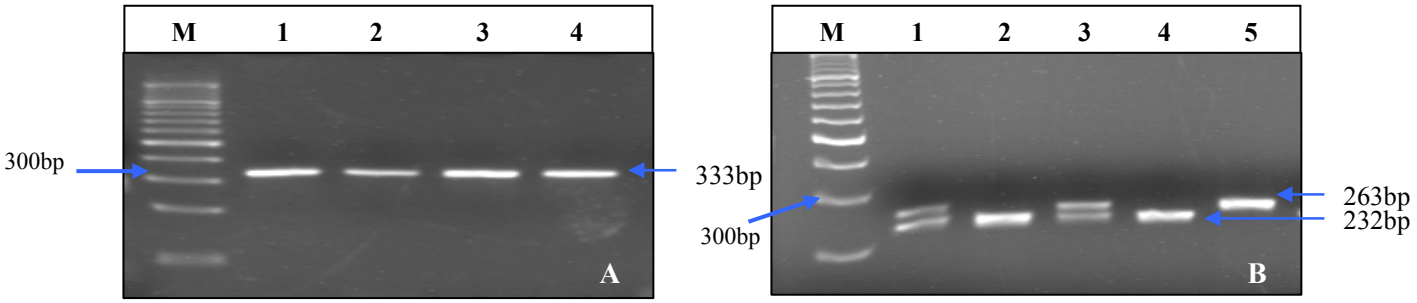



**Figure 4.7: EtBr stained agarose gel electrophoresis for RFLP analysis of CYP1A12C polymorphism. A**- showing PCR amplification of CYP1A12C. **B**-RFLP of CYP1A12C PCR; Lane 2,4-samples with single band (232bp) represent wild type Ile/Ile allele; Lane 5-sample with 263bp represent homozygous variant Val/Val allele. Lane 1,3- samples represent heterozygous Ile/Val allele , M-100bp ladder.

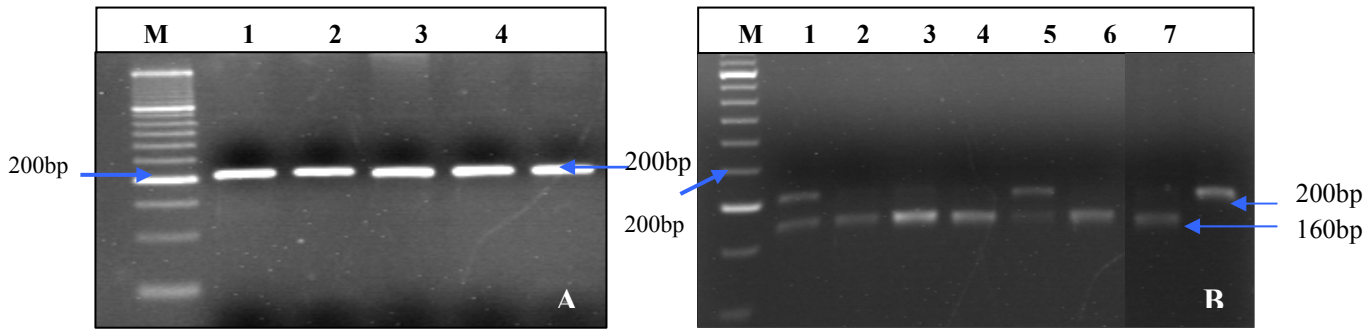



**Figure 4.8: EtBr stained agarose gel electrophoresis showing SULT1A1 polymorphism.** A-showing PCR amplification of SULT1A1 gene. B-RFLP of SULT1A1 PCR; Lane 1 and,5-samples with all three bands (200bp, 160bp, 40bp not visible) represent heterozygous Arg/His allele (GA genotype); Lane 2,3,4,6,7-sample with two band (160bp and 40bp not visible) represent the homozygous Arg/Arg allele.
Lane 8-sample with undigested product (200bp) represents His/His allele. M-50bp ladder.

**Association of genetic factors with lung cancer risk by LR analysis**

The distribution and main effects of genetic factors is summarized in Table 4.7. Genotype distribution of CYP1A1*2A, EPHX1 Tyr113His, SULT1A1 Arg213His and GSTT1 null polymorphism were significantly different in cases from controls (p=0.014, p<0.001, p=0.01 and p=0.04 respectively). Main effects of genotypes in lung cancer susceptibility were evaluated using conditional multivariable logistic regression. Heterozygous genotype in CYP1A1*2A was associated with increased risk (OR=1.69,95% CI=1.11-2.59; p=0.01) whereas heterozygous genotypes in EPHX1 Tyr113His and SULT1A1 Arg213His imparted reduced risk towards lung cancer (OR=0.40;95%C.I=0.25-0.65,p<0.001 and OR=0.51;p=0.33-0.78,p=0.002 respectively). CYP1A1*2A and EPHX1 His139Arg polymorphisms were associated with increased risk to lung cancer in dominant genetic model, whereas EPHX1 Tyr113His and SULT1A1 Arg213His imparted reduced risk in recessive genetic model (Table 4.8).

**Table 4.7: Association of genotypes of xenobiotic metabolizing genes with lung cancer susceptibility**

| Factors | Categories | Cases | | Controls | | OR (95% C.I)[#] | P value |
|---|---|---|---|---|---|---|---|
| **Genetic Factors** | | **n** | **%** | **n** | **%** | | |
| **CYP1A1*2A**[a] | TT | 55 | 29.3 | 122 | 42.1 | 1..0 | |
| | TC | 103 | 54.8 | 124 | 42.8 | **1.69 (1.11-2.59)** | **0.01** |
| | CC | 30 | 16.0 | 44 | 15.2 | 1.53 (0.84-2.78) | 0.15 |
| **CYP1A1*2C** | AA | 122 | 64.9 | 206 | 71.0 | 1..0 | |
| | AG | 56 | 29.8 | 77 | 26.6 | 1.16 (0.75-1.80) | 0.48 |
| | GG | 10 | 5.3 | 7 | 2.4 | 2.18 (0.78-6.09) | 0.13 |
| **EPHX1 Tyr113His**[a] | TT | 82 | 43.6 | 94 | 32.4 | 1.0 | |
| | TC | 51 | 27.1 | 133 | 45.9 | **0.40 (0.25-0.65)** | **<0.001** |
| | CC | 55 | 29.3 | 63 | 21.7 | 1.00 (0.60-1.67) | 0.98 |
| **EPHX1 His139Arg** | AA | 121 | 64.4 | 212 | 73.1 | 1.0 | |
| | AG | 59 | 31.4 | 70 | 24.1 | 1.45 (0.92-2.27) | 0.10 |
| | GG | 8 | 4.3 | 8 | 2.8 | 2.41 (0.79-7.36) | 0.12 |
| **GSTM1** | Wild | 122 | 64.9 | 177 | 61.0 | 1.0 | |
| | Null | 66 | 35.1 | 113 | 39.0 | 0.95 (0.63-1.41) | 0.80 |
| **GSTT1**[a] | Wild | 155 | 82.4 | 217 | 74.8 | 1.0 | |
| | Null | 33 | 17.6 | 73 | 25.2 | 0.62 (0.38-1.02) | 0.06 |
| **GSTP1** | AA | 102 | 54.3 | 179 | 61.7 | 1.0 | |
| | AG | 77 | 41.0 | 96 | 33.1 | 1.46 (0.95-2.23) | 0.07 |
| | GG | 9 | 4.8 | 15 | 5.2 | 1.09 (0.43-2.77) | 0.84 |
| **SULT1A1**[a] | GG | 123 | 65.4 | 153 | 52.8 | 1.0 | |
| | GA | 50 | 26.6 | 116 | 40.0 | **0.51 (0.33-0.78)** | **0.002** |
| | AA | 15 | 8.0 | 21 | 7.2 | 0.87 (0.42-1.82) | 0.72 |

[a] $\chi^2$ significant; p<0.05; [#] ORs adjusted for all environmental factors
Bold number indicate significant p value < 0.05

**Table 4.8: Genotype representation and associations under dominant and recessive model between lung cancer cases and controls**

| Gene | Model | Effect | Reference | OR** (95% CI) | p value |
|---|---|---|---|---|---|
| **CYP1A1*2A** | Dominant | TC+CC | TT | **1.82 (1.20-2.76)** | **0.004** |
| | Recessive | CC | TT+TC | 1.14 (0.66-1.97) | 0.61 |
| **CYP1A1*2C** | Dominant | AG+GG | AA | 1.25 (0.82-1.89) | 0.29 |
| | Recessive | GG | AA+AG | 2.11 (0.74-5.99) | 0.16 |
| **EPHX1 Tyr113His** | Dominant | TC+CC | TT | **0.60 (0.40-0.90)** | **0.01** |
| | Recessive | CC | TT+TC | 1.42 (0.91-2.24) | 0.12 |
| **EPHX1 His139Arg** | Dominant | AG+GG | AA | **1.58 (1.04-2.42)** | **0.03** |
| | Recessive | GG | AA+AG | 2.03 (0.69-5.91) | 0.19 |
| **GSTP1** | Dominant | AG+GG | AA | 1.29 (0.86-1.94) | 0.24 |
| | Recessive | GG | AA+AG | 0.76 (0.30-1.90) | 0.56 |
| **SULT1A1** | Dominant | GA+AA | GG | **0.56 (0.37-0.84)** | **0.006** |
| | Recessive | AA | GG+GA | 1.16 (0.56-2.41) | 0.67 |

Bold number indicate significant p value < 0.05

**Haplotype analysis**

Table 4.9 summarizes the associations between the frequency distributions of the haplotypes in CYP1A1 and EPHX1 genes and the risk of lung cancer. The odds ratios were calculated using the most common haplotype as the reference group. In CYP1A1, "TA" haplotype was the most frequent among both cases and controls and showed significant association. Only CYP1A1-CG haplotype imparted increased risk to lung cancer (OR=1.49;95%CI=1.00-2.21,p=0.04). In EPHX1, the "TA" haplotype was the most common with frequencies of 44.79% and 45.04% in cases and controls respectively. No haplotype was found to be significantly associated with lung cancer risk.

**Table 4.9: Distribution of CYP1A1 and EPHX1 haplotype frequency among lung cancer cases and controls**

| | Case ( 376) | Control (580) | $X^2$ | P value | OR (95%CI);P | D' |
|---|---|---|---|---|---|---|
| | % (n) | % (n) | | | | |
| **CYP1A1 2A*2C** | | | | | | |
| TA | 53.34 (201) | 60.80 (352) | 5.00 | 0.02 | 1.00 | 0.72 |
| TG | 3.31 (12) | 2.65 (16) | 0.21 | 0.64 | 1.31 (0.57-3.00);0.50 | |
| CA | 26.45 (99) | 23.51 (137) | 0.95 | 0.32 | 1.26 (0.91-1.74); 0.15 | |
| CG | 16.90 (64) | 13.04 (75) | 2.94 | 0.08 | **1.49 (1.00-2.21); 0.04** | |
| **EPHX1 Tyr113His * His139Arg** | | | | | | |
| TA | 44.79 (168) | 45.04 (262) | 0.05 | 0.81 | 1.00 | 0.21 |
| TG | 12.39 (47) | 10.31 (59) | 1.57 | 0.20 | 1.23(0.78-1.94); 0.30 | |
| CA | 35.26 (133) | 40.13 (233) | 1.82 | 0.17 | 0.88 (0.65-1.19); 0.42 | |
| CG | 7.56 (28) | 4.52 (26) | 2.64 | 0.10 | 1.67 (0.91-3.06); 0.07 | |

Bold number indicate significant p value < 0.05

**Risk associated with SNPs stratified by smoking**

Since smoking is a well established risk factor to lung cancer and was the strongest independent risk factor in LR, we further stratified the data by smoking status. Distribution and risk associated with genetic factors after stratification is shown in Table 4.10. Heterozygous and homozygous variant genotypes of CYP1A1*2A polymorphism imparted significant risk in non-smokers (OR=2.88;95%CI=1.22-6.81,p=0.016 and OR=4.35;95%CI=1.47-12.84,p=0.008). Also, CYP1A1*2C variant genotype and GSTP1 Ile105Val heterozygous genotype were significantly associated with increased risk in non-smokers (OR=11.81;95%CI=1.24-111.98,p=0.03 and OR=2.40;95%CI=1.15-5.03,p=0.01). Heterozygous genotypes in EPHX1 Tyr113His and SULT1A1 Arg213His were associated with 66% and 55% reduced risk in smokers (OR=0.34;95%CI=0.18-0.63,p=0.001 and OR=0.45;95%CI=0.25-0.80,p=0.007 respectively). However heterozygous genotype in EPHX1 His139Arg conferred significant risk in smokers (OR=1.92;95%CI=1.07-3.45,p=0.02).

**Table 4.10: Main effects of genotypes on lung cancer risk stratified by smoking**

| Polymorphism | Genotype | Smoker | | Non Smoker | |
|---|---|---|---|---|---|
| | | Case/Control (n,%) | OR (95% C.I.),p value* | Case/Control (n,%) | OR (95% C.I.),p value* |
| **CYP1A1*2A** | TT | 44(33.3)/57(41.0) | 1.0 | 11(19.6)/65(28.6) | 1.0 |
| | TC | 74(56.1)/61(43.9) | 1.45(0.84-2.50),0.17 | 29(51.8)/63(41.7) | **2.88(1.22-6.81),0.016** |
| | CC | 14(10.6)/21(15.1) | 0.83(0.36-1.91),0.66 | 16(28.6)/23(15.2) | **4.35(1.47-12.84),0.008#** |
| **CYP1A1*2C** | AA | 86(65.2)/93(66.9) | 1.0 | 36(64.3)/113(74.8) | 1.0 |
| | AG | 40(30.3)/40(28.8) | 1.14(0.65-2.02),0.63 | 16(28.6)/37(24.5) | 1.53(0.67-3.48),0.30 |
| | GG | 6(4.5)/6(4.3) | 1.71(0.43-6.74),0.43 | 4(7.1)/1(0.7) | **11.81(1.24-111.98),0.03** |
| **EPHX1 Tyr113His** | TT | 60(45.5)/41(29.5) | 1.0 | 22(39.3)/53(35.1) | 1.0 |
| | TC | 35(26.5)/71(51.1) | 0.34(0.18-0.63),0.001# | 16(28.6)/62(41.1) | 0.62(0.25-1.54),0.30 |
| | CC | 37(28.0)/27(19.4) | 1.14(0.57-2.29),0.69 | 18(32.1)/36(23.8) | 1.03(0.41-2.56),0.94 |
| **EPHX1 His139Arg** | AA | 80(60.6)/103(74.1) | 1.0 | 41(73.2)/109(72.2) | 1.0 |
| | AG | 48(36.4)/32(23.0) | 1.92(1.07-3.45),0.02 | 11(19.6)/38(25.2) | 0.98(0.41-2.36),0.98 |
| | GG | 4(3.0)/4(2.9) | 1.39(0.31-6.25),0.66 | 4(7.1)/4(2.6) | 4.25(0.54-33.15),0.16 |
| **GSTM1** | WildType | 91(68.9)/86(61.9) | 1.0 | 31(55.4)/91(60.3) | 1.0 |
| | Null | 41(31.1)/53(38.1) | 0.87(0.51-1.48),0.62 | 25(44.6)/60(39.7) | 1.25(0.61-2.54),0.53 |
| **GSTT1** | WildType | 106(80.3)/104(74.8) | 1.0 | 49(87.5)/113(74.8) | 1.0 |
| | Null | 26(19.7)/35(25.2) | 0.75(0.40-1.41),0.37 | 7(12.5)/38(25.2) | 0.48(0.19-1.20),0.11 |
| **GSTP1** | AA | 69(52.3)/77(55.4) | 1.0 | 33(58.9)/102(67.5) | 1.0 |
| | AG | 54(40.9)/55(39.6) | 1.35(0.77-2.36),0.29 | 23(41.1)/41(27.2) | **2.40(1.15-5.03),0.01#** |
| | GG | 9(6.8)/7(5.0) | 1.49(0.49-4.56),0.47 | 0/8(5.3) | NA |
| **SULT1A1** | GG | 84 (63.6)/69 (49.6) | 1.0 | 39(69.6)/84(55.6) | 1.0 |
| | GA | 35 (26.5)/58 (41.7) | 0.45(0.25-0.80),0.007# | 15(26.8)/58(38.4) | 0.54(0.24-1.19),0.13 |
| | AA | 13 (9.8)/12 (8.6) | 1.11(0.45-2.74),0.81 | 2(3.6)/9(6.0) | 0.48(0.09-2.54),0.39 |

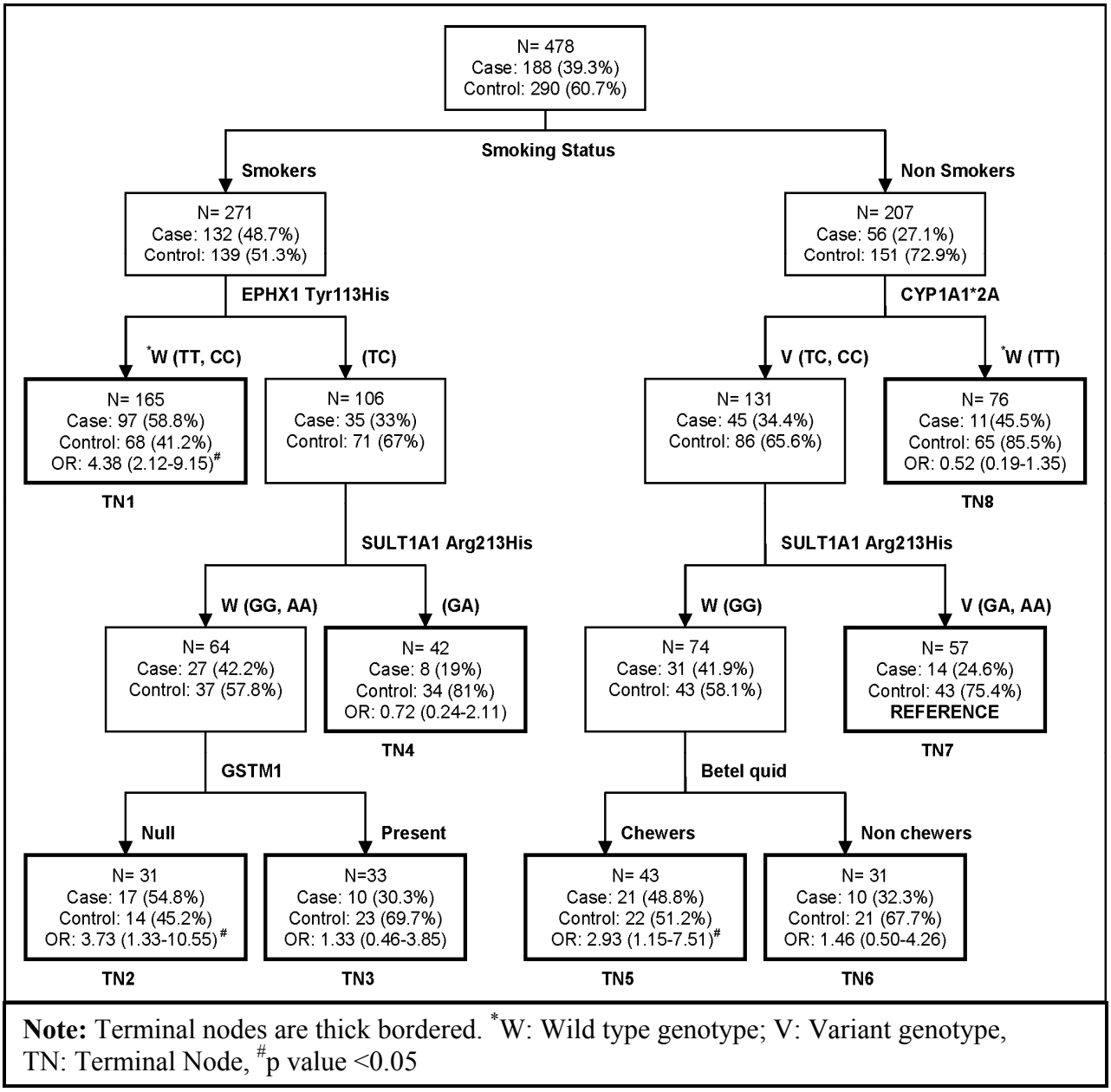
*p values adjusted for tobacco chewing, betel quid chewing and alcohol consumption

# Significant after p-value adjustment for multiple comparision (Sidak correction)

Bold number indicate significant p value < 0.05

**CART analysis**

Figure 4.9 shows the selected CART model constructed on all investigated genetic variants and environmental risk factors. The final tree contained eight terminal nodes. The first split of the root node was on smoking habit, indicating that smoking is the strongest risk factor for lung cancer. Among smokers, the subsequent splits showed interactions between EPHX1 Tyr113His, SULT1A1 Arg213His and GSTM1. In non-smokers first split was on CYP1A1*2A status, which was in concordance with the LR analysis where CYP1A1*2A showed strong association to risk only in nonsmokers. Further interactions were predicted by SULT1A1 Arg213His polymorphism and betel quid status. Terminal node 7, which comprised of least percentage of cases in non-smokers, was taken as reference to calculate OR for other terminal nodes. Among smokers maximum risk was observed for terminal node1 consisting of EPHX1 113TT (Tyr/Tyr) or -113CC (His/His) genotypes (OR=4.38;95%CI=2.12-9.15) and for terminal node 2 with combination of EPHX1 113TC (Tyr/His), SULT1A1 213GG (Arg/Arg) or AA (His/His) and GSTM1 null genotypes (OR= 3.73;95%CI=1.33-10.55, p=0.006). In non-smokers high risk was seen for terminal node 5 comprising of CYP1A1*2A 6235CC or TC, SULT1A1 213GG (Arg/Arg) and betel quid chewing (OR=2.93;95%CI=1.15-7.51, p= 0.01). Parallel to the above, CART analysis on separate data sets of smokers and non-smokers was also performed. However, we did not detect any high-order interaction in these analysis (data not shown).




**Note:** Terminal nodes are thick bordered. *W: Wild type genotype; V: Variant genotype, TN: Terminal Node, #p value <0.05

**Figure 4.9: Classification and regression tree model for xenobiotic metabolizing gene polymorphisms and environmental risk factors.**

**MDR Analysis**

MDR analysis was applied to further explore gene-gene and gene-environment interactions. Best predictive models up to 4 orders of interaction, along with their CVC and TBA are summarized in Table 4.11. The analysis was run separately for total data set and data sets stratified on smoking status. For total data set, smoking was the best one locus model with highest CVC (10/10) and testing accuracy of 0.6114 which was statistically significant ($p<0.001$) determined by 1000 fold permutation testing. For a 2-locus interaction, combination of smoking and EPHX1 Tyr113His was most significant with CVC of 10/10 and TBA of 0.6407 ($p<0.001$). The 3 locus model consisted of smoking, EPHX1 Tyr113His and EPHX1 His139Arg with TBA of 0.6497 ($p<0.001$) and CVC of 10/10. The 4 loci interaction model of smoking, EPHX1 Tyr113His, EPHX1 His139Arg and SULT1A1 Arg213His, was the best model identified, with maximum CVC (10/10) and TBA (0.6503, $p<0.001$). This model had a chi-square value of 66.31 ($p<0.0001$) and an OR of 4.93 (95%CI=3.32-7.33). In smokers the best interaction model was the three loci model consisting of tobacco chewing, EPHX1 Tyr113His and SULT1A1 Arg213His having maximum CVC (10/10) and TBA (0.6436, $p<0.001$) among all models identified. The model imparted 3.5 fold increased risk for lung cancer (95%CI=2.69-7.69). In non-smokers the best model was the three loci model comprising of CYP1A1*2A, GSTP1 Ile105Val and SULT1A1 Arg213His with CVC of 10/10 and TBA of 0.6677 ($p<0.005$) and an OR of 7.32 (95%CI=3.24-16.53).

**Table 4.11: Multifactor dimensionality reduction analysis**

| | No. of Locus | Model | p value ($\chi^2$ test) | TBA | p-value* | CVC | p-value* |
|---|---|---|---|---|---|---|---|
| **Total Data Set** | | | | | | | |
| | 1st order | Smk | p < 0.0001 | 0.6114 | <0.001 | 10 | 0.391 |
| | 2nd order | Smk Ex3 | p < 0.0001 | 0.6407 | <0.001 | 10 | 0.391 |
| | 3rd order | Smk Ex3 Ex4 | p < 0.0001 | 0.6497 | <0.001 | 10 | 0.391 |
| | **4th order**** | **Smk Ex3 Ex4 SULT** | **p < 0.0001** | **0.6503** | **<0.001** | **10** | **0.391** |
| **Smokers** | | | | | | | |
| | 1st order | Ex3 | p < 0.0001 | 0.6228 | 0.012 | 10 | 0.402 |
| | 2nd order | Tbc Ex3 | p < 0.0001 | 0.6105 | <0.02 | 9 | 0.623 |
| | **3rd order**** | **Tbc Ex3 SULT** | **p < 0.0001** | **0.6436** | **<0.001** | **10** | **0.402** |
| | 4th order | Tbc Alc Ex3 SULT | p < 0.0001 | 0.6268 | <0.008 | 7 | 0.846 |
| **Non Smokers** | | | | | | | |
| | 1st order | 2A | p = 0.0019 | 0.6170 | 0.09 | 10 | 0.372 |
| | 2nd order | 2A SULT | p = 0.0004 | 0.5562 | 0.46 | 8 | 0.734 |
| | **3rd order**** | **2A P1 SULT** | **p < 0.0001** | **0.6677** | **<0.005** | **10** | **0.372** |
| | 4th order | 2A 2C P1 SULT | p < 0.0001 | 0.6439 | <0.021 | 10 | 0.372 |

*1,000-fold permutation test. ** Best models selected with maximum cross-validation consistency (CVC) and maximum testing balance accuracy (TBA). Labels: Smk: smoking, Ex3: EPHX1 Tyr113His, Ex4: EPHX1 His139Arg, SULT: SULT1A1 Arg213His, Tbc: tobacco chewing, Alc: alcohol consumption, 2A: CYP1A1*2A, P1: GSTP1 Ile105Val, 2C: CYP1A1*2C.

**False positive report probability (FPRP)**

Table 4.12 shows the FPRPs for the 3 best models obtained from MDR analysis. The 4-loci predictor model on total data set and 3-loci model in smokers showed excellent reliability even when assuming very low prior probabilities (from $10^{-3}$ to $10^{-6}$) for detecting ORs of 1.5 and 2.0. However the best model selected in non smoker category showed true association only at high probability of $10^{-1}$ for detecting OR=1.5 and till $10^{-2}$ for detecting OR=2.0.

**Table 4.12: False positive report probability and odds ratio for best model of MDR analysis**

| | OR (95% CI) p value | OR=1.5 | Prior Probability | | | | | | OR=2.0 | Prior Probability | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Power | $10^{-1}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ | Power | $10^{-1}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
| **Total Data Set Smk Ex3 Ex4 SULT** | 4.93 (3.32-7.33) p < 0.0001 | 0.0001 | **<0.0001** | **<0.0001** | **0.001** | **0.015** | **0.131** | **0.131** | 0.0001 | **<0.0001** | **<0.0001** | **<0.0001** | **<0.0001** | **<0.0001** | **<0.0001** |
| **Smokers Tbc Ex3 SULT** | 4.55 (2.69-7.69) p < 0.0001 | 0.0001 | **0.008** | **0.081** | 0.472 | 0.900 | 0.989 | 0.989 | 0.001 | **0.000** | **0.001** | **0.014** | **0.125** | 0.588 | 0.588 |
| **Non Smokers 2A P1 SULT** | 7.32 (3.24-16.53) p < 0.0001 | 0.0001 | **0.180** | 0.708 | 0.961 | 0.996 | 1.00 | 1.00 | 0.001 | **0.016** | **0.155** | 0.650 | 0.949 | 0.995 | 0.995 |

Prior probabilities ranging from 0.1 to $10^{-6}$, with the estimated statistical power to detect an OR of 1.5 or 2.0 with α level equal to the observed p-value
Bold type indicates the FPRP for the most likely prior probabilities i.e. a noteworthy association at the 0.2 FPRP

## Interaction entropy graphs

After identifying the high-risk combinations using MDR approach, interaction entropy algorithm was applied to interpret relationship between the variables. Graphs were constructed on data set stratified by smoking (Figure 4.10). In smokers, EPHX1 Tyr113His had a large independent effect (4.64%) and a non-additive interaction with tobacco chewing (entropy 1.79%). Considerable entropy was associated with SULT1A1 Arg213His (1.88%) and its interaction with tobacco chewing further removed 1.49% of entropy from case-control group. However we did not detect any non-linear interaction between the two SNPs in the model. We found small percentages of the entropy in case–control status explained by alcohol consumption (0.56%) and tobacco chewing (0.70%) independently, but a large percentage of entropy explained by the interaction between these two environmental factors (2.47%). In non-smokers, CYP1A1*2A showed strongest main effect with entropy removal of 4.7%. GSTP1 Ile105Val too had a strong independent effect (entropy removal=3.28%) and its interaction with SULT1A1 Arg213His further removed 3.02% of entropy. A strong synergistic interaction was observed between SULT1A1 Arg213His and CYP1A1*2C as the combination removed an additional 2.61% of the total entropy.

.

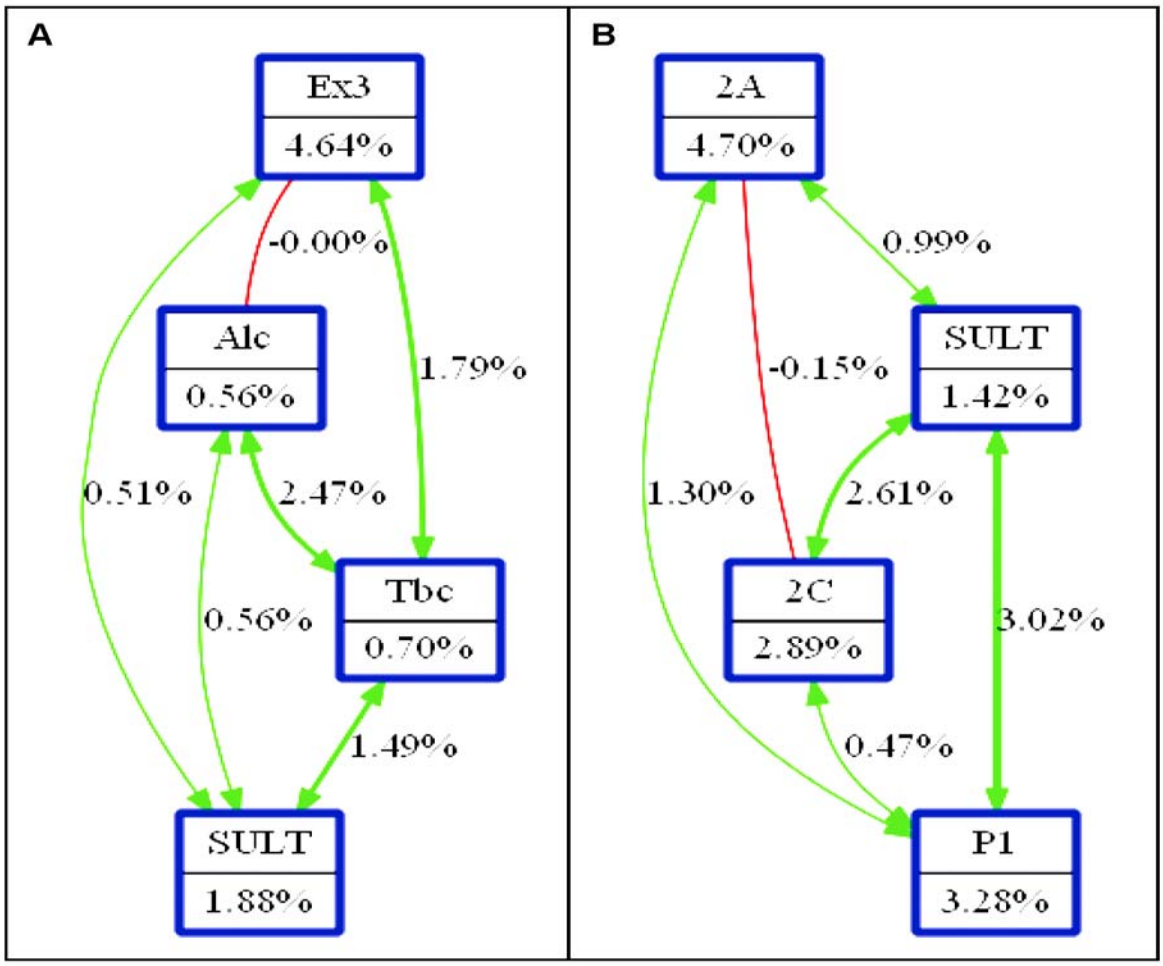



**Figure 4.10: Interaction entropy graphs**.

The interaction model describes the percentage of the entropy (information gain) removed by each variable (main effect: represented by nodes) and by each pairwise combination of attributes (interaction effect: represented by connections). Attributes are selected on the basis of MDR results obtained in case of **(A)** Smokers and **(B)** Non smokers

**Labels:** Ex3: EPHX1 Tyr113His, Alc: alcohol consumption, Tbc: Tobacco chewing, SULT: SULT1A1 Arg213His, 2A: CYP1A1*2A, 2C: CYP1A1*2C, P1: GSTP1 Ile105Val.

## DISCUSSION

The present study used multiple analytical methods to first assess associations and then explore possible interactions of xenobiotic metabolizing genes with environmental factors in risk to lung cancer. The applied data mining approaches have the ability to search and identify interactions regardless of the significance of the main effects. The most significant finding of this study is the consistently identified gene-gene and gene environment interactions by all the three statistical approaches.

Smoking is the primary etiological factor in lung cancer. The same was reflected in the present study as smoking showed strong association in LR, best one factor model in MDR and formed first split in CART. Interaction of EPHX1 Tyr113His and SULT1A1 Arg213His was consistently identified in smokers. Both EPHX1 Tyr113His and SULT1A1 Arg213His conferred reduced risk in smoker subset in LR. The two polymorphisms along with EPHX1 His139Arg formed the best predictor model in MDR analysis in smokers and also formed subsequent splits within smokers in CART. EPHX1 enzyme catabolizes epoxides from PAH into dihydrodiols, which involves generation of more reactive carcinogenic metabolites. Substitution of a variant His allele at codon 113 (EPHX1 Tyr113His) decreases the activity of this enzyme [Hassett et. al, 1994] thereby reduces the risk of cancer. Studies on lung cancer suggest protective effect for His113 (slow type) as compared to Tyr113 (fast type) which imparts increased lung caner risk [Benhamou et. al., 1998, London et. al., 2000, Zhou et. al., 2002]. The variant allele has also been suggested to decrease the risk of ovarian cancer [Lancaster et. al., 1996]. We have earlier reported similar results from the same population in esophageal cancer showing His113 allele to be associated with a significantly reduced risk in smokers [Ihsan et. al., 2010]. Reflecting the same, in CART analysis Terminal node 1 of imparts over 4 fold high risk to smokers possibly due high proportion of the wild Tyr113 homozygous genotype. Sulphonation reaction of SULT1A1 is a detoxification reaction, however it also involves bioactivation of certain procarcinogens, including heterocyclic amines and PAHs to form carcinogen-DNA adduct [Glatt et. al., 1997, Nowell et. al., 2000]. In vitro model studies suggest that substitution of histidine at position 213 in the amino acid sequence is associated with decreased substrate affinity and a lower level of protein [Jones et. al., 1993] which

might protect against chemical carcinogenesis of PAHs in lung cancer [Denissenko et. al., 1997]. Results on association of SULT1A1 Arg213His and risk of cancer are inconsistent, from null association with risk of colorectal cancer [Wong et. al., 2002] and prostate cancer [Steiner et. al., 2000] to increase in risk of breast cancer associated with His213 allele [Zheng et. al., 2001]. Another study on colorectal cancer showed a significantly reduced risk for individuals carrying His213 allele [Bamber et. al., 2001]. A Meta-analysis by Kotnis et. al., 2008 showed a significant protective effect of the polymorphism in seven studies of genitourinary cancers.

Among non-smokers CYP1A1*2A and GSTP1 Ile105Val were the most important polymorphisms identified for lung cancer development. The variant allele of both the polymorphisms conferred significant risk in the non smoking subgroup in LR analysis. Similarly, MDR 3 loci model of CYP1A1*2A, GSTP1 Ile105Val and SULT1A1Arg213His polymorphisms was the best predictor of risk in non-smokers. The CYP1A1 6235T>C MspI (CYP1A1*2A) polymorphism, is associated with higher enzymatic activity towards benzopyrene [Cosma et. al., 1993, Landi et. al., 1994]. Investigations on association between CYP1A1 polymorphisms and lung cancer have yielded equivocal results [Taioli et. al., 1998, Lee et. al., 2008]. Similar to our findings, a study by Taioli et. al., 2003 reported association of CYP1A1*2A variant allele with lung cancer, however after stratification by smoking the association remained confined to non-smokers only. Further, in a pooled analysis of 11 studies on CYP1A1*2C polymorphism in lung cancer, Le Marchand et. al., 2003 found it to be associated with risk in non-smokers, a finding which corroborates our results. Another study by Jose et. al., 2010 on lung cancer found no association of any CYP1A1 polymorphism with smokers. Similar results were reported in colorectal cancer where heterozygous and variant genotypes of both CYP1A1*2A and CYP1A1*2C conferred risk in combinations with NAT2 only among non-smokers [Yoshida et. al., 2007]. In vitro cDNA expression study suggests that GSTP1 with 105Val variant results in a protein with reduced enzyme activity [Ali-Osman et. al., 1997], however it is reported to play an unlikely role for smoking-related cancers [Cote et. al., 2009]. Similar observation has been reported from breast cancer [Zhao et. al., 2001]. Probably the precise role of GSTP1 in

carcinogenesis can be determined by the kind of xenobiotic involved owing to its substrate specificity and affinity [Coles et. al., 2000].

Confirming to its exploratory nature, CART analysis identified two more risk factors, GSTM1 null genotype in smokers and betel quid chewing in non-smokers. The results are quite plausible because both hold functional and biological significance. High risk for smoking related lung cancer has been reported in individuals deficient in GSTM1 [Zhong et. al., 1991, Ketterer et. al., 1992, Jin et. al., 2010]. Smokers with the GSTM1 enzyme have approximately one-third of the risk for lung carcinoma than smokers without the enzyme [Nazar-Stewart et. al., 1993]. There are numerous reports of association between GSTM1 null genotype and smoking in various cancers including esophageal [Jain et. al., 2006], bladder [Rouissi et. al., 2011], colorectal [Chen et. al., 2004] and oral [Buch et. al., 2002]. A recent study by Wen et. al. 2010 showed betel quid chewing increases lung cancer risk in non-smokers, with smoking habit further enhancing the risk. Betel quid chewing is a unique and widespread habit in the north-eastern (NE) region of India. Betel quid is a chewing mixture of whole betel/areca nut wrapped with betel leaves spread with white lime with frequent addition of tobacco. It is known to contain phenolic compounds and alkaloids, besides nitrosamines are formed from an in vivo reaction of betel arecoline, nitrite and thiocynate, all of which act as carcinogens [Awang 1988]. Studies have reported association between betel quid chewing and cancer risk. Significant association of betel quid chewing with risk of oral, stomach [Ihsan et. al., 2011] esophageal [Ihsan et. al., 2010] and breast cancer [Kaushal et. al., 2010] has been reported from the study population. It would be reasonable to assume that the association of betel quid chewing with lung cancer is a result of a complex combination of direct and indirect action of tobacco carcinogens contained in it.

A post-hoc analysis through entropy graph was done to visualize and interpret interaction models identified by MDR. The previously documented main effects of EPHX1 Tyr113His and SULT1A1 Arg213His in smokers and CYP1A1*2A and GSTP1 Ile105Val in non-smokers were evident. Further, synergistic interactions of SULT1A1 Arg213His with GSTP1 Ile105Val and with CYP1A1*2C were observed in non-smokers.

As haplotype are more efficient and informative than separate markers, haplotype association analysis was carried out in CYP1A1 and EPHX1 genes. CG haplotype in CYP1A1 was significantly associated with risk of lung cancer. Noteworthy were results in EPHX1, where frequency of haplotypes among cases was strikingly similar to report published in esophageal cancer from north India [Jain et. al., 2008].

Although both MDR and CART validated LR results, yet they differed in identifying some unique interactions, reflecting different methods followed by each program. Both approaches provide a clear advantage over the traditional LR by identifying non-linear interactions among discrete genetic and environmental attributes. Significant findings of the study are summarized in Figure 4.11. It would be safe to assume a definite association of the commonly recognized factors to lung cancer that might have implications on future studies. Role of CYP1A1*2A polymorphism is evident only among non smokers in all the three methods. LR and CART analyses even showed a gene-dosage effect for the increased lung cancer risk with the increasing number of variant allele in the CYP1A1*2A polymorphism. As aforesaid, this finding provides support to previously published reports [Taioli et. al., 2003, Le Marchand et. al., 2003, Yoshida et. al., 2007, San Jose et. al., 2010]. MDR and CART analysis show epitasis between EPHX1 Tyr113His and SULT1A1 Arg213His polymorphisms exclusively among smokers. Their combined models confer risk to lung cancer however individually both act as protective factors in smokers only. These factors hold their importance as the SNPs are functionally and biologically relevant and have been implicated in the carcinogenesis process in previous studies on various cancers

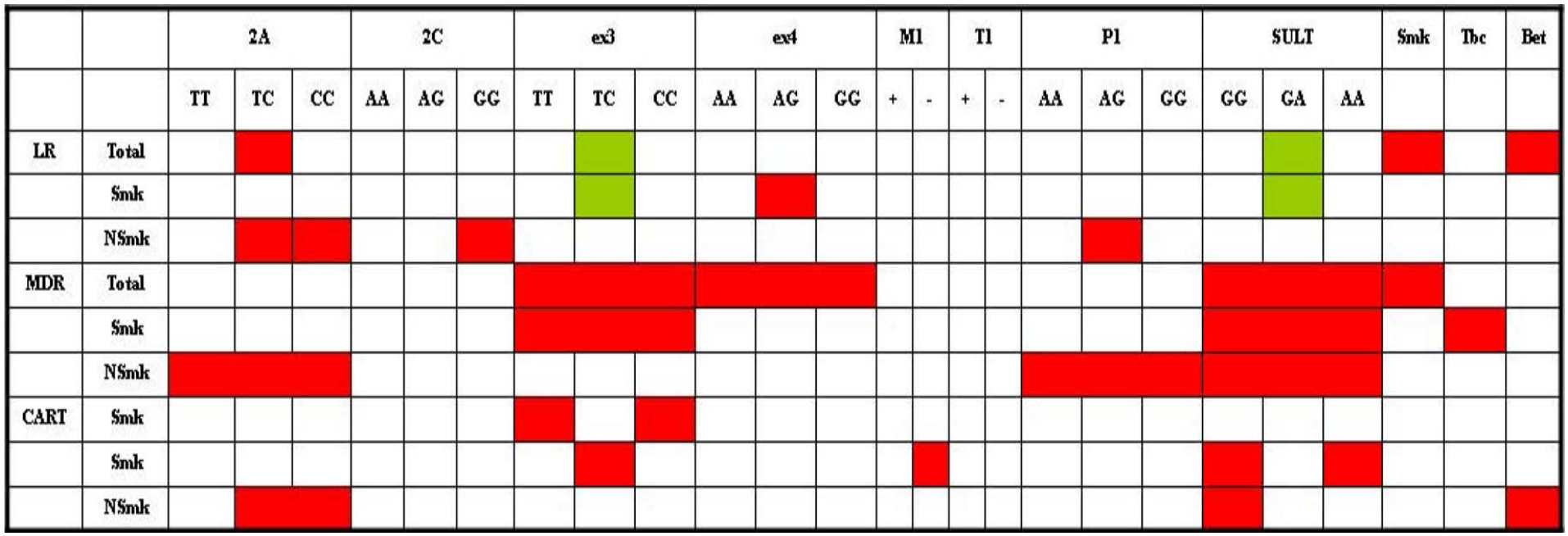



**Figure 4.11: Summarized results for LR, MDR and CART analyses**

Green boxes indicate OR<1. Red boxes indicate OR>1. For MDR and CART significant interactions are shown. LR results should be read individually. Alcohol was excluded as it did not appear significant in any analysis

Major challenge for the identification of true genetic and interactive effects in a multi-factorial study is simultaneous testing of several hypotheses. The three methods of analysis used in this study address the same research hypotheses but differ in terms of their statistical methodologies and analytical approaches. P-value adjustment for multiple testing was performed through SIDAK correction in LR model with the equality as $(1-(1-\alpha)^{1/n})$ where n=4 both in total and stratified analyses. Multiple testing in data mining approaches such as CART and MDR sometimes compromises upon the comparative power. When numerous null hypotheses are being tested yielding higher order interacting combinations the inference drawn from a single erroneous rejection is not an appropriate strategy, rather the proportion of erroneous rejection needs to be controlled. This is achieved by estimation of FPRP. These approaches utilize internal cross-validations and permutation testing of p-value reducing the chances of making type I errors. Both MDR and CART apply cross validation of data before selecting the best model however MDR also uses 1000 fold-permutation testing, to validate its results for minimizing the proportion of false-positives due to multiple testing. The cross validation (5-10 fold) dividing the whole data set into different sets of training and testing set prevents over-fitting and artificial accuracy improvement. Permutation test is considered the gold standard for accurate multiple testing correction. Controlling for false discovery rate (FDR) is a more realistic approach than as compared to concerns raised by the multiple hypothesis testing. This is because FDR is the proportion of incorrect rejection among all such rejections. Likewise, the best models derived from MDR on total data set and smokers set in this study showed good reliability as associations remained robust even at low prior probabilities for FPRP testing. CART analysis was able to define genetic associations with fairly good measures. Correct classification of cases and controls in test data set was approximately 63% for both.

There might be some limitations to this study. The sample size of our study was relatively small, however based on the evidences (OR) provided by our research group on association between GSTs with lung cancer [Yadav et. al., 2010], the minimum sample size determined was 176 at 5% level of significance and 90% power. Polymorphisms of EPHX1 Tyr113His and SULT1A1 Arg213His in cases showed deviation from HWE. After ruling out false positive associations and genotyping errors perhaps population stratification, could be a

reason for this deviation. However, the cases were incident, and thus, the data do not show report or recall bias. Also case-control matching was done in reference to age, gender, and ethnicity, thereby controlling for any confounding effect accounted by these variables.

In conclusion this study highlights that better predictors for lung cancer risk can be obtained through polygenic approaches and exploring gene-environment interactions. The study identified distinct patterns of interaction in smoking and non smoking sub groups. However, the results presented should be treated with caution since this is the first epidemiological evidence identifying the complex relationship between genetic polymorphisms and cancer susceptibility in the studied population. Further studies with large samples in independent populations are warranted to validate the findings of this study.

# *Chapter 5*

# *Investigation on copy number polymorphism of GSTM1 and GSTT1 in susceptibility to lung cancer in a high-risk population from North East India*

## *Introduction*

Tobacco smoking remains the primary etiological factor associated with the development of lung cancer accounting for nearly 80-90% of the disease. Polycyclic aromatic hydrocarbons (PAHs) particularly benzo[a]pyrene (BaP) and nitrogen containing nitrosamines and aromatic amines are main carcinogens present in tobacco smoke that are implicated in lung carcinogenesis. Non-smoking tobacco and betel quid have also been implicated in lung carcinogenesis probably due to their accompanied consumption with smoking [Wen et. al., 2010]. Deleterious effects of tobacco carcinogens are primarily mediated through DNA adduct formations following their activation in the detoxifying pathways. Activated PAHs and N-nitroso compounds produced by phase I xenobiotic metabolizing enzymes are substrates for the *GSTM1* and *GSTT1* phase II enzymes.

Null polymorphism in both *GSTM1* and *GSTT1* correspond to the deletion of the genes which abolishes their enzyme activity. The *GSTM1* is reported to be homozygously deleted in around 53% of the Caucasian and Asian population. However deletion of *GSTT1* was reportedly higher in Asians (47 to 64%) [Raimondi et. al., 2006]. Extensive literature evidences are available on association of *GSTM1* and *GSTT1* genotypes with lung cancer [Raimondi et. al., 2006, Carlsten et. al., 2008]. Most of these literatures compare the "null" genotype with the "non-null" genotype and thus do not distinguish between one and two copy number of the genes. However, studies have reported a trimodal phenotype distribution for both *GSTM1* and *GSTT1* identifying homozygous wild type (+/+), hemizygous (+/-) and null (-/-) genotypes of the genes[Seidegård et. al., 1985, Sprenger et. al., 2000]. These studies suggest a gene dosage effect with three alleles corresponding to fast, intermediate and slow enzyme activity. Enzymatic activity of *GSTT1* has been reported to be varying with the copy number of the gene [Infante-Rivard et. al., 2006]. Sprenger et. al., [2000] in their genotype-phenotype comparison showed correlation of significantly increased enzyme activity in individuals with two copy number of the *GSTT1* compared to those with one copy number. Correspondingly, Roodi et. al., [2004] showed that the relative risk of breast cancer increases with the present allele (+/- and +/+ genotypes) compared with -/- genotype, however this trend was not statistically significant.

Several methods (standard and long-range PCR) in the past have been used for distinguishing *GSTM1* [Roodi et. al., 2004, Buchard et. al., 2007] and *GSTT1* [Sprenger et. al., 2000, Naito et. al., 2006, Buchard et. al., 2007] alleles into three genotypes. These methods were primarily based on the fact that the two genes are flanked by highly homologous regions. Primers flanking these regions were used for detecting "null" and "non-null" genotypes with amplification product ranging from 450bp to 14kb in long arm PCRs. However these PCR based gel electrophoresis genotyping methods were technically difficult and time consuming and did not yield reliable result on degraded and small quantity of DNA thus limiting their use especially in large scale epidemiological and clinical studies. Recent studies have used Taqman based real-time PCR assays to discriminate between the wild-type, hemizygous deletion, and homozygous deletion of the *GSTM1* and *GSTT1* genes [Timofeeva et. al., 2009, Lam et. al., 2009]. Real-time PCR is the best method for candidate copy number detection because of its low screening cost and rapid result generation. Specific assay design along with a validated experimental design and inclusion of controls with reference assays increases the accuracy and reliability.

In our previous report on association of GST polymorphisms, comparing the null genotype (-/-) with combined non-null genotype (+/- and +/+) using traditional multiplex PCR-gel electrophoresis method in high risk North-east Indian population, we showed a significant protective effect of *GSTM1* and *GSTT1* null genotypes in lung cancer [Yadav et. al., 2010]. The present study examines the relationship between *GSTM1* and *GSTT1* gene and lung cancer risk by assessing potential gene dosage effects and gene-environment interactions.

## EXPERIMENTAL METHODS

### Materials

Agarose, Tris base, EDTA, NaCl, SDS, Triton X-100 and other fine chemicals were purchased from Sigma Chemicals, USA. Taq polymerase, dNTPs, MgCl2, was obtained from Invitrogen and MBI fermentas USA. Oligos were synthesized by Microsynth, Switzerland. DNA extraction kit were purchased from Qiagen Sciences, USA.

## Chemicals used

LYSIS BUFFER I: 30mm Tris-Hcl (Ph-8), 5mm EDTA, 50 Mm Nacl; LYSIS BUFFER II: 75mm Nacl, 2mm EDTA (Ph-8); SDS STOCK: 20 gm of SDS dissolved in 80 ml of TDW at $65^0$C. Make up volume up to 100 ml ; PROTEINASE K: 10 mg dissolved in 1 ml of TDW:1% ; AGAROSE: 1gm of agarose dissolved 1% TAE buffer.

## Patient recruitment and sample collection

The study was conducted in 154 histopathologically diagnosed lung cancer cases registered at Dr. Bhubaneswar Borooah Cancer Institute, Guwahati, Civil Hospital, Aizawl, and Sir Thutob Namgyal Memorial Hospital, Gangtok, the collaborating centers in north east India. The sample size was determined based on power calculation methods from evidences provided by our research group on association between GSTs (OR for *GSTT1*=0.47; 95%CI=0.24-0.93; p=0.03) and lung cancer [Yadav et. al., 2010], the minimum sample size calculated was 144 at 5% level of significance and 80% power.

## Inclusion criteria

Incident cases during the period of December 2006 to 2008 and willing to participate in the study were included. 154 voluntary, age (±5 years) and sex matched individuals were selected from the healthy relatives who accompanied cancer patients. This provided a readily available and cooperative source of controls from the same socio-economic background as the cases reducing confounding biases. Patients with only lung as their primary site of cancer were included. Final selected controls were included on the basis of no history of any obvious disease and those not taking any medication at the time of recruitment

## Exclusion criteria

Any subject with history of familial malignancy or pulmonary infectious disease was excluded both from case and control. Patients unwilling or too ill to participate in the study were excluded. Patients who had taken any form of treatment earlier (Secondary cases) were also excluded from the study**.**

**Patient details**

All subjects provided written informed consent for participation in this research which was done under a protocol approved by the institutional ethics committee of Regional Medical Research Centre, North East Region (Indian Council of Medical Research) and participating institutes. Information regarding smoking, usage of tobacco, betel quid and alcohol were obtained from subjects in a standard questionnaire used for all the centers. Smokers, chewers and drinkers were classified into two categories ever and never. For smoking, an individual who had never smoked or smoked less than 100 cigarettes in their lifetime and were not smoking at the time of reporting was considered never smoker or non-smokers. Ever smokers or smokers category included current smokers, and those who had quit within <1 year of reporting [Carlsten et. al., 2008]. As our collaborating centers were public hospitals a large majority of subject belonged to lower to middle socio-economic background. Demographic data and characteristics such as age, sex, smoking habit, usage of tobacco, betel quid and alcohol, were obtained from subjects in a standard questionnaire used for all the centers, in an in-person interview by a trained data collector. A majority of cases and controls were literate with full primary schooling and some upto the college level. The occupational history of the study participants revealed that most of them were farm laborers or engaged in petty jobs and the nature of such jobs did not exposed them to any occupational hazards. Any history of past or present illness was enquired or if undergoing any medication at the time of enrolment.

**Collection of blood samples:** As described in chapter 4

**DNA extraction**: As described in chapter 4

**Quantification of Genomic DNA:** As described in chapter 4

**Agarose Gel Electrophoresis of Extracted DNA Samples:** As described in chapter 4

## Quantitative real-time TaqMan PCR for GSTM1 and GSTT1 copy number determination

*GSTM1* and *GSTT1* genotyping was performed using Taqman Gene Copy Number Assays purchased from Applied Biosystems (Foster City, California). TaqMan gene copy number assays (*GSTM1*:Hs no Hs02595872_cn and *GSTT1*: Hs no Hs00817631_cn) were run simultaneously with a TaqMan Copy Number reference assay (RNase P: Part No. 4403326) in a duplex real-time polymerase chain reaction (PCR) reaction. Each 20µl assay containing 20ng of genomic DNA (5ul) was prepared according to protocols developed by Applied Biosystems for copy number detection. Real-time PCR reactions were run on ABI PRISM 7000 Sequence Detection System. Each reaction was run in duplicates. In addition, a no-template control was also included in each run to rule out any contamination. Universal thermal cycling conditions were used i.e 2 mins at 50°C, 10 mins at 95°C, followed by 40 cycles of 15 secs at 95°C and 60 secs at 60°C. Real-time data was collected by the SDS 1.0 software. The number of copies of the target sequence in each test sample is determined by relative quantitation (RQ) using the comparative CT (ΔΔCT) method. This method measures the CT difference (ΔCT) between target and reference gene, and then compares the ΔCT values of test samples to a calibrator sample which is known to have two copies of the target sequence. The copy number of the target is calculated to be two times the relative quantity. A sample homozygous for *GSTM1* and *GSTT1* wild type allele (2 copy number) was used as calibrator. The samples employed as calibrators were previously analyzed by PCR methods (described below) to confirm possession of two copies of the genes examined.

## Validations of copy number by PCR method

For validation of copy number estimation done through real-time PCR, 30% of the samples were reanalyzed through PCR-gel electrophoresis methods. Copy number detection of *GSTM1* was done through a two step method. Samples showing presence of *GSTM1* gene in multiplex PCR described in our previous report [Yadav et. al., 2010] were reanalyzed for detection of null allele through primer described for *GSTM1* null allele by Buchard et. al., [2007]. Samples showing amplification of the 4748bp null allele were considered as hemizygous genotype (1 copy) and those with no amplification were considered carrying 2 copies of *GSTM1* gene (Figure 5.1). A sample with null *GSTM1* genotype was included as positive control in each PCR. Genotypes of *GSTT1* were detected through multiplex PCR

described by Naito et. al., [2006]. Amplification products of 566bp and 458bp represented null and present alleles of *GSTT1* respectively (Figure 5.2). The results from PCR-gel electrophoresis method were in complete concordance with those from real-time PCR.

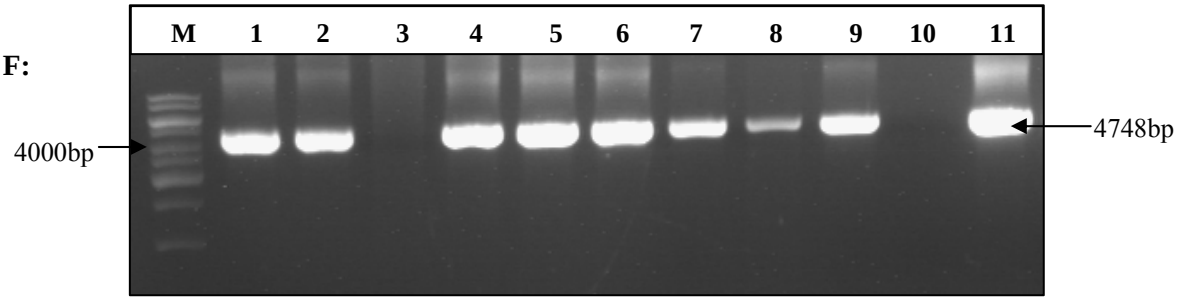



**Figure 5.1:EtBr stained agarose gel electrophoresis for GSTM1 genotyping.** Lanes 1,2,4-9and11 show hemizygous samples (1copy) and lanes 3and10 (no null allele amplification) show samples with 2 copy number of *GSTM1*. M-1kb ladder.

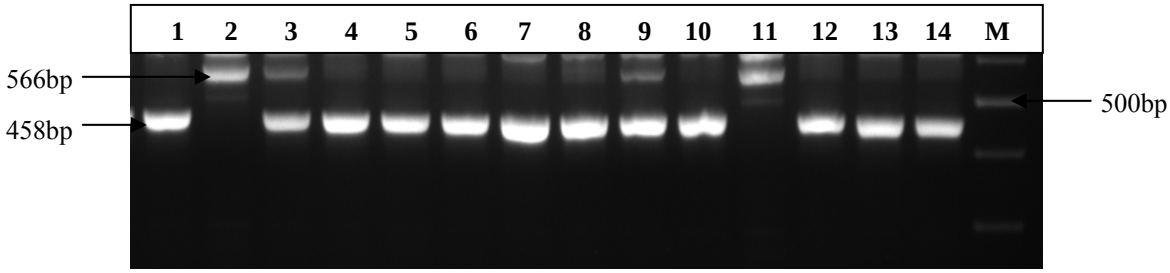



**Figure 5.2:EtBr stained agarose gel electrophoresis for GSTT1 genotyping.** Lanes 1,4-7,10,and12-14 show homozygous present alleles (2 copy number), Lanes 3,8,9 show hemizygous samples (1copy) and Lanes 2and11 (only null allele amplification) show samples with *GSTT1* null. M-500bp ladder.

### Statistical Analysis

The association of *GSTM1* and *GSTT1* genotypes with lung cancer was evaluated by multivariate conditional logistic regression. The association of tobacco smoking, tobacco chewing, betel quid chewing, and alcohol intake with disease development was assessed by chi square/Fisher's exact test. Estimates of cancer risk imparted by *GSTM1* and *GSTT1* genotypes and other covariates such as tobacco smoking, chewing, betel quid chewing, and alcohol were determined by deriving the odds ratio (OR) and corresponding 95% confidence intervals (95% CIs) using univariate and multivariable conditional logistic regression models. To evaluate the putative modifying effects of the GST genotypes on the effects of environmental factors, stratified analysis was performed for subjects positive for

individual risk factors. For all the tests, a two-sided $p<0.05$ was considered statistically significant. The data analysis was performed on STATA 8.0 software.

## RESULTS

### Characteristics of study subjects

A total of 154 lung cancer cases and 154 controls were successfully genotyped for polymorphism in *GSTT1* and *GSTM1*.The distribution of gender and ethnicity was similar for cases and controls. Male were overrepresented in the study compared to female (M/F ratio: 3.05). Mean age of cases and controls was 59.16±9.95 (range 35-80yrs) and 60.39±10.43 (range 38-85yrs) respectively. Distribution of both *GSTM1* and *GSTT1* genotype was in agreement with Hardy-Weinberg equilibrium (HWE) in controls ($p>0.05$), however deviation of *GSTT1* genotypes from HWE was seen in cases ($p=0.01$). The relevant characteristics of the subjects studied are shown in the (Table 5.1).

**Table 5.1: Distribution of demographic variables for lung cancer patients and controls**

| Variables | | Cases n (%) | Controls n (%) | *p-value |
|---|---|---|---|---|
| **Sex** | Male | 38 | 38 | |
| | Female | 116 | 116 | |
| **Age Mean (Range)** | | 59.16±10.04 | 60.03±10.03 | |
| **Smoking status** | Non smokers | 49 (31.8) | 83 (53.9) | |
| | Smokers | 105 (68.2) | 71 (46.1) | **<0.0001** |
| **Tobacco chewing** | Non chewers | 74 (48.1) | 82 (53.2) | |
| | Chewers | 80 (51.9) | 72 (46.8) | |
| **Betel quid chewing** | Non chewers | 35 (22.7) | 66 (42.9) | 0.36 |
| | Chewers | 119 (77.3) | 88 (57.1) | |
| **Alcohol consumption** | Non alcoholic | 114 (74.0) | 109 (70.8) | |
| | Alcoholic | 40 (26.0) | 45 (29.2) | **<0.0001** |
| *$\chi^2$ value | | | | |

## Association of genetic and environmental factors with lung cancer risk

The distribution of environmental risk factors and genotypes in cases and controls and their association with lung cancer risk is summarized in Table 5.2. Risk habits such as smoking, tobacco chewing and betel quid chewing were more common among the cases compared to controls. Smokers constituted 68.2% of the cases and 46.1% of the controls, thus smoking was associated with a significant risk of lung cancer (OR=3.03; 95%CI=1.73-5.31; p<0.001). Betel quid chewing was present in 77.3% of the cases and 57.1% of controls, with the habit conferring greater than 2 fold risk to chewers compared to non-chewers (OR=2.39; 95%CI=1.38-4.16; p=0.002).

Frequency of *GSTM1* wild-type and null alleles in the control population was 0.35 and 0.64 respectively. Distribution of wild type (+/+) two copy, hemizygous deletion (+/-) one copy and homozygous deletion (-/-) null copy of *GSTM1* genotypes was 20.8%, 42.2% and 37.0% in cases and 14.3%, 42.9% and 42.9% in controls. Compared to individuals with two copy (+/+) genotype, the relative risk of lung cancer was 0.73 (95%CI=0.37-1.44; p=0.37) for the hemizygous genotype (+/-) and 0.62 (95%CI=0.31-1.23; p=0.17) for the null genotype (-/-). There was no evidence of gene dosage effect for *GSTM1* ($P_{trend}$=0.13). In contrast *GSTT1* the wild type (+/+) two copy number and hemizygous one copy number genotype was more frequent in cases than controls (27.3% vs 19.5% and 58.4% vs 53.2% respectively). Patients with null genotype conferred 68% (OR=0.32; 95%CI=0.15-0.71;p=0.005) reduced risk compared to patients with two copy number of *GSTT1*. When risk associated with null genotype was compared with one copy number (hemizygous) of the gene it reduced to 51% (OR=0.49; 95%CI=0.25-0.95;p=0.03) (data not shown). Moreover, decreasing copy number of *GSTT1* gene showed a positive dose relationship with lung cancer ($P_{trend}$=0.006). A comparision of our data according to the classical 'null' versus the 'non null' genotype have attenuate the protective effect of *GSTT1* null genotype from 68% to 45% (OR=0.55; 95%CI=0.30-1.00;p=0.05) (Table 5.2).

**Table 5.2: Association of *GSTM1* and *GSTT1* genotypes and environmental risk factors with lung cancer susceptibility**

| | Cases n (%) | Controls n (%) | OR (95% C.I.), p value |
|---|---|---|---|
| **GSTM1** | | | |
| +/+ | 32 (20.8) | 22 (14.3) | Ref |
| +/- | 65 (42.2) | 66 (42.9) | 0.73 (0.37-1.44), 0.37 |
| -/- | 57 (37.0) | 66 (42.9) | 0.62 (0.31-1.23), 0.17 |
| **$P_{trend}$** | | | 0.13 |
| Non-null | 97(62.9) | 88(57.1) | Ref |
| Null | 57(37.0) | 66(42.9) | 0.77(0.47-1.25) 0.29 |
| | | | |
| **GSTT1** [a] | | | |
| +/+ | 42 (27.3) | 30 (19.5) | Ref |
| +/- | 90 (58.4) | 82 (53.2) | 0.66 (0.36-1.22), 0.19 |
| -/- | 22 (14.3) | 42 (27.3) | **0.32 (0.15-0.71), 0.005** |
| **$P_{trend}$** | | | **0.006** |
| Present | 132(85.7) | 112(72.7) | Ref |
| Null | 22(14.3) | 42(27.3) | **0.65(0.47-0.91) 0.01** |
| | | | |
| **Smoking status** [a] | | | |
| Non-smokers | 49 (31.8) | 83 (53.9) | Ref |
| Smokers | 105 (68.2) | 71 (46.1) | **3.03 (1.73-5.31), <0.001** |
| **Tobacco chewing** | | | |
| Non-chewers | 74 (48.1) | 82 (53.2) | Ref |
| Chewers | 80 (51.9) | 72 (46.8) | 0.98 (0.58-1.66) 0.95 |
| **Betel quid chewing** [a] | | | |
| Non chewers | 35 (22.7) | 66 (42.9) | Ref |
| Chewers | 119 (77.3) | 88 (57.1) | **2.39 (1.38-4.16), 0.002** |
| **Alcohol consumption** | | | |
| Non-alcoholic | 114 (74.0) | 109 (70.8) | Ref |
| Alcoholic | 40 (26.0) | 45 (29.2) | 0.67 (0.38-1.19), 0.175 |

[a] $\chi^2$ significant; $p<0.05$
Bold number indicate significant p value < 0.05

**Gene-gene interaction**

To elucidate gene-gene interactions associated with lung cancer, we investigated the role of these polymorphisms in combination (Table 5.3). Interaction of all three *GSTM1* genotype with one or no copy of *GSTT1* gene conferred reduced risk to lung cancer. However most of these interactions were statistically insignificant. Only significant reduced lung cancer risk was observed for individuals with the combined *GSTM1* and *GSTT1* null genotype (OR=0.23; 95%CI=0.06-0.80; p=0.02) which becomes insignificant after Bonferroni correction ($P_{\text{Bonferroni correction}}$=0.12).

**Table 5.3: Joint effect of *GSTM1* and *GSTT1* genotypes and lung cancer risk**

| GSTM1 | GSTT1 | Cases n (%) | Controls n (%) | OR (95% C.I.), p value |
|---|---|---|---|---|
| +/+ | +/+ | 7(21.8) | 4(18.1) | 1.0 |
| | +/- | 17(53.1) | 12(54.5) | 0.99 (0.20-4.92), 0.99 |
| | -/- | 8(25.0) | 6(27.2) | 0.70 (0.11-4.39), 0.70 |
| +/- | +/+ | 21(32.3) | 16(24.2) | 1.0 |
| | +/- | 37(56.9) | 37(56.0) | 0.52 (0.21-1.27), 0.15 |
| | -/- | 7(10.7) | 13(19.6) | 0.31 (0.09-1.09), 0.07 |
| -/- | +/+ | 14(24.5) | 10(15.1) | 1.0 |
| | +/- | 36(63.1) | 33(50.0) | 0.67 (0.24-1.84), 0.44 |
| | -/- | 7(12.2) | 23(34.8) | **0.23 (0.06-0.80), 0.02*** |

* Bold Number indicates significant p value < 0.05

## Gene-environment interaction

To evaluate the potential modifying effect of *GSTM1* and *GSTT1* genes on risk factors, stratified analysis was performed. Genotype distribution and association of *GSTM1* and *GSTT1* in cases and controls positive for smoking, tobacco chewing, betel quid chewing and alcohol consumption are given in Table 5.4. Interaction of risk factors with *GSTM1* and *GSTT1* genes imparted reduced risk with deletion in functional allele (+/- and -/-) compared to individuals with the presence of both allele (+/+). Smokers carrying *GSTT1* null genotype showed significantly reduced risk (OR=0.30; 95%CI=0.10-0.91; p=0.03). Moreover, after stratifying the data further between exclusive smoker and betel nut chewers, protective effect of *GSTT1* null genotype was more pronounced in smokers only (OR=0.030; 95%CI=0.001-0.78; p=0.03, $P_{trend}$=0.006) (Table 5.5). Similar interaction was observed for alcohol, individuals with *GSTT1* null genotype showed significantly reduced risk (OR=0.04; 95%CI=0.003-0.61; p=0.02). Interaction of risk habits with *GSTM1* and *GSTT1* genotypes showed carcinogen specificity. It was interesting to note that even though most of the results were insignificant, yet interactions of risk habits with null genotype almost always yielded more protective effects than interactions with one copy number genotypes for both *GSTM1* and *GSTT1*.

**Table 5.4: Effect modification of *GSTM1* and *GSTT1* genotypes and environmental risk factors on lung cancer risk**

| | Smoking | | Tobacco | | Betel quid | | Alcohol | |
|---|---|---|---|---|---|---|---|---|
| **Genotypes** | **Cases/Controls n (%)** | **[a] OR (95%CI) p-value** | **Cases/Controls n (%)** | **OR (95%CI) p-value** | **Cases/Controls n (%)** | **OR (95%CI) p-value** | **Cases/Controls n (%)** | **OR (95%CI) p-value** |
| **GSTM1** | | | | | | | | |
| +/+ | 22(21.0)/12(16.9) | 1.0 | 14(17.5)/7(9.7) | 1.0 | 23(19.3)/14(15.9) | 1.0 | 7(17.5)/9(20.0) | 1.0 |
| +/- | 46(43.8)/28(39.4) | 0.84(0.34-2.09), 0.71 | 36(45.0)/34(47.2) | 0.52(0.13-2.11), 0.36 | 51(42.9)/39(44.3) | 0.81(0.32-2.04), 0.67 | 18(45.0)/18(40.0) | 2.31(0.48-11.06), 0.29 |
| -/- | 37(35.2)/31(43.7) | 0.75(0.30-1.87), 0.54 | 30(37.5)/31(43.1) | 0.44(0.11-1.75), 0.24 | 45(37.8)/35(39.8) | 0.76(0.27-1.98), 0.63 | 15(37.5)/18(40.0) | 1.58(0.32-7.76), 0.56 |
| **GSTT1** | | | | | | | | |
| +/+ | 28(26.7)/11(15.5) | 1.0 | 19(23.8)/15(20.8) | 1.0 | 25(21)/19(21.6) | 1.0 | 10(25.0)/5(11.1) | 1.0 |
| +/- | 62(59.0)/41(57.7) | 0.59(0.24-1.43), 0.25 | 49(61.3)/35(48.6) | 1.37(0.53-3.55), 0.51 | 76(63.9)/51(58.0) | 1.18(0.54-2.57), 0.66 | 27(67.5)/28(62.2) | 0.75(0.17-3.29), 0.71 |
| -/- | 15(14.3)/19(26.8) | **0.30(0.10-0.91), 0.03**[b] | 12(15.0)/22(30.6) | 0.64(0.20-2.05), 0.46 | 18(15.1)/18(20.5) | 0.62(0.22-1.79), 0.38 | 3(7.5)/12(26.7) | **0.04(0.003-0.61), 0.02**[b] |

[a] OR(95% C.I.),pvalue: Odd Ratio, 95% Confidence Interval

[b] Bold Number indicates significant p value < 0.05

**Table 5.5: Effect modification of *GSTM1* and *GSTT1* genotypes in only smokers and betel chewers**

| Gene | Group | Betel(74) | | | Smoker(43) | | |
|---|---|---|---|---|---|---|---|
| | | ratio | [a]OR (95% C.I.)p | [b]OR (95% C.I.)p | ratio | [a]OR (95% C.I.)p | [b]OR (95% C.I.)p |
| **GSTM1** | +/+ | 7/6 | Ref | Ref | 6/4 | | |
| | +/- | 11/18 | 0.81(0.04-15.36)0.89 | 1.08(0.003-29.9)0.96 | 6/7 | 1.08(0.18-6.40)0.92 | 1.05(0.14-7.31)0.98 |
| | -/- | 15/17 | 6.88(0.44-106.9)016 | 9.37(0.36-243.8)0.17 | 7/13 | 0.16(0.001-1.80)0.13 | 0.07(0.004-1.23)0.07 |
| $P_{trend}$ | | | 0.88 | | | | 0.19 |
| | | | | | | | |
| **GSTT1** | | | | | | | |
| | +/+ | 6/10 | Ref | Ref | 9/2 | Ref | Ref |
| | +/- | 21/23 | 1.84(0.40-8.47)0.43 | 1.70(0.36-7.89)0.49 | 7/13 | 0.22(0.02-2.20)0.19 | 0.14(0.001-1.93)0.14 |
| | -/- | 6/8 | 2.71(0.14-49.68)0.50 | 3.24(0.09-116.2)0.52 | 3/9 | 0.05(0.003-0.97)0.004 | **0.03(0.001-0.78)0.03** |
| $P_{trend}$ | | | | 0.74 | | | **0.006** |

[a]OR (95% C.I.),pvalue: Odd Ratio, 95% Confidence Interval obtained from univariate conditional logistic regression models

[b]OR (95% C.I.),pvalue: Odd Ratio obtained from multivariate conditional logistic regression models (Adjusted for alcohol and tobacco chewing)

## DISCUSSION

Worldwide numerous studies have investigated the association between *GSTT1* and *GSTM1* polymorphism but with conflicting results. Although findings from India showed no significant association of *GSTM1* and *GSTT1* with lung cancer however some suggest a possible interaction with smoking. Sreeja et. al., [2005] showed significant risk associated with *GSTT1* null polymorphism. Similarly, Kumar et. al., [2009] have also found marginally significant risk associated with *GSTT1* null genotype. However, these studies have investigated the risk for lung cancer in individuals with null genotype (no allele) compared with a combined group of individuals with either one or two functional alleles. Thus underestimating the risk without accounting the effect of gene dosage of allele.

In the present study, there was no overall effect of *GSTM1* polymorphism on lung cancer. For *GSTT1* gene, the risk of lung cancer significantly decreased with the deletion of the functional allele i.e from single copy of the gene (0.66) to the null genotype (0.32) compared to the two copy number of gene. However, the defined classical functions of *GSTM1* and *GSTT1* do not support protective roles for the null genotypes but have been precedent in studies on lung cancer particularly from outside India [Reed et. al., 1990, Pemble et. al., 1994, Risch et. al., 2001, Lewis et. al., 2002, Stücker et. al., 2002, Alexandrie et. al., 2004, Sørensen et. al., 2007, López-Cima et. al., 2012]. In one study, Risch et. al., [2001] reported that *GSTT1* null genotype was underrepresented among squamous cell carcinomas. Similarly in two other studies, *GSTM1* null genotype was associated with reduced risk of lung cancer particularly in younger patients of 50-60 years and in squamous cell carcinomas [Lewis et. al., 2002, Sørensen et. al., 2007]. Some other studies have also found moderately decreased although non-significant, risk associated with *GSTT1* null genotype [Stücker et. al., 2002, Alexandrie et. al., 2004, López-Cima et. al., 2012].

In the light of the above results it is easy to speculate a dual role of the GSTs depending upon the conditions of stress. Studies have explained this duality on the basis of metabolite toxicity [Pemble et. al., 1994] and even population genetics [Roodi et. al., 2004]. Human *GSTT1* is an orthologue of the rat GST subunit 5 which has been shown to be mutagenic

in Salmonella typhimurium [Pemble et. al., 1994]. Activation of dihalomethanes resulting in the formation of the glutathione (GSH) metabolic intermediates that might account for this mutagenicity and is believed to be responsible for the carcinogenicity of dichloromethane in mice [Pemble et. al., 1994]. Postulations on population genetics by Roodi et. al., [2004] states that high frequency of *GSTM1* (+/+) genotype in African-American population compared to the white Caucasian in their study is due to a positive selection of the 'beneficial' null genotype during human migration from Africa which results in a high absence of the gene in Caucasians. Besides the above possibilities, it is also important to look for linkage disequilibrium of the GST genes, with other metabolic gene polymorphisms, which could be population specific and might in turn act as ethnic specific risk modifiers.

When analyzing joint effects of the two GST genes, combination of null genotypes of both the genes imparted statistically significant reduced risk though insignificant after Bonferroni correction. This suggests that an increased glutathione conjugation by the present alleles of both genes imparts increased risk to cancer. Combined conjugation activities by GSTs deplete the level of glutathione in the cell impairing xenobiotic defense and thereby exposing it to oxidative damage and induced mutagenesis [Reed et. al., 1990]. The formation of glutathione conjugates generally cause the electrophiles to be less toxic and readily excreted. However this conjugation might also act as transporter molecule by releasing reversibly bound electrophilic compounds.

Tobacco smoking was a strong risk factor in the study. Tobacco is consumed both in smoking and smokeless forms. In India, tobacco is smoked as cigarettes or in the form of bidi, a native cigarette-like stick that consists of tobacco wrapped in a tendu or temburni leaf. Tobacco smoke comprises nearly 60 carcinogenic compounds whereas its unburned form contains 16 identified carcinogens [Hecht et. al., 2003]. Among these, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and polycyclic aromatic hydrocarbons (PAH) are considered to be the most important causative agents for the development of lung cancer. PAH require metabolic activation and subsequent binding to DNA (forming bulky ''PAH-DNA adducts'') to exert their carcinogenic action [Thakker et. al., 1985]. Similar activation of one of the N-nitrosamines, 4-

methylnitrosamino-1-3-pyridyl-1-butanone (NNK), by the P450 system produces metabolites that form methyl and pyridyloxobutyl DNA adducts. Detoxification of these toxic metabolites occur via the action of multiple Phase II enzymes, most notably the glutathione-s-transferases.

In the present study, interaction of smoking with *GSTT1* +/- and -/- genotypes compared to the +/+ genotype were protective in nature indicating a genetic modulation of the risk imparted by smoking. Similar result was reflected in a meta-analysis by Raimondi et. al., [2006], where a negative trend of the odds ratios for *GSTT1* null allele was observed with increasing amount of lifetime smoking for both Caucasians and Asians subjects. Further, studies have indicated adverse association between smoking and lung cancer among individuals with *GSTT1* null genotype particularly in non-smokers. Alexandrie et. al., [2004] found that *GSTT1* null genotype was associated with decreased risk for lung cancer in heavy smokers (OR=0.36;95%CI=0.13-0.99;p=0.004). Although not statistically significant, Wenzlaff et. al., [2005] also found that never smokers with *GSTT1* null genotype with no household environmental tobacco smoke were at one-third the risk of lung cancer compared with *GSTT1* present genotype. Further a possible protective effect of being *GSTT1* null in non-smoker has also been reported by Hou et. al., [2001].

The discrepancy in results could be explained by different ethnic population, differences in categorization of smokers and different habit of dietary compounds. As there are several dietary compounds, particularly intake of crucifereous vegetables that need to be controlled in order to fully elucidate true gene-environment interactions related to lung caner risk. London et. al., [2000] found that individuals with detectable level of Isothiocyanate were at reduced risk of lung cancer with the null genotype of both *GSTM1* and *GSTT1*. Isothiocyanates (ITC), found in cruciferous vegetables, are substrates for GSTs and are associated with reduced cancer risk. The present study lack information on both dietary status and the pack years of the smokers.

There might be some limitations to this study. The sample size of our study was relatively small, however based on the evidences (OR) provided by our research group on association between GSTs with lung cancer [Yadav et. al., 2010], the minimum sample size

determined was 144 at 5% level of significance and 80% power. *GSTT1* genotypes showed deviation from HWE in lung cancer cases. After ruling out false positive associations and genotyping errors perhaps population stratification, could be a reason for this deviation. However, the cases were incident, and thus, the data do not show report or recall bias. Also case-control matching was done in reference to age, gender, and ethnicity, thereby controlling for any confounding effect accounted by these variables. Estimation of interactive OR in this study in some cases yielded small subgroup sizes which limits the reliability of estimating gene-environment effects. Thus these results should be considered empirical observations for further studies on larger number of samples. In summary, trimodular genotypes of *GSTM1* and *GSTT1* were determined and gene dosage effect was observed with *GSTT1* copy number. The direction of our result indicates that null genotype of *GSTT1* may be associated with a reduced risk of lung cancer risk. Furthermore protective effect of *GSTT1* was strongly associated with smokers only.

Our results were in contrast with previous reports on lung cancer by Lam et. al., [2009] and Sorensen et. al., [2007] which observed no significant association of hemizygous and homozygous genotypes of *GSTM1* and *GSTT1* when compared with homozygous wild type genotype. This might be due to the difference in sample size, however in contrast to many of these studies, the homogeneity of our population from an ethnically isolated North-East part of India allowed the statistical detection of the small inherited variations in metabolism. Thus, the present study emphasizes that ethnicity and carcinogen exposure along with trimodal distribution of GST enzymes can be a major determinant of risk of lung cancer. These differences might find implications in drug metabolism and clinical outcomes in the studied population.

# Chapter 6

# Investigation on the role of p53 codon 72 polymorphism and interactions with tobacco, betel quid and alcohol in susceptibility to lung cancer in high risk population from North East India

## *Introduction*

The human T*P53* tumor suppressor gene plays a central role in many cellular processes, regulating cell growth, DNA maintenance and apoptosis. It is an important component of DNA repair machinery in response to DNA damage induced by radiation or adduct formation. This might explain the occurrence of the *p53* gene mutation and alteration in about 50% of all cancers, particularly tobacco related cancers. Studies have shown a relationship between tobacco smoke exposures, carcinogen-DNA adduct formation, tumor specific mutation of T*P53* gene and cancer risk. *p53* gene plays a significant role in the regulation of cellular response to benzo[a]pyrene, one of the most important polycyclic aromatic hydrocarbon (PAH) compounds of tobacco smoke [Xiao and Singh, 2007]. Moreover, recent studies have indicated that there is a strong coincidence in mutational hotspots and sites of preferential formation of PAH adducts along the *p53* gene in lung, laryngeal and head and neck cancers. Besides mutation, polymorphisms in T*P53* gene have also been implicated in the process of tobacco carcinogenesis.

Numerous polymorphism in the wild type *p53* have been reported both in coding and non coding regions [Pietsch et. al., 2006]. Out of the five polymorphisms described in the coding region, polymorphisms in codon 47 and 72 in exon 4 are functionally well characterized. More common of the two, codon 72 polymorphism is a single base substitution of cytosine for guanine, leading to arginine (A72) being replaced by proline (P72) [Pietsch et. al., 2006] that has been reported to be associated with the risk of several cancers [Papadakis et. al., 2000, Tandle et. al., 2001, Wu et. al., 2004, Mitra et. al., 2005, Rogounovitch et. al., 2006]. However, the results are conflicting with Pro/Pro genotype showing association with lung cancer [Kawajiri et. al., 1993, Jin et. al., 1995, Birgander et. al., 1996, Wang et. al., 1999a] breast cancer [Sjalander et. al., 1996, Papadakis et. al., 2000] and gastric cancer [Hiyama et. al., 2002] whereas Arg/Arg genotype being more prevalent in cervical cancer [Storey et. al., 1998, Dokianakis and

Spandidos 2000]. However, no association between either genotype and cancer risk have also been reported for head and neck [Hamel et. al., 2000] and cervical cancer [Minaguchi et. al., 1998]. Literature available from India is limited and inconsistent. Two different studies have reported both Arg/Arg and Pro/Pro genotypes to be associated with risk of lung cancer [Jain et. al., 2005, Sreeja et. al., 2008].

Studies on codon 72 polymorphism have revealed striking ethnic differences [Själander et. al., 1995]. Beckman et. al., 1994 have demonstrated that frequency of *p53* variant allele varies with latitude, increasing in a linear trend as populations near the equator. Thus ethnicity might be related to allelic distribution of the gene and its determinacy in disease involvement; however some studies do refute the ethnicity-risk confounding relationship [Fan et. al., 2000]. North-eastern (NE) part of India, due to its unique, strategic geographic location and the presence of linguistically, culturally and demographically diverse populations is a hotspot for population genetics. Lung cancer is one of the ten leading sites of cancer in NE India with the highest age adjusted rate (AAR) in Mizoram (24.85 in males and 24.72 in females). The area also reports tobacco use in variety of ways of chewing and smoking that are different from the rest of India. Unlike the Western nations or the urban India, use of tobacco or alcohol in crude forms is more prevalent.

High risk to cancer can be an outcome of either environmental and genetic risk factors or a complex interplay of both. Recent literatures have reported *p53* allelic polymorphisms to be possible predisposing factors for tumor development. Lack of data on *p53* codon 72 polymorphism and high incidence of cancers in the north eastern region of India prompted us to explore and evaluate any relevance of this polymorphism in this ethnic population. We carried out a case control study on three commonly occurring cancers i.e. lung, gastric and oral cancer, in north eastern part of India. The role of *p53* codon 72 polymorphism and its interaction with tobacco, betel quid and alcohol use was also analyzed.

## EXPERIMENTAL METHODS

### Materials

Agarose, Tris base, EDTA, NaCl, SDS, Triton X-100 and other fine chemicals were purchased from Sigma Chemicals, USA. Taq polymerase, dNTPs, MgCl2, was obtained from Invitrogen and MBI fermentas USA. Oligos were synthesized by Microsynth, Switzerland. RNA later, DNA and RNA extraction kit were purchased from Qiagen Sciences, USA and Himedia, India.

### Chemicals used

LYSIS BUFFER I: 30mm Tris-Hcl (Ph-8), 5mm EDTA, 50 Mm Nacl; LYSIS BUFFER II: 75mm Nacl, 2mm EDTA (Ph-8); SDS STOCK: 20 gm of SDS dissolved in 80 ml of TDW at $65^0$C. Make up volume up to 100 ml ; PROTEINASE K: 10 mg dissolved in 1 ml of TDW:1% ; AGAROSE: 1gm of agarose dissolved 1% TAE buffer.

### Patient recruitment and sample collection

The study was conducted in 161 histopathologically diagnosed lung cancer cases registered at Dr. Bhubaneswar Borooah Cancer Institute, Guwahati, Civil Hospital, Aizawl, and Sir Thutob Namgyal Memorial Hospital, Gangtok, the collaborating centers in north east India.

### Inclusion criteria

Incident cases during the period of December 2006 to 2008 and willing to participate in the study were included. 282 voluntary, age (±5 years) and sex matched individuals were selected from the healthy relatives who accompanied cancer patients. This provided a readily available and cooperative source of controls from the same socio-economic background as the cases reducing confounding biases. Patients with only lung as their primary site of cancer were included. Final selected controls were included on the basis of no history of any obvious disease and those not taking any medication at the time of recruitment

**Exclusion criteria**

Any subject with history of familial malignancy or pulmonary infectious disease was excluded both from case and control. Patients unwilling or too ill to participate in the study were excluded. Patients who had taken any form of treatment earlier (Secondary cases) were also excluded from the study**.**

**Patient details**

All subjects provided written informed consent for participation in this research which was done under a protocol approved by the institutional ethics committee of Regional Medical Research Centre, North East Region (Indian Council of Medical Research) and participating institutes. Information regarding smoking, usage of tobacco, betel quid and alcohol were obtained from subjects in a standard questionnaire used for all the centers. Smokers, chewers and drinkers were classified into two categories ever and never. For smoking, an individual who had never smoked or smoked less than 100 cigarettes in their lifetime and were not smoking at the time of reporting was considered never smoker or non-smokers. Ever smokers or smokers category included current smokers, and those who had quit within <1 year of reporting [Carlsten et. al., 2008]. As our collaborating centers were public hospitals a large majority of subjects belonged to lower to middle socio-economic background. Demographic data and characteristics such as age, sex, smoking habit, usage of tobacco, betel quid and alcohol, were obtained from subjects in a standard questionnaire used for all the centers, in an in-person interview by a trained data collector. A majority of cases and controls were literate with full primary schooling and some upto the college level. The occupational history of the study participants revealed that most of them were farm laborers or engaged in petty jobs and the nature of such jobs did not exposed them to any occupational hazards. Any history of past or present illness was enquired or if undergoing any medication at the time of enrolment.

**Collection of blood samples:** As described in Chapter 4

**DNA extraction**: As described in Chapter 4

**Quantification of Genomic DNA:** As described in Chapter 4

**Agarose Gel Electrophoresis of Extracted DNA Samples:** As described in Chapter 4

### Genotyping of *p53*

Polymorphism in *p53* was genotyped using PCR-RFLP (polymerase chain reaction-restriction fragment length polymorphism) method. Standard PCR were performed on PTC-200 (MJ Research, USA). The PCR reaction were performed in a volume of 20µl with a final concentration of 1X PCR Buffer (MBI Fermentas), 1.5mM $Mgcl_2$, 200 µM dNTPs, 0.75 unit of Taq polymerase and 500ng of genomic DNA. Negative controls were included in all PCR-runs to prevent misjudging following contamination of samples. PCR amplification consisted of 40 cycles of denaturation at $94^0$C for 45s; annealing at $60^0$C for 45s; $72^0$C for 45s followed by a final extension at $72^0$C for 10 min. PCR products were loaded on 2.5% agarose gel and subjected to gel electrophoresis in 0.5X TBE buffer, stained with ethidium bromide and visualized under UV. Detail of single-nucleotide polymorphism (SNP) is summarized in Table 6.1. Sequence of the primer and their annealing temperatures are given in the Table 6.2.

**Table 6.1: Detail of the single-nucleotide polymorphisms (SNPs) selected for the study**

| Gene | Chr[a] | SNP | Loc | Polymorphism | |
|---|---|---|---|---|---|
| | | | | Nucleotide | Codon |
| P53 | 17p13.1 | rs1042522 | Exon4 | 215C>G | Arg72Pro |

[a]Chromosomal position is based on NCBI Build.

**Table 6.2: Sequence of primers used in the study**

| Gene | Primer sequence | $T^0C$ | PCR (bp) |
|---|---|---|---|
| P53 | 5'-TTGCCGTCCCAAGCAATGGATGA-3'<br>5'-TCTGGGAAGGGACAGAAGATGAC-3' | 60 | 199 |

### RFLP analysis of *p53*

Restriction analysis was performed by digesting the PCR products with 5 units of restriction enzyme BstUI (New England Biolabs, Beverly, MA) at 60°C for 16hrs. (Table 6.3). Heat inactivation of enzyme was done at $80^0$C for 20 minutes after completion of incubation with enzyme. Details of restriction enzyme BstUI is summarized in Table 6.4 and the alleles of each specific sample can be observed as a specific band pattern on the gel (Figure 6.1). Genotyping of 10% of the randomly selected cases and controls were confirmed by sequencing. (Figure 6.2). No discrepancies were observed.

**Table 6.3: Standard protocol used for the RFLP experiment**

| Components | Stock conc. | Working conc. | 1 reaction (µl) |
|---|---|---|---|
| Water | | | 3 |
| Buffer | 10X | 1 | 1.5 |
| Enzyme (BstUI) | 10Units/µl | 5Units | 0.5 |
| PCR product | | | 10.0 |

**Table 6.4: Detail of the RFLP enzymes used for each polymorphism**

| Gene | Enzyme | Site | Incubation Condition | PCR | RFLP product (bp) | |
|---|---|---|---|---|---|---|
| | | | | | Homo wild | Homo variant |
| P53 | BstUI | 5'-CG^CG-3'<br>3'-GC^GC-5' | $37^0$C overnight | 199 | 113+86 | 199 |

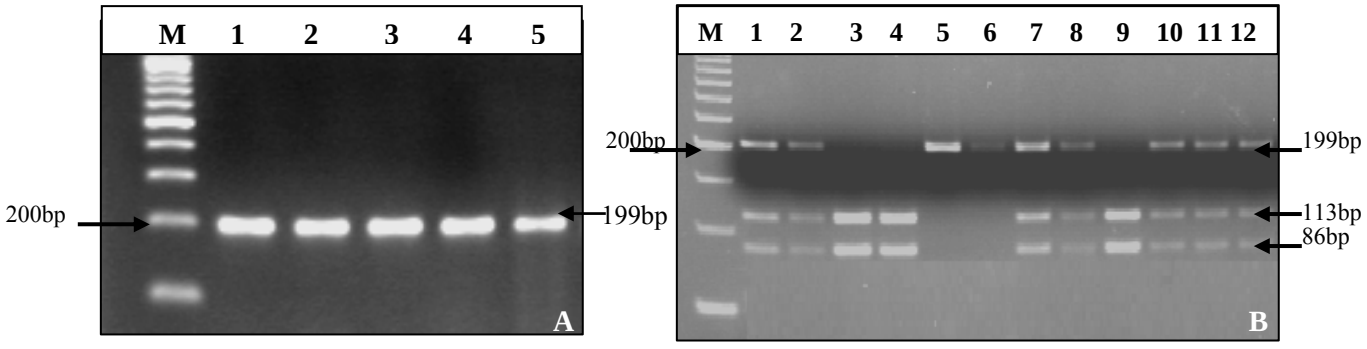



**Figure 6.1: EtBr stained agarose gel electrophoresis for RFLP analysis of *p53*codon72 polymorphism**. A- showing PCR amplification of exon 4 of *p53* gene. B-RFLP of *p53* PCR product; Lane 5 and 6-sample with 199 bp represent homozygous Pro/Pro allele (GG genotype); Lane 1, 2, 7, 8, 10, 11 and 12 - samples with all three bands (199bp, 113bp, 86bp) represent heterozygous Arg/Pro allele (GC genotype); Lane 3, 4 and 9 - samples with two band (113 bp, 86 bp) represent homozygous Arg/Arg allele (CC genotype) . M: A-100bp ladder B- 50bp ladder

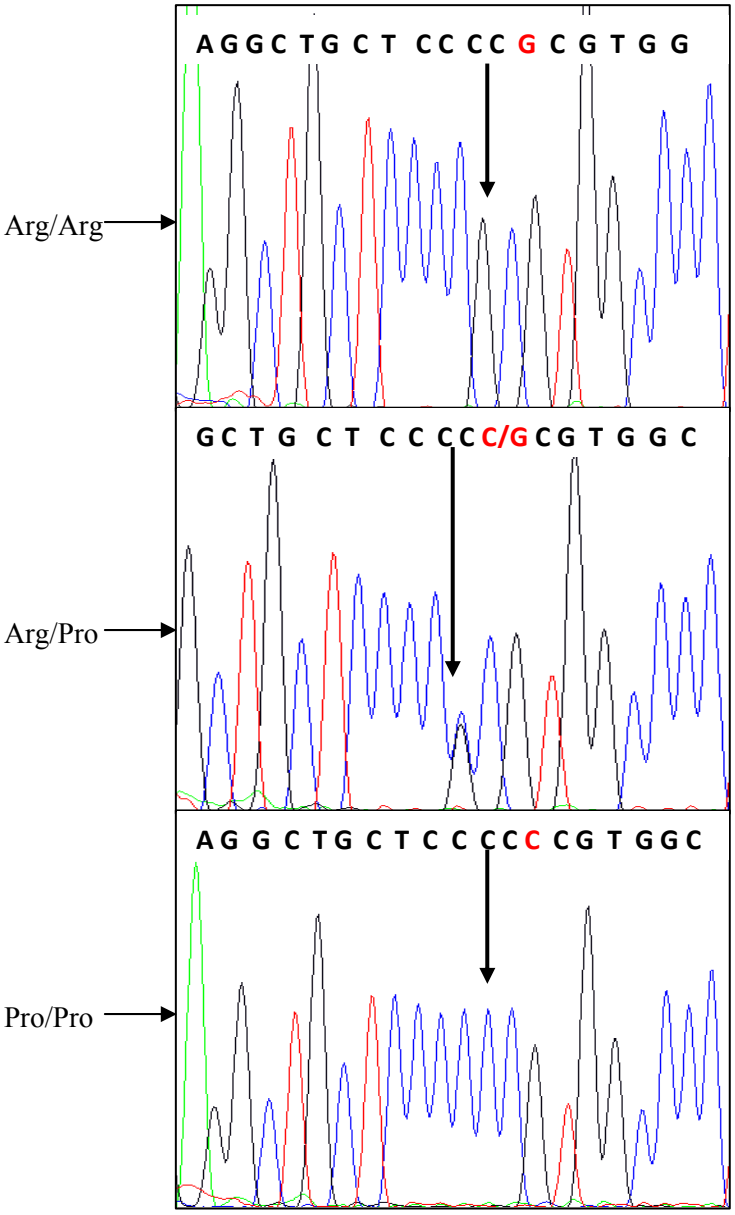



**Figure 6.2: Representative genotypes p53 codon 72 (Arg>Pro) polymorphism by sequencing**

### Statistical analysis

Cases were individually matched with 274 control samples on the basis of age (±5 years), sex and ethnicity, in a case control ratio of approximately 1:2. The association of *p53* codon 72 genotypes with lung cancer was evaluated by multivariable conditional logistic regression in dominant (Pro/Pro and Arg/Pro versus Arg/Arg) and recessive (Pro/Pro versus Arg/Arg and Arg/Pro) genetic models. Armitage trend test was performed to calculate P for trends in additive model (Pro/Pro, Arg/Pro, versus Arg/Arg). The association of tobacco smoking, tobacco chewing, betel quid chewing, alcohol intake with disease outcome was assessed by $\chi 2$/Fisher's exact test. Estimates of risk to cancer, imparted by *p53* genotypes and other covariates as tobacco smoking, chewing, betel quid chewing and alcohol was determined by deriving the odds ratio (ORs) and corresponding 95% confidence intervals (95% CIs) using univariate and multivariable conditional logistic regression models. To evaluate potential modifying effects of *p53* genotypes on the association between various co-variates and cancer risk, cases and controls were tabulated according to the joint distribution of these factors in various possible combinations. Tests for studying interactions were performed for each *p53* genotype with all the considered covariates. OR was the adjusted odds ratio using conditional multivariate logistic regression model. For all the tests a two sided $p<0.05$ was considered statistically significant. The data analysis was performed on STATA 8.0 software.

## RESULT

### Characteristics of study subjects

The distributions of demographic characteristics and potential risk factors are summarized in Table 6.5. Higher percentage of males was seen in cases as well as in controls. The frequency distribution of males and females were 77.1% and 22.9% in cases and 76.2% and 23.85 in controls respectively. Mean age of cases and controls was 60.24±10.77 and 53.21±13.37 respectively. The distribution of *p53* genotype between cases and controls is shown in Table 6.5. The distribution of SNPs both cases and control was in agreement with HWE ($p>0.05$). No significant increase in risk of lung cancer was observed in a univariate or in a multivariable analysis for dominant and recessive models of inheritance Table 6.6. P values obtained from Armitage test for additive model was not significant (data not shown).

**Table 6.5: Distribution of demographic variables and genotypes between cancer cases and controls**

| Variables | Categories | Cases | Controls | OR (95% C.I.) | p value |
|---|---|---|---|---|---|
| | | n (%) | n (%) | | |
| **Sex** | Male | 120 (74.5) | 202 (73.7) | | |
| | Female | 41 (25.5) | 72 (26.3) | | |
| **Age Mean** | | 60.24±10.77 | 53.21±13.37 | | |
| **Smoking status** | Non smokers | 51 (31.7) | 135 (49.3) | 1.0 | |
| | Smokers | 110 (68.3)* | 139 (50.7) | **1.88(1.11-3.19)** | **p=0.018** |
| **Tobacco chewing** | Non chewers | 73 (45.3) | 140 (51.1) | 1.0 | |
| | Chewers | 88 (54.7) | 134 (48.9) | 1.04(0.65-1.67) | p=0.85 |
| **Betel quid chewing** | Non chewers | 28 (17.4) | 105 (38.5) | 1.0 | |
| | Chewers | 133 (82.6)* | 168 (61.5) | **3.54(2.01-6.25)** | **p<0.001** |
| **Alcohol consumption** | Non alcoholic | 118 (73.3) | 196 (71.5) | 1.0 | |
| | Alcoholic | 43 (26.7) | 78 (28.5) | 1.02(0.59-1.75) | p=0.93 |
| **p53 genotypes** | | | | | |
| | Arg/Arg | 38 (23.6) | 64 (23.4) | 1.0 | |
| | Arg/Pro | 86 (53.4) | 141 (51.5) | 1.11(0.63-1.95) | p=0.71 |
| | Pro/Pro | 37 (23.0) | 69 (25.2) | 1.06(0.55-2.04) | p=0.83 |

* $\chi^2$ significant; p<0.05; Bold number indicate significant p value < 0.05

**Table 6.6: Estimate of the effect of the *p53* codon 72 polymorphism on cancer risk modeled with logistic regression**

| Association Model | | Case/Control n (%) | OR (95% CI) |
|---|---|---|---|
| Dominant | Arg/Arg | 38(23.6)/64(23.4) | 1.00 |
| | Arg/Prp and Pro/Pro | 123(76.4)/210(76.6) | 1.09 (0.64-1.86),p=0.73 |
| Recessive | Arg/Arg and Arg/Pro | 124(77.0)/205(74.8) | 1.00 |
| | Pro/Pro | 37(23.0)/69(25.2) | 0.99 (0.58-1.70),p=0.99 |

## Association of genetic factors with lung cancer risk

Distribution and association of risk factors and genotypes in lung cancer cases and controls is given in Table 6.5. Pro/Pro genotype was more frequent in controls (25.2%) than cases (23.0%), however this genotypic distributions was not significant ($\chi2$= 0.27, p=0.87). Risk of cancer was higher for individuals carrying Arg/Pro than Pro/Pro genotypes but the results were not statistically significant (OR= 1.11, 95% CI=0.63-1.95; p=0.71 and OR=1.06, 95% CI=0.55-2.04; p=0.83 for Arg/Pro and Pro/Pro respectively). Frequency of smokers was higher in cases (68.3%) than controls (50.7%) ($\chi2$= 12.824, p<0.0001) and smoking conferred a significant risk (OR= 1.88, 95% CI=1.11-3.19, p=0.018). Distribution of betel quid chewers differed significantly between cases and controls ($\chi2$= 21.15, p<0.0001) and conferred a significant risk of developing lung cancer (OR= 3.54, 95% CI=2.01-6.25, p>0.001). Distribution of tobacco chewers was higher among cases than controls and vice versa for alcohol users (54.7% vs 48.9% and 26.7% vs 28.5%).

## Gene-Environment interactions

Interaction combinations between betel quid chewing and *p53* gene variants were highly significant (Table 6.7). Increased lung cancer risk of upto six fold was observed for all the three genotypes (OR=5.90, 95% CI=1.67-20.81;p=0.006, OR=5.44, 95% CI=1.67-17.75; p=0.005, OR=5.84, 95% CI=1.70-19.97;p=0.005 for Arg/Arg, Arg/Pro and Pro/Pro respectively). Interaction of Arg/Arg genotype with tobacco chewing and alcohol use conferred 64% and 74% less chance of developing the cancer (OR=0.36, 95% CI=0.13-0.97;p=0.04 and OR=0.26, 95% CI=0.07-0.95;p=0.042 respectively). No interaction between smoking and *p53* genotypes rendered significant risk.

**Table 6.7: Gene-Environment interactions odds ratio for *p53* codon 72 genotypes and risk habits for cancer patients**

| Variables | Interactions | Case n (%) | Control n (%) | OR (95% C.I.), p value |
|---|---|---|---|---|
| **Smoking [a]** | | | | |
| | Arg/Arg X Non smoker | 14 (8.7) | 32 (11.7) | 1.0 |
| | Arg/Arg X smoker | 24 (14.9) | 32 (11.7) | 1.58(0.58-4.31),p=0.36 |
| | Arg/Pro X Non smoker | 23 (14.3) | 67 (24.5) | 0.93(0.37-2.33),p=0.88 |
| | Arg/Pro X smoker | 63 (39.1) | 74 (27.0) | 1.97(0.84-4.64),p=0.11 |
| | Pro/Pro X Non smoker | 14 (8.7) | 36 (13.1) | 1.00(0.37-2.69),p=0.99 |
| | Pro/Pro X smoker | 23 (14.3) | 33 (12.0) | 1.78(0.67-4.74),p=0.24 |
| **Tobacco chewing [b]** | | | | |
| | Arg/Arg X Non chewer | 24 (14.9) | 25 (9.1) | 1.0 |
| | Arg/Arg X Chewer | 14 (8.7) | 39 (14.2) | **0.36(0.13-0.97),p=0.04** |
| | Arg/Pro X Non chewer | 34 (21.1) | 75 (27.4) | 0.59(0.27-1.32),p=0.20 |
| | Arg/Pro X Chewer | 52 (32.3) | 66 (24.1) | 0.76(0.33-1.74),p=0.53 |
| | Pro/Pro X Non chewer | 15 (9.3) | 40 (14.6) | 0.49(0.19-1.26),p=0.14 |
| | Pro/Pro X Chewer | 22 (13.7) | 29 (10.6) | 0.82(0.32-2.11),p=0.69 |
| **Betel quid chewing [c]** | | | | |
| | Arg/Arg X Non chewer | 4 (2.5) | 25 (9.1) | 1.0 |
| | Arg/Arg X Chewer | 34 (21.1) | 39 (14.2) | **5.90(1.67-20.81),p=0.006** |
| | Arg/Pro X Non chewer | 19 (11.8) | 53 (19.3) | 1.97(0.56-6.83),p=0.285 |
| | Arg/Pro X Chewer | 67 (41.6) | 88 (32.1) | **5.44(1.67-17.75),p=0.005** |
| | Pro/Pro X Non chewer | 5 (3.1) | 28 (10.2) | 1.37(0.30-6.20),p=0.67 |
| | Pro/Pro X Chewer | 32 (19.9) | 41 (15) | **5.84(1.70-19.97),p=0.005** |
| **Alcohol consumption [d]** | | | | |
| | Arg/Arg X Non alcoholic | 33 (20.5) | 41(15.0) | 1.0 |
| | Arg/Arg X Alcoholic | 5 (3.1) | 23 (8.4) | **0.26(0.07-0.95),p=0.042** |
| | Arg/Pro X Non alcoholic | 57 (35.5) | 106 (38.7) | 0.68(0.35-1.33),p=0.26 |
| | Arg/Pro X Alcoholic | 29 (18.0) | 35 (12.8) | 1.28(0.57-2.87),p=0.54 |
| | Pro/Pro X Non alcoholic | 28 (17.4) | 49 (17.9) | 0.86(0.41-1.82),p=0.70 |
| | Pro/Pro X Alcoholic | 9 (5.6) | 20 (7.3) | 0.64(0.22-1.86),p=0.42 |

a : OR adjusted for tobacco chewing, betel quid chewing and alcohol consumption

b : OR adjusted for tobacco smoking, betel quid chewing and alcohol consumption

c : OR adjusted for tobacco smoking, tobacco chewing and alcohol consumption

d : OR adjusted for tobacco smoking, tobacco chewing and betel quid chewing

Bold number indicate significant p value < 0.05

## DISCUSSION

*P53* is a highly conserved gene with only five polymorphisms being known till date in 11 exons, of which polymorphism of codon 72 is the commonest and most characterized single base substitution of Proline (P72) for Arginine (A72) leading to structural changes in the protein [Matlashewski et. al., 1987, Thomas et. al., 1999]. The polymorphism occurs in the proline rich region of *p53*, which plays a vital role in apoptosis and growth suppression functions, thus indicating that these two polymorphic variants differ in their biological properties [Thomas et. al., 1999]. The P72 variant is a stronger inducer of transcription, probably owing to its stronger affinity to bind to transcription factors and the R72 variant is considered to be a better inducer of apoptosis, thus suppressing transformation more efficiently, than the P72 variant [Thomas et. al., 1999].

Observations on association of *p53* codon 72 polymorphism and cancer are reported to be inconsistent in different ethnic and geographical region with allele frequency varying from 0.45 to 0.78 for Arg and 0.22 to 0.55 for Pro (Table 6.8). Fan et. al., 2000 reported risk of lung cancer to be associated with combined variant of Arg/Pro and Pro/Pro genotypes. Another study associated Pro/Pro genotype carrying lung cancer patients with poorer prognosis than those with Arg/Pro genotype [Wang et. al., 1999b].

**Table 6.8: Frequency of p53 Polymorphism in Lung Cancer: Worldwide Scenario**

| Place of study | N (Case/Control) | Allele freq (Arg) (Case/Control) | Allele freq (Pro) (Case/Control) | First Author |
|---|---|---|---|---|
| Texas | 635/635 | *0.78/- | 0.22/- | Wu et al (2002) |
| Spain | 589/582 | 0.73/0.76 | 0.27/0.24 | Fern´andez-Rubioa et al (2008) |
| India (Delhi) | 40/40 | 0.69/0.45 | 0.31/0.55 | Jain et al (2005) |
| Brazil | 200/264 | 0.68/0.66 | 0.32/0.34 | Honma et al (2008) |
| Massachusetts | 482/510 | 0.65/0.67 | 0.35/0.33 | Fan et al (2000) |
| Japan | 191/152 | 0.65/0.60 | 0.35/0.40 | Murata et al (1996) |
| Chile | 111/133 | 0.58/0.63 | 0.42/0.37 | Caceres et al (2009) |
| Taiwan | 186/152 | 0.56/0.56 | 0.44/0.44 | Wang et al (1999) |
| Baltimore-Washington | 78/72 | 0.55/0.51 | 0.45/0.49 | Weston et al (1992) |

In the present study, no significant effect of the polymorphism on susceptibility to lung, cancer was observed. These findings are concordant with some previous reports spread over different ethnic populations. No association between *p53* variants and lung cancer was observed in African Americans and caucasians in United States by Weston et. al., 1992 and in northwestern Mediterranean population by To-Figueras et. al., 1996. Similarly few other studies found no association between the codon 72 polymorphism and lung cancer [Jin et. al., 1995 and Biros et. al., 2001]. This study reports the Arg/Pro heterozygous genotype to confer greater risk to cancer than Pro/Pro genotype. Literature available, report preferential retention of *p53* codon 72 arginine allele in tumors of patients with Arg/Pro heterozygous germline genotype [Papadakis et. al., 2002]. Furthermore, presence of arginine allele at codon 72 in tumor, related with reduced sensitivity to chemotherapy [Bergamaschi et. al., 2003] and decreased survival in heterozygous breast cancer [Bonafe et. al., 2003].

Betel quid was found to be a major risk factor for lung cancer in this study. Betel quid is chewing mixture of whole betel/areca nut wrapped with betel leaves spread with white lime with frequent addition of tobacco. It is known to contain phenolic compounds and alkaloids. In addition nitrosamines are formed in an in vivo reaction of betel arecoline, nitrite and thiocynate [Awang, 1988]. Betel quid chewing is often not a singular issue, but is coupled with smoking [Wen et. al., 2005], a reason which can be linked to association of betel quid with high risk of lung cancer. A recent study from Taiwan [Wen et. al., 2010] found risk to betel quid chewing in oral, lung, liver, pancreas and other cancers and its combination with smoking attributed to 50% of death among chewers. In addition to betel quid chewing tobacco smoking was also found to confer risk to lung cancer in the present study. In addition to betel quid chewing tobacco smoking was found to confer risk for development of lung cancer. Smoking is a well known causative factor for development of lung cancer. Tobacco smoke contains several thousand compounds including over 60 established carcinogens. PAH and nitrosamines belong to the most potent pulmonary carcinogen known. The most prevalent carcinogen in tobacco smoke are aldehydes and other volatile compounds such as benzene and butadiene.

For studying role of gene-environment interaction that might modify susceptibility of cancers, potential interactions of *p53* with known risk factors was analyzed. When analyzed for *p53* interaction with tobacco chewing, Arg/Arg interaction

with tobacco chewing and alcohol presented as protective factor for lung cancer. Significant risk estimates were observed for interaction of betel quid chewing with all the three genotypes of *p53* for lung cancer. Thus the results indicate a tissue specific and carcinogen specific modulation of cancer risk by *p53* gene and were contrary to the classical roles assumed for the genetic variants of the genes. However, association of the mechanistic relationship of variants with cancer might not be so clear and straight forward, genetic characterizations as linkage disequilibrium of the variants with certain known and unknown functional polymorphisms and epigenetic events could exist [Yi and Lee 2006]. Perhaps investigating the mutational status of tumor and its correlations could possibly provide better understanding.

Analyzing on the basis of previous data of the polymorphism in this ethnic population, sample size of the study gave a power of 80% at 5% level of significance at 6% of the absolute precision. A reasonable fraction of the control came from friends and accompanying family members of the cancer patients. Such selection of control would reduce any confounding bias by reducing background variations with cases. The cases were incident and thus the data does not show report or recall bias. Also case control matching was done in reference to age, gender and ethnicity thereby controlling for any confounding effect on account of these variables. Studies on *p53* codon 72 polymorphism available in Indian population have not explored gene-environment interaction; therefore sample size for such an analysis could not be reliably determined and limit our ability to estimate the interaction effects precisely. The results of interaction analysis should therefore be considered empirical observations for further studies on larger number of samples.

The current study indicates that there is no significant relationship between *p53* codon 72 polymorphism and lung cancer in the high risk north eastern population of India. However, our results show betel quid chewing as a major risk factor for lung cancer. These findings suggest that the role of ethnicity and genetic susceptibility might be underplayed when exposure to carcinogens is high. However, a validation of the results will require its replication in a larger sample size. Taking into account other confounding variables such as dietary habits, environment (working environmental exposures, passive smoking etc) and infections (HPV) can give more conclusive perspective.

# Chapter 7

# Gene expression profile of non-small cell lung cancer in high risk population from North East India

# *INTRODUCTION*

Lung Cancer develops as a result of the progressive accumulation of genetic and epigenetic alterations that enables the evolving populations of premalignant cells to reach the biological endpoints of the malignant transformation. These genetic abnormalities include deletion of tumor-suppressor genes (TSGs) or amplification of oncogenes and epigenetic changes in DNA methylation. Some of these genetic changes can aid prognostic efforts and predictions of metastatic risk or response to certain treatments yet information about a single or a limited number of molecular markers generally fail to provide satisfactory results for a clinical diagnosis of lung cancer. New technologies such as analysis of gene-expression profiles on microarrays enable us to perform comprehensive analyses of gene expression in cancer cells, and can reveal detailed phenotypic and biological information about them. Gene expression profiling of lung cancer has been widely proposed as a powerful method to identify biomarkers for diagnosis, predicting invasion and metastasis through the identification of biomarkers. Systematic analysis of expression levels among thousands of genes is a useful approach to identify unknown molecules involved in the pathways of carcinogenesis and these discoveries can indicate targets for development of novel anti-cancer drugs.

Studies have used gene expression profiling to divide lung carcinomas into several subgroups, to discriminate primary cancer from metastases and predict survival. One study reported persistent molecular signatures characteristic of smoking in adenocarcinoma which could differentiate smokers from non-smokers [Landi et. al., 2008]. In another, Hou et. al. [2010] identified a 17 gene signature for histopathological classification of non small cell lung cancer (NSCLC) and improved prediction of clinical outcome. NSCLC and small cell lung carcinoma (SCLC) exhibit distinct but overlapping patterns of genetic alterations [Taniwaki et. al., 2006]. Some of the alteration that have been convincingly shown to promote the pathogenesis/carinogenesis of NSCLC includes amplification of c-myc, activating mutations in the *K-ras, p53 , EGFR* and inactivation of

*p16* [Gazdar et. al., 1994, Graziano et. al., 1999, Niklinska et. al., 2001, Uematsu et. al., 2003]. However, the molecular mechanisms underlying NSCLC carcinogenesis still remain largely unknown.

There are several reports on gene expression profile in NSCLC from western countries. However, data from India and specifically the high risk north eastern population is lacking. The objective of this study was to identify genes differentially expressed in non small cell lung carcinoma patients in high-risk North East Indian population. Gene expression patterns were evaluated in lung cancer tissues compared with matched normal tissues. Gene ontology and pathway analyses of differentially expressed genes were performed to detect deregulated genes involved in different biological and metabolic processes. To validate microarray data, expression levels of selected candidate up-regulated and down-regulated genes was analyzed by quantitative real-time reverse transcriptase–polymerase chain reaction (RT-PCR).

## EXPERIMENTAL METHODS

### Materials

Agarose, Tris base, EDTA, other fine chemicals were purchased from Sigma Chemicals, USA. Platinum Taq polymerase, dNTPs, MgCl2, was obtained from Invitrogen and MBI Fermentas USA. RNA Later from Ambion (Austin, USA) and RNA extraction kit were purchased from Qiagen, (Hilden, Germany). For microarray experiments ExpressArt® Amino Allyl mRNA amplification kit and Human 40 K A OciChip were purchased from (Ocimum Biosolution, Hyderabad, India). TaqMan Universal PCR Master Mix and Assays-on-Demand Gene Expression probes were purchased from Applied Biosystems (USA).

### Patient recruitment and sample collection

Tissue specimens were obtained from patients undergoing computed tomography (CT) - guided biopsy at Dr. Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam and Civil Hospital, Aizawl between March 2008 and March 2010. Routine histopathology analysis was done to confirm the diagnosis. Tumor tissue were collected in RNA Later (Ambion, Austin, USA), snap-frozen in liquid nitrogen and stored at $-70^0$C until processed.

**Inclusion criteria**

Patients with only lung as their primary site of cancer were included and those not undergoing chemotherapy at the time of recruitment.

**Exclusion criteria:** As described in chapter 4

**Patient details**

A total of 35 lung cancer biopsies were collected. Of this 12 tumor biopsies were collected along with matched ‘‘pathologically normal’’ tissues from distant site on the same lobe. All subjects provided informed consent and the study was done under a protocol approved by the institutional ethics committee. Information on demographic characteristics, such as sex, age, smoking habit, usage of tobacco, betel quid and alcohol, were obtained from subjects in a standard questionnaire.

**RNA Isolation**

Total RNA was isolated from frozen tumor and normal tissue using RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocol. The quality and quantity of the RNA samples were determined with an Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany) and Nano-drop ND-1000 Full –spectrum UV/Vis spectrophotometer (Figure 7.1).

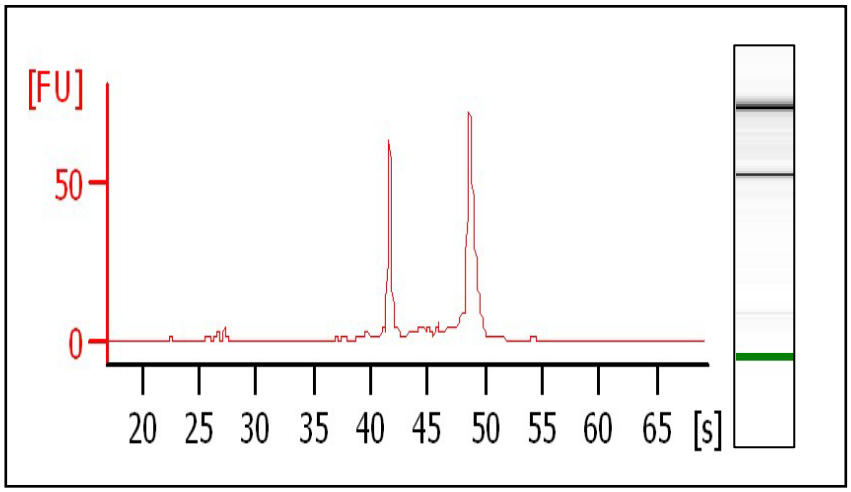



**Figure 7.1: Electropherogram of lung cancer sample showing RNA quality**

## Microarray experiments

Out of the 35 tumor biopsies collected, 5 biopsies with paired normal tissues and with RNA Integrity Numbers (RIN) above 8.0 were chosen for microarray experiments. RNA from corresponding normal tissues were used as reference against 5 tumor tissues. Total RNA from normal tissue from three patients was pooled in one slide and normal tissue from another two patients was pooled for second slide. ExpressArt® Amino Allyl mRNA amplification kit (Ocimum Biosolution, Hyderabad, India) was used for labeling. Individual tumor cRNA and that of the pool controls were labeled with cyanine 3 and hybridized on 'Human 40 K A OciChip' (Ocimum Biosolution, Hyderabad, India) which contained 20160 genes. The labeled and fragmented cRNAs were hybridized at 65˚C for 17 h. For each microarray multiple scans were performed with AFFYMETRIX 428™ array scanner using different photomultiplier amplification settings (PMT gain) for receiving reliable signals from both weak and strong spots with minimal saturation effects and data loss. The ImaGene™ software (BioDiscovery Inc. Los Angeles, CA) was used to calculate the intensity of each spot and the corresponding background from the individual scanner TIFF images. The ImaGene result files are further processed using the MAVI Pro software which combined the ImaGene data from the multiple scans at different PMT gain settings to one dataset per microarray using linear regression analysis. After filtration 19700 genes were retained, which were used for downstream statistical and biological analyses. The missing values were imputed by K-nearest neighbor (KNN) method. The resulting data was subjected to normalization. In order to facilitate comparison across arrays, median centering followed by scale adjustment for each array was performed. Median Absolute Deviation (MAD) scaling technique was used to rescale the data on each array. The ratio of the geometric means of the expression intensities for each gene fragment was calculated and reported in terms of the fold change (up or down) relative to the control. Analysis was done by Genowiz™ Software [Ocimun BIosolution, Hyderabad, India] and R package. The data has been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and assigned series accession number GSE30118.

### Gene ontology and pathway analysis

To determine roles of differentially expressed genes and classify them into functionally significant clusters a list of genes showing significant ($p<0.05$) ≥1.5-fold differential expression between tumor and normal pooled controls were selected and imported into DAVID, the database for annotation, visualization and integrated discovery http://david.abcc.ncifcrf.gov. The genes in the list were mapped to DAVID identifiers, and functionally annotated using the DAVID biological processes and molecular function categories. The one-tailed Fisher exact t-test probability value was used to statistically determine over- or under- representation of classification categories, Bonferroni corrected p-values less than 0.05 were considered significant.

### Ingenuity Pathway Analysis

Whole data set of differentially regulated genes was imported into Ingenuity Pathways Analysis (IPA) (Ingenuity® Systems, http://www.ingenuity.com) as gene identifiers with corresponding expression and p-values. The whole data set was then filtered on basis of ≥1.5 fold differential regulation and p value≤ 0.05. Each identifier from this filtered gene set was mapped to its corresponding gene object in the Ingenuity knowledge base. Global molecular network were developed from information contained in the knowledge base and biological functions that were most significant to these networks were determined. The data set was mined for significant pathways with the IPA library of canonical pathways and represented graphically the molecular relationships between genes and gene products. The intensity of genes (node) color in the networks indicates the degree of downregulation (green) or upregulation (red) of gene expression.

### Real-time PCR for microarray data validation

Microarray data validation was performed for 3 upregulated (*TMSB10, RPS8, PPFIA1*) and 2 downregulated genes (*TNS3, NGFR*), selected on the basis of being differentially regulated in microarray analysis and biological importance. 1 μg of total RNA was reverse transcribed using the High Capacity cDNA Archive Kit (Applied Biosystems). Real Time

analysis was performed with the ABI PRISM 7000 Sequence Detection System by using TaqMan Universal PCR Master Mix and "Assays-on-Demand" gene expression probes (Applied Biosystems; (TMSB10: Hs00363670_m1, RPS8: Hs01374307_g1, PPFIA1: Hs01549000_m1, TNS3: Hs00224228_m1, NGFR: Hs00609976_m1). Real-time PCR was performed in duplicate reactions of 20µl volume each. The mean expression level of target gene was calculated for patients normalized to a house keeping gene 18SrRNA (Part no: 4333760F). The relative gene expression levels were calculated by the $2^{-\Delta\Delta CT}$ method.

## RESULTS

### Patient Characteristics

Demographic characteristics and histological classification of patients is summarized in Table 7.1. The mean age of the patients was 61.04 ± 10.62 years (range 41-75yrs). The patient group comprised 25 male and 10 female and none of the patients received either preoperative radiation or chemotherapy.

**Table 7.1: Demographic and clinical characteristics of lung carcinoma cases**

| FACTORS | CATEGORIES | CASES n (%) |
|---|---|---|
| **Sex** | Male | 25 (71.42) |
| | Female | 10 (28.57) |
| **Histopathology** | Squamous cell carcinoma | 24 (68.57) |
| | Adenocarcinoma | 11 (31.42) |
| **Smoking status** | Non-smokers | 9 (25.7) |
| | Smokers | 26 (74.28) |
| **Tobacco chewing** | Non-chewers | 16 (45.71) |
| | Chewers | 19 (54.28) |
| **Betel quid chewing** | Non chewers | 11 (31.42) |
| | Chewers | 24 (68.57) |
| **Alcohol consumption** | Non-alcoholic | 27 (77.14) |
| | Alcoholic | 8 (22.85) |

## Gene Expression Profile in Lung Tumors

Microarray data were compared for five lung tumor tissues and matched normal tissues pooled in two slides. Out of the total 19700 genes analyzed, 778 genes were significantly ($p<0.05$) differentially expressed. Of this 734 genes were differentially expressed ≥1.5 fold. 311 genes were up-regulated, out of which 74 genes showed 1.5 to 2 fold up-regulation while 210 genes showed 2 to 5 fold difference in expression between tumor and normal tissues. 27 genes had above 5 fold difference in expression. 423 genes showed down-regulated expression, where 92 genes showed 1.5 to 2 fold differences in expression level and 289 genes showed 2 to 5 differences. 42 genes were down-regulated above 5 folds of expression level. Top 25 up and down regulated genes are listed in Table 7.2-7.3. Validation of microarray result was performed in the original tumor set and new set of 30 tumors and 7 control tissues. Consistent with the microarray data, the expression levels of *TMSB10, RPS8* and *PPFIA1* were up-regulated whereas *TNS3* and *NGFR* were down-regulated. The pattern of gene expression (up or down-regulation) initially identified by microarray were confirmed by Real Time PCR (Figure 7.2).

**Table 7.2: Top 25 significantly up-regulated genes between lung cancer tumors and normal control pool**

| S.No. | Accession | Gene symbol | Gene name | Fold change | P-value |
|---|---|---|---|---|---|
| 1 | NM_001016 | RPS12 | ribosomal protein S12 | 16.26 | 0.02 |
| 2 | NM_031924 | RSPH3 | radial spoke 3 homolog (Chlamydomonas) | 15.37 | 0.006 |
| 3 | NM_022764 | MTHFSD | methenyltetrahydrofolate synthetase domain containing | 12.15 | 0.01 |
| 4 | NM_000090 | COL3A1 | collagen, type III, alpha 1 | 11.76 | 0.02 |
| 5 | NM_001021 | RPS17 | ribosomal protein S17 | 8.57 | 0.04 |
| 6 | NM_017707 | ASAP3 | ArfGAP with SH3 domain, ankyrin repeat and PH domain 3 | 8.53 | 0.02 |
| 7 | BC017272 | GRK6 | G protein-coupled receptor kinase 6 | 7.91 | 0.02 |
| 8 | NM_001012 | RPS8 | ribosomal protein S8 | 7.31 | 0.03 |
| 9 | NM_018947 | CYCS | cytochrome c, somatic | 7.05 | 0.004 |
| 10 | NM_001018 | RPS15 | ribosomal protein S15 | 6.57 | 0.02 |
| 11 | NM_024678 | NARS2 | asparaginyl-tRNA synthetase 2, mitochondrial (putative) | 6.51 | 0.009 |
| 12 | NM_138793 | CANT1 | calcium activated nucleotidase 1 | 6.17 | 0.02 |
| 13 | NM_001034 | RRM2 | ribonucleotide reductase M2 polypeptide | 6.16 | 0.03 |
| 14 | NM_004846 | EIF4E2 | eukaryotic translation initiation factor 4E family member 2 | 6.15 | 0.02 |
| 15 | NM_015154 | MESDC2 | mesoderm development candidate 2 | 6.13 | 0.03 |
| 16 | NM_014260 | PFDN6 | prefoldin subunit 6 | 6.06 | 0.03 |
| 17 | NM_201575 | SEZ6L2 | seizure related 6 homolog (mouse)-like 2 | 5.91 | 0.002 |
| 18 | AF155235.1 | NHP2L1 | NHP2 non-histone chromosome protein 2-like 1 (S. cerevisiae) | 5.73 | 0.02 |
| 19 | NM_020689 | SLC24A3 | solute carrier family 24 (sodium/potassium/calcium exchanger), member 3 | 5.30 | 0.04 |
| 20 | NM_033553 | GUCA2A | guanylate cyclase activator 2A (guanylin) | 5.13 | 0.02 |
| 21 | NM_000536 | RAG2 | recombination activating gene 2 | 5.11 | 0.0004 |
| 22 | BC039243 | FGFR2 | fibroblast growth factor receptor 2 | 5.09 | 0.04 |
| 23 | NM_014594 | ZNF354C | zinc finger protein 354C | 4.92 | 0.005 |
| 24 | NM_004990 | MARS | methionyl-tRNA synthetase | 4.90 | 0.04 |
| 25 | NM_005199 | CHRNG | cholinergic receptor, nicotinic, gamma | 4.79 | 0.002 |

**Table 7.3: Top 25 significantly down-regulated genes between lung cancer tumors and normal control pool**

| S.No. | Accession | Gene symbol | Gene name | Fold change | P-value |
|---|---|---|---|---|---|
| 1 | AL359771 | PDPN | podoplanin | 0.02 | 0.005 |
| 2 | BC069331 | ELA2A | elastase 2A | 0.04 | 0.001 |
| 3 | NR_002188 | GBAP | glucosidase, beta; acid, pseudogene | 0.04 | 0.04 |
| 4 | NM_019841 | TRPV5 | transient receptor potential cation channel, subfamily V, member 5 | 0.06 | 0.004 |
| 5 | BC050611 | ERLIN2 | ER lipid raft associated 2 | 0.06 | 0.0008 |
| 6 | NM_004468 | FHL3 | four and a half LIM domains 3 | 0.07 | 0.01 |
| 7 | NM_014723 | SNPH | syntaphilin | 0.07 | 0.04 |
| 8 | BC023974 | NLRX1 | NLR family member X1 | 0.07 | 0.04 |
| 9 | NM_024876 | ADCK4 | aarF domain containing kinase 4 | 0.07 | 0.01 |
| 10 | AF107406 | YWHAB | tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide | 0.08 | 0.03 |
| 11 | NM_006700 | TRAFD1 | TRAF-type zinc finger domain containing 1 | 0.09 | 0.01 |
| 12 | BX640764 | MALAT1 | metastasis associated lung adenocarcinoma transcript 1 (non-protein coding) | 0.09 | 0.04 |
| 13 | NM_014702 | KIAA0408 | KIAA0408 | 0.10 | 0.0003 |
| 14 | AC005041 | LBX2 | ladybird homeobox 2Homo sapiens BAC clone RP11-523H20 from Ch2, complete sequence | 0.10 | 0.003 |
| 15 | NM_019027 | RBM47 | RNA binding motif protein 47 | 0.11 | 0.002 |
| 16 | AF067804 | MLL5 | myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, Drosophila) | 0.11 | 0.03 |
| 17 | NM_018349 | MCTP2 | multiple C2 domains, transmembrane 2 | 0.12 | 0.01 |
| 18 | AL354743 | PRDM16 | PR domain containing 16 | 0.12 | 0.02 |
| 19 | NM_021025 | TLX3 | T-cell leukemia homeobox 3 | 0.13 | 0.04 |
| 20 | NM_022912 | REEP1 | receptor accessory protein 1 | 0.13 | 0.02 |
| 21 | NM_006521 | TFE3 | transcription factor binding to IGHM enhancer 3 | 0.13 | 0.009 |
| 22 | NM_014444 | TUBGCP4 | tubulin, gamma complex associated protein 4 | 0.13 | 0.01 |
| 23 | NM_017553 | INO80 | INO80 homolog (S. cerevisiae) | 0.14 | 0.01 |
| 24 | NM_001126 | ADSS | adenylosuccinate synthase | 0.14 | 0.009 |
| 25 | NM_019886 | CHST7 | carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 7 | 0.14 | 0.01 |

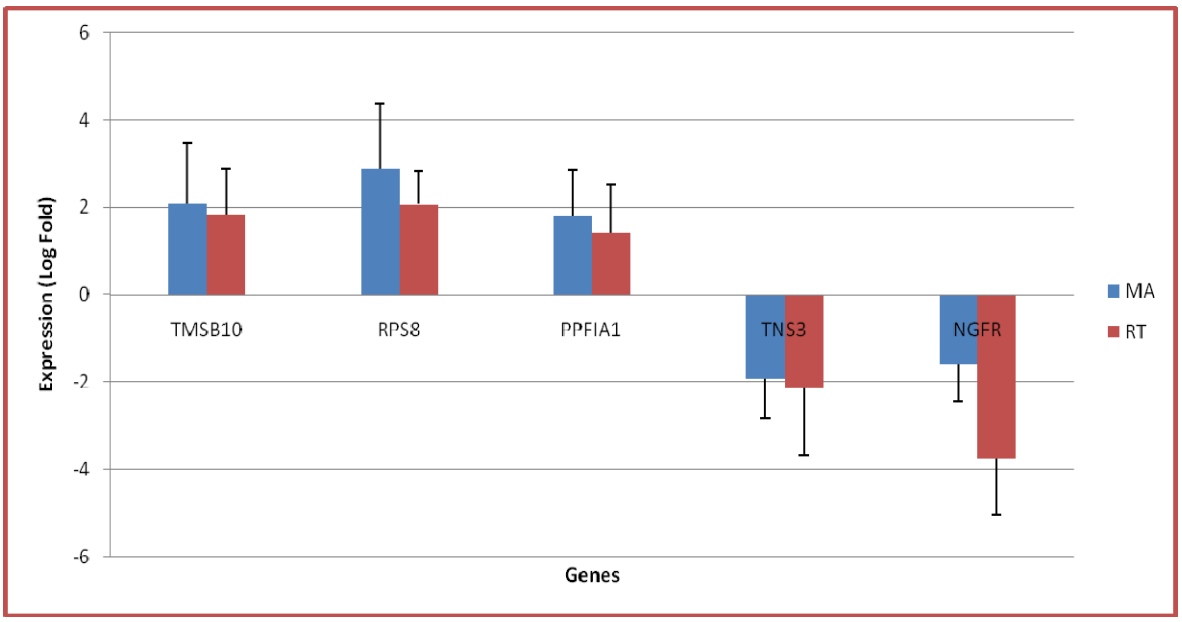



**Figure 7.2: Comparison of gene-expression level by microarray and Real-Time PCR**

### Gene Ontology/Enrichment Analysis

A total of 647 genes were mapped to DAVID identifiers from a list of 734 differentially regulated genes. Genes were clustered on the basis of common annotation terms through the functional annotation clustering tool. Top 10 clusters with an enrichment score of ≥1.3 are represented in Table 7.4. The top most cluster with an enrichment score of 2.03 relates to the epidermal growth factor term consisting of genes like *MUC2, MUC4, HYAL3, FN1* and *FBN2* indicating a role of molecules in extra-cellular matrix. Homeobox related transcription activity genes and transcription cofactor activity genes formed the second and the third clusters. The largest cluster with maximum number of genes (n=60) related to positive regulation of gene expression. The cluster of MAPKKK cascade consisted of 53 genes with most important genes like *MAP3K1, MAPK8IP2, MAPK8IP3* and *CCND3*. Other significant clusters related to genes involved in inflammatory defense response, cell membrane fraction, signal anchors, extra cellular matrix structural constituent and protein complex biogenesis. Kyoto Encyclopedia of Genes and Genomes (KEGG) molecular pathway analysis were performed to determine significant biological pathways. The most significant pathways included the p53 signaling pathway, the neurotrophin pathway and the MAPK signaling pathway (data not shown).

**Table 7.4: Functional enrichment of genes differentially expressed in lung cancer tumors and normal control through annotational clustering in DAVID**

| | Classification Terms | Enrichment Score | No. of components | Genes |
|---|---|---|---|---|
| Cluster 1 | Epidermal Growth Factor | 2.03 | 30 | FBN2, FN1, HSPG2, UMODL1, TGFA, MUC4, HYAL3, CD97, MUC2 |
| Cluster 2 | Homeobox related transcription activity | 1.90 | 57 | PLAG1, POU5F1, TGIF2, YEATS4, YY1, DLX4, FOXA1, ZNF263 |
| Cluster 3 | Transcription cofactor activity | 1.86 | 44 | YY1, NRG1, MYCBP, CREM, ING2, p53 |
| Cluster 4 | Inflammatory/defense response | 1.83 | 47 | FN1, SOD2, NGFR, MAP3K1, IFNG, PDPN, PDGFRA, NCR2, NRG1, SIGIRR |
| Cluster 5 | Cell/membrane fractions | 1.76 | 56 | WISP1, DMD, EDNRB, EMP1, FADS2, PPAP2A, FMO4 |
| Cluster 6 | Signal anchor | 1.66 | 44 | COL3A1, TNFSF14, HS2ST1, PLD1, WWOX, RAP1GAP |
| Cluster 7 | Extra cellular matrix structural constituent | 1.62 | 29 | ADAMTS9, EFEMP1, COL3A1, FBN2, FN1, MUC4, NAV2 |
| Cluster 8 | MAPKKK cascade | 1.50 | 53 | RPS6KA1, MAP3K1, DUSP16, MAPK8IP2, MAPK8IP3, CDC25B, CCND3 |
| Cluster 9 | Positive regulation of gene expression | 1.48 | 60 | SIX1, CDK4, IFNG, HOXC6, OSM, YY1 |
| Cluster 10 | Protein complex biogenesis | 1.38 | 36 | CTTNBP2, CRYAB, YWHAB, IRF7, MDM2, PFDN6 |

**Ingenuity Pathway Analysis (IPA)**

Out of the 18389 genes submitted into IPA, 17535 genes mapped to the IPA database, of which 12802 genes were eligible for network analysis. However after applying a filtering criteria of 1.5 fold differential expression and p value ≤0.05 only 230 genes were eligible for network analysis and 219 genes were eligible for functional/pathway analysis. Relevant statistically significant biological functions were related to cell death, cancer, cell cycle, cellular assembly and organization, cellular compromise and cell morphology (Table 7.5, Figure 7.3). Up-regulated genes associated with cell death, cancer and cell cycle included, *FGFR2, IFNG, RAG2, MAP3K2, TMSB10, CANT1, COL3A1, CXCL9, BRIP1, TOP1* and *AGPAT6*. Down-regulated genes associated with above biological functions were *CD44, NGFR, PDGFRA, CASP10*, *MALAT1, MPL, PDPN, CNN1, ID3* and *ADAMTS9*. Canonical metabolic pathways analysis revealed that the death receptor signaling, aminosugar metabolism, Type I diabetes mellitus signaling, purine metabolism, *BRCA1* in DNA damage response and Myc mediated apoptotic signaling pathways were associated with most DEGs (data not shown).

Out of the 25 networks identified by IPA, 24 networks had a score [-log (p value)] values between 2 and 30, 16 of these had 11–30 focus genes. Top 16 networks with highest focus molecules are summarized in Table 7.6. The analysis highlighted some common terms associated with the significant networks identified. These terms were mainly related to cell death, cell cycle, cellular growth and proliferation, cell signaling, inflammatory and respiratory diseases, DNA replication, gene expression and molecular transport. Networks related to cell death, gene expression, cell cycle and cell signaling are shown in Figure 7.4. Cell death was a major term recognized by IPA under all the three analyses categories i.e. biological function, networks and canonical pathways.

Genes restricted to lung cancer, non small cell lung cancer and tumor formation terms were overlaid with pathways for NSCLC from IPA database. A functional pathway (Figure 7.5) of significant DEG obtained from the microarray result was constructed to identify putative interactions and biomarkers. As is shown in Figure 7.5 there was a strong contribution of *IFNG* and *BCL2* genes.

**Table 7.5: Top ten statistically relevant biological functions in Ingenuity Pathway Analysis**

| Biological Function | -log (p value) | Molecule |
|---|---|---|
| Cell Death | 1.57E-05 | SDHB, **TP73**, PNKP, COL4A3, NRG1, FCER1A, DMD, **BCL2**, PKN2, **TOP1**, HINT1, MPL, **SOD2**, CLK3, **NGFR, GFRA1, INS**, RAG2, **PDGFRA**, HSD11B2, MYBL2, PRDM1, GJB2, CASP10, MAP3K2, **IFNG**, CXCL9, ST8SIA1, **TMSB10**/TMSB4X, YWHAB, **GPR132**, RRM2, SIGLEC12, **FGFR2**, EN2, ID3, EMP1, CIDEA, **SYK**, SLC37A4, **CD44**, GUCA2A, CYCS, **REL**, CTTN |
| Cancer | 2.11E-05 | MALAT1, TPD52L2, **TP73**, PFDN6, FCER1A, PDE4D, **TOP1**, SNPH, **SOD2**, GPR4, OLIG2, PRRX2, EIF1AX, **INS**, ERLIN2, TGIF2, PRDM1, EFNB3, TFE3, BRIP1, CASP10, CXCL9, SLC24A3, RRM2, **FGFR2**, ADAMTS9, GLRX3, CHRNG, **SYK**, RPS15, CYCS, AGPAT6, **REL**, CTTN, CANT1, COL3A1, UBE2I, PDPN, SDHB, TIFA, COL4A3, NRG1, DMD, MRE11A, **BCL2**, HINT1, MPL, **GFRA1**, **NGFR**, RAG2, **PDGFRA**, HSD11B2, RXRB, MTHFD1L, EED, GJB2, PPFIA1, **IFNG**, APOBEC1, **TMSB10**/TMSB4X, **GPR132**, LTBP1, EMP1, PCM1, CNN1, **CD44**, DLEC1, CEP70, PRDM16, SCN3A |
| Cell Cycle | 7.14E-05 | **TP73**, NRG1, MRE11A, MLL5, **BCL2, TOP1, SOD2, NGFR, GFRA1**, **INS, PDGFRA**, MYBL2, TFE3, **IFNG**, ST8SIA1, YWHAB, **GPR132**, **FGFR2**, ID3, FANCL, PCM1, **SYK, CD44**, MXD4, **REL**, ACRBP |
| Cardiovascular System Development and Function | 8.88E-05 | IFNG, TP73, COL4A3, NRG1, GPR132, RRM2, FGFR2, BCL2 |
| Hematological Disease | 1.19E-04 | TP73, TRPV5, FCER1A, PDE4D, ABCA1, TOP1, SOD2, CRADD, OLIG2, INS, KLHL29, BRIP1, UNC13D, CXCL9, RRM2, FGFR2, CHRNG, SYK, REL, AGPAT6, RHAG, UBE2I, TLX3, RBM47, COL4A3, BCAS3, NRG1, C17orf57, CC2D2A, BCL2, TGFBRAP1, MPL, GFRA1, NGFR, RAG2, PDGFRA, LY75, RXRB, MTHFD1L, IFNG, ACBD3, SIGLEC12, KCNH2, FANCL, LTBP1, TNS3, NFIA, MOSC1, PHACTR2, DLEC1, ADAMTSL1 |
| Digestive System Development and Function | 1.42E-04 | IFNG, TP73, PDGFRA, FGFR2, COL3A1, BCL2 |
| Organismal Injury and Abnormalities | 1.77E-04 | UNC13D, COL4A3, UTS2, PDE4D, ABCA1, CHRNG, SLC23A1, MPL, NGFR, SYK, CNN1, INS |
| Cellular Assembly and Organization | 2.85E-04 | MALAT1, IFNG, AP2M1, TMSB10/TMSB4X, GPR132, NRG1, DMD, FHL3, VTI1B, SEC23A, TOP1, NGFR, INS, ARAP1, CD44, CYCS, EPS8L2, EFNB3 |
| Cellular Compromise | 2.85E-04 | IFNG, SDHB, TMSB10/TMSB4X, TP73, FCER1A, FHL3, AQP4, ABCA1, LTBP1, BCL2, TYRP1, CNN1, SYK, SLC37A4, INS, CD44, CYCS |
| Cell Morphology | 3.21E-04 | IFNG, ST8SIA1, TMSB10/TMSB4X, NRG1, GPR132, AQP4, BCL2, SEC23A, MPL, SOD2, NGFR, INS, CD44, ARAP1, EPS8L2, MYBL2, EFNB3, RXRB, CTTN |

List of the statistically relevant top ten over-represented biological functions. Genes in bold are overlapping genes within first three terms. Genes underlined were validated through Real-Time PCR. Fisher's exact test was used to calculate a p-value. Cut-off point of significance is $P < 0.05$, which corresponds to −log (P-value) of 1.3.

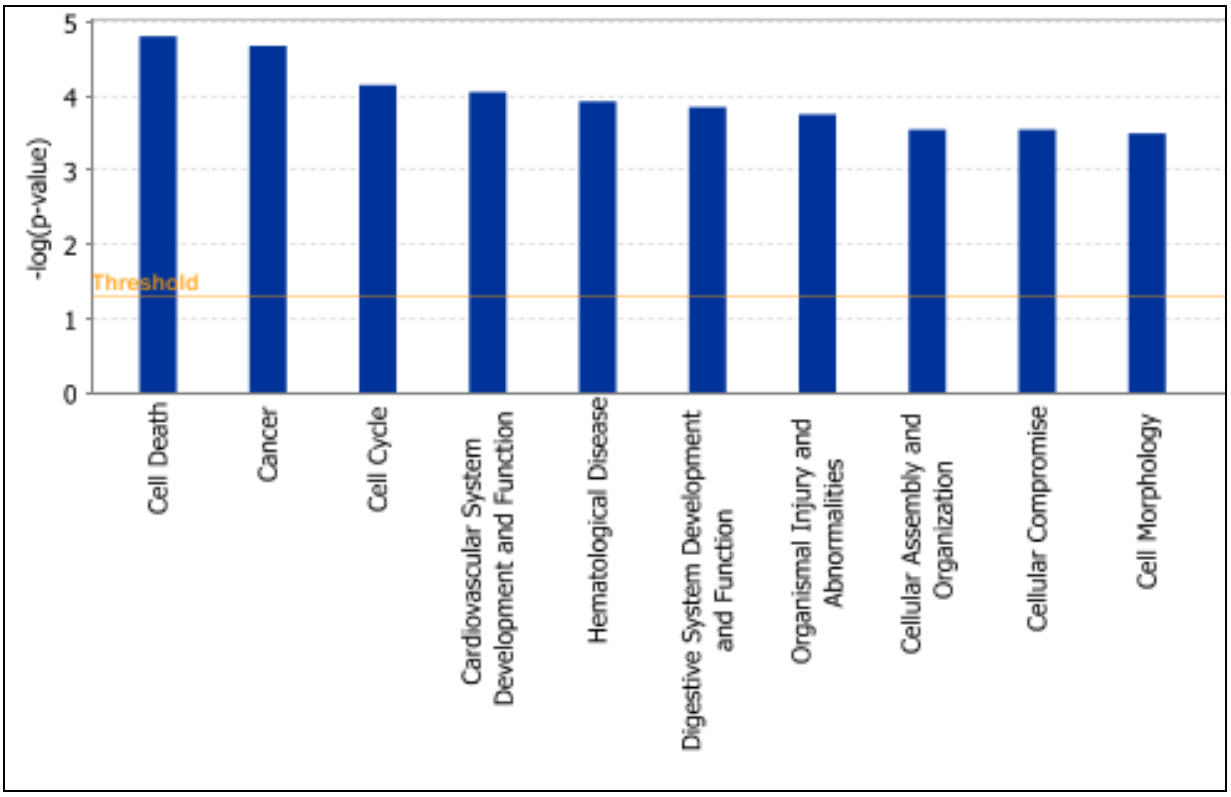



**Figure 7.3: Top enriched bio-functions as determined by Ingenuity Pathway Analysis tool**

Fisher's exact test was used to calculate a p-value. Threshold bar shows cut-off point of significance $P < 0.05$, $-\log(\text{P-value})$ of 1.3.

## Table 7.6: Top significant networks of differentially expressed genes

| | Functions | Score | Focus Molecules | Molecules in Network |
|---|---|---|---|---|
| 1 | Carbohydrate Metabolism, Small Molecule Biochemistry, Lipid Metabolism | 30 | 19 | **ADAMTS9**, Adaptor protein 1, Ap1, **BRIP1**, Calcineurin protein(s), CaMKII, **CASP10**, **CD44, CRADD, CXCL9**, **GLRX3, GNE, HINT1**, **IFN Beta**, IFN TYPE 1, Ikb, IKK (complex), Ikk (family), IL12 (complex), **MAP3K2**, Mek, NFkB (complex), **NGFR, PKN2**, **REL, SLC37A4, ST8SIA1, TIFA**, TIP60, Tlr, **TLX3**, Tnf receptor, **TRAFD1** |
| 2 | Cell Death, Cellular Growth and Proliferation, Tissue Development | 29 | 19 | Akt, **BCAS3, CAMLG**, **CCDC71, CIDEA**, Cyclin A, E2f, **EED, ELK3**, Hdac, HDL, HISTONE, Histone h3, Histone h4, Hsp70, Hsp90, **INS, MARS, MTHFSD**, **NRG1**, **OLIG2**, Rb, **RRM2**, **TFE3**, **TNPO1, TOP1, TP73, TYRP1, UBE2I**, Ubiquitin, **WDR76** |
| 3 | Immunological Disease, Inflammatory Disease, Renal Nephritis | 26 | 17 | **ABCA1**, Alp, Calpain, **CNN1, COL4A3**, collagen, Collagen type I, Collagen(s), **EIF4E2, EIF4ENIF1, EPS8L2, FHL3**, Focal adhesion kinase, growth factor receptor, Integrin, **KCNH2,** LDL**, LTBP1, MYBL2, NRIP2**, Pdgf (complex), PDGF BB, **PDGFRA**, PI3K (complex), **PRRX2**, Ras homolog, Rxr, **RXRB, SDHB**, Tgf beta, **TMEM57, ZNF354C** |
| 4 | Hair and Skin Development and Function, Organ Development, Lipid Metabolism | 24 | 16 | **ADSS, AHNAK, AQP4, COL3A1, CTTN, EFCAB6, EMP1**, ERK1/2, Fc gamma receptor, Fcer1**, FCER1A**, Fgf, **FGFR2**, **GUSB, ID3**, Iga, Ige, IgG, IgG1, IgG2a, IL1, IL23, IL12 (family), **MPL**, P38 MAPK, **PDE4D**, PLC gamma, **PRDM1, PRSS8**, STAT5a/b, **SYK**, SYK/ZAP |
| 5 | Cell Morphology, DNA Replication, Recombination, and Repair, Carbohydrate Metabolism | 24 | 16 | AGTPBP1, **ANO2**, CLTCL1, CP110, **EDEM3**, GSTO1, HMOX2, HNF4A, IRS4, LAS1L, LRP5, **LRRC3, MESDC2, MTMR4, NARS2**, NCK1, NRD1, **ODZ4, PALMD, PKN2, PNKP, PNMA1, RDH11, RSPH3, SEC23A**, SEC23IP, SEC24D, SEC31A, SEMA7A, SETDB1, **SLC23A1, SNX5**, TXNDC9, VPS29, XRCC4 |
| 6 | Organ Development, Reproductive System Development and Function, Genetic Disorder | 22 | 17 | 14-3-3, **AGPAT6*, ATP5B*, BCL2*,** Calmodulin, **CGB (includes others)*, DMD*,** F Actin, **GJB2, GRK6*,** Gsk3, hCG, Insulin, Jnk, **LBH*, LY75, MLXIP*,** Nfat (family), **NLRX1*, PDPN*,** Pka, Pkc(s), PLC, Ras, **SERPINA2**, **SNPH**, Sod, **SOD2**, TCR, Trypsin, **UTS2**, Vegf, **YWHAB** |
| 7 | Cell Signaling, Molecular Transport, Vitamin and Mineral Metabolism | 22 | 15 | **AP2M1, CCRL2,** CELSR1, CHEMOKINE, **CHRNG, CHST7,** Ck2, **CYCS,** ERK, Gpcr, **GPR4, GPR132**, GPR146, **GPR182,** GPR109B, GPRC5B, **IFNG*, MAGOH,** Mapk, MCHR1, **MRE11A*, PFDN6, RAB43,** Rac, RNA polymerase II, RTDR1, RXFP3, RXFP4, **SSR1*,** STAT, **SUMO2,** UTS2R |
| 8 | Gene Expression, Infection Mechanism, Organ Morphology | 22 | 15 | **ACRBP, ADCK4*,** AR, BMF, CDC42EP4, CDK11A/CDK11B, CIDEC, **CTDSP2*,** DUSP3, **EMP1, GJB2, HSD11B2,** IDH2, INSL3, INSR, LPIN1, MAL, NBL1, NR3C1, PNRC1, PRPF6, **SEZ6, SEZ6L2*, SIGLEC12**, **SLC24A3,** TGFB1I1, **TGIF2, THSD4*, TRIM34, TRPV5*,** UBE2K, **WDR6*,** WNT5B |

| | | | | |
|---|---|---|---|---|
| 9 | Nervous System Development and Function, Inflammatory Disease, Respiratory Disease | 22 | 15 | **ACBD3,** ADRB1, AGRN, AGTR1, AKAP6, **ARAP1*,** ATP5A1, **ATP5B*,** ATP5O, **BTN2A3*, C16orf7,** CELSR1, **CFHR5,** DLG4, DNM1L, **FGFR2*, GFRA1,** GRB2, GRIA4, KCNQ1, **KIF26A*, LOC100505503/RPS17, MALAT1,** NCAM1, **PDE4D*, PHACTR2,** PIK3AP1, **POMT1*,** PRKAR1A, PRKAR1B, RTKN2, SDHA, SSTR1, **ZNF7*** |
| 10 | Tumor Morphology, Hematological Disease, Cell Cycle | 20 | 14 | ANK1, **ASAP3*, CANT1,** CTSZ, DOCK4, DUB, **EFNB3, EIF1AX, EN2,** EPO, ESR1, EZR, **FAM70A,** FHL1, GOLGA7, **HINT1,** HRAS, INSL3, **MRPS16,** NOSIP, PINX1, **POP1,** RAC1, **RAG2,** RAP1B, **RHAG*,** RP2, **UMODL1, USP26,** USP32, USP38, USP43, VHL, **ZDHHC9** |
| 11 | Cellular Assembly and Organization, Gene Expression, Cellular Growth and Proliferation | 18 | 13 | ARPC2, **CAND2*, CCDC88B*,** CLCN3, **CNN1*,** CTPS, **ELK3,** EPRS, **ERLIN2*,** FHL1, HNRPDL, IL17RB, KDM5B, KIAA0226, **KIF26A*,** LPCAT3, MAP3K7, MIA3, **MON1A*, MXD4, NHP2L1,** PHF15, PIK3R4, **PMFBP1*, RBM47, REEP1,** Proliferation SAE1, SRF, TGFB1, **TGFBRAP1,** TJP2, TRAF6, UVRAG, VPS33A |
| 12 | Cellular Movement, Carbohydrate Metabolism, Molecular Transport | 18 | 13 | ALAS1, AMD1, APC, APCDD1, AQP9, **CCRL2, CELF3*,** CTSZ, CYTIP, DTL, G0S2, GLA, **GUCA2A,** HNF1A, **HS2ST1,** IL1B, **INO80*, METTL9, MFNG, MLL5*, NFIA,** NFYB, NKX2-2, **ODF2*,** POU5F1, RAB3C, RORB, SCAF11, **SLC37A4*,** TLE1, **TMSB10/TMSB4X*, ZNF462** |
| 13 | Dermatological Diseases and Conditions, Genetic Disorder, Cell-To-Cell Signaling and Interaction | 16 | 12 | ABI3, **AHDC1*,** ANXA11, ARC, ATXN1, C10orf10, CARD9, **CCDC120*,** CCL13, CCL28, CEP72, **COLQ*,** CXCL16, HIVEP1, **HNRNPH3*, KIAA0408,** KRT15, **MOSC1,** MTF1, **PCM1*,** RAB27B, **SLC30A1,** TMC6, **TMC8*,** TNF, **TPD52L2*,** TRAF2, TRIM32, **TRPT1,** TXNRD1, **UNC13D,** XCL1 |
| 14 | Cellular Assembly and Organization, Cellular Function and Maintenance, Gene Expression | 15 | 12 | **APOBEC1,** APOBEC2, BET1, BET1L, BRF1, CABLES1, CCDC99, CMIP, **GBAP1*,** GOLGB1, GOSR1, GOSR2, **HINT1,** IL4, **KIAA1377, LFNG*,** NFATC2IP, **OMP, PDE6B, PET112L*,** RRN3, **SEC22A,** SL1, SRC, STX5, TAF1, TAF1A, TAF1B,**TAF1C*,** TBP, **TNS3*,** TOM1L1, TP53, **VTI1B,** YKT6 |
| 15 | RNA Post-Transcriptional Modification, Protein Synthesis, Cell Cycle | 14 | 11 | AIMP2, **C20orf70,** CDKN2A, CELSR1, DDX56, ELF3, ELF5, **EMP1, FANCL*, FTSJ3*,** HGF, IFNA2, IGLL1/IGLL5, **MRPS24, MTHFD1L*,** MYC, MYCN, NOSIP, **PRDM16,** RPL7, RPL26, RPL35, RPL38, **RPS12, RPS15*,** RPS24, RRM2B, SLK, **TAF9*,** TAF6L, TBC1D4, Tgtp1, TUBGCP3, **TUBGCP4*** |
| 16 | Cell Signaling, Molecular Transport, Vitamin and Mineral Metabolism | 14 | 11 | **CEP70,** CLK2, **CLK3*,** DFFA, DYRK1A, ERC1, FKBP8, FKBP1A, GRM1, ITPR1, ITPR2, ITPR3, **KRT37,** MADD, **MAP3K2*,** NEDD4L, NFKB1, PIK3R4, PNN, **PPFIA1*,** PPFIBP1, RASGRF1, **RPS8*, SCN3A*,** SLC8A1, **SRSF4*,** SRSF6, **TONSL,** TRP, TRPC1, TRPC3, TRPC6, **TRPC7,** YWHAG, **ZNF839*** |

Networks identified in differentially regulated genes analyzed using the IPA tool (version 9.0). Focus genes are shown in bold. Underlined genes are validated through Real-time PCR. The other genes are either absent from the microarray or found not significantly regulated.

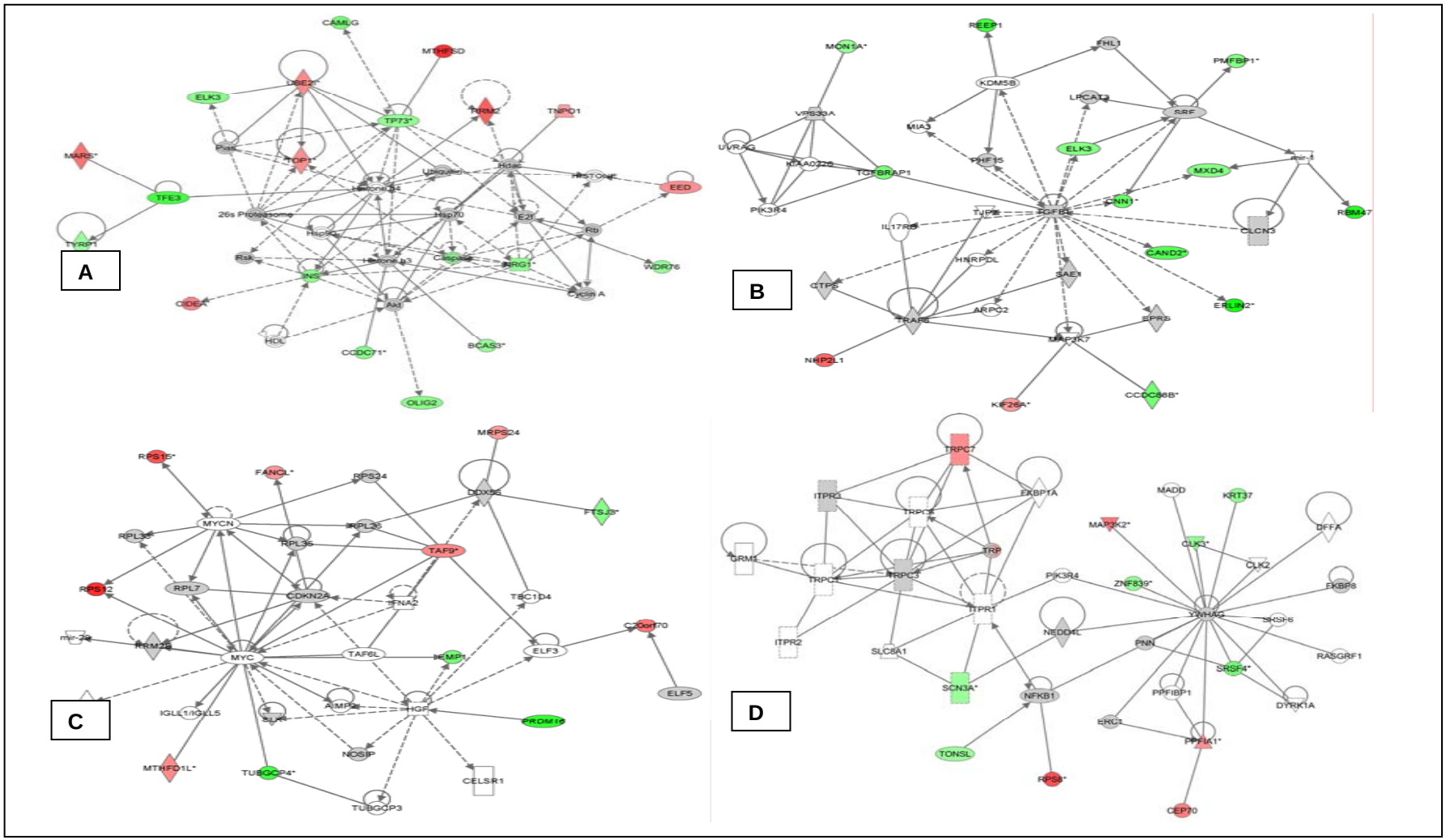



**Figure 7.4: Most significant gene networks of differentially regulated genes in lung cancer.** (A) Cell Death (B) Gene Expression (C) Cell Cycle (D) Cell Signaling. Solid arrows represent known physical interactions, dotted arrows represent indirect interactions. Genes in red showed increased expression in tumor samples while genes in green decreased their expression in tumor. The genes that do not meet the P-value cutoff of 0.05 are shown in gray.

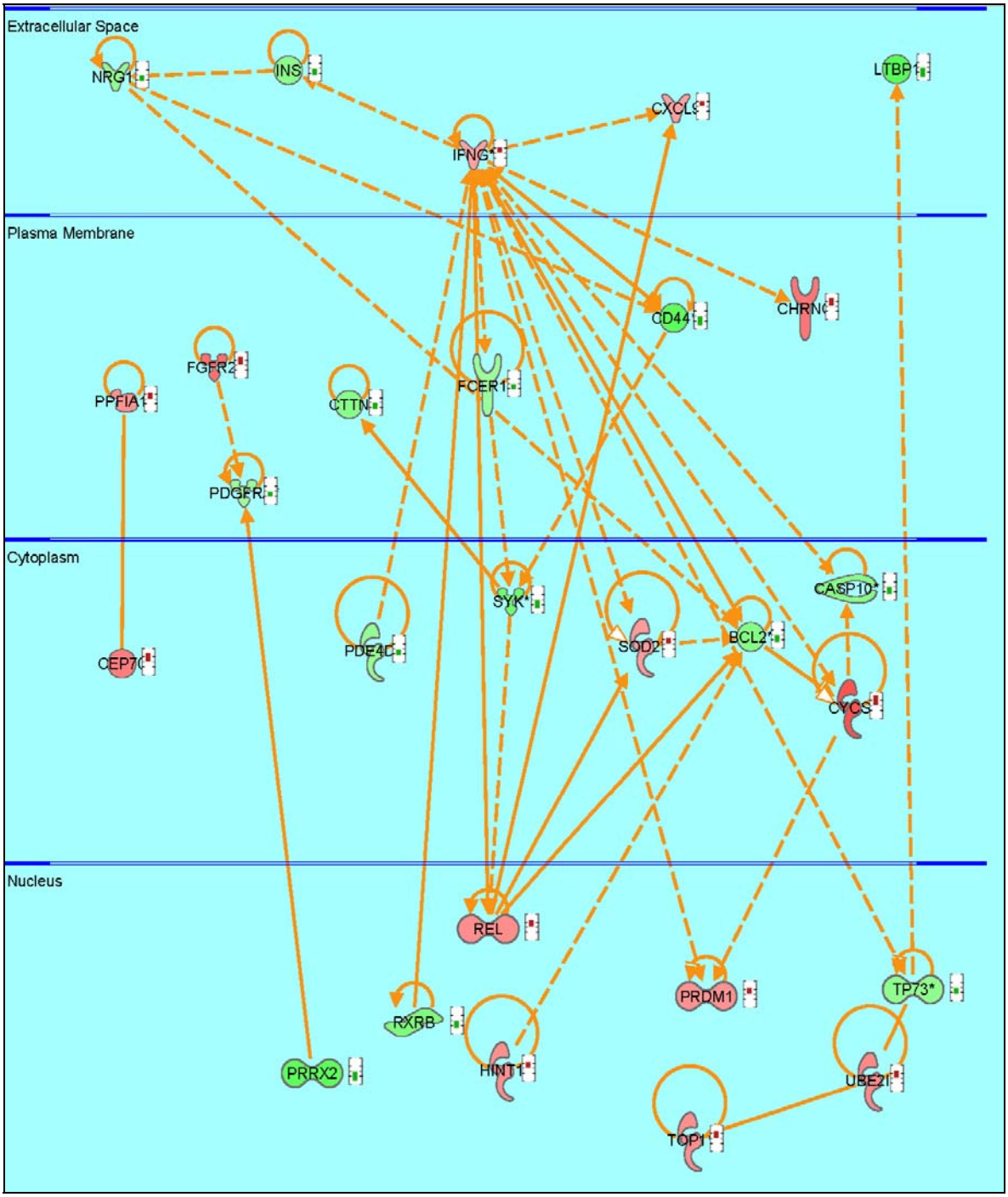



**Figure 7.5: Biological Gene network of differentially regulated genes associated to lung cancer selected as putative biomarkers as determined by IPA.** Genes are positioned in subcellular layout. Solid arrows represent known physical interactions,dotted arrows represent indirect interactions. Genes in red showed increased expression in tumor samples while genes in green decreased their expression in tumor.

## DISCUSSION

NSCLC is a heterogeneous form of lung cancer with complex and diverse molecular and genetic events. In the present study oligonucleotide microarray was used to identify aberrantly expressed genes in NSCLC CT-guided biopsies and matched normal tissues. Lung biopsy has been reported to be a reliable source for gene expression profiling with robust and clinically relevant molecular signatures [Borczuk et. al., 2004].

There was good overall concordance between the microarray and the real time RT-PCR data. Over-expression of *TMSB10* is reported in several cancers such as colon, breast, uterine and ovarian [Santelli et. al., 1999]. It acts as a monomeric actin sequestering protein that regulates actin dynamics and might increases cell motility. As a tumor suppressor it inhibits angiogenesis and its upregulation predisposes a cell to undergo apoptosis [Santelli et. al., 1999, Lee et. al., 2005]**.** Ribosomal proteins (RP) are a major component of ribosomes and play critical roles in protein biosynthesis. Increased expression of *RPS8* along with *RPS24* and *RPL32* has been reported in various tumors including differentiated ovarian, colorectal tumors and astrocytomas [Pogue-Geile et. al., 1991, Welsh et. al., 2000, MacDonald et. al., 2007]. Another up-regulated gene, *PPFIA1* is a member of the liprin protein family which encodes a cytoplasmic protein necessary for focal adhesion, axon guidance and mammary gland development. Over-expression of *PPFIA1* which is associated with enhanced cell spreading and migration is one of the most amplified genes recognized in the 11q13 locus associated with laryngeal, head and neck and breast cancers [Järvinen et. al., 2006, Tan et. al., 2008]. The down-regulated *NGFR* (nerve growth factor receptor) or *p75NTR* in this study is a tumor suppressor gene [Jin et. al., 2007] with role in metastatic suppression and apoptosis induction. Absent or significantly decreased expression of *NGFR* is observed in gastric and hepatocellular carcinomas [Jin et. al., 2007, Yuanlong et. al., 2008]. Role of *TNS3* (Tensin 3) in cancer has largely been unexplored. However, a study on kidney cancer reports downregulation of the Tensin family genes including *TNS3* [Martuszewska et. al., 2009]. Tensin 3 has been shown to anchor integrins to the cytoskeleton, rendering the cell less motile therefore its expression is correlated with negative regulation of migration and longer survival [Martuszewska et. al., 2009]**.**

Many of the pathways identified in DAVID and IPA were related to genes involved in transcriptional activities, gene expression, cellular assembly, cell growth, signaling, and cell death. The most significant cluster in DAVID was epidermal growth factor (EGF) term which primarily consisted of extra-cellular matrix glycoprotein genes (*FBN2, HSPG2*, *UMODL1, TGFA* and *MUC4*) known to harbor EGF-like domains. *MUC4, MUC2, TGFA* and *FN1* modulate downstream cell growth signaling, cell proliferation and migration, mostly through their interactions with the EGF receptors. The second cluster was the homeobox family of transcription factors containing the homeodomain, that directly binds DNA and regulate transcription of genes involved in morphogenesis. The homeobox gene *DLX4*, which is invariably absent in normal tissues but expressed in tumors of lung, breast, prostate and ovary, acts by blocking the anti-proliferative effect of TGF-β [Trinh et. al., 2011]. Another gene *FOXA1*, found to be amplified and over-expressed in esophageal and lung cancer [Lin et. al., 2002], promotes expression of genes associated with metabolic processes, regulation of signaling, and the cell cycle. Concordantly, we found increased expression of *DLX4* and *FOXA1* along with two other transcription factors *POU5F1* and *YY1* whose aberrant expression is associated with cell proliferation, inflammation and increased mortality in NSCLC. Further analysis of the DEGs through functional annotation clustering tool revealed that a strikingly high number of genes (53 genes) were associated with MAPK signaling and its feedback regulation. The expression of several genes involved in the 'MAPKKK' cluster, that is *MAP3K1, MAPK8IP2, MAPK8IP3, RPS6KA1* and *DUSP16* were in agreement with the important role of this pathway in coordinating various cellular processes, such as mitosis, differentiation, proliferation and apoptosis. Deregulation of the classical MAPK pathway (Ras/Raf/MEK/ERK cascade) and aberrant activation of the Wnt signaling pathway have been implicated in NSCLC development and progression in earlier studies also [Uematsu et. al., 2003].

Results from IPA corroborated the DAVID analysis. Classified on the basis of biological function, the top three terms were cell death, cancer and cell cycle and a majority of genes overlapped within these terms. This fuzziness establishes the multi-faceted nature of the genes and improves their chances of discovery. Majority of these overlapping genes like *TP73, TOP1, SOD2, IFNG, CD44, GPR132, FGFR2* and *PDGFR*

are involved in carcinogenesis. Tumor suppressor TP73 is reported to be epigenetically silenced and under-expressed in prostate cancer [Singh et. al., 2007] while expression of Topoisomerase I (*TOP1*) is associated with tumor growth and progression in colorectal cancer [Ataka et. al., 2007]. Superoxide dismutase 2 (*SOD2*), an antioxidation enzyme, has been implicated in different processes of carcinogenesis from cell cycle alterations to inducing senescence [Hempel et. al., 2011].

An interesting finding within this overlapping set was the down-regulation of the genes involved in nervous system development. A neurotrophin receptor *NGFR* (Nerve growth factor receptor) acts as a tumor suppressor which negatively regulates cell growth and proliferation and is associated with a longer disease-free survival [Reis-Filho et. al., 2006]. Another putative tumor suppressor *NRG1* (neuregulin-1) gene is either lost by LOH or is undetectable due to hyper methylation especially in epithelial cancers including colorectal and breast [Oster et. al., 2011, Chua et. al., 2009]. Lastly, *GFRA1* gene (GDNF family receptor alpha 1) interacts with RET to form a signaling complex whose downstream effect regulates cell survival. It shows high expression level in teratomas and neuroblastomas [Bing et. al., 2008, Gimm et. al., 1999].

Among the up-regulated genes not previously associated with lung cancer we found *CLK3* and *PFDN6*. *CLK3* (CDC-like kinase 3), a cell cycle gene and member of protein kinase family is reported to be up-regulated in prostate cancer [Wang et. al., 2005]. *PFDN6*, a subunit of heteromeric prefoldin complex is primarily a chaperone that mediates actin and tubulin protein folding. Increased expression of *PFDN6* implies its role in inflammation and cancer [Ostrov et. al., 2007].

Lastly, we built a functional pathway based on a curated database of molecular interactions reported in the literature using IPA. The interaction pathway included the genes that are oncogenes and putative tumor suppressor with prominent role in cancer death signaling pathways such as *BCL2, CTTN, SOD2, REL, IFNG, TP73* and *SYK*. Of particular interest in the pathway is up-regulated expression of REL gene which encodes c-Rel, a transcription factor and member of the Rel/NFKB family, which also includes *RELA, RELB, NFKB1*, and *NFKB2*. *IFNG*, which directly does not induce *NF-kB*, strongly potentiates the ability of *TNF-α* to induce *NF-kB* nuclear translocation and

stimulate kB-dependent transcription [Cheshire et. al., 1997]. Interestingly, increased expression of *SOD-2* in the pathway is in contrast to the studies that report over-expression of *SOD2* inhibits cellular proliferation both in vitro and in vivo [Kim et. al., 2001]. Ho et al. also demonstrated elevated *SOD-2* levels in primary lung tumors compared with adjacent normal tissue [Ho et. al., 2001]. Similarly, elevated levels of *SOD2* correlated with an increased frequency of invasion and metastasis of gastric and colorectal carcinomas [Toh et. al., 2000, Malafa et. al., 2000]. Increased *SOD2* levels might be a response of tumors to the inflammatory cytokines and growth factors produced as a result of the host anti-tumoral immune reaction. Decreased expression of *BCL-2* as seen in the pathway would theoretically render a cell towards apoptosis. This seems paradoxical in view of the role of *BCL-2* as an inhibitor of apoptosis and whose expression is associated with better prognosis in NSCLC [Pezzella et. al., 1993], breast cancer [Silvestrini et. al., 1994] and colon cancer [Sinicrope et. al., 1995]. However, we cannot rule out the multi-functional role of Bcl-2 in cancer biology that is beyond its classical role in cell survival. Studies on gastric [Saegusa et. al., 1995] and breast adenocarcinomas [Silvestrini et. al., 1994, Gorczyca et. al. 1995] have demonstrated an association between increased *BCL-2* expression and decreased proliferative potential.

A salient finding of the study was the differential expression of 24 zinc finger proteins (ZNP) (10 up-regulated and 14 down-regulated) including a zinc transporter *SLC30A1*. Zinc homeostasis is critical in tumorigenesis and is maintained by transmembrane transporters belonging to the ZIP and ZnT families. ZnT transporters reduce intracellular zinc availability by promoting zinc efflux from cells or into intracellular vesicles [Palmiter et. al., 2003]. Up-regulation of *SLC30A1* indicates low concentration of zinc in tumor cells in the present study. In line with this observation Abnet et al. [2005] showed that high tissue zinc concentration was strongly associated with a reduced risk of developing esophageal tumors. Another study demonstrated inverse relation between dietary intake of zinc and risk of lung cancer [Mahabir et. al., 2007]. Zinc deficiency probably impairs host protective mechanisms against DNA damage ultimately increasing cancer risk [Ho et. al., 2004]. ZNPs form a DNA binding domain and thus are common structural constituent of transcription factors. We found over-expression of *ZNF300* which is shown to promote tumor growth and metastasis through activation of

NF-κB pathway [Wang et. al., 2012]. We also found an over expression of *PRDM1* or *Blimp-1* (B lymphocyte-induced maturation protein) in the study. However, Blimp-1, a known tumor suppressor gene in lymphoid malignancies is found to be frequently inactivated in B-cell lymphomas [Mandelbaum_et. al., 2010]. Other ZNPs like *ZNF281* or *ZBP-99* is a transcriptional repressor of genes including gastrin and ornithine decarboxylase whereas an over-expression of *ZDHHC9* decreases proliferation of colon cancer cells [Zhang_et. al., 2003, Mansilla_et. al., 2007]. We found down-regulated expression of both these genes implying their role in tumor progression.

In summary, the study revealed an alteration in gene expression primarily relating to cell cycle, cell death and MAPK signaling pathway. Also, deregulated expressions of transcription factors belonging to zinc finger proteins and zinc transporters was a characteristic feature of this study. Furthermore, we have identified genes like *TMSB10, RPS, PPFIA1, TNS3*, *NGFR, CLK3* and *PFDN6* for the first time as putative markers for NSCLC. Small sample numbers and hypothesis generating in-silico analysis limit definite conclusions, yet this preliminary analysis in the high risk population confirms to previous reports and presents some novel combinations that can be further explored upon. Further biological and clinical studies, probably including environment factors would be needed in order to assess the potential of these findings as diagnostic and prognostic markers of non small cell lung cancer.

# *Chapter 8*

# *Promoter methylation of p16, RASSF1A, DAPK and GSTP1 genes in lung cancer patients from high risk population of North East India*

## *Introduction*

Lung cancer is a multi-step process involving genetic and epigenetic alterations where resulting DNA damage transforms normal lung epithelial cells into lung cancer. Aberrant DNA methylation in the promoter region of many genes is the most well-defined epigenetic change associated with loss of gene expression which may confer tumor cells of growth advantage [Baylin et. al., 2000, Jones et. al., 1999]. Methylation is considered as an early event in lung tumorigenesis [Belinsky et. al., 2005]. Furthermore, variations in methylation status have been associated with cigarette smoke exposure [Belinsky et. al., 2005, Fujiwara et. al., 2005]. Results from previous studies have reported that promoter methylation of some genes occurred more frequently in lung tumors from smokers, compared with never-smokers [Toyooka et. al., 2003, Divine et. al., 2005, Liu et. al., 2006]. Inactivation of tumor suppressor genes and of genes important in metabolizing carcinogens is essential for lung tumorigenesis.

p16 is the most commonly altered gene in human malignancies and promoter methylation of the gene is an early and frequent event in NSCLC [Belinsky et. al., 1998]. The p16 protein is a molecular component of the retinoblastoma protein (pRB) regulatory pathway which inhibits G1 cyclin-dependent kinases. Thus inactivation of p16 facilitates phosphorylation of pRB that releases it from E2F transcription factor and allows progression of cells into S phase. Ras association domain family 1A functions as a tumor suppressor gene involved in cell apoptosis, genomic stability, and cell cycle regulation. Inactivation of RASSF1A has been reported in progression of lung cancer which is correlated with the hypermethylation of its CpG-island rich promoter region. The death-associated protein kinase (DAPK) is a pro-apoptotic serine/threonine kinase that suppresses tumor growth through apoptosis [Deiss et. al., 1995]. Expression of DAPK is frequently lost in many human cancers, often as a result of silencing by DNA methylation. The glutathione S-transferase P1 (GSTP1) is a phase II xenobiotic metabolizing gene that catalyzes the glutathione conjugation of many hydrophobic and electrophilic compounds.

Inactivation of the GSTP1 gene by promoter hypermethylation has been reported in human neoplasia including prostate, breast, renal, and lung tumors [Esteller et. al., 1998].

Most of the studies on risk and effects of gene methylation in lung cancer have been reported from the Western population, where etiology and genetic factors differ considerably from Asian populations. To our knowledge, there are no reports on methylation and lung cancer from Indian population. The ethnic North Eastern population, which reports as a cancer hub provides an excellent background for exploring the clinical and epidemiological significance of methylation in lung cancer.

The present study was designed to investigate the frequency of promoter hypermethylation in a panel of tumor suppressor genes (p16, RASSF1A, DAPK) and xenobiotic gene (GSTP1) playing important role in lung carcinogenesis. Their interaction with smoking, tobacco and betel quid chewing, and alcohol use were also analyzed. The study was carried out on lung tumor biopsy samples from high-risk North-eastern population of India. The aim of this study was to assess the association of methylation in these genes, reported to be frequently methylated in lung cancer in Western population, with the risk of lung cancer in our ethnic population.

## EXPERIMENTAL METHODS

### Materials

Agarose, Tris base, EDTA, other fine chemicals were purchased from Sigma Chemicals, USA. Platinum Taq polymerase, dNTPs, MgCl2, was obtained from Invitrogen and MBI Fermentas USA. RNA Later from Ambion (Austin, USA) and DNA extraction kit were purchased from Qiagen, (Hilden, Germany). Bisulphate modification of DNA was carried out by EZ methylation kit from Zymo Research Corp. (Orange, CA, USA). Oligos were synthesized by Microsynth, Switzerland.

### Patient recruitment and sample collection

Tissue specimens were obtained from patients undergoing computed tomography (CT) - guided biopsy at Dr. Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam and Civil Hospital, Aizawl between March 2008 and March 2010. Routine histopathology

analysis was done to confirm the diagnosis. Tumor tissue were collected in RNA Later (Ambion, Austin, USA), snap-frozen in liquid nitrogen and stored at -70$^0$C until processed.

**Inclusion criteria**

Patients with only lung as their primary site of cancer were included and those not undergoing chemotherapy at the time of recruitment.

**Exclusion criteria:** As described in Chapter 4

**Patient details**

A total of 57 lung cancer biopsies were collected. All subjects provided informed consent and the study was done under a protocol approved by the institutional ethics committee. Information on demographic characteristics, such as sex, age, smoking habit, usage of tobacco, betel quid and alcohol, were obtained from subjects in a standard questionnaire.

**DNA Isolation**

Genomic DNA was isolated from tumor tissues using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

**Quantification of Genomic DNA:** As described in Chapter 4

**Agarose Gel Electrophoresis of Extracted DNA Samples:** As described in Chapter 4

**Analysis of DNA methylation**

One microgram of genomic DNA was modified using EZ DNA Methylation Kit (Zymo Research Corp) according to the manufacturer's protocol. All bisulfite-converted DNA samples were stored at -20$^0$C until subsequent PCR was performed. Methylation Specific PCR was carried out using 100 ng of bisulfite treated DNA in a PCR mixture containing 16.6 mM ammonium sulfate, 67 mMTris (pH 8.8), 6.7 mM MgCl2, 10mMβ-mercaptoethanol, dNTPs each at 1.25 mM and primers each 1.6 μM in a 25 μl reaction. Two separate primer sets specific for methylated and unmethylated sequences of the genes were used to carry out two separate PCRs. Details of primer sequences, amplification

product base pair used and annealing temperature are given in Table 8.1. For positive and negative controls of the MSP, breast cancer cell lines (MCF-7, MDA-MB-231) or DNA from normal lymphocyte treated with SssI methyltransferase (New England BioLabs, Beverly, MA, USA) as a positive control, untreated normal lymphocyte as negative controls and water with no DNA template as a control for contamination were included in each experiment. After amplification, each PCR product was electrophoresed using a 2% agarose gel, stained with ethidium bromide, and visualized under UV illumination.

**Table 8.1: Details of primer sequences used in MSP**

| Gene | Sequence (5'-3') | Annealing Temperature | Size (bp) |
|---|---|---|---|
| P16 $M_F$ | TTATTAGAGGGTGGGGCGGATCGC | $65^0$C | 150 |
| P16 $M_R$ | GACCCCGAACCGCGACCGTAA | | |
| P16 $U_F$ | TTATTAGAGGGTGGGGTGGATTGT | $60^0$C | 151 |
| P16 $U_R$ | CAACCCCAAACCACAACCATAA | | |
| RASSF1A $M_F$ | GTTGGTATTCGTTGGGCGC | $56^0$C | 160 |
| RASSF1A $M_R$ | GCACCACGTATACGTAACG | | |
| RASSF1A $U_F$ | GGTTGTATTTGGTTGGAGTG | $56^0$C | 180 |
| RASSF1A $U_R$ | CTACAAACCTTTACACACAACA | | |
| DAPK $M_F$ | GGATAGTCGGATCGAGTTAACGTC | $60^0$C | 98 |
| DAPK $M_R$ | CCCTCCCAAACGCCGA | | |
| DAPK $U_F$ | GGAGGATAGTTGGATTGAGTTAATGTT | $60^0$C | 108 |
| DAPK $U_R$ | CAAATCCCTCCCAAACACCAA | | |
| GSTP1 $M_F$ | TTCGGGGTGTAGCGCTCGTC | $59^0$C | 91 |
| GSTP1 $M_R$ | GCCCCAATACTAAATCACGACG | | |
| GSTP1 $U_F$ | GATGTTTGGGGTGTAGTGGTTGTT | $59^0$C | 97 |
| GSTP1 $U_R$ | CCACCCCAATACTAAATCACAACA | | |

$M_F$, forward methylated primer; $M_R$, reverse methylated primer; $U_F$, forward unmethylated primers; $U_R$, reverse unmethylated primers.

### Statistical Analysis

The association of methylation of all gene promoters, demographic, histological and environmental variable with the lung cancer was estimated using Chi-square and Fisher's exact test. Multivariate logistic regression analysis was conducted to assess the effect of demographic and clinic pathological variables on methylation status of the promoter genes. Statistical analysis was performed using SPSS, version 19.0. P value less than 0.05 was considered as significant.

## RESULTS

Promoter hypermethylation of p16, RASSF1A, DAPK and GSTP1 genes was carried out on 57 lung tumor biopsies from high risk NE population. The distribution of clinical and demographic characteristics is summarized in Table 8.2. The study samples consisted of 77.2% males and 22.8% females. The mean age of the patients was 56.29 ± 8. Out of the total 57 LC samples, 24 (42.1%) sample showed hypermethylation of at least one of the gene; 15(26.3%) had two gene methylated, 7 (12%) sample had three gene methylated and only one sample showed methylation of all the four genes. The frequency of methylation for p16 was 45.6% (26/57), for RASSF1A was 40.4% (23/57), for DAPK was 40.4% (23/57) and for GSTP1 was 12.3% (7/57). No statistically significant correlation was found for the methylation status between each gene and were independent of each other ($p>0.05$, Data not shown). Distribution of promoter methylation for p16, RASSF1A, DAPK and GSTP1 did not differ between the age groups, gender, histology, smoking, tobacco chewing and alcohol consumption (P value >0.05). Only, frequency of methylated RASSF1A significantly differed in betel quid chewers and non chewers [72.2 % (39/54) vs. 27.8% (15/54), $p=0.05$].

**TABLE 8.2: Distribution of demographic characteristic and methylation frequencies in lung cancer**

| | | N=57 (%) |
|---|---|---|
| **Age (mean SD)** | | 56.29 ± 8.38 |
| **Age** | < 57 | 30(52.6) |
| | > 57 | 27(47.4) |
| **Gender** | Male | 44(77.2) |
| | Female | 13(22.8) |
| **Histology** | Adeno | 18(31.6) |
| | Squamous | 39(68.4) |
| **Smoking** | Non-smokers | 18(31.6) |
| | Smokers | 39(68.4) |
| **Tobacco chewing** | Non-chewers | 28(49.1) |
| | Chewers | 29(50.8) |
| **Betel quid chewing** | Non chewers | 16(28.1) |
| | Chewers | 41(71.9) |
| **Alcohol consumption** | Non-alcoholic | 40(70.2) |
| | Alcoholic | 17(29.8) |
| **P16** | Unmethylated | 31(54.4) |
| | Methylated | 26(45.6) |
| **RASSF1A** | Unmethylated | 34(59.6) |
| | Methylated | 23(40.4) |
| **DAPK** | Unmethylated | 34(59.6) |
| | Methylated | 23(40.4) |
| **GSTP1** | Unmethylated | 50(87.7) |
| | Methylated | 7(12.3) |
| **Minimum One gene** | Unmethylated | 13(16.0) |
| | Methylated | 68(84.0) |

**Logistic regression analysis of p16, RASSF1A, DAPK and GSTP1 promoter methylation**

Multivariate logistic regression models were employed to control for the potential confounding effects of variables, such as age, gender, tumor histology, smoking status, tobacco chewing, betel quid chewing, and alcohol drinking on methylation status of all the four genes analyzed in the study. Table 8.3 and Table 8.4 present the distribution and OR of multivariate logistic regression estimating the associations between the promoter methylation of the genes and clinicopathological parameters. Association between promoter methylation of p16 gene and smoking was associated with significantly increased risk of LC (OR = 6.68, 95% CI= 1.01–44.2, p= 0.04]. Also, RASSF1A promoter

methylation was associated with betel quid chewing (OR = 7.60, 95% CI= 1.26–45.74, p = 0.02). Smoking is a known risk factor for lung cancer thus increased risk was observed for associations of p16, RASSF1A and GSTP1 promoter methylation with smoking. Surprisingly, association with methylated DAPK yielded decreased risk to lung cancer although the results were not statistically significant (OR=0.49, 95% CI= 0.10-2.38, p=0.37).

**TABLE 8.3: Frequencies of promoter methylation and association with clinicopathological characteristics of lung cancer patients**

| | **Total** | **P16** | | **RASSF1A** | | **DAPK** | | **GSTP1** | |
|---|---|---|---|---|---|---|---|---|---|
| | 57 | U | M | U | M | U | M | U | M |
| | | 31 (%) | 26(%) | 34(%) | 23(%) | 34(%) | 23(%) | 50(%) | 7(%) |
| **Age** | | | | | | | | | |
| < 57 | 30(52.6) | 15(48.4) | 15(57.7) | 21(61.8) | 9(39.1) | 19(55.9) | 11(47.8) | 27(54.0) | 3(42.9) |
| > 57 | 27(47.4) | 16(51.6) | 11(42.3) | 13(38.2) | 14(60.9) | 15(44.1) | 12(52.2) | 23(46.0) | 4(57.1) |
| **Gender** | | | | | | | | | |
| Male | 44(77.2) | 24(77.4) | 20(76.9) | 27(79.4) | 17(73.9) | 25(73.5) | 19(82.6) | 38 (76.0) | 6(85.7) |
| Female | 13(22.8) | 7(22.6) | 6(23.1) | 7(20.6) | 6(26.1) | 9(26.4) | 4(17.4) | 12(24.0) | 1(14.3) |
| **Histology** | | | | | | | | | |
| Adeno | 18(31.6) | 9(29.0) | 9(34.6) | 12(35.3) | 6(26.1) | 12(35.3) | 6(26.1) | 16(32.0) | 2(28.6) |
| Squamous | 39(68.4) | 22(71.0) | 17(65.4) | 22(64.7) | 17(73.9) | 22(64.7) | 17(73.9) | 34(68.0) | 5(71.4) |
| **Smoking** | | | | | | | | | |
| Non-smokers | 18(31.6) | 13(41.9) | 5(19.2) | 12(35.3) | 6(26.1) | 10(29.4) | 8(34.8) | 16(32.0) | 2(28.6) |
| Smokers | 39(68.4) | 18(58.1) | 21(80.8) | 22(64.7) | 17(73.9) | 24(70.6) | 15(65.2) | 34(68.0) | 5(71.4) |
| **Tobacco** | | | | | | | | | |
| Non chewers | 28(49.2) | 17(54.8) | 11(42.3) | 15(44.1) | 13(56.5) | 16(47.1) | 12(52.2) | 26(52.0) | 2(28.6) |
| Chewers | 29(50.8) | 14(45.2) | 15(57.7) | 19(55.9) | 10(43.5) | 18(52.9) | 11(47.8) | 24(48.0) | 5(71.4) |
| **Betel quid** | | | | | | | | | |
| Non chewers | 16(28.1) | 10(32.3) | 6(23.1) | **14(41.2)** | **2(8.7)** | 9(26.5) | 7(30.4) | 12(24.0) | 4(57.1) |
| Chewers | 41(71.9) | 21(67.8) | 20(76.9) | **20(58.8)** | **21(91.3)*** | 25(73.5) | 16(69.6) | 38(76.0) | 3(42.9) |
| **Alcohol** | | | | | | | | | |
| Non alcoholic | 40 70.2) | 25(80.6) | 15(57.7) | 25(80.6) | 15(57.7) | 26(76.5) | 14(60.9) | 36(72.0) | 4(57.1) |
| Alcoholic | 17(29.8) | 6(19.4) | 11(42.3) | 9(26.5) | 8(34.8) | 8(23.5) | 9(39.1) | 14(28.0) | 3(42.9) |

* $\chi^2$ significant; $p<0.05$

**Table 8.4: Logistic regression models of p16, RASSF1A, DAPK and GSTP1 promoter methylation**

| Factors | P16 | RASSF1A | DAPK | GSTP1 |
|---|---|---|---|---|
| **Age** | 0.76(0.22-2.57)0.66 | 3.05(0.85-10.9)0.08 | 1.02(0.32-3.20)0.97 | 1.67(0.26-10.7)0.58 |
| **Gender** | 0.15(0.02-1.19)0.07 | 0.36(0.04-2.82)0.33 | 2.68(0.42-17.1)0.29 | 0.66(0.03-13.1)0.79 |
| **Histology** | 1.26(0.35-4.53)0.72 | 2.02(0.53-7.75)0.30 | 1.39(0.40-4.83)0.59 | 1.37(0.17-10.5)0.76 |
| **Smoking** | **6.68(1.01-44.2)0.04** | 1.74(0.28-10.6)0.54 | 0.49(0.10-2.38)0.37 | 2.02(0.15-25.7)0.58 |
| **Tobacco chewing** | 2.10(0.59-7.49)0.25 | 0.71(0.19-2.73)0.62 | 0.63(0.19-2.10)0.45 | 2.85(0.40-20.0)0.29 |
| **Betel quid chewing** | 0.82(0.20-3.38)0.79 | **7.60(1.26-45.74)0.02** | 0.91(0.24-3.43)0.89 | 0.15(0.02-1.03)0.05 |
| **Alcohol consumption** | 3.76(0.97-1.44)0.05 | 1.37(0.35-5.31)0.64 | 2.13(0.61-7.42)0.23 | 2.10(0.29-14.8)0.45 |

Bold number indicate significant p value < 0.05

## DISCUSSION

Promoter hypermethylation is a known cause in silencing and inactivation of associated genes and play important role in both the development and progression of cancer. Gene promoter hypermethylation is recognized as a major and causal event in lung cancer initiation and progression [Belinsky et. al., 2005]. Presence of a specific pattern of CpG island hypermethylation in human cancers was first reported by Costello et al. [Costello et. al., 2000], and confirmed by Esteller et al. [2001]. However, recent studies have also described variations in methylation with gender and ethnicity [Zhang et. al., 2011, Fraser et. al., 2012]. Recent evidences clearly suggest that divergence of methylation pattern occurs with population difference. This could be in part attributed to complex epistasis or gene environment interactions [Fraser et. al., 2012]. The present study was carried out on samples from ethnic NE population of India characterized by high incidence of cancer and also environmental influences. The NE population of India reports lung cancer incidence nearly ten times that of north India. In particular is the high rate of lung cancer among women in NE population. The area reports tobacco use in variety of ways of chewing and smoking that are different from the rest of India. Fermented betel nut and betel quids, betel nuts wrapped in betel leaf with slaked lime and tobacco, are some unique forms of tobacco consumption in the area [Sharan et. al., 2012]. Thus, we sought to determine the contribution of gene methylation and its interaction with environmental risk factors in NE population and severity of lung cancer.

The frequency of p16 promoter methylation observed in the present study was 45.6%. Frequencies of p16 hypermethylation in lung cancer vary widely from 27% reported in Korean population to 92% in Chinese females. In the current study, smokers, when compared to non-smokers, showed 6.68 times higher risk to p16 promoter methylation. Several studies have described an association between epigenetic alterations of the p16 gene in NSCLCs and tobacco smoking [Belinsky et. al., 1998, Kim et. al., 2001, Cespedes et. al., 2001, Belinsky et. al., 2002]. Kim et. al., [2001] reported that p16 promoter methylation was more common in tumors from smokers than from non-smokers.

Other studies from Asian subcontinent corroborate the findings [Yanagawa et. al., 2003; Nakata et. al., 2006]. Furthermore, results from a meta-analysis by Zhang et. al., [2011] suggested that cigarette smoking leading to p16 hypermethylation was related to the early stage progression of tumorigenesis in lung cancer. Cigarette smoking is known to be causally related to BPDE-DNA adducts that is elevated in the lung tissue of smokers [Lee et. al., 2008]. Also, aberrant methylation of the p16 gene was frequently detected in precursor lesions to lung tumors in rats that were treated with tobacco-specific 4-(methylnitrosamino)-I-(3- pyridyl)-1-butanone [Belinsky et. al., 1998]. Another study by Yanagawa et. al,. [2003] reaffirmed that tobacco smoking leads to inactivation of the p16 gene mainly through the epigenetic mechanism, ultimately increasing the risk of NSCLC.

Previous reports demonstrated that the frequency of RASSF1A methylation varies from 30 and 50% in solid tumors [Burbee et. al., 2001; Agathanggelou et. al., 2005]. In the present study the gene was methylated in 40% of the NE samples. RASSF1A is a candidate tumor suppressor gene at 3p21.3. Although the lack of expression of RASSF1A is common in lung cancer, mutations of RASSF1A are rare, therefore, RASSF1A gene is frequently inactivated in primary lung cancers by the de novo methylation of CpG islands in the promoter region [Dammann et. al., 2000; Burbee et. al., 2001]. Further we report, RASSF1A methylation association with betel quid chewers to be a significant 7 fold risk LC patients. Betel quid, which primarily constitutes phenolic compounds and alkaloids, was found to be a major factor for NE (71.9%) population. A recent study from Taiwan [Wen et. al., 2010] found risk to betel quid chewing in oral, lung, liver, pancreas and other cancers and its combination with smoking attributed to 50% of death among chewers. The molecular genetic mechanisms responsible for betel-associated carcinogenesis are poorly understood. Carcinogens derived from betel quid chewing may induce p53 mutation [Chiang et. al., 1999] and over-expression of c-myc protein [Baral et. al., 1998] with activated ras oncogene and subsequent over-expression of cell cycle regulatory protein, cyclin D1 [Kuo et. al., 1999]. Moreover, aqueous extract of betel-nut is shown to induce DNA-strand breaks and increase cell proliferation in vitro [Wary et. al., 1988]. However, because the RASSF1A is a tumor suppressor gene, it is likely that loss of RASSF1A

expression by promoter methylation does not require Ras activation. Thus, RASSF1A inactivation may induce malignant transformation by some distinct mechanisms other than Ras-mediated antiapoptosis pathway, such as loss of genomic, microtubule stability and cell cycle regulation [Agathanggelou et. al., 2005]. A high frequency of hypermethylation of p14, p15 and p16 was also detected in the precancerous lesions of betel quid chewers in Sri Lanka [Takeshima et al., 2008]. Further, it has been proposed that epigenetic silencing of RASSF1A and p16INK4a gene expressions by promoter hypermethylation may play critical roles in betel nut associated oral cancer [Tran et al., 2005]. Moreover, the absence of MGMT expression associated with promoter hypermethylation has been reported to be related to betel quid chewing in oral cancer [Huang et al., 2010]. The mechanism for betel chewing inducing gene-specific hypermethylation in different cancer remains unclear.

DAPK plays a critical role in apoptosis regulation in tumor development, and is commonly hypermethylated in many cancers [Narayan et. al., 2003; Dulaimi et. al., 2004; Jabłonowski et. al., 2011; Sapari et. al., 2012]. In the present study methylation frequency of DAPK was 40.4% in NE cases. Promoter methylation of the DAPK gene has been reported in 20% to 40% of NSCLC [Tang et al., 2000; Kim et al., 2001]. According to Licchesi et al., [2008] (ref 43), DAPK promoter hypermethylation is absent in normal lung but is present in 39% of the adenocarcinomas. In mouse models, DAPK promoter hypermethylation is detected in almost half of Atypical Alveolar Hyperplasias (AAHs) induced by chronic exposure to 4-methylnitrosamino-1-(3-pyridyl)-1- butanone, a component of tobacco smoke [Pulling et al., 2004].

The rationale of studying GSTP1 methylation was based on reports that GSTP1 is the most abundant GST isoform in human lung [Anttila et. al., 1993, Wang et. al., 2003]. Besides PAHs and other tobacco smoke carcinogens, cisplatin, a common agent used in lung cancer treatment is also a substrate for GSTP1 [Goto et. al., 1999]. Over-expression of the gene is reported in lung tumors and is associated with failure of cancer chemotherapy and low patient survival rates [Waxman et. al., 1990, Morrow et. al., 1990, Tew et. al., 1994]. GSTP1 is found to be hypermethylated in 7–9% of NSCLCs [Esteller

et. al., 1998 and 2001, Zochbauer-Muller 2001]. Interestingly, The present study reports the frequency of methylation in NE population was higher (12.3%) than reported earlier. However, we did not observe any association of GSTP1 promoter methylation with lung cancer in any population or any risk factor. Our results thus corroborate earlier findings that methylation of GSTP1 is a tumor-specific event largely associated with prostate, breast and renal carcinomas [Lee et. al., 1994, Esteller et. al., 1998].

In conclusion, the present study identified association of methylation and interactions with environmental factors in the study populations. Increased risk of LC was associated with promoter methylation of p16 and RASSF1A gene in smokers and betel quid chewers of NE population. These results should be interpreted with cautions as our patients were mixture of smokers, alcoholic, and betel quid chewers. Result could be synergistic effect of all the three risk habits. The genetic pathway responsible for betel-associated LC without the concomitant effects of ethanol and tobacco may be different from the smoker/alcohol drinkers. However, this is the first report from the study populations examining, for lung cancer, methylation frequencies and its association with risk factors based on ethnicity. These results may be useful for the future studies on betel quid and smoking-related epigenetic changes in large number of patients in lung cancer from North East region of India.

# *Chapter 9*

# *Conclusions and Future Scope of Work*

# *Conclusions and Future Scope of Work*

The main focus of the thesis was to assess the role of genetic and epigenetic alterations in the risk assessment of the lung cancer patients from NE region of the India. The following conclusion can be drawn:

- Smoking and betel quid chewing were identified as major environmental risk factors.
- Logistic regression analysis showed that CYP1A1*2A polymorphism was significantly associated with increased lung cancer risk whereas EPHX1 Tyr113His and SULT1A1 Arg213His conferred reduced risk.
- On stratification with smoking, EPHX1 Tyr113His and SULT1A1 Arg213His polymorphisms reduced the risk of lung cancer in smokers, whereas CYP1A1*2A, CYP1A1*2C and GSTP1 Ile105Val imparted increased risk in non-smokers.
- Data mining approaches reveal significant high order gene-gene and gene-environment interaction of xenobiotic metabolizing genes.
- CART analysis identified combination of EPHX1 Tyr/His, SULT1A1 Arg/Arg or His/His and GSTM1 null genotypes while MDR analysis identified combination of tobacco chewing, EPHX1 Tyr113His, and SULT1A1 Arg213His conferring highest risk of lung cancer among smokers.
- Among non-smokers, combination of CYP1A1*2A CC or TC, SULT1A1 Arg/Arg and betel quid chewing through CART analysis and combination of CYP1A1*2A, GSTP1 Ile105Val and SULT1A1 Arg213His through MDR analysis conferred maximum risk of lung caner.

- An increased lung cancer risk was observed for betel quid chewers compared to non-betel quid chewers for p53 codon 72 polymorphism.
- Gene dosage analysis of GSTT1 and GSTM1 genes showed decrease risk of lung cancer associated with hemizygous and null genotype of GSTT1 in smokers.
- Gene expression profiling showed 734 differentially expressed genes (≥1.5 fold, $p<0.05$). Of these, 311 genes were overexpressed and 423 were underexpressed. A signature of 24 (10 up-regulated and 14 down-regulated) differentially expressed zinc finger proteins was identified.
- Gene enrichment analysis and pathway analysis of expression data identified terms related to epidermal growth factor, homeobox related transcription activity terms and MAPK signaling in DAVID and network related to cell death, cancer, cell cycle and cellular assembly and organization in IPA analysis.
- Novel genes *viz*. *TMSB10, RPS, PPFIA1, TNS3, NGFR, CLK3* and PFDN6 were found to be associated with NSCLC.
- Methylation analysis showed promoter of p16 gene was the most frequently methylated followed by RASSF1A, DAPK and GSTP1 and also reports the higher frequency of GSTP1 promoter methylation in the population as compared to earlier studies.
- Results suggest an association of p16 and RASSF1A promoter methylation with smoking and betel quid chewing with increased risk of lung cancer.

**Significance of the Study**

This case-control study attempts to determine the presence of known genetic variations of candidate genes that may be implicated in the pathogenesis of lung cancer. Identifying distinct gene-gene and gene-environment interactions in lung cancer patients with different risk habits help us in understanding the molecular mechanisms involved in the pathophysiology of the disease. In addition to this, it confirms the role of multifactorial

interaction in risk assessment of lung cancer patients, which may help in recognizing those at the risk of developing the disease and may lead to new preventive approaches. The study also highlights some of the opportunities and challenges, which may be encountered in interpreting the value of the role of methylation to improve the management of lung cancer. Last, the expression analysis of NSCLC cases reports the novel markers found in the study which could serve as a putative biomarker for the diagnosis of disease.

**Future Scope of the Thesis**

This is the first study on lung cancer from high-risk region of North East India which explored genetic and epigenetic variations and gene expression profiling of lung cancer associated with environmental risk habits. The study showed increased risk of lung cancer associated with combination of EPHX1 Tyr/His, SULT1A1 Arg/Arg or His/His and GSTM1 null genotypes in smokers and CYP1A1*2A, GSTP1 Ile105Val and SULT1A1 Arg213His in non-smokers. However, much remains to clarify the functional consequences of above combination of gene interactions in lung cancer. These results may serve as good reference for future studies. Epigenetic changes are equally responsible as genetic changes for the development and progression of lung cancer. Higher frequency of promoter methylation for the p16 gene in smoker and RASSF1A gene in betel quid chewers suggest that these observations may be useful for the future study of smoking and betel quid related epigenetic changes in lung carcinogenesis. It will be of interest to further investigate the methylation differences between smokers and betel quid chewers in other genes that have been found to be frequently hypermethylated in lung tumors. Furthermore, novel genes *viz*. *TMSB10, RPS, PPFIA1, TNS3, NGFR, CLK3* and PFDN6 identified in NSCLC cases could serve as a potential diagnostic and therapeutic target for lung cancer. However, future studies on large sample size are needed to prospectively evaluate the validity of these molecular biomarkers at protein level also.

# *References*

# References

Abnet CC, Lai B, Qiao YL, Vogt S, Luo XM, Taylor PR, Dong ZW, Mark SD, Dawsey SM. Zinc concentration in esophageal biopsy specimens measured by x-ray fluorescence and esophageal cancer risk. J Natl Cancer Inst. 2005 Feb 16;97(4):301-6.

Adler V, Yin Z, Fuchs SY, Benezra M, Rosario L, Tew KD, Pincus MR, Sardana M, et al. Regulation of JNK signaling by GSTp. EMBO J. 1999 Mar 1;18(5):1321-34.

Agathanggelou A, Cooper WN, Latif F. Role of the Ras-association domain family 1 tumor suppressor gene in human cancers. Cancer Res. 2005 May 1;65(9):3497-508.

Agathanggelou A, Honorio S, Macartney DP, Martinez A, Dallol A, Rader J, et al. Methylation associated inactivation of RASSF1A from region 3p21.3 in lung, breast and ovarian tumours. Oncogene. 2001 Mar 22;20(12):1509-18.

Alderson MR, Rattan NS, Bidstrup L. Health of workmen in the chromate-producing industry in Britain. Br J Ind Med. 1981 May;38(2):117-24.

Alexandrie AK, Nyberg F, Warholm M, Rannug A. Influence of CYP1A1, GSTM1, GSTT1, and NQO1 genotypes and cumulative smoking dose on lung cancer risk in a Swedish population. Cancer Epidemiol Biomarkers Prev. 2004 Jun;13(6):908-914.

Ali-Osman F, Akande O, Antoun G, Mao JX, Buolamwini J. Molecular cloning, characterization, and expression in Escherichia coli of full-length cDNAs of three human glutathione S-transferase Pi gene variants. Evidence for differential catalytic activity of the encoded proteins. J Biol Chem. 1997 Apr 11;272(15):10004-10012.

Alpízar-Alpízar W, Sierra R, Cuenca P, Une C, Mena F, Pérez-Pérez GI. [Association of the p53 codon 72 polymorphism to gastric cancer risk in a hight risk population of Costa Rica]. Rev Biol Trop. 2005 Sep-Dec;53(3-4):317-324.

Altorki N, Lane ME, Bauer T, Lee PC, Guarino MJ, Pass H, Felip E, Peylan-Ramu N, et al. Phase II proof-of-concept study of pazopanib monotherapy in treatment-naive patients with stage I/II resectable non-small-cell lung cancer. J Clin Oncol. 2010 Jul 1;28(19):3131-7.

Anderson KE, Kliris J, Murphy L, Carmella SG, Han S, Link C, Bliss RL, Puumala S, Murphy SE, Hecht SS. Metabolites of a tobacco-specific lung carcinogen in nonsmoking casino patrons. Cancer Epidemiol Biomarkers Prev. 2003 Dec;12(12):1544-6.

Anguiano A, Nevins JR, Potti A. Toward the individualization of lung cancer therapy. Cancer. 2008 Oct 1;113(7 Suppl):1760-7.

Antequera F, Bird A. CpG islands. EXS 1993;64:169-185

Anttila S, Hirvonen A, Vainio H, Husgafvel-Pursiainen K, Hayes JD, Ketterer B. Immunohistochemical localization of glutathione S-transferases in human lung. Cancer Res. 1993 Dec 1;53(23):5643-8

Ariel M, Robinson E, McCarrey JR, Cedar H. Gamete-specific methylation correlates with imprinting of the murine Xist gene. Nat Genet. 1995 Mar;9(3):312-5.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9.
Ataka M, Ikeguchi M, Yamamoto M, Inoue M, Tanida T, et. al., Topoisomerase I Protein Expression and Prognosis of Patients with Colorectal Cancer. Yonago Acta medica 2007;50:81–87.

Awang MN. Betel quid and oral carcinogenesis.Singapore Med J. 1988 Dec;29(6):589-593.

Bamber DE, Fryer AA, Strange RC, Elder JB, Deakin M, Rajagopal R, Fawole A, Gilissen RA, Campbell FC, Coughtrie MW. Phenol sulphotransferase SULT1A1*1 genotype is associated with reduced risk of colorectal cancer. Pharmacogenetics. 2001 Nov;11(8):679-685.

Bamber DE, Fryer AA, Strange RC, Elder JB, Deakin M, Rajagopal R, Fawole A, Gilissen RA, Campbell FC, Coughtrie MW. Phenol sulphotransferase SULT1A1*1 genotype is associated with reduced risk of colorectal cancer. Pharmacogenetics. 2001 Nov;11(8):679-85.

Baral R, Patnaik S, Das BR. Co-overexpression of p53 and c-myc proteins linked with advanced stages of betel- and tobacco-related oral squamous cell carcinomas from eastern India. Eur J Oral Sci. 1998 Oct;106(5):907-13.

Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R. NCBI GEO: mining millions of expression profiles--database and tools. Nucleic Acids Res. 2005 Jan 1;33 D562-6.

Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum Mol Genet. 2001 Apr;10(7):687-92.

Baylin SB, Herman JG, Graff JR, Vertino PM, Issa JP. Alterations in DNA methylation: a fundamental aspect of neoplasia. Adv Cancer Res. 1998;72:141-96.

Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet. 2000 Apr;16(4):168-74.

Beckman G, Birgander R, Själander A, Saha N, Holmberg PA, Kivelä A, Beckman L. Is p53 polymorphism maintained by natural selection? Hum Hered. 1994 Sep-Oct;44(5):266-70.

Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med. 2002 Aug;8(8):816-24.

Behera D, Balamugesh T. Indoor air pollution as a risk factor for lung cancer in women. J Assoc Physicians India. 2005 Mar;53:190-2.

Behera D, Sharma A, Khanduja KL, Gogna ML. Beta carotene, vitamin-A, and vitamin-C levels in patients with lung cancer. (Abstract) *Lung Cancer* 1998;21(Supp 1):S20

Belinsky SA, Klinge DM, Dekker JD, Smith MW, Bocklage TJ, Gilliland FD, Crowell RE, Karp DD, Stidley CA, Picchi MA. Gene promoter methylation in plasma and sputum increases with lung cancer risk. Clin Cancer Res. 2005 Sep 15;11(18):6505-11.

Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, Baylin SB, Herman JG. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci U S A. 1998 Sep 29;95(20):11891-6.

Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM, Moots PP, Lechner JF, Stidley CA, Crowell RE. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res. 2002 Apr 15;62(8):2370-7.

Belinsky SA. Silencing of genes by promoter hypermethylation: key event in rodent and human lung cancer. Carcinogenesis. 2005 Sep;26(9):1481-7.

Benhamou S, Lee WJ, Alexandrie AK, Boffetta P, Bouchardy C, Butkiewicz D, et al. Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. Carcinogenesis. 2002 Aug;23(8):1343-50.

Benhamou S, Reinikainen M, Bouchardy C, Dayer P, Hirvonen A. Association between lung cancer and microsomal epoxide hydrolase genotypes. Cancer Res. 1998 Dec 1;58(23):5291-5293.

Benvenisty N, Razin A, Reshef L. Developmental changes in the methylation pattern, chromatin conformation and expression of the rat phosphoenolpyruvate carboxykinase gene. Prog Clin Biol Res. 1985;198:185-99.

Bergamaschi D, Gasco M, Hiller L, Sullivan A, Syed N, Trigiante G, Yulug I, Merlano M, Numico G, Comino A, Attard M, Reelfs O, Gusterson B, Bell AK, Heath V, Tavassoli M, Farrell PJ, Smith P, Lu X, Crook T. p53 polymorphism influences response in cancer chemotherapy via modulation of p73-dependent apoptosis. Cancer Cell. 2003 Apr;3(4):387-402.

Berhane K, Widersten M, Engström A, Kozarich JW, Mannervik B. Detoxication of base propenals and other alpha, beta-unsaturated aldehyde products of radical reactions and lipid peroxidation by human glutathione transferases. Proc Natl Acad Sci U S A. 1994 Feb 15;91(4):1480-4.

Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A. 2001 Nov 20;98(24):13790-5.

Bidstrup Pl, Case Ra. Carcinoma of the lung in workmen in the bichromates-producing industry in Great Britain. Br J Ind Med. 1956 Oct;13(4):260-4.

Bing Z, Pasha TL, Lal P, Tomaszewski JE. Expression of glial cell line-derived neurotropic factor receptor alpha-1 in immature teratomas. Am J Clin Pathol. 2008 Dec;130(6):892-6.

Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002 Jan 1;16(1):6-21.

Birgander R, Själander A, Zhou Z, Fan C, Beckman L, Beckman G. p53 polymorphisms and haplotypes in nasopharyngeal cancer. Hum Hered. 1996 Jan-Feb;46(1):49-54.

Boffetta P, Nyberg F. Contribution of environmental factors to cancer risk. Br Med Bull. 2003;68:71-94.

Boffetta P. Involuntary smoking and lung cancer. Scand J Work Environ Health. 2002;28 Suppl 2:30-40.

Bonafé M, Ceccarelli C, Farabegoli F, Santini D, Taffurelli M, Barbi C, Marzi E, Trapassi C, Storci G, Olivieri F, Franceschi C.Retention of the p53 codon 72 arginine allele is associated with a reduction of disease-free and overall survival in arginine/proline heterozygous breast cancer patients. Clin Cancer Res. 2003 Oct 15;9(13):4860-4864.

Borczuk AC, Gorenstein L, Walter KL, Assaad AA, Wang L, Powell CA. Non-small-cell lung cancer molecular signatures recapitulate lung developmental pathways. Am J Pathol 2003;163:1949-60.

Borczuk AC, Shah L, Pearson GD, Walter KL, Wang L, Austin JH, Friedman RA, Powell CA. Molecular signatures in biopsy specimens of lung cancer. Am J Respir Crit Care Med. 2004 Jul 15;170(2):167-74.

Bowman RV, Yang IA, Semmler AB, Fong KM. Epigenetics of lung cancer. Respirology. 2006 Jul;11(4):355-65.

Bray FI, Weiderpass E. Lung cancer mortality trends in 36 European countries: secular trends and birth cohort patterns by sex and region 1970-2007. Int J Cancer. 2010 Mar 15;126(6):1454-66.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet. 2001 Dec;29(4):365-71.

Breitling R, Amtmann A, Herzyk P. Iterative Group Analysis (iGA): a simple tool to enhance sensitivity and facilitate interpretation of microarray experiments. BMC Bioinformatics. 2004 Mar 29;5:34.

Brennan P, Buffler PA, Reynolds P, Wu AH, Wichmann HE, Agudo A, Pershagen G, Jöckel KH, Benhamou S, Greenberg RS, Merletti F, Winck C, Fontham ET, Kreuzer M, Darby SC, Forastiere F, Simonato L, Boffetta P. Secondhand smoke exposure in adulthood and risk of lung cancer among never smokers: a pooled analysis of two large studies. Int J Cancer. 2004 Mar;109(1):125-31.

Breuer RH, Postmus PE, Smit EF. Molecular pathology of non-small-cell lung cancer. Respiration. 2005 May-Jun;72(3):313-30.

Buch SC, Notani PN, Bhisey RA. Polymorphism at GSTM1, GSTM3 and GSTT1 gene loci and susceptibility to oral cancer in an Indian population. Carcinogenesis. 2002 May;23(5):803-807.

Buchard A, Sanchez JJ, Dalhoff K, Morling N. Multiplex PCR detection of GSTM1, GSTT1, and GSTP1 gene variants: simultaneously detecting GSTM1 and GSTT1 gene copy number and the allelic status of the GSTP1 Ile105Val genetic variant. J Mol Diagn. 2007 Nov;9(5):612-617.

Burbee DG, Forgacs E, Zöchbauer-Müller S, Shivakumar L, Fong K, Gao B, Randle D, Kondo M, Virmani A, Bader S, Sekido Y, Latif F, Milchgrub S, Toyooka S, Gazdar AF, Lerman MI, Zabarovsky E, White M, Minna JDEpigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. J Natl Cancer Inst. 2001 May 2;93(9):691-9

Cáceres DD, Quiñones LA, Schroeder JC, Gil LD, Irarrázabal CE. Association between p53 codon 72 genetic polymorphism and tobacco use and lung cancer risk. Lung. 2009 Mar-Apr;187(2):110-115.

Carlsten C, Sagoo GS, Frodsham AJ, Burke W, Higgins JP. Glutathione S-transferase M1 (GSTM1) polymorphisms and lung cancer: a literature-based systematic HuGE review and meta-analysis. Am J Epidemiol. 2008 Apr 1;167(7):759-774.

Carmella SG, Akerkar SA, Richie JP Jr, Hecht SS. Intraindividual and interindividual differences in metabolites of the tobacco-specific lung carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) in smokers' urine. Cancer Epidemiol Biomarkers Prev. 1995 Sep;4(6):635-42.

Carvalho B, Ouwerkerk E, Meijer GA, Ylstra B. High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. J Clin Pathol. 2004 Jun;57(6):644-6

Cassidy A, 't Mannetje A, van Tongeren M, Field JK, Zaridze D, Szeszenia-Dabrowska N, et.al. Occupational exposure to crystalline silica and risk of lung cancer: a multicenter case-control study in Europe. Epidemiology. 2007 Jan;18(1):36-43.

Chaillet JR, Vogt TF, Beier DR, Leder P. Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell. 1991 Jul 12;66(1):77-83.

Chan-Yeung M, Dimich-Ward H. Respiratory health effects of exposure to environmental tobacco smoke. Respirology. 2003 Jun;8(2):131-9.

Chaturvedi HK, Phukan RK, Zoramtharga K, Hazarika NC, Mahanta J. Tobacco use in Mizoram, India: sociodemographic differences in pattern. Southeast Asian J Trop Med Public Health. 1998 Mar;29(1):66-70.

Chen K, Jiang QT, Ma XY, Yao KY, Leng SG, Yu WP, Zhou HG. [Associations between genetic polymorphisms of glutathione S-transferase M1 and T1, smoking and susceptibility to colorectal cancer: a case-control study]. Zhonghua Zhong Liu Za Zhi. 2004 Nov;26(11):645-648.

Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA hypomethylation leads to elevated mutation rates. Nature. 1998 Sep 3;395(6697):89-93.

Cheshire JL, Baldwin AS Jr. Synergistic activation of NF-kappaB by tumor necrosis factor alpha and gamma interferon via enhanced I kappaB alpha degradation and de novo I kappaBbeta degradation. Mol Cell Biol. 1997 Nov;17(11):6746-54.

Chiang CP, Huang JS, Wang JT, Liu BY, Kuo YS, Hahn LJ, Kuo MY. Expression of p53 protein correlates with decreased survival in patients with areca quid chewing and smoking-associated oral squamous cell carcinomas in Taiwan. J Oral Pathol Med. 1999 Feb;28(2):72-6.

Chua YL, Ito Y, Pole JC, Newman S, Chin SF, Stein RC, Ellis IO, Caldas C, O'Hare MJ, Murrell A, Edwards PA. The NRG1 gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene. Oncogene. 2009 Nov 19;28(46):4041-52.

Chung-man Ho J, Zheng S, Comhair SA, Farver C, Erzurum SC. Differential expression of manganese superoxide dismutase and catalase in lung cancer. Cancer Res. 2001 Dec 1;61(23):8578-85.

Churg A, Stevens B. Enhanced retention of asbestos fibers in the airways of human smokers. Am J Respir Crit Care Med. 1995 May;151(5):1409-13.

Cleveland W. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Association*. 1979 74, 829-36.

Cohen O, Inbal B, Kissil JL, Raveh T, Berissi H, Spivak-Kroizaman T, Feinstein E, Kimchi A. DAP-kinase participates in TNF-alpha- and Fas-induced apoptosis and its function requires the death domain. J Cell Biol. 1999 Jul 12;146(1):141-8.

Coles B, Yang M, Lang NP, Kadlubar FF. Expression of hGSTP1 alleles in human lung and catalytic activity of the native protein variants towards 1-chloro-2,4-dinitrobenzene, 4-vinylpyridine and (+)-anti benzo[a]pyrene-7,8-diol-9,10-oxide. Cancer Lett. 2000 Aug 11;156(2):167-175.

Colot V, Rossignol JL. Eukaryotic DNA methylation as an evolutionary device. Bioessays. 1999 May;21(5):402-11.

Cortessis V, Siegmund K, Chen Q, et al. A case-control study of microsomal epoxide hydrolase, smoking, meat consumption, glutathione S-transferase M3, and risk of colorectal adenomas. Cancer Res 2001;61:2381 – 5.

Cosma G, Crofts F, Taioli E, Toniolo P, Garte S. Relationship between genotype and function of the human CYP1A1 gene. J Toxicol Environ Health. 1993 Oct-Nov;40(2-3):309-316.

Costello JF, Frühwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomäki P, Lang JC, Schuller DE, Yu L, Bloomfield CD, Caligiuri MA, Yates A, Nishikawa R, Su Huang H, Petrelli NJ, Zhang X, O'Dorisio MS, Held WA, Cavenee WK, Plass C. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat Genet. 2000 Feb;24(2):132-8.

Cote ML, Chen W, Smith DW, Benhamou S, Bouchardy C, Butkiewicz D, Fong KM, Gené M, Hirvonen A, Kiyohara C, Larsen JE, Lin P, Raaschou-Nielsen O, Povey AC, Reszka E, Risch A, Schneider J, Schwartz AG, Sorensen M, To-Figueras J, Tokudome S, Pu Y, Yang P, Wenzlaff AS, Wikman H, Taioli E. Meta- and pooled analysis of GSTP1 polymorphism and lung cancer: a HuGE-GSEC review. Am J Epidemiol. 2009 Apr 1;169(7):802-814.

Covault J, Abreu C, Kranzler H, Oncken C. Quantitative real-time PCR for gene dosage determinations in microdeletion genotypes. Biotechniques. 2003 Sep;35(3):594-6, 598.

D'Errico A, Malats N, Vineis P, Boffetta P: Review of studies of selected metabolic polymorphisms and cancer. In Metabolic Polymorhisms and Susceptibility to Cancer. Edited by: Vienis P, Malats N, Lang M, D'Errico A, Caporaso N, Boffetta P. IARC Scientific publication No 148. Lyon: International Agency for Research on Cancer; 1999

Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat Genet. 2000 Jul;25(3):315-9.

Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, Deo H, Falk R, et al., Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ. 2005 Jan 29;330(7485):223.

DC, Wain JC, Mark EJ, Kelsey KT. p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. Cancer Res. 2001 Apr 15;61(8):3419-24.

Deiss LP, Feinstein E, Berissi H, Cohen O, Kimchi A Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. Genes Dev. 1995 Jan 1;9(1):15-30.

Demsar J, Zupan B. Orange: From Experimental Machine Learning to Interactive Data Mining. White Paper, Faculty of Computer and Information Science, University of Ljublijana, Ljublijana, Slovenia. Omnipress, Madison, WI 2004.

Denissenko MF, Pao A, Tang M, Pfeifer GP. Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. Science. 1996 Oct 18;274(5286):430-432.

Derks S, Lentjes MH, Hellebrekers DM, de Bruïne AP, Herman JG, van Engeland M. Methylation-specific PCR unraveled. Cell Oncol. 2004;26(5-6):291-9.

Divine KK, Pulling LC, Marron-Terada PG, Liechty KC, Kang T, Schwartz AG, Bocklage TJ, Coons TA, Gilliland FD, Belinsky SA Multiplicity of abnormal promoter methylation in lung adenocarcinomas from smokers and never smokers. Int J Cancer. 2005 Apr 10;114(3):400-5.

Dokianakis DN, Spandidos DA. P53 codon 72 polymorphism as a risk factor in the development of HPV-associated cervical cancer. Mol Cell Biol Res Commun. 2000 Feb;3(2):111-114.

Doll R, Fisher Re, Gammon Ej, Gunn W, Hughes Go, Tyrer Fh, Wilson W. Mortality Of Gasworkers With Special Reference To Cancers Of The Lung And Bladder, Chronic Bronchitis, And Pneumoconiosis. Br J Ind Med. 1965 Jan;22:1-12.

Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. J Epidemiol Community Health. 1978 Dec;32(4):303-13.

Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst. 1981 Jun;66(6):1191-308.

Doll R. Cancer of the lung and nose in nickel workers. Br J Ind Med. 1958 Oct;15(4):217-23.

Drummond SN, De Marco L, Pordeus Ide A, Barbosa AA, Gomez RS. TP53 codon 72 polymorphism in oral squamous cell carcinoma. Anticancer Res. 2002 Nov-Dec;22(6A):3379-3381.

Dulaimi E, Hillinck J, Ibanez de Caceres I, Al-Saleem T, Cairns P. Tumor suppressor gene promoter hypermethylation in serum of breast cancer patients. Clin Cancer Res. 2004 Sep 15;10(18 Pt 1):6189-93.

Dumont P, Leu JI, Della Pietra AC 3rd, George DL, Murphy M. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nat Genet. 2003 Mar;33(3):357-65.

Esteller M, Corn PG, Baylin SB, Herman JG A gene hypermethylation profile of human cancer. Cancer Res. 2001 Apr 15;61(8):3225-9.

Esteller M, Corn PG, Urena JM, Gabrielson E, Baylin SB, Herman JG. Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. Cancer Res. 1998 Oct 15;58(20):4515-8.

Esteller M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene. 2002 Aug 12;21(35):5427-40.

Ezzati M, Henley SJ, Lopez AD, Thun MJ. Role of smoking in global and regional cancer epidemiology: current patterns and data needs. Int J Cancer. 2005 Oct 10; 116(6):963-71

Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. Lancet. 2003 Sep 13;362(9387):847-52.

Fan R, Wu MT, Miller D, Wain JC, Kelsey KT, Wiencke JK, Christiani DC. The p53 codon 72 polymorphism and lung cancer risk. Cancer Epidemiol Biomarkers Prev. 2000 Oct;9(10):1037-1042.

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983 Jan 6;301(5895):89-92

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, et al. (2010) GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer. Available: http://globocan.iarc.fr Accessed 20 June 2011.

Fernández-Rubio A, López-Cima MF, González-Arriaga P, García-Castro L, Pascual T, Marrón MG, Tardón A. The TP53 Arg72Pro polymorphism and lung cancer risk in a population of Northern Spain. Lung Cancer. 2008 Sep;61(3):309-316.

Fraser HB, Lam LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. Genome Biol. 2012 Feb 9;13(2):R8. doi: 10.1186/gb-2012-13-2-r8.

Friberg L, Cederlöf R. Late effects of air pollution with special reference to lung cancer. Environ Health Perspect. 1978 Feb;22:45-66.

Fujiwara K, Fujimoto N, Tabata M, Nishii K, Matsuo K, Hotta K, Kozuki T, Aoe M, Kiura K, Ueoka H, Tanimoto M. Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. Clin Cancer Res. 2005 Feb 1;11(3):1219-25.

Gajalakshmi V, Peto R, Kanaka TS, Jha P. Smoking and mortality from tuberculosis and other diseases in India: retrospective study of 43000 adult male deaths and 35000 controls. Lancet. 2003 Aug 16;362(9383):507-15.

Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, et al. Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A. 2001 Nov 20;98(24):13784-9.

Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL, et al. Metabolic gene polymorphism frequencies in control populations. Cancer Epidemiol Biomarkers Prev. 2001 Dec;10(12):1239-48.

Gazdar AF. The molecular and cellular basis of human lung cancer. Anticancer Res. 1994 Jan-Feb;14(1B):261-7.

Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. Nucleic Acids Res. 2006 Jan 1;34 D322-6.

Germ line polymorphisms of the tumor suppressor gene p53 and lung cancer. Biros E, Kalina I, Kohút A, Stubna J, Salagovic J. Lung Cancer. 2001 Feb-Mar;31(2-3):157-162.

Gilbert L, Elwood LJ, Merino M, Masood S, Barnes R, Steinberg SM, Lazarous DF, et al. A pilot study of pi-class glutathione S-transferase expression in breast cancer: correlation with estrogen receptor expression and prognosis in node-negative breast cancer. J Clin Oncol. 1993 Jan;11(1):49-58.

Gimm O, Gössling A, Marsh DJ, Dahia PL, Mulligan LM, von Deimling A, Eng C. Mutation and deletion analysis of GFR alpha-1, encoding the co-receptor for the GDNF/RET complex, in human brain tumours. Br J Cancer. 1999 May;80(3-4):383-6.

Glatt H, Engelke CE, Pabel U, Teubner W, Jones AL, Coughtrie MW, Andrae U, Falany CN, Meinl W. Sulfotransferases: genetics and role in toxicology. Toxicol Lett. 2000 Mar 15;112-113:341-8.

Glatt H. Sulfation and sulfotransferases 4: bioactivation of mutagens via sulfation. FASEB J. 1997 Apr;11(5):314-321

GLOBOCAN 2002 , 2008 database, International Agency for Research on Cancer, World Health Organization. Available at http://www-dep.iarc.fr/, last accessed January 27, 2010. Available at http://globocan.iarc.fr/. last accessed august 10, 2010.

Goode EL, Potter JD, Bamlet WR, Rider DN, Bigler J. Inherited variation in carcinogen-metabolizing enzymes and risk of colorectal polyps. Carcinogenesis. 2007 Feb;28(2):328-41.

Gorczyca W, Markiewski M, Kram A, Tuziak T, Domagala W. Immunohistochemical analysis of bcl-2 and p53 expression in breast carcinomas: their correlation with Ki-67 growth fraction. Virchows Arch. 1995;426(3):229-33.

Goto S, Iida T, Cho S, Oka M, Kohno S, Kondo T. Overexpression of glutathione S-transferase pi enhances the adduct formation of cisplatin with glutathione in human cancer cells. Free Radic Res. 1999 Dec;31(6):549-58.

Graziano SL, Gamble GP, Newman NB, Abbott LZ, Rooney M, Mookherjee S, Lamb ML, Kohman LJ, Poiesz BJ. Prognostic significance of K-ras codon 12 mutations in patients with resected stage I and II non-small-cell lung cancer. J Clin Oncol. 1999 Feb;17(2):668-75.

Grote HJ, Schmiemann V, Geddert H, Bocking A, Kappes R, Gabbert HE, Sarbia M. Methylation of RAS association domain family protein 1A as a biomarker of lung cancer. Cancer. 2006 Apr 25;108(2):129-34.

Grote HJ, Schmiemann V, Geddert H, Rohr UP, Kappes R, Gabbert HE, Böcking A. Aberrant promoter methylation of p16(INK4a), RARB2 and SEMA3B in bronchial aspirates from patients with suspected lung cancer. Int J Cancer. 2005 Sep 20;116(5):720-5.

Gupta PC. Survey of sociodemographic characteristics of tobacco use among 99,598 individuals in Bombay, India using handheld computers. Tob Control. 1996 Summer;5(2):114-20.

György I, Minarovits J. Epigenetic regulation of lymphoid specific gene sets. Biochem Cell Biol. 2005 Jun;83(3):286-95.

Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. Bioinformatics. 2003 Feb 12;19(3):376-382.

Hainaut P, Pfeifer GP. Patterns of p53 G-->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. Carcinogenesis. 2001 Mar;22(3):367-74.

Hamel N, Black MJ, Ghadirian P, Foulkes WD. No association between P53 codon 72 polymorphism and risk of squamous cell carcinoma of the head and neck. Br J Cancer. 2000 Feb;82(4):757-759.

Hammond EC, Selikoff IJ, Seidman H. Asbestos exposure, cigarette smoking and death rates. Ann N Y Acad Sci. 1979;330:473-90.

Harris JE. Cigarette smoking among successive birth cohorts of men and women in the United States during 1900-80. J Natl Cancer Inst. 1983 Sep;71(3):473-9.

Harris N, Brill E, Shohat O, Prokocimer M, Wolf D, Arai N, Rotter V. Molecular basis for heterogeneity of the human p53 protein. Mol Cell Biol. 1986 Dec;6(12):4650-6.

Hassett C, Aicher L, Sidhu JS, Omiecinski CJ. Human microsomal epoxide hydrolase: genetic polymorphism and functional expression in vitro of amino acid variants. Hum Mol Genet. 1994 Mar;3(3):421-428.

Hassett C, Robinson KB, Beck NB, Omiecinski CJ. The human microsomal epoxide hydrolase gene (EPHX1): complete nucleotide sequence and structural characterization. Genomics. 1994b Sep 15;23(2):433-42.

Hassett, C., Aicher, L., Sidhu, L.A., Omiecinski, C.J., 1994. Human microsomal epoxide

Hayashi SI, Watanabe J, Nakachi K, Kawajiri K. (1991). PCR detection of an A/G polymorphism within exon 7 of the CYP1A1 gene. *Nucleic Acids Res*, 10, 4797

Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. Annu Rev Pharmacol Toxicol 2005;45:51—88.

Hayes JD, McLellan LI. Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress. Free Radic Res. 1999 Oct;31(4):273-300.

Hayes JD, Pulford DJ. The glutathione *S*-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. Crit Rev Biochem Mol Biol 1995;30:445—600.

Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat Rev Cancer. 2003 Oct;3(10):733-744.

Hempel N, Carrico PM, Melendez JA. Manganese superoxide dismutase (Sod2) and redox-control of signaling events that drive metastasis. Anticancer Agents Med Chem. 2011 Feb;11(2):191-201.

Henderson CJ, McLaren AW, Moffat GJ, Bacon EJ, Wolf CR. Pi-class glutathione S-transferase: regulation and function. Chem Biol Interact. 1998 Apr 24;111-112:69-82.

Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med. 2003 Nov 20;349(21):2042-54.

Hirama T, Koeffler HP. Role of the cyclin-dependent kinase inhibitors in the development of cancer. Blood. 1995 Aug 1;86(3):841-54.

Hirayama T. Non-smoking wives of heavy smokers have a higher risk of lung cancer: a study from Japan. Br Med J (Clin Res Ed). 1981 Jan 17;282(6259):183-5.

Hiyama T, Tanaka S, Kitadai Y, Ito M, Sumii M, Yoshihara M, Shimamoto F, Haruma K, Chayama K. p53 Codon 72 polymorphism in gastric cancer susceptibility in patients with Helicobacter pylori-associated chronic gastritis. Int J Cancer. 2002 Jul 20;100(3):304-308.

Ho E. Zinc deficiency, DNA damage and cancer risk. J Nutr Biochem. 2004 Oct;15(10):572-8.

Hollstein M, Shomer B, Greenblatt M, Soussi T, Hovig E, Montesano R, Harris CC. Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation. Nucleic Acids Res. 1996 Jan 1;24(1):141-6.

Honma HN, De Capitani EM, Perroud MW Jr, Barbeiro AS, Toro IF, Costa DB, Lima CS, Zambon L. Influence of p53 codon 72 exon 4, GSTM1, GSTT1 and GSTP1*B polymorphisms in lung cancer risk in a Brazilian population.Lung Cancer. 2008 Aug;61(2):152-162.

Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, van der Leest C, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS One. 2010 Apr 22;5(4):e10312.

Hou SM, Fält S, Nyberg F. Glutathione S-transferase T1-null genotype interacts synergistically with heavy smoking on lung cancer risk. Environ Mol Mutagen. 2001;38(1):83-6.

Houlston RS. Glutathione S-transferase M1 status and lung cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 1999 Aug;8(8):675-82.

Hsu HS, Wen CK, Tang YA, Lin RK, Li WY, Hsu WH, Wang YC. Promoter hypermethylation is the predominant mechanism in hMLH1 and hMSH2 deregulation and is a poor prognostic factor in nonsmoking lung cancer. Clin Cancer Res. 2005 Aug 1;11(15):5410-6.

Hu X, Ji X, Srivastava SK, Xia H, Awasthi S, Nanduri B, Awasthi YC, Zimniak P, Singh SV. Mechanism of differential catalytic efficiency of two polymorphic forms of human glutathione S-transferase P1-1 in the glutathione conjugation of carcinogenic diol epoxide of chrysene. Arch Biochem Biophys 1997(b);545: 32–38

Hu X, O' Donnell R, Srivastava SK, Xia H, Zimniak P, Nanduri B, Bleicher RJ, Awasthi S, Awasthi YC, Ji X, Singh SV. Active site architecture of polymorphic forms of human glutathione S-transferase P1-1 accounts for their enantioselectivity and disparate activity in the glutathione conjugation of 7beta,8alphadihydroxy-9 alpha,10alpha-ox y-7,8,9,10-tetrahydrobenzo(a) pyrene. Biochem Biophys Res Comm 1997(a);235:424–428.

Hu, X., Xia, H., Srivastava, S. K., Herzog, C., Awasthi, Y. C., Ji, X., Zimniak, P., and Singh, S. V. Activity of four allelic forms of glutathione *S*-transferase hGSTP1–1 for diol epoxides of polycyclic aromatic hydrocarbons. Biochem. Biophys. Res. Commun., *238:* 397–402, 1997

Huang SH, Lee HS, Mar K, Ji DD, Huang MS, Hsia KT. Loss expression of O6-methylguanine DNA methyltransferase by promoter hypermethylation and its relationship to betel quid chewing in oral squamous cell carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2010 Jun;109(6):883-9.

Hung RJ, Boffetta P, Brockmöller J, Butkiewicz D, Cascorbi I, Clapper ML, Garte S, et al. CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. Carcinogenesis. 2003 May;24(5):875-82.

Hutt JA, Vuillemenot BR, Barr EB, Grimes MJ, Hahn FF, Hobbs CH, March TH, Gigliotti AP, Seilkop SK, Finch GL, Mauderly JL, Belinsky SA. Life-span inhalation exposure to mainstream cigarette smoke induces lung cancer in B6C3F1 mice through genetic and epigenetic pathways. Carcinogenesis. 2005 Nov;26(11):1999-2009.

IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Lyon, France: International Agency for Research on Cancer; 1987.

IARC. Tobacco smoking and involuntary tobacco smoke. Lyon: International Agency for Research on Cancer. 2004

Ichiba M, Hagmar L, Rannug A, Högstedt B, Alexandrie AK, Carstensen U, Hemminki K. Aromatic DNA adducts, micronuclei and genetic polymorphism for CYP1A1 and GST1 in chimney sweeps. Carcinogenesis. 1994 Jul;15(7):1347-52.

Ihsan R, Chattopadhyay I, Phukan R, Mishra AK, Purkayastha J, Sharma J, Zomawia E, Verma Y, Mahanta J, Saxena S, Kapur S. Role of epoxide hydrolase 1 gene polymorphisms in esophageal cancer in a high-risk area in India. J Gastroenterol Hepatol. 2010 Aug;25(8):1456-1462.

Ihsan R, Devi TR, Yadav DS, Mishra AK, Sharma J, Zomawia E, Verma Y, Phukan R, Mahanta J, Kataki AC, Kapur S, Saxena S. Investigation on the role of p53 codon 72 polymorphism and interactions with tobacco, betel quid, and alcohol in susceptibility to cancers in a high-risk population from North East India. DNA Cell Biol. 2011 Mar;30(3):163-171.

Inbal B, Cohen O, Polak-Charcon S, Kopolovic J, Vadai E, Eisenbach L, Kimchi A. DAP kinase links the control of apoptosis to metastasis. Nature. 1997 Nov 13;390(6656):180-4.

Infante-Rivard C, Weinberg CR, Guiguet M.Xenobiotic-metabolizing genes and small-for-gestational-age births: interaction with maternal smoking. Epidemiology. 2006 Jan;17(1):38-46.v

International Agency for Research on Cancer (IARC). Certain polycyclic aromatic hydrocarbons and heterocyclic compounds. IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Man. Vol 3. Lyon (France): IARC; 1972. p. 45–268.

International Agency for Research on Cancer (IARC). Polynuclear aromatic compounds, part 1, chemical, environmental, and experimental data. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Vol 32. Lyon (France): IARC; 1983. p. 33–451.

International Institute for Population Sciences, Macro International. National Family Health Survey (NFHS-3) 2005–06: India. Mumbai: IIPS; 2007.

International Institute for Population Sciences, World Health Organisation, World Health Organisation(WHO)-India-WR Office. World Health Survey*, 2003 INDIA*. Mumbai: IIPS; 2006

Jabłonowski Z, Reszka E, Gromadzińska J, Wąsowicz W, Sosnowski M. Hypermethylation of p16 and DAPK promoter gene regions in patients with non-invasive urinary bladder cancer. Arch Med Sci. 2011 Jun;7(3):512-6.

Jain M, Kumar S, Rastogi N, Lal P, Ghoshal UC, Tiwari A, Pant MC, Baiq MQ, Mittal B. GSTT1, GSTM1 and GSTP1 genetic polymorphisms and interaction with tobacco, alcohol and occupational exposure in esophageal cancer patients from North India. Cancer Lett. 2006 Oct 8;242(1):60-67.

Jain M, Tilak AR, Upadhyay R, Kumar A, Mittal B. Microsomal epoxide hydrolase (EPHX1), slow (exon 3, 113His) and fast (exon 4, 139Arg) alleles confer susceptibility to squamous cell esophageal cancer. Toxicol Appl Pharmacol. 2008 Jul 15;230(2):247-251.

Jain N, Singh V, Hedau S, Kumar S, Daga MK, Dewan R, Murthy NS, Husain SA, Das BC. Infection of human papillomavirus type 18 and p53 codon 72 polymorphism in lung cancer patients from India. Chest. 2005 Dec;128(6):3999-4007.

Jakulin A, Bratko I. Analyzing attribute dependencies. In Lavrac,N., Gamberger,D., Blockeel,H. and Todorovski,L. (eds) PKDD 2003, Vol. 2838, Springer-Verlag, Cavtat, Croatia, 2003, 229-240.

Järvinen AK, Autio R, Haapa-Paananen S, Wolf M, Saarela M, Grénman R, Leivo I, Kallioniemi O, Mäkitie AA, Monni O. Identification of target genes in laryngeal squamous cell carcinoma by high-resolution copy number and gene expression microarray analyses. Oncogene. 2006 Nov 2;25(52):6997-7008.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011 Mar-Apr;61(2):69-90.

Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev. 2010 Aug;19(8):1893-907.

Jemal A, Thun MJ, Ries LA, Howe HL, Weir HK, Center MM, Ward E, Wu XC, Eheman C, Anderson R, Ajani UA, Kohler B, Edwards BK. Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control. J Natl Cancer Inst. 2008 Dec 3;100(23):1672-94.

Jin H, Pan Y, Zhao L, Zhai H, Li X, Sun L, He L, Chen Y, Hong L, Du Y, Fan D. p75 neurotrophin receptor suppresses the proliferation of human gastric cancer cells. Neoplasia. 2007 Jun;9(6):471-8.

Jin X, Wu X, Roth JA, Amos CI, King TM, Branch C, Honn SE, Spitz MR. Higher lung cancer risk for younger African-Americans with the Pro/Pro p53 genotype. Carcinogenesis. 1995 Sep;16(9):2205-2258.

Jin Y, Xu H, Zhang C, Kong Y, Hou Y, Xu Y, Xue S. Combined effects of cigarette smoking, gene polymorphisms and methylations of tumor suppressor genes on non small cell lung cancer: a hospital-based case-control study in China. BMC Cancer. 2010 Aug 12;10:422.

Jindal S. K, Behera D. Clinical Spectrum Of Primary Lung Cancer-Review Of Chandigarh Experience Of 10 Years. Lung India. 1990; 8(2): 94-98.

Jöckel KH, Ahrens W, Wichmann HE, Becher H, Bolm-Audorff U, Jahn I, Molik B, Greiser E, Timm J. Occupational and environmental hazards associated with lung cancer. Int J Epidemiol. 1992 Apr;21(2):202-13.

Jones AL, Roberts RC, Coughtrie MW. The human phenolsulphotransferase polymorphism is determined by the level of expression of the enzyme protein. Biochem J. 1993 Dec 1;296 ( Pt 2):287-290.

Jones A.L. et al. The human phenolsulphotransferase polymorphism is determined by the level of expression of the enzyme protein. Biochem. J. (1993), 296, 287–290.

Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet. 2002 Jun;3(6):415-28.

Jones PA, Laird PW.Cancer epigenetics comes of age.Nat Genet. 1999 Feb;21(2):163-7.

Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, Tavtigian SV, Stockert E, Day RS 3rd, Johnson BE, Skolnick MH. A cell cycle regulator potentially involved in genesis of many tumor types. Science. 1994 Apr 15;264(5157):436-40.

Kashiwabara K, Oyama T, Sano T, Fukuda T, Nakajima T. Correlation between methylation status of the p16/CDKN2 gene and the expression of p16 and Rb proteins in primary non-small cell lung cancers. Int J Cancer. 1998 Jun 19;79(3):215-20.

Kaushal M, Mishra AK, Raju BS, Ihsan R, Chakraborty A, Sharma J, Zomawia E, Verma Y, Kataki A, Kapur S, Saxena S. Betel quid chewing as an environmental risk factor for breast cancer. Mutat Res. 2010 Dec 21;703(2):143-148.

Kawajiri K, Nakachi K, Imai K, Watanabe J, Hayashi S. Germ line polymorphisms of p53 and CYP1A1 genes involved in human lung cancer. Carcinogenesis. 1993 Jun;14(6):1085-1089.

Kawajiri K, Nakachi K, Imai K, Yoshii A, Shinoda N, Watanabe J. Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene. FEBS Lett. 1990 Apr 9;263(1):131-3.

Kellen E, Hemelt M, Broberg K, Golka K, Kristensen VN, Hung RJ, Matullo G, Mittal RD, Porru S, Povey A, Schulz WA, Shen J, Buntinx F, Zeegers MP, Taioli E. Pooled analysis and meta-analysis of the glutathione S-transferase P1 Ile 105Val polymorphism and bladder cancer: A HuGE-GSEC review. Am J Epidem 2007;165:1221–1230.

Kerr KM, Galler JS, Hagen JA, Laird PW, Laird-Offringa IA. The role of DNA methylation in the development and progression of lung adenocarcinoma. Dis Markers. 2007;23(1-2):5-30.

Ketterer B, Harris JM, Talaska G, Meyer DJ, Pemble SE, Taylor JB, Lang NP, Kadlubar FF. The human glutathione S-transferase supergene family, its polymorphism, and its effects on susceptibility to lung cancer. Environ Health Perspect. 1992 Nov;98:87-94.

Kietthubthew S, Sriplung H, Au WW, Ishida T. The p53 codon 72 polymorphism and risk of oral cancer in Southern Thailand. Asian Pac J Cancer Prev. 2003 Jul-Sep;4(3):209-214.

Kihara M, Kihara M, Noda K. Risk of smoking for squamous and small cell carcinomas of the lung modulated by combinations of CYP1A1 and GSTM1 gene polymorphisms in a Japanese population. Carcinogenesis. 1995 Oct;16(10):2331-6.

Kim DH, Nelson HH, Wiencke JK, Christiani DC, Wain JC, Mark EJ, Kelsey KT. Promoter methylation of DAP-kinase: association with advanced stage in non-small cell lung cancer. Oncogene. 2001 Mar 29;20(14):1765-70

Kim DH, Nelson HH, Wiencke JK, Zheng S, Christiani DC, Wain JC, Mark EJ, Kelsey KT. p16(INK4a) and histology-specific methylation of CpG islands by exposure to tobacco smoke in non-small cell lung cancer. Cancer Res. 2001 Apr 15;61(8):3419-24.

Kim JM, Lee OY, Lee CG, Kwon SJ, Kim KS, Moon W, Koh DH, Lee HL, Yoon BC, Choi HS, Hahm JS, Lee MH, Lee DH. p53 Codon 72 and 16-bp duplication polymorphisms of gastric cancer in Koreans. Korean J Gastroenterol. 2007 Nov;50(5):292-298.

Kim KH, Rodriguez AM, Carrico PM, Melendez JA. Potential mechanisms for the inhibition of tumor cell growth by manganese superoxide dismutase. Antioxid Redox Signal. 2001 Jun;3(3):361-73.

Kishore Chaudhry. Tobacco control in India. Report on 50 years of tobacco control in India http://www.corecentre.co.in/Database/Docs/DocFiles/fiftyyear.pdf . Last accessed on February 17, 2010.

Kitamura E, Igarashi J, Morohashi A, Hida N, Oinuma T, Nemoto N, Song F, Ghosh S, Held WA, Yoshida-Noro C, Nagase H. Analysis of tissue-specific differentially methylated regions (TDMs) in humans. Genomics. 2007 Mar;89(3):326-37.

Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971 Apr;68(4):820-3.

Kok K, Naylor SL, Buys CH. Deletions of the short arm of chromosome 3 in solid tumors and the search for suppressor genes. Adv Cancer Res. 1997;71:27-92.

Kotnis A, Kannan S, Sarin R, Mulherkar R. Case-control study and meta-analysis of SULT1A1 Arg213His polymorphism for gene, ethnicity and environment interaction for cancer risk. Br J Cancer. 2008 Oct 21;99(8):1340-1347.

Kreyberg L. Lung cancer in workers in a nickel refinery. Br J Ind Med. 1978 May;35(2):109-16.

Kumar M, Agarwal SK, Goel SK. Lung cancer risk in north Indian population: role of genetic polymorphisms and smoking. Mol Cell Biochem. 2009 Feb;322(1-2):73-79

Kuo MY, Lin CY, Hahn LJ, Cheng SJ, Chiang CP Expression of cyclin D1 is correlated with poor prognosis in patients with areca quid chewing-related oral squamous cell carcinomas in Taiwan. J Oral Pathol Med. 1999 Apr;28(4):165-9.

Kurihara N, Wada O. Silicosis and smoking strongly increase lung cancer risk in silica-exposed workers. Ind Health. 2004 Jul;42(3):303-14.

Lam TK, Ruczinski I, Helzlsouer K, Shugart YY, Li KE, Clipp S, Strickland PT, Alberg AJ. Copy number variants of GSTM1 and GSTT1 in relation to lung cancer risk in a prospective cohort study. Ann Epidemiol. 2009 Aug;19(8):546-552.

Lam WK, Du YX. Environmental inhaled agents and their relation to lung cancer. In: Loke J (ed.) Pathophysiology and Treatment of Inhalation Injuries, Lung Biology in Human and Diseases, Vol. 34, chapter 10. Marcel Dekker, New York, 1988;423-51.

Lam WK, White NW, Chan-Yeung MM. Lung cancer epidemiology and risk factors in Asia and Africa. Int J Tuberc Lung Dis. 2004 Sep;8(9):1045-57.

Lancaster JM, Brownlee HA, Bell DA, Futreal PA, Marks JR, Berchuck A, Wiseman RW, Taylor JA. Microsomal epoxide hydrolase polymorphism as a risk factor for ovarian cancer. Mol Carcinog. 1996 Nov;17(3):160-162.

Landi MT, Bertazzi PA, Shields PG, Clark G, Lucier GW, Garte SJ, Cosma G, Caporaso NE. Association between CYP1A1 genotype, mRNA expression and enzymatic activity in humans. Pharmacogenetics. 1994 Oct;4(5):242-246.

Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, Mann FE, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PLoS One. 2008 Feb 20;3(2):e1651.

Landi S. Mammalian class theta GST and differential susceptibility to carcinogens: a review. Mutat Res 2000;463:247—83.

Larsen F, Gundersen G, Lopez R, Prydz H. CpG islands as gene markers in the human genome. Genomics. 1992 Aug;13(4):1095-107.

Larsen JE, Minna JD. Molecular biology of lung cancer: clinical implications. Clin Chest Med. 2011 Dec;32(4):703-40.

Lawther Pj, Commins Bt, Waller Re. A Study Of The Concentrations Of Polycyclic Aromatic Hydrocarbons In Gas Works Retort Houses. Br J Ind Med. 1965 Jan;22:13-20.

Le Marchand L, Guo C, Benhamou S, Bouchardy C, Cascorbi I, Clapper ML, Garte S, Haugen A, Ingelman-Sundberg M, Kihara M, Rannug A, Ryberg D, Stücker I, Sugimura H, Taioli E. Pooled analysis of the CYP1A1 exon 7 polymorphism and lung cancer (United States). Cancer Causes Control. 2003 May;14(4):339-346.

Lee AM, Fraumeni JF Jr. Arsenic and respiratory cancer in man: an occupational study. J Natl Cancer Inst. 1969 Jun;42(6):1045-52.

Lee BM, Park KK. Beneficial and adverse effects of chemopreventive agents. Mutat Res. 2003 Feb-Mar;523-524:265-78.

Lee KM, Kang D, Clapper ML, Ingelman-Sundberg M, Ono-Kihara M, Kiyohara C, Min S, Lan Q, et al. CYP1A1, GSTM1, and GSTT1 polymorphisms, smoking, and lung cancer risk in a pooled analysis among Asian populations. Cancer Epidemiol Biomarkers Prev. 2008 May;17(5):1120-6.

Lee SH, Son MJ, Oh SH, Rho SB, Park K, Kim YJ, Park MS, Lee JH. Thymosin {beta}(10) inhibits angiogenesis and tumor growth by interfering with Ras function. Cancer Res. 2005 Jan 1;65(1):137-48.

Lee WH, Morton RA, Epstein JI, Brooks JD, Campbell PA, Bova GS, Hsieh WS, Isaacs WB, Nelson WG. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. Proc Natl Acad Sci U S A. 1994 Nov 22;91(24):11733-7.

Lee WP, Lee CL, Lin HC. Glutathione S-transferase and glutathione peroxidase are essential in the early stage of adriamycin resistance before P-glycoprotein overexpression in HOB1 lymphoma cells. Cancer Chemother Pharmacol. 1996;38(1):45-51.

Lehtiö J, De Petris L. Lung cancer proteomics, clinical and technological considerations. J Proteomics. 2010 Sep 10;73(10):1851-63.

Lewis SJ, Cherry NM, Niven RM, Barber PV, Povey AC. GSTM1, GSTT1 and GSTP1 polymorphisms and lung cancer risk.Cancer Lett. 2002 Jun 28;180(2):165-171.

Li Q, Fraley C, Bumgarner RE, Yeung KY, Raftery AE. Donuts, scratches and blanks: robust model-based segmentation of microarray images. Bioinformatics. 2005 Jun 15;21(12):2875-82.

Licchesi JD, Westra WH, Hooker CM, Herman JG. Promoter hypermethylation of hallmark cancer genes in atypical adenomatous hyperplasia of the lung. Clin Cancer Res. 2008 May 1;14(9):2570-8

Lin L, Miller CT, Contreras JI, Prescott MS, Dagenais SL, Wu R, Yee J, Orringer MB, Misek DE, Hanash SM, Glover TW, Beer DG. The hepatocyte nuclear factor 3 alpha gene, HNF3alpha (FOXA1), on chromosome band 14q13 is amplified and overexpressed in esophageal and lung adenocarcinomas. Cancer Res. 2002 Sep 15;62(18):5273-9.

Lin P, Wang SL, Wang HJ, Chen KW, Lee HS, Tsai KJ, Chen CY, Lee H. Association of CYP1A1 and microsomal epoxide hydrolase polymorphisms with lung squamous cell carcinoma. Br J Cancer. 2000 Feb;82(4):852-7.

Liu L, Wu C, Wang Y, Zhong R, Wang F, Zhang X, Duan S, Lou J, Yu D, Tan W, Yuan J, Wu T, Nie S, Miao X, Lin D. Association of candidate genetic variations with gastric cardia adenocarcinoma in Chinese population: a multiple interaction analysis. Carcinogenesis. 2011 Mar;32(3):336-342.

Liu Y, Lan Q, Siegfried JM, Luketich JD, Keohavong P Aberrant promoter methylation of p16 and MGMT genes in lung tumors from smoking and never-smoking lung cancer patients. Neoplasia. 2006 Jan;8(1):46-51.

Lloyd JW. Long-term mortality study of steelworkers. V. Respiratory cancer in coke plant workers. J Occup Med. 1971 Feb;13(2):53-68.

London SJ, Smart J, Daly AK. Lung cancer risk in relation to genetic polymorphisms of microsomal epoxide hydrolase among African-Americans and Caucasians in Los Angeles County. Lung Cancer. 2000 May;28(2):147-155.

London SJ, Yuan JM, Chung FL, Gao YT, Coetzee GA, Ross RK, Yu MC. Isothiocyanates, glutathione S-transferase M1 and T1 polymorphisms, and lung-cancer risk: a prospective study of men in Shanghai, China. Lancet. 2000 Aug 26;356(9231):724-729.

López-Cima MF, Alvarez-Avellón SM, Pascual T, Fernández-Somoano A, Tardón A. Genetic polymorphisms in CYP1A1, GSTM1, GSTP1 and GSTT1 metabolic genes and risk of lung cancer in Asturias. BMC Cancer. 2012 Sep 27;12:433.

Lorenzo Bermejo J, Hemminki K. Familial lung cancer and aggregation of smoking habits: a simulation of the effect of shared environmental factors on the familial risk of cancer. Cancer Epidemiol Biomarkers Prev. 2005 Jul;14(7):1738-40.

MacDonald TJ, Pollack IF, Okada H, Bhattacharya S, Lyons-Weiler J. Progression-associated genes in astrocytoma identified by novel microarray gene expression data reanalysis. Methods Mol Biol. 2007;377:203-22.

Mahabir S, Spitz MR, Barrera SL, Beaver SH, Etzel C, Forman MR. Dietary zinc, copper and selenium, and risk of lung cancer. Int J Cancer. 2007 Mar 1;120(5):1108-15.

Malafa M, Margenthaler J, Webb B, Neitzel L, Christophersen M. MnSOD expression is increased in metastatic gastric cancer. J Surg Res. 2000 Feb;88(2):130-4.

Malats N, Camus-Radon AM, Nyberg F, Ahrens W, Constantinescu V, Mukeria A, et al. Lung cancer risk in nonsmokers and GSTM1 and GSTT1 genetic polymorphism. Cancer Epidemiol Biomarkers Prev. 2000 Aug;9(8):827-33.

Mandelbaum J, Bhagat G, Tang H, Mo T, Brahmachary M, Shen Q, Chadburn A, Rajewsky K, Tarakhovsky A, Pasqualucci L, Dalla-Favera R. BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. Cancer Cell. 2010 Dec 14;18(6):568-79.

Mansilla F, Birkenkamp-Demtroder K, Kruhøffer M, Sørensen FB, Andersen CL, Laiho P, Aaltonen LA, Verspaget HW, Orntoft TF. Differential expression of DHHC9 in microsatellite stable and instable human colorectal cancer subgroups. Br J Cancer. 2007 Jun 18;96(12):1896-903.

Martuszewska D, Ljungberg B, Johansson M, Landberg G, Oslakovic C, Dahlbäck B, Hafizi S. Tensin3 is a negative regulator of cell migration and all four Tensin family members are downregulated in human kidney cancer. PLoS One. 2009;4(2):e4350.

Matlashewski GJ, Tuck S, Pim D, Lamb P, Schneider J, Crawford LV. Primary structure polymorphism at amino acid residue 72 of human p53. Mol Cell Biol. 1987 Feb;7(2):961-963.

McDonald JW, Taylor JA, Watson MA, Saccomanno G, Devereux TR. p53 and K-ras in radon-associated lung adenocarcinoma. Cancer Epidemiol Biomarkers Prev. 1995 Oct-Nov;4(7):791-3.

McDoniels-Silvers AL, Nimri CF, Stoner GD, Lubet RA, You M. Differential gene expression in human lung adenocarcinomas and squamous cell carcinomas. Clin Cancer Res. 2002 Apr;8(4):1127-38.

McManus ME, Burgess WM, Veronese ME, et al. Metabolism of 2-acetylaminofluorene and benzo(a)pyrene and activation of food-derived heterocyclic amine mutagens by human cytochrome P-450. Cancer Res 1990;50:3367–76.

Mechanic LE, Marrogi AJ, Welsh JA, Bowman ED, Khan MA, Enewold L, Zheng YL, Chanock S, Shields PG, Harris CC. Polymorphisms in XPD and TP53 and mutation in human lung cancer. Carcinogenesis. 2005 Mar;26(3):597-604.

Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, Baylin SB, Sidransky D. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat Med. 1995 Jul;1(7):686-92.

Miller DP, De Vivo I, Neuberg D, Wain JC, Lynch TJ, Su L, Christiani DC. Association between self-reported environmental tobacco smoke exposure and lung cancer: modification by GSTP1 polymorphism. Int J Cancer. 2003 May 10;104(6):758-63.

Minaguchi T, Kanamori Y, Matsushima M, Yoshikawa H, Taketani Y, Nakamura Y. No evidence of correlation between polymorphism at codon 72 of p53 and risk of cervical cancer in Japanese patients with human papillomavirus 16/18 infection. Cancer Res. 1998 Oct 15;58(20):4585-4586.

Mitra S, Misra C, Singh RK, Panda CK, Roychoudhury S. Association of specific genotype and haplotype of p53 gene with cervical cancer in India. J Clin Pathol. 2005 Jan;58(1):26-31.

Miyata M, Kudo G, Lee YH, Yang TJ, Gelboin HV, Fernandez-Salguero P, Kimura S, Gonzalez FJ. Targeted disruption of the microsomal epoxide hydrolase gene. Microsomal epoxide hydrolase is required for the carcinogenic activity of 7,12-dimethylbenz[a]anthracene. J Biol Chem. 1999 Aug 20;274(34):23963-8.

Mohandas T, Sparkes RS, Shapiro LJ. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science. 1981 Jan 23;211(4480):393-6.

Mojtahedi Z, Haghshenas MR, Hosseini SV, Fattahi MJ, Ghaderi A. p 53 codon 72 polymorphism in stomach and colorectal adenocarcinomas in Iranian patients. Indian J Cancer. 2010 Jan-Mar;47(1):31-34.

Monk M, Boubelik M, Lehnert S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development. 1987 Mar;99(3):371-82.

Mooney LA, Bell DA, Santella RM, Van Bennekum AM, Ottman R, Paik M, Blaner WS, Lucier GW, Covey L, Young TL, Cooper TB, Glassman AH, Perera FP. Contribution of genetic and nutritional factors to DNA damage in heavy smokers. Carcinogenesis. 1997 Mar;18(3):503-9.

Moore LE, Huang WY, Chatterjee N, Gunter M, Chanock S, Yeager M, Welch B, Pinsky P, Weissfeld J, Hayes RB. GSTM1, GSTT1, and GSTP1 polymorphisms and risk of advanced colorectal adenoma. Cancer Epidemiol Biomarkers Prev. 2005 Jul;14(7):1823-7.

Morgensztern D, Ng SH, Gao F, Govindan R. Trends in stage distribution for patients with non-small cell lung cancer: a National Cancer Database survey. J Thorac Oncol. 2010 Jan;5(1):29-33.

Morrow CS, Cowan KH. Glutathione S-transferases and drug resistance. Cancer Cells. 1990 Jan;2(1):15-22.

Mulder TP, Verspaget HW, Sier CF, Roelofs HM, Ganesh S, Griffioen G, Peters WH. Glutathione S-transferase pi in colorectal tumors is predictive for overall survival. Cancer Res. 1995 Jun 15;55(12):2696-702.

Murata M, Tagawa M, Kimura M, Kimura H, Watanabe S, Saisho H. Analysis of a germ line polymorphism of the p53 gene in lung cancer patients; discrete results with smoking history. Carcinogenesis. 1996 Feb;17(2):261-264.

Naito M, Goto Y, Ishida Y, Nishio K, Kawai S, Hamajima N. Alternative genotyping method of GSTT1 null/present polymorphism. Expert Rev Mol Diagn. 2006 Nov;6(6):873-877.

Nakachi K, Imai K, Hayashi S, Kawajiri K. Polymorphisms of the CYP1A1 and glutathione S-transferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. Cancer Res. 1993 Jul 1;53(13):2994-9.

Nakata S, Sugio K, Uramoto H, Oyama T, Hanagiri T, Morita M, Yasumoto K. The methylation status and protein expression of CDH1, p16(INK4A), and fragile histidine triad in nonsmall cell lung carcinoma: epigenetic silencing, clinical features, and prognostic significance. Cancer. 2006 May 15;106(10):2190-9.

Nakayama M, Bennett CJ, Hicks JL, Epstein JI, Platz EA, Nelson WG, De Marzo AM. Hypermethylation of the human glutathione S-transferase-pi gene (GSTP1) CpG island is present in a subset of proliferative inflammatory atrophy lesions but not in normal or hyperplastic epithelium of the prostate: a detailed study using laser-capture microdissection. Am J Pathol. 2003 Sep;163(3):923-33.

Narayan G, Arias-Pulido H, Koul S, Vargas H, Zhang FF, Villella J, Schneider A, Terry MB, Mansukhani M, Murty VV. Frequent promoter methylation of CDH1, DAPK, RARB, and HIC1 genes in carcinoma of cervix uteri: its relationship to clinical outcome. Mol Cancer. 2003 May 13;2:24.

National Cancer Registry Programme. Three year report of Population Based Cancer Registries 2006-2008. Indian Council of Medical Research.

Nazar-Stewart V, Motulsky AG, Eaton DL, White E, Hornung SK, Leng ZT, Stapleton P, Weiss NS. The glutathione S-transferase mu polymorphism as a marker for susceptibility to lung carcinoma. Cancer Res. 1993 May 15;53(10 Suppl):2313-2318.

Nazar-Stewart V, Vaughan TL, Stapleton P, Van Loo J, Nicol-Blades B, Eaton DL. A population-based study of glutathione S-transferase M1, T1 and P1 genotypes and risk for lung cancer. Lung Cancer. 2003 Jun;40(3):247-58.

Nelson CP, Kidd LC, Sauvageot J, Isaacs WB, De Marzo AM, Groopman JD, et al. Protection against 2-hydroxyamino-1-methyl-6-phenylimidazo[4,5-b]pyridine cytotoxicity and DNA adduct formation in human prostate by glutathione S-transferase P1. Cancer Res. 2001 Jan 1;61(1):103-9.

Neri M, Taioli E, Filiberti R, Paolo Ivaldi G, Aldo Canessa P, Verna A, et al. Metabolic genotypes as modulators of asbestos-related pleural malignant mesothelioma risk: a comparison of Finnish and Italian populations. Int J Hyg Environ Health. 2006 Jul;209(4):393-8.

Niklinska W, Chyczewski L, Laudanski J, Sawicki B, Niklinski J. Detection of P53 abnormalities in non-small cell lung cancer by yeast functional assay. Folia Histochem Cytobiol. 2001;39(2):147-8.

Nitadori J, Inoue M, Iwasaki M, Otani T, Sasazuki S, Nagai K, Tsugane S. Association between lung cancer incidence and family history of lung cancer: data from a large-scale population-based cohort study, the JPHC study. Chest. 2006 Oct;130(4):968-75.

Notani P, Sanghvi LD. A retrospective study of lung cancer in Bombay. Br J Cancer. 1974 Jun;29(6):477-82.

Notani PN, Shah P, Jayant K, Balakrishnan V. Occupation and cancers of the lung and bladder: a case-control study in Bombay. Int J Epidemiol. 1993 Apr;22(2):185-91.

Nowell S, Ambrosone CB, Ozawa S, MacLeod SL, Mrackova G, Williams S, Plaxco J, Kadlubar FF, Lang NP. Relationship of phenol sulfotransferase activity (SULT1A1) genotype to sulfotransferase phenotype in platelet cytosol. Pharmacogenetics. 2000 Dec;10(9):789-797.

Nowell S, Ambrosone CB, Ozawa S, MacLeod SL, Mrackova G, Williams S, Plaxco J, Kadlubar FF, Lang NP. Relationship of phenol sulfotransferase activity (SULT1A1) genotype to sulfotransferase phenotype in platelet cytosol. Pharmacogenetics. 2000 Dec;10(9):789-97.

Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC TP53 database: new online mutation analysis and recommendations to users. Hum Mutat. 2002 Jun;19(6):607-14.

Oster B, Thorsen K, Lamy P, Wojdacz TK, Hansen LL, Birkenkamp-Demtröder K, Sørensen KD, Laurberg S, Orntoft TF, Andersen CL. Identification and validation of highly frequent CpG island hypermethylation in colorectal adenomas and carcinomas. Int J Cancer. 2011 Dec 15;129(12):2855-66.

Ostrov DA, Barnes CL, Smith LE, Binns S, Brusko TM, Brown AC, Quint PS, Litherland SA, Roopenian DC, Iczkowski KA. Characterization of HKE2: an ancient antigen encoded in the major histocompatibility complex. Tissue Antigens. 2007 Feb;69(2):181-8.

Ott MG, Holder BB, Gordon HL. Respiratory cancer and occupational exposure to arsenicals. Arch Environ Health. 1974 Nov;29(5):250-5.

Otterson GA, Khleif SN, Chen W, Coxon AB, Kaye FJ. CDKN2 gene silencing in lung cancer by DNA hypermethylation and kinetics of p16INK4 protein induction by 5-aza 2'deoxycytidine. Oncogene. 1995 Sep 21;11(6):1211-6.

Palmiter RD, Huang L. Efflux and compartmentalization of zinc by members of the SLC30 family of solute carriers. Pflugers Arch. 2004 Feb;447(5):744-51.

Papadakis ED, Soulitzis N, Spandidos DA. Association of p53 codon 72 polymorphism with advanced lung cancer: the Arg allele is preferentially retained in tumours arising in Arg/Pro germline heterozygotes. Br J Cancer. 2002 Oct 21;87(9):1013-1018.

Papadakis EN, Dokianakis DN, Spandidos DA. p53 codon 72 polymorphism as a risk factor in the development of breast cancer. Mol Cell Biol Res Commun. 2000 Jun;3(6):389-392.

Park JY, Muscat JE, Ren Q, et al. CYP1A1 and GSTM1 polymorphisms and oral cancer risk. Cancer Epidemiol Biomark Prev. (1997). 6, 791-7

Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin. 2005 Mar-Apr;55(2):74-108.

Pemble S, Schroeder KR, Spencer SR, Meyer DJ, Hallier E, Bolt HM, Ketterer B, Taylor JB. Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. Biochem J. 1994 May 15;300 ( Pt 1):271-276.

Perera FP, Mooney LA, Stampfer M, Phillips DH, Bell DA, Rundle A, Cho S, Tsai WY, Ma J, Blackwood A, Tang D; Physicians' Health Cohort Study. Associations between carcinogen-DNA damage, glutathione S-transferase genotypes, and risk of lung cancer in the prospective Physicians' Health Cohort Study. Carcinogenesis. 2002 Oct;23(10):1641-6.

Peto R, Lopez AD, Boreham J, Thun M. Mortality From Smoking in Developed Countries 1950-2000. 2nd ed. Revised June 2006. Available at: http://www.ctsu.ox.ac.uk/_tobacco/SMK_All_PAGES.pdf. International Union Against Cancer (UICC). Geneva: Switzerland. 2006. Last Accessed 7/1/2010.

Petty RD, Nicolson MC, Kerr KM, Collie-Duguid E, Murray GI. Gene expression profiling in non-small cell lung cancer: from molecular mechanisms to clinical application. Clin Cancer Res. 2004 May 15;10(10):3237-48.

Pezzella F, Turley H, Kuzu I, Tungekar MF, Dunnill MS, Pierce CB, Harris A, Gatter KC, Mason DY. bcl-2 protein in non-small-cell lung carcinoma. N Engl J Med. 1993 Sep 2;329(10):690-4.

Piacentini S, Polimanti R, Porreca F, Martínez-Labarga C, De Stefano GF, Fuciarelli M. GSTT1 and GSTM1 gene polymorphisms in European and African populations. Mol Biol Rep. 2011 Feb;38(2):1225-30.

Pietsch EC, Humbey O, Murphy ME. Polymorphisms in the p53 pathway. Oncogene. 2006 Mar 13;25(11):1602-1611.

Pim D, Banks L. p53 polymorphic variants at codon 72 exert different effects on cell cycle progression. Int J Cancer. 2004 Jan 10;108(2):196-9.

Pogue-Geile K, Geiser JR, Shu M, Miller C, Wool IG, Meisler AI, Pipas JM. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. Mol Cell Biol. 1991 Aug;11(8):3842-9.

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet. 1999 Sep;23(1):41-6.

Prise KM, Pinto M, Newman HC, Michael BD. A review of studies of ionizing radiation-induced double-strand break clustering. Radiat Res. 2001 Nov;156(5 Pt 2):572-6.

Pulling LC, Vuillemenot BR, Hutt JA, Devereux TR, Belinsky SA Aberrant promoter hypermethylation of the death-associated protein kinase gene is early and frequent in murine lung tumors induced by cigarette smoke and tobacco carcinogens. Cancer Res. 2004 Jun 1;64(11):3844-8.

Raimondi S, Paracchini V, Autrup H, Barros-Dios JM, Benhamou S, Boffetta P, et al,. Meta- and pooled analysis of GSTT1 and lung cancer: a HuGE-GSEC review. Am J Epidemiol. 2006 Dec 1;164(11):1027-1042

Reed DJ. Glutathione: toxicological implications. Annu Rev Pharmacol Toxicol. 1990;30:603-631.

Reik W, Collick A, Norris ML, Barton SC, Surani MA. Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature. 1987 Jul 16-22;328(6127):248-51.

Reis-Filho JS, Steele D, Di Palma S, Jones RL, Savage K, James M, Milanezi F, Schmitt FC, Ashworth A. Distribution and significance of nerve growth factor receptor (NGFR/p75NTR) in normal, benign and malignant breast tissue. Mod Pathol. 2006 Feb;19(2):307-19.

Risch A, Plass C. Lung cancer epigenetics and genetics. Int J Cancer. 2008 Jul 1;123(1):1-7.

Risch A, Wikman H, Thiel S, Schmezer P, Edler L, Drings P, Dienemann H, Kayser K, Schulz V, Spiegelhalder B, Bartsch H. Glutathione-S-transferase M1, M3, T1 and P1 polymorphisms and susceptibility to non-small-cell lung cancer subtypes and hamartomas. Pharmacogenetics. 2001 Dec;11(9):757-764.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet. 2001 Jul;69(1):138-147

Rogounovitch TI, Saenko VA, Ashizawa K, Sedliarou IA, Namba H, Abrosimov AY, Lushnikov EF, Roumiantsev PO, Konova MV, Petoukhova NS, Tchebotareva IV, Ivanov VK, Chekin SY, Bogdanova TI, Tronko MD, Tsyb AF, Thomas GA, Yamashita S. TP53 codon 72 polymorphism in radiation-associated human papillary thyroid cancer. Oncol Rep. 2006 Apr;15(4):949-956.

Rojas M, Alexandrov K, Cascorbi I, Brockmöller J, Likhachev A, Pozharisski K, Bouvier G, Auburtin G, Mayer L, Kopp-Schneider A, Roots I, Bartsch H. High benzo[a]pyrene diol-epoxide DNA adduct levels in lung and blood cells from individuals with combined CYP1A1 MspI/Msp-GSTM1*0/*0 genotypes. Pharmacogenetics. 1998 Apr;8(2):109-18.

Rojas M, Cascorbi I, Alexandrov K, Kriek E, Auburtin G, Mayer L, Kopp-Schneider A, Roots I, Bartsch H. Modulation of benzo[a]pyrene diolepoxide-DNA adduct levels in human white blood cells by CYP1A1, GSTM1 and GSTT1 polymorphism. Carcinogenesis. 2000 Jan;21(1):35-41.

Roodi N, Dupont WD, Moore JH, Parl FF. Association of homozygous wild-type glutathione S-transferase M1 genotype with increased breast cancer risk. Cancer Res. 2004 Feb 15;64(4):1233-1236.

Rouissi K, Ouerhani S, Hamrita B, Bougatef K, Marrakchi R, Cherif M, Ben Slama MR, Bouzouita M, Chebil M, Ben Ammar Elgaaied A. Smoking and polymorphisms in xenobiotic metabolism and DNA repair genes are additive risk factors affecting bladder cancer in Northern Tunisia. Pathol Oncol Res. 2011 Dec;17(4):879-886.

Ryberg D, Skaug V, Hewer A, Phillips DH, Harries LW, Wolf CR, Ogreid D, Ulvik A, Vu P, Haugen A. Genotypes of glutathione transferase M1 and P1 and their significance for lung DNA adduct levels and cancer risk. Carcinogenesis. 1997 Jul;18(7):1285-9.

Saegusa M, Takano Y, Okayasu I. Bcl-2 expression and its association with cell kinetics in human gastric carcinomas and intestinal metaplasia. J Cancer Res Clin Oncol. 1995;121(6):357-63.

Sakamuro D, Sabbatini P, White E, Prendergast GC. The polyproline region of p53 is required to activate apoptosis but not growth arrest. Oncogene. 1997 Aug 18;15(8):887-98.

San Jose C, Cabanillas A, Benitez J, Carrillo JA, Jimenez M, Gervasini G. CYP1A1 gene polymorphisms increase lung cancer risk in a high-incidence region of Spain: a case control study. BMC Cancer. 2010 Aug 30;10:463.

Sanchez-Cespedes M, Decker PA, Doffek KM, Esteller M, Westra WH, Alawi EA, Herman JG, Demeure MJ, Sidransky D, Ahrendt SA. Increased loss of chromosome 9p21 but not p16 inactivation in primary non-small cell lung cancer from smokers. Cancer Res. 2001 Mar 1;61(5):2092-6.

Sanchez-Cespedes M, Reed AL, Buta M, Wu L, Westra WH, Herman JG, Yang SC, Jen J, Sidransky D. Inactivation of the INK4A/ARF locus frequently coexists with TP53 mutations in non-small cell lung cancer. Oncogene. 1999 Oct 21;18(43):5843-9.

Sanford JP, Clark HJ, Chapman VM, Rossant J. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes Dev. 1987 Dec;1(10):1039-46.

Santelli G, Califano D, Chiappetta G, Vento MT, Bartoli PC, Zullo F, Trapasso F, Viglietto G, Fusco A. Thymosin beta-10 gene overexpression is a general event in human carcinogenesis. Am J Pathol. 1999 Sep;155(3):799-804.

Sapari NS, Loh M, Vaithilingam A, Soong R Clinical potential of DNA methylation in gastric cancer: a meta-analysis. PLoS One. 2012;7(4):e36275

Saracci R, Boffetta P. Interactions of tobacco smoking and other causes of lung cancer. In: Samet JM, ed. Epidemiology of lung cancer. New York, NY: Marcel Dekker, 1994;465-493.

Schoket B, Phillips DH, Kostic S, Vincze I. Smoking-associated bulky DNA adducts in bronchial tissue related to CYP1A1 MspI and GSTM1 genotypes in lung patients. Carcinogenesis. 1998 May;19(5):841-6.

Schwartz AG, Prysak GM, Bock CH, Cote ML. The molecular epidemiology of lung cancer. Carcinogenesis. 2007 Mar;28(3):507-18.

Sciandrello G, Caradonna F, Mauro M, Barbata G. Arsenic-induced DNA hypomethylation affects chromosomal instability in mammalian cells. Carcinogenesis. 2004 Mar;25(3):413-7.

Seidegård J, DePierre JW, Pero RW. Hereditary interindividual differences in the glutathione transferase activity towards trans-stilbene oxide in resting human mononuclear leukocytes are due to a particular isozyme(s). Carcinogenesis. 1985 Aug;6(8):1211-1216.

Serrano M, Hannon GJ, Beach D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. Nature. 1993 Dec 16;366(6456):704-7.

Sharan RN, Mehrotra R, Choudhury Y, Asotra K Association of betel nut with carcinogenesis: revisit with a clinical perspective. PLoS One. 2012;7(8):e42759.

Shekelle RB, Lepper M, Liu S, Maliza C, Raynor WJ Jr, Rossof AH, Paul O, Shryock AM, Stamler J. Dietary vitamin A and risk of cancer in the Western Electric study. Lancet. 1981 Nov 28;2(8257):1185-90.

Shields PG, Caporaso NE, Falk RT, Sugimura H, Trivers GE, Trump BF, Hoover RN, Weston A, Harris CC. Lung cancer, race, and a CYP1A1 genetic polymorphism. Cancer Epidemiol Biomarkers Prev. 1993 Sep-Oct;2(5):481-5.

Shiraishi M, Sekiya T. Change in methylation status of DNA of remaining allele in human lung cancers with loss of heterozygosity on chromosomes 3p and 13q. Jpn J Cancer Res. 1989 Oct;80(10):924-7.

Shivakumar L, Minna J, Sakamaki T, Pestell R, White MA. The RASSF1A tumor suppressor blocks cell cycle progression and inhibits cyclin D1 accumulation. Mol Cell Biol. 2002 Jun;22(12):4309-18.

Shohat G, Shani G, Eisenstein M, Kimchi A. The DAP-kinase family of proteins: study of a novel group of calcium-regulated death-promoting kinases. Biochim Biophys Acta. 2002 Nov 4;1600(1-2):45-50.

Siddique M, Sabapathy K. Trp53-dependent DNA-repair is affected by the codon 72 polymorphism. Oncogene. 2006 Jun 15;25(25):3489-500.

Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013 Jan;63(1):11-30.

Sikora K, Ong G. Molecular biology and respiratory disease. 4. Cancer genes. Thorax. 1990 May;45(5):409-13.

Silvestrini R, Veneroni S, Daidone MG, Benini E, Boracchi P, Mezzetti M, Di Fronzo G, Rilke F, Veronesi U. The Bcl-2 protein: a prognostic indicator strongly related to p53 protein in lymph node-negative breast cancer patients. J Natl Cancer Inst. 1994 Apr 6;86(7):499-504.

Sims P, Grover PL, Swaisland A, Pal K, Hewer A. Metabolic activation of benzo(a)pyrene proceeds by a diol-epoxide. Nature 1974;252:326 – 8.

Singal R, Ginder GD. DNA methylation. Blood. 1999 Jun 15;93(12):4059-70.

Singh AP, Bafna S, Chaudhary K, Venkatraman G, Smith L, Eudy JD, Johansson SL, Lin MF, Batra SK. Genome-wide expression profiling reveals transcriptomic variation and perturbed gene networks in androgen-dependent and androgen-independent prostate cancer cells. Cancer Lett. 2008 Jan 18;259(1):28-38.

Sinha DN, Gupta PC, Dobe M, Prasad VM. Tobacco control in schools of India: review from India Global School Personnel Survey 2006. Indian J Public Health. 2007 Apr-Jun;51(2):101-6.

Sinicrope FA, Hart J, Michelassi F, Lee JJ. Prognostic value of bcl-2 oncoprotein expression in stage II colon carcinoma. Clin Cancer Res. 1995 Oct;1(10):1103-10.

Själander A, Birgander R, Hallmans G, Cajander S, Lenner P, Athlin L, Beckman G, Beckman L. p53 polymorphisms and haplotypes in breast cancer. Carcinogenesis. 1996 Jun;17(6):1313-1316.

Själander A, Birgander R, Kivelä A, Beckman G. p53 polymorphisms and haplotypes in different ethnic groups. Hum Hered. 1995 May-Jun;45(3):144-149.

Smith CA, Harrison DJ. Association between polymorphism in gene for microsomal epoxide hydrolase and susceptibility to emphysema. Lancet. 1997 Aug 30;350(9078):630-3.

Sobin L, Wittekind C, International Union against Cancer. TNM classification of malignant tumours. 6th ed. New York: Wiley-Liss; 2002

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Döhner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer. 1997 Dec;20(4):399-407

Sørensen M, Raaschou-Nielsen O, Brasch-Andersen C, Tjønneland A, Overvad K, Autrup H. Interactions between GSTM1, GSTT1 and GSTP1 polymorphisms and smoking and intake of fruit and vegetables in relation to lung cancer. Lung Cancer. 2007 Feb;55(2):137-144.

Soria JC, Rodriguez M, Liu DD, Lee JJ, Hong WK, Mao L. Aberrant promoter methylation of multiple genes in bronchial brush samples from former cigarette smokers. Cancer Res. 2002 Jan 15;62(2):351-5.

Sozzi G, Sard L, De Gregorio L, Marchetti A, Musso K, Buttitta F, Tornielli S, et al. Association between cigarette smoking and FHIT gene alterations in lung cancer. Cancer Res. 1997 Jun 1;57(11):2121-3.

Sprenger R, Schlagenhaufer R, Kerb R, Bruhn C, Brockmöller J, Roots I, Brinkmann U. Characterization of the glutathione S-transferase GSTT1 deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype-phenotype correlation. Pharmacogenetics. 2000 Aug;10(6):557-565.

Sreeja L, Syamala V, Hariharan S, Madhavan J, Devan SC, Ankathil R. Possible risk modification by CYP1A1, GSTM1 and GSTT1 gene polymorphisms in lung cancer susceptibility in a South Indian population. J Hum Genet. 2005;50(12):618-627.

Sreeja L, Syamala V, Raveendran PB, Santhi S, Madhavan J, Ankathil R. p53 Arg72Pro polymorphism predicts survival outcome in lung cancer patients in Indian population. Cancer Invest. 2008 Feb;26(1):41-46.

Steinberg D, Colla P. (1997) CART- classification and regression trees. San Diego: Salford Systems.

Steiner M, Bastian M, Schulz WA, Pulte T, Franke KH, Röhring A, Wolff JM, Seiter H, Schuff-Werner P. Phenol sulphotransferase SULT1A1 polymorphism in prostate cancer: lack of association. Arch Toxicol. 2000 Jul;74(4-5):222-225.

Storey A, Thomas M, Kalita A, Harwood C, Gardiol D, Mantovani F, Breuer J, Leigh IM, Matlashewski G, Banks L. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. Nature. 1998 May 21;393(6682):229-2234.

Stücker I, Hirvonen A, de Waziers I, Cabelguenne A, Mitrunen K, Cénée S, Koum-Besson E, Hémon D, Beaune P, Loriot MA. Genetic polymorphisms of glutathione S-transferases as modulators of lung cancer susceptibility. Carcinogenesis. 2002 Sep;23(9):1475-1481.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50

Subramanian J, Govindan R. Lung cancer in never smokers: a review. J Clin Oncol. 2007 Feb 10;25(5):561-70.

Sugita M, Geraci M, Gao B, Powell RL, Hirsch FR, Johnson G, Lapadat R, Gabrielson E, et al. Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma. Cancer Res. 2002 Jul 15;62(14):3971-9.

Sul J, Yu GP, Lu QY, Lu ML, Setiawan VW, Wang MR, Guo CH, Yu SZ, Mu L, Cai L, Kurtz RC, Zhang ZF. P53 Codon 72 polymorphisms: a case-control study of gastric cancer and potential interactions.Cancer Lett. 2006 Jul 18;238(2):210-223.

Swafford DS, Middleton SK, Palmisano WA, Nikula KJ, Tesfaigzi J, Baylin SB, Herman JG, Belinsky SA. Frequent aberrant methylation of p16INK4a in primary rat lung tumors. Mol Cell Biol. 1997 Mar;17(3):1366-74.

Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988 Jun 3;240(4857):1285-93.

Szymanowska A, Jassem E, Dziadziuszko R, Borg A, Limon J, Kobierska-Gulida G, Rzyman W, Jassem J. Increased risk of non-small cell lung cancer and frequency of somatic TP53 gene mutations in Pro72 carriers of TP53 Arg72Pro polymorphism. Lung Cancer. 2006 Apr;52(1):9-14.

Tabor HK, Risch NJ, Myers RM. Opinion: candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 2002;3:391–7.

Taioli E, Ford J, Trachman J, Li Y, Demopoulos R, Garte S. Lung cancer risk and CYP1A1 genotype in African Americans. Carcinogenesis. 1998 May;19(5):813-817.

Taioli E, Gaspari L, Benhamou S, Boffetta P, Brockmoller J, Butkiewicz D, Cascorbi I, et al. Polymorphisms in CYP1A1, GSTM1, GSTT1 and lung cancer below the age of 45 years. Int J Epidemiol. 2003 Feb;32(1):60-3.

Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A. 2002 Mar 19;99(6):3740-5.

Takeshima M, Saitoh M, Kusano K, Nagayasu H, Kurashige Y, Malsantha M, Arakawa T, Takuma T, Chiba I, Kaku T, Shibata T, Abiko Y. High frequency of hypermethylation of p14, p15 and p16 in oral pre-cancerous lesions associated with betel-quid chewing in Sri Lanka. J Oral Pathol Med. 2008 Sep;37(8):475-9.

Tan KD, Zhu Y, Tan HK, Rajasegaran V, Aggarwal A, Wu J, Wu HY, Hwang J, Lim DT, Soo KC, Tan P. Amplification and overexpression of PPFIA1, a putative 11q13 invasion suppressor gene, in head and neck squamous cell carcinoma. Genes Chromosomes Cancer. 2008 Apr;47(4):353-62.

Tandle AT, Sanghvi V, Saranath D. Determination of p53 genotypes in oral cancer patients from India. Br J Cancer. 2001 Mar 23;84(6):739-742.

Tang X, Khuri FR, Lee JJ, Kemp BL, Liu D, Hong WK, Mao L. Hypermethylation of the death-associated protein (DAP) kinase promoter and aggressiveness in stage I non-small-cell lung cancer. J Natl Cancer Inst. 2000 Sep 20;92(18):1511-6.

Taniwaki M, Daigo Y, Ishikawa N, Takano A, Tsunoda T, Yasui W, Inai K, Kohno N, Nakamura Y. Gene expression profiles of small-cell lung cancers: molecular signatures of lung cancer. Int J Oncol. 2006 Sep;29(3):567-75.

Tew KD. Glutathione-associated enzymes in anticancer drug resistance. Cancer Res. 1994 Aug 15;54(16):4313-20.

Thakker DR, Yagi H, Levin W, Wood AW, Conney AH, Jerina DM. Polycyclic aromatic hydrocarbons: metabolic activation to ultimate carcinogens. In: Anders MW, editor. Bioactivation of Foreign Compounds. Orlando (FL): Academic Press; 1985. p. 177 – 242.

Thippanna G, Venu K, Gopalakrishnaiah V, Reddy PN, Charan BG. A profile of lung cancer patients in Hyderabad. J Indian Med Assoc. 1999 Sep;97(9):357-9.

Thomas M, Kalita A, Labrecque S, Pim D, Banks L, Matlashewski G. Two polymorphic variants of wild-type p53 differ biochemically and biologically. Mol Cell Biol. 1999 Feb;19(2):1092-1100.

Thun MJ, Hannan LM, Adams-Campbell LL, Boffetta P, et al. Lung Cancer Occurrence in Never-Smokers: An Analysis of 13 Cohorts and 22 Cancer Registry Studies . PLoS Med. 2008 September; 5(9): e185.

Timofeeva M, Jäger B, Rosenberger A, Sauter W, Wichmann HE; KORA Study Group, Bickeböller H, Risch A. A multiplex real-time PCR method for detection of GSTM1 and GSTT1 copy numbers. Clin Biochem. 2009 Apr;42(6):500-509

To-Figueras J, Gene M, Gomez-Catalan J, Galan C, Firvida J, Fuentes M, Rodamilans M, Huguet E, Estape J, Corbella J. Glutathione-S-Transferase M1 and codon 72 p53 polymorphisms in a northwestern Mediterranean population and their relation to lung cancer susceptibility. Cancer Epidemiol Biomarkers Prev. 1996 May;5(5):337-342.

Toh Y, Kuninaka S, Oshiro T, Ikeda Y, Nakashima H, Baba H, Kohnoe S, Okamura T, Mori M, Sugimachi K. Overexpression of manganese superoxide dismutase mRNA may correlate with aggressiveness in gastric and colorectal adenocarcinomas. Int J Oncol. 2000 Jul;17(1):107-12.

Tokuhata Gk, Lilienfeld Am. Familial aggregation of lung cancer in humans. J Natl Cancer Inst. 1963 Feb;30:289-312.

Toyooka S, Maruyama R, Toyooka KO, McLerran D, Feng Z, Fukuyama Y, Virmani AK, Zochbauer-Muller S, Tsukuda K, Sugio K, Shimizu N, Shimizu K, Lee H, Chen CY, Fong KM, Gilcrease M, Roth JA, Minna JD, Gazdar AF. Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. Int J Cancer. 2003 Jan 10;103(2):153-60.

Toyooka S, Tokumo M, Shigematsu H, Matsuo K, Asano H, Tomii K, Ichihara S, Suzuki M, Aoe M, Date H, Gazdar AF, Shimizu N. Mutational and epigenetic evidence for independent pathways for lung adenocarcinomas arising in smokers and never smokers. Cancer Res. 2006 Feb 1;66(3):1371-5.

Toyooka S, Toyooka KO, Maruyama R, Virmani AK, Girard L, Miyajima K, et al. DNA methylation profiles of lung tumors. Mol Cancer Ther. 2001 Nov;1(1):61-7.

Toyooka S, Toyooka KO, Miyajima K, Reddy JL, Toyota M, Sathyanarayana UG, et al. Epigenetic down-regulation of death-associated protein kinase in lung cancers. Clin Cancer Res. 2003 Aug 1;9(8):3034-41.

Tran TN, Liu Y, Takagi M, Yamaguchi A, Fujii H. Frequent promoter hypermethylation of RASSF1A and p16INK4a and infrequent allelic loss other than 9p21 in betel-associated oral carcinoma in a Vietnamese non-smoking/non-drinking female population. J Oral Pathol Med. 2005 Mar;34(3):150-6.

Travis WD, Travis LB, Devesa SS. Lung cancer. Cancer. 1995 Jan 1;75(1 Suppl):191-202.

Trinh BQ, Barengo N, Naora H. Homeodomain protein DLX4 counteracts key transcriptional control mechanisms of the TGF-β cytostatic program and blocks the antiproliferative effect of TGF-β. Oncogene. 2011 Jun 16;30(24):2718-29.

Tsou JA, Shen LY, Siegmund KD, Long TI, Laird PW, Seneviratne CK, Koss MN, Pass HI, Hagen JA, Laird-Offringa IA. Distinct DNA methylation profiles in malignant mesothelioma, lung adenocarcinoma, and non-tumor lung. Lung Cancer. 2005 Feb;47(2):193-204.

Tuck-Muller CM, Narayan A, Tsien F, Smeets DF, Sawyer J, Fiala ES, Sohn OS, Ehrlich M. DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. Cytogenet Cell Genet. 2000;89(1-2):121-8.

Uematsu K, He B, You L, Xu Z, McCormick F, Jablons DM. Activation of the Wnt pathway in non small cell lung cancer: evidence of dishevelled overexpression. Oncogene. 2003 Oct 16;22(46):7218-21.

Vainio H, Boffetta P. Mechanisms of the combined effect of asbestos and smoking in the etiology of lung cancer. Scand J Work Environ Health. 1994 Aug;20(4):235-42.

Vavvas D, Li X, Avruch J, Zhang XF. Identification of Nore1 as a potential Ras effector. J Biol Chem. 1998 Mar 6;273(10):5439-42.

Vertino PM, Spillare EA, Harris CC, Baylin SB. Altered chromosomal methylation patterns accompany oncogene-induced transformation of human bronchial epithelial cells. Cancer Res. 1993 Apr 1;53(7):1684-9.

Vineis P, Alavanja M, Buffler P, Fontham E, Franceschi S, Gao YT, Gupta PC, Hackshaw A, et al. Tobacco and cancer: recent epidemiological evidence. J Natl Cancer Inst. 2004 Jan 21;96(2):99-106.

Vineis P, Anttila S, Benhamou S, Spinola M, Hirvonen A, Kiyohara C, Garte SJ, Puntoni R, Rannug A, Strange RC, Taioli E. Evidence of gene gene interactions in lung carcinogenesis in a large pooled analysis. Carcinogenesis. 2007 Sep;28(9):1902-5.

Vineis P, Veglia F, Benhamou S, Butkiewicz D, Cascorbi I, Clapper ML, Dolzan V, et al. CYP1A1 T3801 C polymorphism and lung cancer: a pooled analysis of 2451 cases and 3358 controls. Int J Cancer. 2003 May 1;104(5):650-7.

Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst. 2004 Mar 17;96(6):434-442.

Wang J, Wang B, Chen X, Bi J. The prognostic value of RASSF1A promoter hypermethylation in non-small cell lung carcinoma: a systematic review and meta-analysis. Carcinogenesis. 2011 Mar;32(3):411-6.

Wang M, Vikis HG, Wang Y, Jia D, Wang D, Bierut LJ, Bailey-Wilson JE, et al. Identification of a novel tumor suppressor gene p34 on human chromosome 6q25.1.Cancer Res. 2007 Jan 1;67(1):93-9.

Wang T, Wang XG, Xu JH, Wu XP, Qiu HL, Yi H, Li WX. Overexpression of the human ZNF300 gene enhances growth and metastasis of cancer cells through activating NF-kB pathway. J Cell Mol Med. 2012 May;16(5):1134-45.

Wang X, Ghosh S, Guo SW Quantitative quality control in microarray image processing and data acquisition. Nucleic Acids Res. 2001 Aug 1;29(15):E75-5.

Wang Y, Shao C, Shi CH, Zhang L, Yue HH, Wang PF, Yang B, Zhang YT, Liu F, Qin WJ, Wang H, Shao GX. Change of the cell cycle after flutamide treatment in prostate cancer cells and its molecular mechanism. Asian J Androl. 2005 Dec;7(4):375-80.

Wang Y, Spitz MR, Schabath MB, Ali-Osman F, Mata H, Wu X. Association between glutathione S-transferase p1 polymorphisms and lung cancer risk in Caucasians: a case-control study. Lung Cancer. 2003 Apr;40(1):25-32

Wang Y, Spitz MR, Tsou AM, Zhang K, Makan N, Wu X. Sulfotransferase (SULT) 1A1 polymorphism as a predisposition factor for lung cancer: a case-control analysis. Lung Cancer. 2002 Feb;35(2):137-42.

Wang YC, Chen CY, Chen SK, Chang YY, Lin P. p53 codon 72 polymorphism in Taiwanese lung cancer patients: association with lung cancer susceptibility and prognosis. Clin Cancer Res. 1999 Jan;5(1):129-134.

Wang YC, Lee HS, Chen SK, Chang YY, Chen CY. Prognostic significance of p53 codon 72 polymorphism in lung carcinomas. Eur J Cancer. 1999 Feb;35(2):226-230.

Wary KK, Sharan RN. Aqueous extract of betel-nut of north-east India induces DNA-strand breaks and enhances rate of cell proliferation in vitro. Effects of betel-nut extract in vitro. J Cancer Res Clin Oncol. 1988;114(6):579-82.

Waxman DJ. Glutathione S-transferases: role in alkylating agent resistance and possible target for modulation chemotherapy--a review. Cancer Res. 1990 Oct 15;50(20):6449-54.

Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockhart DJ, Burger RA, Hampton GM. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci U S A. 2001 Jan 30;98(3):1176-81.

Wen CP, Tsai MK, Chung WS, Hsu HL, Chang YC, Chan HT, Chiang PH, Cheng TY, Tsai SP. Cancer risks from betel quid chewing beyond oral cancer: a multiple-site carcinogen when acting with smoking. Cancer Causes Control. 2010 Sep;21(9):1427-1435.

Wen CP, Tsai SP, Cheng TY, Chen CJ, Levy DT, Yang HJ, Eriksen MP. Uncovering the relation between betel quid chewing and cigarette smoking in Taiwan. Tob Control. 2005 Jun;14 Suppl 1:i16-22.

Wenzlaff AS, Cote ML, Bock CH, Land SJ, Schwartz AG. GSTM1, GSTT1 and GSTP1 polymorphisms, environmental tobacco smoke exposure and risk of lung cancer among never smokers: a population-based study. Carcinogenesis. 2005 Feb;26(2):395-401.

Weston A, Perrin LS, Forrester K, Hoover RN, Trump BF, Harris CC, Caporaso NE. Allelic frequency of a p53 polymorphism in human lung cancer. Cancer Epidemiol Biomarkers Prev. 1992 Sep-Oct;1(6):481-483.

Winton T, Livingston R, Johnson D, Rigas J, Johnston M, Butts C, Cormier Y, Goss G, et al. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. N Engl J Med. 2005 Jun 23;352(25):2589-97.

Wistuba II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, Fondon JW 3rd, Garner HR, et al. High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. Cancer Res. 2000 Apr 1;60(7):1949-60.

Wistuba II, Gazdar AF. Lung cancer preneoplasia. Annu Rev Pathol. 2006;1:331-48.

Wong CF, Liyou N, Leggett B, Young J, Johnson A, McManus ME. Association of the SULT1A1 R213H polymorphism with colorectal cancer. Clin Exp Pharmacol Physiol. 2002 Sep;29(9):754-758.

World Health Organization. Guidelines for Controlling and Monitoring the Tobacco Epidemic. Geneva: World Health Organization, 1998

Wu MT, Chen MC, Wu DC. Influences of lifestyle habits and p53 codon 72 and p21 codon 31 polymorphisms on gastric cancer risk in Taiwan. Cancer Lett. 2004 Mar 8;205(1):61-68.

Wu X, Zhao H, Amos CI, Shete S, Makan N, Hong WK, Kadlubar FF, Spitz MR. p53 Genotypes and Haplotypes Associated With Lung Cancer Susceptibility and Ethnicity. J Natl Cancer Inst. 2002 May 1;94(9):681-690.

Xiao H, Singh SV. p53 regulates cellular responses to environmental carcinogen benzo[a]pyrene-7,8-diol-9,10-epoxide in human lung cancer cells. Cell Cycle. 2007 Jul 15;6(14):1753-1761.

Xu S, Wang Y, Roe B, Pearson WR. Characterization of the human class Mu glutathione S-transferase gene cluster and the GSTM1 deletion. J Biol Chem. 1998 Feb 6;273(6):3517-27.

Yadav DS, Devi TR, Ihsan R, Mishra AK, Kaushal M, Chauhan PS, Bagadi SA, Sharma J, Zamoawia E, Verma Y, Nandkumar A, Saxena S, Kapur S. Polymorphisms of glutathione-S-transferase genes and the risk of aerodigestive tract cancers in the Northeast Indian population. Genet Test Mol Biomarkers. 2010 Oct;14(5):715-723

Yagil Y, Yagil C. Candidate genes, association studies and haplotype analysis in the search for the genetic basis of hypertension. J Hypertens. 2004 Jul;22(7):1255-8.

Yanagawa N, Tamura G, Oizumi H, Takahashi N, Shimazaki Y, Motoyama T. Promoter hypermethylation of tumor suppressor and tumor-related genes in non-small cell lung cancers. Cancer Sci. 2003 Jul;94(7):589-92.

Ye Z, Song H, Higgins JP, Pharoah P, Danesh J. Five glutathione s-transferase gene variants in 23,452 cases of lung cancer and 30,397 controls: meta-analysis of 130 studies. PLoS Med. 2006 Apr;3(4):e91.

Yi SY, Lee WJ. A p53 genetic polymorphism of gastric cancer: difference between early gastric cancer and advanced gastric cancer. World J Gastroenterol. 2006 Oct 28;12(40):6536-6539.

Yoshida K, Osawa K, Kasahara M, Miyaishi A, Nakanishi K, Hayamizu S, Osawa Y, Tsutou A, Tabuchi Y, Shimada E, Tanaka K, Yamamoto M, Takahashi J. Association of CYP1A1, CYP1A2, GSTM1 and NAT2 gene polymorphisms with colorectal cancer and smoking. Asian Pac J Cancer Prev. 2007 Jul-Sep;8(3):438-444.

Youlden DR, Cramb SM, Baade PD. The International Epidemiology of Lung Cancer: geographical distribution and secular trends. J Thorac Oncol. 2008 Aug;3(8):819-31.

Yu KD, Di GH, Fan L, Wu J, Hu Z, Shen ZZ, Huang W, Shao ZM. A functional polymorphism in the promoter region of GSTM1 implies a complex role for GSTM1 in breast cancer. FASEB J. 2009 Jul;23(7):2274-87.

Yuanlong H, Haifeng J, Xiaoyin Z, Jialin S, Jie L, Li Y, Huahong X, Jiugang S, Yanglin P, Kaichun W, Jie D, Daiming F. The inhibitory effect of p75 neurotrophin receptor on growth of human hepatocellular carcinoma cells. Cancer Lett. 2008 Sep 8;268(1):110-9.

Zhang B, Zhu W, Yang P, Liu T, Jiang M, He ZN, Zhang SX, Chen WQ, Chen W. Cigarette smoking and p16INK4α gene promoter hypermethylation in non-small cell lung carcinoma patients: a meta-analysis. PLoS One. 2011;6(12):e28882

Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, Vishwanatha JK, Santella RM, Morabia A. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics. 2011 May;6(5):623-9.

Zhang H, Ahmadi A, Arbman G, Zdolsek J, Carstensen J, Nordenskjöld B, Söderkvist P, Sun XF. Glutathione S-transferase T1 and M1 genotypes in normal mucosa, transitional mucosa and colorectal adenocarcinoma. Int J Cancer. 1999 Apr 20;84(2):135-8.

Zhang X, Diab IH, Zehner ZE. ZBP-89 represses vimentin gene transcription by interacting with the transcriptional activator, Sp1. Nucleic Acids Res. 2003 Jun 1;31(11):2900-14.

Zhang ZY, Fasco MJ, Huang L, Guengerich FP, Kaminsky LS. Characterization of purified human recombinant cytochrome P4501A1-Ile462 and -Val462: assessment of a role for the rare allele in carcinogenesis. Cancer Res. 1996 Sep 1;56(17):3926-33.

Zhao M, Lewis R, Gustafson DR, Wen WQ, Cerhan JR, Zheng W. No apparent association of GSTP1 A(313)G polymorphism with breast cancer risk among postmenopausal Iowa women. Cancer Epidemiol Biomarkers Prev. 2001 Dec;10(12):1301-1320.

Zheng W, Xie D, Cerhan JR, Sellers TA, Wen W, Folsom AR. Sulfotransferase 1A1 polymorphism, endogenous estrogen exposure, well-done meat intake, and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2001 Feb;10(2):89-94.

Zhong S, Howie AF, Ketterer B, Taylor J, Hayes JD, Beckett GJ, Wathen CG, Wolf CR, Spurr NK. Glutathione S-transferase mu locus: use of genotyping and phenotyping assays to assess association with lung cancer susceptibility. Carcinogenesis. 1991 Sep;12(9):1533-1537.

Zhou W, Liu G, Thurston SW, Xu LL, Miller DP, Wain JC, Lynch TJ, Su L, Christiani DC. Genetic polymorphisms in N-acetyltransferase-2 and microsomal epoxide hydrolase, cumulative cigarette smoking, and lung cancer. Cancer Epidemiol Biomarkers Prev. 2002 Jan;11(1):15-21.

Zöchbauer-Müller S, Fong KM, Virmani AK, Geradts J, Gazdar AF, Minna JD. Aberrant promoter methylation of multiple genes in non-small cell lung cancers. Cancer Res. 2001 Jan 1;61(1):249-55.

Zomawia E, Zohmingthanga J, Lalchhanhimi, Zonunsiama, Saia (2010) Population Based Cancer Registry, Mizoram. In.:, Three year report of the Population Based Cancer Registries, 2006–2008. First Report of 20 PBCRs in India. National Cancer Registry Programme, Indian Council of Medical Research, Bangalore.. pp 330–380.

# *Appendix*

# *Appendix*

## Preparation of reagents

### Preparation of 1.2% Formaldehyde-agarose gel (For 50 ml)

Add 0.6 gm Agarose in 37 ml of DEPC $H_2O$. Boil in microwave for 3-5 minute at 700-800 w. Cool at RT and add 5.0 ml of 10X MOPS Buffer + 8.0 ml of 37% Formaldehyde. Shake vigorously and pour in the casting plate. (Gel will solidify within 30-45 min.)

### Preparation of RNA Sample for loading (20µl)

Add solution of Formamide (10µl), Formaldehyde (4µl), DEPC $H_2O$ (2µl), 10X MOPS (2µl), 6x gel lading buffer(2µl), Etbr (0.1 µl) to 1 µg of RNA . Heat it at 650C for 10 minutes. Immediately Plunge the sample tube in ice. Load the 20µl of RNA sample into the Formaldehyde-agarose gel.

### Phosphate Buffer Saline (PBS) (1 litre) pH=7.4

8gm of Sodium Chloride (NaCl), 2gm of Potassium Chloride (KCl), 1.44gm of Disodium hydrogen phosphate ($Na_2HPO_4$) and 0.2gm of Potassium dihydrogen orthophosphate ($KH_2PO_4$), were dissolved in 800ml of dd$H_2O$. pH was set to 7.4 with HCl. Final volume was made up to 1 liter and sterilized by autoclaving at 15lb/ sq.in for 20 minutes and stored at room temperature.

### Paraformaldehyde (PF) pH=7.4

**4% Stock solution (100 ml) :** 4 gms of paraformaldehyde + 50 ml of distilled water + 10 ml of 10x PBS + few drops of 2M NaOH was heated on magnetic stirrer in a fume hood. When the solution became clear it was removed from the heat and after cooling pH was adjusted to 7.2. Finally volume was adjusted with distilled water to 100 ml.

**1% Working solution (10 ml):** 2.5 ml of Stock PF (4%) + 7.5 ml of 1x PBS.

**1M Tris:** 121.1gm of Tris base was dissolved in 800ml of double distilled water and pH set (6.8, 7.4, 8.0) with concentrated HCl. Volume was made up to 1liter and autoclaved.

**0.5M EDTA:** 186.1gm of disodium EDTA.-2$H_2O$ was added in 800ml of double distilled water, stirred vigorously on a stirrer, pH set to 8.0 with NaOH (~20 gm of NaOH pellets) and volume made up to1liter and autoclaved.

**3M sodium acetate:** 204.5gm of $C_2H_3O_2Na$. 3$H_2O$ was dissolved in 400ml of dd$H_2O$, pH set to 5.3 with glacial acetic acid, volume made up to 500 ml and autoclaved.

**10% SDS:** 10gm of electrophoresis grade SDS was dissolved in 70ml of dd$H_2O$, heated at 60°C to dissolve and the volume made up to 100ml.

**Ethidium Bromide (10 mg/ml):** 10mg of ethidium bromide was dissolved in 1ml dd$H_2O$, stored in opaque bottle.

**Calcium Chloride (0.1 M):** 1.47gm of $CaCl_2$.2$H_2O$ was dissolved in 100ml of dd$H_2O$ and sterilized by autoclaving.

**DEPC water:** 0.1% diethylpyrocarbonate was added to 1ltr double distilled water in a fume hood and mixed well. After incubating it for 1hr at 37°C it was autoclaved.

**10 X TAE buffer (Tris acetate, EDTA):** 4.84gm of Tris base in 80ml of dd$H_2O$ was dissolved and 1.2ml of glacial acetic acid and 2ml of 0.5 EDTA pH 8.0 were added. Final volume was made up to 100ml.

**DNA loading dye (6X):** 0.2gm bromophenol blue, 0.2gm of xylene cyanol and 30ml of glycerol were dissolved and volume set to 100 by autoclaved dd$H_2O$.

**10X MOPS Electrophoresis Buffer**

Dissolve 41.8 gm of MOPS (3'N Morpholino Propanesulfonic acid) in 700 ml of sterile 0.1% DEPC (Diethyl pyrocarbonate) water. Adjust the pH to 7.0 with 2N NaOH. Add 20 ml of DEPC-treated 1M sodium acetate and 20 ml of DEPC treated 0.5M EDTA, pH 8.0. Adjust the volume of the solution to 1L with DEPC treated water. Sterilize solutions by autoclaving. Store the buffer at room temperature in dark condition.

# *List of Publications*

# *List of Publications*


1. **Ihsan R**, Chauhan PS, Mishra AK, Singh LC, Sharma JD, Zomawia E, Verma Y, Kapur S, Saxena S. Investigation on copy number polymorphism of GSTM1 and GSTT1 in susceptibility to lung cancer in a high-risk population from North-East India. Indian Journal of Medical Research. (Accepted)

2. Chauhan PS, **Ihsan R**, Mishra AK, Yadav DS, Saluja S, Mittal V, Saxena S, Kapur S. High order interactions of xenobiotic metabolizing genes and P53 codon 72 polymorphisms in acute leukemia.Environ Mol Mutagen. 2012 Oct;53(8):619-30.

3. **Ihsan R**, Chauhan PS, Mishra AK, Yadav DS, Kaushal M, Sharma JD, Zomawia E, Verma Y, Kapur S, Saxena S Multiple analytical approaches reveal distinct gene-environment interactions in smokers and non smokers in lung cancer. PLoS One. 2011;6(12):e29431.

4. **Ihsan R**, Devi TR, Yadav DS, Mishra AK, Sharma J, Zomawia E, Verma Y, Phukan R, Mahanta J, Kataki AC, Kapur S, Saxena S. Investigation on the role of p53 codon 72 polymorphism and interactions with tobacco, betel quid, and alcohol in susceptibility to cancers in a high-risk population from North East India. DNA Cell Biol. 2011 Mar;30(3):163-71.

5. Chauhan PS, **Ihsan R**, Yadav DS, Mishra AK, Bhushan B, Soni A, Kaushal M, Devi TR, Saluja S, Gupta DK, Mittal V, Saxena S, Kapur S. Association of glutathione S-transferase, EPHX, and p53 codon 72 gene polymorphisms with adult acute myeloid leukemia. DNA Cell Biol. 2011 Jan;30(1):39-46.

6. Kaushal M, Mishra AK, Raju BS, **Ihsan R**, Chakraborty A, Sharma J, Zomawia E, Verma Y, Kataki A, Kapur S, Saxena S. Betel quid chewing as an environmental risk factor for breast cancer. Mutat Res. 2010 Dec 21;703(2):143-8.

7. Yadav DS, Devi TR, **Ihsan R**, Mishra AK, Kaushal M, Chauhan PS, Bagadi SA, Sharma J, Zamoawia E, Verma Y, Nandkumar A, Saxena S, Kapur S. Polymorphisms of glutathione-S-transferase genes and the risk of aerodigestive tract cancers in the Northeast Indian population.Genet Test Mol Biomarkers. 2010 Oct;14(5):715-23.

8. **Ihsan R**, Chattopadhyay I, Phukan R, Mishra AK, Purkayastha J, Sharma J, Zomawia E, Verma Y, Mahanta J, Saxena S, Kapur S. Role of epoxide hydrolase 1 gene polymorphisms in esophageal cancer in a high-risk area in India.J Gastroenterol Hepatol. 2010 Aug;25(8):1456-62.

9. Thoudam RD, Yadav DS, Mishra AK, Kaushal M, **Ihsan R**, Chattopadhyay I, Chauhan PS, Sarma J, Zomawia E, Verma Y, Nandkumar A, Mahanta J, Phukan R, Kapur S, Saxena S Distribution of glutathione S-transferase T1 and M1 genes polymorphisms in North East Indians: a potential report.Genet Test Mol Biomarkers. 2010 Apr;14(2):163-9.

***In Communication***

1. **Ihsan R**, Chauhan PS , Raju BS, Sharma J, Zomawia E, Jaiswal A, Gupta K, Kapur S, Saxena S. Study on p16, RASSF1A, DAPK and GSTP1 gene methylation status in Indian lung cancer patients (Communicated In: *Journal of Cancer Research and Clinical Oncology*)

2. **Ihsan R**, Chauhan PS, Sharma J, Zomawia E, Kapur S, Saxena S. Gene expression profile of non-small cell lung cancer in high risk population from North East India (Communicated In: OMICS: A Journal of Integrative Biology

**Abstracts Presented in International/National conferences**

1. **"**Gene expression profile of non-small cell lung cancer in high risk population from North". AACR's first Conference in India "New Horizons in Cancer Research: Biology to Prevention to Therapy" held at Gurgaon, Delhi from December 13-16, 2011.

2. "Association between polymorphism of xenobiotic-metabolizing genes and the risk of acute leukemia". AACR's first Conference in India "New Horizons in Cancer Research: Biology to Prevention to Therapy" held at Gurgaon, Delhi from December 13-16, 2011.

3. "Glutathione S-transferase and Microsomal Epoxide Hydrolase Gene Polymorphisms and Risk of Acute myeloid leukemia". First international conference "Hematologic Malignancies: Bridging the Gap 2010" Singapore City, Singapore, Feb 5-7, 2010.

4. "Significance of TP53 codon 72 polymorphism in lung and breast cancer showing different xenobiotic potential" in Thirteenth Human Genome Meeting (HUGO) 2008, Hyderabad, India from September 27$^{th}$-30$^{th}$ 2008.

5. "Study of Interactions between Glutathione-S-Transferase Metabolic Enzymes and Smoking in Lung Cancer" at IACRCON-2008 and 27th Annual Convention of Indian Association For Cancer Research. Ahmadabad, India from February 6$^{th}$–9$^{th}$, 2008.

*6.* "Assessment of Breast cancer risk: Genotype polymorphism in estrogen synthesizing and metabolizing genes and their contribution in breast cancer susceptibility" at IACRCON-2008 and 27th Annual Convention of Indian Association for Cancer Research. Ahmadabad, India from February 6$^{th}$–9$^{th}$, 2008.

7. Prevalence of Glutathione-S-transferase (GST) polymorphisms in tobacco-associated malignancies in high risk Northeast Indian populatio. at International Symposium on Cancer Biology at National Institute of Immunology New Delhi, November 12-14, 2007.

# *Brief Biography of Candidate*

# *Brief Biography of the Candidate*


**Name** Mrs. Rakhshan Ihsan

**Date of Birth** 13th June 1981

**Postal Address** Lab. No.611,
Tumor Biology Lab,
National Institute of Pathology (ICMR),
Culture, Institute of Pathology,
Safdarjung Hospital Campus,
Post Box 4909, New Delhi-110029, India.

**Phone** 09818940872, 011-26198402-06

**Fax** 011-26198401

**E-mail address** meetrakhshan@yahoo.co.in


**Education Qualification**

| Examination passed | University/ Board | Year | Subjects Studied | Percentage |
|---|---|---|---|---|
| ICSE (Xth) | City Montessori School, Lucknow | 1997 | English, Hindi, Math, Social Science, Science | 88% |
| ISC (XIIth) | City Montessori School, Lucknow | 1999 | English, Math, Physics, Chemistry, Biology | 72% |
| B.Sc (General-ZBC) | University of Lucknow, Lucknow | 2003 | Zoology, Botany, Chemistry | 67% |
| M.Sc (Biochemistry) | Jamia Hamdard, New Delhi | 2005 | Microbiology, Immunology, Biotechnology, Biostatistics, Clinical Biochemistry, Molecular Biology, | 72% |

**Fellowship Awards**

- Qualified CSIR- UGC/ NET (National Eligibility Test), December -2005 for CSIR – JRF (Junior Research Fellowship) in Life Sciences Discipline.
- Qualified Indian Council of Medical Research-Junior Research Fellow, July-2005, ICMR-JRF in Life Sciences Discipline
- Qualified Graduate Aptitude Test Exam (GATE, 92.04 percentile), 2005

**Research Experience**

- Worked as Research Scholar in Molecular Pathology Laboratory, Dept. of Pathology, AIIMS. New Delhi. From May 2005 to November 2005.
- Pursuing Doctoral Program from December 2005 onwards at IOP as CSIR-JRF.
- Pursuing Doctoral Program from August 2008 onwards at IOP as CSIR-SRF.

**Workshop/Training Received**

- Attended the Fourth Workshop on "Genetic Epidemiological Methods for Dissection of Complex Human Traits" organized by TCG-ISI Centre for Population genomics (CpG) from Feb 23-28, 2009 at Kolkatta.
- Attended the "National Workshop on Microarray Technology" held on $16^{th}$ - $18^{th}$ April 2007 organized by Institute of Pathology (ICMR), Safdarjung Hospital Campus, New Delhi, India.

**Abstracts Presented in International/National conferences**

- **"**Gene expression profile of non-small cell lung cancer in high risk population from North". AACR's first Conference in India "New Horizons in Cancer Research: Biology to Prevention to Therapy" held at Gurgaon, Delhi from December 13-16, 2011.

- "Association between polymorphism of xenobiotic-metabolizing genes and the risk of acute leukemia". AACR's first Conference in India "New Horizons in Cancer Research: Biology to Prevention to Therapy" held at Gurgaon, Delhi from December 13-16, 2011.

- "Glutathione S-transferase and Microsomal Epoxide Hydrolase Gene Polymorphisms and Risk of Acute myeloid leukemia". First international conference "Hematologic Malignancies: Bridging the Gap 2010" Singapore City, Singapore, Feb 5-7, 2010.

- "Significance of TP53 codon 72 polymorphism in lung and breast cancer showing different xenobiotic potential" in Thirteenth Human Genome Meeting (HUGO) 2008, Hyderabad, India from September $27^{th}$-$30^{th}$ 2008.

- "Study of Interactions between Glutathione-S-Transferase Metabolic Enzymes and Smoking in Lung Cancer" at IACRCON-2008 and 27th Annual Convention of Indian Association For Cancer Research. Ahmadabad, India from February $6^{th}$–$9^{th}$ , 2008

- "Assessment of Breast cancer risk: Genotype polymorphism in estrogen synthesizing and metabolizing genes and their contribution in breast cancer susceptibility" at IACRCON-2008 and 27th Annual Convention of Indian Association For Cancer Research. Ahmadabad, India from February $6^{th}$–$9^{th}$, 2008.

- Prevalence of Glutathione-S-transferase (GST) polymorphisms in tobacco-associated malignancies in high risk Northeast Indian populatio. at International Symposium on Cancer Biology at National Institute of Immunology New Delhi, November 12-14, 2007.

# *Brief Biography of the Supervisor*

## *Brief Biography of the Supervisor*

| | | |
|---|---|---|
| Name | : | **Dr. (Mrs.) Sunita Saxena** |
| Date of Birth | : | 16th September, 1952 |
| Designation | : | **Director** |
| Address | : | Institute of Pathology-ICMR<br>Safdarjang Hospital Campus,<br>Post Box No.4909,<br>New Delhi – 110029 |

Academic Qualifications : **M.B.B.S., D.C.P., M.D.(Path)**

| Degree | Institute | Year | Remarks |
|---|---|---|---|
| M.B.B.S. | M.L.N.Medical College, Allahabad | 1974 | |
| D.C.P. (Clinical Pathology | L.L.R.M.Medical College Meerut | 1978 | **Received Gold Medal & Merit Certificate** |
| M.D. (Pathology) | -do- | 1981 | |

Details of Employment

| Post | Duration | Institute |
|---|---|---|
| Research Officer | April 1981 to Dec.1985 | Institute of Pathology, New Delhi. |
| Sr. Research Officer | Jan.1986 to Jan.1991 | -do |
| Asst. Director | 9th Jan.1991-2nd May, 1994 | -do- |
| Deputy Director | 2nd May, 1994 to 2nd May,'99 | -do- |
| Deputy Director (Sr.Gr) | 3rdMay, 99 to 8thApril, 2002 | |
| Deputy Director (Sr.Gr) & Officer In charge | 9thApril, 2002to 13th Dec., 2004 | -do- |
| Director | 14th Dec., 2004 till date | -do- |

Areas of Specialization : **Molecular Oncology, Oncopathology**

Areas of Interest **Breast Tumors,**

**Tobacco Associated cancers**

**Genito urinary cancers**

**Membership of National and International bodies**

**International: -** Life Member International Union against Cancer (U.I.C.C.)

**National:** Life Member: Indian Association of Cancer Research (IACR).

Life Member: Indian Association of Pathologists and Microbiologists (IAPM).

Life Member: Association for the promotion of DNA fingerprinting and other DNA technologies (ADNAT)

Life Member: Proteomic Society of India

Life member: Human Genomic Organization (HUGO)

**Trainings Received:-**

1. Trained for **'Culture of fastidious cells and modern techniques of cell manipulation'** at National Facility for animal Tissue and cell culture, Pune in Dec., 1991.

2. Trained in **'Genetic mutation detection techniques for BRCA 1 and BRCA2 genes in genomic DNA of Breast cancer patients'** at Unit Genetic Epidemiology at International Agency for Research on Cancer, Lyon, France in 1998.

3. Attended **Hands-on Training Course on Proteomics and DNA Micro arrays** held from 25th February to 10th March 2003 at CCMB, Hyderabad.

**Fellowships:-**

1. Awarded **WHO fellowship** to work on **Renal Pathology** at **St. Vincent's Hospital, Melbourne, Australia**, 1987.

2. Awarded **Yamigawa Yoshida Fellowship of U.I.C.C.** to work on "Mutation analysis of BRCA1 and BRCA2 genes in Breast Cancer Patients**"** at unit of Genetic Epidemiology **International Agency for Research on Cancer, Lyon, France** in 1998.

**Awards and Honors Received:-**

1. Awarded **Gold Medal** and **Merit Certificate** for securing highest marks in **Diploma in Clinical Pathology** (D.C.P.).
2. **K. C. Basu Mullick award** for best research work by **Indian Association of Pathologists and Microbiologists for year 2008.**

3. Received '**NOVARTIS ORATION AWARD 2006**" of Indian Council of Medical Research for her work on Breast cancer on 18th Sept. 2009.

4. Elected **Fellow of National Academy of Medical Sciences in 2010.**

5. Elected **Fellow of Indian College of Pathologist in 2010**

6. Paper entitled "Flow cytometric analyses of Th1 and Th2 cytokine production as a parameter of immunologic dysfunction in patients with superficial transitial cell Carcinoma" **received special appreciation award for the Best Poster presentation at 24th Annual Convention of Indian Association for Cancer Research & International Symposium of Human Papilloma virus and cervical cancer held at ICPO from 9th – 12th Feb., 2005.**

7. **Best Poster Award for poster entitled** "What Androgen Receptor CAG repeats polymorphism and p53 mutations/polymorphisms have to do with prostate cancer risk and progression? at "XXXIII Annual Conference of the ISHG and International Symposium on Genetics Revisited: the Genomics and Proteomics Advantage", Vishakhapatnam, India from 11-13 Feb 2008.

**Research Grants received**:-

1. Awarded research grant to work on **"Genetics of Breast Cancer in Indian Women"** by Indo-French Centre for Promotion of Advanced Research, **New Delhi (1999-2002).**

2. Awarded Research grant from Deptt. of Science and Technology for proposal entitled **" A new approach to the management of superficial bladder cancer. Role of in vitro Cytotoxicity assessment and immunologic enhancement"** (2000-2003).

3. **"Comprehensive study of carcinoma esophagus at North-East India-Multidiscipline approach"**. Collaborative, Multicentric ICMR Task Force project (2004-2007).

4. **"Microsatellite instability in androgen receptor gene and mismatch repair system in Prostate cancer in Indian males"** – Department of Science Technology (2005-2008).

5. **"Role of tobacco use in causation of cancer in north-east India"** – Indian Council of Medical Research Task Force project (2005-2008).

6. **"Effects of pesticide exposure in causation of cancer in north-east India"** – Indian Council of Medical Research Task Force project (2005-2008).

7. **"Establishment of Cell lines from Primary Breast Cancer"** – Indian Council of Medical Research. Task force project (2007-2010).

8. **"Study on Gene Expression and Hypermethylation Profiles in Early Onset Breast Cancer"** Department of Biotechnology (2008-2011)

9. "**Characterization of host immune factors associated with progression of superficial TCC of bladder by microarray analysis"** Indian Council of Medical Research (2009-2012)

10. "**Immunogenetic profile of Nasopharyngeal Cancer in a high prevalence region of Northeast India"** Department of Biotechnology (2010-2013)

11. **"Comparative study of Genetic, Clinical and Epidemiological Factors of Breast Cancer in Rural and Urban Area of India"** Indian Council of Medical Research Task force project (2009-2012).

12. **"Epigenetic studies in esophageal cancer in high risk region of Northeast India"** Department of Biotechnology, Twining Program for NER (2011-2013)

13. "**Genome wide Analysis of Genetic alterations in patients with Esophageal Cancer from Northeast India using Single Nucleotide Polymorphism arrays"** Indian Council of Medical Research (2011-2013)

14. **"Study on miRNA signatures associated with Breast cancer stem like cells (CSC) and their role in drug response"** Indian Council of Medical Research (2012-2015)

15. **"Targeted sequencing of Breast cancer specific genes in early onset breast carcinoma."** Indian Council of Medical Research (2012-2015)

**Academic Experience:-**

a. Diplomat of National Board Examinations (N.B.E.) – Institute is accredited for running this program since 1993.
    - **Core member, Supervisor & examiner**.
    - Guided 20 DNB dissertations as supervisor and co supervisor.
    - Nominated as Inspector, Examiner for theory & Practical examinations and paper setter.

b. Ph.D. - **Supervisor/Mentor**
   - **Supervisor** of student registered under GGSIPS University, Delhi, BITS, Pilani.
   - External Examiner for Ph.D. candidates of AIIMS, PGI, Chandigadh, Agra University, BITS, Pilani.
   - 4 students completed Ph.D and 6 are registered.

c. Member of project Review Committee of Divisions of **Non Communicable Diseases of Indian Council of Medical Research.**

d. Member of project Review Committee of Divisions of **Basic Medical Sciences of Indian Council of Medical Research.**

e. Member of **Scientific Advisory committee of Institute Of Cytology and Preventive Oncology, Noida**

f. Reviewer of papers of **Indian Journal of Medical Research, Human Mutation, Cancer Immunology Immunotherapy, World Journal of Surgical Oncology**, **BMC Cancer, Cancer Detection and Prevention, British Journal of Urology International, Indian Journal of Medical Research**

g. Member of **Scientific Advisory committee of National Jalma Institute for Leprosy and other Mycobacterial diseases, Agra.**

h. Member of **Scientific Advisory committee of Regional Medical Research Centre, Dibrugadh**

i. Member of **Scientific Advisory Group of Publication & Information Division Of Indian Council Of Medical Research**

j. **Chairperson of Institutional Ethical Committee of Safdarjang Hospital, New Delhi.**

k. Member of **Technical committee of Indian Council of Medical Research**

l. Member of **DBT sponsored DSMB on Curcumin Trial in Cancer Cervix**

m. **Senate member of BITS, Pilani.**

n. Nominated as the expert member of **Task force on Leprosy at ICMR.**

o. Nominated as the **member of Scientific Advisory Group** for creating a new centre for Environmental Health and Bhopal Gas Tragedy at Bhopal.

p. Nominated as the **member of the Data Safety Monitoring Board** (DSMB) on "BASANT Clinical Trial" of DBT, New Delhi.

q. Appointed as **Appraiser and Inspector** by National Board of Examination for assessment of DNB students and institutes.
r. Dr. Sunita Saxena has been nominated as the **expert member** of " ICMR-ICAR Joint Task force on the Epidemiology of Human and Animal Brucellosis".

s. Dr. Sunita Saxena has been nominated as **nodal officer** for getting ICMR university status.

**International Conferences attended.**

1. Presented a paper on **"Pattern of lymphokines in minimal change Nephrotic syndrome"** in 5th Asia Pacific Congress of Nephrology held in New Delhi during 9-12th Dec., 1992.

2. Presented paper on **'Role of Proto-oncogene, Growth Factor Receptor and Steroid Hormones on Malignant Human Mammary Epithelial Cancer Cells in vitro and vivo"** in XVI, International Cancer Congress (U.I.C.C.) at New Delhi, 30th-5th Nov., 1994.

3. **'Stage A carcinoma of Prostate'** paper presented at first conference of Nephrology, Urology and Transplantation Society of SAARC Countries held at A.I.I.M.S. , New Delhi during 24th-26th March, 1995.

4. **'Mutation profile of BRCA 1 / 2 genes in Indian patients'** paper presented at XV Asia Pacific Cancer Congress held at Chennai during Dec. 12-15, 1999.

5. **'Mutation profile of BRCA 1 / 2 mutations in worldwide population. The MAGIC project'** paper presented in meeting of American Association of Cancer Research AACR, 2001, LA, USA.

6. Attended the 7th International Symposium on Molecular Basis of Predictive Oncology and Intervention Strategies' and presented a paper **BRCA1 and BRCA2 Genes in Indian Breast Cancer Patients** held at Nice, France from 7th 10th Feb. 2004.

7. Attended the **"UICC World Cancer Congress and Centre for Disease Control and Prevention (CDC)"** held during 8th to 13th July, 2006 at Washington DC, U.S.A and presented paper **"Study of candidate genes associated with Breast Cancer Susceptibility in the Indian Women".**

8. Attended the NCRI Cancer Conference held at International Convention Centre in Birmingham, UK from 30th September - 3rd October 2007 and presented paper (oral and poster) entitled **"Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis"**

9. Attended World Cancer Congress-2008 held in Shanghai, China during 12th-17th June 2008 and presented paper **"Differential gene expression in familial and tobacco associated esophageal cancers in north-east region of India".**

10. Attended HUGO's 13th Human Genome Meeting, September 27 - 30 2008, Hyderabad. and presented papers-
    - **Prognostic value of TP53 Codon 72 polymorphism in oral cancer and stomach cancer in high risk region of India**
    - **Differential gene expression profile of stomach and oral cancer in high risk region of India.**
    - **Differential expression of MAPK and GPCR pathway in esophageal cancer of North-east region of India**
    - **Significance of TP53 codon 72 polymorphism in breast and lung cancer showing different xenobiotic potential spectrum**

11. Attended **First Symposium on HPV Vaccination in the Asia Pacific and Middle East Region held at Seoul, Korea** during1st to 3rd June'09.

12. Visited **University of Minnesota, USA as member of expert team of Indian Scientists on Cancer** for collaborative research projects.

13. Presented papers in **Conference of the Organisation for Oncology and Translational Research (OOTR), 6th Annual Conference on 26 and 27 February, 2010 at Kyoto Japan**
    - Genetic alterations in patients with esophageal cancer from high-risk region in India by SNP array. Sujala Kapur, Indranil Chattopadhyay, Rupkumar Phukan, Joydeep Purkayastha, Vikki Marshal, Amal Kataki, Jagdish Mohanta, David Bowtell, Sunita Saxena
    - Genome-wide analysis of genetic alterations in breast cancer patients from Northeast India using 10K SNP arrays. Sunita Saxena, Mishi Kaushal, Indranil Chatterjee, A. Bhatnagar, Chintamani, D. Bhatnagar, Sujala Kapur

14. Presented papers in **20th Asia Specific Cancer Conference, November 12-14, 2009,Japan.**
    - GENOME-WIDE ANALYSIS OF GENETIC ALTERATIONS IN ESOPHAGEAL SQUAMOUS CELL CARCINOMA BY SNP ARRAY. Sujala Kapur, Indranil Chattopadhyay, Rupkumar Phukan, Joydeep Purkayastha, Vikki Marshal, Amal Kataki, Jagdish Mohanta, David Bowtell, Sunita Saxena.
    - GENOME-WIDE ANALYSIS OF DNA COPY NUMBER VARIATIONS IN INDIAN BREAST CANCER PATIENTS USING HIGH-DENSITY SNP ARRAYS. Sunita Saxena, Mishi Kaushal Wasson, Indranil Chatterjee, A Bhatnagar, Dr Chintamani, D Bhatnagar, Sujala Kapur.

15. Presented paper entitled "**Molecular Profile of Esophageal Cancer in High Risk Region of India"** in 10th World Congress of OESO held at Boston, USA during 28-31 Aug'2010. (Abstract J. Clin. Gastroenterology. Vol: 45 (2), 2011.

16. Presented paper entitled "**Genomic alterations in breast cancer patients from Northeast India using 10K SNP arrays"** in BMC group conference **Beyond the Genome: The true gene count, human evolution and disease genomics** at Harvard Medical School, Boston, USA during11th-13th Oct'2010

17. Presented paper entitled "**Betel Quid Chewing A Risk Factor For Breast Cancer: Study Of Genomic Alterations"** at 16th Human Genome Meeting 2012 held at Sydney, Australia during 11th-14th March, 2012

**Invited Speaker**

1. **"Morphological patterns of Childhood Nephrotic Syndrome. Ultrastructure and Immunohistologic study"** at meeting of Delhi State Chapter of Indian Association of Pathologists & Microbiologists held in Feb. 1985.

2. **"Patterns of Lymphokines in Minimal changes nephritic Syndrome"** at meeting of Delhi State chapter of IAPM held in July, 1991.

3. **"Study of Differentiating markers in Breast Cancer" and 'Pathobiology of Prostate tumors'** in the workshop held on **"An overview of tumor Biology"** at Institute of Pathology in May 1993 in collaboration with Tata Memorial Hospital and Cancer Research Institute, Bombay.

4. **"Pathobiology of Prostate Malignancies"** at IX annual conference of Delhi regional Chapter of IAPM, 1994 held at Ram Manohar Lohia Hospital, New Delhi.

5. **"Tumor Markers in Paediatric Malignancies – Clinical Application"** in National Seminar on Paediatric malignancies in Feb. 96 at Safdarjang Hosptial, New Delhi.

6. **"Breast Cancer – Diagnosis"** popular lecture for 85th Centenary Celebration of ICMR in May'96 at Institute of Pathology, New Delhi.

7. **"Determinants of cell behavior in Breast Cancer"** in Sept.'97 at Delhi Breat Group meeting at Batra Hospital, New Delhi.

8. **"Genetic Predisposition of Breast Cancer in Indian Women"** talk given at Institute of Cytology and Preventive Oncology in Jan.'2000.

9. **"Genetic Predisposition of Breast Cancer in Indian Women – Clinical Significance"** invited talk at meeting of Delhi Breast Group in Feb.'2000.

10. **"Bioinformatics : Opportunities and Challenges for New Millennium"** at ICMR-WHO workshop on use of informatics in Biomedical Research, 13-15th Dec., 2000, New Delhi.

11. **"Breast Cancer : Genetics, Risks and Strategies"** in CME Programme on Oncopathology at A.H. Regional Cancer Center, Cuttack, 17-18th March, 2001.

12. **"Genetics and Prognostic Markers in Colorectal Cancer"** in XVIII Annual Conference of Association of Surgeons of India (Delhi Chapter), 24th March 2001, New Delhi.

13. **"Breast cancer Genetics : Risk assessment to prognostic implication"** in XI UP Chapter of Indian Association of Pathologists and Microbiologists held at LLRM Medical College, Meerut on 6-7th Oct. 2001.

14. **"Determinants of cell behavior in Breast cancer"** at meeting of Delhi State Chapter of IAPM held on Ist Dec. 2001 at Safdarjang Hospital, New Delhi.

15. Delivered a plenary talk on **"Early Onset Breast Cancer in Indian women and Genetic Susceptibility and Molecular characterstics"** on **the 24th Annual Convention of Indian Association for Cancer Research (IACR)** & International Symposium on Human Paillomavirus and Cervical Cancer at ICPO, NOIDA from 9th to 12th February, 2005.

16. Attended Symposium on "Biological and clinical relevance of placenta" and delivered talk on "**Hormones & Receptors in Placenta – Role in Fetal development Trophoblastic neoplasms"** at Department of Anatomy, Vardhman Mahavir Medical College & Safdarjang Hospital, New Delhi on 15th February, 2006.

17. Delivered talk on "**Study of candidate Genes Associated with Breast Cancer Susceptibility in Indian Women"** at International Symposium on Preventive and Predictive Molecular Diagnostics held on Januarly 21-22, 2006 at Dhirubhai Ambani Life Sciences, Mumbai.

18. Invited as a guest speaker in CME entitled "Ancillary Techniques in Anatomic Pathology from digital imaging to confocal imaging and laser microdissection" at 56th Annual Conference of the Indian Association of Pathologists and Microbiologists [APCON 2007] held at PGIMER, Chandigarh from 26th -29th November 2007 and delivered a talk on **" Molecular Biology of Cancer and Laser Capture Microscopy"**

19. Delivered a talk on "**Breast Cancer risk factors in North-East Indian women"** in Breast Con-2008 held at Guwahati on 7th March 2008.

20. Chaired session on Non-communicable diseases during conference on **"Show casing Science by Indian Women Scientist"** held on 8th – 10th March, 2008 and delivered a talk on **"Genetic Profile of Breast Cancer in Indian women".**

21. Delivered a talk on **"Genome-wide approach to identify prognostic markers for Esophageal Cancer"** in seminar on 'Prognostic and Predictive Factors in Cancer Management' at Foundation Day celebration & Workshop on Brachytherapy, Department of Radiotherapy at Convention Center, CSM Medical University, Lucknow on 14th December 2008.

22. Delivered a talk on **"Breast cancer in Indian women : Risk and prevention"** in 32nd Session of Indian – Social Science Congress (ISSC) held at Department of Biotechnology, Jamia Millia Islamia University, New Delhi on 18th December 2008.

23. Invited to deliver a talk on **"Understanding molecular biology of cancer using Genomic approaches"** at 63rd IAPM Kerala Chapter Meeing & 6th National CME in Pathology during 14-15th Feb. 2009 organised by Department of Pathology, Amrita Institute of Medical Sciences, Kochi.

24. Invited as **Chief Guest** to attend Conference on **"Emerging Trends in Life Sciences Research"** organized by BITS, Pilani and delivered Key-note address on **"Genome-wide approach to identify biomarkers for Esophageal cancer in North East India"** on 6th March 2009.

25. Delivered a talk on "**Molecular biology of Cancer by Genome-wide approaches"** at CME Pathology held on 18th Nov' 2009 at Maulana Azad Medical College, New Delhi. Invited as **Guest faculty** to give talk on **" Esophageal Cancer in North East region- Contribution of genetic vis-à-vis environmental factors "** at 29th Annual Convention of Indian Association of Cancer Research held at Amrita Institute of Medical Sciences, Cochin during 21st-23rd Feb.10.

26. Invited to give a talk on **"Understanding molecular biology of cancer using Genomic approaches"** at **National Symposium on Current Trends in Genomics and Proteomics** organized by Deshbandhu Gupta College, University of Delhi during 4th-5th Feb'10.

27. Invited as Guest faculty to give talk on**" Trends of Cancer in Indian Women : Can I prevent cancer"** for the International Conference on**" Empowering Women in Developing Countries through better health care and Nutrition"** held at BITS, Pilani during 22nd -24th April' 10. Dr Sunita Saxena was invited to deliver talk on **" Role of Electronmicroscopy in Renal Pathology"** at the International Conference entitled "**Renal Pathology for the Nephrologists**" organized by Department of Pathology, G.B. Pant Hospital, New Delhi during 22nd & 23rd January, 2011.

28. Dr. Saxena was invited as faculty member to1st Indo-USA initiative on "**Translational Cancer Prevention and Biomarkers workshop 2011**" held at Mazumdar-Shaw cancer Center,Bangaluru during 13th to 16th February, 2011 and
    1. Chaired the session on "**Prevention of Breast Cancer"**
    2. Acted as panel discussant in session on "**Genetic and Population Epidemiology**"and presented the talk "**Identification of Geo-ethnic variation in North-East India and their association with cancer risk**."

29. Invited to give talk on "**Applications of Cancer Genomics**" in the workshop "New Frontiers in Medicine" organized by Deptt. Of Pathology, M.L.N.Medical College, Allhabad on 10th Feb'2011.

30. Delivered talk on "**Applications of Cancer Genomics**" at workshop on "**Hands on Training in basic Molecular Biology Techniques**" at the Institute of Pathology, New Delhi during 1st to 4th March, 2011, as part of ICMR Centenary Celebrations.

31. Invited as Chief Guest at **National conference on "Current Trends in Advanced Biomedical Technology (CTA B-II)** organized by Department of Biosciences, Nehru Arts and Sciences College, Coimbatore, during 20th to 21st September, 2011 and delivered Key note address on "**Understanding molecular biology of cancer using Genomic approaches"**

32. Invited to deliver talk on "**Genome wise approach to identify Cancer Biomarker : Role of Microarray Technology**" in **1st Biennial Conference of IASN** organized by Agra Medical College in association of National JALMA Institute for Leprosy and Other Microbacterial Disease (ICMR), Agra held in Agra during 30th November – 1st December, 2011.

33. Invited to deliver talk on "**Application of Tissue Microarray in Pathology**"in "**Symposium on Antibody – based proteomics**" at annual meeting of **International Academy of Pathologists** held at Govt. Medical College Patiala during 1st-4th Dec. 2011.

34. Invited to deliver talk on "**Applications of Tissue Microarray in Pathology and Research**" in "**Workshop on Applications of Molecular Biology in Cancer Diagnostics"** organized by Dr. B. Borooah Cancer Institute Guwahati and National Institute of Pathology, New Delhi during 28th Feb. -1st March 2012 at BBCI Guwahati

**Publications: -**

Research Articles**:**

1. Chandra M., **Saxena, S.**, Dave P.K., Kaza R. M. and Saxena, H.M.K.: **Proliferative myositis (a pseudosarcomatous lesion of skeletal muscle) occurring in an infant**. Ind. J. Path. & Microbiol : 26, 213, 1983.

2. Nagar P., **Saxena S.** (nee Sinha). Pratap V.K. and Mehrotra M.L.: **Malignant haemangiopericytoma of the uterus**. **A case report with brief review of literture**. J. of Obst. Gynaec. Of India : 34,753, 1984.

3. **Saxena S.**, Andal, A. and Saxena H.M.K.: **Ultrastructure study of minimal change nephrotic syndrome – a clinico-morphologic correlation**. Ind. J. Med. Res. 82: 171, 1985.

4. **Saxena S.,** Andal A, and Saxena H.M.K. **Stereomicroscopic examination of kidney tissue for rapid identification of glomerulus**. Nephron: 45: 249, 1987.

5. **Saxena S.,** Mehrotra M.L.: **Host tissue response in soft tissue sarcomas.** Ind. J. Path. & Microbiol. 30:97, 1987.

6. **Saxena S.,** Andal A, and Saxena H.M.K. **Idiopathic nephrotic syndrome of childhood: Ultrastructural immunohistologic and Clinicocomorphologic correlation.** Ind. J. Path. & Microbiol. 31 (3) 195, 1988.

7. Andal A, **Saxena S**, Chellani H.K. and Sharma S. **Pure Mesangioproliferative Glomerulonephritis. A Clinicomorphologic analysis and its possible role in morphological transition of minimal change lesion to Focal glomerulosclerosis**. Nephron: 51(3): 314, 1989.

8. **Saxena S,** Davies D.J., Krisner R.L.G. **Thin basement membrane in minimally abnormal glomeruli.** J. Clin. Pathol. 43: 32, 1990.

9. **Saxena S.**, Andal A, Saxena R.K., Sharma S, Chandra M, Saxena H.M.K. **Immune status of children suffering from Minimal change nephrotic syndrome**. Ind. J. Path. & Microbiol. 35(3) 171, 1992.

10. **Saxena. S.**, Davies D.J., **Glomerular alterations in Idiopathic haematuria– Ultrastructural and Morphometric analysis**. Ind. J. Path. & Microbiol. 35(4), 326-332, 1992.

11. **Saxena. S.**, Andal. A., Sharma. S, Saxena H.M.K., Chandra M. **Immunomodulation by measles vaccine in children with Minimal change nephrotic syndrome**. Indian J. of Nephrology 2, 141-146, 1992

12. Verma. A.K., Tandon, R.., **Saxena. S.**, Pandy, J., Talib. V.H. **Aspiration Cytology of maxillary myxoma**. Diagnostic Cytopathology 9(2), 202-204, 1993.

13. **Saxena S**, Mital. A, Andal A,: **Pattern of interleukins in MCNS of childhood**. Nephron 65(1) 56-61, 1993

14. **Saxena S.**, Bhargawa R., Mohanty N.K., Talwar M: **Primary adenocarcinoma of the urinary bladder. A case report with review of literature** Ind J Pathol and Microbiol. 37(4), 453, 1994.

15. Saha T.K., Jolly B B., Mohanty N.K., **Saxena S.**, Dawson. L. **Multiple stones in Ectopic megaureter with Dysgenetic kidney – A case report**. Ind. J. Nephrol. 4(2). 61, 1994.

16. **Saxena S**. **Cytokine growth factors and childhood nephrotic syndrome**. Jr. of Nephrol. Vol. 8(6), 287, 1995.

17. Mohanty NK, Jolly BB, **Saxena S**, Dawson L. **Squamous cell carcinoma of peripheral urethrostomy**. Urol. Int. 1995, 55: 118-119.

18. Mohanty NK, Jolly BB, Talwar M, **Saxena S**, Dawson L. **Aspergillosis kidney. A case report**. Indian Jr. of Nephrol. 6(2), 56-58, 1996.

19. **Saxena S**, Jain A K, Pandey K K, Dewan A K. **Study on role of Steroid Hormone Receptors, Growth factor/receptors and Proto-oncogenes on behavior of Human Mammary Epithelial cancer cells in vitro.** Pathobiology 65(2), 75-82, 1997.

20. **Saxena S**, Mohanty N K, Talwar M, Jain A K. **Screening of Prostate Cancer in males with prostatism**. Ind. J. of Path & Microbiol. 40(4), 441-450, 1997.

21. Mohanty N K, Gulati P, **Saxena S**. **Role of interferon $\alpha$-2b in the prevention of superficial carcinoma of bladder recurrence**. Urol. Intern.59: 194-196, 1997.

22. Mohanty N.K., Jha AK, **Saxena S**, Kumar S., Arora RP. **Ten years experience with Adujuvant Intravesical Immunotherapy in management of superficial transitional cell carcinoma of Urinary bladder – A review**. Ind. J. of urology 2001, 17, 127.

23. **Saxena S**, Jain A.K., Bhatnagar D. **Study of events leading to cellular Senescence to Human Mammary epithelial cancer cells in vitro**. Indian. J. Cancer. 38: 103-116, 2001

24. **Saxena S**., Beena KR, Bansal A, Bhatnagar A. **Emperipolesis: Significance of an unusual phenomenon in common breast malignancy**. Acta Cytologica. 46: 883-886, 2002

25. **Saxena S**, Szabo C, Barjhoux H, Chopin S, Siniliniova O, Lenoir G, Goldgar D, Bhatnagar D**. BRCA 1 and BRCA 2 in Indian Breast Cancer Patients.** Human mutation. 20 (6): 473-74, 2002.

26. Chintamani, Sharma R D, Bardan R, Singhal V, **Saxena S,** Bansal A **Sweat gland Adenocarcinoma – a rare clinical dilemma**. World J Sur. Oncol. 1: 13, 2003

27. Bharat R, Burra U, Vidyadharan G, **Saxena S**. **Morphogical spectrum of cysticercus cellulose on cytology in case of malnutrished child**. J Cytol. 21 (2): 95-06, 2004

28. **Saxena S,** Bansal A, Mohil R S, Bhatnagar D. **Metaplastic carcinoma of the breast-A rare breast tumor.** Ind J Pathol and Microbiol. 47(2): 217-220, 2004

29. **C**hintamani, Shankar M, Singhal V, Singh J P, **Saxena S**. **Squamous cell carcinoma developing in the scar of fournier's gangrene-case report.** BMC Cancer. 4:16, 2004.

30. Bharat R, **Saxena S,** Burra U. **Fine needle aspiration cytology of Dermato fibrosarcoma protuberans**. J Cytol. 21(3), 2004

31. **C**hintamani, Singhal V, Singh J P, Bansal A, **Saxena S**, Lyall A. **Is drug induced cytotoxicity a good predictor of response to new adjuvantchemotherapy in breast cancer? A prospective clinical study.** BMC Cancer. Aug 13; 4(1): 48, 2004

32. Mukherji A, Madholia V, Malhotra S, Singh P, Rekhi B, **Saxena S**, Aggarwal Y, Bhowmik K.T. **Multiple Myeloma Of The Breast –An Unusual Case Of Multiple Myeloma Of The Breast With Pathological Fractures Of Humerus and Femur**. Jr. Of Clinical Radiotherapy and Oncology.4 (4):27-30, 2004.

33. Mohanti. N.K, **Saxena . S**, Goyal .N.K, Singh. U.P, Arora .R.P **Delayed Cystectomy for T1G3 TCC of Urinary Bladder Managed initially by TURBT & Intravesical Immunotherapy (BCG+Interferron)—Rationale & our results.** Indian Journal of Urology, 2004, vol. 20: 2.

34. **C**hintamani, Singhal V, Singh J P, Bansal A, **Saxena S**. **Half versus full vacuum suction drainage after modified radical mastectomy for breast cancer-A prospective randomized clinical trial** (ISRCT N24484328). BMC Cancer. 5:11, 2005**.**

35. Rekhi B, Bansal A, Bhatnagar D, Bhatnagar A, **Saxena S. Cytomorphological study of soft tissue neoplasms: role of fluorescent immunocytochemistry in diagnosis** Cytopathology, 16( 5) :219-26, 2005.

36. Rekhi B, **Saxena S,** Chintamani**. Gastric outlet obstruction and cutaneous metastasis in Adenocarcinoid Tumor of Stomach- Unusual presentations with cytologic and Ultrastructural findings.** Indian J Cancer. ,42(2):99-101,2005

37. Ghai R, Rekhi B, **Saxena S**, Kapoor S. **An unusual presentation of Primary Lympoma of the Thyroid in a young male patient—A case report.** I.J.P.M.,48(3): 385-387,2005

38. Murthy NS, Chaudhary K, Saxena **S**. **Trends in Incidence of Cervical Cancer – Indian Scenario.** Euro. J Can Prev. 2005 Dec; 14(6):513-8.

39. A. Agarwal, S. Verma, U. Burra, NS Murthy, NK Mohanty and S. Saxena Ca"**Flow Cytometric analysis of Th1 and Th2 cytokines in PBMCs as a parameter of immunological dysfunction in patients of Superficial Transitional cell carcinoma of bladder".** Cancer Immunology and Immunotherapy. 2006, 55(6), 734-743.

40. N.K.Mohanty, **Sunita Saxena**, Uday Pratap Singh, Neeraj K.Goyal, R.P.Arora "Lycopene **as a chemoprevetvive agent in the treatment of High Grade Prostate Intraepithelial Neoplasia.**" Urol Oncol. 2005 Nov-Dec; 23(6):383-5.

41. Chintamani, Jai Prakash Singh, Mahesh K Mittal, **Sunita Saxena**, Anju Bansal, Ashima Bhatia, Pranjal Kulshreshtha **The role of P--glycoprotein expression in predicting response to neoadjuvant chemotherapy in breast cancer-a prospective clinical study.** World Journal of Surgical Oncology 2005 **3**:61

42. **Sunita Saxena** ; Bharat Rekhi ; Anju Bansal ; Ashok Bagga ; Chintamani C and N.S.Murthy: **Clinico-morphological patterns of Breast Cancers Including family history in a Delhi hospital, India- A Cross-sectional study** World Journal of Surgical Oncology 2005, **3:**67

43. Burra UK, Singh A, **Saxena S**. **Eccrine porocarcinoma (malignant eccrine poroma): a case report.** Dermatol Online J. 2005 Aug 1; 11(2):17.

44. **Sunita Saxena**, Usha Agrawal, Abhilasha Agarwal*, Saurabh Verma, NS Murthy*, NK Mohanty **"Adjuvant Intravesical Therapy Based on In Vitro Cytotoxicity Assay In Management Of Superficial Transitional Cell Cancer of Urinary Bladder"** BJU International 2006 Vol. 98 (5), 1012

45. Bharat Rekhi, **Sunita Saxena "New Pot-pourri of Markers related to Invasive Breast Cancer"** JIMSA 2006 19(1), Jan-Mar.

46. Rekhi B, **Saxena S "Cytomorphology of Basal Cell Type Of Solid Ameloblastoma-A Case Report"** J Of Cytology 2006;23:83-85

47. **Sunita Saxena**, Anurupa Chakraborty, Mishi Kaushal Sanjeev Kotwal, Dinesh Bhatnagar, RS Mohil, Chintamani Chintamani, AK Aggarwal, Veena Sharma, PC Sharma, Gilbert Lenior and David Goldgar, Csilla Szabo **Contribution of germline BRCA1 and BRCA2 sequence alterations in to breast cancer in Northern India.** BMC Medical Genetics 2006, 7:75.

48. **Chatterjee I, Kapur S,** Mahanta J, Phukan RK, Barooah MN, Kataki AC, Purkayastha J**, Saxena S. Gene Expression profile in oesophageal cancer patients of Northeast region of India**. J Cancer Res. Ther, 2(1): S 20, 2006.

49. Singh A, **Saxena S**. **"Infiltrating Duct Carcinoma Of Breast, Metastatic to Axillary Lymph Nodes Harbouring Primary Tuberculous Lymphadinitis"** Pathology Oncology Research,12(3),2006

50. Singh Avninder, Amar Bhatnagar, Usha Agrawal and **Sunita Saxena**. **Isolated splenic metastasis from colorectal mucinous carcinoma: a case report** International Journal of Gastrointestinal Cancer 2006;37(2-3):98-101

51. N S Murthy, Usha K Burra, K Chaudhry, and **S Saxena**" **Trends in incidence of breast cancer-Indian Scenario".** Europian Jr. Of Cancer Care. doi:10.1111/j.1365-2354.2006.

52. Anurupa Chakraborty[1], N.S. Murthy[2], Chintamani[3], D Bhatnagar[3], R.S. Mohil[3], A. Bhatnagar[3], P.C. Sharma[4]**, Sunita Saxena[1] CYP 17 gene polymorphism and its association with high-risk North-Indian breast cancer patients"** Journal of Human Genetics 52(2):159-165,2007

53. IndranilChatterjee, SujalaKapur, JoydeepPurkayastha, Rupkumar Phukan, Amal Kataki, Jayanta Mahanta, **Sunita Saxena**. **Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis.** World Jr of Gastroentrology 2007; 13(9):1438-1444.

54. Chintamani, Binita P Jha, Anju Bansal, **Sunita Saxena** and Dinesh Bhatnagar **The expression of mismatched repair genes and their correlation with clinicopathological parameters and response to neo-adjuvant chemotherapy in breast cancer** International Seminars in Surgical Oncology.20074:5.

55. Chintamani, Pranjal Kulshreshtha, Nidhi Sugandhi, Anju Bansal, Dinesh Bhatnagar and **Sunita Saxena. Is an aggressive approach justified in the management of an aggressive cancer-the squamous cell carcinoma of thyroid?** International Seminars in Surgical Oncology 2007, 4:8 doi:10.1186/1477-7800-4-8

56. Chintamani, Rohan Khandelwal, Aliza Mittal, Sai Saijanani, Amita Tuteja, Anju Bansal, Dinesh Bhatnagar and **Sunita Saxena. Qualitative and quantitative dermatoglyphic traits in patients with breast cancer: a prospective clinical study**
BMC Cancer 2007, **7:**44

57. Chintamani, Vinay Singhal, Anju Bansal, Dinesh Bhatnagar and **Sunita Saxena. Isolated colostomy site recurrence in rectal cancer -two cases with review of literature** World Journal of Surgical Oncology 2007 **5**:52

58. Tyagi I, Agarwal U, Amitabh V, Jain A K, **Saxena S. Thickness of Glomerular and Tubular basement membranes in preclinical and clinical stages of Diabetic Nephropathy.** Indian Jr of Nephrology 2008; 18(2):60-65.

59. Singh, A., Kapur, S. and Saxena, S. **Cytokeratins and gastrointestinal cancer: A brief review.** Gastroenterol Today 12 (2008) 115.

60. Sharma, M., Chintamani, Saxena, S. and Agarwal, U. **Squamous cell carcinoma arising in unilateral Warthin's tumor of parotid gland.** J Oral Maxillo Facial Pathol 12 (2008).

61. Anurupa Chakraborty[1], A.K Mishra[1], Abha Soni[1], Thodum Regina[1], D Bhatnagar,[2] A Bhatnagar[2], Chintamani,[2] Sunita Saxena[1] **VDR gene polymorphism(s) and breast cancer risk in North Indian Population.** Cancer Detection and Prevention **32** (2009) pp. 386-394

62. Chattopadhyay I, Phukan R, Vasudevan M**,** Singh A, Purkayastha J, Hewitt S, Kataki A, Mahanta J, Kapur S, Saxena S; **"Molecular profiling to identify molecular mechanism in esophageal cancer with familial clustering"** Oncology Reports 21:1135-1146,2009

63. Chintamani, T. Aeron, M. Mittal, D. Bhatnagar, U. Agarwal, S. Saxena **Are the structures preserved in functional neck dissections truly preserved functionally? – A prospective study of patients with head and neck cancer at a tertiary cancer care center** Oral Oncology Supplement, Volume 3, Issue 1, July 2009, Page 175

64. Murthy, N.S., Chaudhry, K., Nadayil, D., Agarwal, U.K. and Saxena, S. **Changing trends in incidence of breast cancer: Indian Scenario.** Indian J Cancer 46 (2009) 73.

65. Avninder Singh, ; Sujala Kapur,; Indranil Chattopadhyay,; Joydeep Purkayastha,; Jagannath Sharma,; Ashwani Mishra,; Stephen M. Hewitt; Sunita Saxena,. **Cytokeratin immunoexpression in esophageal squamous cell carcinoma of high-risk populationin Northeast India.** Applied Immunohistochemistry & Molecular Morphology , 17(5):419-424,Oct.2009

66. Mishi Kaushal, Indranil Chattopadhyay, Rupkumar Phukan, Joydeep Purkayastha, Jagadish Mahanta, Sujala Kapur, Sunita Saxena. **Contribution of germline BRCA2 sequence alterations to risk of familial esophageal cancer in high-risk area of India.** Disease of the Esophagus. DOI:10.1111/j.1442-2050.2009.00975.x (published online),2010:23(1) 71-5.

67. Agrawal A, Agrawal U, Verma S, Mohanty N.K and SaxenaS. **Serum Th1 and Th2 cytokine balance in patients of superficial transitional cell carcinoma of bladder pre and post intravesical combination immunotherapy.** Immunopharmacology and Immunotoxicology.2010:32(2)348-56

68. Regina Devi T, Yadav DS, Mishra AK, Kaushal M, Ihsan R, Chattopadhayay I, Chauhan P, Sarma J, Zomawia E, Verma Y, Nandkumar A, Mahanta J, Phukan R, Kapur S, Saxena S: **Distribution of Glutathione S- transferase T1 and M1 genes polymorphisms in North East Indians. A potential report.** Genetic Testing and Molecular Biomarkers. 14(2);163-169,2010

69. Chintamani Chintamani, Rohan Khandelwal, Megha Tandon, Yashwant K, Pranjal Kulshreshtha, Tushar Aeron, Dinesh Bhatnagar, Anju Bansal, Sunita Saxena **Carcinoma developing in a fibroadenoma in a woman with a family history of breast cancer: a case report and review of literature** Cases Journal 2009, 2:9348

70. Chattopadhyay I, Singh A, Phukan R, Purkayastha J, Kataki A, Mahanta J, Saxena S, Kapur S .**Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma exposed to tobacco and betel quid from high-risk area in India** Mutation Research - Genetic Toxicology and Environmental Mutagenesis 696 (2010), pp. 130-138

71. Rakshan I, Chattopadhyay I, Phukan R, Mishra A K, Purkayastha J, Sharma J, Zomawia E, Verma Y, Nandkumar A, Mahanta J, Saxena S, Kapur S. **Role of EPHX1 gene polymorphisms in esophageal cancer of high-risk area in India.** Jr. of Gastroenterology and Hepatology 2010 Aug; 25(8):1456-62.

72. Usha Agrawal, Ashwani K Mishra, Payal Salgia, Saurabh Verma, Nayan K Mohanty, **Sunita Saxena. Role of Tumor Suppressor and Angiogenesis Markers in Prediction of Recurrence of Non Muscle Invasive Bladder Cancer.** Pathology and Oncology Research 17(1); 91-101,2011,

73. Dhirendra Singh. Yadav, Thoudam Regina Devi, Rakhshan Ihsan, AK Mishra, Mishi Kaushal, Indranil Chattopadhyay, Pradeep Singh Chauhan, Jagannath Sharma, Eric Zomawia, Yogesh Verma, A. Nandkumar, Jagadish Mahanta, Rupkumar Phukan, Sunita Saxena., Sujala Kapur. **Polymorphisms of Glutathione-S-transferase (GST) genes and the risk of aerodigestive cancers in Northeast Indian population.** Genetic Testing and Molecular Biomarker 14(5);1-9,2010

74. Pradeep Singh Chauhan, Rakhshan Ihsan, Dhirendra Singh Yadav, Ashwani Kumar Mishra, Bharat Bhushan , Abha Soni, Mishi Kaushal, Thoudam Regina Devi, Sumita Saluja, Dipendra Kumar Gupta, Vishakha Mittal, Sunita Saxena,Sujala Kapur. **Association of GST, EPHX and p53 codon 72 gene polymorphism with adult acute myeloid leukaemia.** DNA and Cell Biology 20 Jan; 30(): 39-46.

75. Chintamani**,** Pranjal Kulshreshtha, Anurupa Chakraborty, L C Singh, Ashwani K Mishra, Dinesh Bhatnagar, Sunita Saxena**. Androgen receptor status predicts response to chemotherapy, not risk of breast cancer in Indian women** World Journal of Surgical Oncology 2010, **8**:64

76. Kaushal M, Mishra AK,Raju BS, Ihsan R, Chakraborty A, Sharma J, Zomawia E, Verma Y, Kataki A, Kapur S. Saxena S: **Betel quid chewing as an environmental risk factor for breast cancer.** Mutation Research - Genetic Toxicology and Environmental Mutagenesis 703(2010), 143-148

77. Rakhshan Ihsan, Thoudam Regina Devi, Dhirendra Singh Yadav, Ashwani Kumar Mishra, Jagannath Sharma, Eric Zomawia, Yogesh Verma, Rupkumar Phukan, Jagadish Mahanta, Amal Chandra Kataki, Sujala Kapur, Sunita Saxena. **Investigation on the role of p53 codon 72 polymorphism and interactions with tobacco, betel quid and alcohol in susceptibility to cancers in a high risk population from north east India** DNA and Cell Biology 2011 March; 30(3): 163-171

78. .Chintamani, Jp Singh, Megha Tandon, Rohan Khandelwal, Tushar Aeron, Sidharth Jain, Nikhil Narayan, Rahul Bamal, Yashwant Kumar, S Srinivas, Sunita Saxena **Vulval elephantiasis as a result of tubercular lymphadenitis: two case reports and a review of the literature.** Journal of Medical Case Reports 2010 **4**:369

79. Abha Soni, Anju Bansal, L C Singh, Ashwani Kumar Mishra, Thoudam Regina, N K Mohanty, Sunita Saxena. **Gene expression profile and mutational analysis of DNA mismatch repair genes in carcinoma prostate in Indian population.** OMICS: 2011 Feb 24 (Epub ahead of print)

80. Simmy Soni, Gayatri Rath, Chandra Prakash Prasad, Sudha Salhan, Arun Kumar Jain, Sunita Saxena. **Fas-Fas System in Molar Pregnancy.** American Journal of Reproductive Immunology. (In Press)

81. Simmy Soni, Gayatri Rath, Chandra Prakash Prasad, Sudha Salhan, Sunita Saxena, Arun Kumar Jain. **Apoptosis and Bcl-2 protein expression in Human Placenta over the course of Normal Pregnancy**. Anatomia. Hisologia Embryologia.39(2010): 426-431

82. Chintamani**,** Megha Tandon**,** Ashwini Mishra**,** Usha Agarwal and **Sunita Saxena. Sentinel lymph node biopsy using dye alone method is reliable and accurate even after neo-adjuvant chemotherapy in locally advanced breast cancer- a prospective study.** World Journal of Surgical Oncology 2011, 9:19

83. Agarwal S, Agrawal U, Mohanty NK, Saxena S. **Multilocular Cystic Renal Cell Carcinoma: A case report of rare entity.** Arch Pathol Lab Med—Vol135, March 2011

84. L C Singh, Anurupa Chakraborty, Ashwani K Mishra, Thoudam Regina Devi, Nidhi Sugandhi[2], Chintamani, Dinesh Bhatnagar, Sujala Kapur, **Sunita Saxena[*]"Study on predictive role of AR and EGFR family genes with response to Neo-adjuvant Chemotherapy in Locally Advanced Breast Cancer in Indian women"** Medical Oncology 2012, 29(2) : 539-546

85. Singh A, Mishra A K, Ylaya K, Hewitt S M, Sharma K C, Saxena S. **Wilms Tumor-1, Claudin-1 and Ezrin are useful Immunohistochemical markers that helps to distinguish Schwannoma from Fibroblastic Meningiomas.** Pathol. Oncol. Res. DOI 10.1007/s12253-011-9456-x

86. Anju Bansal, Abha Soni, Punita Rao, LC Singh, Ashwini Mishra, N K Mohanty, Sunita Saxena. **Implication of DNA repairs genes in prostate carcinogenesis in Indian men.** Indian J Med Res 136, October 2012, pp 622-632

87. Ihsan R, Chauhan PS, Mishra AK, Yadav DS, Kaushal M, Sharma JD, Zomawia E, Verma Y, Kapur S, Saxena S. **Multiple Analytic Approaches reveal distinct Gene- Environment interactions in Smokers in Lung Cancers. PLoS One**, 6(12): e29431, 2011.

88. Bansal A, Bhatnagar A, Saxena S. **Metastasizing granular cell ameloblastoma.** J Oral Maxillofac Pathol 2012;16:122-4.

89. Abha Sony, Anju Bansal, Aswini Kumar Mishra, Jyotsna Batra, L.C. Singh, Anurupa Chakraborty, Dhirendra Singh Yadav, N. K. Mohanty, Sunita Saxena **"Association of Androgen Receptor, Prostate Specific Antigen and CYP19 gene polymorphisms to Prostate Carcinoma and Benign Prostatic Hyperplasia in North Indian population"** Genetic Testing and Molecular Biomarker ( In Press)

90. Mishra AK, Agrawal U, Negi S, Bansal A, Bhatnagar A, Bhatnagar D, Chintamani, Mohil R, Saxena S. **Study on expression of AR in Breast Cancer and its correlation with other steroid receptors and growth factors**. Indian J Med Res 135, June 2012, pp 843-852

91. Chauhan PS, Ihsan R, MishraAK,Yadav DS, Saluja S, Mittal V, Saxena S, and Kapur S: **High Order Interactions of Xenobiotic Metabolizing Genes and P53 Codon 72 Polymorphisms in Acute Leukemia.** Environmental and Molecular Mutagenesis (2012)

92. Mishi Kaushal, Ashwani. K. Mishra, Jaganath Sharma, Eric Zomawia , Amal Kataki, Sujala Kapur, Sunita Saxena **Genomic alterations in breast cancer patients in betel quid** 1 **and non betel quid chewers. PLoS One (Accepted)**

93. Rakhshan Ihsan, Pradeep Singh Chauhan, Ashwani Kumar Mishra, L.C. Singh, Jagannath Dev Sharma, Eric Zomawia, Yogesh Verma, Sujala Kapur, Sunita Saxena **Investigation on copy number polymorphism of GSTM1 and GSTT1 in susceptibility to lung cancer in a high-risk population from North-East India.** Indian Journal of Medical Research. (Accepted)

**Books-**

1. **Mishra AK, Chakraborty A, Saxena S: Significance of Vitamin D receptor polymorphisms in Breast Cancer—Multinomial Logistic Regression Analysis in Vitamin D,: Nutrition, Side effects and supplements. Editor: Stephanie R. Malone**

2. **Sunita Saxena: Molecular and Genetic Aspects of Lung Cancer. Chapter 13 pp 141-148 In Pathological and Occupational Lung Health. Editors: V.K.Vijyan, H.K.Tazelaar and Ritu Kulshrestha**