Gene Expression Profiling and Genetic Variations in Oral Cancer Associated with Tobacco Consumption

SYNOPSIS OF THE THESIS

Submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

By

DHIRENDRA SINGH YADAV

Under the Supervision of

Dr. Sujala Kapur

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN) INDIA 2012
**Synopsis**

**Background:**

Oral cancer is the sixth most common cancer worldwide, accounting for 4% of cancers in men and 2% of cancers in women (Napier and Speight, 2008). Annually, over 300,000 new cases of oral cancer are diagnosed all over the world where the majority of them are diagnosed in the advanced stages III or IV. Such data make the oral cancer an important public health problem which is responsible for 3% to 10% of cancer mortality worldwide. There is up to a 20-fold geographical variation in the incidence rates of oral cancer (Warnakulasuriya, 2009). The areas with high incidence rates for oral cancer are found in the South and South East Asia. It has been estimated that of the 390,000 oral and oro-pharyngeal cancers that occur annually worldwide, 58% occur in south and south-east Asia (Nair et al., 2004). In India oral cancer is the most common cancer among men and ranks third among women (Soya et al., 2007), with age-standardized incidence rates of 12.8 and 7.5 per 100,000 population respectively (Nair et al., 2004). Northeastern states in India have reported a very high prevalence of aero-digestive tract cancers including oral cancer when compared with other regions of India (Bhattacharjee et al., 2006; Phukan et al., 2004). Prevalence of oral cancer is highest in the Kamrup district of Assam (ICMR-Report, 2006). In this region there is widespread chewing habit of tobacco with peculiarly fermented betel nut (Phukan et al., 2001). This study is based on oral cancer samples collected from three regional collaborating centers (Guwahati, Sikkim and Aizawl) of Northeast (NE) India. These are the Population Based Cancer Registry (PBCR) centers located in NE India which were in collaboration with our Institute for five years (2005-2010), and provided samples for this study. Knowledge of cancer genetics is rapidly improving our understanding of cancer biology, helping to identify at-risk individuals, furthering the ability to characterize malignancies, establishing treatment tailored to the molecular fingerprint of the disease, and leading to the development of new therapeutic modalities.

The etiology of oral cancer is multi-factorial with genetic, environmental, medical, and lifestyle factors interacting to produce malignant stage. The lifestyle behaviors associated to oral cancer with convincing evidence are tobacco use, betel quid chewing, alcohol drinking, low fruit and vegetable consumption. These factors may act separately or
synergistically. Worldwide, 20-30% of oral cancer cases are attributed to tobacco consumption (Balaram et al., 2002; Hashibe et al., 2007; Rahman et al., 2005). 50% of men and almost 90% of women cases are attributed to frequent betel quid chewing with or without tobacco consumption in areas where chewing prevalence is high (Balaram et al., 2002). 7–19% of total cases of the world are attributed to heavy alcohol drinking (Hashibe et al., 2007) and 10–15% cases are attributed to micronutrient deficiency (WHO-FAO-Report, 2007; WHO-Report, 2002). Worldwide, smoking accounts for 42% of deaths from cancers of the oral cavity (including the pharynx) and heavy alcohol consumption for 16% of the deaths; the corresponding percentages in high income countries are about 70% and 30%, respectively (Danaei et al., 2005; Jemal et al., 2011). Smokeless tobacco products and betel quid with or without tobacco are the major risk factors for oral cavity cancer in India, and other neighboring countries (Sri Lanka, Pakistan, Bangladesh and Taiwan) (Jayalekshmi et al., 2009; Wen et al., 2010). The rise in the incidence rate of oral cancer in India may have been in part due to the increased consumption of tobacco, betel quid and alcohol. However not all tobacco and alcohol consumers develop premalignant or frankly malignant diseases of the oral cavity. The quantitative absorption, distribution, metabolism, and excretion of carcinogenic tobacco constituents depend on the activity and efficiency of metabolic and enzymatic detoxification pathways. The enzymatic detoxification process is mainly divided into two phases. Phase I involves activation of toxic compounds mainly by oxidation into more reactive intermediate products that are neutralized and conjugated by phase II family of enzymes. Cytochrome P450 (CYP) family of enzymes are the major phase I enzymes which usually converts tobacco constituents into more active intermediate compounds which are further detoxified by phase II family of enzymes such as glutathione-S-transferase (GST), NAD(P)H dehydrogenase quinone 1 (NQO1), and N-acetylationtransferases (NAT). The resultant water-soluble and less-toxic product can easily be eliminated from the body. The role of genetic factors including single nucleotide polymorphism (SNP) of genes associated with activation and detoxification of toxic compounds is conflicting (Buch et al., 2008; Gattas et al., 2006; Hatagima et al., 2008; Zhuo et al., 2009). As the Northeast Indian population is exposed to high levels of carcinogenic compounds, it was hypothesized that less efficient detoxification mechanism due to polymorphic variants of genes encoding detoxification enzymes may explain high incidence of oral cancer in this region. Moreover p53 mutations have been associated with reduced repair and enhanced cytotoxicity in cell damaged by benzo(α)pyrene diol epoxide-DNA adducts. In the current study, the association
of oral cancer in a high risk region of Northeast India was investigated for total of eight polymorphisms present in seven genes CYP1A1 (Msp1 and Nco1), GSTT1, GSTM1, GSTP1, NAT2, NQO1 and codon 72 polymorphism of p53 gene.

It is evident that oral cancer is a multi-factorial disease influenced by both aetiological factors and ethnicity; as a result molecular profiles of oral cancer vary throughout the world (Ambatipudi et al., 2011; Paterson et al., 1996; Shah and Singh, 2006). High throughput methods such as cDNA and oligonucleotide microarrays are increasingly being used to systematically compare molecular features of individual cancers to key clinical parameters. Previous studies have documented the importance of genetic alterations affecting known oncogene and tumour suppressor genes in the development of oral cancer (Bettendorf et al., 2004; Tsantoulis et al., 2007). For detection of genetic alterations, sequencing has historically been considered as the golden standard because it reveals the exact location and the type of mutation. DNA sequencing represents a single method to forecast a broad range of biological events. However this method was slow and tedious. Methodological development in the form of next generation sequencing can provide better insights for possible therapeutic options as it can reveal broad range of genetic aberrations including mutations at nucleotide level (such as SNPs, insertions or deletions) involved in carcinogenesis. This method can also provide possible functional/structural changes in resulted protein using appropriate softwares in cases of known aberrations and provide a basis for further analysis in cases of finding novel genetic alterations.

Genome wide screening can be either quantitative as in the case of gene expression profiling by microarray or qualitative as in the case of mutation detection by next generation sequencing. Altered gene expression profiling may be attributed to several types of genetic alterations such as base changes, insertions, deletions, chromosomal translocations, inversions, loss of heterozygosity and copy number variations. To understand genetic alterations in large number of genes, precise, fast and cost-effective methods have been developed to analyze several types of genetic alterations in a single experiment. In recent years there has been a revolution in sequencing methods in the form of next generation sequencing (NGS) technologies capable of producing millions of DNA sequence reads in a single run. This is rapidly changing the landscape of cancer research with the potential to accelerated analysis of genomes, transcriptomes and interactomes which become inexpensive, routine and widespread (Shendure and Ji, 2008). NGS has enabled whole-
genome analysis with essentially unlimited resolution (Stankiewicz and Lupski, 2010). Using next generation sequencing clinically relevant molecular signatures can be discovered which may be of great significance in understanding the underlying biology of cancer, providing the ability to answer questions with unimaginable speed which may lead us towards identification of molecular targets in carcinogenesis and subsequent development of innovative methods for diagnosis, prognosis and therapy of cancer.

The discovery of mutations (hallmarks of cancer) that are involved in carcinogenesis is a crucial step in cancer research which is tremendously facilitated by NGS technology useful for both targeted and genome-wide screening. For the present study targeted re-sequencing of 169 functionally relevant and potentially important genes was done to analyze mutational changes using Illumina’s high throughput solexa sequencing technology. Our aim was to identify specific molecular signatures involved in oral carcinogenesis as well as to identify potential biomarkers for oral cancer predisposition, progression and therapeutic manipulation.

As the deregulated expression of genes lies at the origin of tumors, its measurement using microarray technology can be very helpful to model or predict the clinical behavior of malignancies. Many studies have shown that cancer diagnosis based on microarray data can effectively integrate the fundamental processes underlying carcinogenesis into the clinical decision-making process (Wong and Wang, 2008). Microarray technology has made it possible to examine the expression of many genes over multiple developmental stages or different experimental conditions (Guo et al., 2007). In the current study differential gene expression analysis was done on tumour tissue of oral cancer patients and matched normal tissue distant from the tumour site. Significantly deregulated genes were selected from microarray data and validated by real-time RT PCR. This is the first study to provide gene expression profiling and genetic variations of oral cancer associated with betel quid and tobacco consumption in a high-risk region of India.

**Gap in Existing Research:**

In Northeast India, cancer incidences and causative factors responsible for them were not well documented till few years ago. First report of National Cancer Registry Programme (NCRP) of ICMR for Northeastern region of country which emphasized on cancer incidence and patterns of cancer in six population based cancer registries (PBCRs) of the Northeastern
region (for the two year period 1 Jan 2003 to 31 December 2004) revealed very high incidence of tobacco associated cancers in this region as compared to other parts of the country. In the older established registries (other regions of India) age adjusted incidence rate (AAR) for all anatomical sites has been around 100 per 100,000. The results of this report by NCRP for Northeastern states were notable, in the sense that incidence rates of well over 100 per 100,000 persons have been recorded in five of the eight registry areas identified for describing the incidence and patterns of cancer. Highest AAR for cancer incidence was reported from Aizawl district (AAR: 277.2 in males and 231.5 in females) of Mizoram state followed by Kamrup urban district (AAR: 177.2 in males and 154.1 in females) of Assam state of Northeast India (ICMR-Report, 2006). This region has reported a very high prevalence of aero-digestive tract cancers compared with other regions of India (Bhattacharjee et al., 2006; Phukan et al., 2004). Prevalence of oral cancer is highest in the Kamrup district of Assam (ICMR-Report, 2006). However there is lack of genetic studies on patients with oral cancer from high risk Northeast region of India. The present study is based on the samples collected from three centers [Guwahati (Assam); Gangtok (Sikkim) and Aizawl (Mizoram)] of Northeast India.

High frequency of oral cancer in Indian subcontinent is mainly attributed to tobacco chewing which attributes for more than 66% of the total oral cancer cases in India (IARC-Report, 2010). The surveys carried out by National Family Health Survey (NFHS) and Global Youth Tobacco Survey (GYTS) conducted in 18 states in India have reported an alarmingly high prevalence of chewable tobacco products in younger Indian population (average prevalence of tobacco chewing was 29% compared to 13% for smoking). The highest prevalence of tobacco chewing within India was observed in Northeastern states ranging from 47-63% (Kuruvilla, 2008). This is one of the major contrasting features of Indian population as compared with western countries where tobacco smoking is more prevalent compared to chewing (Kuruvilla, 2008).

Despite recent advancement in the treatment, imaging and diagnosis of oral carcinoma, a 5-years survival and mortality rate for this cancer is still at 50% (Tanaka and Ishigamori, 2011). The survival of oral cancer patients remains very low, mainly due to the fact that it is often revealed when it has metastasized to another location, most likely the lymph nodes of the neck. About half of the patients affected will die within five years of diagnosis, while surviving patients may be left with severe aesthetic and functional
compromise. Given the poor prognosis associated with oral cancer and lack of genetic studies available from high risk Northeast region of India, there was need to elucidate the molecular determinants and critical signal pathways underlying oral carcinogenesis which may further lead to identification of novel diagnostic and therapeutic targets.

**Aims and Objectives:**

The present study was undertaken with the following objectives:

**Objective 1:** *To identify the role of polymorphism of major genes coding for xenobiotic metabolizing phase I and phase II enzymes in oral cancer susceptibility*

The role of polymorphisms of genes responsible for detoxification of xenobiotics and *p53* codon 72 were investigated by PCR-RFLP methods in 235 oral cancer cases and 289 healthy controls from high incidence region of oral cancer in Northeast India. To identify the role of polymorphisms in genes coding for phase I enzymes, *CYP1A1*<sup>2A</sup> and *CYP1A1*<sup>2C</sup> genes were selected. Polymorphisms in genes *GSTT1, GSTM1, GSTP1, NQO1* and *NAT2* involved in phase II pathway of detoxification were studied to identify their role in oral carcinogenesis.

**Objective 2:** *To identify genetic variations (SNPs, Insertions and Deletions) in tumor samples of patients with oral cancer*

Targeted re-sequencing of 169 functionally relevant and potentially important genes was performed in 25 samples of oral cancer from Assam (Guwahati) region of Northeast India using Illumina-Solexa’s Next Generation Sequencing platform and sequencing by synthesis approach.

**Objective 3:** *To study differential gene expression profiles of oral squamous cell carcinoma using microarray technology*

Differential gene expression profiling was performed by cDNA microarray to identify differentially expressed genes in 5 pair of samples of oral cancer and corresponding matched controls.
Synopsis

**Objective 4:** *To validate the microarray data for specific genes of interest using Real Time PCR*

Validation of gene expression data of microarray was performed in specific genes of interest [four up-regulated (*PDPN, IL8, COPS5* and *INHBA*) and four down regulated (*KRT4, DOCK8, SPRR3* and *MAL*)] in 27 samples of oral cancer with respect to a pooled control by quantitative real time RT PCR using TaqMan probe based assay.

**Results and Discussion:**

**Objective 1:** *Role of SNPs in Xenobiotic Metabolizing Genes (Encoding Phase I and Phase II Detoxification Enzymes) and p53 Codon 72 Polymorphism in Oral Cancer*

Eight polymorphisms in seven genes [*CYP1A1* (*MspI* and *NcoI*), *GSTT1*, *GSTM1*, *GSTP1*, *NAT2*, *NQO1* and Codon 72 of *P53*] were analyzed using PCR-RFLP and correlated with risk factors of oral cancer. Tobacco chewing as well as smoking were found to impart a significant risk for oral cancer (AOR=2.78, 95% CI=1.71-4.51, and AOR=1.57, 95% CI=1.05-2.35 respectively) with tobacco chewers at higher risk compared to smokers. Frequency of betel quid chewers and alcohol consumers were also higher in cases (80% and 37% respectively) compared with controls (66% and 30% respectively) but when adjusted with other factors the difference was statistically insignificant. *GSTT1* and *GSTM1* null genotypes and the variant genotypes of *CYP1A1* *2A*, *CYP1A1* *2C*, and *p53 codon* 72 were not found to be associated with oral cancer risk. Homozygous variant genotypes of *NAT2* (AA) were significantly higher in cases (14%) compared to controls (9%) [OR=1.83, 95% CI=1.01-3.31, P=0.04]. However when adjusted with other risk factors this risk became statistically insignificant. Frequency distribution of *NQO1* genotypes Pro/Pro, Pro/Ser and Ser/Ser was 45%, 35% and 20% in cases and 60%, 27% and 30% in controls respectively. This difference was statistically significantly different for heterozygous (AOR=1.64, 95% CI=1.04-2.58) as well as for homozygous (AOR=1.81, 95% CI=0.98-3.32) variant genotypes. When data was analyzed in different geographic regions of NE India, the *GSTT1* null genotype and homozygous variant genotypes of *GSTP1* were found to impart significant risk for oral cancer (AOR= 2.58, 95% CI 1.01–6.61, p= 0.05, and AOR=3.14, 95% CI 0.94-10.49, p=0.06 respectively) in samples obtained from the Assam region. Further, variant genotypes of *CYP1A1* *2A* were found to impart significant risk for heterozygous (AOR = 4.55, 95% CI...
Synopsis

0.88-23.36, p=0.07) as well as for homozygous genotypes (AOR= 6.38, 95% CI 1.10-40.83, p=0.05) for oral cancer in Sikkim population of NE India. Gene-environment interaction analysis revealed that the variant genotypes of \textit{NQO1} did not interact statistically with tobacco consumption habits. It is possible that although variant genotypes of \textit{NQO1} may play an important role in the genetic susceptibility to oral cancer, its pathway is unrelated to the detoxification mechanism of tobacco constituents.

**Objective 2: Detection of Genetic Instability/ Genomic Alterations using Targeted Next Generation Sequencing in Oral Squamous Cell Carcinoma**

Several types of genetic alterations such as base changes, insertions, deletions, chromosomal translocations, inversions, loss of heterozygosity and copy number variations may be involved in oral carcinogenesis. In this study targeted re-sequencing of 169 functionally relevant and potentially important genes showed 96 SNPs (50 novel and 46 known SNPs) and 46 InDels (29 novel and 17 known InDels). Most of the known SNPs with high read depths (>50) were located in tumor suppressor genes such as \textit{RB1}, \textit{FHIT}, \textit{FAT1}, \textit{FAT2} and \textit{VHL}. SNPs detected in \textit{RB1}, \textit{FHIT} and \textit{FAT1} were located in the intronic regions of the gene while those in \textit{ATM}, \textit{VHL}, \textit{IL12B}, and \textit{MET} were located in 3'UTR.

Novel SNPs with significant read depths (>25), and with non-synonymous type variations which may lead to structural and functional changes in resulting protein were observed in \textit{FAT1}, \textit{FAT2}, \textit{TP53}, \textit{NOTCH2}, \textit{CDH3}, \textit{ATM} and \textit{MET}. Synonymous type variations were observed in \textit{APC} and \textit{IL12B} genes and those present in non-coding regions were observed in or near to \textit{EGFR}, \textit{STAT5B}, \textit{CDK5} and \textit{MYCL1} genes.

The known deletions observed in our study were present in non-coding regions of the gene, 4 of them were present in 3'UTR (\textit{TSC1}, \textit{FAT1}, \textit{MAP2K6}, and \textit{ERBB4}), two at 5'UTR (\textit{BMP4}, and \textit{SLC22A18}) and one in intronic region of \textit{BRCA1}. UTR regions of a gene have significant role in regulation of gene activity, thus its alteration may have an adverse effect on gene activity. The known deletion with highest read depth (272) was present in UTR region of \textit{TSC1} gene (rs34947162; rs115091888). \textit{TSC1} plays a central role in regulating cell survival and proliferation signaling pathways. A total of 11 novel deletions have been observed in oral cancer cases in our study. Novel deletion with the highest read depth (58) was present in \textit{MSH6} gene (NM_000179 at position 48033455). Other important genes with novel deletions include \textit{IGF1R}, \textit{BRCA2}, \textit{TSC2} and \textit{PAK3}.
Of the known insertions observed in our study, 4 were present in UTR regulatory regions of *APC, SMAD2, RHOB* and *NBL1* genes while the remaining 6 were located in intronic regions of *ADH6, PDGFRA, BRIP1, FAT2, DLG2 and KLK8*. The insertion with highest read depth (102) was that of base A at position 112180228 in *APC* gene (rs11432316; rs79379053). This gene encodes a tumor suppressor protein that acts as an antagonist of the Wnt signaling pathway.

Insertions and deletions (InDels) of bases are among highly damaging mutations. The affected genes in our study may be responsible for oral carcinogenesis as they may result in abrupt changes in expression of genes or in the structural or functional activities of the encoded protein.

**Objective 3: Differential Gene Expression Profiling of Oral Squamous Cell Carcinoma and Adjacent Noncancerous Tissues with cDNA Microarray in North East Indian Population**

Differential gene expression profiling was done by cDNA microarray to identify differentially expressed genes in oral cancer. Six hundred and thirty four differentially expressed (247 up-regulated and 387 down-regulated) genes were identified. Most significantly up-regulated pathways were related to ribosomal activity (*RPL38, RPSA, RPL6, RPS3, RPS20, RPS6 and RPS7*), Neuroactive ligand-receptor interaction (*GRM8, GRM4, NTSRI and P2RX7*), ECM-receptor interaction (*LAMC2, ITGB1, GP6 FN1 and COL2A1*) and Aminoacyl-tRNA biosynthesis (*LARS2, AARS and WARS*). Functional annotation clustering of up-regulated genes using web-based *DAVID* analysis also showed most significantly enriched cluster to be associated with ribosomal activity, ribosome biogenesis, and translation (involving genes such as *COPS5, IDO1, KYNU* and *RPS7*). Second cluster of up-regulated genes was related to inflammatory response, defense response and response to wounding (involving genes such as *IL8, INHBA* and *PDPN*).

Most significantly downregulated pathways (when analyzed by *Genowiz*) were related to cell junctions (involving 12 genes *KRT34, DSG1, KRT2, KRT15, KRT36, VIM, KRT4, KRT10, KRT78, KRT13, ACTB and KRT33A*), and Valine, leucine and isoleucine degradation (*ALDH3A1, BCAT2, HIBCH, BCKDHB, ACAT2, ALDH2* and *ALDH3A2*). *DAVID* analysis revealed clusters with highest enrichment score of down-regulated genes to be associated with structural constituents of cytoskeleton (involving genes such as *CRYAB,*...
Synopsis

KRT4, TPM1 and KRTAP5-9), epidermal cell differentiation and Keratinisation (involving genes such as ALDH3A2, JUN, KRT10, KRT13, KRT15, KRT2, KRT34, TP63, KRT4, KRTAP5-9 and SPRR3).

Objective 4: Validation of Gene Expression Profiling of OSCC by Quantitative real time RT-PCR

Eight significantly deregulated genes [four up-regulated (PDPN, IL8, COPS5 and INHBA) and four down regulated (KRT4, DOCK8, SPRR3 and MAL)] were selected from microarray data and validated by real-time RT PCR. Our study revealed several genes such as DOCK8 and SPRR3 which were reported for the first time to be associated with oral carcinogenesis. Recent reports have also suggested significance of KRT4 and MAL to work as possible prognostic and therapeutic markers for oral carcinogenesis and DOCK8 and SPRR3 may be further investigated for their association with oral carcinogenesis.

To summarize, this is the first study to provide gene expression profiling and genetic variations of oral cancer associated with betel quid and tobacco consumption habits in a high-risk region of Northeast India. Polymorphic study revealed that although variant genotypes of NQO1 may play an important role in the genetic susceptibility to oral cancer, its pathway may be unrelated to the detoxification mechanism of tobacco constituents. On gene expression profiling analysis, NQO1 was found to be significantly up-regulated in our study. Thus, it appears that it may be the variant form of NQO1 and not its expression level, which may be responsible for the higher risk of oral cancer in this region. Genes such as FAT1, TSC, GAS7 and APC showed high level of genomic instability (including known and novel variations) in OSCC. Furthermore GAS7 which is functionally involved in chemokine-induced migration in the immune system was also found to be significantly down-regulated in our gene expression profiling study. Thus these genes may work as useful prognostic and therapeutic targets in OSCC.

Future Scope of Work:

The discovery of genetic variations involved in carcinogenesis is a crucial step in cancer research which is tremendously facilitated by microarray and next generation sequencing technology, useful for both targeted and genome-wide screening. This is the first study on oral cancer from high-risk region of Northeast India which provided genetic
variations and gene expression profiling of oral cancer associated with betel quid and tobacco consumption habits. Our study revealed a large number of mutational changes including known SNPs, novel SNPs, known insertions, novel insertions, known deletions and novel deletions. These genetic variations provide a rich source of information which may be further investigated for their role to work as possible diagnostic, prognostic and therapeutic markers for oral carcinogenesis. Furthermore several differentially expressed genes such as DOCK8 and SPRR3 were reported for the first time to be associated with oral carcinogenesis. Gene expression profiling of oral cancer may help to unlock the molecular basis of phenotype, response to treatment and heterogeneity of disease. Thus findings on genetic instability and gene expression profiling of oral cancer from this study will provide foundations for future research and may help in the development of molecular signatures for accurate prediction of clinical outcome and improved therapeutic strategies for patients with oral cancer.

References:


Synopsis


Gene Expression Profiling and Genetic Variations in Oral Cancer Associated with Tobacco Consumption

THESIS

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

DHIRENDRA SINGH YADAV

Under the Supervision of

Dr. Sujala Kapur

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASHTHAN) INDIA
2012
BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
PILANI (RAJASHTHAN)

CERTIFICATE

This is to certify that the thesis entitled “Gene Expression Profiling and Genetic Variations in Oral Cancer Associated with Tobacco Consumption” submitted by DHIREDRA SINGH YADAV (ID No 2006PHXF021P) for award of Ph. D. Degree of the Institute, embodies original work done by him under my supervision.

Signature in full of the Supervisor: (Dr. SUJALA KAPUR)
Name in capital block letters: SCIENTIST ‘E’/ DEPUTY DIRECTOR
Designation: National Institute of Pathology (ICMR)
Place: New Delhi

Date:
Acknowledgement

It is a pleasure to thank those who made this thesis possible. First and foremost, I would like to express my deepest gratitude and regards to my Supervisor Dr. Sujala Kapur, for her unflinching encouragement, guidance and support in various ways from the initial to the final level of this work, enabled me to develop an understanding of the subject. Her valuable criticism, positive attitude, great patience and friendly guidance helped me to develop clarity to my ideas and better writing skills which certainly would lead me to become a better science person. I am indebted to her more than she knows.

I express my heartfelt gratitude to Dr. Sunita Saxena, Director, National Institute of Pathology (ICMR), for letting me join this institute and providing me facilities to pursue research work in the Institute. I am thankful to her for consistent guidance, help and encouragement and particularly for her keen insight, acute sense of perception and refusal to settle for nothing less than the best.

I am immensely thankful to Prof. B. N. Jain, (Vice-Chancellor), Prof. R. Prakash (Ex-Dean), Prof. A. Das (Ex-Dean), Research and Consultancy Division (RCD), BITS, Pilani, for providing me this opportunity to pursue the off-campus Ph.D of the Institute.

I express my deep gratitude to Doctoral Advisory Committee member, and present Dean Prof. S. K. Verma, (Dean-RCD), BITS, Pilani for his consistent official support, timely response, encouragement and providing valuable suggestions for improvement in my thesis.

I am thankful to my DAC member, Dr. Sanjeev Kumar, Department of Biological Sciences, BITS, Pilani, for investing his valuable time to go through my draft thesis and providing several valuable suggestions that immensely helped in improving the quality of final Ph.D. thesis.

I would also like to thank Dr. Hemanth Jadav, Mr. Sharad Shrivastava, Mr. Dinesh Kumar, Ms. Monica Sharma, Mr. Gunjan Soni, Mr. Amit Kumar and Ms. Sunita Bansal, nucleus members of RCD, BITS, Pilani, without whose cooperation, fast response on any query, and guidance, it would not have been possible for me to pursue such goal oriented research during each of the past few semesters. I also express my gratitude to the office staff of RCD whose secretarial assistance helped me in submitting the various evaluation documents in time.

I would like to thank Dr. Usha Agrawal, Scientist E, National Institute of Pathology (ICMR), for her useful suggestions, encouragement and positive attitude which enhanced my self confidence to face the difficulties.
I am thankful to Dr. Anju Bansal and Dr. B. S. A. Raju for their valuable support, suggestions and help in various forms. I am very much thankful to Dr. L. C. Singh for his keen support and guidance in performing various realtime PCR and sequencing related experiments. I am also very much thankful to Dr. A. K. Mishra for always been there for all questions, big and small related to statistics. Without your statistical expertise I would have been at a loss.

My thanks giving would be incomplete without mentioning my past and present lab mates Dr. Anurupa Chakraborty, Dr. Bharat Bhusan, Dr. Anand Verma, Dr. Indranil Chatterji, Mrs. Abha Soni, Ms. Thodam Regina, Mr. Pradeep Singh Chauhan, Mrs. Mishi Kaushal, Mrs. Rakhshan Ikshan, Mrs. Shanti Latha, Mrs. Shreshtha Malviya, Ms. Meena Lakhanpal, for their encouraging, suggestions, support and cooperation.

I could always rely on Mrs. Saratha, Administrative officer, National Institute of Pathology, for her help in countless instances and express my deep felt thanks to her. I wish to express my gratitude to the staff of the administrative section, library, computer and store section at Institute of Pathology for all the support they provided me.

I sincerely thank to lab members Mrs. Valsamma Mathew and Mr. Jagdish Pant for their caring and supportive attitude. I also thank to Mr. Mohan Singh, Mr. Jagat and Mr. Sajid Hussain for their assistance and help during the research work.

For this type of research, data was essential and many people helped me with this and I would like to thank them wholeheartedly. Hundreds of oral cancer patients and their friends and relatives have filled out questionnaires for me and without their generosity there would be nothing to work with. I would like to thank Dr. A. C. Kataki, Director, Dr. B.B.C.I. Guwahati, Assam; Dr. Jagannath Sharma, Senior Pathologist, Dr. B.B.C.I. Guwahati; Dr. Yogesh Verma, Senior Consultant, Sir T.N.M. Hospital Gangtok, Sikkim; Dr. Erick Zamoivia, Senior Consultant, Civil Hospital, Aizawl, Mizoram. And thanks to supporting staff at these centres Mr. Bhumi, Mr. Najmul Hque, Mss. Sunita and Mr. David who helped in collection of samples from these centres.

This thesis work was not possible without financial support from University Grant Commission (NET-JRF), Indian council of medical research (ICMR, SRF), and National Institute of Pathology, India for proving me research fellowship and funding for research work from ICMR. Thus their immense support is gratefully acknowledged.

I wish to thank my best friends Dhiraj, Shashank, Regina, Anas, Abhishek, Aapurb, Abha, Shubhra, and Sanjay for helping me get through the difficult times, and for all the emotional support, entertainment, and caring they provided during this whole period of research.
Finally, this thesis would not have been possible without the confidence, endurance and support of my family. My family has always been a source of inspiration and encouragement. Most importantly, I wish to thank my parents (Smt. Kiran Yadav and Shri. R. S. L. Yadav) whose love, teachings and support have brought me this far. During these long years of academic degree their patience and tolerance was really incredible. I would also like to thank to my elder brothers Satyendra and Kaushalendra; Sister in laws (Bhabhi Ji) Mrs. Anita and Mrs. Meena; my loving nieces Tejashwita and Yashashwi; nephews Krishna and Saurya for providing a loving environment which helped me keep going.

Last but not least, God Almighty, thank you for the wonderful life that you have given me and the great set of human beings that you have made me friends with, I thank.

Date: (Dhirendra Singh Yadav)

Place: New Delhi
Dedicated

To

My Parents and Teachers
Abstract

Oral cancer is the sixth most common cancer worldwide, accounting for 4% of cancers in men and 2% of cancers in women. Annually, over 300,000 new cases of oral cancer are diagnosed all over the world where the majority of them are diagnosed in the advanced stages III or IV. Such data make the oral cancer an important public health problem which is responsible for 3% to 10% of cancer mortality worldwide. There is up to a 20-fold geographical variation in the incidence rates of oral cancer. In India, oral cancer is the most common cancer among men and ranks third among women, with age-standardized incidence rates of 12.8 and 7.5 per 100,000 population respectively. Northeastern states in India have reported a very high prevalence of oral cancer. In this region there is widespread chewing habit of tobacco with peculiarly fermented betel nut. This study is based on oral cancer samples collected from three regional collaborating centers (Guwahati, Sikkim and Aizawl) of Northeast (NE) India. These are the Population Based Cancer Registry (PBCR) centers located in NE India which were in collaboration with our Institute for five years (2005-2010), and provided samples for this study. Knowledge of cancer genetics is rapidly improving our understanding of cancer biology, helping to identify at-risk individuals, furthering the ability to characterize malignancies, establishing treatment tailored to the molecular fingerprint of the disease, and leading to the development of new therapeutic modalities. Given the poor prognosis associated with oral cancer, there is an urgent need to elucidate the molecular determinants and critical signal pathways underlying the malignant transformation of precancerous to cancerous tissue which may lead to identification of novel diagnostic and therapeutic targets.

The cell of origin of oral squamous cell carcinoma (OSCC) is the oral keratinocyte. Oral cancer, as any cancer, is caused by DNA mutation, often spontaneous but increased by exposure to any of a range of mutagens; chemical (such as polycyclic aromatic hydrocarbons, nitrosamines, aldehydes and ketones), physical (such as ultraviolet radiations) or microbial (such as Human papilloma virus, Epstein-barr virus and Hepatitis C virus). The metabolism of environmental carcinogens depends on the efficiency of enzymatic detoxification pathways. The role of polymorphisms of genes responsible for detoxification of xenobiotics and p53 codon 72 were investigated in this study in relation to oral carcinogenesis. Two hundred and thirty five oral cancer cases and 289 healthy controls from high incidence region of oral cancer in NE India were included in this study. Eight polymorphisms in seven genes
[CYP1A1 (Msp1 and Nco1), GSTT1, GSTM1, GSTP1, NAT2, NQO1 and Codon 72 of P53] were analyzed using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). These genes and their polymorphisms were selected on the basis of their reported role in detoxification of various carcinogenic compounds present in tobacco and betel-quid. In earlier studies, polymorphic variants of these genes in certain specific codons have been reported to code for enzymes with less or nil efficiency of detoxification.

In the current study tobacco chewing as well as smoking were found to impart a significant risk for oral cancer (AOR=2.78, 95% CI=1.71-4.51, and AOR=1.57, 95% CI=1.05-2.35 respectively) with tobacco chewers at higher risk compared to smokers. Frequency of betel quid chewers and alcohol consumers were also higher in cases (80% and 37%, respectively) compared with controls (66% and 30%, respectively) but when adjusted with other risk factors (alcohol, betel-quid, tobacco chewing, tobacco smoking and polymorphisms of various genes) the difference was statistically insignificant. GSTT1 and GSTM1 null genotypes and the variant genotypes of CYP1A1*2A, CYP1A1*2C (codes for key enzymes in the phase I bioactivation of xenobiotics), and p53 codon72 were not found to be associated with oral cancer risk. Homozygous variant genotypes of NAT2 (AA) were significantly higher in cases (14%) compared to controls (9%) [OR=1.83, 95% CI=1.01-3.31, P=0.04]. However when adjusted with other risk factors (alcohol, betel-quid, tobacco chewing, tobacco smoking and polymorphisms of various genes) this risk became statistically insignificant. Combined frequency distribution of NQO1 genotypes Pro/Pro, Pro/Ser and Ser/Ser from all three geographical regions in this study, was 45%, 35% and 20% in cases and 60%, 27% and 30% in controls respectively. This difference was statistically significantly different for heterozygous (AOR=1.64, 95% CI=1.04-2.58, p=0.03) as well as for homozygous (AOR=1.81, 95% CI=0.98-3.32, p=0.06) variant genotypes. When data was analyzed in different geographic regions of North East India, the GSTT1 null genotype and homozygous variant genotypes of GSTP1 were found to impart significant risk for oral cancer (AOR= 2.58, 95% CI 1.01–6.61, p= 0.05, and AOR=3.14, 95% CI 0.94:10.49, p=0.06 respectively) in samples obtained from the Assam region. Further, variant genotypes of CYP1A1*2A were found to impart significant risk for heterozygous (AOR = 4.55, 95% CI 0.88-23.36, p=0.07) as well as for homozygous genotypes (AOR= 6.38, 95% CI 1.10-40.83, p=0.05) for oral cancer in Sikkim population of NE India. Gene-environment interaction analysis revealed that the variant genotypes of NQO1 did not interact statistically with tobacco consumption habits. It is possible that although variant genotypes of NQO1 may play
an important role in the genetic susceptibility to oral cancer, its pathway is unrelated to the detoxification mechanism of tobacco constituents. On gene expression profiling analysis, \textit{NQO1} was found to be significantly up-regulated in our study. Thus, it appears that it may be the variant form of \textit{NQO1} and not its expression level, which may be responsible for the higher risk of oral cancer in this region.

Oral cancer is a multi-factorial disease and influenced by both aetiological factors and ethnicity; as a result molecular profiles of oral cancer vary throughout the world. Altered gene expression profiling may be attributed to several types of genetic alterations such as base changes, insertions, deletions, chromosomal translocations, inversions, loss of heterozygosity and copy number variations. In the current study targeted re-sequencing of 169 functionally relevant and potentially important genes was done to analyze mutational changes. These genes were selected by literature survey on the basis of their significant role in tobacco associated cancers. These genes are either tumor suppressor genes or they have role in inflammation. In our study 4837 exon regions of interest were analyzed with average read depth of 47.14. A large number of mutational changes were observed including 96 SNPs (50 novel and 46 known SNPs) and 46 Insertions and deletions (29 novel InDels and 17 known InDels). Gene ontology (GO) analysis showed that most of the known SNPs with high read depths (>50) were located in tumor suppressor genes such as \textit{RB1}, \textit{FHIT}, \textit{FAT1}, \textit{FAT2} and \textit{VHL}. SNPs detected in \textit{RB1}, \textit{FHIT} and \textit{FAT1} were located in the intronic regions of the gene while those in \textit{ATM}, \textit{VHL}, \textit{IL12B}, and \textit{MET} were located in 3'UTR. Though these are non-coding regions, earlier studies have reported their involvement in regulation of gene activity, thus they may have functional relevance in oral carcinogenesis.

Novel SNPs with significant read depths (>25), and with non-synonymous type variations which may lead to structural and functional changes in resulting protein were observed in tumor suppressors such as \textit{FAT1}, \textit{FAT2}, and \textit{TP53}; cell cycle regulators such as \textit{NOTCH2}, \textit{ATM}, and \textit{CDH3} or receptors of growth factor such as \textit{MET}. Synonymous type variations were observed in \textit{APC} (tumor suppressor) and \textit{IL12B} (cytokine) genes and those present in non-coding regions were observed in or near to \textit{EGFR}, \textit{STAT5B}, \textit{CDK5} and \textit{MYC} genes.

Insertions and deletions (InDels) of bases are among highly damaging mutations and responsible for carcinogenesis as they may result in abrupt changes in expression of genes or in the structural or functional activities of the encoded protein. The known deletions observed
in our study were present in non-coding regions of the gene, 4 of them were present in 3’UTR (TSC1, FAT1, MAP2K6, and ERBB4), two at 5’UTR (BMP4, and SLC22A18) and one in intronic region of BRCA1. UTR regions of a gene have significant role in regulation of gene activity, thus its alteration may have an adverse effect on gene activity. The known deletion with highest read depth (272) was present in UTR region of TSC1 gene (rs34947162; rs115091888). TSC1 plays a central role in regulating cell survival and proliferation signaling pathways. A total of 11 novel deletions have been observed in oral cancer cases in our study. Novel deletion with the highest read depth (58) was present in MSH6 gene (NM_000179 at position 48033455). Other important genes with novel deletions include IGF1R, BRCA2, TSC2 and PAK3.

Of the known insertions observed in our study, 4 were present in UTR regulatory regions of APC, SMAD2, RHOB and NBL1 genes while the remaining 6 were located in intronic regions of ADH6, PDGFRA, BRIP1, FAT2, DLG2 and KLK8. The insertion with highest read depth (102) was that of base A at position 112180228 in APC gene (rs11432316; rs79379053). This gene encodes a tumor suppressor protein that acts as an antagonist of the Wnt signaling pathway.

Alteration of gene expression was done by cDNA microarray to identify genes differentially expressed in OSCC associated with prevalent risk factors in that region such as tobacco and betel quid chewing. Six hundred and thirty four differentially expressed (247 upregulated and 387 down-regulated) genes were identified in OSCC tissues compared to normal oral mucosa distant from the tumor site. Most significantly up-regulated pathways were related to ribosomal activity (RPL38, RPSA, RPL6, RPS3, RPS20, RPS6, and RPS7), Neuroactive ligand-receptor interaction (GRM8, GRM4, NTSR1, and P2RX7), ECM-receptor interaction (LAMC2, ITGB1, GP6 FN1, and COL2A1), Aminoacyl-tRNA biosynthesis (LARS2, AARS, WARS). Functional annotation clustering of up-regulated genes using web-based DAVID analysis also showed most significantly enriched cluster to be associated with ribosomal activity, ribosome biogenesis, and translation (involving genes such as COPS5, IDO1, KNYU, and RPS7). Second cluster of up-regulated genes was related to inflammatory response, defense response and response to wounding (involving genes such as IL8, INHBA, and PDNP).

Most significantly downregulated pathways in our study (when analyzed by Genowiz) were related to cell junctions (involving 12 genes KRT34, DSG1, KRT2, KRT15, KRT36,
VIM, KRT4, KRT10, KRT78, KRT13, ACTB, and KRT33A), and Valine, leucine and isoleucine degradation (ALDH3A1, BCAT2, HIBCH, BCKDHB, ACAT2, ALDH2, and ALDH3A2). While DAVID analysis revealed clusters with highest enrichment score of downregulated genes to be associated with structural constituents of cytoskeleton (involving genes such as CRYAB, KRT4, TPM1, KRTAP5-9), epidermal cell differentiation, and Keratinisation (involving genes such as ALDH3A2, JUN, KRT10, KRT13, KRT15, KRT2, KRT34, TP63, KRT4, KRTAP5-9, and SPRR3).

Eight significantly deregulated genes [four up-regulated (PDPN, IL8, COPS5, and INHBA) and four down regulated (KRT4, DOCK8, SPRR3, and MAL)] were selected from microarray data and validated by real-time RT PCR. Our study revealed several genes such as DOCK8 and SPRR3 which were reported for the first time to be associated with oral carcinogenesis. Recent reports have also suggested significance of KRT4 and MAL to work as possible prognostic and therapeutic markers for oral carcinogenesis and DOCK8 and SPRR3 may be further investigated for their association with oral carcinogenesis.

Genes such as FAT1, TSC, GAS7 and APC showed high level of genomic instability (including known and novel variations) in OSCC. Furthermore GAS7 which is functionally involved in chemokine-induced migration in the immune system was also found to be significantly downregulated in our gene expression profiling study. Thus these genes may be useful prognostic and therapeutic targets in OSCC.

This is the first study to provide gene expression profiling and genetic variations of oral cancer associated with betel quid and tobacco consumption habits in a high-risk region of North-East India. Thus, our findings on genetic instability and gene expression profiling of oral cancer will provide foundations for future research and may help in the development of molecular signatures for accurate prediction of clinical outcome and improved therapeutic strategies for patients with oral cancer.
# Contents

List of Figures | I
---|---
List of Tables | V
List of Abbreviations | VII

1 Introduction | 1
2 Aims and Objectives | 9
3 Review of Literature | 12
   Cancer | 13
   Oral cancer | 13
   The Global distribution and the Indian scenario of oral cancer | 15
   Age and sex distribution | 18
   Survival and Mortality | 19
   Risk factors for oral cancer | 20
   Treatment and main complications of oral cancer | 25
   Genes and Genetics of Oral Cancer | 27
   Role of carcinogen detoxification pathways in oral cancer | 35
   The tumour microenvironment | 45
   Signaling interactions in the tumour microenvironment | 46
   Hallmarks of cancer and possible therapeutic targeting | 48
   Analyzing genomewide aberrations in cancer | 49
   Next Generation Sequencing Technology | 49
   Gene expression profiling using microarray | 55
   Real time reverse transcription PCR | 60
4 Role of SNPs in Xenobiotic Metabolizing Genes (Encoding Phase I and Phase II Detoxification Enzymes) and p53 Codon 72 Polymorphism in Oral Cancer | 66
   Introduction | 67
   Materials and Methods | 72
   Results | 85
   Discussion | 93
| 5 | Detection of Genomic Instability/ Genomic alterations using Targeted Next Generation Sequencing in OSCC | 100 |
|   | Introduction | 101 |
|   | Materials and Methods | 103 |
|   | Results | 109 |
|   | Discussion | 123 |

| 6 | Differential Gene Expression Profiling of Oral Squamous Cell Carcinoma and Adjacent Noncancerous Tissues with cDNA Microarray in North East Indian Population | 135 |
|   | Introduction | 136 |
|   | Materials and Methods | 138 |
|   | Results | 142 |
|   | Discussion | 167 |

| 7 | Validation of Gene Expression Profiling of OSCC by Quantitative Real time RT-PCR | 173 |
|   | Introduction | 174 |
|   | Materials and Methods | 176 |
|   | Results | 178 |
|   | Discussion | 184 |

| 8 | Conclusions and Future Scope of Work | 189 |

| 9. | References | 195 |

| 10. | List of Publications | 236 |
| 11. | Biography of Candidate | 241 |
| 12. | Biography of Supervisor | 245 |
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig 3.1</td>
<td>Countries with high incidence and mortality with oral cancer</td>
<td>16</td>
</tr>
<tr>
<td>Fig 3.2</td>
<td><em>Age-Standardized Oral Cavity Cancer Incidence Rates by Sex and World Area</em></td>
<td>17</td>
</tr>
<tr>
<td>Fig 3.3</td>
<td>Figure showing India as one of the highest affected countries with oral cancer</td>
<td>20</td>
</tr>
<tr>
<td>Fig 3.4</td>
<td>Postulated causative factors and mechanisms implicated in the induction of oral squamous cell carcinoma due to Consumption of tobacco, betel quid, pan masala and gutkha</td>
<td>21</td>
</tr>
<tr>
<td>Fig 3.5</td>
<td>Risk factors of OSCC development</td>
<td>22</td>
</tr>
<tr>
<td>Fig 3.6</td>
<td>State wise prevalence of tobacco consumption in India</td>
<td>23</td>
</tr>
<tr>
<td>Fig 3.7</td>
<td>Treatment of oral squamous cell carcinoma, and main complications</td>
<td>26</td>
</tr>
<tr>
<td>Fig 3.8</td>
<td>Phase I and Phase II enzymatic detoxification in human</td>
<td>36</td>
</tr>
<tr>
<td>Fig 3.9</td>
<td>The GSTM1 gene is part of the Mu-class GST gene cluster at 1p13.3</td>
<td>39</td>
</tr>
<tr>
<td>Fig 3.10</td>
<td>The GSTT1 gene is part of the Theta-class GST gene cluster at 22q11.2</td>
<td>39</td>
</tr>
<tr>
<td>Fig 3.11</td>
<td>Overview of GSTP1 gene at 11q13 locus, mRNA, and protein</td>
<td>40</td>
</tr>
<tr>
<td>Fig 3.12</td>
<td>Proposed mechanism of stabilization of p53 via a protein-protein interaction with NQO1</td>
<td>43</td>
</tr>
<tr>
<td>Fig 3.13</td>
<td>The Cells of the Tumor Microenvironment</td>
<td>45</td>
</tr>
<tr>
<td>Fig 3.14</td>
<td>Signaling interactions in the tumour microenvironment during malignant progression</td>
<td>46</td>
</tr>
<tr>
<td>Fig 3.15</td>
<td><em>Intracellular Signaling Networks Regulate the Operations of the Cancer Cell</em></td>
<td>47</td>
</tr>
<tr>
<td>Fig 3.16</td>
<td>Hallmarks of cancer and possible therapeutic targeting</td>
<td>48</td>
</tr>
<tr>
<td>Fig 3.17A-D</td>
<td><em>Illumina-Solexa Sequencing by Synthesis method for Next Generation Sequencing</em></td>
<td>52-54</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>Fig 3.18</td>
<td>Six Steps in a Microarray Experiment</td>
<td></td>
</tr>
<tr>
<td>Fig 3.19</td>
<td>Real-time PCR chemistries</td>
<td></td>
</tr>
<tr>
<td>Fig 3.20</td>
<td>The principle of the 5’ nuclease (TaqMan) assay</td>
<td></td>
</tr>
<tr>
<td>Fig 3.21</td>
<td>qRT-PCR amplification curves plot fluorescence signal versus cycle number</td>
<td></td>
</tr>
<tr>
<td>Fig 4.1</td>
<td>Estimation of the quality of genomic DNA on EtBr stained 0.8% agarose gel</td>
<td></td>
</tr>
<tr>
<td>Fig 4.2</td>
<td>2% Agarose gel showing multiplex PCR for GSTT1 and GSTM1</td>
<td></td>
</tr>
<tr>
<td>Fig 4.3A</td>
<td>2% agarose gel stained with EtBr showing PCR product of GSTP1</td>
<td></td>
</tr>
<tr>
<td>Fig 4.3B</td>
<td>3.5% agarose gel showing RFLP fragments of GSTP1</td>
<td></td>
</tr>
<tr>
<td>Fig 4.4A</td>
<td>2% agarose gel showing PCR product of CYP1A1*2A</td>
<td></td>
</tr>
<tr>
<td>Fig 4.4B</td>
<td>3.5% agarose gel showing RFLP fragments of CYP1A1*2A</td>
<td></td>
</tr>
<tr>
<td>Fig 4.5A</td>
<td>2% agarose gel showing PCR product of CYP1A1*2C</td>
<td></td>
</tr>
<tr>
<td>Fig 4.5B</td>
<td>3.5% agarose gel showing RFLP fragments of CYP1A1*2C</td>
<td></td>
</tr>
<tr>
<td>Fig 4.6A</td>
<td>2% agarose gel showing PCR product of NAT2</td>
<td></td>
</tr>
<tr>
<td>Fig 4.6B</td>
<td>3.5% agarose gel showing RFLP fragments of NAT2</td>
<td></td>
</tr>
<tr>
<td>Fig 4.7A</td>
<td>2% agarose gel showing PCR product of NQO1</td>
<td></td>
</tr>
<tr>
<td>Fig 4.7B</td>
<td>3.5% agarose gel showing RFLP fragments of NQO1</td>
<td></td>
</tr>
<tr>
<td>Fig 4.8A</td>
<td>2% agarose gel showing amplified product of codon 72 of p53</td>
<td></td>
</tr>
<tr>
<td>Fig 4.8B</td>
<td>3.5% agarose gel showing RFLP fragments of codon 72 of p53</td>
<td></td>
</tr>
<tr>
<td>Fig 5.1</td>
<td>On-array DNA capture workflow used for next generation sequencing by Illumina</td>
<td>104</td>
</tr>
<tr>
<td>Fig 5.2</td>
<td>Bioanalyzer profiles of sonicated samples</td>
<td>105</td>
</tr>
<tr>
<td>Fig 5.3</td>
<td>Process of library preparation</td>
<td>105</td>
</tr>
<tr>
<td>Fig 5.4</td>
<td>Bioanalyzer profile of amplified product (ePCR1 and ePCR2)</td>
<td>106</td>
</tr>
<tr>
<td>Fig 5.5</td>
<td>Analysis flow for identifying genomic variations</td>
<td>107</td>
</tr>
<tr>
<td>Fig 5.6</td>
<td>Diagramatic representation of genomic aberrations observed in OSCC</td>
<td>111</td>
</tr>
<tr>
<td>Fig 5.7</td>
<td>Bar diagram of Known SNPs observed in OSCC (Read depth versus Genes)</td>
<td>112</td>
</tr>
<tr>
<td>Fig 5.8</td>
<td>Bar diagram of Novel SNPs observed in OSCC (Read depth versus Genes)</td>
<td>112</td>
</tr>
<tr>
<td>Fig 5.9</td>
<td>Bar diagram of Known and Novel Deletions in OSCC (Read depth vs Genes)</td>
<td>113</td>
</tr>
<tr>
<td>Fig 5.10</td>
<td>Bar diagram of Known and Novel Insertions in OSCC (Read depth vs Genes)</td>
<td>114</td>
</tr>
<tr>
<td>Fig 6.1</td>
<td>RNA integrity checked by Agilent Bio-analyser and by agarose gel electrophoresis</td>
<td>139</td>
</tr>
<tr>
<td>Fig 6.2</td>
<td>Experimental design for microarray experiment</td>
<td>140</td>
</tr>
<tr>
<td>Fig 6.3</td>
<td>Boxplots for raw and log2-transformed intensities for samples</td>
<td>143</td>
</tr>
<tr>
<td>Fig 6.4</td>
<td>Boxplots for each sample after centering and scaling the respective intensity data</td>
<td>144</td>
</tr>
<tr>
<td>Fig 6.5</td>
<td>Volcano plot showing distribution of up and down regulated genes</td>
<td>145</td>
</tr>
<tr>
<td>Fig 6.6</td>
<td>Scatter plot showing the distribution of up and down regulated genes</td>
<td>145</td>
</tr>
<tr>
<td>Fig 6.7</td>
<td>Heatmap showing the relatedness within samples and within genes</td>
<td>147</td>
</tr>
<tr>
<td>Fig 6.8</td>
<td>Network-1, of down-regulated genes analysed by IPA</td>
<td>159</td>
</tr>
<tr>
<td>Fig 6.9</td>
<td>Network-2, of down-regulated genes analysed by IPA</td>
<td>160</td>
</tr>
<tr>
<td>Fig 6.10</td>
<td>Network-3, of down-regulated genes analysed by IPA</td>
<td>161</td>
</tr>
<tr>
<td>Fig 6.11</td>
<td>Network-4, of down-regulated genes analysed by IPA</td>
<td>162</td>
</tr>
<tr>
<td>Fig 6.12</td>
<td>Network-5, of down-regulated genes analysed by IPA</td>
<td>163</td>
</tr>
</tbody>
</table>
Fig 6.13  Network-1, of up-regulated genes  analysed by IPA  
Fig 6.14  Network-2, of up-regulated genes  analysed by IPA  
Fig 6.15  Network-3, of up-regulated genes  analysed by IPA  
Fig 6.16  Network-4, of up-regulated genes  analysed by IPA  
Fig 6.17  Network-5, of up-regulated genes  analysed by IPA  

Fig 7.1  Bar diagram showing fold change of gene expression RT-PCR for PDPN  
Fig 7.2  Bar diagram showing fold change of gene expression by RT-PCR for IL8  
Fig 7.3  Bar diagram showing fold change of gene expression by RT-PCR for COPS5  
Fig 7.4  Bar diagram showing fold change of gene expression by RT-PCR for INHBA  
Fig 7.5  Bar diagram showing fold change of gene expression by RT-PCR for KRT4  
Fig 7.6  Bar diagram showing fold change of gene expression by RT-PCR for MAL  
Fig 7.7  Bar diagram showing fold change of gene expression by RT-PCR for SPRR3  
Fig 7.8  Bar diagram showing fold change of gene expression by RT-PCR for DOCK8  
Fig 7.9  Comparison of gene expression level by microarray data with that of real time data  
Fig 7.10 Correlation between gene expression of eight genes by real time PCR and cDNA microarray
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Table Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Primers and restriction enzymes used and analysis of their products</td>
<td>75</td>
</tr>
<tr>
<td>4.2</td>
<td><em>Reaction components of multiplex PCR for GSTT1 and GSTM1</em> genotyping</td>
<td>76</td>
</tr>
<tr>
<td>4.3</td>
<td><em>Demographic distribution of various life style factors in oral cancer and control population</em></td>
<td>87</td>
</tr>
<tr>
<td>4.4</td>
<td><em>Demographic distribution of various genetic factors in oral cancer and control population</em></td>
<td>88</td>
</tr>
<tr>
<td>4.5</td>
<td>Risk estimates for various lifestyle risk factors</td>
<td>89</td>
</tr>
<tr>
<td>4.6</td>
<td>Risk estimates for polymorphisms of CYP1A12A, CYP1A12C, NAT2 and NQO1 genes with oral cancer</td>
<td>90</td>
</tr>
<tr>
<td>4.7</td>
<td>Risk estimates for polymorphisms of GSTT1, GSTM1 and GSTP1 genes with oral cancer</td>
<td>91</td>
</tr>
<tr>
<td>4.8</td>
<td>Frequency distribution and risk estimates for p53 codon 72 polymorphisms in oral cancer</td>
<td>92</td>
</tr>
<tr>
<td>5.1</td>
<td><em>Demographic and clinical characteristics of oral squamous cell carcinoma cases used for detection of genomic alterations by NGS</em></td>
<td>109</td>
</tr>
<tr>
<td>5.2</td>
<td><em>Genomic Variations in Oral Cancer Cases in a Nutshell</em></td>
<td>110</td>
</tr>
<tr>
<td>5.3</td>
<td>Known SNPs in Oral Cancer</td>
<td>115</td>
</tr>
<tr>
<td>5.4</td>
<td>Novel SNPs in oral cancer</td>
<td>117</td>
</tr>
<tr>
<td>5.5A</td>
<td>Known Deletions in Oral Cancer</td>
<td>119</td>
</tr>
<tr>
<td>5.5B</td>
<td>Novel Deletions in Oral Cancer</td>
<td>120</td>
</tr>
<tr>
<td>5.6A</td>
<td>Known Insertions in Oral Cancer</td>
<td>121</td>
</tr>
<tr>
<td>5.6B</td>
<td>Novel Insertions in Oral Cancer</td>
<td>122</td>
</tr>
<tr>
<td>6.1</td>
<td><em>Demographic and clinical characteristics of oral squamous cell carcinoma cases used for differential gene expression profiling of OSCC</em></td>
<td>142</td>
</tr>
<tr>
<td>6.2A</td>
<td>List of top 25 Significantly Down-regulated Genes in OSCC</td>
<td>150</td>
</tr>
<tr>
<td>6.2B</td>
<td>List of top 25 Significantly Up-regulated Genes in OSCC</td>
<td>151</td>
</tr>
</tbody>
</table>
6.3A Significantly Up-regulated Pathways (analyzed using Genowiz\textsuperscript{TM} software) 152

6.3B Significantly Down-regulated Pathways (analyzed using Genowiz\textsuperscript{TM} software) 153

6.4A Functional Annotation Clustering of Up-regulated Genes by (analyzed using DAVID 6.7 software) 154

6.4B Table 6.4B: Functional Annotation Clustering of Down-regulated Genes (analyzed using DAVID 6.7 software) 155

6.5A Top 5 networks of down-regulated genes and their functional categories as revealed by IPA analysis 156

6.5B Top 5 networks of up-regulated genes and their functional categories as revealed by IPA analysis 156

6.6A Downregulated pathways by IPA (Molecular and cellular function) 157

6.6B Up-regulated Pathways by IPA (Molecular and cellular function) 157

6.7A Top 5 Canonical Pathways of Downregulated Genes 158

6.7B Top 5 Canonical Pathways of Up-regulated Genes 158

7.1 Information on the eight genes examined by Real-Time PCR 177

7.2 Demographic and clinical characteristics of oral squamous cell carcinoma cases used for validation by real time RT-PCR 178
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAR</td>
<td>Age adjusted incidence rate</td>
</tr>
<tr>
<td>AI</td>
<td>Allelic Imbalance</td>
</tr>
<tr>
<td>BBCI</td>
<td>Bhubaneswar Boorah Cancer Institute</td>
</tr>
<tr>
<td>BQ</td>
<td>Betelquid</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>CYP</td>
<td>Cytochrome P-450</td>
</tr>
<tr>
<td>CT</td>
<td>Computerized Tomography</td>
</tr>
<tr>
<td>CTRT</td>
<td>Chemotherapy-Radiotherapy</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene diamine tetra acetic acid</td>
</tr>
<tr>
<td>EtBr</td>
<td>Ethidium Bromide</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>HPV</td>
<td>Human Papilloma virus</td>
</tr>
<tr>
<td>ICMR</td>
<td>Indian Council of Medical Research</td>
</tr>
<tr>
<td>IARC</td>
<td>International Agency for Research on Cancer</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilo base</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss of Heterozygocity</td>
</tr>
<tr>
<td>MIAME</td>
<td>Minimum information about microarray experiment</td>
</tr>
<tr>
<td>MSI</td>
<td>Microsatellite Instability</td>
</tr>
<tr>
<td>Mg</td>
<td>Microgram</td>
</tr>
<tr>
<td>ML</td>
<td>Microlitre</td>
</tr>
<tr>
<td>Mg</td>
<td>Milligram</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger Ribonucleic Acid</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>NAT</td>
<td>N-acetyletransferases</td>
</tr>
<tr>
<td>NCRP</td>
<td>National Cancer Registry Programme</td>
</tr>
<tr>
<td>NE</td>
<td>North East</td>
</tr>
<tr>
<td>NGS</td>
<td>Next Generation Sequencing</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>NQO1</td>
<td>NAD(P)H dehydrogenase, quinone 1</td>
</tr>
<tr>
<td>OD</td>
<td>Optical Density</td>
</tr>
<tr>
<td>OSCC</td>
<td>Oral Squamous Cell Carcinoma</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PBCR</td>
<td>Population Based Cancer Registry</td>
</tr>
<tr>
<td>PET</td>
<td>Positron Emission Tomography</td>
</tr>
<tr>
<td>RIN</td>
<td>RNA Integrity Number</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction Fragment Length Polymorphism</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal Ribonucleic acid</td>
</tr>
<tr>
<td>RT</td>
<td>Room Temperature</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse Transcriptase- Polymerase Chain Reaction</td>
</tr>
<tr>
<td>SCC</td>
<td>Squamous Cell Carcinoma</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>TBE</td>
<td>Trizma-base Boric Acid EDTA</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris (hydroxymethyl) amino acid</td>
</tr>
<tr>
<td>TNM</td>
<td>Tumour size, Nodal metastasis, Distant metastases</td>
</tr>
<tr>
<td>TSG</td>
<td>Tumour Suppressor Gene</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>UICC</td>
<td>International Union Against Cancer</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Cancer has overtaken heart disease as the major cause of death in the world in 2011, part of a trend that will more than double global cancer cases with an estimated 12 million deaths by 2030 (WHO-Report, 2009). Cancer is not just one disease, but a generic term used to denote a group of more than two hundred diseases in which cells in a part of the body begin to grow out of control. Growing out of control and invading other tissues are what makes a cell a cancer cell. Cancer cells display uncontrolled growth (division beyond the normal limits), invasion (intrusion on and destruction of adjacent tissues), and sometimes metastasis (spread to other locations in the body via lymph or blood). These abnormal characteristics of cells are attributed to multiple genomic alterations, including point mutations, translocations, gene amplifications, epigenetic modifications, deletions, aberrant splicing, and altered gene expression. Oral squamous cell carcinoma (OSCC), one of the most common forms of oral cancer is usually associated with environmental carcinogen and other lifestyle risk factors (Tsantoulis et al., 2007). About half of the patients affected will die within five years of diagnosis, while surviving patients may be left with severe esthetic and/or functional compromise. One of the most striking features of oral cancer is the disfiguration of the face, which adds further importance for its study and treatment (Silverman, 2001).

Oral cancer is the sixth most common cancer worldwide, accounting for 4% of cancers in men and 2% of cancers in women (Napier and Speight, 2008). Annually, over 3,00,000 new cases of oral cancer are diagnosed all over the world where the majority of them are diagnosed in the advanced stages III or IV. Such data make the oral cancer an important public health problem which is responsible for 3% to 10% of cancer mortality worldwide. There is up to a 20 fold geographical variation in the incidence rates of oral cancer (Warnakulasuriya, 2009). The areas with high incidence rates for oral cancer are found in the South and South East Asia. It has been estimated that of the 390,000 oral and oro-pharyngeal cancers that occur annually worldwide, 58% occur in south and south-east Asia (Nair et al., 2004). In India oral cancer is the most common cancer among men and ranks third among women (Soya et al., 2007), with age-standardized incidence rates of 12.8 and 7.5 per 100,000 population respectively (Nair et al., 2004). Northeastern states in India
have reported a very high prevalence of aero-digestive tract cancers when compared with other regions of India (Bhattacharjee et al., 2006; Phukan et al., 2004). Prevalence of oral cancer is highest in the Kamrup district of Assam (ICMR-Report, 2006). In this region there is widespread chewing habit of tobacco with peculiarly fermented betel nut (Phukan et al., 2001). Betel nut contains arecoline, which can produce 3-methyl nitrosamine propionitrile, a potent carcinogen, and safrole-like DNA adduct that have been shown to be genotoxic and mutagenic. In addition, when betel nut is fermented, it is contaminated by fungi, leading to production of carcinogenic aflatoxins, which further add to the risk for these cancers in this region (Chattopadhyay et al., 2007).

The cell of origin of OSCC is the oral keratinocyte. OSCC, as any other cancers, is caused by DNA mutation, often spontaneous but increased by exposure to any of a range of mutagens – chemical, physical or microbial (Scully and Bagan, 2009). The various changes in the DNA can progress from a normal keratinocyte to a pre-malignant or a potentially malignant keratinocyte that is characterized by an ability to proliferate in a less-controlled fashion than normal. The cells become autonomous and a true cancer results, characterized by invasion across the epithelial basement membrane and, ultimately, metastasis to lymph nodes, bone, brain, liver and other sites (Scully and Bagan, 2009). Early indicators of oral cancer are oral leukoplakia and submucous fibrosis with transformation rate of 2-12% to frank malignancies (Anantharaman et al., 2007). Knowledge of cancer genetics is rapidly improving our understanding of cancer biology, helping to identify at-risk individuals, furthering the ability to characterize malignancies, establishing treatment tailored to the molecular fingerprint of the disease, and leading to the development of new therapeutic modalities. Given the poor prognosis associated with oral leukoplakia, and the difficulties in distinguishing it from cancer lesions, there is an urgent need to elucidate the molecular determinants and critical signal pathways underlying the malignant transformation of precancerous to cancerous tissue, and thus to identify novel diagnostic and therapeutic targets.

The etiology of oral cancer is multi-factorial with genetic, environmental, medical, and lifestyle factors interacting to produce malignant stage. The lifestyle behaviors associated to oral cancer with convincing evidence are tobacco use, betel quid chewing, alcohol drinking, low fruit and vegetable consumption. These factors act separately or synergistically. Worldwide, 20–30% of oral cancer cases are attributable to tobacco usage (Smoking and/or Chewing) (Balaram et al., 2002; Hashibe et al., 2007; Rahman et al., 2005). 50% of men and
Introduction

Almost 90% of women cases are attributed to frequent betel quid chewing with or without tobacco consumption in areas where chewing prevalence is high (Balaram et al., 2002). 7–19% cases of total cases of the world are attributed to heavy alcohol drinking (Hashibe et al., 2007; Room et al., 2005), 10–15% cases attribute to micronutrient deficiency (WHO-FAO-Report, 2007; WHO-Report, 2002), and about 3% cases are attributed to human papillomavirus (HPV) infection (Parkin and Bray, 2006).

Carcinogenicity is dose-dependent and magnified by multiple exposures to risk factors. Conversely, low and single exposures of these factors do not significantly increase oral cancer risk. These lifestyle related behaviors have common characteristics: (i). They are widespread worldwide: one billion men, 250 million women smoke cigarettes, 600–1200 million people chew betel quid, two billion consume alcohol, unbalanced diet is common amongst developed and developing countries; (ii). They were already used by human forerunners millions of years ago because they were helpful in overcoming conditions such as cold, hunger and famine; (iii). Their use was seasonal and limited by low availability, in contrast with the pattern of consumption of the modern era, characterized by routine, heavy usage, for recreational activities and with multiple exposures; (iv). Their use in small doses is not disadvantageous for the human body and activates the dopaminergic reward system of the brain, thus giving instant pleasure, ‘liking’ (overconsumption) and ‘wanting’ (craving). For these reasons, effective public health measures aimed at preventing oral cancer and other lifestyle-related conditions fail to realize their final goal to eradicate these lifestyles (Petti, 2009).

Worldwide, smoking accounts for 42% of deaths from cancers of the oral cavity (including the pharynx) and heavy alcohol consumption for 16% of the deaths; the corresponding percentages in high income countries are about 70% and 30%, respectively (Danaei et al., 2005; Jemal et al., 2011). Smokeless tobacco products and betel quid with or without tobacco are the major risk factors for oral cavity cancer in India, and other neighboring countries (Sri Lanka, Pakistan, Bangladesh and Taiwan) (Jayalekshmi et al., 2009; Wen et al., 2010). The rise in the incidence rate of oral cancer in India may have been in part due to the increased consumption of tobacco, betel quid and alcohol. However not all tobacco and alcohol consumers develop premalignant or frankly malignant diseases of the oral cavity. The quantitative absorption, distribution, metabolism, and excretion of carcinogenic tobacco constituents depend on the activity and efficiency of metabolic and enzymatic detoxification pathways. The enzymatic detoxification process is mainly divided
Introduction into two phases. Phase I involves activation of toxic compounds mainly by oxidation into more reactive intermediate products that are neutralized and conjugated by phase II family of enzymes. Cytochrome P450 (CYP) family of enzymes are the major phase I enzymes which usually converts tobacco constituents into more active intermediate compounds which are further detoxified by phase II family of enzymes such as glutathione-S-transferase (GST), NAD(P)H dehydrogenase quinone 1 (NQO1), and N-acetytransferases (NAT). The resultant water-soluble and less-toxic product can easily be eliminated from the body. The role of genetic factors including single nucleotide polymorphism (SNP) of genes associated with activation and detoxification of toxic compounds is conflicting (Bartsch et al., 1999; Buch et al., 2008; Chen et al., 2001; Duarte et al., 2008; Evans et al., 2004; Gattas et al., 2006; Hahn et al., 2002; Hatagima et al., 2008; Kietthubthew et al., 2001; Losi-Guembarovski et al., 2008; Marques et al., 2006; Zhuo et al., 2009). As the northeast Indian population is exposed to high levels of carcinogenic compounds, it was hypothesized that less efficient detoxification mechanism due to polymorphic variants of genes encoding detoxification enzymes may explain high incidence of oral cancer in this region. Moreover p53 mutations have been associated with reduced repair and enhanced cytotoxicity in cell damaged by benzo (α) pyrene diol epoxide-DNA adducts. In the current study, the association of oral cancer in a high risk region of northeast India was investigated for total of eight polymorphisms present in seven genes CYP1A1 (Msp1 and Nco1), GSTT1, GSTM1, GSTP1, NAT2, NQO1 and codon 72 polymorphism of TP53 gene.

OSCC typically evolves from normal epithelium through dysplasia, carcinoma in situ finally to the invasive carcinoma stage. During this tumorigenesis, cumulative genetic alterations including microsatellite instability (MSI) and loss of heterozygocity (LOH) occur. Loss of heterozygosity in a cell is the loss of normal function of one allele of a gene in which the other allele was already inactivated. Microsatellite instability (MSI) is a condition manifested by damaged DNA due to defects in the normal DNA repair process. Sections of DNA called microsatellites, which consist of a sequence of repeating units of 1-6 base pairs in length, become unstable and can shorten or lengthen. LOH which represents the suppressor phenotype, seem to be more common than MSI in OSCC. Although both types of microsatellite alterations have been correlated with clinicopathological features of OSCC, only LOH seems to have prognostic value. The predictive value of both MSI and LOH towards radiotherapy and chemotherapy is debatable. Biggest challenges however remain in the methodological problems connected with these types of investigations (De-Schutter et al.,
Introduction

In a well planned study by De-Schutter et al who used automatic fragment analysis as the preferred technique to assess MSI and LOH, in which they used large panel of microsatellite markers and compared their sensitivity, with strict cutoff values for LOH detection. That study resulted in a very low (around 1%) percentage of MSI, suggesting that indeed the prevalence of MSI in OSCC (or Head and Neck SCC) has been overestimated in literature, partly due to the use of non-optimal techniques. Based on this experience, a role for MSI as prognostic or predictive marker in this tumor type seems highly unlikely (De-Schutter et al., 2007). On the other hand, De-Schutter et al. suggested the clinical importance of the detection of LOH with the use of microsatellite markers. As LOH at certain loci may be indicative for the loss of a tumor suppressor gene (TSG), therapeutic options would mainly be directed towards re-expression of the involved gene, which is the goal of several gene therapy trials. However, re-expression therapies are mainly experimental and still face a lot of difficulties (De-Schutter et al., 2007). For detection of genetic alterations, sequencing has historically been considered as the golden standard because it reveals the exact location and the type of mutation. DNA sequencing represents a single method to forecast a broad range of biological events. However this method was slow and tedious. Next generation sequencing can provide better insights for possible therapeutic options as it can provide broad range of genetic aberrations including mutations at nucleotide level (such as SNPs, insertions or deletions) involved in carcinogenesis. This method can also provide possible functional/structural changes in resulted protein using appropriate softwares in cases of known aberrations and provide a basis for further analysis in cases of finding novel genetic alterations.

It is evident that oral cancer is a multi-factorial disease influenced by both aetiological factors and ethnicity; as a result molecular profiles of oral cancer vary throughout the world (Ambatipudi et al., 2011; Paterson et al., 1996; Shah and Singh, 2006). High throughput methods such as cDNA and oligonucleotide microarrays are increasingly being used to systematically compare molecular features of individual cancers to key clinical parameters. Previous studies have documented the importance of genetic alteration affecting known oncogene and tumour suppressor genes in the development of oral cancer (Bettendorf et al., 2004; Tsantoulis et al., 2007). Genome wide screening can be either quantitative as in the case of gene expression profiling by microarray or qualitative as in the case of mutation detection by next generation sequencing.
Altered gene expression profiling may be attributed to several types of genetic alterations such as base changes, insertions, deletions, chromosomal translocations, inversions, loss of heterozygosity and copy number variations. To understand genetic alterations in large number of genes, precise, fast and cost-effective methods have been developed to analyze several genetic alterations in a single experiment. In recent years there has been a revolution in sequencing methods in the form of next generation sequencing (NGS) technologies capable of producing millions of DNA sequence reads in a single run. This is rapidly changing the landscape of cancer research with the potential to accelerated analysis of genomes, transcriptomes and interactomes which become inexpensive, routine and widespread (Shendure and Ji, 2008). NGS has enabled whole-genome analysis with essentially unlimited resolution (Stankiewicz and Lupski, 2010). Using next generation sequencing clinically relevant molecular signatures can be discovered which may be of great significance in understanding the underlying biology of cancer, providing the ability to answer questions with unimaginable speed which may lead us towards identification of molecular targets in carcinogenesis and subsequent development of innovative methods for diagnosis, prognosis and therapy of cancer.

The discovery of mutations (hallmarks of cancer) that are involved in carcinogenesis is a crucial step in cancer research which is tremendously facilitated by NGS technology useful for both targeted and genome-wide screening. For the present study targeted re-sequencing of 169 functionally relevant and potentially important genes was done to analyze mutational changes using Illumina’s high throughput solexa sequencing technology. Our aim was to identify specific molecular signatures involved in oral carcinogenesis as well as to identify potential biomarkers for oral cancer predisposition, progression and therapeutic manipulation.

As the deregulated expression of genes lies at the origin of tumors, its measurements using microarray technology can be very helpful to model or predict the clinical behavior of malignancies. Many studies have shown that cancer diagnosis based on microarray data can effectively integrate the fundamental processes underlying carcinogenesis into the clinical decision-making process (Wong and Wang, 2008). Microarray technology has made it possible to examine the expression of many genes over multiple developmental stages or different experimental conditions (Guo et al., 2007). Though expression microarrays are powerful and increasingly more widely used investigative, diagnostic, and prognostic molecular biological tools, there are technical aspects to using expression microarrays that
can produce results erroneously representing either suppression or over-expression of specific genes. For example, false negativity can result from low expression levels, transcript drop-out attributable to inefficient priming of specific mRNA(s), poor adhesion of DNA to the slide, and splice variants with sequences not included on the array. Conversely, sources of false positivity include repetitive nucleotide elements, poly (A) tails, and sequence homology between functionally different transcripts, an inappropriately chosen reference standard, and high background levels due to nonspecific binding of nucleotides to the microarray slides. However, since these sources of error remain a potential source of confounding data, confirmation of expression microarray results before proceeding to undertake more elaborate, gene-specific experiments based on array results is important (True and Feng, 2005). In the current study differential gene expression analysis was done on tumour tissue of oral cancer patients and matched normal tissue distant from the tumour site. Significantly deregulated genes were selected from microarray data and validated by real-time RT PCR. This is the first study to provide gene expression profiling and genetic variations of oral cancer associated with betelquid and tobacco consumption in a high-risk region of India.
Chapter 2

AIMS AND OBJECTIVES
Aims and Objectives

Aims:

A. To identify the role of polymorphism of major genes coding for xenobiotic metabolizing phase I and phase II enzymes in oral cancer susceptibility

B. To identify genetic variations (SNPs, Insertions and Deletions) in tumor samples of patients with oral cancer

C. To study differential gene expression profiles of oral squamous cell carcinoma using microarray technology

D. To validate the microarray data for specific genes of interest by using Real Time PCR

To achieve the above mentioned aims, the present study was undertaken in patients with oral cancer from a high risk region of India with the following specific objectives:

Objective – 1

To identify the role of polymorphisms in $CYP1A1*2A$ and $CYP1A1*2C$ genes encoding for phase I detoxification enzymes that are involved in the first step of detoxification pathways.

Objective - 2

To identify the role of polymorphism of $GSTT1$, $GSTM1$, $GSTP1$, $NQO1$ and $NAT2$ genes encoding for phase II detoxification enzymes and polymorphism in the codon 72 of $p53$ gene. Phase II enzymes are involved in second step of detoxification pathways by neutralizing the activated intermediate metabolic products of various carcinogens.

Objective - 3

To identify genetic instability (SNPs, Insertions and Deletions) in OSCC cases by next generation targeted re-sequencing using Illumina-Solexa platform and sequencing by synthesis approach.
Objective - 4

To study differential gene expression profiles of OSCC as compared to paired matched control by cDNA microarray.

Objective - 5

To validate the specific genes of interest from the above microarray data by using quantitative real time RT PCR.
Chapter 3

REVIEW OF LITERATURE
Cancer

Cancer is not just one disease, but a generic term used to encompass a group of more than two hundred diseases in which cells in a part of the body begin to grow out of control. Cancer cells can also invade (grow into) other tissues, something that normal cells cannot do. Growing out of control and invading other tissues are what makes a cell a cancer cell. DNA, the hereditary material is in every cell and directs and controls all its actions. In a normal cell, when DNA gets damaged the cell either repairs the damage or the cell dies. In cancer cells, the damaged DNA is not repaired, but the cell does not die like it should. Cancer cells often travel to other parts of the body, where they begin to grow and form new tumors that replace normal tissue (metastasis).

Cancer is a leading cause of death worldwide. The global burden of cancer continues to increase largely because of aging, growth of the world population and an increasing adoption of cancer-causing behaviors, particularly tobacco consumption, in economically developing countries. Based on the GLOBOCAN 2008 estimates, about 12.7 million cancer cases and 7.6 million cancer deaths (around 13% of all deaths) are estimated to have occurred in 2008; of these, 56% of the cases and 64% of the deaths occurred in the economically developing world (Jemal et al., 2011). Deaths from cancer worldwide are projected to continue rising, with an estimated 12 million deaths in 2030 (WHO-Report, 2009). While incidence rates for all cancers combined in economically developed countries are nearly twice as high as in economically developing countries in both males and females, mortality rates for all cancers combined in developed countries are 21% higher in males and only 2% higher in females. Such disparities in incidence and mortality patterns between developed and developing countries reflects, for a given cancer, regional differences in the prevalence and distribution of the major risk factors, detection practices, and/or the availability and use of treatment services (Jemal et al., 2011).

Oral cancer

Oral squamous cell carcinomas (OSCCs), amount to more than 90% of malignant tumors of the oral cavity and constitutes the most malignant tumors of the head and neck
Revie\$ of Literature

(Diez-Perez et al., 2011; Tsantoulis et al., 2007). This aggressive epithelial neoplasm is associated with severe morbidity and <50% five year survival, despite advances in surgical treatments and in radio- and chemo-therapy (Diez-Perez et al., 2011). The cell of origin of OSCC is the oral keratinocyte. OSCC, as any other cancers, is caused by DNA mutation, often spontaneous but increased by exposure to any of a range of mutagens - chemical, physical or microbial. The various changes in the DNA can progress from a normal keratinocyte to a pre-malignant or a potentially malignant keratinocyte that is characterized by an ability to proliferate in a less controlled fashion than normal. The cells become autonomous and a true cancer results, characterized by invasion across the epithelial basement membrane and, ultimately, metastasis to lymph nodes, bone, brain, liver and other sites. As in other parts of the upper aero-digestive tract, there is strong and synergistic association of oral cancer with tobacco consumption and alcohol abuse. In some regions of the world particularly the Indian subcontinent, oral cancer is among the most frequent malignancies, largely due to tobacco consumption habits (Leon-Barnes, 2005).

Histopathology:

Oral cancer is an invasive epithelial neoplasm which can be classified as spindle-cell carcinoma, papillary squamous cell carcinomas, adenosquamous carcinoma, acantholytic squamous cell carcinomas and carcinoma cuniculatum, with more than 90% of malignant lesions being SCC (Johnson et al., 1993; Warnakulasuriya, 2009). Squamous differentiation, seen as keratinisation with ‘pearl’ formation and an invasive growth is a prerequisite for the diagnosis of squamous cell carcinoma. Invasion is seen as a breach of the basement membrane and extension into the underlying stroma. The degree of keratinisation, cellular and nuclear pleomorphism and mitotic activity provide a basis for the grading of OSCC. These tumours are classified as grade 1 which is well differentiated, grade 2 being moderately differentiated and poorly differentiated carcinomas termed as grade 3 (Pindborg et al., 1985). The well-differentiated squamous cell carcinoma resembles the normal squamous epithelium with extensive keratinisation and pearl formation. Moderately-differentiated squamous cell carcinomas show less keratinisation and more nuclear pleomorphism and mitotic activity. The poorly-differentiated ones show markedly pleomorphic cells with minimal keratinisation and prominent mitotic activity with several atypical mitotic figures. Most squamous cell carcinomas are moderately differentiated. Most cases of squamous cell carcinoma present no diagnostic problems; however, sometimes a pseudo-epitheliomatous hyperplasia overlying a granular cell tumour in necrotising
sialometaplasia and in papillary hyperplasia of palate may cause dilemmas to the histopathologist (WHO-IARC-Report, 2011).

GLOBAL DISTRIBUTION AND INDIAN SCENARIO OF ORAL CANCER:

Global Distribution:

Oral cancer is the 6th most common cancer worldwide, accounting for 4% of cancers in men and 2% of cancers in women (Napier and Speight, 2008), which is a serious and growing health problem in many parts of the world. Cancers of the oral cavity accounted for 3,00,000 cases, with almost two-thirds of these cases occurring in South and South East Asia (Nair et al., 2004; Warnakulasuriya, 2009). There is up to a 20-fold geographical variation in the incidence rates of oral cancer (Warnakulasuriya, 2009). The areas with high incidence rates for oral cancer are found in the South and South East Asia (e.g. India, Sri Lanka, Pakistan and Taiwan), parts of western (e.g. France) and Eastern Europe (e.g. Hungary, Slovakia and Slovenia), parts of Latin America and the Caribbean (e.g. Brazil, Uruguay and Puerto Rico) and in Pacific regions (e.g. Papua New Guinea and Melanesia) (Warnakulasuriya, 2009).

An estimated 2,63,900 new cases and 1,28,000 deaths from oral cavity cancer (including lip cancer) occurred in 2008 worldwide. Generally, the highest oral cavity cancer rates are found in Melanesia, South-Central Asia, and Central and Eastern Europe and the lowest in Africa, Central America, and Eastern Asia for both males and females (Figures 3.1 and 3.2) (Jemal et al., 2011; Warnakulasuriya, 2009).
Fig 3.1: Incidence and mortality of oral cancer. Regions with red colour indicate countries with high incidence and mortality while those with green colour indicate regions with low incidence and mortality of oral cancer (Warnakulasuriya, 2009).
Indian Scenario:

India has one of the highest rates of oral cancer in the world, caused by tobacco consumption. In India, Oral cancer is the leading cancer site in men and is ranks as the
number three cancer in females with 75,000 to 80,000 new cases of oral cancer each year (comprises 25-30% of total malignancies in India). Age-standardized incidence rates per 100,000 population in India were estimated to be 12.8 in men and 7.5 in women (Ferlay et al., 2010; Nair et al., 2004; Parkin et al., 2010). High prevalence of oral cancer in India is attributed to the influence of region-specific epidemiological factors, especially tobacco and betel quid chewing (Tsantoulis et al., 2007). Tobacco use is responsible for more than 90% percent of oral cancers in men and 60% among women and is responsible for 90% of oral cancer deaths in males (Cinciripini and McClure, 1998; Surgeon-General’s-Report, 1989). All forms of tobacco - cigarettes, pipes, cigars, and smokeless tobacco - have been implicated in the development of oral cancers (NCI-NIH-Report, 2011).

In Northeast India, cancer incidences and causative factors responsible for them were not well documented till few years ago. First report of National Cancer Registry Programme (NCRP) of ICMR for northeastern region of country which emphasized on cancer incidence and patterns of cancer in six population based cancer registries (PBCRs) of the Northeastern region (for the two year period 1 Jan 2003 to 31 December 2004) revealed very high incidence of tobacco associated cancers in this region as compared to other parts of the country. In the older established registries (other regions of India) age adjusted incidence rate (AAR) for all anatomical sites has been around 100 per 100,000. The results of this report by NCRP for Northeastern states were notable, in the sense that incidence rates of well over 100 per 100,000 persons have been recorded in five of the eight registry areas identified for describing the incidence and patterns of cancer. Highest AAR for cancer incidence was reported from Aizawl district (AAR: 277.2 in males and 231.5 in females) of Mizoram state followed by Kamrup urban district (AAR: 177.2 in males and 154.1 in females) of Assam state of Northeast India (ICMR-Report, 2006). This region has reported a very high prevalence of aero-digestive tract cancers compared with other regions of India (Bhattacharjee et al., 2006; Phukan et al., 2004). Prevalence of oral cancer is highest in the Kamrup district of Assam (ICMR-Report, 2006). The present study is based on the samples collected from three centres [Guwahati (Assam); Gangtok (Sikkim) and Aizawl (Mizoram)] of Northeast India.

**Age and sex distribution:**

In most countries around the world, oral cancer is more common in men than in women. The reported sex differences are attributable to heavier indulgence in risk habits by
men and exposure to sunlight (for lip cancer) as a part of outdoor occupations. The ratio of males to females diagnosed with oral cancer, however, has declined over the decades and is now about 1.5:1 (Warnakulasuriya, 2009). The risk of developing oral cancer increases with age and the majority of cases occur in people aged 40 or over. In India, peak occurrence of oral cancers for men is in the 50-59 years age group while for women it is in the 60-69 years age group (Sanghvi and Krishnamurthy, 1986).

**Anatomic sites:**

Tongue is the most common site for intraoral cancer among European and the US populations, amounting to 40–50% of oral cancers. However, in parts of the world where tobacco or betel quid chewing is prominent, cancers of the buccal mucosa and retromolar trigone are common (Noonan and Kabani, 2005). In India, buccal mucosa is one of the most commonly affected sites probably due to continuous carcinogenic exposure with tobacco and betel quid chewing, a popular habit in India. Other intraoral sites for mouth cancer include tongue, floor of mouth, gingivae and palate (Warnakulasuriya, 2009).

**Survival and Mortality:**

Despite recent advancement in the treatment, imaging and diagnosis of oral carcinoma, a 5-years survival and mortality rate for this cancer is still at 50%. The survival of oral cancer patients remains very low, mainly due to the fact that it is often revealed when it has metastasized to another location, most likely the lymph nodes of the neck (Tanaka and Ishigamori, 2011). Prognosis at this stage of discovery is significantly worse than when it is caught in a localized intraoral area. Besides the metastasis, at these later stages, the primary tumor has had time to invade deep into local structures. About half of the patients affected will die within five years of diagnosis, while surviving patients may be left with severe aesthetic and functional compromise. One of the most striking features of oral cancer is the disfiguration of the face, which adds further importance for its study and treatment. Oral cavity cancer mortality rates among males decreased significantly in most countries, including those of Europe and Asia, over the past decades (Garavello et al., 2010). But rates continued to increase in several Eastern European countries, including Hungary and Slovakia (Garavello et al., 2010). The increase in females in most European countries largely reflects the ongoing tobacco epidemic (Garavello et al., 2010). This contrasts with the decreasing trends at all ages in both males and females in the United States and United Kingdom, where the tobacco epidemic began and declined earlier (DeLancey et al., 2008;
Edwards et al., 2010; Garavello et al., 2010). However, incidence rates for oral cancer sites related to HPV infections, such as the oropharynx, tonsil, and base of the tongue, are increasing in young adults in the United States and in some countries in Europe (Chaturvedi et al., 2008; Conway et al., 2006; Shiboski et al., 2005) which is hypothesized to be in part due to changes in oral sexual behavior (D'Souza et al., 2009; Marur et al., 2010). In Indian continent high incidence and mortality of oral cancer is mainly attributed to high prevalence of tobacco consumption specially tobacco chewing.

Fig 3.3: India is one of the highest affected countries with oral cancer. Estimated age standardized mortality rate per 100,000 is denoted by different colours, red colour indicates regions with highest mortality rates (Globocan, 2008).

**RISK FACTORS OF ORAL CANCER:**

The etiology of oral cancer is multi-factorial which is influenced by age, sex, race, local environmental factors, diet and genetics. These factors interact to produce a given malignancy. Consequently, the incidence of cancer and cancer types vary depending on these variable factors (Davis, 2011). For instance in India oral cancer is the most common cancer among men and ranks third among women. Knowledge of cancer genetics is rapidly improving our understanding of cancer biology, helping to identify at-risk individuals, further enhancing the ability to characterize malignancies, establishing treatment tailored to the
molecular fingerprint of the disease, and leading to the development of new therapeutic modalities. As a consequence, this expanding knowledge base has implications for all aspects of cancer management, including prevention, screening, and treatment.

Fig 3.4: Postulated causative factors and mechanisms implicated in the induction of oral squamous cell carcinoma due to Consumption of tobacco, betel quid, pan masala and gutkha (Nair et al., 2004).

Most oral cancer cases and deaths are due to both individual predisposition, linked to specific genetic characteristics, and exposure to carcinogens, caused by lifestyle behaviors. Specifically, 20–30% overall cases are attributable to tobacco/bidi smoking (Balaram et al., 2002; Hashibe et al., 2007; Rahman et al., 2005), 50% of men and almost 90% of women cases are attributed to frequent betel quid without tobacco chewing in areas where chewing prevalence is particularly high (Balaram et al., 2002), 7–19% cases to heavy alcohol drinking (Hashibe et al., 2007; Room et al., 2005), 10–15% cases to micronutrient deficiency (WHO-FAO-Report, 2007; WHO-Report, 2002), and also 3% cases to human papillomavirus (HPV) infection, generally (but not exclusively) associated to sexual behaviour (Parkin and Bray, 2006). In addition, exposure to two or more of these factors has a synergistic effect in increasing oral cancer risk (Applebaum et al., 2007; Boccia et al., 2008; Subapriya et al., 2007). Tobacco consumption either smoking or chewing, betelquid consumption, alcohol use and HPV infections are the major risk factors for oral cavity cancer, with smoking and
alcohol having synergistic effects (Blot et al., 1988; Hashibe et al., 2009). The contribution of each of these risk factors to the burden varies across regions (D'Souza et al., 2009; IARC-Monographs, 2004; IARC-Monographs, 2007; Surgeon-General's-Report, 2004).

**Fig 3.5:** Risk factors for oral cancer development. Life style related risk factors are the major cause for oral cancer (Scully and Bagan, 2009).

**Tobacco:**

India is the second largest producer and consumer of tobacco in the world after China (Kuruvilla, 2008). Polycyclic aromatic hydrocarbons (PAHs), nitrosamines, aldehydes and ketones are the major carcinogens present in tobacco (Hecht, 2003). However the concentrations of these compounds vary depending upon the nature of tobacco use. Smokeless tobacco is rich in nitrosamines while due to high temperatures at the burning tip, tobacco smoke contains pyrolysis products (Anantharaman et al., 2007; Hecht, 2003). These compounds are capable of generating reactive oxygen species (ROS) by direct reaction or metabolic activation (Yin et al., 2001).

**Tobacco Chewing:**

Smokeless tobacco products and betel quid with or without tobacco are the major risk factors for oral cavity cancer in India, and other neighboring countries (Sri Lanka, Pakistan, Bangladesh and Taiwan) (IARC-Report, 2004; Jayalekshmi et al., 2009; Wen et al., 2010). However high frequency in Indian subcontinent is mainly attributed to tobacco chewing (IARC-Report, 2010). Chewing tobacco attributes for more than 66% of the total oral cancer
cases in India. The surveys carried out by National Family Health Survey (NFHS) and Global Youth Tobacco Survey (GYTS) conducted in 18 states in India have reported an alarmingly high prevalence of chewable tobacco products in younger Indian population (average prevalence of tobacco chewing was 29% compared to 13% for smoking). The highest prevalence of tobacco chewing within India was observed in Northeastern states ranging from 47-63%. This is one of the major contrasting features of Indian population as depicted in figure 3.6 compared with western countries where tobacco smoking is more prevalent compared to chewing (Kuruvilla, 2008).

**Tobacco smoking**

Worldwide, smoking accounts for 42% of deaths from cancers of the oral cavity (including the pharynx). The corresponding percentage in high-income countries is about 70% (Danaei et al., 2005). When tobacco smoke is inhaled, 25% of the nicotine reaches the brain in about seven seconds. Brain levels of nicotine fall rapidly and the tobacco user experiences craving for a further tobacco intake within 30 min. Nicotine functions by binding to nicotinic acetylcholine receptors, causing increased heart rate, vasoconstriction, and alertness (WHO-Report, 2008). Tobacco consumption habits vary within India also with people from Assam state having high tobacco chewing habit compared to smoking (Rani et al., 2003) (Fig 3.6).

---

**Fig 3.6:** State wise prevalence of tobacco consumption in India. Most of the Northeastern states have high prevalence of tobacco chewing habits (Rani et al., 2003).
Chewing habits of Arecanut, Betelquid, Pan Masala and Gutkha:

Areca nut has been declared a known human carcinogen by an IARC expert group. It has been estimated that, worldwide, around 600,000,000 people chew areca nut (Nelson and Heischober, 1999). A causal association between tobacco and betel quid (BQ) chewing habits and oral mucosal diseases such as leukoplakia, oral submucous fibrosis and oral cancer has been established and heavy users have a significantly increased mortality rate.

The BQ is a mixture of areca nut (Areca catechu), catechu (Acacia catechu) and slaked lime (calcium oxide and calcium hydroxide) wrapped in a betel leaf (Piper betle). Condiments, sweetening agents and spices may be added according to individual preferences. In India, most habitual chewers of BQ add tobacco. In some countries, such as Papua New Guinea and China, tobacco is not added (Nair et al., 2004). BQ chewing has been related mainly to oral, pharyngeal and oesophageal cancer (IARC-Report, 2004).

Pan masala is basically a preparation of areca nut, catechu, cardamon, lime and a number of natural and artificial perfuming and flavouring materials. Gutkha is a variant of pan masala, in which in addition to these ingredients flavoured chewing tobacco is added. Both products are often sweetened to enhance the taste. Aggressively advertised and marketed, often claimed to be safer products, they are consumed by the very young and old alike, particularly in India, but also among migrant populations from these areas worldwide. These products have been strongly implicated in the recent increase in the incidence of oral submucous fibrosis, especially in the very young, even after a short period of use. This precancerous lesion, which has a high rate of malignant transformation, is extremely debilitating and has no known cure.

Alcohol consumption:

Regular alcohol consumption is associated with an increased risk for oral cancer. Such association is dose-dependent. Indeed, among individuals consuming 4–5 drinks daily, the risk for cancer of the oral cavity is 2-3 fold higher than among non-drinkers (Room et al., 2005; Seitz and Stickel, 2007). Overall, 7-19% oral cancer cases are attributable to heavy alcohol drinking (Hashibe et al., 2007; Room et al., 2005). Worldwide, heavy alcohol consumption accounts for 16% of the deaths; the corresponding percentage in high-income countries is about 30% (Danaei et al., 2005). According to the World Health Organization estimates, there are almost two billion people worldwide who consume alcohol and almost 80
Million with diagnosable alcohol abuse disorders. As for the various groups of drinks, beers account for 37%, spirits/liqueurs for 33% and wines for 30% of overall alcohol consumption.

Alcohol and tobacco which work synergistically account for 75% of disease burden of oral malignancies in Europe, the Americas and Japan. For the highest levels of consumption to the lowest ones, relative risks from 70 to over 100 have been shown. Most of the rise in western countries in recent years has been attributed to rising alcohol consumption in Northern Europe and rises in tobacco consumption in parts of Southern Europe. Significant risk increases have also been reported amongst nondrinking smokers and to a lesser extent, nonsmoking heavy drinkers. Studies that have attempted to estimate a difference between wine, beer and hard liquors generally indicate that heavy consumption of all types of alcoholic beverage confers risk, the differences in risk estimates being largely due to sociocultural correlates of drinking patterns in various populations (Leon-Barnes, 2005).

Oral cancer risk among alcohol drinkers further increases for tobacco chewing/smoking and/or betel quid chewing. However, shadows of uncertainty on carcinogenicity of alcoholic drinks originate from data on ‘never tobacco users’ showing no significant increase in oral cancer risk among alcohol drinkers, irrespective of drinking, duration and frequency (Hashibe et al., 2007). Acetaldehyde is responsible for the oral carcinogenic effect of ethanol, owing to its multiple mutagenic effects on DNA. In addition, ethanol is not the only carcinogen present in alcoholic drinks, other minor components, such as nitrosamines, acrylamide, oxidized polyphenols are classified as probable carcinogenic to humans, with animal experiments showing mutagenic activity on oral epithelial cells (Jagerstad and Skog, 2005; Petti, 2009).

**Human papillomavirus (HPV):**

HPVs, especially those genotypes of known high oncogenic potential in uterine cervix and skin such as HPV 16 and 18, are found in a variable but small proportion of oral cancers. Recent studies have suggested that HPV may be responsible for a small fraction of oral cancers (Leon-Barnes, 2005).

**Treatment and main complications of oral squamous cell carcinoma:**

It is generally accepted that prognosis is best in early carcinomas, especially those that are well-differentiated and not metastasised: unfortunately, most OSCC are diagnosed at a late stage of disease. Oral squamous cell carcinoma treatment is still often surgery, with
radiotherapy, and chemotherapy which are influenced by a number of other factors, especially the balance between positive outcomes and adverse effects (Fig. 3.7). According to TNM classification of the International Union against Cancer (UICC) tumour size (T), nodal metastases (N) and distant metastases (M) relates well to overall survival rate i.e. the earlier the tumour, the better the prognosis and the less complicated is the treatment. Surgery remains the most well established mode of initial definitive treatment for a majority of patients with OSCC (Shah and Gil, 2009). The particular surgical approach is influenced by site, location, size, depth of infiltration and proximity to bone. Metastases from OSCC, when present, will occur in cervical lymph nodes in almost 80% of patients, and cervical lymphadenectomy has played an important role in the management of OSCC for over a century. Radiotherapy plays a key role in the management of early-stage and locally advanced SCC either alone or, more frequently combined with surgery and/or chemotherapy (Mazeron et al., 2009; Scully and Bagan, 2009). Attempts to improve the efficacy of radiotherapy, whilst maintaining acceptable toxicities, include altered fractionated radiotherapy or concomitant chemo-radiotherapy (CT-RT). Potential imaging techniques to detect residual and recurrent locoregional disease after CT-RT are serial CT (computerised tomography) or MRI (magnetic resonance imaging) and FDG-PET (positron emission tomography [PET] scanning with the tracer fluorine-18, fluorodeoxyglucose [FDG]), and possibly diffusion MRI and PET/CT.

Fig 3.7: Treatment of oral squamous cell carcinoma, and main complications associated with therapy. Complications associated with chemotherapy and radiotherapy is their non-invasiveness with side effects such as mouth ulcers and xerostomia while Surgery has limitation to have its local effects only with cosmetic or functional defects (Scully and Bagan, 2009).
GENES AND GENETICS OF ORAL CANCER:

The development of cancer is a result of the accumulation of genetic and epigenetic changes within cells. These genotypic alterations can influence hundreds of genes, leading to phenotypic changes in critical cellular functions such as resistance to cell death, increased proliferation, induction of angiogenesis, and the ability to invade and metastasize. The mechanisms underlying these genetic and epigenetic aberrations include, but are not limited to, genomic instability through chromosomal rearrangement, amplification, deletion, methylation, and mutation. These genetic alterations have been shown to contribute directly to cancer development and progression and are central to understanding the biology of oncogenes and tumor suppressor genes (TSG) as well as the phenotypes they regulate.

Loss of heterozygosity (LOH) and microsatellite instability or allelic imbalance (AI):

Loss of heterozygosity (LOH) and allelic imbalance (AI), have been relevant targets in cancer research. AI may occur when one copy of a polymorphic marker (with two slightly different alleles) is lost (LOH) or amplified (allelic gain). The term LOH is commonly used to describe this process, but as allelic gain occurs very frequently, and may be more common, AI describes the process more accurately. AI occurs at loci across the genome at low frequency and at higher frequency at 3p (3p24-25, 3p21, 3p13-14), 9p21 (p16), 17p13 (p53) and 8p22-24, with loci at 13q14, 18q and 21q being implicated in some studies. The markers utilized in these studies evolved as more informative markers, and those showing higher frequency of AI (indicating the position of relevant genes), were discovered (Califano et al., 1996; Cheng and Wright, 2005; Garnis et al., 2004; Lingen et al., 2011; Mao et al., 1996; Rosin et al., 2000; Roz et al., 1996; Tabor et al., 2003). At present, the identity of the relevant sequences at many of these loci is not known such that regulatory sequences as well as oncogenes or TSG may reside here. With respect to AI and dysplasia, initial studies (prior to 2002) revealed AI at many loci in different chromosomes (Califano et al., 1996; El-Naggar et al., 1998; Jiang et al., 2001; Partridge et al., 2001; Tabor et al., 2001). In general, there is a trend for lesions with more disturbance in cellular organization and architecture to harbor more genetic changes at 3p and 9p (Epstein et al., 2003; Tabor et al., 2001; Tsui et al., 2008; Zhang et al., 2005). However, not all studies confirm this observation and AI at 3p and 9p may not result in any phenotypic change in the oral epithelium that can be detected by light microscopy (Jiang et al., 2001; Kayahara et al., 2001). Additional chromosomes studied for LOH and AI are 8p, 8q, 11p, 17p, and 18q; however, the facts in favour of the predictive
value of these loci are inconclusive (Bremmer et al., 2008; Bremmer et al., 2005; Chen et al., 2005; El-Naggar et al., 1998; Lingen et al., 2011; Tabor et al., 2001). Traditional methods for discovering useful markers for diagnosis and prognosis of cancer are labour intensive and time consuming. In recent years genome wide screening for identification of genetic aberrations using high throughput methods such as next generation sequencing technology have been introduced. These may provide a deeper insight into the genetic changes responsible for carcinogenesis.

**Aneuploidy:**

Chromosomal instability often leads to imbalanced DNA content and the generation of near-diploid or aneuploid clones. Aneuploidy may result from gene dose imbalance, loss of TSG, gain of tumor promoting genes or oncogenes, or formation of fusion genes that leads to increased survival and proliferation advantage. Approximately 50-60% of oral cancers are aneuploid with one study reporting a figure of 90% (Abou-Elhamd and Habib, 2007; Diwakar et al., 2005; Torres-Rendon et al., 2009b). Aneuploidy in OSCC has also been shown to be associated with higher incidence of local recurrence and lymph node metastases (Baretton et al., 1995; Hemmer et al., 1999; Rubio-Bueno et al., 1998). Conversely, another study did not find a correlation between dysplasia and aneuploidy and the development of OSCC (Seoane et al., 1998). Thus there are limited and conflicting data regarding aneuploid dysplasias and the likelihood of progressing to OSCC (Diwakar et al., 2005; Kahn et al., 1994; Klanrit et al., 2007). Thus at the present time there is insufficient data to determine whether aneuploidy can be used as a biomarker for predicting the development of OSCC (Lingen et al., 2011).

**miRNA:**

miRNA are 20-22 nucleotide-long members of the non-coding RNA family, which adds another layer of gene regulation that is altered as cancer develops. They may be present as intergenic transcription units or found in the intronic sequences of protein-coding genes. Functional studies have identified that miRNAs act as conventional tumor suppressors or as oncogenes, and affect the translation or stability of target mRNA. Most are negative regulators of gene expression and have fundamental roles in biologic processes with this function being dysregulated as cancer develops. miR-21, miR-181b, and miR-345 were found to be consistently increased in oral dysplasia and are associated with lesion severity (Cervigne et al., 2009). These regulatory sequences may have therapeutic potential as many of them influence multiple pathways that are dysregulated in cancer. Presently, there is
limited information regarding the expression of miRNAs in oral dysplasia. There are insufficient evidences available to delineate recommendations regarding the clinical utility of miRNA expression and the prediction of whether a dysplastic lesion will progress to OSCC (Lingen et al., 2011).

**Epigenetic events:**

Epigenetic changes involve modifications of DNA and histones that are not coded in the DNA sequence although these changes are heritable (Egger et al., 2004). Three systems are involved: DNA hypermethylation, RNA-associated post-transcriptional silencing, and histone modifications. Of these, DNA methylation has been studied in OSCC. In normal tissues, unmethylated cytosine is found in high densities in CpG islands (areas with high concentration of cytosine and guanine) that map close to a promoter region in 40% of mammalian genes (Egger et al., 2004). This unmethylated state is associated with a high rate of transcriptional activity. Hypermethylation of TSG, mediated through the enzyme DNA methyltransferase, results in stable transcriptional silencing of tumor suppressor activity. This process has been detected in OSCC and is a hallmark of many other cancers as well. In OSCC, hypermethylation of p16 occurs in 50–73% of cases and p15 in 60% of cases (Goldenberg et al., 2004; Kato et al., 2006; Kulkarni and Saranath, 2004; Wong et al., 2003). Hypermethylation (as well as point mutation and deletion) of p16 (locus on 9p21) probably abrogates its activity via the p16/Rb/cyclin D1 tumor suppressor pathway (Goldenberg et al., 2004; Kresty et al., 2002). Some of the other important genes shown to be hypermethylated in OSCC include RARB2, CDH1, MGMT, DAPK1, RARB2 (Ha and Califano, 2006; Kato et al., 2006; Kulkarni and Saranath, 2004; Youssef et al., 2004). However, these studies do not correlate hypermethylated states with recurrence of OSCC or with progression to invasive OSCC. At this time, there is insufficient evidence to determine if hypermethylation can be used as a predictive biomarker for the progression of dysplastic lesions (Lingen et al., 2011).

**Telomerase regulation:**

Telomeres are specialized areas of the distal end of chromosomes composed of chromatin formed by tandem repeats of the sequence TTAGGG bound to specific telomere-binding proteins. They are progressively shortened with each cell division, ultimately resulting in aging and senescence of cells (Pannone et al., 2007). As telomere loss limits lifespan of cells, the loss also reduces the probability of cancer development. Telomerase is
an enzyme that directs the synthesis and maintenance of these telomeres and is composed of hTR (human telomerase RNA, the RNA template), hTEP1 or TP1 (telomerase-associated protein 1) and hTERT (human telomerase reverse transcriptase) (Pannone et al., 2007). Cancer cells are able to stabilize telomeres by activating telomerase, thereby bypassing senescence and facilitating cell immortalization (Shay and Wright, 2010). Depending on the assay utilized, telomerase activity is noted in 67-100% of OSCC (Pannone et al., 2007; Sumida et al., 1998). Enhanced telomerase expression is seen in 50-100% of moderate and severe dysplasia (Chen et al., 2007; Kim et al., 2001; Liao et al., 2000; Miyoshi et al., 1999; Zhang and Zhang, 1999).

Studies showed that OSCC, OSCC margins, and dysplastic lesions have similar expression of telomerase activity (Fujita et al., 2004; Yajima et al., 2004). Liao et al. and Miyoshi et al. showed that telomerase activity increases from 0-50% in mild to 50-100% in moderate-to-severe dysplasia, and was highly expressed in OSCC, suggesting that the acquisition of activity was part of multi-step carcinogenesis (Liao et al., 2000; Miyoshi et al., 1999). One study did not find differences in expression between mild, moderate and severe dysplasia and OSCC (Chen et al., 2007). Mutirangura et al. demonstrated that non-dysplastic leukoplakias that progressed to OSCC were also associated with increased telomerase activity (Mutirangura et al., 1996).

Compared with normal mucosa, telomerase activity is increased in dysplasia and OSCC. The activation of telomerase appears to be a late change during progression, but the frequency of increased telomerase activity varies greatly from study to study. There are no studies that have attempted to correlate telomerase activity and progression with OSCC. At this time, there is insufficient evidence to determine if increased telomerase activity can be used as a predictive biomarker for dysplastic lesions.

**Proliferation markers:**

It is generally accepted that increased cell proliferation is associated with the progression in the multistep process of carcinogenesis. Immunohistochemical methods of detecting proliferation markers, such as proliferating cell nuclear antigen (PCNA), minichromosome-maintenance protein 2 (MCM2) and Ki-67, have been widely used as possible indicators of genetic abnormalities typical of malignant progression. Ki67 antigen is one of the best known proliferation markers as its expression is induced in proliferating cells
(G1, S, G2 phase), but not in resting cells (G0 phase). MCM2 is expressed throughout the cell cycle, including cells leaving G0 to enter into the early G1 phase, distinguishing them from Ki67. PCNA, another marker frequently used as a measure of proliferation, is an essential factor both for replication and for repair of DNA. Dysregulation of Ki67, PCNA, and MCM2 protein expression has been observed in OSCC (Fourati et al., 2009; Iamaroon et al., 2004; Torres-Rendon et al., 2009a; Watanabe et al., 2010; Xie et al., 1999). There is conflicting evidence regarding their possible role as prognostic markers for OSCC (Kodani et al., 2003; Szelachowska et al., 2006; Xie et al., 1999).

**p53 Protein:**

p53 is a tumour suppressor gene located on chromosome 17p13 which plays a major role in cell-cycle progression, cellular differentiation, DNA repair and apoptosis, and is regarded as a guardian of the genome. Loss of p53 function diminishes the regulation of cell cycle arrest and apoptosis, thereby altering the ability of cells to respond to stress or damage (such as DNA damage, hypoxia, and oncogene activation). This can subsequently lead to genomic instability and the accumulation of additional genetic alterations. p53 is the most commonly inactivated TSG in human cancer including OSCC (Vousden and Lane, 2007). Various genetic events can lead to inactivation of p53 including mutation, inactivation through interaction with a viral protein of oncogenic’ HPV subtype, (such as HPV16 or HPV18), or through loss of one allele as a result of LOH (Gonzalez-Moles et al., 2000; Nagpal et al., 2002; Olshan et al., 1997). In normal cells, p53 protein levels are low due to the wild-type protein’s short half life and are essentially undetectable by immunohistochemistry (IHC) (Smeenk and Lohrum, 2010). Stabilizing mutations may cause an increased half-life for the protein, which frequently results in increased expression of mutant p53 in neoplastic cells. Association of p53 with other proteins that protect against degradation has also been shown to be responsible for the over-expression of p53. IHC expression of a mutant p53 protein has been correlated with increased risk for secondary tumors, early recurrence, metastatic spread, and resistance to chemotherapy or radiation therapy (Shin et al., 1996; Temam et al., 2000; Warnakulasuriya et al., 2000). Poeta et al, 2007 reported that inactivation of p53 in OSCC is associated with reduced survival after surgical treatment (Poeta et al., 2007). However, in view of the heterogeneity of laboratory
techniques as well as limited clinical data of various studies, the value of the p53 as a biomarker in patients with OSCC is still controversial.

**Receptor tyrosine kinase pathways:**

Several signal transduction pathways are frequently altered in cancer and share common nodes and interact as a network. Their modification can affect cell survival, cell proliferation, morphology, and angiogenesis. Comprehension of the underlying pathways governing the progression of oral premalignant lesions is thus of utmost importance. A number of growth factors, including platelet-derived growth factor (PDGF), epidermal growth factor (EGF), fibroblast growth factor (FGF), nerve growth factor (NGF), and transforming growth factor-α (TGF-α) family members, signal by inducing dimerization and activation of receptors that are protein tyrosine kinases. Regulation of normal epithelium by growth factors such as TGF-α or EGF is dependent on the expression of the corresponding receptors on the target cell.

Receptor tyrosine kinase pathway (EGFR/TGF-α) plays an important role in cell proliferation, apoptosis, invasion, angiogenesis, and metastasis. Via the tyrosine kinase cascade, the receptor tyrosine kinase (also known as Type I receptor tyrosine kinases or ErbB tyrosine kinase receptors) has many downstream signaling targets that are associated with carcinogenesis. Development, growth, and survival of OSCC are highly dependent upon the EGFR signaling pathway. Increased expression of EGFR and TGF-α is observed in most OSCC, and expression correlates with poor prognosis (Ciardiello and Tortora, 2003).

EGFR signaling also appears to be important at the stage of oral premalignancy. For example, Grandis et al. found increased expression of TGF-α and EGFR mRNA expression in both dysplasias and OSCC (Grandis and Tweardy, 1993). Furthermore, amplification of EGFR in premalignancy has also been described. Specifically, Nagatsuka et al. reported EGFR amplification in epithelial dysplasia and carcinoma in situ, and this amplification increased with the histologic grade of dysplasia (Nagatsuka et al., 2001). Similarly, using IHC, several investigators have reported increased expression of both EGFR and TGF-α in premalignant lesions (Christensen, 1998; Srinivasan and Jewell, 2001). Therefore, coexpression of TGF-α and EGFR may provide an early marker for the onset of epithelial dysplasia preceding OSCC. In a recent study, Taoudi-Benchekroun et al. assessed whether EGFR expression and gene copy number changes might predict the risk of progression of oral
leukoplakia to oral SCC. They reported that increased EGFR gene copy number in oral leukoplakias was associated with an increased risk of developing OSCC (Taoudi-Benchekroun et al., 2010). Increased EGFR and TGF-α are observed in both premalignancy and OSCC. Further studies are required to validate the utility of EGFR as a predictive biomarker.

**Phosphoinositide 3-kinase (PI3K)/AKT pathway:**

Once activated, EGFR stimulates a number of downstream signaling events, namely the Ras/Raf/mitogen activated protein kinase (MAPK) signaling pathway, the transcription factor signal transducer and activator transcription, and the PI3K/AKT pathway, which in turn contributes to the malignant growth, and metastatic potential of oral cancer (Molinolo et al., 2009). PI3K is a lipid kinase that phosphorylates structural components of the cell membrane such as the inositol of phosphatidyl-1D-myoinositol (PI) at the 3-position, and is known to be closely involved in carcinogenesis (Massarelli et al., 2005).

mRNA expression of PI3K class III was reported to be 2.5-11 times greater in dysplastic mucosa and OSCC compared with normal tissue which was further validated by IHC by demonstrating the presence of p-AKT-positive cells only in dysplastic and early cancerous lesions (Watanabe et al., 2009). This finding is supported by Kaur et al., who assessed PI synthase expression by IHC in clinical specimens from oral leukoplakias without dysplasia, with dysplasia (mild, moderate and severe) and OSCCs. They reported increased PI synthase expression to be an early event in oral tumorigenesis, further sustained during the development and progression of OSCC (Kaur et al., 2010). AKT activation has been shown as an early event in oral preneoplastic lesions, and its expression is correlated with poor outcome in oral cancer patients (Massarelli et al., 2005).

**ERK/MAPK pathway**

The extracellular-signal regulated kinases (ERK /MAPKs) pathway is critically involved in the regulation of cell differentiation, proliferation, and survival (Mishima et al., 2002). MAPKs are activated by phosphorylation on two sites within the kinase domain and activated forms phosphorylate serine/threonine residues present on effector kinases. The MAPKs include two mammalian isoforms (ERK1, p44MAPK and ERK2, 42MAPK), which are translocated to the nucleus upon activation by growth factors such as EGF, NGF, and PDGF (Marshall, 1995). Extracellular-signal regulated kinases/mitogen activated protein
kinases play a central role in mitogenic signaling, which is a cascade of phosphorylation reactions involving cell surface receptor, Ras, Raf, and MEK or protein kinase C, Raf, and MEK (Cobb and Goldsmith, 1995; Gutkind, 1998). Activation of ERK1/2-MAPK pathway is often the result of the stimulation of EGFR signaling, with previous studies showing that in OSCC the Ras/RAF/MAPK pathway may be either constitutively activated due to gain in functional mutations in ras genes or may be activated downstream from the persistent autocrine or paracrine stimulation of EGFR and other growth factor receptors, namely FGF (Tsui et al., 2009). Furthermore, there is some evidence that alterations in this pathway may help identify a subset of dysplastic lesions that are more likely to progress to OSCC. Additional studies are required to determine their diagnostic and prognostic utility.

**Cyclin D1 pathway:**

The *CCND1* gene encodes the cyclin D1 protein which is a key regulator of the G1 phase of the cell cycle. Deregulation of the cell cycle is linked to carcinogenesis, specifically, the deregulation of G1 to S phase progression. The transition from G1 into S phase is regulated by CDKs, CDK4 and CDK6, in protein complexes with cyclin D1 (Tsui et al., 2009). Cyclin D1 catalyzes the phosphorylation of Rb, which then releases the transcriptional factor E2F that will activate a number of downstream genes necessary for cell cycle progression. Therefore, overexpression of the protein accelerates the G1 phase transition, whereas inhibition of cyclin D1 results in cell cycle arrest. Overexpression of cyclin D1 is the result of gene rearrangement and gene amplification and is often present in OSCC.

Turatti et al reported that the major components AP-1 transcriptional factors (c-Jun and c-Fos) and cyclin D1 are altered in dysplastic epithelium and OSCC, with cyclin D1 expression increasing with the degree of histologic differentiation from normal to moderate dysplasia and OSCC (Turatti et al., 2005). Ye et al. reported that the *CCND1* P241P polymorphism was significantly associated with a 2.5-fold increased risk of oral premalignant lesion (Ye et al., 2008b). In a case-control study Huang et al. showed that individuals with one or more copies of the *CCND1* G870A variant A-allele had an increased risk of oral premalignant lesion development. These findings support the hypothesis that this polymorphism may be a susceptibility factor for OSCC (Huang et al., 2006). Alterations in the Cyclin D1 pathway are present in both OSCC and dysplasia. However, there is currently insufficient evidence to determine whether these alterations could be used as predictive markers for OSCC.
Vascular endothelial growth factor (VEGF) pathway:

Angiogenesis is an essential phenotype in both physiologic and pathologic settings including tumor formation. The angiogenic phenotype is one of the first recognizable phenotypic changes observed in both experimental models as well as in human OSCC, suggesting that angiogenesis markers may hold promise for diagnosis and prevention (Carlile et al., 2001; Macluskey et al., 2000; Pazouki et al., 1997). The VEGF family is thought to be one of the factors that play a central role in the induction of blood vessel growth. VEGF acts by increasing vessel permeability and enhancing endothelial cell proliferation, migration and differentiation (Tae et al., 2000). The biologic effects of the VEGF ligands are mediated through their binding to members of the VEGF receptor family (VEGFR-1, VEGFR-2, VEGFR-3). Vascular endothelial growth factor expression is increased in both dysplasia and HNSCC (Inoue et al., 1997; Li et al., 2005). With respect to premalignancy, Johnstone et al. reported a significant up-regulation of VEGF during progression from normal oral mucosa to dysplasia and OSCC (Johnstone and Logan, 2007). Conversely, Denhart et al. reported that only 50% of premalignant lesions and 75% of OSCC expressed VEGF (Denhart et al., 1997; Johnstone and Logan, 2007), implying that 50% of the premalignant and 25% of the malignant lesions in this study were inducing angiogenesis via an alternative mechanism that did not seem to involve VEGF. Similarly, Tae et al. found that levels of VEGF in premalignant and malignant oral tissue were lower than in normal tissue (Tae et al., 2000). Finally, Hasina et al. reported that OSCC demonstrate angiogenic heterogeneity that had an impact on targeted anti-angiogenic therapy (Hasina et al., 2008). The current data suggest that there is heterogeneity with respect to the expression of VEGF in both dysplasia and OSCC. These findings suggest that selection of a single angiogenic factor / pathway biomarker may have limited ability to predict which lesions may or may not progress to OSCC.

Role of Carcinogen detoxification pathways in Oral Cancer:

The human body is continuously exposed to a wide array of xenobiotics in one’s lifetime, from food components to environmental toxins to pharmaceuticals, and these xenobiotics may interact deleteriously with an organism, causing toxic and sometimes carcinogenic effects. The ability to survive the threat posed by endogenously produced or environmental xenobiotic compounds probably represents a biological adaptation fundamental to survival (Hayes and McLellan, 1999). Our body has developed complex
enzymatic mechanisms to detoxify these substances. These mechanisms exhibit significant individual variability, and are affected by environment, lifestyle, and genetic influences (Sheehan et al., 2001).

Non-reactive xenobiotic compounds could be biotransformed in two phases: functionalization, which uses oxygen to form a reactive site, and conjugation, which results in addition of a water-soluble group to the reactive site. These two steps, functionalization and conjugation, are termed Phase I and Phase II detoxification, respectively. The result is the biotransformation of a lipophilic compound, not able to be excreted in urine, to a water-soluble compound able to be removed in urine. Therefore, detoxification is not one reaction, but rather a process that involves multiple reactions and multiple players (Fig 3.8).

![Fig 3.8: Phase I and Phase II enzymatic detoxification in human](image)

**Enzyme Systems Involved in Detoxification:**

**The Phase I System:** The Phase I detoxification system, composed mainly of the cytochrome P450 supergene family of enzymes, is generally the first enzymatic defense against foreign compounds. Most pharmaceuticals are metabolized through Phase I biotransformation. In a typical Phase I reaction, a cytochrome P450 enzyme (CYPP450) uses oxygen and, as a cofactor, NADH, to add a reactive group, such as a hydroxyl radical. As a consequence of this step in detoxification, reactive molecules, which may be more toxic than the parent molecule, are produced. If these reactive molecules are not further metabolized by Phase II conjugation, they may cause damage to proteins, RNA, and DNA within the cell. Several studies have shown evidence of associations between induced Phase I and/or
decreased Phase II activities and an increased risk of disease. The major P450 enzymes involved in metabolism of drugs or exogenous toxins are the CYP1A1, CYP1A2, CYP2D6, CYP2C, and the CYP3A4 enzymes.

The Phase II System: Phase II conjugation reactions generally follow Phase I activation, resulting in a xenobiotic that has been transformed into a water-soluble compound that can be excreted through urine or bile. Several types of conjugation reactions are present in the body, including glucuronidation, sulfation, and glutathione and amino acid conjugation. These reactions require cofactors which must be replenished through dietary sources. Much is known about the role of Phase I enzyme systems in metabolism of pharmaceuticals as well as their activation by environmental toxins and specific food components. However, the role of Phase I detoxification in clinical practice has received less consideration. The contribution of the Phase II system has received lesser attention both in academic research circles and in clinical practice. And, little is currently known about the role of the detoxification systems in metabolism of endogenous compounds.

Recently, antiporter activity (p-glycoprotein or multidrug resistance) has been defined as the Phase III detoxification system. Antiporter activity is an important factor in the first pass metabolism of pharmaceuticals and other xenobiotics. The antiporter is an energy-dependent efflux pump, which pumps xenobiotics out of a cell, thereby decreasing the intracellular concentration of xenobiotics. Antiporter activity in the intestine appears to be co-regulated with intestinal Phase I CYP3A4 enzyme. This observation suggests the antiporter may support and promote detoxification. Possibly, its function of pumping nonmetabolized xenobiotics out of the cell and back into the intestinal lumen may allow more opportunities for Phase I activity to metabolize the xenobiotic before it is taken into circulation. Two genes encoding antiporter activity have been described: the multi-drug resistance gene 1 (MDR1) and multi-drug resistance gene 2 (MDR2).

CYP1A1: One of the most important groups of metabolic enzymes involved in the detoxification of a wide range of toxic compounds is the cytochrome P450 (CYP) super family. This superfamily is subdivided into a number of families and subfamilies, based on nucleotide sequence homology where genes within a family have a minimum of 40% sequence identity (Smith et al., 1998). CYP1A1 is a key enzyme in the phase I bioactivation of xenobiotics (Nebert, 1991). It contributes to the aryl hydrocarbon hydroxylase activity, catalyzing the first step in the metabolism of a number of polycyclic aromatic hydrocarbons.
These include tobacco carcinogen, benzo (a) pyrene and several other tobacco related procarcinogens such as nitrosamines and aromatic amines. They are metabolized to their ultimate DNA-binding forms. CYP1A1 gene is expressed in many epithelial tissues especially in buccal mucosa, which is responsible for the in situ activation of tobacco carcinogens (Nair et al., 1999; Sam et al., 2010; Vondracek et al., 2001). The mutation at the 3’ flanking region of the CYP1A1 gene determines three different genotypes, called m1/m1, which is homozygotes for the wild-type allele and does not have the restriction site for MspI, m1/m2, and m2/m2, which are the heterozygotes and the homozygotes, respectively, for the mutant allele and that have the site for MspI. Another mutation at position 4889 in exon 7 was found to be linked to MspI. This mutation leads to isoleucine/valine substitution in exon 7 (Hayashi et al., 1991).

**Glutathione S-transferases (GSTs):**

Glutathione S-transferases (GSTs) constitute a superfamily of ubiquitous, multifunctional enzymes, which play a key role in cellular detoxification, protecting macromolecules from attack by reactive electrophiles (Strange et al., 2001). The GSTs catalyze the conjugation of the tripeptide glutathione (GSH) to a wide variety of exogenous and endogenous chemicals with electrophilic functional groups (e.g. products of oxidative stress, environmental pollutants, and carcinogens), thereby neutralizing their electrophilic sites, and rendering the products more water-soluble (Hayes and Pulford, 1995). Based on sequence homology and immunological cross-reactivity, human cytosolic GSTs have been grouped into seven families, designated GST α, μ, π, δ, θ, ω, and ψ. Four members of the GST genes (GSTM1, GSTT1, GSTP1, and GSTM3) display polymorphisms that have been associated with increased risks for certain cancers (Board et al., 2000; Parl, 2005). In view of the importance of GSTs in cellular detoxification of carcinogens, genetic variants of GSTT1, GSTM1, and GSTP1 have been studied with respect to cancer risk.

**GSTM1:**

The GSTM subfamily is encoded by a 100-kb gene cluster at 1p13.3 arranged as 5’-GSTM4-GSTM2- GSTM1-GSTM5-GSTM3-3’ (Pearson et al., 1993; Xu et al., 1998) (Fig. 3.8). Deletion of the GSTM1 gene (GSTM1*0) results in the null (-/-) genotype. The GSTM1*0 deletion is caused by a homologous recombination involving the left and right 4.2-kb repeats (Xu et al., 1998). Analysis of 20 GSTM1*0 alleles from unrelated individuals in a study showed the same recombination pattern, which results in a 16-kb deletion containing
the entire GSTM1 gene. The GSTM1 gene is excised relatively precisely leaving the adjacent GSTM2 and GSTM5 genes intact. A missense single nucleotide polymorphism also occurs in the GSTM1 gene, i.e. nucleotide 534 G/C (172 Lys/Asn, corresponding to GSTM1*A and GSTM1*B, respectively), which does not appear to affect the enzyme function (Widersten et al., 1991).

![Diagram of GST gene cluster]

**Fig 3.9: The GSTM1 gene is part of the Mu-class GST gene cluster at 1p13.3 (Parl, 2005)**

**GSTT1:**

The GSTT subfamily consists of two genes, GSTT1 and GSTT2, which are located at 22q11.2 and separated by about 50 kb (Coggan et al., 1998; Parl, 2005; Whittington et al., 1999). Both genes have five exons with identical intron/exon boundaries but share only 55% amino acid identity. The deletion of the GSTT1 gene does not include GSTT2 (Coggan et al., 1998). Analysis of a 1197 kb section containing the GSTT1 and GSTT2 genes revealed extensive homologies e.g. two 18 kb regions, HA3 and HA5, with >90% homology flanking GSTT1. HA3 and HA5 contained two identical 403-bp repeats, which were identified as deletion/junction regions of the GSTT1 null allele (Sprenger et al., 2000). Similar to GSTM1*0, the GSTT1*0 deletion is most likely caused by a homologous recombination event involving the left and right 403-bp repeats. The recombination results in a 54-kb deletion containing the entire GSTT1 gene *(Fig. 3.9).*

![Diagram of GST gene cluster]

**Fig 3.10: The GSTT1 gene is part of the Theta-class GST gene cluster at Human 22q11.2 (Parl, 2005).**
The majority of polymorphisms affecting genes involved in carcinogen metabolism are single nucleotide polymorphisms. Deletions are less common and the complete absence of a gene in the form of a null allele is rare. It is for this reason that the GSTM1 and GSTT1<sup>-/-</sup> genotypes have been studied extensively for various cancers by different scientific groups involving cases from different geographical regions.

**GSTP1:**

The single GSTP1 gene at 11q13 is 2.8 kb long and contains seven exons (Kano et al., 1987; Parl, 2005) (Fig. 3.10). Several single nucleotide polymorphisms have been described in the GSTP1 gene. Two of the polymorphisms result in amino acid substitutions in codons 104 (Ile/Val) and 113 (Ala/Val) in exons 5 and 6, respectively (Board et al., 1989) (Fig. 3.10). Both amino acids 104 and 113 affect substrate specificity to the point of distinguishing between planar and non-planar substrates (Ji et al., 1999).

![Fig 3.11: Overview of GSTP1 gene at 11q13 locus, mRNA, and protein (Parl, 2005)](image)

Since GSTs have overlapping substrate specificities, deficiency of an individual GST isoenzyme may be compensated by other isoforms. Therefore, simultaneous determination of all GST genotypes appears to be a prerequisite for reliable interpretation of the role of the GST family in cancer development.

**NAT:**

The arylamine N-acetyltransferases (NATs) catalyse the acetyl transfer from acetyl coenzyme A to an aromatic amine, heterocyclic amine or hydrazine compound and are involved in the metabolism of a variety of different compounds that we are exposed to on a daily basis. Many drugs and chemicals found in the environment, such as those in cigarette
smoke, car exhaust fumes and in foodstuffs, can be either detoxified by NATs and eliminated from the body or bioactivated to metabolites that have the potential to cause toxicity and cancer. As a result, the levels of NATs in the body have important consequences with regard to an individual’s susceptibility to xenobiotics induced toxicities and cancers.

Two NAT isoenzymes have been identified in humans, namely NAT1 and NAT2, which are the products of distinct genetic loci, designated \textit{NAT1} and \textit{NAT2}, respectively (Blum \textit{et al.}, 1990). Each consists of an intronless open reading frame of 870 base pairs (Grant \textit{et al.}, 1989). The two genes are 87\% homologous and are located at 8p22, (Blum \textit{et al.}, 1990; Hickman \textit{et al.}, 1994) a chromosomal region commonly deleted in human cancers (Butcher \textit{et al.}, 2002). While the entire transcript of \textit{NAT1} is derived from a single exon, and that of \textit{NAT2} is derived from the protein encoding exon together with a second noncoding exon of 100 bp located about 8 kb upstream of the translation start site (Blum \textit{et al.}, 1990; Ebisawa and Deguchi, 1991).

Individuals who were homozygous for \textit{NAT2} polymorphisms had a slow acetylator phenotype, individuals heterozygous for \textit{NAT2} polymorphisms had an intermediate acetylator phenotype, and individuals who lacked \textit{NAT2} polymorphisms had a rapid acetylator phenotype. The frequency of the slow acetylator phenotype varies considerably among ethnic groups, (Evans, 1989) and this is due to the differing frequencies of the polymorphisms that correspond to the slow acetylator alleles.

Historically, NAT1 was thought to be genetically invariant or monomorphic in nature. Western blots for NAT1 showed that low activity was due to a parallel decrease in NAT1 protein content, indicating that slow acetylator status was a result of a decrease in the amount of a functionally normal enzyme rather that the presence of a protein with altered acetylation capacity (Butcher \textit{et al.}, 2002). Altered risk with either the slow or rapid phenotype has been observed for bladder, colon and breast cancer, systemic lupus erythematosis, diabetes, Gilbert’s disease, Parkinson’s disease and Alzheimer’s disease. These associations imply a role for environmental factors that are metabolised by the NATs, in particular NAT2, in each disorder (Butcher \textit{et al.}, 2002).

\textbf{NQO1:}

NAD(P)H: quinine oxidoreductase-1 (NQO1) which was first called DT-diaphorase, is a flavoprotein, known to catalyze two electron reduction of a broad range of substrates
(Lind et al., 1990). The broad substrate specificity has also been explained by structural studies demonstrating the presence of a highly plastic active site that can accommodate a range of structures (Faig et al., 2001).

A polymorphism of C to T in the 609th codon of NQO1 DNA leads to the formation of the NQO1*2 allele, which is markedly weak in its biochemical activity. The NQO1 T allele has only 2-4% enzymatic activity in comparison to its wild-type form. Cells homozygous for the polymorphic NQO1 allele (T/T) express NQO1 mRNA, but they have no detectable NQO1 protein because the mutant NQO1 protein is rapidly degraded by the proteasomal system (Siegel et al., 2001). This C to T substitution causes a proline to serine change in the 187th amino acid location of the NQO1 protein sequence. The heterozygous phenotype for the T allele for NQO1 gene has 3-fold weaker enzyme activity, and that homozygous for T allele has almost complete enzyme activity loss (Zhang et al., 2003). NQO1 gene expression is stimulated by oxidative compounds such as PAHs and also antioxidants (Long et al., 2001).

Since NQO1 is expressed at high levels throughout many human solid tumors, compounds efficiently bioactivated by NQO1 have been designed for the therapy of tumors rich in NQO1 (Beall et al., 1995; Faig et al., 2001; Winski et al., 1998; Winski et al., 2001). Currently, a new NQO1-targeted aziridinylbenzoquinone, RH1, (Winski et al., 1998) is undergoing phase 1 clinical trials. Induction of NQO1 has been demonstrated to protect against the cytotoxicity, mutagenicity and carcinogenicity of many compounds.

NQO1 as Component of Stress Response:

Stabilization of p53 Studies with proteins typically considered as metabolic enzymes suggest that these proteins may have additional roles outside the range of their normal metabolic functions. For example, glutathione-S-transferase associates with c-Jun N-terminal kinase leading to inhibition of kinase activity and modulation of signaling and cellular proliferation (Adler et al., 1999; Ruscoe et al., 2001; Wang et al., 2001).

Few studies reported that NQO1 may influence the stability of the tumor suppressor protein p53 by inhibiting its degradation (Asher et al., 2001; Asher et al., 2002a; Asher et al., 2002b). In these studies the authors hypothesized that the NQO1-mediated conversion of NADH to NADP promoted stabilization of p53. A significant study has shown a direct physical interaction between p53 and NQO1 (Anwar et al., 2003) (Figure 3.12).
Fig 3.12: Proposed mechanism of stabilization of p53 via a protein-protein interaction with NQO1 (Anwar et al., 2003).

**Regulation of Detoxification Activities:**

Specific detoxification pathways may be induced or inhibited depending on the presence of various dietary or xenobiotic compounds, the age and sex of the individual, genetics, and lifestyle habits, such as smoking (Goldberg, 1996; Meyer et al., 1990; Park et al., 1996; Vesell, 1979). Furthermore, various diseases can also influence activity of the enzymes. Inhibition of these enzymes can occur by competition between two or more compounds for the same detoxifying enzyme. Increased toxic exposure may lead to inhibition of detoxification of a number of compounds by simply overwhelming the systems and competing for detoxification enzyme activities (Liska, 1998). Mono-functional inducers, such as polycyclic hydrocarbons from cigarette smoke and aryl amines from charbroiled meats, result in dramatic induction of the CYP1A1 and CYP1A2 enzymes, leading to a substantial increase in Phase I activity, with little or no induction of Phase II enzymes (Guengerich, 1984). Similarly, glucocorticoids and anti-convulsants induce CYP3A4 activity, and ethanol, acetone, and isoniazid induce CYP2E1 (Park et al., 1996; Wacher et al., 1995). Induction of these activities without co-induction of Phase II activities may lead to an uncoupling of the Phase I and Phase II balance of activity and, therefore, a higher level of reactive intermediates, which can cause damage to DNA, RNA, and proteins (Elangovan et al., 1994; Park et al., 1996). The multifunctional inducers include many of the flavonoid molecules found in fruits and vegetables. For example, ellagic acid found in red grape skin has been shown to induce several Phase II enzymes while decreasing Phase I activity (Manson et al., 1997). Garlic oil, rosemary, soy, cabbage, brussels sprouts, fruits and vegetables all contain...
compounds that can induce several Phase II enzyme activities (Appelt and Reicks, 1997; Guengerich, 1984; Ip and Lisk, 1997; Manson et al., 1997; Offord et al., 1995; Pantuck et al., 1979; Park et al., 1996). In general, this increase in Phase II supports better detoxification in an individual and helps to promote and maintain a healthy balance between Phase I and Phase II activities and explain the ability of fruits and vegetables to protect against many cancers (Elangovan et al., 1994; Guengerich, 1984; Liska, 1998; Manson et al., 1997; Park et al., 1996).

**Codon 72 of p53 Gene:**

P53 tumour suppressor gene is an important component of DNA repair machinery in response to DNA damage induced by radiation or adducts formation. Mutations in p53 have been reported to be associated with reduced genomic repair capacity and enhanced cytotoxicity in cells damaged by benzo(a)pyrene diol epoxide-DNA adducts (Wani et al., 2000). 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 p53 gene and cancer risk.

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 codons 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 (Mitra et al., 2005; Papadakis et al., 2000; Rogounovitch et al., 2006; Tandle et al., 2001; Wu et al., 2004). Reports on codon 72 of p53 gene available from India is limited and inconsistent and the results are conflicting whereas no association was reported with oral cancer (Nagpal et al., 2002; Tandle et al., 2001). However another study showed carriers of Arg/Arg genotype at higher risk for oral cancer (Katiyar et al., 2003).

Thus study of genes [CYP1A1 (Msp1 and Nco1), GSTT1, GSTM1, GSTP1, NAT2 and NQO1 genes] encoding for phase I and phase II detoxifying enzymes alongwith codon 72 of p53 gene polymorphism will enhance our understanding about oral carcinogenesis and may provide some answers for high incidence of oral cancer in Indian population particularly in northeast region of India.
The Tumor Microenvironment:

Tumors have increasingly been recognized as organs whose complexity approaches and may even exceed that of normal healthy tissues. An assemblage of diverse cell types constitutes most solid tumors. Both the parenchyma and stroma of tumors contain distinct cell types and subtypes that collectively enable tumor growth and progression. Notably, the immune inflammatory cells present in tumors can include both tumor-promoting as well as tumor-killing subclasses. The multiple stromal cell types create a succession of tumor microenvironments that change as tumors invade normal tissue and thereafter seed and colonize distant tissues. The abundance, histologic organization, and phenotypic characteristics of the stromal cell types, as well as of the extracellular matrix (hatched background), evolve during progression, thereby enabling primary, invasive, and then metastatic growth. The surrounding normal cells of the primary and metastatic sites, shown only schematically, likely also affect the character of the various neoplastic microenvironments (Hanahan and Weinberg, 2011) (Figure 3.13).

**Fig 3.13:** Different type of cells of a typical solid tumor microenvironment and their arrangement and complexity at various stages of tumour development (Hanahan and Weinberg, 2011)
Signaling Interactions in the Tumor Microenvironment during Malignant progression:

The intracellular signaling within the tumor microenvironment is not static but changes during tumor progression as a result of reciprocal signaling interactions between cancer cells of the parenchyma and stromal cells that convey the increasingly aggressive phenotypes that underlie growth, invasion, and metastatic dissemination. Certain organ sites (sometimes referred to as “fertile soil” or “metastatic niches”) can be especially permissive for metastatic seeding and colonization by certain types of cancer cells, as a consequence of local properties that are either intrinsic to the normal tissue or induced at a distance by systemic actions of primary tumors. Cancer stem cells may be variably involved in some or all of the different stages of primary tumorigenesis and metastasis (Hanahan and Weinberg, 2011) (Figure 3.14).

![Signaling interactions between various cancerous cell types in the tumor microenvironment during malignant progression](image)

*Fig 3.14: Signaling interactions between various cancerous cells in the tumor microenvironment during malignant progression* (Hanahan and Weinberg, 2011).
Intracellular Signaling Networks Regulate the Operations of the Cancer Cell:

A complex integrated circuit operates within normal cells and is reprogrammed to control hallmark capabilities within cancer cells. Separate subcircuits, depicted in the following figure in differently colored fields, are specialized to orchestrate the various capabilities. The intracellular integrated circuit can be segmented into distinct sub-circuits, each of them is specialized to support a discrete cell-biological property in normal cells and is reprogrammed in order to implement a hallmark capability in cancer cells.

An additional dimension of complexity involves considerable interconnections and thus crosstalk between the individual subcircuits. For example, certain oncogenic events can affect multiple capabilities, as illustrated by the diverse effects that prominent oncogenes, such as mutant RAS and upregulated MYC, have on multiple hallmark capabilities (e.g., proliferative signaling, energy metabolism, angiogenesis, invasion, and survival) (Hanahan and Weinberg, 2011) (Figure 3.15).

Figure 3.15: Intracellular signaling networks which regulate the operations of the cancer cell (Hanahan and Weinberg, 2011).
Hallmarks of cancer and possible therapeutic targeting:

Hallmarks of cancer have been defined by ‘Hanahan and Weinberg’ as acquired functional capabilities that allow cancer cells to survive, proliferate, and disseminate; these functions are acquired in different tumor types via distinct mechanisms and at various times during the course of multistep tumorigenesis. Acquisition of these hallmarks is made possible by two enabling characteristics. Most prominent is the development of genomic instability in cancer cells, which generates random mutations including chromosomal rearrangements; among these are the rare genetic changes that can orchestrate hallmark capabilities. A second enabling characteristic involves the inflammatory state of premalignant and frankly malignant lesions that is driven by cells of the immune system. Summary of hallmarks of cancer and their possible therapeutic approaches have been shown in the (Hanahan and Weinberg, 2011) Figure 3.16.

Fig 3.16: Hallmarks of cancer (written in blue text) and possible therapeutic targeting (in text boxes) (Hanahan and Weinberg, 2011).
Analyzing genome-wide aberrations in cancer using high throughput methods:

Advances in the molecular-genetic analysis of cancer cell genomes have provided the most compelling demonstrations of function-altering mutations and of ongoing genomic instability during tumor progression. One type of analysis - comparative genomic hybridization (CGH) documents the gains and losses of gene copy number across the cell genome; in many tumors, the pervasive genomic aberrations revealed by CGH provide clear evidence for loss of control of genome integrity. Importantly, the recurrence of specific aberrations (both amplifications and deletions) at particular sites in the genome indicates that such sites are likely to harbor genes whose alteration favors neoplastic progression (Korkola and Gray, 2010).

More recently, with the advent of efficient and economical DNA-sequencing technologies, higher-resolution analyses have become possible. Early studies are revealing distinctive patterns of DNA mutations in different tumor types (source [http://cancergenome.nih.gov/](http://cancergenome.nih.gov/)). Genomewide sequencing promises to clarify the prevalence of ostensibly random mutations scattered across cancer cell genomes. Recurring genetic alterations may point to a causal role of particular mutation in tumor pathogenesis. It is evident that the defects in genome maintenance and repair are selectively advantageous and therefore instrumental for tumor progression, because they accelerate the rate at which evolving premalignant cells can accumulate favorable genotypes. As such, genome instability is clearly an enabling characteristic that is causally associated with the acquisition of hallmark capabilities.

Next Generation Sequencing Technology:

Next-generation sequencing (NGS) broadly describes those technologies that share the ability to massively parallel sequence millions of DNA templates. NGS is arguably one of the most significant technological advances in the biological sciences of the last 30 years. The terms second-generation and third-generation sequencing are also used synonymously to describe the evolution of sequencing technology from the first-generation dideoxy ‘Sanger’ sequencing. To achieve massive parallel sequencing, second-generation platforms employ the clonal amplification of DNA templates on a solid support matrix followed by cyclic sequencing. The second generation sequencing platforms have advanced rapidly to the point
that several genomes can now be sequenced simultaneously in a single instrument run in under two weeks. The shift to single molecule PCR-free protocols and cycle-free chemistry is broadly characteristic of the progression to third-generation platforms (Meldrum et al., 2011; Schadt et al., 2010).

Medical research has warmly welcomed the technology and the cancer field is at the forefront of these efforts given the genetic aspects of the disease. World-wide efforts to catalogue mutations in multiple cancer types are underway and this is likely to lead to new discoveries that will be translated to new diagnostic, prognostic and therapeutic targets. NGS is now maturing to the point where it is being considered by many laboratories for routine diagnostic use. The sensitivity, speed and reduced cost per sample make it a highly attractive platform compared to other sequencing modalities (Meldrum et al., 2011).

Immediate and significant impact will come from either replacement or expansion of existing technologies for genetic screening purposes. Some striking examples of its clinical use include prenatal testing for the detection of chromosomal aneuploidy in foetal DNA, (Chiu and Lo, 2010), identification of rare genetic variants associated with monogenic Mendelian disorders (Lupski et al., 2010; Sobreira et al., 2010) and efficient detection of either inherited or somatic mutations in cancer genes (Link et al., 2011; Welch et al., 2011).

As cancer is a genetic disease driven by heritable or somatic mutations, new DNA sequencing technologies will have a significant impact on the detection, management and treatment of disease. Next-generation sequencing is enabling worldwide collaborative efforts, such as the International Cancer Genome Consortium (ICGC) (Hudson et al., 2010) and The Cancer Genome Atlas (TCGA) project, (http://cancergenome.nih.gov) to catalogue the genomic landscape of thousands of cancer genomes across many disease types (Meldrum et al., 2011).

The predominant application of NGS in a clinical setting will undoubtedly be resequencing of genomic DNA. Whole genome sequencing (WGS) simply provides the ultimate genetic survey of an individual’s genome or cancer genome where a detailed map of single nucleotide variations (SNV), indels, complex structural rearrangements and copy number changes can be attained in a single assay (Pleasance et al., 2009).

Targeted DNA enrichment methods allow even higher genome throughput at a reduced cost per sample. Targeted enrichment strategies feeding into NGS are finding
importance in both research and clinical diagnostic fields. Targeted enrichment before sequencing can reduce costs, allow higher coverage over regions of interest and potentially simplify the bioinformatic interpretation of NGS data (Meldrum et al., 2011). Whilst single targets for single therapies are currently the norm, it is very likely that future treatments will rely more on therapies directed to multiple targets to avoid relapses common to these treatment modalities (Meldrum et al., 2011).

**Illumina-Solexa Sequencing by Synthesis method for Next Generation Sequencing:**

The sequencing technology developed by Solexa, subsequently acquired by Illumina, is, like other high throughput sequencing methods, based on first fragmenting the DNA followed by sequencing the fragments and then reconstructing the full sequence in a computer. After fragmentation, adaptor sequences are ligated to the fragment ends. The fragments are then distributed randomly onto a surface already covered with small pieces of DNA complementary to the adaptor sequences (Figure 3.17A).

![Fig 3.17A](http://seqanswers.com/forums/showthread.php?t=21)

**Fig 3.17A:** 1) Randomly fragment genomic DNA and ligate adaptors to both ends of the fragments. 2) Bind single stranded fragments randomly to the inside surface of the flow cell channels. 3) Add unlabelled nucleotides and enzyme to initiate solid-phase bridge amplification. (http://seqanswers.com/forums/showthread.php?t=21)
The fragments get attached to the surface and the fragments are multiplied using a technique called bridge amplification, to form small clusters of single stranded fragments. The clusters are formed spontaneously because of the fact that the newly produced copies of the fragment get attached in close proximity to the original fragment. After the bridge amplification is done, there is no single fragment sparsely distributed over the surface, but instead densely packed clusters of fragments, each cluster consisting of many single stranded copies of the same fragment (Figure 3.17B).
unused nucleotides are washed away and a laser is used to scan the surface which makes the fluorescent dyes emit light of different colors, one color for each type of nucleotide. A camera records the color of the light being emitted by each cluster of fragments, and this is where the actual sequencing takes place. (Figure 3.17C)

Fig 3.17 C: 7) First chemistry cycle: to initiate the first sequencing cycle. Add all four labelled reversible terminators, primers and DNA polymerase enzyme to the flow cell. 8) After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster. 9) Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell. (http://seqanswers.com/forums/showthread.php?t=21).

Since the nucleotides are reversible terminators it is easy to remove the terminator part and also the dyes, and repeat the process. The process of adding nucleotides, imaging and removing the terminator is called a cycle. Each cycle results in one determined nucleotide from all fragments. The number of cycles is currently limited to 150 for the GenomeAnalyzer IIx and 100 for the HiSeq2000 instruments respectively. The maximum number of cycles is limited by imperfections in the chemistry causing an increased uncertainty in the measurements as the number of cycles increases. This results in a reduction of the quality of the determined bases for each cycle added.
Each cycle results in a number of images, one for each color multiplied by the number required to cover the whole surface. The images are then analyzed in two steps. First the clusters and their intensities are identified and written to intensity files, and then the bases are determined from the intensities. Each base is assigned quality value ranging from 2 to 40, depending on the certainty in the basecalling. ([https://www.uppnex.uu.se/uppnex-book/technologies/solexa-sequencing](https://www.uppnex.uu.se/uppnex-book/technologies/solexa-sequencing)).

Fig 3.17D: 10) After laser excitation; collect the image data as before. Record the identity of the second base in each cluster. 11). Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time. 12). Align data, compare to a reference, and identify sequence differences. ([http://seqanswers.com/forums/showthread.php?t=21](http://seqanswers.com/forums/showthread.php?t=21))

**Advantages and disadvantages of NGS technology:**

With the evolution of sequencing technologies, there is much hope given to the promise of genomics and the impact that technology platforms will have on our understanding, diagnosis and treatment of diseases. With this development research in cancer genomics is entering a period of great promise and also of great expectation. However, it should be remembered that the sequencing technology methods are only one aspect of a routine diagnostic molecular pathology laboratory. Sequencing technologies have been used
for some time as a clinical diagnostic tool and it has taken at least a decade for the support systems required for data analysis and interpretation to be developed, scrutinized and validated for clinical diagnostic use. At the same time, ongoing sequencing efforts will continue to generate massive quantities of data. A major challenge in cancer genomics is the standardization, storage, and public availability of these data (Haimovich). In some aspects they still fall short of the requirements of a diagnostic laboratory for a number of applications. Nevertheless, during this time the limitations of Sanger sequencing systems have become well understood and commercial software tools are readily available to assist with these tasks.

From a technical perspective some problems related to NGS will include higher error rates, fundamental platform differences, the selection of appropriate quality values and data handling. In many respects these issues are not unique to NGS but are likely to be exaggerated by its use. The validation of new sequencing platforms will be achieved through retrospective comparison to previously analysed samples and will also require validation of the software used by the platforms and or supplied by third party vendors or open access software. Exhaustive prospective comparisons will be required since retrospective testing is unlikely to correlate entirely with data obtained from newly designed NGS panels (Meldrum et al., 2011).

Clinical data interpretation will require much greater demands both within the laboratory and in the clinic. Protocols for dealing with NGS data that guide what and how particular information will be reported and conveyed to the clinician will need to be established.

**Gene Expression Profiling using Microarray:**

High-throughput, genome-wide analytical technologies are now commonly used in all fields of medical research. The most commonly applied of these technologies, gene expression microarrays, have been shown to be both accurate and precise when properly implemented (Bhattacharya and Mariani, 2009). Gene expression microarrays provide a wealth of information on gene expression patterns and cancer pathways with potential for (1) cancer diagnosis, prognosis, and prediction of therapeutic responsiveness, (2) discovering new cancer subtypes and (3) identifying cancer-associated (signalling) molecular markers and their complex interactions (Wang et al., 2008b).
The advantage of microarray analysis of gene expression is that it can be used to discover some genes that were previously thought to be unrelated to a physiologic or pathologic event. During the last decade, applications of microarray in cancer investigation have shifted from molecular profiling, identifying previously undiscovered cancer types, predicting outcomes of cancer patients, revealing metastasis signatures of solid tumors, to guiding the use of therapeutics. The role of cancer genomic signatures has evolved through three phases. In the first phase, genomic signatures were described in stored cancer specimens and dubbed as molecular portraits of cancer. When gene expression profiles were carefully correlated with sufficient clinical information of cancer patients, new subgroups of cancers with distinct outcomes were revealed. In studies of the second phase, validation of cancer signatures was emphasized and commonly performed with independent groups of cancer specimens or independent data set. In the third phase, cancer genomic signatures have been further expanded beyond depicting the molecular portrait of cancer to predicting patient outcomes and guiding the use of cancer therapeutics. Cancer genomic signatures have become an essential part of a new generation of cancer clinical trials. It is advocated that, in future clinical trials of cancer therapy, the cancer specimens of each participant should be tested for currently available predictor genomic signatures, so that the most effective treatment with the least adverse effects for each patient can be identified. Then, participants can be sorted to an appropriate study group (Wang and Chao, 2007).

The microarray technology allows simultaneous analysis of the global gene expression in cells and tissues. In tumor diseases, important transcriptional changes of genes have been unravelled by this method. Microarray analysis is routinely able to identify biomarkers correlated with survival and reveal pathways underlying pathogenesis and invasion (Lexe et al., 2009) The resulting gene expression profiles can be used for the prediction of diagnosis, prognosis or therapeutic outcome, as well as for the identification of novel drug targets (Kuner et al., 2009).

**Microarray Technology:**

Microarray methods were initially developed to study differential gene expression using complex populations of RNA (Lipshutz et al., 1999). Refinements of these methods now permit the analysis of copy number imbalances and gene amplification of DNA (Pollack et al., 1999) and have recently been applied to the systematic analysis of expression at the protein level (Haab, 2001). Many of the guiding principles of global analysis using
microarrays are, in principle, applicable at the RNA, DNA, or protein level (Macgregor and Squire, 2002). Any microarray study typically involves six steps (Figure 3.18):

1. Manufacturing of Microarrays
2. Experimental Design and choice of Reference
3. Target preparation and Hybridization
4. Image acquisition and Quantification
5. Databases and Normalization
6. Statistical Analysis and Data Mining

Fig 3.18: Six Steps in a Microarray Experiment (Macgregor and Squire, 2002)
1. Manufacturing of Microarrays:

Spotted arrays are manufactured using xyz robots that use hollow pins to deposit cDNA (PCR products) or short oligonucleotides onto specially coated glass microscope slides (Schena et al., 1995). Spot sizes range between 80 and 150 nm in diameter, and arrays that contain up to 80 000 spots can be obtained. Gene sequences to be arrayed are selected from several public databases, which contain resources to access well-characterized genes and expressed sequence tags (ESTs) representative of genes of unknown function (Fodor et al., 1991). The clones chosen are amplified from appropriate cDNA libraries by PCR and purified before spotting on the solid support.

Because these arrays can be spotted with thousands of sequenced expressed genes and ESTs of unknown function, they offer the potential for the discovery of new genes and defining their role in disease. One disadvantage of spotted arrays is that they provide information only on the relative gene expression between specific cells or tissue samples as opposed to direct quantification of RNA expression (Macgregor and Squire, 2002).

2. Experimental Design and choice of Reference:

In a case - control study, two samples from a single individual, e.g., tumor tissue and healthy tissue, are compared directly. Because patient variability and genetic heterogeneity are key issues in microarray data analysis, the case - control design is an excellent solution when feasible.

3. Target preparation and Hybridization:

Both total RNA and mRNA can be used for microarray experiments and allow the attainment of high-quality data with a high degree of confidence. High-quality RNA is crucial for successful microarray experiments. Different standard RNA extraction methodologies have been used successfully, and the choice of protocol is largely a question of personal experience. Quantitative and qualitative evaluation of the RNA obtained can be carried out by standard techniques, such as agarose gel electrophoresis, but is limited by the relatively large amounts of sample required. More recently, assessment of RNA quality and quantity has been greatly facilitated by the use of microcapillary- based devices such as the Agilent Bioanalyzer (Agilent Technologies), which can be used with as little as 5 ng of total RNA (Macgregor and Squire, 2002).
For standard microarray experiments, the isolated RNA is reverse-transcribed into
target cDNA in the presence of fluorescent (generally Cy3-dNTP or Cy5-dNTP) or
radiolabeled deoxynucleotides ([33P]- or [32P]-dCTP). After purification and denaturation, the
labeled targets are hybridized to the microarrays at a temperature determined by the
hybridization buffer used. After hybridization, the arrays are washed under stringent
conditions to remove nonspecific target binding and are air-dried (Macgregor and Squire,
2002).

4. Image acquisition and Quantification:

Microarray image processing uses differential excitation and emission wavelengths of
the two fluors to obtain a scan of the array for each emission wavelength, typically as two 16-
bit grayscale TIFF images. These images are then analyzed to identify the spots, calculate
their associated signal intensities, and assess local background noise. Most image acquisition
software packages also contain basic filtering tools to flag spots such as extremely
lowintensity spots, ghosts spots (where background is higher than spot intensity), or damaged
spots (e.g., dust artifacts). These results allow an initial ratio of the evaluated
channel/reference channel intensity to be calculated for every spot on the chip. The products
of the image acquisition are the TIFF image pairing and a quantified data file that has not yet
been normalized (Macgregor and Squire, 2002).

5. Databases and Normalization:

The quantity of data generated in a microarray experiment typically requires a
dedicated database system to store and organize the microarray data and images. The first
role of a local microarray database is the storage and annotation (description of experimental
parameters) of microarray experiments by the investigator who designed and carried out the
microarray experiments. To answer this need, the Minimal Information about a Microarray
Experiment, (MIAME) standard, has been proposed by the MGED (http://www.mged.org)
organization as a series of criteria that should be used when defining microarray experiment
parameters. Normalization is a process that scales spot intensities such that the normalized
ratios provide an approximation of the ratio of gene expression between the two samples
(Macgregor and Squire, 2002).

6. Statistical Analysis and Data Mining:

Analysis of large gene expression data sets is a new area of data analysis with its own
unique challenges. Data mining methods typically fall into one of two classes: supervised and
unsupervised. In unsupervised analysis, the data are organized without the benefit of external
classification information. Hierarchical clustering (Eisen et al., 1998), Kmeans clustering (Tavazoie et al., 1999), or self-organizing maps (Tamayo et al., 1999; Tavazoie et al., 1999) are examples of unsupervised clustering approaches that have been widely used in microarray analysis (Alizadeh et al., 2000; Eisen et al., 1998; Macgregor and Squire, 2002; Tamayo et al., 1999; Tavazoie et al., 1999).

Real-time Reverse Transcription PCR:

To validate the expression profiles of the target genes obtained through microarray experiments, one commonly used method is quantitative real-time reverse transcription-PCR (qRT-PCR). qRT-PCR has become the benchmark for the detection and quantification of RNA targets and is firmly established as a mainstream research technology (Ginzinger, 2002). Its potential for high-throughput, together with regular introduction of enhanced or novel chemistries, more reliable instrumentation and improved protocols, has also seen the development of qRT-PCR-based clinical diagnostic assays.

Advantages of qRT-PCR over conventional RT-PCR:

Quantitative real time RT-PCR assays have several significant advantages over conventional RT-PCR (Bustin and Mueller, 2005; Orlando et al., 1998):

(i) They use fluorescent reporter molecules to monitor the production of amplification products during each cycle of the PCR, and the combination of the DNA amplification and detection steps into one homogeneous assay obviates the requirement for post-PCR processing;

(ii) Their wide dynamic range allows the analysis of samples differing in target abundance by orders of magnitude;

(iii) There is little inter-assay variation, which helps generate reliable and reproducible results.

(iv) Fluorescence based qRT-PCR realizes the inherent quantitative capacity of PCR based assays making it a quantitative rather than a qualitative assay (Bustin and Mueller, 2005; Halford et al., 1999).
**Principle of qRT-PCR:**

Following the RT of RNA into cDNA, it requires a suitable detection chemistry to report the presence of PCR products, an instrument to monitor the amplification in realtime and appropriate software for quantitative analysis (Bustin and Mueller, 2005; Wittwer *et al.*, 1997). qRT-PCRs are characterized by the point in time during cycling when amplification of a PCR product is first detected. The higher the starting copy number of the nucleic acid target, the sooner a significant increase in fluorescence is observed.

**The Chemistry of real time PCR:**

Generally, real-time PCR chemistries consist of special fluorescent probes in the PCR (Figure. 3.19). Several types of probes exist, including DNA-binding dyes like EtBr or SYBR green I, hydrolysis probes (5’-nuclease probes), hybridization probes, molecular beacons, sunrise and scorpion primers, and peptide nucleic acid (PNA) light-up probes. Each type of probe has its own unique characteristics, but the strategy for each is simple. They must link a change in fluorescence to amplification of DNA (Valasek and Repa, 2005).

SYBR green I binds to the minor groove of dsDNA, emitting 1,000-fold greater fluorescence than when it is free in solution (Figure. 3.19A). Therefore, the greater the amount of dsDNA present in the reaction tube, the greater the amount of DNA binding and fluorescent signal from SYBR green I. Thus any amplification of DNA in the reaction tube is measured. Other dsDNA-specific dyes (e.g., BEBO, YOYO-1, TOTO-1, etc.) have also been described but are not as widely used. The primary concern with the usage of any of these sequence independent dsDNA-binding probes is specificity. To help ensure specificity, the dissociation curve of the amplified product can be analyzed to determine the melting point. If there are two or more peaks, it suggests that more than one amplified sequence was obtained, and the amplification was not specific for a single DNA target (Valasek and Repa, 2005).

Hydrolysis probes (also called 5’-nuclease probes because the 5’-exonuclease activity of DNA polymerase cleaves the probe) offer an alternative approach to the problem of specificity (Figure. 3.19B). These are likely the most widely used fluorogenic probe format (Mackay, 2004) and are exemplified by TaqMan probes. In terms of structure, hydrolysis probes are sequence specific dually fluorophore-labeled DNA oligonucleotides. One
fluorophore is termed the quencher and the other is the reporter. When the quencher and reporter are in close proximity, that is, they are both attached to the same short oligonucleotide, the quencher absorbs the signal from the reporter. This is an example of fluorescence resonance energy transfer (also called Förster transfer) in which energy is transferred from a “donor” (the reporter) to an “acceptor” (the quencher) fluorophore. During amplification, the oligonucleotide is broken apart by the action of DNA polymerase (5’-nuclease activity) and the reporter and quencher separate, allowing the reporter’s energy and fluorescent signal to be liberated. Thus destruction or hydrolysis of the oligonucleotide results in an increase of reporter signal and corresponds with the specific amplification of DNA. Examples of common quencher fluorophores include TAMRA, DABCYL, and BHQ, whereas reporters are more numerous (e.g., FAM, VIC, NED, etc) (Valasek and Repa, 2005). Hydrolysis probes afford similar precision as SYBR green I (Wilhelm et al., 2003) but they give greater insurance regarding the specificity because only sequence-specific amplification is measured. In addition, hydrolysis probes allow for simple identification of point mutations within the amplicon using melting curve analysis (Valasek and Repa, 2005).

**Fig 3.19: Real-time PCR chemistries (Valasek and Repa, 2005).**  
**A:** SYBR green I fluoresces (absorbing light of 480-nm wavelength and emitting light of 520-nm wavelength) when associated with dsDNA.  
**B:** other detection formats often utilize compatible fluorophores. Shown in this example is the Taqman probe, which contains a reporter fluorophore (R) that emits at a wavelength absorbed by the quencher fluorophore (Q). During PCR amplification, the DNA polymerase (Taq) cleaves the probe, thus liberating the reporter from the quencher and allowing for measurable fluorescence.
The RT step synthesizes a cDNA copy of the RNA template. After denaturation, primers and probe anneal to their targets. The probe contains a reporter dye at the 5’ end and a quencher (Q) at its 3’ end. During the polymerization step, the 5’ nuclease activity of the Taq polymerase displaces and cleaves the probe. This physically separates the reporter dye and quencher dyes, resulting in reporter fluorescence. The increase in signal is directly proportional to the number of molecules released during that cycle. Accumulation of PCR products is detected directly by monitoring the increase in fluorescence of the reporter dye.
**Characteristics of qRT-PCR amplification curves:**

The curves for three samples, run in duplicate, are shown in the figure 3.21 below. Ct values are indicated by arrows and represent the cycle fractions where the instrument can first reliably detect fluorescence derived from the amplification reaction. The fluorescence signal during the initial cycles of the PCR is below the instrument’s detection threshold and defines the baseline for the amplification plot. An increase in fluorescence above the threshold indicates the detection of accumulated PCR product. The key parameter Ct is defined as the fractional cycle number from clinical samples at which the fluorescence passes a fixed threshold chosen either by the instrument or by the operator. A plot of the log of initial target copy number for a set of standards versus Ct is a straight line. The amount of target in an unknown sample is quantified by measuring the Ct and using the standard curve to determine starting copy number (Bustin and Mueller, 2005).

![Fig 3.21: qRT-PCR amplification curves plot fluorescence signal versus cycle number (Bustin and Mueller, 2005).](image)

Quantitative results obtained by qRT-PCR are not only more informative than qualitative data, but simplify assay standardization and quality management. qRT-PCR is being utilized increasingly in novel clinical diagnostic assays, since it can be automated and performed on fresh or archived formalin-fixed, paraffin-embedded tissue samples. The outcome of these analyses might accelerate the application of basic research findings into
daily clinical practice through translational research and may have an impact on foreseeing the clinical outcome, predicting tumour response to specific therapy, identification of new prognostic biomarkers, discovering targets for the development of novel therapies and providing further insights into tumour biology.
Chapter 4

Role of SNPs in Xenobiotic Metabolizing Genes (Encoding Phase I and Phase II Detoxification Enzymes) and $p53$

Codon 72 Polymorphism in Oral Cancer
INTRODUCTION:

It has been estimated that of the 390,000 oral and oro-pharyngeal cancers that occur annually worldwide, 58% occur in south and south-east Asia (Nair et al., 2004). In India, oral cancer is the most common cancer among men and ranks third among women (Soya et al., 2007), with age-standardized incidence rates per 100,000 population being 12.8 and 7.5 respectively (Nair et al., 2004). However there are regional differences within India. Prevalence of aerodigestive tract cancers including oral cancer was reported to be highest in some northeastern (NE) regions (Bhattacharjee et al., 2006; ICMR-Report, 2006; Phukan et al., 2004). In the Assam region there is widespread chewing habit of tobacco with peculiar fermented betel nut. Betelnut contains arecoline which can produce 3-methylnitrosaminopropionitrile (MNPN), a potent carcinogen and safrole like DNA adduct that have been shown to be genotoxic and mutagenic. The widespread chewing habit of tobacco with peculiarly fermented betel nut may further add to the risk for oral cancer in this region (Phukan et al., 2001).

India is the second largest producer and consumer of tobacco in the world after China. Polycyclic aromatic hydrocarbons (PAHs), nitrosamines, aldehydes and ketones are the major carcinogens present in tobacco (Hecht, 2003). However the concentrations of these compounds vary depending upon the nature of tobacco use. Smokeless tobacco is rich in nitrosamines while due to high temperatures at the burning tip, tobacco smoke contains pyrolysis products (Anantharaman et al., 2007; Hecht, 2003). These compounds are capable of generating reactive oxygen species (ROS) by direct reaction or metabolic activation (Yin et al., 2001). Increased levels of ROS and down regulation of ROS scavengers and antioxidant enzymes are associated with various cancers (Waris and Ahsan, 2006). The enzymatic detoxification process is mainly divided into three phases. Phase I involves activation of toxic compounds predominantly by oxidation into more reactive intermediates that are neutralized and conjugated by phase II family of enzymes such as glutathione-s-transferase (GST) (Guengerich, 1990; Sheehan et al., 2001). The resultant water soluble and less toxic conjugated product can easily be eliminated from the cell by Phase III transport mechanisms for the elimination of conjugates.
The genetic predisposition or host susceptibility to various carcinogens is regulated by interactions between genetic host factors and carcinogens in the ambient environment such as those in the tobacco chewing, smoking, diet and ambient air (Raunio et al., 1995). Majority of the environmental carcinogenic chemicals do not produce their biological effects per se, but require metabolic activation to get converted into their respective reactive electrophilic intermediates which interact with cellular macromolecules (Tatemichi et al., 1999). Most of the carcinogenic chemicals are converted to reactive electrophilic metabolites by the cytochrome P-450 (CYP) superfamily of phase I enzymes. Subsequently, the phase II enzymes detoxify these intermediates by conjugation reactions (Guengerich and MacDonald, 1990; Sheehan et al., 2001). Some of the major phase II enzymes are glutathione S-transferases (GSTs), N-acetyltransferases (NAT), and NAD(P)H-quinone oxidoreductase 1 (NQO1).

\textbf{CYP1A1} is aromatic hydrocarbon hydroxylase, which catalyzes the first oxidative step in the metabolism of many substrates including aromatic polycyclic aromatic hydrocarbons (PAHs) like benzo (\textalpha) pyrene, a constituent of tobacco (Siraj et al., 2008). The metabolic products are usually highly reactive oxygen species and more carcinogenic than the parent compounds which with the help of phase II enzymes are detoxified for excretion (Terry et al., 2003). Large differences in the frequency of \textit{CYP1A1} polymorphic variants have been reported among different ethnic groups (Fragoso et al., 2005). Out of four \textit{CYP1A1} polymorphic variants described so far, only two have been extensively studied in relation to cancer risk because of their functional relevance with carcinogenesis. Polymorphism for the \textit{CYP1A1}*2A (rs4646903) is thymidine to cytosine substitution at the 3’ end of the non-coding region of the gene, and that for \textit{CYP1A1}*2C (rs1048943) is a point mutation in exon 7 resulting in a substitution of isoleucine with valine. These changes ultimately results in an altered enzyme activity and was shown to be associated with cancer (Raunio et al., 1995; Terry et al., 2003).

\textbf{GSTs} are the phase II family of enzymes which neutralize the reactive metabolic products of phase I detoxification by conjugation mechanism. The detoxification efficiency of GST enzymes is determined by the presence, amount and nature of the isoenzymes coded by \textit{GSTT1}, \textit{GSTM1} and \textit{GSTP1} genes. The allelic polymorphism of \textit{GSTT1} and \textit{GSTM1} are characterized by the deletion of a part of the gene. \textit{GSTP1} polymorphism is a single base pair substitution where adenine is replaced by guanine resulting in an amino acid change in which isoleucine (I105) is replaced by valine (V105) (Coles and Kadlubar, 2003; Watson et al.,
Electrophilic compounds are reported to be detoxified less efficiently in individuals with null genotypes of \textit{GSTT1} and \textit{GSTM1} or variant genotypes of \textit{GSTP1} (Ile/Val and Val/Val) as compared to those with wild type genotype (Bolt and Thier, 2006). The presence of \textit{GSTT1} and \textit{GSTM1} null genotypes have been reported to be associated with increased risk for several cancers including skin, lung, bladder, prostate, colorectal and oral cancers (Gao et al., 2002; Jain et al., 2006). However, several other reports have failed to confirm this association (Buch et al., 2002; Sobti et al., 2008). In fact, \textit{GSTT1} null genotype had been reported to be a protective factor for oral cancer in a central Indian population (Anantharaman et al., 2007). Polymorphic variants of \textit{GSTP1} have also been reported to increase the risk of various cancers (Hirvonen, 1999; Rebbeck, 1997). Previous studies of gene polymorphisms and risk for tobacco-associated cancers have suggested that the polymorphisms in \textit{GSTT1}, \textit{GSTM1} and \textit{GSTP1} increase cancer risk in tobacco consumers (Singh et al., 2008; Soya et al., 2007).

However, the prevalence of polymorphism in \textit{GSTs} genes in oral cancer patients from Northeastern region of India is not well known. The individual difference in susceptibility to chemically induced carcinomas may possibly be attributed to the genetic differences in the activation or detoxification of carcinogens due to polymorphic variants of \textit{GSTs}. In the current study, the association of tobacco, betel quid habits and polymorphism of \textit{GSTT1}, \textit{GSTM1} and \textit{GSTP1} genes with oral cancer was evaluated to find out if this could explain the unusually high prevalence of oral cancer in NE region of India.

\textbf{NAT2} catalyzes the acetyl transfer from acetyl coenzyme A to an aromatic amine, heterocyclic amine or hydrazine compound (Butcher et al., 2002). Thus, this enzyme is involved in the metabolism of several environmental toxins and drugs. Slow or fast acetylation phenotypes are results of sequence variations in the \textit{NAT2} which result in the production of NAT proteins with variable enzyme activity or stability (Brockton et al., 2000; Siraj et al., 2008). The \textit{NAT2} acetylation polymorphism is very important in clinical toxicology because it plays a major role in activation and/or deactivation of a large number of aromatic amines and hydrazine compounds (Hein et al., 2000). The \textit{NAT2} alleles described so far may contain up to four of the 10 reported mutations. It has been observed that some mutations consistently reduce acetylation activity (e.g. T$^{341}$C). The functional effect of variant genotypes on detoxification mechanism is due to impairment of the protein translation or stability; messenger RNA levels are not altered (Brockton et al., 2000). Interethnic differences in the frequency of \textit{NAT2} genotypes associated with fast or intermediate
acetylation have been reported so their importance may also vary between different ethnic groups (Anitha and Banerjee, 2003).

**NQO1** formally called DT-diaphorase is a cytosolic enzyme catalyzing a 2 electron reduction. NQO1 can reduce quinone compounds to hydroquinones and prevent their participation in redox cycling and thus in oxidative stress which ultimately protect against the carcinogenicity of quinone compounds (Siraj *et al.*, 2008; Yin *et al.*, 2001). A genetic polymorphism in *NQO1* is a C to T point mutation at bp 609 of exon 6 which codes for proline to serine substitution in the amino acid sequence of the protein which ultimately results in a loss of NQO1 activity (Traver *et al.*, 1997; Yin *et al.*, 2001). The relationship between *NQO1* genetic polymorphism and cancer risk is controversial in different ethnic groups. In addition to the well-documented action in reducing quinone compounds and preventing the formation of reactive oxygen species, NQO enzymes, especially NQO1 also possess other important biological activities. These include anti-inflammatory effects, direct scavenging of superoxide anion radicals, and stabilization of p53 and other tumor suppressors (Zhu and Li, 2011).

**Codon 72 of P53 Gene:**

p53 mutations have been reported to be associated with reduced genomic repair capacity and enhanced cytotoxicity in cells damaged by reactive oxygen species generated during detoxification process or benzo(a)pyrene diol epoxide-DNA adducts (Wani *et al.*, 2000). 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 (Mitra *et al.*, 2005; Papadakis *et al.*, 2000; Rogounovitch *et al.*, 2006; Tandle *et al.*, 2001; Wu *et al.*, 2004). Reports on codon 72 of *p53* gene available from India is limited and inconsistent and the results are conflicting whereas no association was reported with oral cancer (Nagpal *et al.*, 2002; Tandle *et al.*, 2001). However another study showed carriers of Arg/Arg genotype at higher risk for oral cancer (Katiyar *et al.*, 2003).

Studies on codon 72 polymorphism have revealed striking ethnic differences (Sjalander *et al.*, 1995). It has been reported that frequency of *p53* variant allele varies with
latitude, increasing in a linear trend as populations near the equator (Beckman et al., 1994). Thus ethnicity might be related to allelic distribution of the gene and its disease causing effect; however some studies disprove the ethnicity-risk confounding relationship (Fan et al., 2000). North-easter (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. 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 oral cancer in the north eastern region of India prompted us to explore and evaluate any relevance of this polymorphism in this ethnic population.

Early indicators of oral cancer such as leukoplakia and submucous fibrosis are reported to be associated with tobacco and alcohol consumption with a transformation rate of 2% - 12% to frank malignancies (Anantharaman et al., 2007). However not all tobacco and alcohol consumers develop premalignant or frankly malignant diseases of the oral cavity. The role of genetic factors including polymorphism of genes associated with activation and detoxification of toxic compounds is conflicting. The quantitative absorption, distribution, metabolism, and excretion of carcinogenic tobacco constituents depend on the efficiency of metabolic and enzymatic detoxification pathways. Thus study of genes encoding for phase I and phase II detoxifying enzymes along with codon 72 of p53 gene polymorphism may provide some answers and further insights for high incidence of oral cancer in Indian population. In the current study, we have investigated the association between eight polymorphisms [present in CYP1A1 (Msp1 and NcoI), GSTM1, GSTT1, GSTP1, NAT2 and NQO1 genes encoding for xenobiotics metabolizing enzymes and one polymorphism in codon 72 of p53 gene] and oral cancer in a high risk population of Northeast India.
MATERIALS AND METHODS:

Collection of Samples:
Two hundred and thirty five histopathologically confirmed oral squamous cell carcinoma cases and 289 healthy volunteer controls were included in this study from the collaborating center in Northeast India during the period 2006 to 2009. The patients were diagnosed at three different tertiary health facilities of NE India, including Dr. Bhubneshwar Borooah Cancer Institute, Guwahati, Assam; Sir T.N.M. Hospital, Gangtok, Sikkim and Civil Hospital, Aizawl, Mizoram. All study subjects provided informed consent for participation in this research, which was done under a protocol approved by the Institutional Ethics Committee of various institutes as per guidelines. Questionnaires containing information on age, sex, region of origin, occupation, duration and type of tobacco, betel quid and alcohol consumption habits were recorded. Three to 5mL of peripheral blood was collected in tubes containing ethylenediaminetetraacetic acid (EDTA), stored at -70°C freezer and transported under frozen conditions for processing. DNA was isolated using Qiagen Blood DNA Isolation kit (Qiagen GmbH) and stored at -30°C till further analysis. In an earlier study the distribution of wild type genotype among cases and controls is reported as 72.9% and 70.0% respectively with an odds ratio 2.83. (Chatterjee et al., 2009) For testing of hypothesis at this odds ratio in case control study at 5% level of significance with 90% power a minimum of 148 samples are required. The sample size considered in the present study is well above this estimated number.

Isolation of Genomic DNA from peripheral Blood:
Isolation of genomic DNA from blood was carried out using phenol/CHCl₃ method (Sambrook and Russell, 2001).

Reagents used in DNA isolation:
1. Lysis buffer I
   a. Tris-HCl 30 mM (pH 7.4)
   b. EDTA 5 mM (pH 8.0)
   c. NaCl 50 mM
2. Lysis buffer II
   a. NaCl 75 mM
   b. EDTA 2 mM (pH 8.0)
3. Proteinase K (10 mg/ml stock), working concentration (100 µg/ml)
4. SDS 20%
5. Tris saturated Phenol (Tris pH 7.5)
6. Chloroform: Isoamyl alcohol (24:1)
7. Ethanol 70%
8. Absolute ethanol
9. Sodium Acetate 3 M (pH 5.5)
10. TE: Tris 10 mM, EDTA 1 mM (pH 8.0)
11. Tris Borate EDTA (TBE) (pH 8.3) buffer
   For 1L of 10X
   108 g Tris
   55 g Boric acid
   20 ml of 0.5 mM EDTA
   DDW for volume adjustment
12. 6X Gel loading buffer
   0.25% (w/v) bromophenol blue
   0.25% (w/v) xylene cyanol
   30% (v/v) glycerol

**Protocol followed for DNA isolation:**

Three to 4 ml blood was taken from stored and frozen samples in 15 ml centrifuge tube. To this was added equal volume of blood cell lysis buffer I mixed gently and stored at -20°C for 1-2 h. Sample, after removal from -20°C, was immediately transferred to the water bath maintained at 65°C for 10 min. Cells were pelleted at 10,000 rpm for 10 minutes at 15°C. The supernatant was discarded and the lymphocyte cell pellet was suspended in 3 ml of lyses buffer II. Proteinase K (100 µg/ml) and SDS (2%) were added, incubated for 4h at 37°C for complete digestion. To the lysate added equal volume of Tris saturated phenol (pH 8), mixed and centrifuged at 10,000 rpm for 10 minutes at 15°C. The aqueous upper phase (upper layer) containing DNA was carefully transferred to a new tube. Equal volume of Tris saturated phenol (pH 8) and Chloroform: Isoamyl alcohol (24:1), was added, mixed gently and centrifuged at 10,000 rpm for 10 minutes at 15°C. The supernatant having aqueous phase
was transferred carefully to new centrifuge tube. The step was repeated with equal volume of Chloroform: Isoamyl alcohol (24:1). DNA was precipitated from the aqueous phase by adding equal volume of chilled isopropyl alcohol (may add 1/20th volume of 3M sodium acetate, pH 5.5) and pelleted by centrifuging at 10,000 rpm for 15 minutes. The precipitated DNA was spooled out in a micro centrifuge tube and washed with 70% ethanol for removal of traces of salt. It was finally centrifuged at 12,000 rpm for 5 min to obtain a pellet which was vacuum dried and resuspended in 1X TE buffer (100-200 µl) and stored at -20°C.

**Qualitative and Quantitative Estimation of Extracted Genomic DNA:**

The phenol-chloroform extracted genomic DNA from the blood of normal and oral cancer patients were checked for their quality and quantity in an ethidium bromide stained 0.8% agarose gel (0.8 gm agarose in 1X TBE). The DNA was visualized by a UV transilluminator. High quality DNA obtained was evident from the presence of a single intact band without any smearing or degradation (*Figure 4.1*). Concentration of genomic DNA was determined by using nanodrop spectrophotometer, which was in the range of 200-500 ng/µl.

*Fig 4.1: Estimation of the quality of genomic DNA on EtBr stained 0.8% agarose gel. Wells 1-7 indicates genomic DNA isolated from blood samples of patients with oral cancer.*

**GENOTYPING:**

Genotyping for *GSTP1, CYP1A1, NAT2, NQO1* and *codon 72 of p53* genes was done by the PCR-RFLP methods (table 4.1). The PCR products were analyzed in 2.0% agarose gel while fragments of RFLP were analysed in 3.5% agarose gel as discussed before (Blazej-Rubis, 2005; Siraj et al., 2008). Genotyping procedures were validated by reanalyzing 10% of the randomly selected cases and controls. The primer sequences, restriction enzyme and the product sizes are shown in table 4.1.
Table 4.1: Primers and restriction enzymes used and analysis of their products

<table>
<thead>
<tr>
<th>Gene</th>
<th>rs Number</th>
<th>Primer Sequence</th>
<th>Product Length</th>
<th>Rest. Enz.</th>
<th>Genotypes (RFLP fragments)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A1*2A</td>
<td>rs464903</td>
<td>F-5'-CAGTGAAGAGTGTTAGCC GCT-3'</td>
<td>340bp</td>
<td>MspI</td>
<td>WW (340bp) VV (200, and 140bp)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R-5'-TACGTTAGCTGCTGCTGCTG-3'</td>
<td></td>
<td></td>
<td>VV (200, and 140bp)</td>
</tr>
<tr>
<td>CYP1A1*2C</td>
<td>rs1048943</td>
<td>F-5'-GAAGGCTGGTCCAAGCTGCTTGCTG-3'</td>
<td>263bp</td>
<td>NcoI</td>
<td>II (232, and 31bp) IV (263, 232, and 31 bp)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R-5'-CTACCGGATGCTAGGCTAGGCTAGG-3'</td>
<td></td>
<td></td>
<td>VV (263bp)</td>
</tr>
<tr>
<td>GSTT1</td>
<td>NA</td>
<td>F-5'-TTTCCTACTGGTCTACACATCTC-3'</td>
<td>459 bp</td>
<td>NA</td>
<td>Wild type (459bp present) GSTT1 Null (No band)</td>
</tr>
<tr>
<td>GSTM1</td>
<td>NA</td>
<td>F-5'-GAAGGCTGGTCCAAGCTGCTTGCTG-3'</td>
<td>219 bp</td>
<td>NA</td>
<td>Wild type (219 bp present) GSTM1 Null (No band)</td>
</tr>
<tr>
<td>β-Globin</td>
<td>NA</td>
<td>F-5'-CAACTTCATCCAGGTCACC-3'</td>
<td>268 bp</td>
<td>NA</td>
<td>Used as positive internal control for PCR amplification</td>
</tr>
<tr>
<td>GSTP1</td>
<td>rs 947894</td>
<td>F-5'-CCAGTGAACGTGTCTGCTGCTG-3'</td>
<td>189 bp</td>
<td>BsmAI</td>
<td>Ile/Ile (189 bp) Ile/Val (189, 148, and 41 bp) Val/Val (148, and 41 bp)</td>
</tr>
<tr>
<td>NAT2</td>
<td>rs 179930</td>
<td>F-5'-CCAGTGAACGTGTCTGCTGCTG-3'</td>
<td>421 bp</td>
<td>TaqI</td>
<td>GG (170, 139, and 112bp GA (282, 170, 139, and 112bp) AA (282, and 139 bp)</td>
</tr>
<tr>
<td>NQO1</td>
<td>rs 1800566</td>
<td>F-5'-CAACTTCATCCAGGTCACC-3'</td>
<td>237 bp</td>
<td>HinfI</td>
<td>Pro/Pro (188, and 85bp) Pro/Ser (188, 151, 85, and 37bp) Ser/Ser (151, 85, and 37bp)</td>
</tr>
<tr>
<td>Codon 72 of p53</td>
<td>rs 1042522</td>
<td>F-5'-CTGCGTCTCCAGGAAAGCAGGAGATGAC-3'</td>
<td>199bp</td>
<td>BstUI</td>
<td>Pro/Pro (199 bp band) Arg/Pro (199, 113 and 86bp) Arg/Arg (113 and 86bp bands)</td>
</tr>
</tbody>
</table>

W, Wild type; V, Variant type; rs, RefSNP; Ref., Reference; Rest. Enz., Restriction enzyme, NA; Not applicable

**Genotyping of GSTT1 and GSTM1:**

A multiplex PCR method was used to detect the presence or absence of the GSTT1 and GSTM1 genes in the genomic DNA samples of patients and controls. Twenty five µl of PCR mixture was prepared by mixing 2.5µl of 10x Taq buffer, 1µl of 25mM MgCl2, 0.2µl of 25mM dNTP mix, 0.6 µl of each forward and reverse primers (10 pM), 100 - 200 ng of template DNA and 1.5 unit of Taq polymerase (M/s Fermentas, Vilnius, Lithuania) (Table 4.2). The primers were synthesized from M/s Microsynth, Lindau, Germany. Primer pairs were 5'-TTTCCTACTGGTCTACACATCTC-3' and 5'-TAAGGGCTGGTCCAAGCTGCTTGCTG-3' for GSTT1, 5'-GAACTCCCTGAAAAGCTGCTGCTG-3' and 5'-CAACTTCATCCAGGTCACC-3' for GSTM1 and 5'-TAAGGGCTGGTCCAAGCTGCTTGCTG-3' and 5'-GAACTCCCTGAAAAGCTGCTGCTG-3' for β-globin. β-globin (268 bp fragment) was used as an internal control to ensure PCR amplification if the samples had null genotypes of GSTM1 and GSTT1. To test for contamination of genomic DNA from other sources during the experiment, negative controls (PCR master mix without template) were included in every PCR run. PCR was done by denaturation at 94°C for 4 min, followed by 20 cycles of
denaturation at 93°C for 1 min; annealing at 60°C for 1 min; extension at 72°C for 1 min with additional 15 cycles of denaturation at 93°C for 1 min; annealing at 50°C for 1 min; 72°C for 1 min and a final extension at 72°C for 10 min. The PCR products were electrophoresed in 2% agarose gels containing ethidium bromide, prepared and run in 0.6X TBE buffer. The absence of 459 bp band indicated GSTT1 null genotype and the absence of 219 bp band indicated GSTM1 null genotype (Figure 4.2). Approximately 10% of samples were randomly selected and repeated for genotyping.

**Table 4.2: Reaction components of multiplex PCR for GSTT1 and GSTM1 genotyping**

<table>
<thead>
<tr>
<th>Components</th>
<th>Concentrations</th>
<th>Final Volume (for 25 µl rxn)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stock</td>
<td>Working Conc.</td>
</tr>
<tr>
<td>10 X buffer</td>
<td>10X</td>
<td>1X</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>25mM</td>
<td>1.0mM</td>
</tr>
<tr>
<td>dNTPs</td>
<td>25mM</td>
<td>0.2mM</td>
</tr>
<tr>
<td>GSTT1 (Forward Primer)</td>
<td>10pm/µl</td>
<td>0.2pm/µl</td>
</tr>
<tr>
<td>GSTT1 (Reverse Primer)</td>
<td>10pm/µl</td>
<td>0.2pm/µl</td>
</tr>
<tr>
<td>GSTM1 (Forward Primer)</td>
<td>10pm/µl</td>
<td>0.2pm/µl</td>
</tr>
<tr>
<td>GSTM1 (Reverse Primer)</td>
<td>10pm/µl</td>
<td>0.2pm/µl</td>
</tr>
<tr>
<td>β-Globin (Forward Primer)</td>
<td>10pm/µl</td>
<td>0.2pm/µl</td>
</tr>
<tr>
<td>β-Globin (Reverse Primer)</td>
<td>10pm/µl</td>
<td>0.2pm/µl</td>
</tr>
<tr>
<td>Taq Polymerase</td>
<td>5units/µl</td>
<td>1.5 units/rxn</td>
</tr>
<tr>
<td>DNA Template</td>
<td>100-200 ng/µl</td>
<td>-</td>
</tr>
<tr>
<td>MQ H₂O</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig 4.2:** 2% Agarose gel showing multiplex PCR for GSTT1 and GSTM1 in samples of patients with oral cancer. Lane 1: 100 bp DNA ladder; Lanes 2, 3: Cases positive for both GSTT1 and GSTM1 genotypes; Lanes 4, 5: Cases Null for GSTM1; Lane 6: Cases Null for both GSTT1 and GSTM1.
Genotyping of GSTP1:

Polymorphic variants of GSTP1 were detected by PCR-Restriction Fragment Length Polymorphism (RFLP). Twenty five µl of PCR mixture was prepared by mixing 2.5µl of 10x Taq buffer, 2 µl of 25 mM MgCl₂, 1.25µl of 10 mM dNTP mix, 1.25µl of each forward (5'-CCAGTGACTGTGTTGATC-3') and reverse (5'-CAACCCTGGTGAGATGCTC-3') primers (10 pM) for GSTP1, 50-100 ng of template DNA and 1 Unit of Taq Polymerase. Cycling conditions were: initial denaturation at 94°C for 3 min; followed by 35 cycles of 94°C for 1 min, 58°C for 30 sec, 72°C for 30 sec, and a final extension at 72°C for 10 min. The PCR product of GSTP1 was 189 bp in size (Figure: 4.3A). After testing for the amplification of PCR products (in 2% agarose gel, ten µl of PCR product was digested using BsmA1 restriction enzyme (M/s Fermentas, Vilnius, Lithuania) in a reaction volume of 30 µl by overnight incubation at 37°C. The products were separated by electrophoresis in 3.5 % agarose gel in 0.6X TBE. Based on the band pattern, three genotypic variants were identified. The wild type genotype [Ile/Ile (A/A)], completely undigested was represented by a single band at 189 bp. The genotypic variant [Val/Val (G/G)], was completely digested and yielded two bands of 148 bp and 41 bp. The digested product that yielded all the three bands represented the heterozygous genotype [Ile/Val (A/G)] (Figure: 4.3B). A positive known control sample that had earlier been identified as Val/Val (G/G) variant of GSTP1 was included in all experiments. Approximately 10% of samples were randomly selected and repeated for genotyping.
Genotyping of CYP1A1*2A and CYP1A1*2C

The CYP1A1 Msp1 polymorphism was detected by amplification with primers, 5'-CAG TGA AGA GGT GTA GCC GCT -3' and 5'- TAG GAG TCT TGT CTC ATG CCT-3', followed by restriction enzyme analysis with Msp1 according to manufacturer’s instructions. Homozygous wild-type (WW) produced a 340 bp band; homozygote variant (VV) produced two bands of 200 bp and 140 bp, while heterozygote (WV) samples exhibited a digestion pattern of all three bands (Figure 4.4A and Figure 4.4B).

CYP1A1 C2455 A>G gene polymorphism was determined using primers, 5'-GAA AGG CTG GGT CCA CCC TCT-3' and 5'-CCA GGA AGA GAA AGA CCT CCC AGC
GGG CCA-3', followed by restriction enzyme analysis with NcoI according to manufacturer's instructions. The wild-type allele (Ile) was identified by the presence of 263 bp fragment, whereas bands of 232 bp and 31 bp represented the variant allele (Val). Heterozygous samples (Ile/Val) showed both of the three fragments of 263 bp and 232 bp (Figure 4.5A and Figure 4.5B).

---

**Figure 4.5A**: 2% agarose gel showing PCR product of CYP1A1*2A. Lane 1: 100 bp DNA Ladder; Lanes 2-8: Bands of PCR of CYP1A1*2A for blood genomic DNA samples of patients with oral cancer.

**Figure 4.5B**: 3.5% agarose gel showing RFLP fragments of CYP1A1*2A gene using restriction enzyme MspI. Lane 1: 50 bp DNA Ladder; Lanes 5, 6, 10, 11, 12: Wild type genotypes (WW); Lanes 2, 3, 7, 8, 9, 13: Heterozygous genotypes (WV); Lane 4: Homozygous variant genotypes (VV).
Role of SNPs in Xenobiotic Metabolizing Genes and P53 codon 72 in Oral Cancer

Fig 4.5A: 2% agarose gel showing PCR product of CYP1A1*2C. Lane 1: 100 bp DNA Ladder; Lanes 2-6: Bands of PCR product of CYP1A1*2C for blood genomic DNA samples of oral cancer from Northeast region of India.

Fig 4.5B: 3.5% agarose gel showing RFLP fragments of CYP1A1*2C gene, using restriction enzyme Nco1. Lane 1: 100 bp DNA Ladder; Lanes 2-13: Wild type genotypes (II); Lanes 3, 12: Heterozygous genotypes (IV).
Genotyping of NAT2

NAT2 was amplified using primers, 5'-CCT GGA CCA AAT CAG GAG AG-3' and 5'-ACA CAA GGG TTT ATT TTG TTC C-3', followed with restriction enzyme analysis with TaqI according to the manufacturer’s instructions. The NAT2 wild-type genotypes (GG) produced three fragments of 170 bp, 139 bp and 112 bp, while homozygote mutation (AA) was identified by the presence of two fragments of 282 bp and 139 bp, and heterozygote (GA) samples showed presence of all four fragments (Figure 4.6 A and Figure 4.6B.)

Fig 4.6A: 2% agarose gel showing PCR product of NAT2 Lane 1: 100bp Ladder; Lanes 2-6: PCR products of NAT2 gene for blood genomic DNA samples of patients with oral cancer from Northeast region of India.

Fig 4.6B: 3.5% agarose gel showing RFLP fragments of NAT2 gene using restriction enzyme Taq1. Lane 1: 100bp Ladder; Lanes 4, 9: Wild type genotypes (GG); Lanes 2, 3, 6, 8: Heterozygous genotypes (GA); Lanes 5, 7: Homozygous variant genotypes (AA)
Genotyping of NQO1

Genotyping of NQO1 polymorphism was performed using PCR amplification with the primers, 5'-AGT GGC ATT CTG CAT TTC TGT G-3' and 5'-GAT GGA CTT GCC CAA GTG ATG-3' followed with restriction enzyme analysis with Hinfl according to the manufacturer’s instructions. The wild type (Pro/Pro) is identified by the presence of two bands, a 188 bp and 85 bp band, NQO1 homozygote mutation (Ser/Ser) carries a C-T substitution to form a Hinfl site and is identified by the presence of three bands of 151 bp, 85 bp and 37 bp, while NQO1 heterozygote (Pro/Ser) has both alleles and is identified by the presence of all four bands when digested with Hinfl. (Figure 4.7A and Figure 4.7B.)
Genotyping of Codon 72 of P53 Gene:

Each PCR reaction mixture (20µl) contained 0.2µM of each primer, 1.5 mM MgCl2, 0.2mM each dNTP, 0.75 unit of Taq polymerase and 500ng of genomic DNA. Reaction mixtures were preincubated for 10 min at 94°C. PCR conditions were 94°C for 45s and 60°C for 45s, followed by 72°C for 45s for 40 rounds. 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. The digested products were electrophoresed through a 2.5% agarose gel and stained with ethidium bromide. The Genotyping of 10% of the randomly selected cases and controls were repeated for confirmation. Representative genotypes were sequenced. No discrepancies were observed. Arg/Arg (wild genotype) resulted in the presence of two bands of 113 bp and 86 bp each and the Pro/Pro (homozygous mutant genotype) resulted in a single uncut band of 199 bp. PCR results were evaluated without knowledge of case and control status of the sample.

Fig 4.8A: 2% agarose gel showing amplified product of codon 72 of p53 gene in oral cancer patients. Lane 1: 50bp Ladder; Lanes 2-8: bands of PCR product of 199 bp

Fig 4.8 B: 3.5% agarose gel showing RFLP fragments of codon 72 of p53 gene using restriction enzyme BstUI. Lane 1 represents 50 bp ladder; Lane 2,3,8,9,11,12,13,14,15 and 16: Arg/Pro; Lane 4,5,10,17,19, 20: Arg/Arg; Lane 6,7, 18: Pro/Pro
STATISTICAL ANALYSIS:

Cases in each group were individually matched with controls (± 5yrs), sex and ethnicity. Hardy-Weinberg equilibrium (HWE) test was done to compare the difference between the observed and expected frequencies of genotypes. The risk for oral cancer with tobacco use (ever/never), chewing (ever/never), smoking (ever/never), alcohol consumption (alcoholics/non-alcoholics) and the polymorphisms of CYP1A1, GSTP1, NAT2, NQO1, codon 72 of p53 (wild type/variant) and GSTT1 (present/null), GSTM1 (present/null) were estimated by applying conditional logistic regression analysis and the results were interpreted in terms of adjusted odds ratios (AORs) and 95% confidence intervals (95% CIs). The adjusted estimates for the specific variable (tobacco/chewing/smoking/alcohol) were in relation to other variables as considered for analysis in the regression model. Table 4.6 and table 4.7, depicts the results of the dominant genetic model adjusted for individual level characteristics (tobacco/chewing/smoking/alcohol). A two sided p<0.05 was considered statistically significant. Data was analyzed using STATA 10 version software.
RESULTS:

Oral cancer was found more often in males [181/235 (77%)], hence a similar frequency of matched male controls [220/289 (76%)] were included in this study. The mean age ± standard deviation was 54.9 ± 12 for cancer cases and 56.2 ± 12 for matched controls. Frequency of smokers was significantly higher in cases (62%) than in controls (50%) and individuals who smoked were at approximately two fold higher risk (AOR=1.57, 95% CI=1.05-2.35, p=0.027) for developing oral cancer. Frequency of tobacco chewing was also significantly higher in oral cancer cases (74%) compared to controls (50%) and conferred a three fold risk for the development of oral cancer (AOR=2.78, 95% CI=1.71-4.51, p=<0.001). Although the frequency of betel quid chewers and alcohol consumers were higher in cases (80% and 37% respectively) compared with controls (66% and 30% respectively), but the difference was not found statistically significant (Table 4.3 and Table 4.5).

The distribution of \textit{CYP1A1*2A}, \textit{CYP1A1*2C}, \textit{NAT2} and \textit{NQO1} genotypes between cases and controls is shown in table 4.4. Deviation from Hardy-Weinberg equilibrium (HWE) was seen in frequency distribution of \textit{NQO1} genotypes in cases ($\chi^2=14.43$, p=<0.01) and controls ($\chi^2=29.81$, p=<0.01) and in controls of \textit{NAT2} genotypes ($\chi^2=7.87$, p=0.03). However all other genotypes were in HWE for both cases and controls.

Frequencies of WW, WV and VV genotypes of \textit{CYP1A1*2A} gene were 34%, 50% and 16% in cases and 38%, 45% and 17% in controls. Frequencies of Ile/Ile, Ile/Val and Val/Val genotypes of \textit{CYP1A1*2C} gene were 69%, 29% and 2% in cases and 67%, 30% and 3% in controls. No significant difference was observed for any of the two genetic polymorphisms between cases and controls (Table 4.6). However when we analysed samples geographically, percentage of homozygous variant genotypes (VV) was significantly higher in cases (42%) as compared to controls (17%) in samples obtained from Sikkim (Table 4.4). Thus CYP1A1 homozygous variant genotypes were found to impart six fold risk (AOR=6.38, 95% CI=1.10-40.83, p=<0.05) for oral cancer development in Sikkim population (Table 4.6).

Frequency distribution of \textit{NAT2} genotypes GG, GA and AA was 39 %, 47% and 14% in cases and 40%, 51% and 9% respectively in controls (Table 4.4). Variant genotypes of \textit{NAT2} were not found to impart risk for oral cancer development in northeast region of India (Table 4.6).
Frequency distribution of \textit{NQO1} genotypes Pro/Pro, Pro/Ser and Ser/Ser was 45%, 35% and 20% in cases and 60%, 27% and 13% in controls respectively (\textit{Table 4.4}). The heterozygous (Pro/Ser) as well as homozygous (Ser/Ser) variants of \textit{NQO1} imparted significant risk for oral cancer (AOR=1.64, 95% CI=1.04-2.58, p=0.03; AOR=1.81, 95% CI=0.98-3.32, p=0.06 respectively). When analysis was done geographically (centrewise), homozygous variant genotypes of \textit{NQO1} were found to impart three fold risk in Guwahati (Assam) population (AOR=2.57, 95% CI=1.14-5.81, p=0.024 for variant genotypes) (\textit{Table 4.6}).

The frequency of \textit{GSTT1} and \textit{GSTM1} null genotype was 34% (79/235) and 47% (110/235) in samples obtained from patients with oral cancer and 29% (84/289) and 42% (122/289) in controls respectively (\textit{Table 4.4}). When adjusted for other variables under consideration no significant association was found for \textit{GSTM1} and \textit{GSTT1} null genotype independently or in combination with oral cancer risk. However when analyzed geographically, GSTT1 null genotypes was found to create two fold risk in Assam population (AOR=2.04, 95% CI=1.07-3.87, p=0.029) (\textit{Table 4.7}).

The frequency of the variant genotypes of \textit{GSTP1} (heterozygous \textit{AG} and homozygous \textit{GG}) was higher in samples of patients with oral cancer (45%; 105/235) as compared to those with controls (40%; 115/289) (\textit{Table 4.4}). Homozygous variant genotypes of \textit{GSTP1} were found to impart risk (OR=2.91, 95% CI 1.19-7.08, p=0.02) for the occurrence of oral cancer (\textit{Table 4.7}).

Frequency distribution of \textit{p53} genotypes Arg/Arg, Arg/Pro and Pro/Pro was 25.9%, 54.3%, 19.8% in cases and 22.7%, 51.4%, 25.9% in controls. The heterozygous genotype was associated with higher risk for oral cancer (OR$_2$= 1.04, 95% CI=0.57-1.89; p=0.89) whereas Pro/Pro genotype appeared to be a protective factor (OR$_2$= 0.82, 95% CI= 0.40-1.68; p=0.60), however both of these results were not significant statistically (\textit{Table 4.8}).
### Table 4.3: Demographic distribution of various life style factors in oral cancer and control population

<table>
<thead>
<tr>
<th>Factors</th>
<th>NE Combined</th>
<th>Guwahati</th>
<th>Sikkim</th>
<th>Aizawl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls N=289</td>
<td>Cases N=235</td>
<td>Controls N=169</td>
<td>Cases N=160</td>
</tr>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Male</td>
<td>220 (76)</td>
<td>181 (77)</td>
<td>139 (82)</td>
<td>130 (81)</td>
</tr>
<tr>
<td>Female</td>
<td>69 (24)</td>
<td>54 (23)</td>
<td>30 (18)</td>
<td>30 (19)</td>
</tr>
<tr>
<td>Mean age ±SD (Age Range)</td>
<td>56.2±12 (21-85yrs)</td>
<td>54.9±12 (18-85yrs)</td>
<td>52.9±12 (21-80yrs)</td>
<td>54.5±12 (18-85yrs)</td>
</tr>
<tr>
<td>Tobacco Smoker</td>
<td>145 (50)</td>
<td>145 (62)</td>
<td>83 (49)</td>
<td>96 (60)</td>
</tr>
<tr>
<td>Tobacco Chewers</td>
<td>145 (50)</td>
<td>174 (74)</td>
<td>96 (57)</td>
<td>126 (79)</td>
</tr>
<tr>
<td>Betel Quid Chewers</td>
<td>192 (66)</td>
<td>187 (80)</td>
<td>135 (80)</td>
<td>140 (88)</td>
</tr>
<tr>
<td>Alcohol</td>
<td>86 (30)</td>
<td>86 (37)</td>
<td>18 (11)</td>
<td>54 (34)</td>
</tr>
</tbody>
</table>

N, Total number of cases or control in a particular region;
n, number of cases or controls belonging to a particular category
*% of males was higher in case group so similar ratio of male/females in controls was included.
Table 4.4: Demographic distribution of various genetic factors in oral cancer and control population

<table>
<thead>
<tr>
<th>Genes</th>
<th>Genotypes</th>
<th>NE Combined</th>
<th>Guwahati</th>
<th>Sikkim</th>
<th>Aizawl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Controls N=289</td>
<td>Cases N=235</td>
<td>Controls N=169</td>
<td>Cases N=58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>CYP1A12A</td>
<td>WW</td>
<td>110(38)</td>
<td>80(34)</td>
<td>67(40)</td>
<td>65(41)</td>
</tr>
<tr>
<td></td>
<td>WV</td>
<td>130(45)</td>
<td>118(50)</td>
<td>78(46)</td>
<td>79(49)</td>
</tr>
<tr>
<td></td>
<td>VV</td>
<td>49(17)</td>
<td>37(16)</td>
<td>24(14)</td>
<td>16(10)</td>
</tr>
<tr>
<td>CYP1A12C</td>
<td>II</td>
<td>193(67)</td>
<td>163(69)</td>
<td>118(70)</td>
<td>115(72)</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>86(30)</td>
<td>68(29)</td>
<td>49(29)</td>
<td>43(27)</td>
</tr>
<tr>
<td></td>
<td>VV</td>
<td>10(3)</td>
<td>4(2)</td>
<td>2(1)</td>
<td>2(1)</td>
</tr>
<tr>
<td>NAT2</td>
<td>GG</td>
<td>115(40)</td>
<td>91(39)</td>
<td>58(34)</td>
<td>58(36)</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>148(51)</td>
<td>110(47)</td>
<td>91(54)</td>
<td>75(47)</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>26(9)</td>
<td>34(14)</td>
<td>20(12)</td>
<td>27(17)</td>
</tr>
<tr>
<td>NQO1</td>
<td>Pro/Pro</td>
<td>173(60)</td>
<td>105(45)</td>
<td>96(57)</td>
<td>65(41)</td>
</tr>
<tr>
<td></td>
<td>Pro/Ser</td>
<td>77(27)</td>
<td>83(35)</td>
<td>54(32)</td>
<td>60(37)</td>
</tr>
<tr>
<td></td>
<td>Ser/ser</td>
<td>39(13)</td>
<td>47(20)</td>
<td>19(11)</td>
<td>35(22)</td>
</tr>
<tr>
<td>GSTT1</td>
<td>Null</td>
<td>84(29)</td>
<td>79(34)</td>
<td>24(14)</td>
<td>48(30)</td>
</tr>
<tr>
<td>GSTM1</td>
<td>Null</td>
<td>122(42)</td>
<td>110(47)</td>
<td>56(33)</td>
<td>69(43)</td>
</tr>
<tr>
<td>GSTP1</td>
<td>AA</td>
<td>174(60)</td>
<td>130(55)</td>
<td>101(60)</td>
<td>87(54)</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>100(35)</td>
<td>88(38)</td>
<td>60(35)</td>
<td>61(38)</td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>15(5)</td>
<td>17(7)</td>
<td>8(5)</td>
<td>12(8)</td>
</tr>
</tbody>
</table>

N, Total number of cases or control in a particular region;
n, number of cases or controls belonging to a particular category
Table 4.5: Risk estimates for various lifestyle risk factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>OR</th>
<th>Combined NE Region</th>
<th>Guwahati</th>
<th>Sikkim</th>
<th>Aizawl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(95% CI) p Value</td>
<td>(95% CI) p Value</td>
<td>(95% CI) p Value</td>
<td>(95% CI) p Value</td>
</tr>
<tr>
<td><strong>T. Smoking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonsmokers</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Smokers</td>
<td></td>
<td><strong>1.56 (1.11-2.20)</strong> p=0.01*</td>
<td><strong>1.49 (0.97-2.28)</strong> p=0.067*</td>
<td>1.62 (0.75-3.49) p=0.22</td>
<td>1.62 (0.63-4.14) p=0.31</td>
</tr>
<tr>
<td>AOR</td>
<td></td>
<td><strong>1.57 (1.05-2.35)</strong> p=0.027**</td>
<td>1.51 (0.92-2.50) p=0.10</td>
<td>0.86 (0.27-2.71) p=0.79</td>
<td>1.63 (0.23-11.55) p=0.62</td>
</tr>
<tr>
<td><strong>T. Chewing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non chewers</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chewers</td>
<td></td>
<td><strong>2.97 (1.98-4.47)</strong> p=&lt;0.001*</td>
<td><strong>2.87 (1.70-4.85)</strong> p=&lt;0.001*</td>
<td><strong>4.78 (1.63-13.97)</strong> p=0.004*</td>
<td>2.15 (0.82-5.58) p=0.12</td>
</tr>
<tr>
<td>AOR</td>
<td></td>
<td><strong>2.78 (1.71-4.51)</strong> p=&lt;0.001**</td>
<td><strong>2.84 (1.44-5.58)</strong> p=0.002**</td>
<td>3.03 (0.58-15.66) p=0.87</td>
<td>5.09 (0.73-35.51) p=0.10</td>
</tr>
<tr>
<td><strong>B.Q. Chewing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non chewers</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chewers</td>
<td></td>
<td><strong>1.93 (1.29-2.89)</strong> p=0.001*</td>
<td>1.7 (0.94-3.22) p=0.075</td>
<td>4.72 (1.56-14.26) p=0.006</td>
<td>1.13 (0.45-2.83) p=0.78</td>
</tr>
<tr>
<td>AOR</td>
<td></td>
<td>1.50 (0.92-2.45) p=0.1</td>
<td>1.40 (0.62-3.14) p=0.41</td>
<td>3.16 (0.48-20.77) p=0.23</td>
<td>0.76 (0.13-4.53) p=0.77</td>
</tr>
<tr>
<td><strong>Alcohol</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non Drinkers</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Drinkers</td>
<td></td>
<td>1.38 (0.95-2.01) p=0.09</td>
<td>0.92 (0.56-1.47) p=0.76</td>
<td>1.54 (0.66-3.60) p=0.32</td>
<td><strong>5.39 (1.76-16.53)</strong> p=0.003*</td>
</tr>
<tr>
<td>AOR</td>
<td></td>
<td>0.88 (0.56-1.38) p=0.57</td>
<td>0.73 (0.40-1.35) p=0.32</td>
<td>0.66 (0.15-2.81) p=0.57</td>
<td>2.21 (0.13-36.52) p=0.58</td>
</tr>
</tbody>
</table>

**OR:** Crude odds ratio  
**AOR:** Adjusted odds ratio (Adjusted for age, sex and all other risk factors under consideration including polymorphic variant genotypes)  
* Significant association (Highlighted in bold)  
** Remained significant after p-value adjustment for multiple comparisons (Sidak Correction) – (Bold & Italicized values)
Table 4.6: Risk estimates for polymorphisms of CYP1A12A, CYP1A12C, NAT2 and NQO1 genes with oral cancer

<table>
<thead>
<tr>
<th>Genotype</th>
<th>OR</th>
<th>Combined NE Region</th>
<th>Guwahati</th>
<th>Sikkim</th>
<th>Aizawl</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(95% CI) p Value</td>
<td>(95% CI) p Value</td>
<td>(95% CI) p Value</td>
<td>(95% CI) p Value</td>
</tr>
<tr>
<td>CYP1A12A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WW</td>
<td></td>
<td>OR (0.89-2.02) p=0.15</td>
<td>1.07 (0.66-1.73) p=0.79</td>
<td>2.72 (0.84-8.85) p=0.096</td>
<td>2.27 (0.79-6.54) p=0.13</td>
</tr>
<tr>
<td>WV</td>
<td></td>
<td>1.35 (0.92-2.43) p=0.11</td>
<td>0.98 (0.54-1.79) p=0.95</td>
<td>4.55 (0.88-23.36) p=0.07</td>
<td>2.26 (0.33-15.35) p=0.40</td>
</tr>
<tr>
<td>VV</td>
<td></td>
<td>1.12 (0.67-1.86) p=0.67</td>
<td>0.64 (0.30-1.38) p=0.26</td>
<td>5.97 (1.58-22.58) p=0.008</td>
<td>0.43 (0.11-1.68) p=0.22</td>
</tr>
<tr>
<td>WV+VV</td>
<td></td>
<td>1.24 (0.87-1.86) p=0.21</td>
<td>0.96 (0.61-1.53) p=0.87</td>
<td>3.59 (1.17-10.98) p=0.025</td>
<td>1.23 (0.50-3.03) p=0.65</td>
</tr>
<tr>
<td>CYP1A12C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td>1.00 (0.68-1.48) p=0.98</td>
<td>0.86 (0.53-1.39) p=0.54</td>
<td>0.81 (0.37-1.80) p=0.61</td>
<td>0.97 (0.33-2.91) p=0.96</td>
</tr>
<tr>
<td>VV</td>
<td></td>
<td>0.94 (0.59-1.50) p=0.79</td>
<td>0.79 (0.43-1.46) p=0.45</td>
<td>0.63 (0.19-2.14) p=0.46</td>
<td>3.08 (0.43-21.89) p=0.26</td>
</tr>
<tr>
<td>IV+VV</td>
<td></td>
<td>0.96 (0.66-1.42) p=0.86</td>
<td>0.86 (0.54-1.39) p=0.55</td>
<td>0.79 (0.37-1.69) p=0.54</td>
<td>0.80 (0.28-3.23) p=0.69</td>
</tr>
<tr>
<td>NAT2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td></td>
<td>0.96 (0.67-1.39) p=0.83</td>
<td>0.83 (0.52-1.33) p=0.44</td>
<td>1.07 (0.48-2.39) p=0.87</td>
<td>1.42 (0.48-4.18) p=0.53</td>
</tr>
<tr>
<td>GA</td>
<td></td>
<td>0.93 (0.61-1.43) p=0.74</td>
<td>0.99 (0.56-1.76) p=0.97</td>
<td>1.23 (0.35-4.35) p=0.75</td>
<td>3.00 (0.24-37.89) p=0.39</td>
</tr>
<tr>
<td>AA</td>
<td></td>
<td>1.83 (1.01-3.31) p=0.04</td>
<td>1.30 (0.66-2.52) p=0.44</td>
<td>0.81 (0.07-9.52) p=0.87</td>
<td>3.49 (0.76-16.11) p=0.11</td>
</tr>
<tr>
<td>GA+AA</td>
<td></td>
<td>1.61 (0.81-3.20) p=0.18</td>
<td>1.22 (0.56-2.66) p=0.62</td>
<td>1.35 (0.01-211) p=0.90</td>
<td>2.77 (0.17-45.68) p=0.48</td>
</tr>
<tr>
<td>NQO1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro/Pro</td>
<td></td>
<td>1.67 (1.12-2.49) p=0.012</td>
<td>1.63 (1.00-2.65) p=0.049</td>
<td>1.51 (0.43-5.36) p=0.52</td>
<td>2.51 (0.93-6.79) p=0.07</td>
</tr>
<tr>
<td>Pro/Ser</td>
<td></td>
<td>1.64 (1.04-2.58) p=0.03</td>
<td>1.36 (0.75-2.44) p=0.31</td>
<td>1.76 (0.28-11.16) p=0.55</td>
<td>3.15 (0.55-18.11) p=0.19</td>
</tr>
<tr>
<td>Ser/Ser</td>
<td></td>
<td>1.84 (1.08-3.13) p=0.025</td>
<td>2.83 (1.45-5.52) p=0.002</td>
<td>1.20 (0.45-3.23) p=0.71</td>
<td>lesser count - NA</td>
</tr>
<tr>
<td>Pro/Ser+Ser/Ser</td>
<td>1.81 (0.98-3.32) p=0.06</td>
<td>2.57 (1.14-5.81) p=0.024</td>
<td>1.56 (0.30-8.22) p=0.59</td>
<td>lesser count - NA</td>
<td></td>
</tr>
<tr>
<td>Pro/Ser+Ser/Ser</td>
<td>1.72 (1.20-2.47) p=0.003</td>
<td>1.91 (1.22-2.99) p=0.005</td>
<td>1.30 (0.54-3.10) p=0.55</td>
<td>1.85 (0.74-4.62) p=0.19</td>
<td></td>
</tr>
<tr>
<td>AOR</td>
<td></td>
<td>1.55 (0.96-2.28) p=0.025</td>
<td>1.61 (0.98-2.64) p=0.05</td>
<td>1.74 (0.63-4.85) p=0.28</td>
<td>2.04 (0.68-6.16) p=0.20</td>
</tr>
</tbody>
</table>

Note - Abbreviation for significance and OR are the same as those mentioned under footnotes for table 4.3
Table 4.7: Risk estimates for polymorphisms of GSTT1, GSTM1 and GSTP1 genes with oral cancer

<table>
<thead>
<tr>
<th>Gene</th>
<th>Combined NE Region</th>
<th>Guwahati</th>
<th>Sikkim</th>
<th>Aizawl</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSTT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>OR</td>
<td>OR (95% CI) p Value</td>
<td>OR (95% CI) p Value</td>
<td>OR (95% CI) p Value</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Null</td>
<td>OR</td>
<td>1.25 (0.87-1.79) p=0.23</td>
<td>2.41 (1.40-4.14) p=0.002</td>
<td>0.72 (0.29-1.81) p=0.49</td>
</tr>
<tr>
<td></td>
<td>AOR</td>
<td>1.28 (0.84-1.95) p=0.25</td>
<td>2.04 (1.07-3.87) p=0.029</td>
<td>0.76 (0.17-3.47) p=0.73</td>
</tr>
<tr>
<td>GSTM1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>OR</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Null</td>
<td>OR</td>
<td>1.22 (0.86-1.73) p=0.25</td>
<td>1.60 (1.01-2.55) p=0.047</td>
<td>0.83 (0.39-1.76) p=0.64</td>
</tr>
<tr>
<td></td>
<td>AOR</td>
<td>1.18 (0.80-1.76) p=0.40</td>
<td>1.60 (0.90-2.89) p=0.11</td>
<td>0.87 (0.28-2.82) p=0.84</td>
</tr>
<tr>
<td>GSTP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>OR</td>
<td>1.13 (0.79-1.64) p=0.49</td>
<td>1.21 (0.76-1.93) p=0.42</td>
<td>0.75 (0.32-1.77) p=0.51</td>
</tr>
<tr>
<td></td>
<td>AOR</td>
<td>1.14 (0.74-1.74) p=0.55</td>
<td>1.31 (0.74-2.30) p=0.35</td>
<td>1.56 (0.44-5.56) p=0.49</td>
</tr>
<tr>
<td>GG</td>
<td>OR</td>
<td>1.60 (0.75-3.43) p=0.22</td>
<td>1.87 (0.72-4.88) p=0.20</td>
<td>0.60 (0.08-4.22) p=0.61</td>
</tr>
<tr>
<td></td>
<td>AOR</td>
<td>2.91 (1.19-7.08) p=0.02</td>
<td>3.14 (0.94-10.49) p=0.06</td>
<td>1.81 (0.12-28.54) p=0.67</td>
</tr>
<tr>
<td>AG+GG</td>
<td>OR</td>
<td>1.19 (0.83-1.69) p=0.34</td>
<td>1.28 (0.82-2.01) p=0.27</td>
<td>0.75 (0.32-1.75) p=0.50</td>
</tr>
<tr>
<td></td>
<td>AOR</td>
<td>1.33 (0.90-1.94) p=0.15</td>
<td>1.32 (0.81-2.15) p=0.26</td>
<td>1.19 (0.41-3.43) p=0.74</td>
</tr>
</tbody>
</table>

OR: Crude odds ratio
AOR: Adjusted odds ratio (Adjusted for age, sex and all other risk factors under consideration including polymorphic variant genotypes)
* Significant association - (Highlighted in Bold)
** Remained significant after p-value adjustment for multiple comparisons (Sidak Correction) – (Bold & Italicized values)
Table 4.8: Frequency distribution and risk estimates for p53 codon 72 polymorphisms in oral cancer

<table>
<thead>
<tr>
<th>p53 Genotypes</th>
<th>Cases</th>
<th>Controls</th>
<th>OR</th>
<th>Estimated risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% ( n/N)</td>
<td>% ( n/N)</td>
<td>OR</td>
<td>OR (95% CI) p Value</td>
</tr>
<tr>
<td>Arg/Arg</td>
<td>25.9% (30/116)</td>
<td>22.7% (63/278)</td>
<td>OR</td>
<td>1.0</td>
</tr>
<tr>
<td>Arg/Pro</td>
<td>54.3% (63/116)</td>
<td>51.4% (143/278)</td>
<td>OR</td>
<td>0.83(0.47-1.45), p=0.52</td>
</tr>
<tr>
<td>Pro/Pro</td>
<td>19.8% (23/116)</td>
<td>25.9% (72/278)</td>
<td>AOR</td>
<td>1.04(0.57-1.89), p=0.89</td>
</tr>
</tbody>
</table>

N= Number of samples included in the study;
n= number of samples possessing a particular genotype

OR: crude odds ratio

AOR: adjusted for tobacco smoking, tobacco chewing, betel quid chewing and alcohol consumption
DISCUSSION:

The intensity and duration of response to exposure of environmental toxins is dependent on its concentration in the target organ, which is determined by the quantitative absorption, distribution, metabolism, and excretion of that toxic compound (Nakajima and Aoyama, 2000). Further, inter-individual effects of toxic compounds are based on pathophysiological factors and environmental interactions as well as on genetic characteristics (Cascorbi, 2006). Of the various factors determined by pharmacokinetics, recent research is focusing on the metabolism of environmental toxins (Nakajima and Aoyama, 2000). Majority of chemical carcinogens require metabolic activation before interacting with cellular macromolecules to initiate cancer (Raunio et al., 1995). It has been found that the inherited differences in the capacity of xenobiotics metabolizing enzymes are important factors in determining the genetic susceptibility to various malignancies (Hung et al., 2005). In our study, we have reported GSTT1 null genotype to be a significant risk factor for the occurrence of oral cancer (Yadav et al., 2010). This finding gave an interesting lead to explain the high prevalence of oral cancer in this region. In current study the association between seven polymorphisms present in six genes [CYP1A1 (Msp1 and Nco1), NAT2, NQO1, GSTT1, GSTM1, and GSTP1] encoding for xenobiotics metabolizing enzymes and risk of oral cancer was analyzed from three different centres of north east India. As the population in NE India has unique betel quid and tobacco consumption habits, we hypothesized that polymorphisms of genes responsible for detoxification of xenobiotics may be associated with the high risk of oral cancer in this region.

The association of oral cancer with tobacco consumption have been well documented (Cho and Purohit, 2006). Chewing tobacco attributes for more than 66% of the total oral cancer cases in India. The surveys carried out by National Family Health Survey (NFHS) and Global Youth Tobacco Survey (GYTS) conducted in 18 states in India have reported an alarmingly high prevalence of chewable tobacco products in younger Indian population (average prevalence of tobacco chewing was 29% compared to 13% for smoking). The highest prevalence of tobacco chewing within India was observed in Northeastern states ranging from 47 - 63%. This is one of the major contrasting features of Indian population compared with western countries where tobacco smoking is more prevalent compared to chewing (Kuruvilla, 2008). Our results indicate that tobacco use in any form; chewing or smoking is a strong risk factor for oral cancer. Moreover our study suggested that the relative risk for oral cancer was higher in tobacco chewers (AOR=2.78, 95% CI=1.71-4.51,
p<=0.001) as compared to tobacco smokers (AOR=1.57, 95% CI=1.05-2.35, p=0.027). This has also been reported in previous studies (Dikshit and Kanhere, 2000; Hamada et al., 1991; Nandakumar et al., 1990; Sankaranarayanan, 1990; Znaor et al., 2003). It is known that approximately 90% of oral cancers are squamous cell carcinomas arising from buccal mucosa, the site which remains exposed directly to tobacco constituents for longer duration while chewing (Kuruvilla, 2008). The longer duration of carcinogenic exposure to buccal mucosa gives higher chances for absorption of tobacco components leading to mutagenic changes in underlying cells. Moreover, the particulate nature of chewable tobacco products may cause local trauma and injury to the buccal mucosa which may further enhance the absorption of tobacco components into the squamous cell lining. This justifies tobacco chewing habit as the major risk factor conferring the highest rates of oral premalignant and evident malignancies in India. There is inter-individual variation in detoxification efficiency of carcinogenic compounds that in turn is dependent on polymorphic forms of detoxifying genes.

Most of the known polymorphisms are functionally neutral, some affect function of the coded protein or the regulation of the gene expression which ultimately may affect some metabolic pathways (Sameer et al., 2010). Polymorphism in the genes that code for detoxification enzymes may alter expression or function thus increasing or decreasing the activation or detoxification of carcinogenic compounds (Olshan et al., 2000). The inherited differences in the capacity of xenobiotics metabolizing enzymes have been found to be an important factor that determines the variable susceptibility to carcinogenesis (Hung et al., 2005).

CYP1A1 is a key enzyme involved in the phase I bio-activation of a wide range of environmental toxins and carcinogens. CYP1A1 gene is expressed in many epithelial tissues especially in buccal mucosa (Sam et al., 2010). The polymorphic sites located in CYP1A1 gene have been reported to be associated with genetic susceptibility to several types of cancers (Fragoso et al., 2005). Several earlier studies that have analyzed the role of CYP1A1 polymorphism with susceptibility to oral carcinoma, have reported inconsistent results (Zhuo et al., 2009). Our results suggest that variant genotypes of CYP1A1 may not be a risk factor for oral cancer in NE Indian population. To our knowledge, the only known study to investigate the role of CYP1A1 polymorphic variants on the same population in patients with oral leukoplakia showed similar results (Chatterjee et al., 2009). A meta-analysis conducted by Zhuo et. al also found no association of CYP1A1 variant genotypes with the development
Role of SNPs in Genes Encoding Phase I and Phase II enzymes in OSCC (Zhuo et al., 2009). When analysis was done geographically variant genotypes of CYP1A1*2A were found to impart risk for oral cancer in patients from Sikkim region.

Lack of GSTT1 and GSTM1 isoenzyme activity or differences in the activity and distribution of allelic variants of GSTP1 have earlier been implicated in increased cancer risk following exposure to environmental carcinogens. Of these, GSTT1 is responsible for the biotransformation of the constituents of tobacco smoke such as alkyl halides and its derivatives such as monohaloethanes, ethylene oxide, benzo(α)pyrene diol epoxide, and acrolein (Pemble et al., 1994; Rebbeck, 1997). GSTM1 subfamily metabolises lipid peroxidation products, DNA hydroperoxides and polyaromatic hydrocarbons (PAH) such as benzo [alpha] pyrene (Jain et al., 2006; Lear et al., 2000; Ye et al., 2004). The GSTP1 enzyme is widely expressed in tumor cells and is responsible for the detoxification of benzo(α)pyrene diol epoxide and acrolein present in cigarette smoke. The GSTP1 isoform is also known to metabolize tobacco related carcinogens with elimination of the oxidative products of thymidine or uracil propenal (Matthias et al., 1998). Polymorphism of the GSTT1 and GSTM1 genes results in deletion of their loci with subsequent loss of specific enzymatic functional activity and reduced ability to detoxify potentially toxic substances. Polymorphism of GSTP1 gene shows a single base pair substitution where adenine is replaced by guanine resulting in aminoacid isoleucine (I105) being replaced by valine (V105) (Coles and Kadlubar, 2003; Watson et al., 1998). Since GST genes are involved in the detoxification of tobacco constituents, there is a possibility that the genetic polymorphisms of these enzymes may be a risk factor for the widespread occurrence of tobacco-associated oral cancer in NE Indians.

Earlier studies from different regions of the world have reported a higher risk for the occurrence of several cancers in patients with GSTT1 and GSTM1 null genotypes. However, many other studies have reported conflicting results. GSTM1 null genotype has been reported as a risk factor for oral cancer (Duarte et al., 2008; Gattas et al., 2006). This is in contrast to other reports where no significant association of GSTM1 null genotype was found with oral cancer risk (Losi-Guembarovski et al., 2008). In fact, there are reports that have shown GSTM1 null genotype as protective factor for oral cancer (Hatagima et al., 2008). GSTT1 null genotype has been reported as a risk factor for oral cancer (Bartsch et al., 1999; Duarte et al., 2008; Jourenkova-Mironova et al., 1999), whereas no significant association of GSTT1 null genotype had been reported with oral cancer risk in other studies (Hatagima et al., 2008; Kietthubthew et al., 2001; Losi-Guembarovski et al., 2008). As reported for GSTM1 null
Role of SNPs in Genes Encoding Phase I and Phase II enzymes in OSCC

GSTT1 null genotype has also been reported as a protective factor for some cancers such as head and neck cancer (Evans et al., 2004), bladder cancer (Kim et al., 2002) and breast cancer (Garcia-Closas et al., 1999).

A review of Indian studies also showed different results for different ethnic groups for association with GST polymorphism. For example, GSTM1 null genotypes were reported as a significant risk for oral cancer in the western Indian population (Anantharaman et al., 2007; Buch et al., 2002), whereas no risk was reported in oral cancer in a north Indian population (Sharma et al., 2006). GSTT1 null genotype was reported to be as risk factor for oral cancer in a north Indian population (Sharma et al., 2006; Singh et al., 2008). However, no significant risk was reported in a study on oral cancer in a western Indian population (Buch et al., 2002). GSTP1 variants have been reported as a risk for oral cancer in an East Indian population (Sikdar et al., 2004). Moreover data from different geographical regions of India show large variation in different ethnic groups in healthy population (Thoudam et al., 2010).

In the current study, GSTT1 null genotype was not found to be associated with risk of oral cancer when NE population was taken as one group. However, analysis of GST polymorphisms in different geographic regions of NE India showed GSTT1 null genotype to be a significant risk factor for oral cancer in Guwahati (Assam) region of NE India. GSTTM1 null genotypes showed significant risk for oral cancer in Assam population but risk vanished when adjusted with other factors. In the current study, the homozygous variant genotypes GSTP1 (Val/Val) were found to be significantly associated with oral cancer when northeast was taken as a single group as well as for Assam region. Several earlier studies have reported conflicting results of GSTP1 polymorphism, with both risk factor (Miller et al., 2003; Sreeja et al., 2008) and no association (Reszka et al., 2003; Sobti et al., 2008) having been reported. In addition epigenetic factors such as hypermethylation of promoter region of GSTP1 gene may lead to downregulated gene expression and reduced activity of the enzyme. Methylation of GSTP1 promoter region has earlier been found to be associated with some cancers, particularly prostate cancers, where it has been used for its early diagnosis and prognosis (Duffy et al., 2009). However no such significant association has so far been reported for oral cancer.

Exposure to the type and amount of environmental toxins is variable not only in different geographic regions, but also in different ethnic groups within the same geographic region due to variations in their dietary, social and cultural habits. Although the samples...
Role of SNPs in Genes Encoding Phase I and Phase II enzymes in OSCC

included in our study belonged to a common geographical region of India, the inhabitants of this region are of different ethnic origin. Since the ethnically different population inhabiting this region of India has presumably been exposed to shared environmental factors such as pesticide exposure and high level of tobacco and betel quid consumption, we have analyzed the data of different racial composition separately as well as a combined group from NE India. The inconsistency in results of association of gene polymorphism with different population groups may be due to different ethnicity or interaction between different environmental and genetic factors. For example, individuals who inherit the GSTT1 enzyme can produce a mutagenic and carcinogenic metabolite of industrial chemical dichloromethane following conjugation with glutathione (Pemble et al., 1994). Although GSTs are enzymes that are synthesized mainly in the liver (Jefferies et al., 2003) however, the localization and concentration of different classes of GSTs in the cytosol of different organs are variable. Moreover, some other properties of GST enzymes, such as affinity towards substrate and isoelectric focusing are also variable (Awasthi et al., 1994). The distribution of GST enzymes in different organs may also vary with the age and sex of different individuals. These factors may lead to variation in the carcinogenic concentration of toxins in different tissues, and a variable role of GST genotypes in different populations exposed to different environmental carcinogens as has been found in our study.

Acetylation is an important route of biotransformation for highly mutagenic and carcinogenic aromatic amines to which humans are routinely exposed via tobacco smoking, cooked foods and other sources (Katoh et al., 1998). Metabolism of carcinogenic aromatic amines is complex and many potential pathways exist, of which \( N \)-hydroxylation by cytochrome P-450 oxidases followed by \( O \)-acetylation by \( N \)-acetyle transferases was suggested as a possible route for activation in oral tissues (Katoh et al., 1998). In earlier studies variant genotypes of \( NAT2 \) were reported to implicate risk for developing oral cancer, (Buch et al., 2008; Marques et al., 2006) while few other studies reported no association (Chen et al., 2001; Hahn et al., 2002). In our study, homozygous variant genotypes of \( NAT2 \) were found to be higher in cases (14%) as compared to controls (9%) but the difference was statistically insignificant (AOR=1.61, 95% CI=0.81-3.20, p=0.18). This disparity in results by different study groups may be explained by interethnic variations in genetic composition and environmental factors.

Another phase II, detoxifying enzyme NQO1 protects cells from oxidative stress and carcinogens present in tobacco by direct scavenging of quinone substrates and inhibiting the
formation of CYP1A1-generated metabolites and subsequent binding to DNA (Sameer et al., 2010). In addition to its well established detoxifying enzymatic function which is critical for maintenance of low concentrations of reactive oxygen species, it is also involved in stabilization of p53 tumour suppressor. It was reported that NQO1 deficient mice show reduced p53 induction and apoptosis, impaired NF-kB function and increased susceptibility to chemically induced tumors (Fagerholm et al., 2008). NF-kB was reported to play a significant role in tobacco associated cancers (Rahlan et al., 2009). NQO1 gene polymorphism at nucleotide 609 results in a diminished NQO1 detoxifying activity and is associated with susceptibility to various cancers such as bladder (Pandith et al., 2011) gastric, (Malik et al., 2011) lung, (Eom et al., 2009; Wenzlaff et al., 2005) and ovarian cancer (Olson et al., 2004). Only few relatively small studies have examined its role in squamous cell carcinoma of head and neck. As far as oral cancer is concerned, we found no literature in pubmed search that has investigated the role of NQO1 polymorphisms in association with oral cancer. In our study NQO1 variant genotypes were found to confer a significant risk for the development of oral cancer in both heterozygous (Pro/Ser, AOR=1.64, 95% CI=1.04-2.58, p=0.03) as well as in homozygous variants (Ser/Ser, AOR=1.81, 95% CI=0.98-3.32, p=0.06). NQO1 genotypes showed deviation from HWE. This may be attributed to the fact that the population in this study comprised of geographically isolated and culturally or religiously non-communicating strata that generally do not intermarry with other caste or religion (Ihsan et al., 2011). Case control matching was done in reference to the age, gender and ethnicity, thereby controlling any confounding effect on account of these variables. When analysis was done geographically, homozygous variant genotypes of NQO1 conferred significant risk (Ser/Ser, AOR=2.57, 95% CI=1.14-5.81, p=0.024) for oral cancer in Assam population only.

P53 is a highly conserved gene with only five polymorphisms are well known 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. The P72 variant is a stronger inducer of transcription, probably owing to its stronger affinity to bind to transcription factors and the A72 variant is considered to be a better inducer of apoptosis, thus suppressing transformation more efficiently, than the P72 variant.
Role of SNPs in Genes Encoding Phase I and Phase II enzymes in OSCC

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 0.40 to 0.78 for Arg and 0.22 to 0.58 for Pro (Ihsan et al., 2011).

Considerable number of studies has reported no difference in distribution of any genotype between cases of oral cancer and controls (Hamel et al., 2000; McWilliams et al., 2000; Shen et al., 2002). In the present study, no significant effect of the polymorphism of codon 72 of $p53$ gene on susceptibility to oral cancer was seen. These findings are concordant with some previous reports spread over different ethnic populations. No association between $p53$ variants and oral cancer were observed in earlier studies by (Drummond et al., 2002; Kietthubthew et al., 2003) in Brazilian and Thai population. However, most of these studies on oral cancer constituted small sample size ranging from 58 to 97. Study on oral cancer available from India also report similar distribution of $p53$ genotypes between cases and control (8%, 72%, 19% and 14% 65%, 20% for Arg/Arg, Arg/Pro and Pro/Pro in cases and controls respectively) (Tandle et al., 2001).

In the current study tobacco chewing as well as smoking found to impart a significant risk for oral cancer with tobacco chewers being at higher risk compared to smokers. Frequency of betel quid chewers and alcohol consumers were also higher in cases compared with controls but the difference was statistically insignificant. However when centrewise risk was investigated alcoholics were at risk for oral cancer in Aizawl population, but this risk vanishes when it is adjusted with other risk factors. Out of eight genetic polymorphisms studied, variant genotypes of $NQO1$ (homozygous as well as heterozygous variants) and homozygous variants of $GSTP1$ were found to impart risk for oral carcinogenesis. However when centerwise analysis was performed, variant genotypes of $NQO1$, $GSTP1$ and $GSTT1$ were found to impart significant risk for oral cancer in Guwahati population while variant genotypes of $CYP1A1*2A$ imparted risk for oral cancer in Sikkim population. Gene enviromental interacrtion analysis showed that the variant genotypes of $NQO1$ did not interact statistically with tobacco consumption habits. Although variant genotypes of $NQO1$ may play an important role in the genetic susceptibility to oral cancer, it’s pathway appears to be unrelated to the detoxification mechanism of tobacco constituents. In our gene expression profiling study $NQO1$ was found to be significantly up-regulated. $NQO1$ has been reported to play a significant role in stabilisation of $p53$ tumour suppressor gene, which may be a possible route of its involvement in oral carcinogenesis.
Chapter 5

Detection of Genetic Instability/ Genomic Alterations using Targeted Next Generation Sequencing in Oral Squamous Cell Carcinoma
INTRODUCTION:

Oral cancer is the sixth most common malignancy in humans, and the mortality rate remains high at approximately 50% with a particularly poor 5-year survival rate which has not improved significantly in the last 40 years. In India oral cancer is the most common cancer among men and ranks third among women (Soya et al., 2007), with age-standardized incidence rates per 100,000 population to be 12.8 and 7.5 respectively (Nair et al., 2004). Northeastern states in India have reported a very high prevalence of aerodigestive tract cancers compared with other regions of India (Bhattacharjee et al., 2006; Phukan et al., 2004). Prevalence of oral cancer is highest in Kamrup district of Assam (ICMR-Report, 2006). The present study is based on the samples collected from high risk population of Assam region of northeast India.

It is evident that genomic aberrations are hallmarks of cancer in which large number of mutational changes are involved in the development and progression of cancer. Genome wide screening can be either quantitative as in the case of gene expression profiling by microarray or qualitative as in the case of mutation detection by next generation sequencing. Altered gene expression profiling may be attributed to several types of genetic alterations such as base changes, insertions, deletions, chromosomal translocations, inversions, loss of heterozygosity and copy number variations. For detection of genetic alterations, sequencing has historically been considered as the golden standard because it reveals the exact location and the type of mutation; however this method was slow and tedious. So to understand genetic alterations in large number of genes, it was highly desirable to develop methods that are precise, fast, cost-effective and can help us in simultaneous analysis of several genetic alterations in a single experiment. Fortunately in recent years we have seen revolution in sequencing methods in the form of next generation sequencing (NGS) technologies with the help of which cancer research will lead towards accelerated discovery of molecular signatures involved in carcinogenesis. Using this technology, the inexpensive production of large volumes of data in a single experiment is the primary advantage over conventional methods. Using next generation sequencing clinically relevant molecular signatures can be discovered which may be of great significance in understanding the underlying biology of cancer,
providing the ability to answer questions with unimaginable speed which may lead us towards identification of molecular targets in carcinogenesis and subsequent development of innovative methods for diagnosis, prognosis and therapy of cancer.

Oral cancer is a multi-factorial disease which is influenced by both aetiological factors and ethnicity as a result molecular profiles of oral cancer vary throughout the world (Ambatipudi et al., 2011; Paterson et al., 1996; Shah and Singh, 2006). The discovery of mutations that are involved in carcinogenesis is a crucial step in cancer research which is tremendously facilitated by NGS technology which may be useful for both targeted and genome-wide screening. For the present study of targeted re-sequencing of 169 functionally relevant genes, Illumina-Solexa platform has been used. The present study is one of the first reports of a genome-wide targeted sequencing of candidate genes using next generation sequencing technology with aim at identifying molecular targets involved in carcinogenesis of OSCC associated with chewing of tobacco and betelquid from India.
MATERIALS AND METHODS:

For the present study of targeted re-sequencing of functionally relevant 169 genes was performed using Illumina-Solexa platform, more specifically on the GAIIx instrument. This platform involves ‘sequencing by synthesis’ approach using reversible dye terminator chemistry. 72 bp singleton sequence reads were generated and data analyzed including alignment, assembly, and variation discovery.

Selection of patients and collection of samples:

Tissue samples from 25 patients undergoing surgical treatment for oral cancer at Dr B. Barooah Cancer Hospital, Guwahati from 2006 to 2009 were included in this study. For these patients, within 10 - 15 minutes of surgical removal the tissue was collected in PBS from the tumour site as well as normal appearing site distant from the tumour area. One part of the tissue sample was used for histopathologic processing. Samples stored in PBS were frozen at -70°C till further processed. All 25 samples had a confirmed histopathologic diagnosis of OSCC. Detailed questionnaire with specific information regarding dietary, smoking, alcohol consumption habits and family history of cancer was completed for all patients. Informed consent was obtained from all the patients to use their specimens and clinicopathologic data for this study. Approval for this study had earlier been obtained from the Institutional Human Ethics Committee.

Twenty five samples that showed good quality of DNA in the tumor tissue as well as in the normal appearing tissue distant from the tumor site were selected for experiments. Only samples from patients with confirmed diagnosis of OSCC who gave history of tobacco and betel quid chewing were selected for mutational analysis of selected genes by next generation sequencing experiments to maintain uniformity of the experimental design.
Detection of Genomic Alteration in OSCC using NGS

Next Generation Sequencing Methodology:

List of Genes, Exons, Genomic regions / Choose from Catalog DNA Capture designs → Capture Microarray (60mer Probes)

Genomic DNA ↓

Fragmentation (Sonication / Shearing) ↓

Repair ends ↓

Ligate adapters

Hybridization ↓

Elution of Captured material ↓

Enrichment analysis by qPCR ↓

Captured DNA ready for Sequencing

Fig 5.1: On-array DNA capture workflow used for next generation sequencing by Illumina

Capture array design:

1 x 244 K Agilent capture array comprising 60 mer tiling probes was designed for exonic regions of interest with covering some flanking regions as well. The repetitive regions were removed by RepeatMasker filtering and unique probes selected to avoid nonspecific binding.

Library construction:

The OSCC DNA samples and normal DNA samples were each pooled in equal concentrations to generate two different sets of pooled samples. 10 micrograms of genomic DNA from each pooled sample was made up to 200 ul with nuclease free water and sonicated using a VibraCell (12 pulses of 10s on and 10s off @ 20% amplitude) to fragment size ranging between 100 to 800 bp (Figure 5.2). The resulting fragmented DNA was cleaned up using QIAquick columns (QIAGEN). The size distribution was checked by running aliquots of the samples on Agilent Bioanalyzer 7500 Nano chips. Subsequently, DNA was subjected to a series of enzymatic reactions that repair frayed ends, phosphorylate the fragments, and add a single nucleotide ‘A’ overhang using Illumina recommended reagents for multiplexed paired end library preparation (Figure 5.3). After ligating Illumina adaptors, ~300 bp fragments were size selected by gel electrophoresis and purified. Multiple PCR amplifications were performed for the ligated products (enrichment PCR) so as to obtain ~10 micrograms of amplicons per sample. Specific barcodes were also introduced during the PCR
Detection of Genomic Alteration in OSCC using NGS

(Oral cancer library was barcoded with Illumina index 5 and Normal library with Illumina index 8).

Fig 5.2: Bioanalyzer profiles of sonicated samples

Fig 5.3: Process of genomic DNA library preparation
Hybridization and Elution:

The OSSC and normal libraries were each pooled in equal amounts to a total of 20 micrograms and hybridized on Agilent 244k Microarray (AMADID: 027271 and AMADID: 027271) following standard protocol recommended by Agilent for 65h at 65°C (Hodges et al., 2009). After standard washing procedures, the slides were reassembled with nuclease free water (Ambion) and exposed to high temperature (95°C for 10 min). DNA eluted in nuclease free water was recovered using a syringe. PCR was carried out with the eluted DNA in several replicates and cleaned up using QIA quick columns (QIAGEN).

Comparison of ePCR1 and ePCR2 products for target regions showed enrichment of target region. Captured samples showed early amplification over ePCR1 samples signifying enrichment. On the other hand the non-target regions were not detected in ePCR2 samples when compared with ePCR1 (Figure 5.4).

Sequencing:

The captured samples were sequenced using Illumina GAIIX Analyzer. Single end 72 base pair reads were generated further quality control was performed using QC tool SeqQC (Genotypic, India). The coverage was found ~25x against the reference sequence (Human Genome 19). Using 20 phred quality for high quality cut off, we found that 88-97% of the bases were of high quality and 88-97% of reads were of high quality across four samples.
Only high quality reads were passed for further analysis. Sequence was downloaded from UCSC (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/).

**Data Analysis:**

![Diagram of Analysis Flow](attachment:analysis_flow.png)

**Fig 5.5: Analysis flow for identifying genomic variations**

Alignment of sequences was done using alignment software BWA version 0.5.7 to perform gapped alignment of Illumina sequences against the reference sequence. With various parameters used for alignment, (Maximum number of gap opens = 2, Maximum number of gap extensions= 10). SNP calling was performed using Samtools version 0.1.11 with following parameters:

- Minimum read depth = 5
- Maximum read depth = 255
- Minimum mapping quality = 20
- Minimum neighboring quality = 20
- Window size around potential indels = 5
- Maximum number of SNPs in a window = 2
Indels were detected using Samtools version 0.1.11 from the gapped alignment performed using BWA. Following were the parameters used for Indel filtering:

- Read depth at Indel bases $\geq 8$
- % of reads representing Indel $\geq 50$

Functional analysis was performed using GenoSNP which assigned SNPs characteristics like synonymous, nonsynonymous, missense, as well as location like 3’UTR, 5’UTR, Intronic. It also predicts the impact of mutation at amino acid level.
RESULTS:

Demographic and clinical characteristics of oral cancer samples:

Twenty five OSCC samples were included in the present study. All patients were male and gave a history of tobacco and betel quid chewing (Table 5.1).

Table 5.1: Demographic and clinical characteristics of oral squamous cell carcinoma cases (all males) included for this study.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age</th>
<th>BQ</th>
<th>Tobacco chewing</th>
<th>Tobacco smoking</th>
<th>Alcohol</th>
<th>Grade</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC 4</td>
<td>40</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>OC 5</td>
<td>60</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>OC 6</td>
<td>63</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 13</td>
<td>69</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>OC 17</td>
<td>55</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 19</td>
<td>48</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>OC 26</td>
<td>55</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>OC 28</td>
<td>70</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>OC 34</td>
<td>68</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 35</td>
<td>52</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 40</td>
<td>45</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 41</td>
<td>38</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 70</td>
<td>60</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>OC 81</td>
<td>50</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 98</td>
<td>55</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>OC 102</td>
<td>45</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 105</td>
<td>58</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>OC 113</td>
<td>55</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 118</td>
<td>42</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>OC 135</td>
<td>52</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 140</td>
<td>50</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>OC 149</td>
<td>32</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>OC 150</td>
<td>42</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>OC 153</td>
<td>74</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>OC 159</td>
<td>56</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G2</td>
<td>3</td>
</tr>
</tbody>
</table>

++ = Frequently; + = Occasionally; - = Never
G1 = Well differentiated squamous cell carcinoma;
G2 = Moderately differentiated squamous cell carcinoma;
G3 = Poorly differentiated squamous cell carcinoma
**Analysis of genetic alterations in oral cancer:**

In oral squamous cell carcinoma the total number of SNPs observed (at >=5 read depth and >=50 percentage variation), were 96 (81 heterozygous SNPs and 15 homozygous SNPs). Out of these 46 were known dbSNPs and 50 were novel SNPs (Unannotated). Total 46 InDels were observed (28 insertions and 18 deletions). Number of known InDels was 17 (10 insertions and 7 deletions) while novel InDels observed were 29 (18 insertions and 11 deletions) in oral cancer cases (**Table 5.2, Figure 5.6**).

**Table 5.2: Variations in Oral Cancer Cases in a Nutshell**

<table>
<thead>
<tr>
<th>Read Depth</th>
<th>&gt;=5</th>
<th>&gt;=50</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Percentage variation for InDels</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Number of SNPs observed</td>
<td>96</td>
<td>-</td>
</tr>
<tr>
<td>Number of Heterozygous SNPs</td>
<td>81</td>
<td>-</td>
</tr>
<tr>
<td>Number of Homozygous SNPs</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>Number of Known (dbSNP) SNPs</td>
<td>46</td>
<td>-</td>
</tr>
<tr>
<td>Number of Novel SNPs (Unannotated)</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>Total number of InDels observed</td>
<td>46</td>
<td>-</td>
</tr>
<tr>
<td>Number of Known Insertions observed</td>
<td>28</td>
<td>-</td>
</tr>
<tr>
<td>Number of Novel Insertions observed</td>
<td>18</td>
<td>-</td>
</tr>
<tr>
<td>Number of known Deletions observed</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>Number of Novel Deletions observed</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td><strong>Statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Reads(uniq read)</td>
<td>2897971</td>
<td>-</td>
</tr>
<tr>
<td>aligned Reads</td>
<td>666809</td>
<td>-</td>
</tr>
<tr>
<td>Percentage reads aligned</td>
<td>23.01</td>
<td>-</td>
</tr>
<tr>
<td>Target Sequence Length</td>
<td>1060956</td>
<td>-</td>
</tr>
<tr>
<td>Total Target covered</td>
<td>942860</td>
<td>-</td>
</tr>
<tr>
<td>%Total Target covered</td>
<td>88.87</td>
<td>-</td>
</tr>
<tr>
<td>%Total Target covered with atleast 5X Read Depth</td>
<td>85.42</td>
<td>-</td>
</tr>
<tr>
<td>%Total Target covered with atleast 10X Read Depth</td>
<td>79.63</td>
<td>-</td>
</tr>
<tr>
<td>%Total Target covered with atleast 15X Read Depth</td>
<td>72.91</td>
<td>-</td>
</tr>
<tr>
<td>%Total Target covered with atleast 20X Read Depth</td>
<td>66.05</td>
<td>-</td>
</tr>
<tr>
<td>Average Read Depth</td>
<td>47.14</td>
<td>-</td>
</tr>
<tr>
<td>Optimized Average Read Depth</td>
<td>47.12</td>
<td>-</td>
</tr>
</tbody>
</table>
Fig 5.6: Chromosome-wise diagramatic representation of genomic aberrations observed in OSCC. Outermost layer represents Known SNPs: Homozygous type represented by blue and heterozygous type by red. Second from outside represents novel SNPs: Homozygous type represented by orange while Heterozygous in green. Third layer from outside represents Insertions: Known (light blue), Novel (light red). Innermost circle is for Deletions: Known deletion (grey), Novel deletion (yellow).
Known SNPs observed in oral cancer:

Maximum number of known SNPs observed in OSCC were in intronic regions of the genes (18), followed by in UTRs (14). Known SNPs with read depth more than 50% were located in the genes RBL1 (rs3020646), FAT1 (rs172903), EGFR (rs17337023), ATM (rs4585), MET (rs41737, rs41738), FHT (rs13067835), VHL (rs1642742), IL12B (rs3212227), IL1RN (rs2234677), FAT2 (rs1469680), IGF1R (rs951715), MADD (rs1051006) and EGF (rs6825106) (Table 5.3, Figure 5.7).

![Graph of known SNPs observed in OSCC](image)

**Fig 5.7: Known SNPs observed in OSCC (SNPs with read depth >50 included in the list)**

Novel SNPs observed in oral cancer:

In the present study 50 novel SNPs in oral cancer have been observed. Novel SNPs with maximum number of read depths (>25) were located in the genes APC (NM_001127511), EGFR (Non-coding), FAT1 (NM_005245), STAT5B (Non-coding), CDK5 (Non-coding), TP53 (NM_001126115), NOTCH2 NM_024408, FAT2 (NM_001447), IL12B (NM_002187), CDH3 (NM_001793), ATM (NM_000051), MET (NM_001127500), MYCL1 (Non-coding) (Table 5.4, Figure 5.8).

![Graph of novel SNPs observed in OSCC](image)

**Fig 5.8: Novel SNPs observed in OSCC (SNPs with read depth >20 included in the list)**
Detection of Genomic Alteration in OSCC using NGS

Deletions in genes observed in oral cancer:

Eighteen deletions (7 known and 11 novel) were observed in OSCC cases which were not present in controls. Among known deletions, deletion with the highest read depth (272) was observed in regulatory region of TSC1 gene (rs34947162; rs115091888) followed by in FAT1 (rs34700250; rs71652217), BMP4 (rs77966378; rs111393992), MAP2K6 (rs66753968), ERBB4 (rs34156748), BRCA1 (rs8176144; rs74395723) and SLC22A18 (rs77164179) (Table: 5.5A).

Novel deletion with the highest read depth (58) was present in MSH6 gene (NM_000179), followed by IGF1R (NM_000875), BRCA2 (NM_000059), TSC2 (NM_001114382), PAK3 (NM_001128166), GRLF1 (NM_004491), FBLN1 (NM_006485), CCND2 (NM_001759), DLG3 (NM_001166278), RASSF1 (NM_007182), and BARD1 (NM_000465) (Table: 5.5B, Figure 5.9).

Insertions in genes observed in oral cancer:

Twenty eight insertions (10 known and 18 novel) were observed in oral cancer cases only which were not present in the controls. Known insertions in the decreasing order of their read depths that is their pathological significance were present in the genes APC (rs11432316; rs79379053), ADH6 (rs5860571), SMAD2 (rs111850625), PDGFRA (rs3830355; rs72599396), RHOB (rs116662870), BRIPl (rs79494688), NBL1 (rs77253948), FAT2 (rs75548276), DLG2 (rs11464149; rs79205739) and KLK8 (rs35747818) (Table: 6A).
Novel insertions with higher significance on the basis of their read depths (>25) were present in the genes *IGF1R* (NM_000875), *RB1* (NM_000321), *PA2G4* (NM_006191), *GRLF1* (NM_004491), *CDK2AP1* (NM_004642), *CDH3* (NM_001793), *PIK3CA* (NM_006218), *DLG4* (NM_001128827) and *APC* (NM_000038). Complete list of novel insertions is provided in the table (Table: 6B).

![Fig 5.10: Known and Novel Insertions observed in various genes (Read depth >15) in OSCC (Read depth versus Genes).](image)
Detection of Genomic Alteration in OSCC using NGS

Table 5.3 – Known SNPs observed in Oral Cancer

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Chr. Location</th>
<th>rsID</th>
<th>dbSNP allele</th>
<th>Type of variation</th>
<th>Reference Base</th>
<th>SNP</th>
<th>Read Depth</th>
<th>SNP Ratio</th>
<th>Zygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB1</td>
<td>13q14.1</td>
<td>rs3020646 C/T</td>
<td>Intron</td>
<td>C T</td>
<td>138 T(138)</td>
<td>Homozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAT1</td>
<td>4q35.2</td>
<td>rs172903 C/G</td>
<td>Intron</td>
<td>C G</td>
<td>123 G(122)</td>
<td>Homozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>7p11.2</td>
<td>rs17337023 A/T</td>
<td>coding-synonymous;reference;near-gene-3</td>
<td>T W</td>
<td>123 A(66)/T(57)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>11q22.3</td>
<td>rs4585 G/T</td>
<td>utr-3</td>
<td>G K</td>
<td>95 T(51)/G(44)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>7q31.2</td>
<td>rs41737 A/G</td>
<td>coding-synonymous;reference</td>
<td>G R</td>
<td>94 G(53)/A(40)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FHIT</td>
<td>3p14.2</td>
<td>rs13067835 C/T</td>
<td>Intron</td>
<td>T Y</td>
<td>83 T(51)/C(32)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VHL</td>
<td>3p25.3</td>
<td>rs1642742 C/T</td>
<td>utr-3</td>
<td>G A</td>
<td>78 A(78)</td>
<td>Homozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL12B</td>
<td>5p25.3</td>
<td>rs3212227 A/C</td>
<td>utr-3</td>
<td>T G</td>
<td>73 G(72)</td>
<td>Homozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL1RN</td>
<td>11q22.3</td>
<td>rs2234677 A/G</td>
<td>utr-3</td>
<td>G R</td>
<td>71 G(42)/A(29)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAT2</td>
<td>5q33.1</td>
<td>rs1469680 A/G</td>
<td>near-gene-5</td>
<td>T Y</td>
<td>66 C(62)/T(4)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGF1R</td>
<td>15q26.3</td>
<td>rs951715 A/G</td>
<td>Intron</td>
<td>G R</td>
<td>65 A(48)/G(17)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADD</td>
<td>11p11.2</td>
<td>rs1051006 A/G</td>
<td>missense;reference</td>
<td>G R</td>
<td>64 A(33)/G(31)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>4q25</td>
<td>rs6825106 A/G</td>
<td>Intron</td>
<td>G R</td>
<td>57 A(39)/G(17)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>7q31.2</td>
<td>rs41738 A/G</td>
<td>Intron</td>
<td>A R</td>
<td>51 G(27)/A(24)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLCC2A18</td>
<td>11p15.4</td>
<td>rs3764896 A/G</td>
<td>near-gene-3</td>
<td>G R</td>
<td>49 A(31)/G(18)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUL2</td>
<td>10p11.21</td>
<td>rs16935840 A/G</td>
<td>coding-synonymous;reference</td>
<td>C Y</td>
<td>46 T(26)/C(20)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAT1</td>
<td>15q26.3</td>
<td>rs28423024 C/G</td>
<td>Intron;near-gene-5</td>
<td>G S</td>
<td>40 G(31)/C(8)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERBB4</td>
<td>2q34</td>
<td>rs10932374 A/G</td>
<td>utr-3</td>
<td>G R</td>
<td>37 G(21)/A(16)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RASSF1</td>
<td>3p21.31</td>
<td>rs2073498 A/C</td>
<td>utr-5;reference;missense</td>
<td>C M</td>
<td>37 A(19)/C(18)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGF1R</td>
<td>15q26.3</td>
<td>rs74534003 A/G</td>
<td>utr-3</td>
<td>G R</td>
<td>35 G(30)/A(5)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSH2</td>
<td>2p21</td>
<td>rs3732183 A/G</td>
<td>Intron</td>
<td>G R</td>
<td>34 A(18)/G(16)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDKN1A</td>
<td>6p21.2</td>
<td>rs1059234 C/T</td>
<td>utr-3</td>
<td>G Y</td>
<td>34 T(18)/G(16)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP53</td>
<td>17p13.1</td>
<td>rs2909430 A/G</td>
<td>Intron;utr-5</td>
<td>C T</td>
<td>31 T(31)</td>
<td>Homozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL1RN</td>
<td>11q23.1</td>
<td>rs2234679 C/G</td>
<td>utr-5</td>
<td>G S</td>
<td>30 G(21)/C(9)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLNC</td>
<td>7q32.1</td>
<td>rs7787924 A/G</td>
<td>Intron</td>
<td>G R</td>
<td>30 A(20)/G(10)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPHB2</td>
<td>1p36.12</td>
<td>rs309497 A/G</td>
<td>Intron</td>
<td>C Y</td>
<td>29 C(21)/T(8)</td>
<td>Heterozygous</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.3 continue

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Chr. Location</th>
<th>rsID</th>
<th>dbSNP allele</th>
<th>Type of variation</th>
<th>Reference Base</th>
<th>SNP</th>
<th>Read Depth</th>
<th>SNP Ratio</th>
<th>Zygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATM</td>
<td>11q22.3</td>
<td>rs79075295</td>
<td>A/G</td>
<td>missense;reference</td>
<td>G</td>
<td>R</td>
<td>29</td>
<td>A(15)/G(14)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>GADD45A</td>
<td>1p31.3</td>
<td>rs607375</td>
<td>C/G</td>
<td>near-gene-3</td>
<td>G</td>
<td>C</td>
<td>21</td>
<td>C(21)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>ATM</td>
<td>11q22.3</td>
<td>rs664143</td>
<td>C/T</td>
<td>Intron</td>
<td>A</td>
<td>G</td>
<td>20</td>
<td>G(20)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>RB1</td>
<td>13q14.2</td>
<td>rs520342</td>
<td>C/T</td>
<td>Intron</td>
<td>C</td>
<td>Y</td>
<td>20</td>
<td>C(11)/T(9)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>TP53</td>
<td>1p31.1</td>
<td>rs1042522</td>
<td>C/G</td>
<td>reference;missense;near-gene-5</td>
<td>G</td>
<td>S</td>
<td>20</td>
<td>C(19)/G(1)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>SART1</td>
<td>11q24.2</td>
<td>rs754532</td>
<td>C/T</td>
<td>utr-3</td>
<td>G</td>
<td>R</td>
<td>18</td>
<td>A(13)/G(5)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>STAT5A</td>
<td>17q21.2</td>
<td>rs2293155</td>
<td>C/T</td>
<td>Intron</td>
<td>A</td>
<td>G</td>
<td>17</td>
<td>G(17)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>GADD45B</td>
<td>19p13.3</td>
<td>rs14384</td>
<td>C/T</td>
<td>Intron</td>
<td>A</td>
<td>G</td>
<td>17</td>
<td>T(11)/C(6)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>RHOBTB2</td>
<td>18p13.1</td>
<td>rs2430811</td>
<td>C/T</td>
<td>Intron</td>
<td>A</td>
<td>G</td>
<td>16</td>
<td>G(16)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>NA</td>
<td>17q21.31</td>
<td>rs80357258</td>
<td>A/G</td>
<td>NA</td>
<td>T</td>
<td>K</td>
<td>16</td>
<td>T(12)/G(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>ERBB4</td>
<td>1p34</td>
<td>rs2289086</td>
<td>A/G</td>
<td>Intron</td>
<td>T</td>
<td>C</td>
<td>15</td>
<td>C(15)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>APC</td>
<td>5q22.2</td>
<td>rs41116</td>
<td>C/T</td>
<td>utr-3</td>
<td>T</td>
<td>Y</td>
<td>15</td>
<td>C(14)/T(1)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>BCAR1</td>
<td>1q23.1</td>
<td>rs16957558</td>
<td>A/C/G/T</td>
<td>missense;reference</td>
<td>C</td>
<td>Y</td>
<td>15</td>
<td>T(11)/G(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>ESR1</td>
<td>6q25.1</td>
<td>rs2273207</td>
<td>A/G</td>
<td>Intron</td>
<td>A</td>
<td>R</td>
<td>13</td>
<td>G(9)/A(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>CUL1</td>
<td>10p11.21</td>
<td>rs17582954</td>
<td>C/T</td>
<td>Intron</td>
<td>T</td>
<td>Y</td>
<td>11</td>
<td>C(9)/T(2)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>AKT1</td>
<td>14q32.33</td>
<td>rs10138227</td>
<td>C/T</td>
<td>intron:utr-5</td>
<td>C</td>
<td>T</td>
<td>11</td>
<td>T(11)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>1p12</td>
<td>rs6685892</td>
<td>A/T</td>
<td>coding-synonymous;reference</td>
<td>A</td>
<td>T</td>
<td>10</td>
<td>T(10)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>DLEC1</td>
<td>3q22.2</td>
<td>rs116954440</td>
<td>G/T</td>
<td>missense;reference</td>
<td>T</td>
<td>K</td>
<td>9</td>
<td>G(8)/T(1)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>RASSF3</td>
<td>12q14.2</td>
<td>rs1797683</td>
<td>A/G</td>
<td>near-gene-3</td>
<td>T</td>
<td>Y</td>
<td>9</td>
<td>C(5)/T(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>GAS7</td>
<td>17p13.1</td>
<td>rs16958968</td>
<td>A/G</td>
<td>utr-3</td>
<td>G</td>
<td>A</td>
<td>9</td>
<td>A(9)</td>
<td>Homozygous</td>
</tr>
</tbody>
</table>
Table 5.4: Novel SNPs observed in oral cancer

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Chromosome</th>
<th>Position</th>
<th>RefSeq ID</th>
<th>Type of variation</th>
<th>Reference Base</th>
<th>SNP</th>
<th>Read Depth</th>
<th>SNP Ratio</th>
<th>Zygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>chr5</td>
<td>112173679</td>
<td>NM_001127511</td>
<td>Synonymous</td>
<td>T</td>
<td>Y</td>
<td>93</td>
<td>T(68)/C(25)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>EGFR</td>
<td>chr7</td>
<td>55273892</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>86</td>
<td>T(57)/G(29)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>FAT1</td>
<td>chr4</td>
<td>187542876</td>
<td>NM_005245</td>
<td>Non-Synonymous</td>
<td>C</td>
<td>Y</td>
<td>62</td>
<td>C(44)/T(18)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>STAT5B</td>
<td>chr17</td>
<td>40353883</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>45</td>
<td>T(30)/A(15)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>CDK5</td>
<td>chr7</td>
<td>150751400</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>44</td>
<td>T(26)/G(16)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>TP53</td>
<td>chr17</td>
<td>7572980</td>
<td>NM_001126115</td>
<td>Non-Synonymous</td>
<td>T</td>
<td>K</td>
<td>43</td>
<td>T(22)/G(21)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>chr1</td>
<td>120469214</td>
<td>NM_024408</td>
<td>Non-Synonymous</td>
<td>C</td>
<td>Y</td>
<td>38</td>
<td>C(28)/T(10)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>MYC2</td>
<td>chr5</td>
<td>150885598</td>
<td>NM_001447</td>
<td>Non-Synonymous</td>
<td>T</td>
<td>K</td>
<td>31</td>
<td>T(16)/G(15)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>IL12B</td>
<td>chr5</td>
<td>158749446</td>
<td>NM_002187</td>
<td>Synonymous</td>
<td>C</td>
<td>M</td>
<td>31</td>
<td>C(24)/A(7)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>CDH3</td>
<td>chr16</td>
<td>68732199</td>
<td>NM_001793</td>
<td>Non-Synonymous</td>
<td>A</td>
<td>M</td>
<td>30</td>
<td>A(21)/C(9)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>ATM</td>
<td>chr11</td>
<td>108114752</td>
<td>NM_000051</td>
<td>Non-Synonymous</td>
<td>T</td>
<td>W</td>
<td>29</td>
<td>T(16)/A(13)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>MET</td>
<td>chr7</td>
<td>116380999</td>
<td>NM_001127500</td>
<td>Non-Synonymous</td>
<td>T</td>
<td>K</td>
<td>26</td>
<td>T(15)/G(11)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>MYCL1</td>
<td>chr1</td>
<td>40361910</td>
<td>NA</td>
<td>Non-coding</td>
<td>C</td>
<td>Y</td>
<td>26</td>
<td>T(16)/C(10)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>ING1</td>
<td>chr13</td>
<td>111367548</td>
<td>NA</td>
<td>Non-coding</td>
<td>G</td>
<td>R</td>
<td>24</td>
<td>G(13)/A(11)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>ADH6</td>
<td>chr4</td>
<td>100124767</td>
<td>NA</td>
<td>Non-coding</td>
<td>C</td>
<td>M</td>
<td>24</td>
<td>A(14)/C(10)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>FLNC</td>
<td>chr7</td>
<td>128483410</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>22</td>
<td>T(14)/G(8)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>GRLF1</td>
<td>chr19</td>
<td>47506818</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>21</td>
<td>G(12)/T(9)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>BARD1</td>
<td>chr2</td>
<td>215661754</td>
<td>NA</td>
<td>Non-coding</td>
<td>G</td>
<td>S</td>
<td>21</td>
<td>G(17)/C(3)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>E2F1</td>
<td>chr20</td>
<td>32266183</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>21</td>
<td>T(12)/G(9)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>DLEC1</td>
<td>chr3</td>
<td>38138049</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>20</td>
<td>T(12)/G(8)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>FGFR4</td>
<td>chr5</td>
<td>176516710</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>19</td>
<td>T(11)/G(8)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>FGFR4</td>
<td>chr6</td>
<td>178516720</td>
<td>NA</td>
<td>Non-coding</td>
<td>T</td>
<td>K</td>
<td>19</td>
<td>T(12)/G(7)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>DBC1</td>
<td>chr9</td>
<td>121976170</td>
<td>NA</td>
<td>Non-coding</td>
<td>G</td>
<td>R</td>
<td>17</td>
<td>A(11)/G(6)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>BRCA1</td>
<td>chr17</td>
<td>41197722</td>
<td>NM_007305</td>
<td>Non-Synonymous</td>
<td>T</td>
<td>K</td>
<td>16</td>
<td>T(11)/G(5)</td>
<td>Heterozygous</td>
</tr>
</tbody>
</table>
### Table 5.4 Novel SNPs observed in oral cancer (Continue)

<table>
<thead>
<tr>
<th></th>
<th>Gene</th>
<th>Chromosome</th>
<th>Location</th>
<th>Reference Transcript</th>
<th>Type</th>
<th>Codon</th>
<th>Amino Acid Change</th>
<th>Position</th>
<th>Heterozygous Alleles</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>GAS7</td>
<td>chr17</td>
<td>9821388</td>
<td>NM_201433</td>
<td>Non-Synonymous</td>
<td>A</td>
<td>M</td>
<td>15</td>
<td>A(8)/C(7)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>27</td>
<td>GAS7</td>
<td>chr17</td>
<td>9821397</td>
<td>NM_201433</td>
<td>Non-Synonymous</td>
<td>A</td>
<td>M</td>
<td>15</td>
<td>C(8)/A(7)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>28</td>
<td>ALDH2</td>
<td>chr12</td>
<td>11229295</td>
<td>NM_201433</td>
<td>Non-Synonymous</td>
<td>NA</td>
<td>C</td>
<td>15</td>
<td>C(8)/A(7)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>29</td>
<td>BRMS1</td>
<td>chr11</td>
<td>66108330</td>
<td>NM_015399</td>
<td>Synonymous</td>
<td>C</td>
<td>Y</td>
<td>14</td>
<td>C(9)/T(5)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>30</td>
<td>DLEC1</td>
<td>chr3</td>
<td>38087168</td>
<td>NM_007337</td>
<td>Synonymous</td>
<td>T</td>
<td>W</td>
<td>13</td>
<td>T(10)/A(3)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>31</td>
<td>IGF1R</td>
<td>chr15</td>
<td>99434907</td>
<td>NM_201433</td>
<td>Non-Synonymous</td>
<td>NA</td>
<td>T</td>
<td>13</td>
<td>T(8)/C(5)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>32</td>
<td>IGF1R</td>
<td>chr15</td>
<td>99434913</td>
<td>NM_201433</td>
<td>Non-Synonymous</td>
<td>NA</td>
<td>G</td>
<td>13</td>
<td>G(8)/C(5)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>33</td>
<td>CDKN1A</td>
<td>chr6</td>
<td>36654967</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>T</td>
<td>13</td>
<td>T(11)/C(2)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>34</td>
<td>AR</td>
<td>chrX</td>
<td>66931553</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>G</td>
<td>13</td>
<td>G(11)/A(2)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>35</td>
<td>SLC22A18</td>
<td>chr11</td>
<td>29433551</td>
<td>NM_183233</td>
<td>Synonymous</td>
<td>T</td>
<td>K</td>
<td>11</td>
<td>T(8)/C(5)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>36</td>
<td>LOC100287196</td>
<td>chr7</td>
<td>45151166</td>
<td>XM_002342718</td>
<td>Synonymous</td>
<td>C</td>
<td>T</td>
<td>11</td>
<td>T(11)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>37</td>
<td>HRAS</td>
<td>chr11</td>
<td>532481</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>A</td>
<td>11</td>
<td>A(9)/T(2)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>38</td>
<td>CDH3</td>
<td>chr16</td>
<td>68732556</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>C</td>
<td>11</td>
<td>C(7)/A(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>39</td>
<td>CDH3</td>
<td>chr16</td>
<td>68732561</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>G</td>
<td>11</td>
<td>G(8)/A(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>40</td>
<td>CDH3</td>
<td>chr16</td>
<td>68732568</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>A</td>
<td>11</td>
<td>A(9)/T(2)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>41</td>
<td>EPHB2</td>
<td>chr1</td>
<td>23111308</td>
<td>NM_014559</td>
<td>Non-Synonymous</td>
<td>T</td>
<td>Y</td>
<td>10</td>
<td>C(6)/T(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>42</td>
<td>EPHB2</td>
<td>chr1</td>
<td>23111311</td>
<td>NM_014559</td>
<td>Non-Synonymous</td>
<td>A</td>
<td>M</td>
<td>10</td>
<td>A(5)/C(5)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>43</td>
<td>NBL1</td>
<td>chr1</td>
<td>19981475</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>G</td>
<td>10</td>
<td>G(8)/A(2)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>44</td>
<td>EBBB2</td>
<td>chr17</td>
<td>37871479</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>T</td>
<td>10</td>
<td>T(6)/G(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>45</td>
<td>FBLN1</td>
<td>chr22</td>
<td>45937296</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>C</td>
<td>10</td>
<td>A(10)</td>
<td>Homozygous</td>
</tr>
<tr>
<td>46</td>
<td>LAT5</td>
<td>chr6</td>
<td>15000168</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>G</td>
<td>10</td>
<td>G(8)/T(2)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>47</td>
<td>DLEC1</td>
<td>chr3</td>
<td>38158010</td>
<td>NM_007337</td>
<td>Synonymous</td>
<td>T</td>
<td>K</td>
<td>9</td>
<td>G(5)/T(4)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>48</td>
<td>KLKG1</td>
<td>chr12</td>
<td>9162444</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>G</td>
<td>8</td>
<td>G(5)/A(2)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>49</td>
<td>TSC2</td>
<td>chr16</td>
<td>2098847</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>C</td>
<td>8</td>
<td>C(6)/A(1)</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>50</td>
<td>IL8</td>
<td>chr4</td>
<td>74609345</td>
<td>NM_014559</td>
<td>Non-coding</td>
<td>NA</td>
<td>G</td>
<td>8</td>
<td>G(6)/A(2)</td>
<td>Heterozygous</td>
</tr>
</tbody>
</table>
### Table 5.5A: Known Deletions observed in Oral Cancer

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Chromosome position</th>
<th>Location</th>
<th>Reference Position</th>
<th>rsID</th>
<th>No of Variations</th>
<th>InDells</th>
<th>Read Depth</th>
<th>Read Showing Variations</th>
<th>Percentage Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TSC1</td>
<td>Tuberous sclerosis 1 protein isoform 1</td>
<td>9q34.13</td>
<td>utr-3</td>
<td>135771333</td>
<td>rs34947162;rs115091888</td>
<td>1</td>
<td>A</td>
<td>272</td>
<td>140</td>
<td>51.4705</td>
</tr>
<tr>
<td>2</td>
<td>FAT1</td>
<td>Homo sapiens FAT tumor suppressor homolog 1</td>
<td>4q35.2</td>
<td>utr-3</td>
<td>187509284</td>
<td>rs34700256;rs71652217</td>
<td>1</td>
<td>A</td>
<td>53</td>
<td>35</td>
<td>66.03773585</td>
</tr>
<tr>
<td>3</td>
<td>BMP4</td>
<td>Bone morphogenetic protein 4</td>
<td>14q22.2</td>
<td>utr-5</td>
<td>54416438</td>
<td>rs77966378;rs111393992</td>
<td>1</td>
<td>A</td>
<td>51</td>
<td>28</td>
<td>54.90196078</td>
</tr>
<tr>
<td>4</td>
<td>MAP2K6</td>
<td>Mitogen-activated protein kinase kinase 6</td>
<td>17q24.3</td>
<td>utr-3</td>
<td>67538341</td>
<td>rs66753968</td>
<td>1</td>
<td>A</td>
<td>32</td>
<td>20</td>
<td>62.50</td>
</tr>
<tr>
<td>5</td>
<td>ERBB4</td>
<td>V-erb-a erythroblastic leukemia viral oncogene</td>
<td>2q34</td>
<td>utr-3</td>
<td>212243613</td>
<td>rs34156748</td>
<td>2</td>
<td>TT</td>
<td>20</td>
<td>17</td>
<td>85.00</td>
</tr>
<tr>
<td>6</td>
<td>BRCA1</td>
<td>Breast cancer 1, early onset isoform 4</td>
<td>17q21.31</td>
<td>intron</td>
<td>41249364</td>
<td>rs8176144;rs74395723</td>
<td>1</td>
<td>A</td>
<td>18</td>
<td>18</td>
<td>100.00</td>
</tr>
<tr>
<td>7</td>
<td>SLC22A18; SLC22A18AS</td>
<td>Solute carrier family 22</td>
<td>11p15.4</td>
<td>utr-5</td>
<td>2920954</td>
<td>rs77164179</td>
<td>1</td>
<td>A</td>
<td>15</td>
<td>15</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Table No. 5.5 B: Novel Deletions observed in Oral Cancer

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Chromosome position</th>
<th>Reference Position</th>
<th>RsID</th>
<th>No of Variations</th>
<th>InDell s</th>
<th>Read Depth</th>
<th>Read Showing Variations</th>
<th>Percentage Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MSH6</td>
<td>MutS homolog 6</td>
<td>2p16.3</td>
<td>4803345</td>
<td>NM_000179 5</td>
<td>2 AG</td>
<td>58</td>
<td>29</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>IGF1R</td>
<td>Insulin-like growth factor 1 receptor precursor</td>
<td>15q26.3</td>
<td>9950108</td>
<td>NM_000875 8</td>
<td>2 CT</td>
<td>49</td>
<td>33</td>
<td>67.34693878</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>BRCA2</td>
<td>Breast cancer 2, early onset</td>
<td>13q13.1</td>
<td>3295071</td>
<td>NM_000059 6</td>
<td>1 C</td>
<td>25</td>
<td>17</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>TSC2</td>
<td>Tuberous sclerosis 2 isoform 1</td>
<td>16p13.3</td>
<td>2127672</td>
<td>NM_0011143 82</td>
<td>1 G</td>
<td>24</td>
<td>13</td>
<td>54.16666667</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PAK3</td>
<td>p21-activated kinase 3 isoform a</td>
<td>Xq23</td>
<td>1102569</td>
<td>NM_0011281 66</td>
<td>1 C</td>
<td>22</td>
<td>13</td>
<td>59.09090909</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>FBLN1</td>
<td>Fibulin 1 isoform A precursor</td>
<td>22q13.31</td>
<td>4592373</td>
<td>NM_006485 4</td>
<td>1 G</td>
<td>18</td>
<td>10</td>
<td>55.55555556</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>GRLF1</td>
<td>Glucocorticoid receptor DNA binding factor 1</td>
<td>19q13.32</td>
<td>4750596</td>
<td>NM_004491 5</td>
<td>1 G</td>
<td>18</td>
<td>15</td>
<td>83.33333333</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CCND2</td>
<td>Cyclin D2</td>
<td>12p13.32</td>
<td>4410520</td>
<td>NM_001759 1</td>
<td>1 T</td>
<td>16</td>
<td>8</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>DLG3</td>
<td>Synapse-associated protein 102 isoform a</td>
<td>Xq13.1</td>
<td>6967504</td>
<td>NM_0011662 78</td>
<td>1 C</td>
<td>13</td>
<td>13</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>RASSF1</td>
<td>Ras association domain family 1 isoform C</td>
<td>3p21.31</td>
<td>5037454</td>
<td>NM_007182 2</td>
<td>1 G</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>BARD1</td>
<td>BRCA1 associated RING domain 1</td>
<td>2q35</td>
<td>2156572</td>
<td>NM_000465 60</td>
<td>1 C</td>
<td>8</td>
<td>5</td>
<td>62.5</td>
<td></td>
</tr>
<tr>
<td>Sl. No.</td>
<td>Gene Symbol</td>
<td>Gene Name</td>
<td>Chromosome position</td>
<td>Location</td>
<td>Reference Position</td>
<td>rsID</td>
<td>Number of Inserted Bases</td>
<td>Inserted Bases</td>
<td>Read Depth</td>
<td>Read Showing Insertions</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------------------------------------</td>
<td>---------------------</td>
<td>----------</td>
<td>-------------------</td>
<td>----------------------------</td>
<td>--------------------------</td>
<td>-----------------</td>
<td>------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>1</td>
<td>APC</td>
<td>adenomatous polyposis coli</td>
<td>5q22.2</td>
<td>utr-3</td>
<td>112180 228</td>
<td>rs11432316;rs7937053</td>
<td>1</td>
<td>A</td>
<td>102</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>ADH6</td>
<td>Alcohol dehydrogenase 6</td>
<td>4q23</td>
<td>intron</td>
<td>100134 713</td>
<td>rs5860571</td>
<td>1</td>
<td>T</td>
<td>95</td>
<td>76</td>
</tr>
<tr>
<td>3</td>
<td>SMAD2</td>
<td>Sma- and Mad-related protein 2</td>
<td>18q21.1</td>
<td>utr-3</td>
<td>453610 16</td>
<td>rs111850625</td>
<td>4</td>
<td>TTAT</td>
<td>76</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>PDGFRA</td>
<td>platelet-derived growth factor</td>
<td>4q12</td>
<td>intron</td>
<td>551519 58</td>
<td>rs3830355;rs72599396</td>
<td>1</td>
<td>A</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>RHOB</td>
<td>ras homolog gene family, member B</td>
<td>2p24.1</td>
<td>utr-3;utr-5</td>
<td>206489 33</td>
<td>rs116662870</td>
<td>1</td>
<td>T</td>
<td>32</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>BRIP1</td>
<td>BRCA1 interacting protein C-terminal helicase 1</td>
<td>17q23.2</td>
<td>intron</td>
<td>598575 99</td>
<td>rs79494688</td>
<td>1</td>
<td>A</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>NBL1</td>
<td>neuroblastoma, suppression of</td>
<td>1p38.13</td>
<td>utr-3; near- gene-5</td>
<td>199841 00</td>
<td>rs77253948</td>
<td>1</td>
<td>C</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>FAT2</td>
<td>FAT tumor suppressor 2 precursor</td>
<td>5q33.1</td>
<td>intron</td>
<td>150887 184</td>
<td>rs75548276</td>
<td>3</td>
<td>AGA</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>DLG2</td>
<td>chapsyn-110 isoform 4</td>
<td>11q14.1</td>
<td>intron</td>
<td>831801 69</td>
<td>rs11464149;rs79205739</td>
<td>1</td>
<td>A</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>KLK8</td>
<td>kallikrein 8 isoform 1 preproprotein</td>
<td>19q13.41</td>
<td>intron</td>
<td>515009 18</td>
<td>rs35747818</td>
<td>1</td>
<td>C</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>
### Table 5.6 B: Novel Insertions observed in Oral Cancer

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Chromosome position</th>
<th>Reference Position</th>
<th>rsID</th>
<th>Number of Inserted Bases</th>
<th>Inserted Bases</th>
<th>Read Depth</th>
<th>Read Showing Insertions</th>
<th>Percentage Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IGF1R</td>
<td>Insulin-like growth factor 1 receptor precursor</td>
<td>15q26.3</td>
<td>99478505</td>
<td>NM_00087</td>
<td>1</td>
<td>C</td>
<td>58</td>
<td>49</td>
<td>84.48276</td>
</tr>
<tr>
<td>2</td>
<td>RB1</td>
<td>Retinoblastoma-associated protein variant protein</td>
<td>13q14.2</td>
<td>48921994</td>
<td>NM_00032</td>
<td>1</td>
<td>A</td>
<td>48</td>
<td>29</td>
<td>60.41667</td>
</tr>
<tr>
<td>3</td>
<td>PA2G4</td>
<td>ErbB3-binding protein 1</td>
<td>12q13.2</td>
<td>56507345</td>
<td>NM_00619</td>
<td>1</td>
<td>T</td>
<td>32</td>
<td>19</td>
<td>59.375</td>
</tr>
<tr>
<td>4</td>
<td>GRLF1</td>
<td>Glucocorticoid receptor DNA binding factor 1</td>
<td>19q13.32</td>
<td>47504535</td>
<td>NM_00449</td>
<td>1</td>
<td>A</td>
<td>32</td>
<td>16</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>CDK2AP1</td>
<td>CDK2-associated protein 1</td>
<td>12q24.31</td>
<td>123749787</td>
<td>NM_00179</td>
<td>1</td>
<td>T</td>
<td>31</td>
<td>17</td>
<td>54.83871</td>
</tr>
<tr>
<td>6</td>
<td>CDH3</td>
<td>Cadherin-3</td>
<td>16q22.1</td>
<td>68711974</td>
<td>NM_00112</td>
<td>1</td>
<td>G</td>
<td>31</td>
<td>21</td>
<td>67.74194</td>
</tr>
<tr>
<td>7</td>
<td>PIK3CA</td>
<td>Phosphoinositide-3-kinase</td>
<td>3q26.32</td>
<td>178952463</td>
<td>NM_00032</td>
<td>1</td>
<td>G</td>
<td>27</td>
<td>15</td>
<td>55.55556</td>
</tr>
<tr>
<td>8</td>
<td>DLG4</td>
<td>Post-synaptic density protein 95 isofrm 2</td>
<td>17p13.1</td>
<td>7096891</td>
<td>NM_00112</td>
<td>1</td>
<td>A</td>
<td>30</td>
<td>19</td>
<td>63.3333</td>
</tr>
<tr>
<td>9</td>
<td>APC</td>
<td>Adenomatous polyposis coli</td>
<td>5q22.2</td>
<td>112176071</td>
<td>NM_00003</td>
<td>1</td>
<td>T</td>
<td>27</td>
<td>15</td>
<td>55.55556</td>
</tr>
<tr>
<td>10</td>
<td>NF2</td>
<td>Neurofibromin 2 isofrm 2</td>
<td>22q12.2</td>
<td>30090908</td>
<td>NM_18182</td>
<td>1</td>
<td>A</td>
<td>22</td>
<td>13</td>
<td>59.09091</td>
</tr>
<tr>
<td>11</td>
<td>BRIP1</td>
<td>BRCA1 interacting protein C-terminal helicase 1</td>
<td>17q23.2</td>
<td>59760965</td>
<td>NM_03204</td>
<td>1</td>
<td>T</td>
<td>21</td>
<td>12</td>
<td>57.14286</td>
</tr>
<tr>
<td>12</td>
<td>GNAS1</td>
<td>Growth arrest-specific 1 precursor</td>
<td>9q21.33</td>
<td>89561020</td>
<td>NM_00204</td>
<td>1</td>
<td>A</td>
<td>16</td>
<td>16</td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>KLK8</td>
<td>Kallikrein 8 isofrm 1 preproprotein</td>
<td>19q13.41</td>
<td>51503266</td>
<td>NM_00719</td>
<td>1</td>
<td>T</td>
<td>15</td>
<td>10</td>
<td>66.66667</td>
</tr>
<tr>
<td>14</td>
<td>MAP2K1</td>
<td>Mitogen-activated protein kinase kinase 1</td>
<td>15q22.31</td>
<td>66774045</td>
<td>NM_00204</td>
<td>1</td>
<td>G</td>
<td>14</td>
<td>11</td>
<td>78.57143</td>
</tr>
<tr>
<td>15</td>
<td>MAP3K1</td>
<td>MAP-kinase activating death domain-containing</td>
<td>11p11.2</td>
<td>47291817</td>
<td>NM_00368</td>
<td>1</td>
<td>G</td>
<td>12</td>
<td>12</td>
<td>100</td>
</tr>
<tr>
<td>16</td>
<td>BRMS1</td>
<td>Breast cancer metastasis suppressor 1 isofrm 2</td>
<td>11q13.2</td>
<td>66106296</td>
<td>NM_00539</td>
<td>1</td>
<td>T</td>
<td>11</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>17</td>
<td>CDKN2B</td>
<td>Cdk2-associated protein 1</td>
<td>9p21.3</td>
<td>22005862</td>
<td>NM_00493</td>
<td>1</td>
<td>T</td>
<td>11</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>18</td>
<td>ATM</td>
<td>Ataxia telangiectasia mutated isofrm 1</td>
<td>11q22.3</td>
<td>108205903</td>
<td>NM_00005</td>
<td>1</td>
<td>A</td>
<td>9</td>
<td>7</td>
<td>77.77778</td>
</tr>
</tbody>
</table>
DISCUSSION:

DNA sequencing represents a single method by which a broad range of biological events can be forecasted. The recent development of instruments capable of producing millions of DNA sequence reads in a single run is rapidly changing the landscape of cancer research with the potential to accelerated analysis of genomes, transcriptomes and interactomes. Using this technology in the present study, we have analyzed 25 cases of oral squamous cell carcinoma and revealed large number of genetic variations involved in carcinogenesis. To the best of my knowledge, this is the first study in oral cancer using NGS for identifying mutational changes involved in oral carcinogenesis.

**Known SNPs observed in our study:**

Most of the known SNPs with high read depths (>50) were located in tumor suppressor genes such as *RB1*, *FHIT*, *FAT1*, *FAT2* and *VHL*. SNPs detected in *RB1*, *FHIT* and *FAT1* are located in the intronic regions of the gene while those in *ATM*, *VHL*, *IL12B*, and *MET* were located in 3’UTR. Though these are non-coding regions, various studies have reported their significant functions associated with diseases, thus they might be of functional relevance with the process of oral carcinogenesis. Additionally the GO analysis have revealed various biological functions of these genes which shows possible relevance of these genes with the processes of carcinogenesis.

Of the known SNPs observed in our study, 6 SNPs which resulted in miss-sense mutations were present in *ATM*, *MADD*, *TP53*, *BCAR1*, *RASSF1* and *DLEC1* genes. SNP present in *ATM* gene (rs79075295) is a G to A variation resulting in amino acid change from R (AGA) to K (AAA). *ATM* is thought to play a key role in the caretaking of the overall genome stability, and its mutations have been implicated in human cancers. However, the role of *ATM* variations in oral carcinogenesis is largely unexplored. The protein encoded by this gene belongs to the PI3/PI4-kinase family. This protein is an important cell cycle checkpoint kinase that phosphorylates; thus, it functions as a regulator of a wide variety of downstream proteins, including tumor suppressor proteins p53 and *BRCA1*, checkpoint kinase CHK2, checkpoint proteins RAD17 and RAD9, and DNA repair protein *NBS1*. *ATM* is required for the repair of DNA double strand breaks that arise endogenously or following oxidative stress (Segal-Raz *et al.*, 2011; Woodbine *et al.*, 2011). *ATM* activates the pentose phosphate pathway promoting anti-oxidant defence and DNA repair (Cosentino *et al.*, 2011). Truncating and missense mutations in the *ATM* gene, which cause insufficient DNA damage
surveillance, allow damaged cells to proceed into mitosis, which eventually results in increased cancer susceptibility (Dombernowsky et al., 2008). Gene expression profile study of OSCC cases showed that over-expression of ATM gene played an important role in oral carcinogenesis (He et al., 2008). Another polymorphism (rs189037) in this gene using traditional PCR-RFLP method showed polymorphic variants of ATM to be associated with oral cancer susceptibility (Bau et al., 2010). This polymorphic variant observed in our study (rs79075295) using NGS was reported for the first time to be associated with oral carcinogenesis.

The SNP present in MADD gene (rs1051006) is a V (GTG) to M (ATG) change. GO analysis revealed MADD gene to be involved in regulation of apoptosis; cell surface receptor linked signaling pathway and activation of MAPK activity. This variant was also first time reported to be involved in oral carcinogenesis.

SNP present in RASSF1 gene (rs2073498) is a C to A variation resulting in amino acid change from A (GCT) to S (TCT). GO analysis revealed this gene to be involved in protein stabilization; positive regulation of protein ubiquitination; negative regulation of cell cycle arrest; signal transduction; cell cycle arrest and Ras protein signal transduction. Loss or altered expression of this gene has been associated with the pathogenesis of a variety of cancers, which suggests the tumor suppressor function of this gene. The encoded protein was found to interact with DNA repair protein XPA. The protein was also shown to inhibit the accumulation of cyclin D1, and thus induce cell cycle arrest. In few earlier reports locus 3p where this RASSF1 gene is located was reported to be frequently deleted in oral cancer (Tran et al., 2005; Tsui et al., 2008; Yamamoto et al., 2007). However, this variation is first time reported to be associated with oral cancer.

SNP present in p53 gene (rs1042522) is a G to C variation, which resulted in amino acid change from P to R due to change in codon from CCC to CGC. This gene encodes tumor protein p53, which responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutants of p53 that frequently occur in a number of different human cancers fail to bind the consensus DNA binding site, and hence cause the loss of tumor suppressor activity. Variety of TP53 mutations have been reported each with its own biological and clinical implications (Lindenbergh-van der Plas et al., 2011). Mutations in the p53 gene frequently occur in many cancers and are present in 50-60% of head and neck squamous cell carcinoma (HNSCC).
Detection of Genomic Alterations in OSCC using NGS (Golusinski et al., 2011), which may occur early or late in the development of carcinogenesis (Ogmundsdottir et al., 2009). The present SNP (rs1042522) was reported to be present in other diseases and cancers but to the best of our knowledge, this is first time reported to be associated with oral carcinogenesis.

SNP observed in BCAR1 gene (rs16957558) is a C to A or T variation resulting in R to H or L change in amino acid. BCAR1 is a Src family kinase substrate involved in various cellular events, including migration, survival, transformation, and invasion (Sawada et al., 2006). Additionally GO analysis revealed involvement of BCAR1 gene in cell adhesion; cell proliferation; actin filament organization; epidermal growth factor receptor signaling pathway; regulation of apoptosis; regulation of cell growth and regulation of cell migration. Genetic variations in this gene were reported to play role in different cancers but this is the first report showing association of this SNP (rs16957558) with oral carcinogenesis.

K type SNP observed in DLEC1 gene (rs116954440) with unknown functional details. DLEC1 gene is located in the 3p22-p21.3 chromosomal segment that is commonly deleted in various carcinomas. GO analysis revealed its function as negative regulator of cell proliferation. Genetic alterations in this gene were reported to be associated with various cancers such as lung, gastric, hepatocellular carcinoma (Qiu et al., 2008; Ying et al., 2009; Zhang et al., 2010). The only known report in oral carcinoma showed transcriptional repression of DLEC1 to be associated with the depth of tumor invasion in oral squamous cell carcinoma (Chan et al., 2010). However no mutational report was found for this gene with oral cancer thus it is the first report showing this variation (rs116954440) to be associated with oral carcinogenesis.

**Novel SNPs observed in our study:**

Novel SNPs observed in oral cancer with more significance (read depths >25), with non-synonymous type variations which may lead to structural and functional changes in resulting protein were observed in FAT1, FAT2, TP53, NOTCH2, CDH3, ATM and MET, synonymous type variations were observed in APC and IL12B genes and those present in non-coding regions were observed in or near to EGFR, STAT5B, CDK5 and MYCL1 genes.

Novel and non-synonymous SNPs present in FAT1 and FAT2 genes were of Y and K type respectively. Tumor suppressor encoded by FAT gene is known for controlling cell proliferation. The gene product is a member of the cadherin superfamily, a group of integral
membrane proteins. Its product probably functions as an adhesion molecule and/or signaling receptor, and is likely to be important in developmental processes and cell communication.

*FAT1* was also reported to be involved in the control of cell migration and invasion of oral squamous cell carcinoma through the localization of β-catenin (Nishikawa *et al.*, 2011). GO analysis showed *FAT1* to be involved in biological processes of cell adhesion; cell-cell signaling, anatomical structure morphogenesis and homophilic cell adhesion. Human FAT2 is reported to be localized at immature adherens junctions in epidermal keratinocytes (Matsui *et al.*, 2007). Additionally GO analysis showed role of *FAT2* in biological process; cell adhesion; and homophilic cell adhesion. Results from another study using array CGH identified mutations in FAT as an important factor in the development of oral cancer (Nakaya *et al.*, 2007). But the novel variations which were identified in our study were not reported earlier, so considering their biological relevance these genetic changes may work as useful predictive and prognostic markers for oral carcinogenesis.

Another novel and non-synonymous SNP (NM_001126115) observed at location 7572980 in *p53* gene is a K type variation. Biological significance of *p53* is very well known so identification of this variation in *p53* gene may further enhance our understanding about oral carcinogenesis.

A non-synonymous and novel SNP (NM_024408 at 120469214 position) was identified in *NOTCH2* gene also. Notch family members play a role in a variety of developmental processes by controlling cell fate decisions. The Notch signaling network is an evolutionarily conserved intercellular signaling pathway which regulates interactions between physically adjacent cells. The GO analysis revealed involvement of *NOTCH2* in cell cycle arrest; cell growth; cell differentiation; induction of apoptosis; negative regulation of cell proliferation and regulation of transcription. TNF signaling activates *NOTCH2* that sensitizes endothelial cells to apoptosis via modulation of the key apoptosis regulator survivin (Quillard *et al.*, 2009). Genetic variations in *NOTCH2* were reported to be associated with cancers of breast (Fu *et al.*, 2010), colon (Chu *et al.*, 2009) and glial brain tumors (Boulay *et al.*, 2007). But to the best of our knowledge this is the first report showing mutational variations in this gene to be associated with oral carcinogenesis.

*CDH3* gene is a classical cadherin from the cadherin superfamily. The encoded protein is a calcium-dependent cell-cell adhesion glycoprotein. In our study we found a novel and non-synonymous M type SNP in this gene (NM_001793 at position 68732199) which
was not reported earlier. This gene is located in a six-cadherin cluster in a region on the long arm of chromosome 16 that is involved in loss of heterozygosity events in breast and prostate cancer. In addition, there are few studies which suggested aberrant expression of protein encoded by this gene could constitute a hallmark of aggressive biological behaviour in oral squamous cell carcinoma have (Lo-Muzio et al., 2005; Lo-Muzio et al., 2004). Mutational variation in this gene is first time reported in our study which may play a significant role oral carcinogenesis.

Novel and non-synonymous SNP (W type at 108114752 position) observed in ATM gene. Importance of this gene in relation to carcinogenesis has been discussed earlier in known SNPs. This protein and the closely related kinase ATR are thought to be master controllers of cell cycle checkpoint signaling pathways that are required for cell response to DNA damage and for genome stability (Hadian and Krappmann, 2011). ATM dictates purine, pyrimidine, and urea cycle pathways through the regulation of adenosine monophosphate (AMP) activated protein kinase (AMPK), a major sensor and regulator of cellular energy homeostasis (Cheema et al., 2011). Through binding to p53, EDD (E3 identified by differential display) actively inhibits p53 phosphorylation by ATM and plays a role in ensuring smooth G(1)/S progression (Ling and Lin, 2011). ATM activates the pentose phosphate pathway promoting anti-oxidant defence and DNA repair (Cosentino et al., 2011). Studies indicate that the mutation patterns of p53, ATM and CDKN2A support the oncogene-induced DNA replication stress model, which attributes genomic instability and p53 and ATM mutations to oncogene-induced DNA damage (Negrini et al., 2010). In addition to activating DNA damage signaling, ATM may avert chromosome translocations by preventing excessive loading of recombinational repair proteins onto translocation breakpoint hotspots (Sun et al., 2010). Truncating and missense mutations in the ATM gene, which cause insufficient DNA damage surveillance, allow damaged cells to proceed into mitosis, which eventually results in increased cancer susceptibility (Dombernowsky et al., 2008). Gene expression profile study of OSCC cases showed that over-expression of ATM gene played an important role in oral carcinogenesis (He et al., 2008). In a study polymorphic variants of ATM allele were found to be correlated with oral cancer susceptibility (Bau et al., 2010). However the variant observed in our study was reported for the first time to be associated with oral carcinogenesis.

Non-synonymous SNP located in MET gene (NM_001127500 at position 116380999) is a K type variation. The proto-oncogene MET product is the hepatocyte growth factor
receptor and encodes tyrosine-kinase activity. The MET promotes cancer development by promoting angiogenesis, proliferation, enhanced cell motility, resistance to apoptosis, invasion, and eventual metastasis (Ma et al., 2003; Mascarenhas et al., 2010). In few reports on MET gene showed its association with metastasis and progression of oral squamous cell carcinoma (Klosek et al., 2004; Zhao et al., 2011). However this mutational variation is first time reported in oral carcinogenesis and may play as an important predictive marker for oral cancer.

**Known deletions observed in our study:**

Insertions and deletions (InDels) of bases are among highly damaging mutations and responsible for carcinogenesis as they may result in abrupt changes in expression of genes or in the structural or functional activities of the encoded protein. The known deletions observed in our study were present in non-coding regions of the gene, 4 of them were present in 3’UTR (TSC1, FAT1, MAP2K6, and ERBB4), two at 5’UTR (BMP4, and SLC22A18) and one in intronic region of BRCA1. UTR regions of a gene have significant role in regulation of gene activity, thus its alteration may badly affect gene activity. The known deletion with highest read depth (272) was present in UTR region of TSC1 gene (rs34947162; rs115091888). TSC1 which codes for hamartin plays a central role in regulating cell survival and proliferation signaling pathways (Au et al., 2008). Tuberous sclerosis tumor suppressors TSC1 and TSC2 form a protein complex that integrates and transmits cellular growth factor and stress signals to negatively regulate checkpoint kinase TOR activity (Ellisen, 2005). It was reported that deregulation of the TSC-mTOR pathway may cause not only tumor development but also metabolic disorders such as diabetes and its complications (Inoki, 2008). TSC1 mutation possibly has a causative role in the initiation or progression of some bladder tumors and this process is possibly related to the functional loss of p27 (Adachi et al., 2003). Deletion in regulatory region of this will cause aberrant gene expression which may be one of the responsible factors for oral carcinogenesis and was reported for the first time by present study.

**FAT1** which was discussed earlier showed presence of known deletion also (rs34700250; rs71652217) in addition to known and novel SNPs, and insertion of high significance within this gene in the present study. Thus our study suggests high importance of FAT gene and found this as one of the significantly altered genes in oral carcinogenesis. Another known deletion of significant importance in oral cancer was present in BMP4 gene.
An earlier study on breast cancer cells suggested that BMP4 is an important regulator of key phenotypic characteristics of cancer cells, cell growth, cell migration, and invasion (Ketolainen et al., 2010). To the best of my knowledge this mutational change is first time reported to be associated with oral carcinogenesis. Other known deletions observed in oral cancer were located in UTR region of MAP2K6 (rs66753968). Biological functions revealed by GO analysis showed role of this gene in cell cycle arrest; activation of MAPK activity; positive regulation of apoptosis; signal transduction; DNA damage induced protein phosphorylation. ERBB4 gene also showed a deletion at 3'UTR region (rs34156748). GO analysis revealed that this gene play functional role in positive regulation of cell migration; cell proliferation; regulation of transcription; signal transduction; positive regulation of anti-apoptosis; and in positive regulation of epithelial cell proliferation. Thus it may be one of important finding related to oral carcinogenesis. A less studied gene in which deletion observed was SLC22A18 (rs77164179). Mutations in this gene have been earlier reported in Wilms' tumor and lung cancer. Another known deletion was found in the intronic region of BRCA1 (rs8176144; rs74395723), a well known tumor suppressor gene involved in carcinogenesis. These all genes have biological relevance with carcinogenesis, so identification of these mutational changes will enhance our knowledge about oral carcinogenesis and these markers may work as useful prognostic and therapeutic markers for oral cancer.

**Novel deletions observed in our study:**

A total of 11 novel deletions have been observed in oral cancer cases in our study. Novel deletion with the highest read depth (58) was present in MSH6 gene (NM_000179 at position 48033455), The encoded protein of MSH6 gene combines with MSH2 to form a mismatch recognition complex that functions as a bidirectional molecular switch that exchanges ADP and ATP as DNA mismatches are bound and dissociated. In earlier reports mutations in MSH6 were found to be associated with some cancers such as hereditary nonpolyposis colon cancer (HNPCC) and endometrial cancer, hereditary prostate cancer (Bauer et al., 2011) and colorectal cancer (Lim et al., 2010). However it is the novel deletion observed in MSH6 gene in oral cancer cases in our study.

Another significant novel deletion was observed in IGF1R (NM_000875 at position 99501088). This gene encodes receptor which binds insulin-like growth factor with a high affinity. It has tyrosine kinase activity. The insulin-like growth factor I receptor plays a
critical role in transformation events. It is highly overexpressed in most malignant tissues where it functions as an anti-apoptotic agent by enhancing cell survival. Aberrant expression of this gene was reported to be associated with various cancers. A recent study identified role of IGF-1R in primary and metastatic undifferentiated carcinoma of the head and neck and suggested it a possible target for immunotherapy (Friedrich et al., 2010). Thus this deletion may be of great significance with regard to oral carcinogenesis.

Other novel deletions were found to be located in BRCA2 (NM_000059 at position 32950716), TSC2 (NM_001114382 at position 2127672), PAK3 (NM_001128166 at position 110256975), FBLN1 (NM_006485 at position 45923734), GRLF1 (NM_004491 at position at 47505965), CCND2 (NM_001759 at position 4410520), DLG3 (NM_001166278 at position 69675040), RASSF1 (NM_007182 at position 50374542), and BARD1 (NM_000465 at position 215657260). Most of these genes have biological relevance with carcinogenesis, so identification of these mutational changes will enhance our understanding about carcinogenesis and may work as useful prognostic and predictive markers and therapeutic targets for oral carcinogenesis.

**Known Insertions observed in our study:**

Of the known insertions observed in our study, 4 were present in regulatory regions of APC, SMAD2, RHOB and NBL1 genes remaining 6 were located in intronic regions of ADH6, PDGFRA, BRIP1, FAT2, DLG2 and KLK8. The insertion with highest read depth (102) was that of base A at position 112180228 in APC gene (rs11432316; rs79379053). This gene encodes a tumor suppressor protein that acts as an antagonist of the Wnt signaling pathway. APC also has a role in the G2/M transition, potentially through association with topo IIalpha (Wang et al., 2008a). It can negatively regulate cell cycle progression through inhibition of DNA replication by direct interaction with DNA (Qian et al., 2008). Alternations in the APC gene mutations are involved in tumor growth and in tumor progression (De Filippo et al., 2002). APC induces apoptosis and thus plays a role in tumor suppression (Steigerwald et al., 2005). In addition to its known role in beta-catenin transcriptional signaling APC also play a role in cell adhesion (Faux et al., 2004). Truncating APC mutations have dominant effects on cell proliferation, spindle checkpoint control, survival and chromosome stability (Tighe et al., 2004). Wnt-1, beta-catenin and APC expressions were related to the differentiation of oral squamous cell carcinoma (Zhang and Gao, 2005). Inactivation of the APC gene plays a minor role in the carcinogenesis of oral
Another important known insertion was reported at 45361016 location in SMAD2 gene (rs111850625). The protein encoded by this gene belongs to the SMAD, a family of proteins similar to the gene products of the Drosophila gene 'mothers against decapentaplegic' (Mad) and the C. elegans gene Smo. SMAD proteins are signal transducers and transcriptional modulators that mediate multiple signaling pathways. This protein mediates the signal of the transforming growth factor (TGF)-beta, and thus regulates multiple cellular processes, such as cell proliferation, apoptosis, and differentiation. This protein is recruited to the TGF-beta receptors through its interaction with the SMAD anchor for receptor activation (SARA) protein. In response to TGF-beta signal, this protein is phosphorylated by the TGF-beta receptors. The phosphorylation induces the dissociation of this protein with SARA and the association with the family member SMAD4. The association with SMAD4 is important for the translocation of this protein into the nucleus, where it binds to target promoters and forms a transcription repressor complex with other cofactors. Differential roles of Smad2 and Smad3 have been reported in the regulation of TGF-β1-mediated growth inhibition and cell migration in pancreatic ductal adenocarcinoma cells (Ungefroren et al., 2011). The expression of p-Smad2 is associated with malignant phenotype and poor prognosis in patients with advanced gastric carcinoma (Shinto et al., 2010). In a recent study expression of Smad6 together with Smad2 were suggested to work as prognostic factors, independent of nodal status in oral SCC after curative resection (Mangone et al., 2010). A novel missense mutation of SMAD2, located in exon 8 at codon 276 TCG (ser) to TTG (leu), was identified in head and neck squamous cell carcinoma cell line SCC-15 (Qiu et al., 2007). Disruption of the TGF-beta 1-Smad2 signaling pathway may lead to the resistance of TGF-beta 1 growth-inhibitory effect on oral squamous cell carcinoma (Peng et al., 2006). However this type of insertion was first time reported in this gene associated with OSCC and may be further analyzed for its importance as a diagnostic or therapeutic marker.

Another gene RHOB, in which insertion observed at 20648933 (rs116662870), encodes protein which is reported to work as tumour suppressor protein (Connolly et al. 2010). In a study it was reported to be involved in the process of inducing apoptosis related to genotoxic stress (Srougi and Burridge, 2011). Up-regulation of RhoB significantly inhibited heat stress-induced apoptosis and elevated transcriptional activity of NF-kappaB (Li et al., 2005). RhoB is essential for DNA damage-induced apoptosis in neoplastically transformed squamous cell carcinoma (Rivero et al., 2008). However this base insertion was first time reported to be associated with oral carcinogenesis.
Detection of Genomic Alteration in OSCC using NGS

cells (Liu et al., 2001). Thus an insertion in this gene may alter its functions and thus may be playing a significant role in oral carcinogenesis.

*NBL1* gene also showed insertion at 19984100 location (rs77253948). Functions of this gene are not very well known and it is very less studied gene in relation to carcinogenesis. On searching pubmed only two studies in relation to carcinogenesis were found which showed it to be associated with prostate and pancreatic cancers (Olakowski et al., 2009; Shaikhibrahim et al., 2011).

Another known insertion was observed at location 100134713 in *ADH6* gene (rs5860571), alcohol dehydrogenase 6 (class V), this gene encodes class V alcohol dehydrogenase, which is a member of the alcohol dehydrogenase family. Members of this family metabolize a wide variety of substrates, including ethanol, retinol, other aliphatic alcohols, hydroxysteroids, and lipid peroxidation products. This gene has been reported to play role in various cancers.

Known insertion (rs3830355; rs72599396) was also observed in *PDGFRA*, (platelet-derived growth factor receptor, alpha polypeptide), which encodes a cell surface tyrosine kinase receptor for members of the platelet-derived growth factor family. These growth factors are mitogens for cells of mesenchymal origin.

*BRIP1* also showed a known insertion (rs79494688), protein encoded by this gene is a member of the RecQ DEAH helicase family and interacts with the BRCT repeats of breast cancer, type 1 (*BRCA1*). The bound complex is important in the normal double-strand break repair function of breast cancer, type 1 (*BRCA1*). This gene may be a target of germline cancer-inducing mutations.

*FAT2* (rs75548276), gene is the second identified human homolog of the Drosophila fat gene, which encodes a tumor suppressor essential for controlling cell proliferation. The gene product is a member of the cadherin superfamily, a group of integral membrane proteins characterized by the presence of cadherin-type repeats. This protein most likely functions as a cell adhesion molecule and controlling cell proliferation. It has been discussed earlier also and showed importance in various cancers.

Another known insertion (rs11464149; rs79205739) was present in *DLG2* (discs, large homolog 2), that encodes a member of the membrane-associated guanylate kinase (MAGUK) family protein and an insertion (rs35747818) was also reported in *KLK8*,
Kallikreins which are a subgroup of serine proteases having diverse physiological functions. Growing evidence suggests that many kallikreins are implicated in carcinogenesis and some have potential as novel cancer and other disease biomarkers.

**Novel Insertions observed in our study:**

Total 18 novel insertions have been reported in our study present in 18 different genes (table 5.6B). Novel insertion with highest read depth (49) was observed in *IGF1R* (insulin-like growth factor 1 receptor) gene at ‘99478505’ location. This receptor binds insulin-like growth factor with a high affinity. It has tyrosine kinase activity. It plays a critical role in transformation events. It was reported to be highly overexpressed in most malignant tissues where it functions as an anti-apoptotic agent by enhancing cell survival. IGF-1R was also suggested a biomarker for the stage and risk of carcinogenesis during neoplastic initiation and progression along the colorectal normal mucosa-polyp-cancer sequence (Shan et al., 2011). Overexpression of *IGF-1R* was reported to play an important role in the carcinogenesis and development of laryngeal squamous cell carcinoma (Liu et al., 2010). It was also reported to play role in primary and metastatic undifferentiated carcinoma of the head and neck and Friedrich et al suggested it a possible target of immunotherapy (Friedrich et al., 2010). Disruption of IGF-1R signaling increases TRAIL-induced apoptosis (Karasic et al., 2010). Casa et al reported that IGF-1R pathway played a key role in cancer therapeutic resistance (Casa et al., 2008).

Another novel insertion was observed in *RB1* gene, the protein encoded by this gene is a negative regulator of the cell cycle and was the first tumor suppressor gene discovered. The encoded protein also stabilizes constitutive heterochromatin to maintain the overall chromatin structure. The active, hypophosphorylated form of the protein binds transcription factor E2F1. Defects in this gene are reported to be cause of various cancers.

Another gene *PA2G4* (proliferation-associated 2G4) in which novel insertion was reported, encodes an RNA-binding protein that is involved in growth regulation. This protein is present in pre-ribosomal ribonucleoprotein complexes and may be involved in ribosome assembly and the regulation of intermediate and late steps of rRNA processing. This protein can interact with the cytoplasmic domain of the ErbB3 receptor and may contribute to transducing growth regulatory signals. This protein is also a transcriptional co-repressor of androgen receptor-regulated genes and other cell cycle regulatory genes through its interactions with histone deacetylases. This protein has been implicated in growth inhibition
and the induction of differentiation of human cancer cells. Other novel insertions were reported in various genes most of them have functional relevance with carcinogenesis. But being reported for the first time these aberrations have high importance and will make a foundation for further research in oral carcinogenesis.

Mutations are hallmark of cancers and detection of mutations is crucial in our understanding of the disease. Next generation sequencing has demonstrated its enormous potential for identification of mutational changes. Oral cancer patients included in our study have high exposure to carcinogenic compounds as most of our cases have chewing habit of tobacco and betel quid as a consequence carcinogenic compounds remain directly in contact with buccal mucosa for a longer duration. Thus it is expected to have high number of mutational changes in these cases as revealed by our study also in which we found large number of InDels in addition to known and novel SNPs, most of these variations reported for the first time to be associated with oral carcinogenesis. Thus identification of these mutational changes may be of great significance in understanding the underlying biology of oral carcinogenesis, as well as in the design of clinically useful therapeutic biomarkers for oral cancer.
Chapter 6

Differential Gene Expression Profiling of Oral Squamous Cell Carcinoma and Adjacent Noncancerous Tissues with cDNA Microarray in North East Indian Population
INTRODUCTION:

Molecular diagnostics is a rapidly advancing field in which insights into disease mechanisms are being elucidated by use of new gene-based biomarkers. Until recently, diagnostic and prognostic assessment of diseased tissues and tumors relied heavily on indirect indicators that permitted only general classifications into broad histologic or morphologic subtypes and did not take into account the alterations in individual gene expression. Global expression analysis using microarrays now allows for simultaneous interrogation of the expression of thousands of genes in a high-throughput fashion and offers unprecedented opportunities to obtain molecular signatures of the state of activity of diseased cells and patient samples. Several studies have shown that cancer diagnosis based on microarray data can effectively reveal the crucial processes underlying carcinogenesis (Wong and Wang, 2008).

OSCC is the sixth most common malignancy in humans, and the mortality rate remains high at approximately 50% with a particularly poor 5-year survival rate which has not improved significantly in the last 40 years. In India oral cancer is the most common cancer among men and ranks third among women (Soya et al., 2007), with age-standardized incidence rates per 100,000 population to be 12.8 and 7.5 respectively (Nair et al., 2004). Early indicators of oral cancer are oral leukoplakia and submucous fibrosis with transformation rate of 2-12% to frank malignancies (Anantharaman et al., 2007). In the western countries, smoking and alcohol consumption are considered to be the main risk factors while in India, smokeless tobacco products and betel quid (BQ) with or without tobacco are the major risk factors for oral cavity cancer (Jemal et al., 2011). Buccal mucosa represents the primary site for oral squamous cell carcinoma (OSCC) among chewers of tobacco and betel quid, contrary to tongue cancer, which represents the primary site of cancers in Western countries where cigarette smoking and heavy alcohol consumption are the main causative factors. Prevalence of aerodigestive tract cancers including oral cancer was reported to be highest in some northeastern (NE) regions of India (Bhattacharjee et al., 2006;
ICMR-Report, 2006; Phukan et al., 2004). In this region there is widespread chewing habit of tobacco with peculiarly fermented betel nut, which may further add to the risk for oral cancer in this region (Phukan et al., 2001).

Oral cancer is a highly aggressive malignancy involving multistep carcinogenesis process and requires accumulation and interplay of a series of molecular events. It is well established that in the process of oral carcinogenesis large number of genes are involved in the development and progression of cancer that leads to the functional changes in cells such as cell proliferation, evasion of apoptosis, angiogenesis, tissue invasion and metastasis. Understanding the genetic processes and biological pathways involved in the development of OSCC might lead to valuable information that might improve disease classification, early detection and diagnosis, as well as therapeutic planning and drug development (Nagpal and Das, 2003; Reibel, 2003). Microarrays represent a promising tool that makes it possible to explore the expression profile of thousands of genes simultaneously, at the RNA level (Otero-Rey et al., 2004; Russo et al., 2003). In the literature, there are several microarray studies on OSCCs with promising findings (Choi and Chen, 2005; Otero-Rey et al., 2004; Russo et al., 2003). Although the influence of life-style factors such as tobacco, alcohol use and nutrition are important to consider in the causation of OSCCs, there is a wide inter-individual differences in susceptibility to chemical carcinogens.

So to discover reliable prognostic markers and to identify molecular targets in OSCC for subsequent innovative therapies, high throughput analysis of genetic changes is required to find out some reliable answers for one of the biggest health problems of India. As the deregulated expression of genes is the major factor responsible for carcinogenesis, analysis of gene expression profiling using microarray technology may lead us towards identification of functionally relevant pathways the molecular targets in carcinogenesis and subsequent development of innovative methods for diagnosis, prognosis and therapy of cancer. In the present study we did gene expression profiling in tobacco associated oral cancer cases using microarray followed by the validation of the significantly deregulated genes by quantitative real time RT-PCR method in northeast Indian population.
MATERIALS AND METHODS:

Selection of patients and collection of samples:

One hundred and seventy patients with oral cancer registered at Dr B. Barroah Cancer Hospital, Guwahati from 2006 to 2009 were included in this study; tissue samples from 32 patients undergoing surgical treatment for oral cancer were included in gene expression profiling study. Within 10-15 minutes of surgical removal the tissue was collected in RNA later (Ambion, USA) from the tumour site as well as normal appearing site distant from the tumour area. One part of the tissue sample was used for histopathogic processing. Samples stored in RNA Later were frozen at -70°C till further processed. All 32 samples had a confirmed histopathologic diagnosis of OSCC. Detailed questionnaire with specific information regarding dietary, smoking, alcohol consumption habits and family history of cancer was completed for all patients. Informed consent was obtained from all the patients to use their specimens and clinicopathologic data for this study. Approval for this study had earlier been obtained from the Institutional Human Ethics Committee.

Isolation of RNA from tissue samples:

Five paired tissue samples obtained from tumor tissue and normal appearing buccal mucosa were selected for microarray experiments. Total RNA was isolated using RNeasy mini kit (Qiagen, Valencia, CA) following manufacture’s instructions. RNA quality was checked by on agarose gel. RNA integrity number was determined using the RNA 6000 Nano LabChip on the Agilent 2100 Bioanalyzer (Figure 6.1). Good quality of RNA was confirmed by the presence of discrete bands of 28s and 18s RNA. Quantity of RNA was determined by the NanoDrop® ND-1000 UV-Vis spectrophotometer.

Five samples that showed RNA quantity to be > 500 ng and RNA integrity number of > 8 in the tumour tissue as well as in the normal appearing tissue distant from the tumour site were selected for microarray experiments. Only samples from patients with confirmed diagnosis of OSCC who gave history of tobacco and betel nut chewing were selected for microarray experiments to maintain uniformity of the experimental design. Validation of gene expression by real-time RT-PCR assay was performed in 27 pair of samples.
Differential Gene Expression Profiling of OSCC using Microarray

![Ladder with bands of 28s and 18s](Image)

**Fig 6.1**: A. RNA integrity analysed by Agilent Bio-analyser showed clear bands of 28s and 18s. B. Quality of RNA also analysed on agarose gel. C. Electropherogram by bioanalyser illustrating the discrete bands of 28s and 18s.

**Microarray experiments**

Equal quantity of RNA samples obtained from normal buccal mucosa were pooled as shown in **Figure 6.2**. ExpressArt® Amino Allyl mRNA amplification Kits (Ocimum Biosolution, Hyderabad, India) was used for labeling cRNA following manufacturer’s protocol. Differential gene expression of each of the five tumor tissues was compared with the pooled normal controls. For this, each cRNA from tumour samples and pooled controls were labeled with cyanine 3 and hybridized on human 'OciChip A' chip (Ocimum Biosolution, Hyderabad, India) which contained 20160 genes. The labeled and fragmented cRNAs were hybridized at 65°C for 17 h.
Fig 6.2: Experimental design: \( S_{1-5} \) and \( N_{1-5} \) indicate samples (\( S \)) and their corresponding normal (\( N \)) tissue RNA respectively. Cy-3\( S_{1-5} \) indicates the Cy-3 labeled tumor RNA samples. Sl.066 Normal; Sl.071 Normal; Sl.017 Tumor; Sl.018 Tumor; Sl.085 Tumor; Sl.072 Tumor; Sl.079 Tumor indicated barcode of microarray chips

**Microarray image acquisition and data analysis:**

Hybridized arrays were scanned at 5 \( \mu \)m resolution on an AFFYMETRIX 428\textsuperscript{TM} Array Scanner at 100\% laser power and 30\% PMT at 532 nm for Cy3-labeled samples. The resulting TIFF images were analyzed by R package and Genowiz\textsuperscript{TM} Software (Ocimun Biosolution, India). The expression data was filtered by removing the blank spots and controls spots on the chip and 19700 probes of the 20160 were used for further analysis. The data obtained by image processing showed positively skewed distribution for each array. In order to have across array comparison, the data was normalized using \( \log_2 \) transformation. Median centering and median absolute deviation (MAD) scaling was performed on each array. The simple pre-processing adjusted the mean intensity levels of each array to zero.

A threshold p-value of 0.09 was fixed such that all probes having p-value less than the threshold were declared as statistically significantly expressed across the two conditions. Moreover, the fold change for each probe was calculated as the difference between the mean log transformed intensity levels, thereby resulting into log fold change. The statistically significant probes with log fold change either greater than +1 or less than -1 (equivalent to 2-fold change) were considered to be biologically significant. The expression data on the up
and down regulated probes on seven experimental samples (five from tumour samples and two from normal controls) was considered for hierarchical clustering. Two-way hierarchical clustering was used with Euclidean distance as a measure of proximity and average linkage method were used to determine clustering.

**Gene Enrichment Analysis**

The probes obtained through the comparison were studied for their over abundance in different Gene Ontology (GO) terms as well as Pathways. The terms could be categorized into biological process, molecular function and cellular component. Fisher's exact test was used to determine the significance of the GO term. If a term was significant with say $p < 0.05$, then it was implied that it was enriched with genes. Accordingly, the biological relevance of the term and the associated genes could be explored.

**Functional annotation clustering using DAVID 6.7**

Functional annotation clustering was performed by using Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7), a web-based publically available bio-informatics tool ([www.david.abcc.ncifcrf.gov](http://www.david.abcc.ncifcrf.gov)).

**Pathway Analysis and Generation of Pathways:**

Analysis of relevant pathways was done using Genowiz™ software designed for gene expression data analysis which annotates genes and classifies them into functional categories (Gene Ontology).

Another important software used in our study for pathway analysis and generation of pathways using gene expression data was Ingenuity IPA software ([http://www.ingenuity.com/products/pathways_analysis.html](http://www.ingenuity.com/products/pathways_analysis.html)) which delivers a rapid assessment of the signaling and metabolic pathways, molecular networks, and biological processes that are most significantly perturbed in the dataset of interest.
RESULTS:

Clinical and epidemiological information of enrolled patients:

Five OSCC samples were compared with normal pooled buccal mucosa tissue were used for differential gene expression profiling using microarray. All patients were male and gave a history of tobacco consumption and betel nut chewing. None of them has family history of cancer (Table 6.1).

Table 6.1: Demographic and clinical characteristics of oral squamous cell carcinoma in male cases

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age</th>
<th>Betel Quid Consumption</th>
<th>Tobacco Chewing</th>
<th>Tobacco Smoking</th>
<th>Alcohol</th>
<th>Grade</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>70</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
</tbody>
</table>

++ = Frequently; + = Occasionally; - = Never
G1 = Well differentiated squamous cell carcinoma;
G2 = Moderately differentiated squamous cell carcinoma;
G3 = Poorly differentiated squamous cell carcinoma

Gene Expression Profiling by Oligonucleotide Microarray:

Differential gene expression profiling of tobacco and betel quid chewing associated OSCC samples from northeast Indian patients was done by hybridizing the cDNA from pairwise normal and tumor samples to Human OciChip 40kA (Ocimum Biosolutions, Hyderabad). The intensity for each gene on each array was obtained by referring to 'Signal Mean Intensity'. The empty spots and control probes were removed from each data set. This resulted into 19700 probes for downstream analysis.

To understand the raw intensity profiles for samples, a boxplot of raw values was obtained (Figure 6.3A). The boxplot for each sample clearly indicates a high positive skewness requiring transformation. A log2-transformation was used that resulted into normal
Differential Gene Expression Profiling of OSCC using Microarray

intensity profiles for all the samples. The boxplots after log2- transformation are shown in (Figure 6.3B).

Fig 6.3(A-B): Boxplots for raw and log2-tranformed intensities for samples. After transformation the data for each sample shows nearly symmetric expression profile (Normal distribution).
The profiles for normal samples are similar but intensity distribution for new sample is on a higher side as compared to the previous normal sample. In order to facilitate comparison across arrays, median centering and median absolute scaling was carried out for each array. This data was further used for down stream analysis. The resulting distribution profiles are shown in Figure 6.4.

Fig 6.4: Boxplots for each sample after centering and scaling the respective intensity data. Figure shows that the intensity distribution is quite uniform across the samples after centering and scaling of respective data.
A volcano plot was obtained to view the distribution of up and down-regulated genes (Figure 6.5). The same data has also been represented as scatter plot (Figure 6.6).

Fig 6.5: Volcano plot showing distribution of up and down regulated genes. Genes significantly up-regulated are shown by red colour dots while green colour dots showed significantly downregulated genes.

Fig 6.6: Scatter plot showing the distribution of up and down regulated genes. Genes significantly up-regulated are shown by red colour dots while green colour dots showed significantly downregulated genes.
Differential expression analysis:

The normalized intensity dataset was considered to obtain genes that are differentially expressed between the two conditions i.e. normal and tumor. To determine statistical significance of difference in the expression levels of each gene across two conditions, t-test for independent samples was used. The p-values for genes were obtained and those with value less than 0.09 were declared as having statistically significant difference in the two conditions. Further, a fold change (FC) criterion was used to select biologically meaningful genes. In the present context, log\(_2\) fold change of ±1 (equivalent to fold change of 2.0) was set and the genes that are statistically significant and having log\(_2\) fold change either above +1 or below -1 were declared as differentially expressed genes. Following these criteria a set of 634 differentially expressed (247 upregulated and 387 down-regulated) genes were identified. Of these top 25 highly significant up and down regulated genes are listed in Table 6.2A and Table 6.2 B.

Gene enrichment analysis:

Differentially expressed genes were categorized using the Gene Ontology database into known or probable functional categories on the basis of biological processes and molecular function using various softwares. Top 10 up and down regulated pathways/functional categories of differentially expressed genes using Genowiz™ software are shown in Table 6.3A and Table 6.3 B.

Hierarchical clustering

Two way hierarchical clustering was used with Euclidean distance as a measure of proximity and average linkage method were used to determine clustering. Both the samples and probes were clustered simultaneously and visualized through a heatmap (Figure 6.7). The heatmap reveals that samples have been appropriately grouped based on the gene subset.
Fig 6.7: Heatmap showing the relatedness within samples and within genes. It is evident from the map that samples have been appropriately grouped based on the gene subset. Red colour up-regulation and yellow represents similar expression in both tumor and normal tissue.
Functional annotation clustering using DAVID 6.7

Functional annotation clustering was performed by using Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7), a web-based publically available bio-informatics tool (www.david.abcc.ncifcrf.gov). Cluster 1 (total genes = 57) of up-regulated genes with highest enrichment score of 4.45, belongs to the functional category of structural constituents of ribosome, ribosomal biogenesis and translation activity. Cluster 2 (n=26), of up-regulated genes with enrichment score 2.74 belongs to Inflammatory response, defense response and response to wounding. Other important clusters of upregulated genes are shown in Table 6.4A.

Cluster 1 of down-regulated genes (n=28) with highest enrichment score of 4.05, is related to the structural constituents of cytoskeleton and keratin filament. Cluster 2 of down-regulated genes (n=20) with enrichment score 3.77 belongs to epidermal cell differentiation and keratinisation. Other clusters with enrichment score more than 1.5 belongs to categories such as cellular response to reactive oxygen species, Cell-cell adhesion and Glucosidase activity (Table 6.4B).

Pathway Analysis using IPA software (Ingenuity Analysis):

Ingenuity pathway analysis was done for differentially expressed genes with the following settings and filter summary:

Reference set: Ingenuity Knowledge Base (Genes Only), Relationship to include: Direct and Indirect Includes Endogenous Chemicals Optional Analysis: My Pathways My List

Filter Summary:

Considered only relationships where confidence = Experimentally Observed. This software gives interaction between genes which have been reported earlier experimentally. Ingenuity Pathways Analysis networks are displayed graphically as nodes (genes/gene products) and edges (the biological relationships between the nodes). Nodes are displayed using various shapes that represent the functional class of the gene product. The node color represents the fold change (green = upregulation in this case). The higher the fold increase, the more intense the color of green. Edges are displayed with various labels that describe the nature of the relationship between the nodes (e.g., B for binding, T for transcription). The
length of an edge reflects the evidence supporting that node-to-node relationship, in that edges supported by more articles from the literature are shorter.

**Top 5 Significantly Altered Networks:**

Significantly down-regualted networks observed by IPA belongs to functional categories such as ‘hair and skin development and function’, posttranslational modification, cellular compromise, protein folding etc (*Table 6.5A*). Networks of functional pathways of upregaulted genes belongs to categories such as ‘Cell to cell signaling and interaction’, cell death, cellular compromise, DNA replication, molecular transport, nucleic acid metabolism etc (*Table 6.5B*).

**Top 5 Significantly Altered Biological Functions:**

Significantly downregulated biological functions as analyzed by ‘Ingenuity pathway analysis’ belongs to functional category of Cell to cell signaling and interaction, Drug metabolism, Molecular transport, Small molecular biochemistry and Cell morphology (*Table 6.6A*). Significantly up-regulated biological functions include Cell death, Cellular growth and proliferation, Cell to cell signaling and interaction, Cellular assembly and organization and Cell cycle (*Table 6.6B*).

**Top 5 Significantly Altered Canonical Pathways:**

Significantly down-regulated canonical pathways include functions related to Stilbene, Coumarine and Lignin Biosynthesis, Valine, Leucine and Isoleucine Degradation, Signaling by Rho family GTPases, Tight Junction Signaling, and RhoGDI Signaling (*Table 6.7 A*). Significantly up-regulated canonical pathways include functions related to EIF2 Signaling, Regulation of eIF4 and p70S6K signaling; G-Protein coupled Receptor Signaling, Antigen presentation pathway and TREM1 signaling (*Table 6.7 B*).
<table>
<thead>
<tr>
<th>S. No.</th>
<th>Accession No.</th>
<th>Chromosome Location</th>
<th>Gene symbol</th>
<th>Gene Name</th>
<th>Log_FC</th>
<th>p-value (gene specific)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NM_002272.1</td>
<td>12q13.13</td>
<td>KRT4</td>
<td>Keratin 4</td>
<td>-3.89</td>
<td>0.0258</td>
</tr>
<tr>
<td>2</td>
<td>BC000458.2</td>
<td>2cen-q13</td>
<td>MAL</td>
<td>Mal, T-cell differentiation protein</td>
<td>-3.52</td>
<td>4.00E-004</td>
</tr>
<tr>
<td>3</td>
<td>NM_019843.2</td>
<td>22q11.2</td>
<td>EIF4ENIF1</td>
<td>Eukaryotic translation initiation factor 4E nuclear import factor 1</td>
<td>-3.01</td>
<td>0.0205</td>
</tr>
<tr>
<td>4</td>
<td>NM_005416.1</td>
<td>1q21-q22</td>
<td>SPRR3</td>
<td>Small proline-rich protein 3</td>
<td>-2.93</td>
<td>0.0106</td>
</tr>
<tr>
<td>5</td>
<td>NM_006717.1</td>
<td>7q11.23</td>
<td>GTF2I</td>
<td>General transcription factor II, i</td>
<td>-2.82</td>
<td>0.0667</td>
</tr>
<tr>
<td>6</td>
<td>BC075796.1</td>
<td>3p22.1</td>
<td>MOBP</td>
<td>Myelin-associated oligodendrocyte basic protein</td>
<td>-2.25</td>
<td>0.0021</td>
</tr>
<tr>
<td>7</td>
<td>BC000551.2</td>
<td>3q21.3</td>
<td>MGLL</td>
<td>Monoglyceride lipase</td>
<td>-1.92</td>
<td>0.0091</td>
</tr>
<tr>
<td>8</td>
<td>AY640415.1</td>
<td>15q22.1</td>
<td>TPM1</td>
<td>Tropomyosin 1 (alpha)</td>
<td>-1.85</td>
<td>0.0475</td>
</tr>
<tr>
<td>9</td>
<td>NM_001885.1</td>
<td>11q22.3-q23.1</td>
<td>CRYAB</td>
<td>Crystallin, alpha B</td>
<td>-1.75</td>
<td>0.0074</td>
</tr>
<tr>
<td>10</td>
<td>NM_015157.1</td>
<td>11q23.3</td>
<td>PHLD1</td>
<td>Pleckstrin homology-like domain, family B, member 1</td>
<td>-1.69</td>
<td>0.0478</td>
</tr>
<tr>
<td>11</td>
<td>NM_033199.3</td>
<td>3p21.3</td>
<td>UCN2</td>
<td>Urocortin 2</td>
<td>-1.69</td>
<td>0.0362</td>
</tr>
<tr>
<td>12</td>
<td>NM_052854.1</td>
<td>11p11.2</td>
<td>CREB3L1</td>
<td>cAMP responsive element binding protein 3-like 1</td>
<td>-1.67</td>
<td>0.0322</td>
</tr>
<tr>
<td>13</td>
<td>NM_203447.1</td>
<td>9p24.3</td>
<td>DOCK8</td>
<td>Dedicator of cytokinesis 8</td>
<td>-1.67</td>
<td>0.0299</td>
</tr>
<tr>
<td>14</td>
<td>NM_144670.2</td>
<td>12p13.31</td>
<td>A2ML1</td>
<td>Alpha-2-macroglobulin-like 1</td>
<td>-1.65</td>
<td>0.0189</td>
</tr>
<tr>
<td>15</td>
<td>NM_000691.3</td>
<td>17p11.2</td>
<td>ALDH3A1</td>
<td>Aldehyde dehydrogenase 3 family, member A1</td>
<td>-1.58</td>
<td>0.0085</td>
</tr>
<tr>
<td>16</td>
<td>NM_002705.3</td>
<td>16p13.3</td>
<td>PPL</td>
<td>Periplakin</td>
<td>-1.57</td>
<td>0.0037</td>
</tr>
<tr>
<td>17</td>
<td>NM_025054.3</td>
<td>8q13</td>
<td>VCP1P1</td>
<td>Valosin containing protein (p97)/p47 complex interacting protein 1</td>
<td>-1.53</td>
<td>0.0559</td>
</tr>
<tr>
<td>18</td>
<td>NM_000382.2</td>
<td>17p11.2</td>
<td>ALDH3A2</td>
<td>Aldehyde dehydrogenase 3 family, member A2</td>
<td>-1.5</td>
<td>0.0606</td>
</tr>
<tr>
<td>19</td>
<td>NM_032028.2</td>
<td>1p35-p34</td>
<td>TSSK3</td>
<td>Testis-specific serine kinase 3</td>
<td>-1.48</td>
<td>0.0694</td>
</tr>
<tr>
<td>20</td>
<td>NM_001261.2</td>
<td>9q34.1</td>
<td>CDK9</td>
<td>Cyclin-dependent kinase 9</td>
<td>-1.47</td>
<td>0.0783</td>
</tr>
<tr>
<td>21</td>
<td>NM_005553.2</td>
<td>11q13.5</td>
<td>KRTAP5-9</td>
<td>Keratin associated protein 5-9</td>
<td>-1.45</td>
<td>0.0166</td>
</tr>
<tr>
<td>22</td>
<td>NM_032268.3</td>
<td>16q23.1</td>
<td>ZNRF1</td>
<td>Zinc and ring finger 1</td>
<td>-1.33</td>
<td>0.0279</td>
</tr>
<tr>
<td>23</td>
<td>NM_000452.2</td>
<td>6q23</td>
<td>ARG1</td>
<td>Arginase, liver</td>
<td>-1.32</td>
<td>0.0879</td>
</tr>
<tr>
<td>24</td>
<td>BC003385.1</td>
<td>17q21.33</td>
<td>SPOP</td>
<td>Speckle-type POZ protein</td>
<td>-1.31</td>
<td>0.0012</td>
</tr>
<tr>
<td>25</td>
<td>NM_001307.3</td>
<td>17p13.1</td>
<td>CLDN7</td>
<td>Claudin 7</td>
<td>-1.31</td>
<td>0.0138</td>
</tr>
</tbody>
</table>
### Table 6.2 B: List of top 25 Significantly Up-regulated Genes in OSCC

<table>
<thead>
<tr>
<th>No.</th>
<th>Accession</th>
<th>Chrom. Location</th>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Log FC</th>
<th>p-Value (Gene Specific)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NM_032036.2</td>
<td>1q32.12</td>
<td>IFI27L2</td>
<td>Interferon, alpha-inducible protein 27-like 2</td>
<td>3.85</td>
<td>0.0862</td>
</tr>
<tr>
<td>2</td>
<td>Z15009.1</td>
<td>1q25-q31</td>
<td>LAMC2</td>
<td>Laminin, gamma 2</td>
<td>3.24</td>
<td>0.0629</td>
</tr>
<tr>
<td>3</td>
<td>NM_000924.2</td>
<td>1q13</td>
<td>PDE1B</td>
<td>Phosphodiesterase 1B, calmodulin-dependent</td>
<td>2.22</td>
<td>0.0525</td>
</tr>
<tr>
<td>4</td>
<td>NM_006643.2</td>
<td>9q34.3</td>
<td>SDCCAG3</td>
<td>Serologically defined colon cancer antigen 3</td>
<td>2.16</td>
<td>0.0129</td>
</tr>
<tr>
<td>5</td>
<td>NM_003937.2</td>
<td>2q22.2</td>
<td>KYNU</td>
<td>Kynureninase (L-kynurenine hydrolase)</td>
<td>1.84</td>
<td>0.0146</td>
</tr>
<tr>
<td>6</td>
<td>NM_002192.2</td>
<td>7p15-p13</td>
<td>INHBA</td>
<td>Inhibin, beta A</td>
<td>1.61</td>
<td>0.0314</td>
</tr>
<tr>
<td>7</td>
<td>AL359771.27</td>
<td>1p36.21</td>
<td>PDPN</td>
<td>Podoplanin</td>
<td>1.59</td>
<td>0.0053</td>
</tr>
<tr>
<td>8</td>
<td>NG_005169.1</td>
<td>3q13.13</td>
<td>PPIAP15</td>
<td>Peptidylprolyl isomerase A (cyclophilin A) pseudogene 15</td>
<td>1.54</td>
<td>0.035</td>
</tr>
<tr>
<td>9</td>
<td>NM_005005.1</td>
<td>8q13.3</td>
<td>NDUFB9</td>
<td>NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9</td>
<td>1.47</td>
<td>0.0032</td>
</tr>
<tr>
<td>10</td>
<td>NM_005848.2</td>
<td>4q13-q21</td>
<td>IL8</td>
<td>Interleukin 8</td>
<td>1.37</td>
<td>0.0298</td>
</tr>
<tr>
<td>11</td>
<td>NM_005849.1</td>
<td>16p12.2</td>
<td>IGSF6</td>
<td>Immunoglobulin superfamily, member 6</td>
<td>1.36</td>
<td>0.0158</td>
</tr>
<tr>
<td>12</td>
<td>XM_497621.1</td>
<td>19q12</td>
<td>LOC126</td>
<td>Similar to peptidyl-Pro cis trans isomerase</td>
<td>1.34</td>
<td>0.0158</td>
</tr>
<tr>
<td>13</td>
<td>NM_031458.1</td>
<td>3q21</td>
<td>PARP9</td>
<td>Poly (ADP-ribose) polymerase family, member 9</td>
<td>1.27</td>
<td>0.0418</td>
</tr>
<tr>
<td>14</td>
<td>AY460334.1</td>
<td>5q32-q33.1</td>
<td>TCOF1</td>
<td>Treacher Collins-Franceschetti syndrome 1</td>
<td>1.25</td>
<td>0.0469</td>
</tr>
<tr>
<td>15</td>
<td>NM_002164.3</td>
<td>8p12-p11</td>
<td>IDO1</td>
<td>Indoleamine 2,3-dioxygenase 1</td>
<td>1.22</td>
<td>0.0651</td>
</tr>
<tr>
<td>16</td>
<td>XM_060887.3</td>
<td>1p34.1</td>
<td>LOC128</td>
<td>Similar to peptidyl-Pro cis trans isomerase</td>
<td>1.19</td>
<td>0.0695</td>
</tr>
<tr>
<td>17</td>
<td>NM_004658.1</td>
<td>12q23-q24</td>
<td>RASAL1</td>
<td>RAS protein activator like 1 (GAP1 like)</td>
<td>1.15</td>
<td>0.0798</td>
</tr>
<tr>
<td>18</td>
<td>NM_030800.1</td>
<td>15q22.31</td>
<td>C15orf44</td>
<td>Chromosome 15 open reading frame 44</td>
<td>1.14</td>
<td>0.0365</td>
</tr>
<tr>
<td>19</td>
<td>NM_001011.3</td>
<td>2p25</td>
<td>RPS7</td>
<td>Ribosomal protein S7</td>
<td>1.13</td>
<td>0.0919</td>
</tr>
<tr>
<td>20</td>
<td>NM_016186.1</td>
<td>14q32.13</td>
<td>SERPIN10</td>
<td>Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 10</td>
<td>1.11</td>
<td>0.0539</td>
</tr>
<tr>
<td>21</td>
<td>NM_004296.3</td>
<td>14q24.3</td>
<td>RGS6</td>
<td>Regulator of G-protein signaling 6</td>
<td>1.09</td>
<td>0.0788</td>
</tr>
<tr>
<td>22</td>
<td>AL832151.1</td>
<td>22q11.2</td>
<td>MED15</td>
<td>Mediator complex subunit 15</td>
<td>1.07</td>
<td>0.0725</td>
</tr>
<tr>
<td>23</td>
<td>NM_181670.2</td>
<td>12q23.1</td>
<td>ANK3IB</td>
<td>Ankyrin repeat and sterile alpha Motif domain containing 1B</td>
<td>1.06</td>
<td>0.0998</td>
</tr>
<tr>
<td>24</td>
<td>NM_006837.2</td>
<td>8q13.1</td>
<td>COPS5</td>
<td>COP9 constitutive photomorphogenic homolog subunit 5 (Arabidopsis)</td>
<td>1.06</td>
<td>0.0276</td>
</tr>
<tr>
<td>25</td>
<td>NM_004833.1</td>
<td>1q22</td>
<td>AIM2</td>
<td>Absent in melanoma 2</td>
<td>1.06</td>
<td>0.0198</td>
</tr>
</tbody>
</table>
Pathway/Functional category analysis using Genowiz™ software:

**Table 6.3A: Significantly Up-regulated Pathways:**

<table>
<thead>
<tr>
<th>Pathway/Functional Category (*)</th>
<th>P value</th>
<th>Cluster of Genes Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Ribosome (7/94)</td>
<td>0.0002</td>
<td>RPL38, RPSA, RPL6, RPS3, RPS20, RPS6, <strong>RPS7</strong></td>
</tr>
<tr>
<td>2 Neuroactive ligand-receptor interaction (10/256)</td>
<td>0.0018</td>
<td><strong>GRM8, GRM4, NTSR1, P2RX7, TAAR2, FSHR, PARD3, NMBR, AVPR1B, TSHR</strong></td>
</tr>
<tr>
<td>3 ECM-receptor interaction (5/84)</td>
<td>0.004</td>
<td><strong>LAMC2, ITGB1, GP6 FN1, COL2A1</strong></td>
</tr>
<tr>
<td>4 Aminoacyl-tRNA biosynthesis (3/39)</td>
<td>0.015</td>
<td><strong>LARS2, AARS, WARS</strong></td>
</tr>
<tr>
<td>5 Focal adhesion (7/203)</td>
<td>0.0159</td>
<td><strong>VAV3, LAMC2, GRB2, ITGB1, BIRC3, FN1, COL2A1</strong></td>
</tr>
<tr>
<td>6 Bladder cancer (3/42)</td>
<td>0.018</td>
<td><strong>IL8, DAPK1, CDKN1A</strong></td>
</tr>
<tr>
<td>7 Proteasome (3/48)</td>
<td>0.025</td>
<td><strong>PSMB4, POMP, PSMB5</strong></td>
</tr>
<tr>
<td>8 Small cell lung cancer (4/86)</td>
<td>0.025</td>
<td><strong>LAMC2, ITGB1, BIRC3, FN1</strong></td>
</tr>
<tr>
<td>9 Cell adhesion molecules (CAMs) (5/133)</td>
<td>0.028</td>
<td><strong>VCAN, CD6, ITGB1, HLA-C, HLA-F</strong></td>
</tr>
<tr>
<td>10 Tryptophan metabolism (3/51)</td>
<td>0.029</td>
<td><strong>IDO1, KYNU, WARS</strong></td>
</tr>
</tbody>
</table>

Note: 1. * Number in brackets represent genes found significantly up regulated in the pathway out of total number of genes involved in that pathways

2. Highlighted (represented in bold) genes are those which appeared in the list of top 25 differentially up-regulated genes
### Table 6.3B: Significantly Down-regulated Pathways (analyzed using Genowiz™ software):

<table>
<thead>
<tr>
<th>Pathway/Functional category (*)</th>
<th>P value</th>
<th>Cluster of Genes Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cell junctions (12/141)</td>
<td>0.00004</td>
<td>KRT34, DSG1, KRT2, KRT15, KRT36, VIM, KRT4, KRT10, KRT78, KRT13, ACTB, KRT33A</td>
</tr>
<tr>
<td>2 Valine, leucine and isoleucine degradation (7/46)</td>
<td>0.00006</td>
<td>ALDH3A1, BCAT2, HIBCH, BCKDH, ACAT2, ALDH2, ALDH3A2</td>
</tr>
<tr>
<td>3 3-Chloroacrylic acid degradation (4/14)</td>
<td>0.0003</td>
<td>ALDH3A1, ADH7, ALDH2, ALDH3A2</td>
</tr>
<tr>
<td>4 Urea cycle and metabolism of amino groups (5/28)</td>
<td>0.0004</td>
<td>ARG1, ALDH3A1, GATM, ALDH2, ALDH3A2</td>
</tr>
<tr>
<td>5 Glycolysis / Gluconeogenesis (6/64)</td>
<td>0.0023</td>
<td>ENO3, ALDH3A1, ADH7, ALDH2, GPI, ALDH3A2</td>
</tr>
<tr>
<td>6 Tyrosine metabolism (5/46)</td>
<td>0.0029</td>
<td>ALDH3A1, ADH7, MIF, TYRP1, COMT</td>
</tr>
<tr>
<td>7 Butanoate metabolism (4/39)</td>
<td>0.0093</td>
<td>ALDH3A1, ACAT2, ALDH2, ALDH3A2</td>
</tr>
<tr>
<td>8 Pyruvate metabolism (4/42)</td>
<td>0.0117</td>
<td>ALDH3A1, ACAT2, ALDH2, ALDH3A2</td>
</tr>
<tr>
<td>9 Tight junction (7/135)</td>
<td>0.0203</td>
<td>CLDN6, MYL2, CLDN14, CLDN7, PPP2R1A, MYH2, ACTB</td>
</tr>
<tr>
<td>10 Valine, leucine and isoleucine biosynthesis (2/11)</td>
<td>0.0251</td>
<td>VARS2, BCAT2</td>
</tr>
</tbody>
</table>

**Note:**
1. * Number in brackets represent genes found significantly down-regulated in the pathway out of total number of genes involved in that pathways.
2. Highlighted (represented in bold) genes are those which appeared in the list of top 25 differentially down-regulated genes.
Table 6.4A: Functional Annotation Clustering of Up-regulated Genes by (analyzed using DAVID 6.7 software)

<table>
<thead>
<tr>
<th>Classification Term</th>
<th>Enrichment Score</th>
<th>No. of Components</th>
<th>P value</th>
<th>Benjamini</th>
<th>Cluster of Genes Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AARS, ANXA1, BFSP1, BIRC5, CCT4, DKN1A, COL19A1, COL2A1, COP5S, DLG1, FN1, GRB2, HAL, HNRNPL, HNRNPu, HNRPD1, HPCAL1, IDO1, KRT14, KYNU, LARS2, NOP10, NUPL2, ODC1, PARK7, PDE1B, PDLM3, PKP1, POMP, PRKAR1B, PSMB5, RBM5, RPL10, RPL23, RPL38, RPL6, RPL7A, RPLP0, RPS14, RPS20, RPS3, RPS6, RPS7, RPSA, RTCD1, SCLY, SNRNP70, SPTAN1, STAT1, TARBP2, UGP2, UF1, UTP14A, YAV3, WARS, WDFY1</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>2.74</td>
<td>26</td>
<td>0.0009</td>
<td>0.05</td>
<td>ANXA1, APOA4, CCNL2, CCR2, CCR3, CFH, CRP, FN1, GP6, HLA-C, IDO1, IL8, INHBA, ITGB1, KYNU, LTF, P2RX7, P2RX7, PAGE1, PDPN, PPBP, SERPINA10, SOD2, SP100, TLR1, VCAN</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>2.04</td>
<td>22</td>
<td>0.008</td>
<td>0.060</td>
<td>AARS, DDX39, HNRNPL, HNRNPu, HNRPD1, HPCAL1, LARS2, NOP10, RBM5, RPL7A, RPLP0, RPS14, RPS20, RPS3, RPS6, RPS7, RPSA, RTCD1, SNRNP70, TARBP2, UTP14A, WARS</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>1.58</td>
<td>15</td>
<td>0.0011</td>
<td>0.826</td>
<td>B2M, CFH, FN1, GP6, HFE, HFE, HLA-C, HLA-C, HLA-C, HLA-F, IGF6, INHBA, ITGB1, POMP, TLR1, VCAN</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>1.41</td>
<td>68</td>
<td>0.007</td>
<td>0.344</td>
<td>ACCN3, ADCY9, ANKS1B, ANXA1, APOA4, AVPR1B, C1ORF116, CCNL2, CCR2, CCR3, CCR4, CNM4, COL2A1, COX8A, DLG1, DLL1, FSHR, FST, GP6, GRB2, GRM4, GRM8, HFE, HLA-C, HLA-F, HSD3B1, IDO1, IFI27, IGF6, IL13RA2, IL15RA, IL8, INHBA, ITGB1, LTF, LY6K, MAGI2, NMBR, NTSR1, P2RX7, PAGE1, PARD3, PARK7, PDE1B, PDPN, PPBP, PRKAR1B, PTPN2, PTPRR, PTPRZ1, RARG, RGS6, RPS14, RPS6, RPSA, SLC38A2, SLC7A9, SLC9A8, SOD2, STAT1, STRN, TAAR2, TLR1, TMPRSS11E, TRPV3, TSHR, TXNDC12, VAV3</td>
</tr>
<tr>
<td>Cluster 6</td>
<td>1.40</td>
<td>57</td>
<td>0.048</td>
<td>0.832</td>
<td>ACCN3, ADCY9, AVPR1B, BIRC5, C1ORF116, CCNL2, CCR2, CCR3, CCR4, CNM4, COX8A, DLG1, DLL1, FSHR, FST, GP6, GRB2, GRM4, GRM8, HFE, HLA-C, HLA-F, HSD3B1, IDO1, IFI27, IL13RA2, IL8, INHBA, ITGB1, NMBR, NQO1, NTSR1, P2RX7, PARD3, PARK7, PDE1B, PDPN, PPBP, PRKAR1B, PSMB4, PSMB5, PTPN2, PTPRR, PTPRZ1, RARG, RGS6, RPSA, SLC38A2, SLC7A9, SOD2, STAT1, STRN, TAAR2, TARB2, TLR1, TMPRSS11E, TSHR</td>
</tr>
</tbody>
</table>

Note: Highlighted genes are those which appeared in the list of top 25 differentially up-regulated genes.
### Table 6.4B: Functional Annotation Clustering of Down-regulated Genes (analyzed using DAVID 6.7 software)

<table>
<thead>
<tr>
<th>Classification Terms</th>
<th>Enrichment Score</th>
<th>No. of components</th>
<th>p-Value</th>
<th>Benjamini</th>
<th>Cluster of Genes Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>Keratin filament, Cytoskeleton structural constituents</td>
<td>4.05</td>
<td>28</td>
<td>0.00002</td>
<td>0.003</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>Epidermal cell differentiation, Keratinisation</td>
<td>3.77</td>
<td>20</td>
<td>0.00006</td>
<td>0.035</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>Lipid and amino acid metabolic processes</td>
<td>2.14</td>
<td>20</td>
<td>0.003</td>
<td>0.270</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>Cell fraction, membrane fraction</td>
<td>2.11</td>
<td>35</td>
<td>0.005</td>
<td>0.243</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>Cellular response to reactive oxygen species</td>
<td>2.01</td>
<td>24</td>
<td>0.0002</td>
<td>0.096</td>
</tr>
<tr>
<td>Cluster 6</td>
<td>Oxidation reduction</td>
<td>1.87</td>
<td>20</td>
<td>0.003</td>
<td>0.188</td>
</tr>
<tr>
<td>Cluster 7</td>
<td>Glucosidase activity</td>
<td>1.72</td>
<td>3</td>
<td>0.015</td>
<td>0.743</td>
</tr>
<tr>
<td>Cluster 8</td>
<td>Cell-cell adhesion</td>
<td>1.60</td>
<td>23</td>
<td>0.002</td>
<td>0.213</td>
</tr>
<tr>
<td>Cluster 9</td>
<td>Cytoskeletal part</td>
<td>1.58</td>
<td>60</td>
<td>0.062</td>
<td>0.625</td>
</tr>
</tbody>
</table>

**Note:** Highlighted genes are those which appeared in the list of top 25 differentially down-regulated genes.
Pathways Analysis by IPA software (Ingenuity Pathway Analysis):

**TOP NETWORKS:**

**Table 6.5A: Top 5 networks of down-regulated genes and their functional categories as revealed by IPA analysis**

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Associated Network Functions</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.1</td>
<td>Embryonic development, hair and skin development and function, organ development</td>
<td>61</td>
</tr>
<tr>
<td>P.2</td>
<td>Posttranslational modification, cellular compromise, Protein folding</td>
<td>37</td>
</tr>
<tr>
<td>P.3</td>
<td>Small molecule biochemistry, cell cycle, carbohydrate metabolism</td>
<td>36</td>
</tr>
<tr>
<td>P.4</td>
<td>Cell to cell signaling interaction, Nervous system development and function, Behavior</td>
<td>36</td>
</tr>
<tr>
<td>P.5</td>
<td>Inflammatory response, cell death, cell to cell signaling and interaction</td>
<td>29</td>
</tr>
</tbody>
</table>

**Table 6.5B: Top 5 networks of up-regulated genes and their functional categories as revealed by IPA analysis**

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Associated Network Functions</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.1</td>
<td>Cell to cell signaling and interaction, Tissue development, hematological system development and function</td>
<td>37</td>
</tr>
<tr>
<td>P.2</td>
<td>Cell death, cellular compromise, DNA replication, Recombination and repair</td>
<td>35</td>
</tr>
<tr>
<td>P.3</td>
<td>Cell signaling, molecular transport, Nuclied acid metabolism</td>
<td>30</td>
</tr>
<tr>
<td>P.4</td>
<td>Cellular assembly and organization, DNA replication, Recombination and repair, Cellular function and maintenance</td>
<td>27</td>
</tr>
<tr>
<td>P.5</td>
<td>Cell to cell signaling and interaction, Hematological system development and function, Immune cell trafficking</td>
<td>26</td>
</tr>
</tbody>
</table>
**TOP ALTERED BIOLOGICAL FUNCTIONS (Analyzed by IPA):**

**Table 6.6A: Downregulated pathways by IPA (Molecular and cellular function):**

<table>
<thead>
<tr>
<th>Functional Category</th>
<th>p-Value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cell to cell signaling and interaction</td>
<td>4.73E-05 – 1.70E-02</td>
<td>24</td>
</tr>
<tr>
<td>2 Drug metabolism</td>
<td>4.73E-05 – 1.70E-02</td>
<td>3</td>
</tr>
<tr>
<td>3 Molecular transport</td>
<td>4.73E-05 – 1.70E-02</td>
<td>50</td>
</tr>
<tr>
<td>4 Small molecular biochemistry</td>
<td>4.73E-05 – 1.70E-02</td>
<td>52</td>
</tr>
<tr>
<td>5 Cell morphology</td>
<td>5.69E-05 – 1.70E-02</td>
<td>31</td>
</tr>
</tbody>
</table>

**Table 6.6B: Up-regulated Pathways by IPA (Molecular and cellular function):**

<table>
<thead>
<tr>
<th>Functional Category</th>
<th>p-Value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cell death</td>
<td>1.89E-06 - 1.15E-02</td>
<td>66</td>
</tr>
<tr>
<td>2 Cellular growth and proliferation</td>
<td>1.90E-06 - 1.109E-02</td>
<td>66</td>
</tr>
<tr>
<td>3 Cell to cell signaling and interaction</td>
<td>1.82E-05 – 1.11E-02</td>
<td>41</td>
</tr>
<tr>
<td>4 Cellular assembly and organization</td>
<td>1.82E-05 – 9.46E-03</td>
<td>27</td>
</tr>
<tr>
<td>5 Cell cycle</td>
<td>5.08E-05 – 1.15E-02</td>
<td>13</td>
</tr>
</tbody>
</table>
### TOP CANONICAL PATHWAYS (As analyzed by IPA):

**Table 6.7A: Top 5 Canonical Pathways of Downregulated Genes**

<table>
<thead>
<tr>
<th>Functional Category</th>
<th>p-Value</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Stilbene, Coumarine and Lignin Biosynthesis</td>
<td>5.13E-07</td>
<td>6/74 (0.081)</td>
</tr>
<tr>
<td>2 Valine, Leucine and Isoleucine Degradation</td>
<td>1.54E-05</td>
<td>8/108 (0.074)</td>
</tr>
<tr>
<td>3 Signaling by Rho family GTPases</td>
<td>4.7E-05</td>
<td>14/256 (0.055)</td>
</tr>
<tr>
<td>4 Tight Junction Signaling</td>
<td>7.4E-05</td>
<td>11/164 (0.067)</td>
</tr>
<tr>
<td>5 RhoGDI Signaling</td>
<td>2.16E-04</td>
<td>11/201 (0.055)</td>
</tr>
</tbody>
</table>

**Table 6.7B: Top 5 Canonical Pathways of Up-regulated Genes**

<table>
<thead>
<tr>
<th>Functional Category</th>
<th>p-Value</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 EIF2 Signaling</td>
<td>2.04E-07</td>
<td>13/199 (0.065)</td>
</tr>
<tr>
<td>2 Regulation of eIF4 and p70S6K signaling</td>
<td>2.53E-04</td>
<td>8/170 (0.047)</td>
</tr>
<tr>
<td>3 G-Protein coupled Receptor Signaling</td>
<td>9.07E-04</td>
<td>15/530 (0.028)</td>
</tr>
<tr>
<td>4 Antigen presentation pathway</td>
<td>1.13E-03</td>
<td>4/43 (0.093)</td>
</tr>
<tr>
<td>5 TREM1 signaling</td>
<td>3.44E-03</td>
<td>4/66 (0.061)</td>
</tr>
</tbody>
</table>
Networks Identification and Design by IPA (Ingenuity Pathway Analysis):

When differentially expressed genes were analyzed using IPA (Ingenuity pathway analysis), first network of genes which is shown below (Figure 6.8) was found to be associated with functions such as hair and skin development and organ development. Important genes present within this network include KRT4, KRT10, KRT13, ING1 and SPRR3.

![Network-1](image)

**Fig 6.8:** Network-1, of down-regulated genes which is functionally associated with embryonic development, hair and skin development, and organ development. IPA analysis displayed graphically as nodes (genes/gene products) and edges (the biological relationships between the nodes). Nodes are displayed using various shapes that represent the functional class of the gene product. The length of an edge reflects the evidence supporting that node-to-node relationship, in that edges supported by more articles from the literature are shorter.
Second important network of downregulated genes which is shown in Figure 6.9 was found to be associated with functions such as posttranslational modification, cellular compromise, and protein folding. Important genes present within this network include CRYAB, DOCK8, and RPL6.

Fig 6.9: Network-2, of down-regulated genes which is functionally associated with posttranslational modification, cellular compromise, protein folding
Third important network of downregulated genes which is shown Figure 6.10 was found to be associated with functions such as small molecule biochemistry, cell cycle, carbohydrate metabolism. Important genes present within this network include ALDH3A1, ALDH3A2, PTPN5 etc.

Fig 6.10: Network-3, of down-regulated genes which is functionally associated with Small molecule biochemistry, cell cycle, carbohydrate metabolism
Fourth network of downregulated genes which is shown Figure 6.11 was found to be associated with functions such as Cell to cell signaling interaction; Nervous system development and function; Behavior. Genes present within this network include MAL, ARG1, DSG1 etc.

Fig 6.11: Network-4, of down-regulated genes which is functionally associated with Cell to cell signaling interaction, Nervous system development and function, Behavior
Fifth important network of downregulated genes which is shown Figure 6.12 was found to be associated with functions such as inflammatory response, cell death, cell to cell signaling and interaction. Genes present within this network include GAS7, NFkB, TLR10 etc.

Networok of upregulated genes (Figure 6.13) with highest score was found to be associated with Cell to cell signaling and interaction, Tissue development, hematological system development and function. This network include genes such as RPL7A, RPS6, RPL23, Rnr, DCN, VAV3, PTPRR, C1q, CRP, FST.
Netwrok of upregualted genes (Figure 6.14) with second highest score was found to be associated with Cell death, cellular compromise, DNA replication, Recombination and repair. This network include genes such as COPS5, NQO1, RPL6, KRT14 etc.

Fig 6.14: Upregulated Network-2, which is functionally associated with Cell death, cellular compromise, DNA replication, Recombination and repair.

Third important netwrok of upregualted genes which is shown in Figure 6.15 was functionally associated with Cell signaling, Molecular transport, and Nucleic acid metabolism. This network includes genes such as DOCK5, GRM4, GRM8, CCT4, PDPN etc.

Fig 6.15: Upregulated Network-3, which is functionally associated with Cell signaling, Molecular transport, and Nucleic acid metabolism.
Upregulated network-4, is functionally associated with Cellular assembly and organization, DNA replication, Recombination and repair, Cellular function and maintenance. Genes involved in this network include *IL8, RASAL1, CCNL2* etc.

Fig 6.16: Upregulated Network-4, which is functionally associated with Cellular assembly and organization, DNA replication, Recombination and repair, Cellular function and maintenance.
Upregulated Netwrok-5, which is functionally associated with Cell to cell signaling and interaction, Hematological system development and function, Immune cell trafficking. Genes include IFNβ, IDO1, TLR1, HLA-C, HLA-F, B2M, POMP, PSMB5 etc

![Diagram of network-5](image)

**Fig 6.17:** Upregulated Netwrok-5, which is functionally associated with Cell to cell signaling and interaction, Hematological system development and function, Immune cell trafficking - Include IFN Beta, IDO1, TLR1, NFKB complex, Tlr, Interferon alpha, IFNBeta, HLA-C, HLA-F, B2M etc

IPA analysis for identifying altered biological function revealed five most significantly altered and downregulated biological functional categories belong to cell to cell signaling and interaction; Drug metabolism, Molecular transport, Small molecular biochemistry and Cell morphology. Significantly up-regulated biological functions are associated with Cell death, Cellular growth and proliferation, Cell to cell signaling and interaction, Cellular assembly and organization and Cell cycle.
DISCUSSION:

In this study, we analyzed gene expression profiles of OSCC cases from Guwahati (Assam) of Northeast region of India. We identified a large number of genes, including genes previously associated with tumorigenesis and new candidate genes. Six hundred thirty four differentially expressed genes were identified between tumors and normal control in which 247 (39%) were upregulated and 387 (61%) were downregulated. Significantly differentially expressed genes include genes which were reported to be associated with oral carcinogenesis by various studies as well as there are some genes such as Dock8 and SPRR3 which are reported for the first time by the present study to be associated with oral carcinogenesis.

When pathway analysis was done using Genowiz\textsuperscript{TM} software several functionally relevant pathways were found to be significantly altered in our study. Most significantly upregulated pathway is related to ribosomal activity. Functional annotation clustering of upregulated genes using web-based DAVID analysis also showed most significantly enriched cluster to be associated with ribosomal activity, ribosome biogenesis, and translation.

Ribosome biogenesis and translation can be simplified as the processes of generating ribosomes and their use for decoding mRNA into a protein (Brina \textit{et al.}, 2011). Cell growth and proliferation demands new protein synthesis and ribosomes provide the basis for protein production and are vital for cell growth and survival. Precise regulation of ribosome biogenesis is fundamental to maintain normal cell growth and proliferation, and accelerated ribosome biogenesis is associated with malignant transformation (Chan \textit{et al.}, 2011). Indeed, in proliferating cells the amount of cell constituents must be increased to ensure that daughter cells have the necessary complement for survival and normal functioning (Conlon and Raff, 1999; Derenzini and Ploton, 1991; Thomas, 2000). This is accomplished by increased synthesis of proteins that, in turn, is induced by an upregulation of the rate of ribosome production. These alterations are usually done to meet the needs of cancer cells, (Maggi and Weber, 2005; Ruggero and Pandolfi, 2003), which are usually characterized by high protein synthesis and rRNA transcription rates. Up-regulation of ribosome biogenesis has been associated with alterations in cell cycle, cell proliferation and cell growth and often contributes to increased susceptibility to cancer (Freed \textit{et al.}, 2010; Montanaro \textit{et al.}, 2008; Ruggero and Pandolfi, 2003).

The ultimate products of ribosome biogenesis, the ribosomes, act as the central players of the translation of mRNAs into proteins. Deregulation of mRNA translation has
already been demonstrated to trigger neoplastic transformation. Up-regulated ribosome biogenesis rate might be responsible for changes in the balance of the translational processes, with an increased quantity of ribosomes making it easier to translate those mRNAs (whose product might be involved in tumorigenesis) but that were not, or to a lesser degree, translated in normal conditions (Montanaro et al., 2008). Several ribosomal proteins are overexpressed in a variety of tumors, (Bilanges and Stokoe, 2007; Maggi and Weber, 2005; Ruggero and Pandolfi, 2003), but it is not yet clear if these alterations are a cause or a result of tumorigenesis (Freed et al., 2010). RPS and RPL group of genes which were found to be up-regulated in our study (RPS7, RPS6, RPL38, RPSA, RPL6, RPS3, and RPS20), were reported earlier also for their functional relevance while functions of genes such as KNYU and AARS were not well studied in relation to carcinogenesis. RPS7 was earlier reported to be involved in the maturation of the 18S rRNA and is required for the assembly of the 40S subunit. RPS7 can also modify p53 activity (Freed et al., 2010). Thus the genes involved in this pathway can be further studied for their functional relevance with oral carcinogenesis.

Second significantly up-regulated pathway in OSCC which was identified in our study was neuroactive ligand-receptor interaction. Expression study of this pathway was not very well reported in relation to carcinogenesis. However it was found to be altered in few cancers such as hepatocellular carcinoma (Wang et al., 2010) and esophageal squamous cell carcinoma (Chattopadhyay et al., 2007). Thus genes found to be up-regulated in our study (GRM4, GRM8, NTSR1, P2RX7, TAAR2, FSHR, PARD3, NMBR, AVPR1B, and TSHR) and involved in this pathway may be further studied for their functional relevance with oral carcinogenesis. Overexpression of these genes involved in neuroactive ligand-receptor interaction was first time reported to be associated with oral cancer.

Another up-regulated pathway is ECM-receptor interaction (with gene components of LAMC2, ITGB1, GP6, FN1 and COL2A1). The extracellular matrix (ECM) consists of a complex mixture of structural and functional macromolecules and serves an important role in tissue and organ morphogenesis and in the maintenance of cell and tissue structure and function. Specific interactions between cells and the ECM are mediated by transmembrane molecules, mainly integrins and perhaps also proteoglycans, CD36, or other cell-surface-associated components. These interactions lead to a direct or indirect control of cellular activities such as adhesion, migration, differentiation, proliferation, and apoptosis. In addition, integrins function as mechanoreceptors and provide a force-transmitting physical link between the ECM and the cytoskeleton. Integrins are a family of glycosylated,
heterodimeric transmembrane adhesion receptors that consist of noncovalently bound alpha- and beta-subunits. Integrating signals from the ECM via the cell surface into the nucleus is an essential feature of multicellular life and often malfunctions in cancer (Willier et al., 2011). Overexpression of ECM pathway was also reported to be associated with carcinogenesis but in very few recent reports (Krupp et al., 2011; Tsoi et al., 2011) and in cancers such as colorectal cancer (Takemasa et al., 2011), gastric cancer (Yang et al., 2007), clear cell renal cell carcinoma (Zhou et al., 2011), endometrial cancer (Du et al., 2008) and one study showed its overexpression to be associated with head and neck squamous cell carcinomas (Roman et al., 2008). Thus its association with OSCC will provide considerable insights about oral carcinogenesis.

Role of aminoacyl-tRNA synthetases (AARSs) have been unexplored in association with carcinogenesis, mostly because many researchers assumed that they were simply 'housekeepers' that were involved in protein synthesis (Ibba and Soll, 2000). AARSs are essential proteins of all living organisms. The function of aminoacyl-tRNA synthesis is to precisely match amino acids with tRNAs containing the corresponding anticodon. This is achieved by direct attachment of an amino acid to the corresponding tRNA by an aminoacyl-tRNA synthetase, although intrinsic proofreading and extrinsic editing are also essential in several cases (Ibba and Soll, 2000). This is completed by virtue of matching the nucleotide triplet of anticodon with cognate amino acid (Skupinska et al., 2009). Thus they ensure the reliability of transfer of genetic information from the DNA into the protein. AARSs are also engaged in the other crucial cellular processes. So the disturbance of function of any of them often causes serious disorders. Genes encoding different AARSs have been highly expressed or modified in association with a variety of cancers (Park et al., 2008). However genes involved in this pathway and observed to be overexpressed in our study (LARS2, AARS, WARS) were also reported by some other studies such as LARS2 in association with tongue squamous cell carcinoanoma (Ye et al., 2008a). Thus these genes may play an important role in oral carcinogenesis and may be further studied for their significance to work as prognostic markers for OSCC.

Upon detachment from the extracellular matrix, tumor epithelial cells and tumor-associated endothelial cells are capable of gaining survival benefits, and hence contribute to the process of metastasis. The focal-adhesion complex formation recruits the association of key adaptor proteins such as FAK (focal-adhesion kinase). Signalling downstream of growth factor receptors and integrins converge on the ubiquitously expressed non-receptor tyrosine
Differential Gene Expression Profiling of OSCC using Microarray

kinase FAK. FAK is involved in endothelial cell proliferation, migration and survival, is upregulated in many cancers and has recently been shown to control tumour angiogenesis. Indeed, FAK inhibitors are presently being developed for the treatment of cancer (Lechertier and Hodivala-Dilke, 2011). Thus genes observed to be overexpressed in this pathway (VAV3, LAMC2, GRB2, ITGB1, BIRC3, FN1, COL2A1) in our study will enhance our understanding about oral carcinogenesis.

Other than the above discussed pathways, significantly upregulated genes were reported to be associated with functional category/pathway of bladder cancer (IL8, DAPK1, CDKN1A), proteasome (PSMB4, POMP, PSMB5), small cell lung cancer (LAMC2, ITGB1, BIRC3, FN1), cell adhesion molecules (VCAN, CD6, ITGB1, HLA-C, HLA-F), and tryptophan metabolism pathway (IDO1, KYNU, WARS). Many of these genes such as IL8, LAMC2, and ITGB1 have been reported earlier to be associated with oral carcinogenesis while some others such as DAPK1, VCAN, and KYNU were reported for the first time to be overexpressed in OSCC in our study.

Most significantly downregulated pathway in our study (when analyzed by Genowiz) was related to cell junctions (Involving 12 genes KRT34, DSG1, KRT2, KRT15, KRT36, VIM, KRT4, KRT10, KRT78, KRT13, ACTB, and KRT33A) while DAVID analysis revealed clusters with highest enrichment score of down-regulated genes to be associated with structural constituents of cytoskeleton, epidermal cell differentiation, and Keratinisation.

Cell junctions are links between neighbouring epithelial cells and are important for barrier formation. Recent evidence suggests that cell junctions also participate in signal transduction mechanisms that regulate epithelial cell proliferation, gene expression, differentiation and morphogenesis (Matter and Balda, 2007). It is now well-established that dysregulation of these functions contributes to initiation and progression of cancer. The cytoskeleton is a complex of detergent-insoluble components of the cytoplasm playing critical roles in cell motility, shape generation, and mechanical properties of a cell. Fibrillar polymers-actin filaments, microtubules, and intermediate filaments- are major constituents of the cytoskeleton, which constantly change their organization during cellular activities (Svitkina, 2009). Thus changes in the expression levels of genes involved in cell junctions and constituents of cytoskeleton may provide further insights about oral carcinogenesis.

The epidermis is the outermost layer of the body and protects it from environmental toxic exposure and damage. This crucial function is sustained by a continuous process of self-
Differential Gene Expression Profiling of OSCC using Microarray

renewal involving the carefully balanced proliferation and differentiation of progenitor cells constantly replacing the mature cells at the surface of the epidermis. Genetic changes in the signalling pathways controlling keratinocyte proliferation and differentiation disrupt this balance and lead to pathological changes including carcinogenesis (Kern et al., 2010).

Keratins are the intermediate filament proteins which are predominantly expressed in the epithelial cells. Keratins are essential for maintaining structural integrity and there are evidences indicating that they also exert non-mechanical functions (Galarneau et al., 2007). Most of the studies which evaluate the status of keratins in clinical samples of the oral cavity are based on the identification of their presence and localization by immunohistochemistry using monoclonal antibodies. Recent study by Fulzele A, et al. reported aberrant expression of KRT13 to be associated with oral cancer (Fulzele et al., 2012). Another study showed association of KRT4 with tongue squamous cell carcinoma (Ye et al., 2008a). In our study genes related to keratinization such as KRT2, KRT4, KRT10, KRT13, and KRT15 were found to be significantly downregulated. Thus aberrant expression of these genes (especially of KRT4), are associated with OSCC and may work as important biomarkers for OSCC.

Branched-chain amino acids (BCAAs) are a group of essential amino acids comprising valine, leucine, and isoleucine. A low ratio of plasma BCAAs to aromatic amino acids is a physiological hallmark of liver cirrhosis, and BCAA supplementation was originally devised with the intention of normalizing amino acid profiles and nutritional status (Kawaguchi et al., 2011). Genes related to Valine, leucine and isoleucine metabolism (biosynthesis and degradation) were found to be downregulated in our study (ALDH3A1, BCAT2, HIBCH, BCKDHB, ACAT2, ALDH2, ALDH3A2, VARS2, BCAT2). BCAA are not just structural constituents of proteins, but also have "Pharmacologic" properties, known for several years: BCAA are catabolized mainly in muscle; can be oxidized with energy production, being nitrogen donors for other amino acids; regulate protein synthesis and degradation; modulate metabolism of neuroactive mediators. These properties make the clinical use of BCAA particularly suitable in critical conditions such as liver cirrhosis, sepsis, surgical or nonsurgical trauma, acute renal failure, acute pancreatitis and cancer in which energy production from conventional substrates is altered and, at the same time, reduction of protein catabolism and enhancement of synthetic processes is advisable (Chiarla et al., 1997). Increased consumption of BCAAs was reported in various illnesses such as sepsis, diverse injuries and heart diseases and cancer (Szpetnar et al., 2004). Thus their defective metabolism
may play an important role in oral carcinogenesis and may be further investigated for its functional relevance with oral cancer.

Other significantly downregulated pathways/functional terms revealed by \textit{Genowiz}^{TM} analysis include genes involved in tight junctions; 3-Chloroacrylic acid degradation; urea cycle and metabolism of amino groups; glycolysis/gluconeogenesis; and metabolism of tyrosine, butanoate and pyruvate. \textit{DAVID} analysis revealed few more functional terms for which significant gene enrichment clustering was found. These terms of downregulated genes include ‘cellular response to reactive oxygen species; Lipid and aminoacid metabolic processes; oxidation-reduction; and glucosidase activity.

In our study significantly downregulated canonical pathways as identified by IPA analysis are ‘biosynthesis of stilbene, coumarine and lignin; Valine, Leucine and Isoleucine Degradation; Signaling by Rho family GTPases; Tight Junction Signaling; and RhoGDI Signaling. Significantly up-regulated canonical pathways include EIF2 Signaling; Regulation of eIF4 and p70S6K signaling; G-Protein coupled Receptor Signaling; Antigen presentation pathway; and TREM1 signaling.

Our study is the first report on gene expression profiling in OSCC involving cases from a high risk region of northeast India. This study provided gene expression signatures for OSCC and identified several pathways and candidate genes. Some of these pathways were reported earlier in relation to oral carcinogenesis while many others are reported for the first time in our study to be associated with oral carcinogenesis. Identification of such genes and functional pathways which were previously not linked to oral carcinogenesis will provide the foundation for further functional validation of these pathways and specific candidate genes involved in OSCC.
Chapter 7

Validation of Gene Expression Profiling of OSCC by
Quantitative Real time RT-PCR
INTRODUCTION:

Though expression microarrays are powerful and increasingly more widely used investigative, diagnostic, and prognostic molecular biological tools, there are technical aspects to using expression microarrays that can produce results erroneously representing either under- or overexpression of specific genes. For example, false negativity can result from low expression levels, transcript drop-out (attributable to inefficient priming of specific mRNAs), poor adhesion of DNA to the slide, and splice variants with sequences not included on the array. Conversely, sources of false positivity include repetitive nucleotide elements, poly (A) tails, and sequence homology between functionally different transcripts, an inappropriately chosen reference standard, and high background levels due to nonspecific binding of nucleotides to the microarray slides. However, since these sources of error remain a potential source of confounding data, confirmation of expression microarray results before proceeding to undertake more elaborate, gene-specific experiments based on array results is important (True and Feng, 2005).

The real-time reverse transcription PCR (real-time RT-PCR) has dramatically changed the field of measuring gene expression. Real-time PCR is the technique of collecting data throughout the PCR process as it occurs, thus combining amplification and detection into a single step. This is achieved using a variety of different fluorescent chemistries that correlate PCR product concentration to fluorescence intensity (Higuchi et al., 1993). Reactions are characterized by the point in time (or PCR cycle) where the target amplification is first detected. This value is usually referred to as cycle threshold (Ct), the time at which fluorescence intensity is greater than background fluorescence. Consequently, the greater the quantity of template in the starting material, the faster a significant increase in fluorescent signal will appear, yielding a lower Ct (Heid et al., 1996).

There are many benefits of using real-time PCR over other methods to quantify gene expression. It can produce quantitative data with an accurate dynamic range of 7 to 8 log orders of magnitude and does not require post-amplification manipulation (Morrison et al., 1998). Real-time PCR assays are 10,000 to 100,000 fold more sensitive than RNase
Validation of Gene Expression Profiling of OSCC by Real time RT-PCR

... protection assays (Wang and Brown, 1999), 1000 fold more sensitive than dot blot hybridization (Malinen et al., 2003), and can even detect a single copy of a specific transcript (Palmer et al., 2003). In addition, real-time PCR assays can reliably detect gene expression differences as small as 23% between samples (Gentle et al., 2001) and have lower coefficients of variation (cv; SYBR® Green at 14.2%; TaqMan® at 24%) than end point assays such as band densitometry (44.9%) and probe hybridization (45.1%) (Schmittgen et al., 2000). Real-time PCR can also distinguish between messenger RNAs (mRNAs) with almost identical sequences, requires much less RNA template than other methods of gene expression analysis, and can be relatively high-throughput. Because of the reliability of real-time PCR we have chosen it as a supporting technique to validate and better quantitate the expression of most interesting candidate genes from microarray data.
MATERIALS AND METHODS:

Selection of patients and collection of samples:

Tissue samples from patients with oral cancer and matched normal controls were collected as discussed before in earlier chapter on microarray.

Isolation of RNA from tissue samples:

Validation of gene expression by real-time RT-PCR assay was completed in 27 pair of oral cancer samples and matched normal samples.

Validation of microarray results by quantitative real-time RT-PCR analysis.

One microgram of tumor and pooled normal RNA was reverse transcribed into cDNA with random primers (High Capacity cDNA archive kit, Applied Biosystems, Foster City, CA). Real time RT-PCR reactions were performed using an ABI Prism 7000 sequence detection system (Applied Biosystems). Primers and TaqMan probes of eight target genes and an internal control gene TBP were purchased as assays-on-demand from Applied Biosystems (Table 7.1). The thermal cycling conditions included an initial denaturation step at 95°C for 10 min, 40 cycles at 95°C for 15 sec and 60°C for one min. The $2^{\Delta\Delta CT}$ method was used to calculate relative changes in gene expression determined from real-time quantitative PCR experiments. Validation of microarray results was done in 27 OSCC cases and pooled matched controls by real time RT-PCR.
### Table 7.1: Information on the eight genes examined by real time RT PCR

<table>
<thead>
<tr>
<th></th>
<th>Gene</th>
<th>Gene Bank ID</th>
<th>Location</th>
<th>Status</th>
<th>Putative function</th>
<th>Assay ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PDPN</td>
<td>AL359771.27</td>
<td>1p36.21</td>
<td>Up</td>
<td>Translation</td>
<td>Hs00366764_m1</td>
</tr>
<tr>
<td>2</td>
<td>IL8</td>
<td>NM_000584.2</td>
<td>4q13.3</td>
<td>Up</td>
<td>IL8 receptor binding</td>
<td>Hs00174103_m1</td>
</tr>
<tr>
<td>3</td>
<td>COPS5</td>
<td>NM_006837.2</td>
<td>3q22</td>
<td>Up</td>
<td>Transcription co-activator activity</td>
<td>Hs00272789_m1</td>
</tr>
<tr>
<td>4</td>
<td>INHBA</td>
<td>NM_002192.2</td>
<td>7p14.1</td>
<td>Up</td>
<td>Growth and differentiation</td>
<td>Hs0170103_m1</td>
</tr>
<tr>
<td>5</td>
<td>KRT4</td>
<td>NM_002272.1</td>
<td>12q13.13</td>
<td>Down</td>
<td>Cell Junctions</td>
<td>Hs00361611_m1</td>
</tr>
<tr>
<td>6</td>
<td>DOCK8</td>
<td>NM_203447.1</td>
<td>9p24.3</td>
<td>Down</td>
<td>Intracellular signaling Network</td>
<td>Hs00298892_m1</td>
</tr>
<tr>
<td>7</td>
<td>SPRR3</td>
<td>NM_005416.1</td>
<td>1q21.3</td>
<td>Down</td>
<td>Promotes apoptosis</td>
<td>Hs00707014_s1</td>
</tr>
<tr>
<td>8</td>
<td>MAL</td>
<td>BC000458.2</td>
<td>2q11.1</td>
<td>Down</td>
<td>Metastasis suppressor</td>
<td>Hs00271304_m1</td>
</tr>
</tbody>
</table>
RESULTS:
Twenty seven oral squamous cell carcinoma samples and corresponding matched normal controls were included in our study. More than 90% of cases were chewers of betel-quid and tobacco. Demographic details with histopathological grades are given in table 7.2.

Table 7.2: Demographic and clinical characteristics of oral squamous cell carcinoma samples of male patients were used in validation of genes expression by real time PCR

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age</th>
<th>BQ</th>
<th>Tobacco Chewing</th>
<th>Tobacco Smoking</th>
<th>Alcohol</th>
<th>Grade</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 OC 11</td>
<td>32</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>2 OC 12</td>
<td>65</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>3 OC 13</td>
<td>69</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>4 OC 14</td>
<td>60</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>5 OC 15</td>
<td>63</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>6 OC 16</td>
<td>55</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>7 OC 17</td>
<td>55</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>8 OC 18</td>
<td>47</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>9 OC 19</td>
<td>16</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>10 OC 20</td>
<td>55</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>11 OC 21</td>
<td>55</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>12 OC 22</td>
<td>45</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>13 OC 23</td>
<td>55</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>14 OC 24</td>
<td>70</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>15 OC 25</td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>16 OC 26</td>
<td>52</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>+</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>17 OC 27</td>
<td>60</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>18 OC 28</td>
<td>68</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>19 OC 29</td>
<td>52</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>20 OC 30</td>
<td>70</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>21 OC 31</td>
<td>50</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>22 OC 32</td>
<td>28</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>G3</td>
<td>4</td>
</tr>
<tr>
<td>23 OC 33</td>
<td>55</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>24 OC 34</td>
<td>45</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>25 OC 35</td>
<td>60</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
<tr>
<td>26 OC 36</td>
<td>55</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>G1</td>
<td>2</td>
</tr>
<tr>
<td>27 OC 37</td>
<td>50</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>G2</td>
<td>3</td>
</tr>
</tbody>
</table>

++ = Frequently (Regular habit, atleast some amount daily);
+ = Occasionally (Consumes occasionally);
- = Never user
G1 = Well differentiated squamous cell carcinoma;
G2 = Moderately differentiated squamous cell carcinoma;
G3 = Poorly differentiated squamous cell carcinoma
Validation of Gene Expression by real time RT-PCR:

Total eight genes were selected from microarray data for their validation by real time RT-PCR on the basis of their being significantly and differentiated expressed in microarray data. Out of these eight genes 4 were up-regulated (PDPN, IL8, COPS5, and INHBA) and 4 down-regulated (KRT4, DOCK8, SPRR3, and MAL) in microarray study. Validation of gene expression of these eight genes of interest was completed in 27 samples of OSCC. Expression of these genes in most of the cases was similar to that observed in microarray study.

**Fig 7.1:** Real time RT-PCR showed gene expression of PDPN to be up-regulated consistently in all tumor samples as expected from microarray data except in three samples which showed its down-regulation compared to control sample.

**Fig 7.2:** IL8 gene expression by real time RT-PCR found it to be up-regulated consistently in all tumor samples as expected from microarray data except in two samples which showed its down-regulation compared to control sample.
Fig 7.3: Real time RT-PCR showed COPS5 to be up-regulated in most of the OSCC samples compared to normal controls.

Fig 7.4: Gene expression of INHBA was consistently up-regulated in all samples except in two in which it’s expression is downregulated compared to the control.
Fig 7.5: Down-regulation of KRT 4 was consistently observed in almost all samples validated by real time RT-PCR which is similar to the results obtained by microarray observation.

Fig 7.6: Real-time RT-PCR showed down-regulation of MAL to be consistently observed in almost all OSCC samples validated by real time RT-PCR.
Validation of Gene Expression Profiling of OSCC by Real time RT-PCR

**Fig 7.7:** Downregulation of SPRR3 was observed in almost all OSCC samples validated by real time RT-PCR

**Fig 7.8:** Real-time RT-PCR showed down-regulation of DOCK8 in all samples except two
Validation of Gene Expression Profiling of OSCC by Real time RT-PCR

Fig 7.9: Comparison of mean of gene expression level by microarray data with that of real time data for eight genes (4 Down-regulated or supressed genes are DOCK8, SPRR3, MAL, KRT4 and 4 Up-regulated or induced genes are IL8, INHBA, COPS5 and PDPN) selected for the validation of microarray data. Negative values on y axis indicates suppression of genes while induction is shown by positive values of log2 of fold change.

Figure 7.10: Correlation between gene expression of eight genes by real time PCR and cDNA microarray in oral cancer cases compared with control from Northeast region of India. The correlation coefficient is 0.9448 which indicates a strong correlation between the two.
DISCUSSION:

*PDPN* (*Podoplanin*) gene encodes a type-I integral membrane glycoprotein with diverse distribution in human tissues. The physiological function of this protein may be related to its mucin-type character. The specific function of this protein has not been determined but it has been proposed as a marker of lung injury. Podoplanin promotes directional persistence of motility in epithelial cells, a feature that requires CD44, and that both molecules cooperate to promote directional migration in SCC cells (Martin-Villar *et al*., 2011). Data from another study showed that podoplanin expression promoted tumor cell motility in vitro and, unexpectedly, increased tumor lymphangiogenesis and metastasis to regional lymph nodes in vivo, without promoting primary tumor growth (Cueni, *et al*., 2010). The membrane glycoprotein podoplanin is expressed by several types of human cancers and might be associated with their malignant progression. Its exact biological function and molecular targets are unclear, however. Importantly, high cancer cell expression levels of podoplanin were found to be correlated with lymph node metastasis and reduced survival times in a large cohort study of oral squamous cell carcinoma patients (Cueni *et al*., 2010). High podoplanin expression was also found to be associated with esophageal squamous-cell carcinoma (Rahadiani *et al*., 2010). Expressions of podoplanin in Oral lichen planus were significantly associated with malignant transformation risk.

In our study podoplanin was found to be significantly up-regulated and earlier studies have also reported its function to be related with the process of tumorigenesis such as migration of epithelial cell similar suggested podoplanin that further functional study of these genes might reveal the potential role of this as a biomarker for risk assessment of oral malignant transformation in patients with oral cancer. In addition to our study overexpression of PDPN protein was found to be associated with several other studies on oral premalignant/malignant cancers (Kawaguchi *et al*., 2008; Martin-Villar *et al*., 2005). Another study concluded that podoplanin expression correlates with cervical lymph node metastases and clinical outcome (Kreppel *et al*. 2010) and another study suggested it as a novel biomarker for oral squamous cell carcinomas that might be involved in migration/invasion. A study suggested that podoplanin induces an alternative pathway of tumor cell invasion in the absence of epithelial-mesenchymal transition (Wicki *et al*., 2006). Thus over-expression of podoplanin may work as a novel biomarker for oral squamous cell carcinomas that might be involved in migration/invasion.
The protein encoded by **IL8 (Interleukin 8)** gene is a member of the CXC chemokine family. This chemokine is one of the major mediators of the inflammatory response. It functions as a chemoattractant, and is also a potent angiogenic factor. This gene is believed to play a role in the pathogenesis of bronchiolitis, a common respiratory tract disease caused by viral infection. This gene and other ten members of the CXC chemokine gene family form a chemokine gene cluster in a region mapped to chromosome 4q. Benzo(a)pyrene induces oxidative stress-mediated interleukin-8 production in human keratinocytes via the aryl hydrocarbon receptor signaling pathway (Tsuji et al., 2009). Overexpression of IL-8 promotes tumor growth, metastasis, chemoresistance and angiogenesis (Ning et al., 2010). High-expression of IL-8 was reported to be significantly associated with risk of other cancers such as nasopharyngeal carcinoma (Liao et al., 2010), non-small cell lung cancer (Liu et al., 2010), and bladder cancer (Mahmoud et al., 2009). Thus findings from our study suggested that over-expression of IL8 may work as biomarker for oral carcinogenesis.

**COPS5** also known as JAB1; SGN5; or MOV-34. The protein encoded by this gene is one of the eight subunits of COP9 signalosome, a highly conserved protein complex that functions as an important regulator in multiple signaling pathways. This protein is reported to be involved in the degradation of cyclin-dependent kinase inhibitor CDKN1B/p27Kip1. It is also known to be an coactivator that increases the specificity of JUN/AP1 transcription factors. It was reported that COPS5 promotes cell growth by decreasing p27 level (Malicet et al., 2006). Overexpression of COPS5 was reported to contribute to tumor progression in various cancers such as pancreatic cancer (Fukumoto et al., 2006), laryngeal squamous cell carcinomas (Dong et al., 2005) and hepatocellular carcinoma (Patil et al., 2005). There are few reports which showed over expression of COPS5 to be associated with oral squamous cell carcinoma (Harada et al., 2006; Shintani et al., 2003). Thus our study suggests that overexpression of COPS5 may work as prognostic marker for oral carcinogenesis.

The **INHBA (Inhibin Beta A)** subunit joins the alpha subunit to form a pituitary FSH secretion inhibitor. Inhibin has been shown to negatively regulate gonadal stromal cell proliferation and have tumor-suppressor activity. It is proposed that inhibin may be both a growth/differentiation factor and a hormone. Furthermore, the beta A subunit forms a homodimer, activin A, and also joins with a beta B subunit to form a heterodimer, activin AB, both of which stimulate FSH secretion. It was found that activin A play role in human embryonic stem cell proliferation and differentiation (Tsai et al., 2009).
Upregulated INHBA expression may promote cell proliferation and was found to be associated with various cancers such as lung adenocarcinoma (Seder et al., 2009b), tongue cancer (Ye et al., 2008a), esophageal adenocarcinoma (Seder et al., 2009a; Yoshinaga et al., 2008). In a recently published report, Increased INHBA expression was found to be significantly correlated with the diameter of cancer and depth of tumor invasion in patients with gastric cancer. And in that study it was reported that patients with higher expression levels of INHBA had a shorter disease-free survival rate (Wang et al., 2012). In a gene expression profiling study involving head and neck squamous cell carcinoma cell lines, higher expression of INHBA was found to be associated with HNSCC (Shimizu et al., 2007). Thus our findings suggested that over-expression of INHBA in oral squamous cell carcinomas may contribute to tumor progression and it may work as a significant prognostic marker for OSCC.

KRT4 is a member of the keratin gene family. The type II cytokeratins consist of basic or neutral proteins which are arranged in pairs of heterotypic keratin chains coexpressed during differentiation of simple and stratified epithelial tissues. This type II cytokeratin is specifically expressed in differentiated layers of the mucosal and esophageal epithelia. Mutations in this gene have been associated with White Sponge Nevus, characterized by oral, esophageal, and anal leukoplakia (Chao et al., 2003; Zhang et al., 2009).

Epithelial tissues function to protect the organism from physical, chemical, and microbial damage and are essential for survival. To perform this role, epithelial keratinocytes undergo a well-defined differentiation program that results in the expression of structural proteins which maintain the integrity of epithelial tissues and function as a protective barrier. Keratin proteins comprise the predominant cytoskeletal component of these epithelia. Keratin filaments are attached to the plasmamembrane via desmosomes, and together these structural components form a three-dimensional array within the cytoplasm of epithelial cells and tissues (Presland and Dale, 2000). Various studies showed down-regulation of KRT4 to be associated with esophageal cancer (Huang et al., 2007), tongue squamous cell carcinoma (Ye et al., 2008a), and with an ovarian cancer cell line (Sun et al., 2009). In our study, KRT4 was found to be significantly down-regulated in both microarrays as well as in real time RT-PCR study which suggested down-regulation of KRT4 gene may be used as a prognostic and therapeutic marker for OSCC.
The protein encoded by MAL gene is a highly hydrophobic integral membrane protein belonging to the MAL family of proteolipids. The protein has been localized to the endoplasmic reticulum of T-cells and is a candidate linker protein during T-cell signal transduction. The protein plays a role in the formation, stabilization and maintenance of glycosphingolipid-enriched membrane microdomains.

Down-regulation of MAL gene was reported to be associated with various cancers such as esophageal squamous cell carcinoma (Kazemi-Noureini et al., 2004; Mimori et al., 2007) laryngeal carcinoma (Jiang et al., 2009), tongue squamous cell carcinoma (Ye et al., 2008a) head and neck cancer (Beder et al., 2009). A study suggested that the epigenetic inactivation of MAL can contribute to human epithelial cell carcinoma and may be served as a biomarker in HNSCC (Beder et al., 2009; Cao et al., 2010). Thus our findings and the data from various other studies strongly suggest the MAL gene as a metastasis-suppressor candidate for OSCC.

Small proline rich repeat protein 3 (SPRR3), a member of the SPRR family of cornified envelope precursor proteins, which is expressed during epithelial cell differentiation (Zhang et al., 2008). It’s functional details are not very well known and it has not been reported earlier in association with OSCC however downregulation of SPRR3 was found to be associated with malignant transformation of the healthy esophageal mucosa into esophageal squamous cell carcinoma (Berhane et al., 1994; Chen et al., 2000). Thus to the best of my knowledge our study is the first report showing association of aberrant gene expression of SPRR3 to be associated with oral carcinogenesis. Thus it may work as a useful biomarker for oral carcinogenesis.

DOCK8 (Dedicator of cytokinesis 8) gene encodes a member of the DOCK180 family of guanine nucleotide exchange factors. Guanine nucleotide exchange factors interact with Rho GTPases and are components of intracellular signaling networks. DOCK8 was reported to be involved in processes that affect the organisation of filamentous actin (Ruusala and Aspenstrom, 2004). A study on squamous cell carcinoma of the lung showed frequent silencing of chromosome 9p, which involved homozygous DOCK8 deletion at 9p24.3 (Kang et al., 2010). In another study under-expression of DOCK8 was found to be associated with hepatocellular carcinoma (Saelee et al., 2009). Our study is the first report showing association of downregulation of DOCK8 with oral carcinogenesis. Thus DOCK8 might be
of interest for further study of the pathophysiology of OSCC as potential targets for therapeutic measures.

Genes selected for validation by real time RT-PCR have been validated successfully in all samples of OSCC. Genes such as $DOCK8$ and $SPRR3$ were reported for the first to be associated with oral carcinogenesis and other genes have very limited reports of their association with oral carcinogenesis. These may be useful biomarkers for chewing of tobacco and betelquid associated oral cancers. This is the first report on gene expression profiling of oral cancer from this high risk region of India. Thus further study in relation to detailed functional aspects of these genes will enhance our understanding about oral carcinogenesis.
Chapter 8

Conclusions and Future Scope of Work
Objective 1: Role of SNPs in Xenobiotic Metabolizing Genes (Encoding Phase I and Phase II Detoxification Enzymes) and p53 Codon 72 Polymorphism in Oral Cancer

The quantitative absorption, distribution, metabolism, and excretion of carcinogenic tobacco constituents depend on the activity and efficiency of metabolic and enzymatic detoxification pathways. The role of polymorphisms of genes responsible for detoxification of xenobiotics and p53 codon 72 were investigated in 235 oral cancer cases and 289 healthy controls from high incidence region of oral cancer in Northeast India. Eight polymorphisms in seven genes [CYP1A1 (Msp1 and Nco1), GSTT1, GSTM1, GSTP1, NAT2, NQO1 and Codon 72 of P53] were analyzed using PCR-RFLP and correlated with risk factors of oral cancer. Tobacco chewing as well as smoking were found to impart a significant risk for oral cancer (AOR=2.78, 95% CI=1.71-4.51, and AOR=1.57, 95% CI=1.05-2.35 respectively) with tobacco chewers at higher risk compared to smokers. Frequency of betel quid chewers and alcohol consumers were also higher in cases (80% and 37% respectively) compared with controls (66% and 30% respectively) but when adjusted with other factors the difference was statistically insignificant. GSTT1 and GSTM1 null genotypes and the variant genotypes of CYP1A1*2A, CYP1A1*2C, and p53 codon72 were not found to be associated with oral cancer risk. Homozygous variant genotypes of NAT2 (AA) were significantly higher in cases (14%) compared to controls (9%) [OR=1.83, 95% CI=1.01-3.31, P=0.04]. However when adjusted with other risk factors this risk became statistically insignificant. Frequency distribution of NQO1 genotypes Pro/Pro, Pro/Ser and Ser/Ser was 45%, 35% and 20% in cases and 60%, 27% and 30% in controls respectively. This difference was statistically significantly different for heterozygous (AOR=1.64, 95% CI=1.04-2.58) as well as for homozygous (AOR=1.81, 95% CI=0.98-3.32) variant genotypes. When data was analyzed in different geographic regions of NE India, the GSTT1 null genotype and homozygous variant genotypes of GSTP1 were found to impart significant risk for oral cancer (AOR= 2.58, 95% CI 1.01–6.61, p= 0.05, and AOR=3.14, 95% CI 0.94-10.49, p=0.06 respectively) in samples obtained from the Assam region. Further, variant genotypes of CYP1A1*2A were found to impart significant risk for heterozygous (AOR = 4.55, 95% CI 0.88-23.36, p=0.07) as well as for homozygous genotypes (AOR= 6.38, 95% CI 1.10-40.83, p=0.05) for oral cancer in Sikkim population of NE India. Gene-environment interaction analysis revealed that the
variant genotypes of *NQO1* did not interact statistically with tobacco consumption habits. It is possible that although variant genotypes of *NQO1* may play an important role in the genetic susceptibility to oral cancer, its pathway is unrelated to the detoxification mechanism of tobacco constituents.

**Objective 2: Detection of Genetic Instability/ Genomic Alterations using Targeted Next Generation Sequencing in Oral Squamous Cell Carcinoma**

Several types of genetic alterations such as base changes, insertions, deletions, chromosomal translocations, inversions, loss of heterozygosity and copy number variations may be involved in oral carcinogenesis. In this study targeted re-sequencing of 169 functionally relevant and potentially important genes showed 96 SNPs (50 novel and 46 known SNPs) and 46 InDels (29 novel and 17 known InDels). Most of the known SNPs with high read depths (>50) were located in tumor suppressor genes such as *RB1, FHIT, FAT1, FAT2* and *VHL*. SNPs detected in *RB1, FHIT* and *FAT1* were located in the intronic regions of the gene while those in *ATM, VHL, IL12B*, and *MET* were located in 3’UTR.

Novel SNPs with significant read depths (>25), and with non-synonymous type variations which may lead to structural and functional changes in resulting protein were observed in *FAT1, FAT2, TP53, NOTCH2, CDH3, ATM* and *MET*. Synonymous type variations were observed in *APC* and *IL12B* genes and those present in non-coding regions were observed in or near to *EGFR, STAT5B, CDK5* and *MYCL1* genes.

The known deletions observed in our study were present in non-coding regions of the gene, 4 of them were present in 3’UTR (*TSC1, FAT1, MAP2K6*, and *ERBB4*), two at 5’UTR (*BMP4*, and *SLC22A18*) and one in intronic region of *BRCA1*. UTR regions of a gene have significant role in regulation of gene activity, thus its alteration may have an adverse effect on gene activity. The known deletion with highest read depth (272) was present in UTR region of *TSC1* gene (rs34947162; rs115091888). *TSC1* plays a central role in regulating cell survival and proliferation signaling pathways. A total of 11 novel deletions have been observed in oral cancer cases in our study. Novel deletion with the highest read depth (58) was present in *MSH6* gene (NM_000179 at position 48033455). Other important genes with novel deletions include *IGF1R, BRCA2, TSC2* and *PAK3*.

Of the known insertions observed in our study, 4 were present in UTR regulatory regions of *APC, SMAD2, RHOB* and *NBL1* genes while the remaining 6 were located in intronic regions
Conclusions and Future Scope of Work of ADH6, PDGFRα, BRIP1, FAT2, DLG2 and KLK8. The insertion with highest read depth (102) was that of base A at position 112180228 in APC gene (rs11432316; rs79379053). This gene encodes a tumor suppressor protein that acts as an antagonist of the Wnt signaling pathway.

Insertions and deletions (InDels) of bases are among highly damaging mutations. The affected genes in our study may be responsible for oral carcinogenesis as they may result in abrupt changes in expression of genes or in the structural or functional activities of the encoded protein.

Objective 3: Differential Gene Expression Profiling of Oral Squamous Cell Carcinoma and Adjacent Noncancerous Tissues with cDNA Microarray in North East Indian Population

Alteration of gene expression was done by cDNA microarray to identify differentially expressed genes in oral cancer. Six hundred and thirty four differentially expressed (247 upregulated and 387 down-regulated) genes were identified. Most significantly up-regulated pathways were related to ribosomal activity (RPL38, RPSA, RPL6, RPS3, RPS20, RPS6, RPS7), Neuroactive ligand-receptor interaction (GRM8, GRM4, NTSR1, P2RX7), ECM-receptor interaction (LAMC2, ITGB1, GP6 FN1, COL2A1) and Aminoacyl-tRNA biosynthesis (LARS2, AARS, WARS). Functional annotation clustering of up-regulated genes using web-based DAVID analysis also showed most significantly enriched cluster to be associated with ribosomal activity, ribosome biogenesis, and translation (involving genes such as COPS5, IDO1, KYNU, RPS7). Second cluster of up-regulated genes was related to inflammatory response, defense response and response to wounding (involving genes such as IL8, INHBA, and PDPN).

Most significantly downregulated pathways (when analyzed by Genowiz) were related to cell junctions (involving 12 genes KRT34, DSG1, KRT2, KRT15, KRT36, VIM, KRT4, KRT10, KRT78, KRT13, ACTB, KRT33A), and Valine, leucine and isoleucine degradation (ALDH3A1, BCAT2, HIBCH, BCKDHB, ACAT2, ALDH2, ALDH3A2). DAVID analysis revealed clusters with highest enrichment score of down-regulated genes to be associated with structural constituents of cytoskeleton (involving genes such as CRYAB, KRT4, TPM1, KRTAP5-9), epidermal cell differentiation and Keratinisation (involving genes such as ALDH3A2, JUN, KRT10, KRT13, KRT15, KRT2, KRT34, TP63, KRT4, KRTAP5-9,SPRR3).
Objective 4: Validation of Gene Expression Profiling of OSCC by Quantitative Real-time RT-PCR

Eight significantly deregulated genes [four up-regulated (PDPN, IL8, COPS5, INHBA) and four down regulated (KRT4, DOCK8, SPRR3, MAL)] were selected from microarray data and validated by real-time RT PCR. Our study revealed several genes such as DOCK8 and SPRR3 which were reported for the first time to be associated with oral carcinogenesis. Recent reports have also suggested significance of KRT4 and MAL to work as possible prognostic and therapeutic markers for oral carcinogenesis and DOCK8 and SPRR3 may be further investigated for their association with oral carcinogenesis.

To summarize, this is the first study to provide gene expression profiling and genetic variations of oral cancer associated with betel quid and tobacco consumption habits in a high-risk region of Northeast India. Polymorphic study revealed that although variant genotypes of NQO1 may play an important role in the genetic susceptibility to oral cancer, its pathway is unrelated to the detoxification mechanism of tobacco constituents. On gene expression profiling analysis, NQO1 was found to be significantly up-regulated in our study. Thus, it appears that it may be the variant form of NQO1 and not its expression level, which may be responsible for the higher risk of oral cancer in this region. Genes such as FAT1, TSC, GAS7 and APC showed high level of genomic instability (including known and novel variations) in OSCC. Furthermore GAS7 which is functionally involved in chemokine-induced migration in the immune system was also found to be significantly downregulated in our gene expression profiling study. Thus these genes may be useful prognostic and therapeutic targets in OSCC.

Future Scope of Work:

The discovery of genetic variations involved in carcinogenesis is a crucial step in cancer research which is tremendously facilitated by microarray and next generation sequencing technology, useful for both targeted and genome-wide screening. This is the first study on oral cancer from high-risk region of Northeast India which provided genetic variations and gene expression profiling of oral cancer associated with betel quid and tobacco consumption habits.

Our study revealed a large number of mutational changes including novel SNPs, known SNPs, known insertions, novel insertions, known deletions and novel deletions. These genetic variations provide a rich source of information which may be further investigated for
their role to work as possible diagnostic, prognostic and therapeutic markers for oral carcinogenesis. Furthermore several differentially expressed genes such as \textit{DOCK8} and \textit{SPRR3} were reported for the first time to be associated with oral carcinogenesis. Gene expression profiling may help to unlock the molecular basis of phenotype, response to treatment and heterogeneity of disease.

Thus findings on genetic instability and gene expression profiling of oral cancer from this study will provide foundations for future research and may help in the development of molecular signatures for accurate prediction of clinical outcome and improved therapeutic strategies for patients with oral cancer.


epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol Biomarkers Prev* 9, 29-42.


References


References


References


References


expression may promote cell proliferation and is associated with poor survival in lung adenocarcinoma. *Neoplasia* 11, 388-396.


References


List of Publications

Research Articles in International Journals:


PMID: [22731640](https://www.ncbi.nlm.nih.gov/pubmed/22731640)


PMID: [22930568](https://www.ncbi.nlm.nih.gov/pubmed/22930568)

8. **Yadav DS**, Mishra AK, Sharma JD, Katakai AC, Saxena S, Kapur S. “Genetic Polymorphisms of CYP1A1, NQO1 and NAT2 genes and risk of oral cancer in Northeast India”. (Manuscript communicated to ‘Cancer Investigation’).


Abstracts in International/National Conferences:


4. **Yadav DS**, Devi TR, Mishra AK, Saxena S, Kapur S. “Genetic polymorphisms of CYP1A1, NQO1 and NAT2 and risk of oral cancer in Northeast India: Tobacco consumption as risk modulator” in proceedings of HUGO’s 15th Human Genome Meeting on “Genomics of Human Diversity and Heritable Disorders” Held in Dubai from 14th-17th March 2011.


6. Agrawal U, **Yadav DS** Sharma M., Saxena S. “Role of Vitamin D receptor polymorphism and retinoid expression in breast carcinoma” in 6th Asia Pacific IAP Congress, held at Le Meridian Resorts & Convention Centre, Kochi, Kerala, India on 20th-23rd August 2009.


Biography of Candidate

(Research Profile & Co-curricular Activities)

Academic Qualifications:

M. Sc. (Biotechnology) – Indian Institute of Technology (IIT), Roorkee – 2001 -2003
CSIR-UGC NET – Qualified for Junior Research Fellowship (JRF), in June 2004
CSIR-UGC NET – Qualified for Lecturership (LS), in December 2003
GATE-2004 (Life Sc.) – Qualified with All India Rank 135 (Perc. Score 97.22) in Feb 2004

Research and Teaching Experiences:

1. 1st Jan 2003- 31st May 2003
   ‘Binding studies of daunomycin to calf-thymus DNA by fluorescence and absorption spectroscopy’: Dissertation work, the compulsory part of M.Sc. final semester under Prof. Ritu Barthwal, Deptt of Biotechnology IIT, Roorkee.

2. 25th August 2004 to 31st August 2005
   Worked as Lecturer in the Department of Biotechnology in Ch. Charan Singh Post Graduate College, Heonra, Etawah.

3. 1st September 2005 to 31st August 2007
   Worked as UGF-JRF in National Institute of Pathology (ICMR) under the supervision of Dr. Sujala Kapur, (Scientist-E). Research work leading to Ph.D. entitled ‘Genetic variation and gene expression profiling of oral cancer’.

4. 1st September 2007 to 31st August 2010
   Worked as UGC-SRF in National Institute of Pathology (ICMR) under the supervision of Dr. Sujala Kapur (Scientist-E).

5. 1st September 2010 to 6th April 2012
   ICMR-SRF in National Institute of Pathology (ICMR).

6. 10th April 2012 to till date
   Junior Scientific Officer (through UPSC recruitment) in Central Forensic Institute (CFI/CFSL), Directorate of Forensic Science Services, Ministry of Home Affairs.
**Trainings Received:**


2. Hands on training on Florence In-situ Hybridization (FISH) in “National Workshop on Molecular Cytogenetics: Cancer Cytogenetics (Haematological malignancy) by FISH” conducted by Department of Reproductive Biology, All India Institute of Medical Sciences (AIIMS), New Delhi from November 29 to December 1, 2010.

3. Two months summer training under the supervision of Dr. Rajkumar at Institute of Nuclear Medicine and Allied Sciences (INMAS), Defense Research and Development Organization (DRDO), Delhi. Work was on gamma rays induced stress proteins. During the tenure, I learnt techniques like protein isolation, PAGE and blotting etc June –July 2002.

**Conferences/Workshops attended:**

1. Attended “American Association of Cancer Research (AACR), International Conference on New Horizons in Cancer Research: Biology to prevention to therapy” December 13-16, 2011, Gurgaon, Delhi

2. Participated in “National Workshop on Molecular Cytogenetics: Cancer Cytogenetics (Haematological malignancy) by FISH” conducted by Department of Reproductive Biology, All India Institute of Medical Sciences (AIIMS), New Delhi from November 29 to December 1, 2010.


5. Attended 13th Human genome meeting (HGM 2008) on “Genomics and the Future of Medicine” organized by CSIR and HUGO international meetings held in Hyderabad on September 27-30 2008.

7. Attended “International symposium on cancer biology” which was held in National Institute of Immunology (NII), New Delhi on November 14-16, 2007.

8. Attended “National workshop on microarray technology” which was held in Institute of pathology (ICMR), New Delhi on April 16-18 2007.

9. Attended “21st annual conference of Indian Association of Pathologists and Microbiologists”, Delhi Chapter organized by Institute of Pathology (ICMR) and Safdarjung Hospital and V.M.M. College held on 16 April 2006.


M.Sc. Dissertation Co-Supervised:
During my Ph.D., I co-supervised 5 postgraduate students for their M.Sc. dissertation projects, which is compulsory part of their M.Sc. curriculum.


2. Ratnam Prasad – M.Sc. ((Biotechnology) student from Jiwaji University Gwalior. She worked with me from 1st Feb 2009 to 10th July 2009


Extracurricular Activities:
1. I was active in sports since my school days and was winner in various events of athletics.
2. I won 2 Gold, 2 Silver and 3 Bronze medals in sports meets of IIT Roorkee during my postgraduation.
3. Was selected as Sports Secretary of IIT Roorkee, and honored by the sports association of IIT Roorkee.
4. Was the **member of Himalayan Explorer’s Club** (HEC), an adventurous sports club of IIT Roorkee, and was the member of various teams like Rock climbing training tour, trekking etc.

5. Represented IIT Roorkee in inter IIT Sports meet in various events of athletics.

Place: 
Signature:

Date: 10-09-2012 
Name: **DHIRENDRA SINGH YADAV**
**Brief Biography of Supervisor**

Name: Dr. Sujala Kapur, M.D.,
Designation: Scientist E/Deputy Director
Address: Institute of Pathology (ICMR)
Safdarjung Hospital Campus,
New Delhi – 110029.
Telephone: 2619 8402-06
Fax: 2619 8401
Email: sujalakapur@gmail.com

**CURRENT RESEARCH INTERESTS**

- **Gene expression, copy number variations and high throughput sequencing using microarrays, Affymetrix and Illumina based platforms and validation by quantitative real time RT-PCR**
- Molecular biomarkers in hematopoietic lymphoid malignancies.
- Flow cytometric assays to analyse chemotherapeutic sensitivity in leukaemic cells

**Education (Post-graduation onwards & Professional Career)**

<table>
<thead>
<tr>
<th>Institution</th>
<th>Place</th>
<th>Degree</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lady Hardinge Medical College</td>
<td>New Delhi</td>
<td>MBBS</td>
<td>1973</td>
</tr>
<tr>
<td>S.N. Medical College</td>
<td>Agra</td>
<td>MD (Path)</td>
<td>1985</td>
</tr>
<tr>
<td>Institute of, Pathology ICMR</td>
<td>New Delhi</td>
<td>Research Officer</td>
<td>1988-1993</td>
</tr>
<tr>
<td>Institute of, Pathology ICMR</td>
<td>Senior Research Officer</td>
<td>1993</td>
<td></td>
</tr>
<tr>
<td>Institute of, Pathology ICMR</td>
<td>Senior Research Officer</td>
<td>1995-1999</td>
<td></td>
</tr>
<tr>
<td>Institute of, Pathology ICMR</td>
<td>Assistant Director</td>
<td>1999-2005</td>
<td></td>
</tr>
<tr>
<td>Institute of, Pathology ICMR</td>
<td>Deputy Director &amp; Scientist E</td>
<td>2005 to date</td>
<td></td>
</tr>
</tbody>
</table>


Extramural projects

- Immunogenetic profile of nasopharyngeal cancer in a high-prevalence region of Northeast India (Collaborative, Multicentric, DBT Project) (1).
- Genome-wide analysis of genetic alterations in patients with esophageal cancer from Northeast India using single nucleotide polymorphism arrays (Collaborative, Multicentric, ICMR Task Force Project) (3).
- Tobacco Use and Pesticide Exposure in causation of carcinoma oesophagus and other cancers at North East India (Collaborative, Multicentric, ICMR Task Force Projects) (3) (completed).
- Flow cytometric assays to evaluate prognosis and patient’s response to chemotherapy in acute leukemia (Extramural ICMR Project) (1) (completed).

Fellowships/Award

- Awarded Scientific Fellowship, Kiel University, Germany, 1995 & 1998
- Awarded INSA Visiting Fellowship, 2004
- Awarded In-country WHO Fellowship, 2004

Training Programmes

- Training in Molecular Pathology, Kiel, Germany
- Training in Microarray Techniques at Department of Biochemistry, Indian Institute of Science, Bangalore and Ann Arbor, Michigan.
- Training in Laboratory Genetics at Department of Anatomy, All India Institute of Medical sciences, New Delhi.

Membership of Professional Societies

- Member, IAPM
- Member, IAPM, Delhi Chapter
- Member, Indian association of Cancer research
- Member, HUGO

Thesis Guide for DNB/PhD/MSc Students

DNB

- Cytochemical, immunocytochemical, flow cytometric analysis and endothelial-mesenchymal interaction in acute leukemia, acute and chronic myeloid disorders and lymphoproliferative disorders (3).
- Histopathological analysis, cytokeratin expression, angiogenic factors and Helicobacter pylori associated changes in upper gastrointestinal lesions (3).

Ph.D. Thesis Completed

- Differential gene expression pattern in oesophageal carcinoma. 2009

Ongoing Ph.D. Thesis

- Gene expression profile and genetic variation in oral cancer associated with tobacco consumption.
- Molecular biomarkers in hematopoietic lymphoid malignancies and their association with pesticides.

M.Sc.

- p53 status in acute leukemia
- Expression pattern of drug-resistance related proteins in acute leukemic cells
- Flow cytometric toxicity assay in patients with acute leukemia
<table>
<thead>
<tr>
<th>Year</th>
<th>S.No.</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>11</td>
<td>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. Chattopadhyay I, Singh A, Phukan RK,</td>
</tr>
<tr>
<td>Year</td>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Year</th>
<th>Number</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
</table>

**Chapters in Books**
