Molecular Profile of Esophageal Cancer Patients in North East Region of India

THESIS

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

By

SRI INDRANIL CHATTOPADHYAY

Under the Supervision of
Dr. Sujala Kapur

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
PILANI (RAJASTHAN) INDIA
2009

Format for Certificate from the candidate working without the benefit of a supervisor

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

CERTIFICATE

This thesis is submitted under Regulation 8.20 (a) of the Academic Regulations for Doctoral Programmes which allows a faculty member of the Institute/Professional to do Ph.D. research without the benefit of a supervisor.

This is to certify that the thesis entitled **Molecular Profile of Esophageal Cancer Patients in North East Region of India** which is submitted for award of Ph.D. Degree of the Institute embodies my original work.

Signature in full: _____

Name in block letters: SRI INDRANIL CHATTOPADHYAY

IDNo: **2006PHXF022P**

Designation: Senior Research Fellow

Date:

Format for certificate from Thesis Supervisor to be incorporated in the Thesis

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

CERTIFICATE

This is to certify that the thesis entitled **Molecular Profile of Esophageal Cancer Patients in North East Region of India** and submitted by SRI INDRANIL CHATTOPADHYAY ID No 2006PHXF022P for award of Ph. D. Degree of the Institute embodies original work done by him under my supervision.

Signature in full of the Supervisor:
Name in capital block letters: Dr. SUJALA KAPUR
Designation: Scientist E

Date:

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and regards to Supervisor Dr. Sujala Kapur, Scientist E & Deputy Director, Tumor Biology group, Institute of Pathology (Indian Council of Medical Research), Safdarjangh Hospital Campus, New Delhi, for her guidance, help, support and encouragement throughout my Ph.D. programme.

I would also like to express my sincere gratitude and regards to Dr. Sunita Saxena, Director, Institute of Pathology (Indian Council of Medical Research), Safdarjangh Hospital Campus, New Delhi, for her guidance, help, support and encouragement throughout my Ph.D. programme.

I am immensely thankful to Vice-chancellor BITS, Pilani for providing me this opportunity to pursue the off-campus Ph.D of the Institute. I express my gratitude to Prof. Ravi Prakash, Dean, Research and Consultancy Division (RCD), BITS, Pilani for his constant official support, encouragement, and making the organization of my research work through the past few years easy.

I thank Dr. Hemanth Jadav, Dinesh Kumar, Ms.Monica Sharma, and Sharad Shrivastava, nucleus member of RCD and Dr. Ashis Das and Dr. S.K.Verma of Biological Science Division of BITS, without whose cooperation and guidance it would not have been possible for me to pursue the goal oriented research during each of the past few semesters.

I would like to thank Dr. Joydeep Purkayastha, Senior Surgeon, Dr. Taposhi Roy, Senior Research Fellow and Mr. Nipon Das, Project Assistant, BBCI, Guwahati, Assam for their help in esophageal tumor samples collection.

I am grateful to Dr. Jaganath Sharma, Senior Pathologist, BBCI, and Dr. Amal Kataki, Director, BBCI, Guwahati, Assam for histopathological analysis.

I wish to express my warm and sincere thanks to Dr. Avninder Pal Singh, Scientist C, Tumor Biology group, Institute of Pathology (Indian Council of Medical Research), Safdarjangh Hospital Campus, New Delhi for his help in tissue microarray preparation.

I am thankful to Dr. Raju, Scientist B, and Dr. L.C. Singh, Technical Officer, Tumor Biology group, Institute of Pathology (Indian Council of Medical Research), Safdarjangh Hospital Campus, New Delhi and Dr. R.K. Phukan, Scientist C, RMRC, Dibrugarh, Assam for their help and support.

My thanks giving would be incomplete without mentioning my past and present lab mates Ms. Mishi Kaushal, Ms.Rakshan Ishan, Ms. Regina Thodam, Ms. Shantilata, Mr. Pradeep Singh Chauhan, Mrs. Anurupa Chakraborty, and Mr. Dhirendra Singh Yadav for their encouraging support and cooperation.

No words will be enough to express my feelings to all those suffering patients who willingly donated their valuable blood and tissue for the study.

Senior Research Fellowship and financial assistance for performing experiments provided by the prestigious organization Indian Council of Medical Research, New Delhi is gratefully acknowledged.

I strongly believe that the blessings and the approach given by all especially Supreme Lord, my revered mentor Dr. Sujala Kapur and Director of Institute of Pathology Dr. Sunita Saxena shall help me in my future endeavor and professional excellence.

Sri Indranil Chattopadhyay

Abstract

The development of esophageal cancer is a leading example in which environmental carcinogens in addition to geographic and genetic factors appear to play major etiologic roles. Esophageal cancer occurs at very high frequencies in certain parts of China, Iran, South Africa, Uruguay, France, Italy and in some regions of India. The highest incidence of this cancer in India has been reported from Assam (AAR of 33/100,000) in the North-east region (NE) where it is the second leading cancer in men and third leading cancer in women. There is wide spread use of tobacco and betel quid with fermented areca nuts in this geographic region. Further familial aggregation of esophageal cancer has also been reported from this region of India. However, molecular profiling of esophageal cancer in this region has so far not been investigated.

In the current study, GeneChip® Human Mapping 10K Array Xba 142 2.0 was applied to determine common aberrations (deletion or amplification) and copy number alterations in 20 pairs of matched germ-line and tumor DNA obtained from 20 patients with esophageal squamous cell carcinoma from a high-risk region of India where tobacco use and alcohol consumption are widespread and the users of these two substances are also betel quid chewers. Frequent amplifications were found on chromosomes arms 1p36.13, 1q21.1, 2p14, 3q28, 3q27, 3q26.1, 5p15.2, 5q11.2, 6p25.3, 7q11.21, 9q31.3, and17p13.1 and frequent deletions were found on chromosome arms 3p, 4p, 5q, 8p, 9p, 11q, 13q, 17p, and 18q in ESCC of high-risk Indian population. The genes located at amplified regions of chromosomes were mostly involved in MAPK signaling pathway and inflammation. The genes located at deleted regions of chromosomes were mostly involved in Wnt signaling pathway, Focal/cell adhesion, and metabolism of xenobiotics

by cytochrome P450 pathways which were reported to be associated with tobacco associated carcinogenesis. These findings suggest that the gains and losses of chromosomal regions may contain ESCC-related oncogenes and tumor suppressor genes and provide important theoretic information for identifying and cloning novel ESCC-related oncogenes and tumor suppressor genes.

cDNA microarray gene expression analysis was done to identify the genes differentially expressed in esophageal cancer associated with prevalent risk factors such as tobacco use and betel quid chewing in a high-risk Indian population. 127 differentially expressed genes (87 up-regulated and 40 down regulated) were identified in tumor tissue. Genes involved in dimethylallyltransferase activity and farnesyltransferase activity (*CTLA4*), cation antiporter activity (*SLC9A2*) and cation transporter activity (*KCNN2*, *SLC30A4, KCNJ15, CACNA2D3*), G-protein coupled receptor (GPCR) activity (*GPR87, NPY*), MAPK signaling pathway (*FGF12*), and protein serine or threonine kinase activity (*GRK4*) were significantly up regulated. Genes involved in anti-apoptosis activity (*BIRC1*) and cellular proliferation (*EGR2*) were also significantly activated. Genes involved in structural constituent of the ribosome (*RPL32*, *RPS4X),* structural constituent of cytoskeleton (*KRT17*, *KRT4*, *PLA2G1B*), cysteine protease inhibitor activity (*CSTB*, *CSTA*), anti-oxidant activity (*PRDX6*), acyl groups transferase activity (*TGM3*), and translation elongation (*EEF1A1*) were significantly down-regulated. Genes involved in humoral immune response (*CD24*) and base-excision repair (*MPG*) were also significantly down regulated.

Since there is wide spread use of tobacco and fermented betel quid among patients with ESCC in high risk area of India, exposure to tobacco and betel quid

constituents may contribute to the development and progression of ESCC in this area by facilitating the deregulation of genes involved in these pathways. Several genes that showed alteration in our study have also been reported from a high incidence area of esophageal cancer in China. This indicates that molecular profiles of esophageal cancer in these two different geographic locations are highly consistent.

In the current study of 317 cases of esophageal cancer, 92 (29%) cases had a family history of esophageal and/or other cancers. To identify the genes and molecular functional pathways involved in esophageal cancer, we analyzed gene expression profile of esophageal tumor tissue from patients having family history of esophageal cancer by cDNA microarray. Three hundred and fifty differentially expressed genes (26 up-regulated and 324 down-regulated) were identified in familial ESCC cases. Genes involved in humoral immune response (*PF4*), extracellular matrix organization (*COL4A4*), metabolism of xenobiotics *(EPHX1*), TGF-β signaling (*SMAD1*) and calcium signaling pathways (*VDAC1)* were down-regulated, and genes involved in regulation of actin cytoskeleton (*WASL*), neuroactive ligand receptor interaction (*GRM3*), Toll-like receptor (*CD14*), B-cell receptor (*IFITM1*) and insulin signaling pathways (*FOXO1A)* were up- regulated in familial ESCC cases. Validation of differential expression of subset of genes by QRT-PCR and tissue microarray in familial and non familial cases showed no significant difference in expression of these genes in two groups suggesting familial clustering occurs as result of sharing of common environmental factors. Gene expression profiling of clinical specimens from well characterized populations that have familial clustering of cancer identified molecular mechanism associated with progression of esophageal cancer.

An earlier study from China and Iran suggested that mutations in *BRCA2* gene may play a role in the etiology of familial ESCC. However, the frequency of *BRCA2* gene germline mutations and its contribution to risk of familial aggregation of ESCC in high-risk region of India are not known. Screening for mutations of the *BRCA2* gene in the germline DNA was done for twenty familial and eighty non-familial ESCC patients. Our results suggest that germ-line variants (K2729N and I3412V) of *BRCA2* gene that have earlier been reported from China and Iran may play a role in familial aggregation of ESCC in high risk region of India.

Table of Contents

List of Tables

List of Figures

37	Regression plots for fold change by microarray (Y-axis*)* and quantitative real-time PCR assay (X-axis) for *ARG1* (A), *CD14* (B), *PF4* (C), *MAPK7* (D) and *EPHX1* (E).	176-178
38	Photomicrograph of tissue microarray cores from esophageal tumor biopsies obtained from familial ESCC patients showing negative immunostaining for KRT4 (A) and Collagen IV (D) and positive immunostaining for VEGF (B) and NF-kB (C).	183
39	Schematic illustration of differentially expressed genes involved in molecular mechanism of esophageal tumorigenesis. ↑ indicates up-regulated genes and ↓indicates down regulated genes. The key differentially expressed genes (*CD14, ARG1, EPHX1, MAPK7, PF4, COL4A4*, and *CK4*) were validated by RT-PCR/ tissue microarray.	184
40	*CD14* which is involved in Toll-Like Receptor Signaling Pathway significantly up-regulated in ESCC.	190
41	*FOXO1* which is involved in Insulin Signaling Pathway significantly up-regulated in our study.	191
42	Status of family history of cancer among esophageal squamous cell carcinoma patients from high-risk area of India registered at BBCI during the year of 2005-2006.	207
43	Risk estimates of lifestyle cancer risk factors (e.g., tobacco chewing, smoking, betel quid chewing and family history) in all esophageal cancer patients by Logistic regression model using SPSS 15.0 version software (SPSS, Chicago, IL, USA) package. CI indicates confidence intervals. X- Axis denotes lifestyle cancer risk factors and Y-axis denotes Odd ratio. The odds ratio (OR) and 95% confidence interval (CI) were calculated using a logistic regression model and adjusted for age and gender.	208
44	Risk estimates of lifestyle cancer risk factors (e.g., tobacco chewing, smoking, and betel quid chewing) among familial esophageal cancer patients by Logistic regression model using SPSS 15.0 version software (SPSS, Chicago, IL, USA) package. CI indicates confidence intervals. X-Axis denotes lifestyle cancer risk factors and Y-axis denotes Odd ratio. The odds ratio (OR) and 95% confidence interval (CI) were calculated	209

	using a logistic regression model and adjusted for age and gender.	
45	Heteroduplex analysis showed alteration in exon 18 & 27C of *BRCA2* gene	210-211
46	Sequence chromatograms of missense mutation as determined by automated sequence analysis. Missense mutation in exon 27 (3A) and exon 18 (3B) in *BRCA2* gene. Arrow indicates the position of changes in nucleotide sequences of *BRCA2* gene.	212-213

List of Abbreviations

ADC: Adenocarcinoma

AAR: Age-adjusted incidence rate

BBCI: Bhubanasewar Boorah Cancer Institute

CN: Copy number

CNAT: Copy Number Analysis Tool

CGH: Comparative genome hybridization

ESCC: Esophageal squamous cell carcinoma

EDTA: Ethylene Diamine Tetraaceticacid

EB: Elution Buffer

GCOS: Gene Chip Operating System

HRR: Homologous Recombination Repair

HDX: Heteroduplex

LOH: Loss of heterozygosity

MIAME: Minimum information about microarray experiment.

MDE: Mutation Detection Enhancement

NE: North-east

NCRP: National cancer registry programme

RIN: RNA Integrity Number

RFLP: Restriction Fragment Length Polymorphism

SCC: Squamous cell carcinoma

SNP: Single nucletotide polymorphism

TMA: Tissue Microarray

TBE: Trizma-base Boric Acid EDTA

Chapter 1

INTRODUCTION

INTRODUCTION

Cancer is a highly complex disease which can encompass multiple genomic alterations, including point mutations, translocations, gene amplifications, epigenetic modifications, deletions, aberrant splicing, and altered gene expression. These changes may be inherited or somatically acquired during progression from a normal to a malignant transformed cell. Among human cancer, esophageal carcinogenesis appears to be a complex multi-step process with a multi-factorial etiology, where environmental, geographical and genetic factors appear to play major role [1, 2].

Esophageal cancer is among the ten most common malignancies worldwide and ranks as the sixth leading cause of death from cancer. It constitutes 7% of all gastrointestinal cancers and is one of the most lethal of all cancers [3]. Globally esophageal cancer ranks as the sixth most common cancer among males and ninth most common cancer among females. However, in India, it is the second most common cancer among males and the forth most common cancer among females [4]. The incidence of esophageal cancer varies greatly between developed and developing countries and a 50-fold difference has been observed between high and low-risk populations. It is more common in Asian than in western countries. The esophageal cancer belt is a geographic area of high incidence, which stretches from north-central China westward through Central Asia to northern Iran [3]. A high frequency of esophageal cancer has been reported from certain parts of China, Iran, South Africa, Uruguay, France and Italy [5]. The uneven geographical distribution of esophageal cancer reflects the influence of local environmental conditions, lifestyle factors and genetic predisposition in the development of this cancer [6].

Esophageal cancer exists in two main forms with distinct etiological and pathological characteristics, squamous cell carcinoma (SCC) and adenocarcinoma. More than 90% of esophageal cancers worldwide are SCCs, although adenocarcinomas are more prevalent in the USA. SCC is often preceded by increased proliferation of esophageal epithelial cells leading to basal cell hyperplasia, dysplasia, and carcinoma in situ [5]. Several predisposing factors that lead to esophageal squamous cell carcinoma (ESCC) include tylosis, lye ingestion, achalasia and Plummer-Vinson syndrome [7].

Tobacco and alcohol, known to be involved in the carcinogenesis of esophageal cancer, have been reported to interact in a multiplicative way in the etiology of this neoplasm [8]. Tobacco smoking and alcohol drinking are strongly associated with the risk of esophageal SCC and to a lesser degree with the risk of esophageal adenocarcinoma. Reports from Southern India suggest that ESCC occurs in more than 80% of cases in chronic tobacco smokers that is further potentiated by heavy use of alcohol and additional prevailing risk factors, including nutritional factors and vitamin deficiencies [1]. Areca nut chewing, or betel quid chewing, a common habit in Southeast Asia, has been found recently by epidemiological studies from India, Taiwan and Thailand, to increase the risk of developing esophageal cancer by 4.7–13.3-fold, although other exogenous risk factors may be also have contributed to the development of the disease [8]. In South America, Central and Southeast Asia, the consumption of hot beverages is also considered as a risk factor. Many other factors such as dietary nitrosamine, micronutrient deficiency, HPV, opium, and spicy food have been suggested as possible additional factors [6]. ESCC in the Indian population is associated with poor

nutritional status, low socioeconomic conditions, bidi smoking and consumption of smokeless tobacco products, besides alcohol drinking and cigarette smoking [4].

Population-based and hospital –based cancer registries in India, have reported the highest incidence of esophageal cancer in Assam in the north east region of the country, followed by Bangalore and Bombay [9]. Assam, the largest state in this region, has a diverse population structure comprising many ethnic and linguistic groups, each having its own life styles, cultures and customs. The dietary patterns are also quite different from the rest of the country. The ethnic groups who had migrated several centuries ago from South China, Thailand, Myanmar and other Southeast Asian countries are now living as indigenous people. With the diverse tribal and non-tribal mix of life styles and cultures, this region is the paradise for epidemiologists who seek the etiological clues from the geographic variations in the occurrence of the various diseases and the relationship between the genetic and environmental factors in this area. This is especially true in the case of esophageal cancer. Betel nut chewing with or without tobacco has been shown to be independently associated with the development of esophageal cancer in Assam [9]. More than 80% of the people of Assam chew typical raw (green) ripe betel nuts after being specially fermented underground for more than a month. The impact of these differences in risk factors for ESCC in North-east Indian population, in comparison with Chinese and Western population, on the molecular pathogenesis of the disease remains to be determined.

Within these high-risk regions, studies have shown a strong tendency toward familial aggregation, suggesting that genetic susceptibility, in conjunction with potential environmental exposures, may be involved in the etiology of this cancer [10]. The

aggregation of esophageal cancer in families is a long-observed and well-documented phenomenon, but it is still subject to explore. Genetic predisposition is based on individual types of carcinogen activation or detoxification and DNA repair in a population at risk [6]. A positive family history was found to be associated with an increased risk of esophageal cancer in several case-control and cohort studies in China, Iran and Japan [11]. The familial aggregation of esophageal cancer among the Turkmen population in northern Iran may reflect the influence of environmental factors operating on individuals who are already genetically susceptible [12]. Environmental factors are probably the main contributor to the familial aggregation of cervical, lung and upper aerodigestive tract cancer [13]. The role of genetics in the development of esophageal cancer is also supported by previous reports of familial aggregation of esophageal cancer and by one segregation analysis of esophageal pedigrees that suggested an autosomal recessive Mendelian inheritance pattern. Preliminary studies have shown high frequencies of loss of heterozygosity (LOH), characteristic patterns of gene expression and significant differences in both LOH and gene expression by family history in ESCC tumors from a high-risk population in neighboring Shanxi provience of China [14]. Molecular profiles in ESCC are highly consistent and expression patterns in familial cases were different from those in sporadic cases of high-risk area of China [10]. Esophageal cancer was significantly more common in the first degree relatives of the cases than the relatives of unaffected controls in the Turkmen population of Iran. This suggests that genetic factors contribute significantly to the high incidence rate of esophageal cancer among Turkmen [15]. Familial clustering of cancer has been one of

the main avenues to the understanding of cancer etiology and the signal to the involvement of heritable genes [16].

Previous studies of ESCC have shown a high frequency of allelic loss on chromosome 13q, infrequent somatic mutations in *BRCA2*, and a suggested association between a positive family history (FH+) of upper gastrointestinal cancer and germ line *BRCA2* mutations [17]. An earlier study from China and Iran suggested that *BRCA2* may play a role in the etiology of ESCC. Recent studies reveal that the BRCA2 protein is required for maintenance of chromosomal stability in mammalian cells and functions in the biological response to DNA damage, as evidenced by the finding that mutations in *BRCA2* lead to chromosomal instability due to defects in the repair of double-strand and single-strand DNA breaks [17, 18]. Several studies have made efforts to understand the role of genetics in the etiology of SCC of the esophagus by attempting to identify molecular events associated with the development of precursor and invasive lesions.

Molecular studies of human esophageal tumors have revealed frequent genetic abnormalities. Regardless of patient origin and suspected etiological factors, genetic changes that are consistently observed in esophageal SCC are: (i) alterations in tumor suppressor genes, specifically *p53*, leading to altered DNA replication and repair, cell proliferation and apoptosis; (ii) disruption of the G_1/S cell cycle checkpoint and loss of cell cycle control; (iii) alterations in oncogene function leading to deregulation of cell signaling cascades [5].

Over expression of Cyclin B1 in ESCC has been reported to induce tumor cell invasive growth and metastasis [19]. *PTEN* may play a role in carcinogenesis and progression of ESCC in a high incidence area of northern China [20]. Most commonly

used serum markers of ESCC include cytokeratin 19-fragment, squamous cell carcinoma antigen, and carcinoembryonic antigen, which show poor sensitivity of 43.9%, 26.8% and 17% respectively. Despite advancement in therapy, the overall 5-year survival rate of esophageal cancer still remains less than 30%. This has been largely attributed to the lack of an early diagnostic marker with desired sensitivity and specificity [7]. Additional markers that are known to be over expressed in early stages of ESCC include ALCAM, BPA-2, MMP13, MMP7 and MMP9. Coexpression of *MMP7, MMP9, and MMP13* has also been associated with poor outcome in ESCC [7]. It has been previously reported that E-cadherin is commonly down regulated by CpG island hypermethylation in ESCC that predicts the poor prognosis of patients [21, 22]. Reduced expression of TGM3 may play an important role in esophageal carcinogenesis [23, 24]. Altered Ras function can contribute to human esophageal SCC development [5]. Up-regulation of Aurora-A expression may reflect the malignant behavior of ESCC and may prove useful information as a prognostic factor for ESCC patients [25, 26]. Over expression of oncogenes *Fra-1* and Neogenin and cell cycle-related genes *Id-1* and *CDC25B* were reported in ESCC [27]. *VEGF-C* may also play a role in tumor progression via lymphangiogenesis and angiogenesis in human esophageal carcinoma [28, 29, 30]. It has been demonstrated that COX-2 is expressed in the majority of esophageal SCCs and COX-2 derived prostaglandins play an important role in the regulation of proliferation and apoptosis of esophageal tumor cells [31, 32]. *EMP3* may be a tumor suppressor gene at the last step of esophageal carcinogenesis [33]. Down-regulation of RUNX3 may play a role in disease progression of ESCC [34]. mRNA and protein level of CDC25B and LAMC2 were over expressed in esophageal squamous cell carcinomas and premalignant

lesions in subjects from a high-risk population in China [35]. A loss of FHIT expression is associated not only with alcohol-induced esophageal carcinogenesis, but also with multicentric carcinogenesis [36]. Stefin A plays an important role in the growth, angiogenesis, invasion, and metastasis of esophageal cancer suggesting that stefin A may be useful in its therapy [37].

An earlier study has reported that Arachidonic acid metabolism pathway and its altered expression may contribute to esophageal squamous cell carcinogenesis [38]. Detection of DDH over expression in ESCC cells would then provide an alternative link between chronic inflammation and carcinogenesis of esophagus, and, possibly, the disease progression of ESCC [39]. Survivin mRNA expression in ESCC has been reported as a good biomarker for identifying patients with high risk of cancer recurrence [40]. It has been reported earlier that OPN is associated with esophageal tumorigenesis and progression, but not patients' survival [41]. Up-regulation of *MAP3K3, AKAP13, ZnT7* and *TG2* and deregulation of genes associated with zinc homeostasis were reported in ESCC of low-risk region of India [4]. Studies on differential global gene expression profiling in ESCCs using cDNA and oligonucleotid arrays have been performed in Chinese, Japanese and Western population. These studies revealed the complex nature of molecular changes associated with development of ESCC and suggested deregulation of differentiation associated genes and arachiodonic acid metabolism in ESCC [4].

Chromosomal or gene copy number alterations are one of the main important mechanisms that perturb normal gene function by inducing changes reflected in gene expression [42]. Gene amplifications and deletions frequently contribute to tumorigenesis [43]. Earlier studies have shown a very high level of chromosomal instability in ESCC

from this high-risk area of China. Frequent allelic deletions and other genetic abnormalities affecting individual tumor suppressor genes have been detected in these tumors. Chromosomal regions with frequent allelic loss may point to major susceptibility genes that will assist in understanding molecular events involved in esophageal carcinogenesis and serve as the basis for the development of markers for genetic susceptibility testing and screening for the early detection of cancer [44]. Genomic alterations include various types of mutations, translocations, and copy number alterations. The last category involves chromosomal regions with either more than two copies (amplifications), one copy (heterozygous deletions), or zero copies (homozygous deletions) in the cell. Genes contained in amplified regions are natural candidates for cancer-causing oncogenes, while those in regions of deletion are potential tumor-suppressor genes. Thus, the localization of these alterations in cell lines and tumor samples is a central aim of cancer research [45]. Comparative Genome Hybridization (CGH) and loss of heterozygosity (LOH) analysis are standard approaches used to characterize copy number changes on a whole-genome or candidate gene level. LOH studies and conventional comparative genomic hybridization (CGH) analyses have demonstrated genetic complexity in ESCC and have identified multiple recurrent copy number alterations, namely gains of 1q, 2q, 3q, 5p, 7p, 7q, 8q, 11q, 12p, 12q, 14q, 17q, 20p and 20q. Amplifications of regions harboring oncogenes e.g. 7p12 (*EGFR*), 8q24 (*MYC*), 17q21 (*FGFR*) and 11q13 (*CCND1, FGF4/3, and EMS1*) have consistently been observed. Losses, albeit at a lower frequency than gains, have recurrently involved 3p, 5q, 9p, 13q, 18q and 21q and include target genes such as *FHIT, APC, RB1* and *CDKN2A* [46]. Frequent deletions were found on chromosome arms 1p, 3p, 4p, 5q, 8p, 9p, 9q, 11q,

13q, 16p, 17p, 18q, 19p, and 19q in ESCC [47]. Some of the changes identified, e.g. gain of 8q24, 11q13, 12p and 20q12 and loss of 3p have been associated with poor prognosis, but genetic alterations and biological characteristics have so far had a limited impact on clinical prognostication and treatment [46].

Traditional techniques, such as microsatellite analysis and BAC-CGH, are laborious, have limited whole-genome resolution, and represent challenges to data standardization initiatives. Gene Chip oligonucleotide arrays provide a standard platform to enable data standardization, as well as higher resolution, to detect smaller changes and map genetic boundaries. Furthermore, single nucleotide polymorphism (SNP) arrays offer significant advantages over traditional methods by combining copy number, LOH, and SNP genotyping analysis into a single assay, thereby enabling the detection of copy neutral LOH events.

Cancer development is a complex multi-step process, involving various genetic and epigenetic changes. Progress of phenotypes from normal to advanced carcinoma is controlled by a transcriptional hierarchy that coordinates the action of hundreds of genes [48]. During the past decade, increasingly research results have distinguished how these genomic perturbations drive cancer cell survival by alteration the mechanism for cell cycle control, DNA repair, differentiation, apoptosis, tumor vascularization, and metabolism. By improving the understanding of these molecular mechanisms, scientists have gained insight into the initiation of cancer, its progression and its sensitivity to different therapeutics. Conventional approaches investigating one or several candidate genes at a time can not show the whole story of carcinogenesis. The generation of vast amounts of DNA sequence information, coupled with advances in technologies

developed for the experimental use of such information, allows the description of biological processes from a view of global genetic perspective. However, little is known about the exact expression changes in each stage of tumorigenesis, which will help us to identify the exact series of events that leads to the initiation and progression of cancer development. As a step toward understanding the complicated changes between normal and malignant cells, this report focused on gene expression profile variations among normal and abnormal esophageal epithelium tissues. To illustrate the mechanisms controlling malignant changes at molecular level may provide a further understanding of tumorigenesis, as well as new approaches in early detection and treatment of esophageal cancer. One such technology, DNA microarray, permits simultaneous monitoring of thousands of genes [48]. 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. cDNA microarray technology is a useful tool to discover genes frequently in esophageal neoplasia and provides novel clues to diagnosis, early detection and intervention of SCC [49].

The gene expression profile of esophageal cancer in a high incidence region of India where tobacco use and alcohol consumption are widespread and the users of these two substances are also betel quid chewers, has so far not been investigated. In the current study, cDNA microarray gene expression analysis was done to identify the genes differentially expressed in esophageal cancer associated with prevalent risk factors such as tobacco use and betel quid chewing in a high-risk Indian population. The aim was to analyze genes and pathways that may be involved in tobacco and betel quid chewing

related esophageal malignancies. This is the first study to provide gene expression profile of esophageal cancer in this region of India. In addition, we established a high-resolution chromosomal instability profile by analyzing germ-line and tumor DNA with a GeneChip® Human Mapping 10K Array Xba 142 2.0 to identify common chromosomal aberrations that occur in patients with ESCC from north-east area of India.

The genetic and molecular factors responsible for familial clustering of esophageal cancer in NE region of India are not well delineated. The use of high throughput genomic technologies on large numbers of samples from well-characterized populations that are known to have familial clustering of cancer may lead to identification of molecular biomarker for cancer predisposition, progression and target for therapeutic manipulation beside elucidating the molecular functional pathways involved in carcinogenesis. An earlier study from China and Iran suggested that mutations in *BRCA2* gene may play a role in the etiology of familial ESCC. However, the frequency of *BRCA2* gene germ-line mutations and its contribution to risk of familial aggregation in high-risk region of India are not known. Screening for germ-line mutations in *BRCA2* gene has been carried out in familial ESCC cases and non-familial ESCC cases to confirm whether variants identified in cases are associated with elevated risk of familial ESCC in high-risk region of India.

In the current study, molecular signature of patients with esophageal cancer from high risk area of India have been studied by gene expression profiling and SNP array to understand molecular pathogenesis of esophageal cancer and to identify potential biomarkers, if any, for cancer predisposition, progression and therapeutic manipulation.

Chapter 2

AIMS & OBJECTIVES

Aims and Objectives

Aims:

A. To identify significant genomic alterations responsible for the emergence and progression of esophageal cancer in a high-risk region of India and to investigate the role of gene-environment interaction.

B. To identify specific genes or pathways related to esophageal squamous cell carcinoma in high-risk region of India.

C. Identification of potential new biomarkers for early detection of progression and/or treatment targets.

To achieve the projected aims, the present study was undertaken with the following objectives:

Objective no. 1

Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma of high-risk area in India by SNP array.

Objective no. 2

Gene expression profile of esophageal squamous cell carcinoma and its association with tobacco and betel quid consumption in high-risk area of India by cDNA microarray analysis.

Objective no. 3

Molecular profiling to identify molecular mechanism in esophageal cancer with familial clustering.

In the current study, molecular signature of ESCC from high risk area of India has been studied by gene expression profiling in esophageal cancer patients with family

history of esophageal cancer with the aim to elucidate molecular pathogenesis of esophageal cancer in these patients.

Objective no. 4

Contribution of germ line *BRCA2* sequence alterations to risk of familial aggregation of esophageal cancer in high-risk area of India.

Chapter 3

Review of Literature

A. Global scenario of esophageal cancer:

Esophageal cancer ranks as the sixth most common cancer among males and ninth most common cancer among females globally [4]. Overall, it is the eighth most common cancer worldwide, responsible for 462,000 new cases in 2002 (4.2% of all cancer cases), and sixth most common cause of death from cancer with 386,000 deaths (5.7% of all cancer cases). It constitutes 7% of all gastrointestinal cancers and is one of the most lethal of all cancers [3]. It has a very poor survival: 16% of the cases in the United States and 10% in Europe survive at least five years [50]. Esophageal cancer is a particularly aggressive tumor with poor prognosis [51]. Despite advances in multimodality therapy, due to late diagnosis and poor efficacy of treatment, the prognosis for patients with esophageal squamous cell carcinoma (ESCC) still remains poor with an average 5-year survival of less than 10% globally and less than 12% in India [4].

An uneven geographical distribution in the occurrence of esophageal cancer reflects the influence of local environmental conditions, lifestyle and genetic predisposition in the development of the cancer [6, 3]. The incidence of esophageal cancer varies greatly between populations, with a greater than 50-fold difference observed between rates in high and low-risk populations. It is more common in Asia than in western countries. The esophageal cancer belt is a geographic area of high incidence, which stretches from north-central China westward through Central Asia to northern Iran [3]. The high-incidence areas include the Henan and Shanxi provinces of China, the Caspian littoral of Iran, South Africa, Normandy in France, Italy and some areas (Uruguay) in South America [6, 5]. Esophageal cancer is more common in males in most areas— the sex

ratio is 7:1 in Eastern Europe, for example— although in the high-risk areas of Asia and Africa, the sex ratio is much closer to unity [50]. In case of male and female, the incidence and mortality rate of esophageal cancer is higher in developing countries (Fig.1) [50].

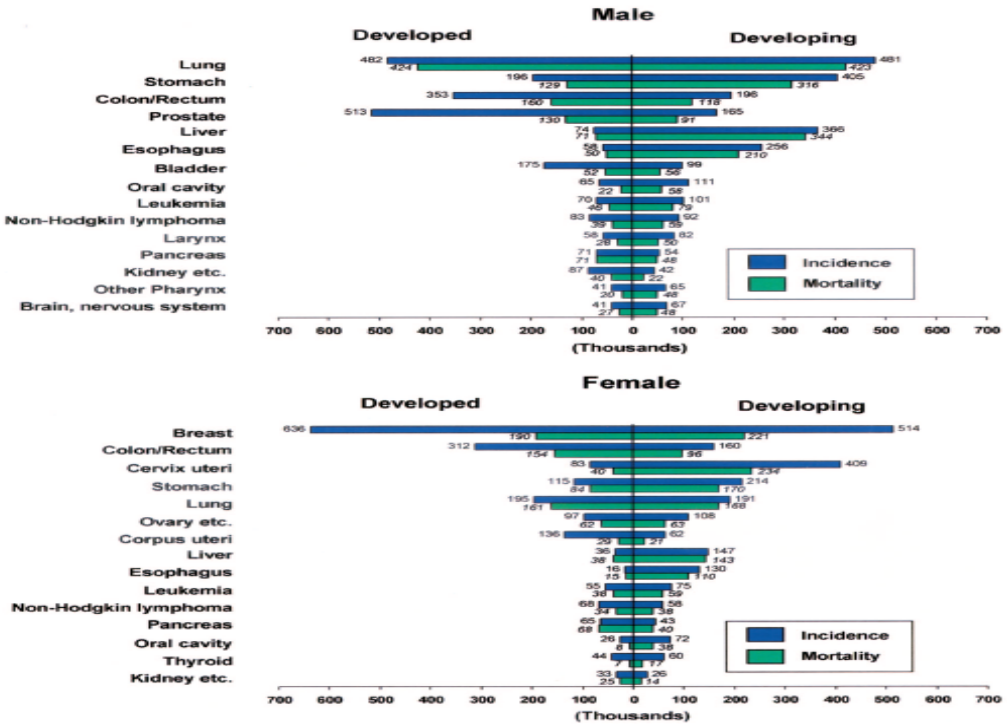


Figure 1: Incidence and mortality rate of esophageal cancer in male and female of developed and developing countries. (Global Cancer Statistics, 2002: CA Cancer J Clin. 2005 Mar-Apr; 55(2):74-108)

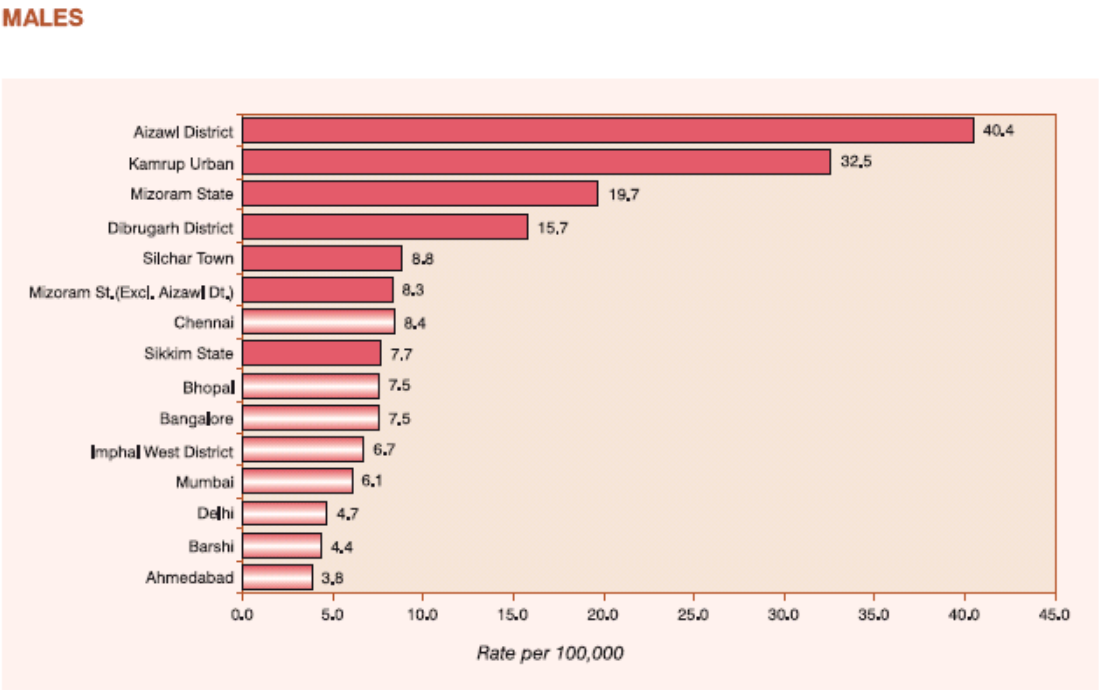
B. Histological subtypes of esophageal cancer:

The two major histological subtypes of esophageal cancer are squamous cell carcinoma (SCC) and adenocarcinoma (ADC). More than 90% of esophageal cancers worldwide are SCCs, although adenocarcinomas are more prevalent in the USA. The principal precursor lesion of esophageal SCC is epithelial dysplasia. Microscopically these lesions represent an accumulation of atypical cells with nuclear hyperchromasia, abnormally clumped chromatin and loss of polarity. There is evidence from prospective studies that esophageal SCC probably develops through a progressive sequence from mild to severe dysplasia, carcinoma *in situ* and, finally, invasive carcinoma. These tumors frequently present as fungating, ulcerating or infiltrating lesions in the esophageal epithelium. Microscopically, esophageal SCCs range from well-differentiated tumors that exhibit keratinization, moderate nuclear atypia and minimal necrosis to poorly differentiated tumors with a high mitotic index and large areas of necrosis. A large majority of patients with cancer of the esophagus present with advanced metastatic disease. The prognosis for such cases is poor; the overall 5 year survival rate of patients with metastatic disease is <10% [5].

C. Indian scenario of esophageal cancer:

It is the second most common cancer among males and the forth most common cancer among females in India [4]. Cancer data from both population-based and hospital –based cancer registries in India, showed the highest incidence of esophageal cancer to occur in Assam in the north east region of the country, followed by Bangalore and Bombay (NCRP) (Fig.2) [9]. Assam is the major state in this region with 71% population share of the region having diverse population structure comprising many ethnic and

linguistic groups each having its own life styles, cultures and customs. The dietary patterns are also quite different from the rest of the country. The ethnic groups who had migrated several centuries ago from South China, Thailand, Myanmar and other Southeast Asian countries are now living as indigenous people. With the diverse tribal and non-tribal mix of life styles and cultures, this region is the paradise for epidemiologists who seek the etiological clues from the geographic variations in the occurrence of the disease and the relationship between the genetic and environmental factors in these areas. This is especially true in the case of cancer esophagus.



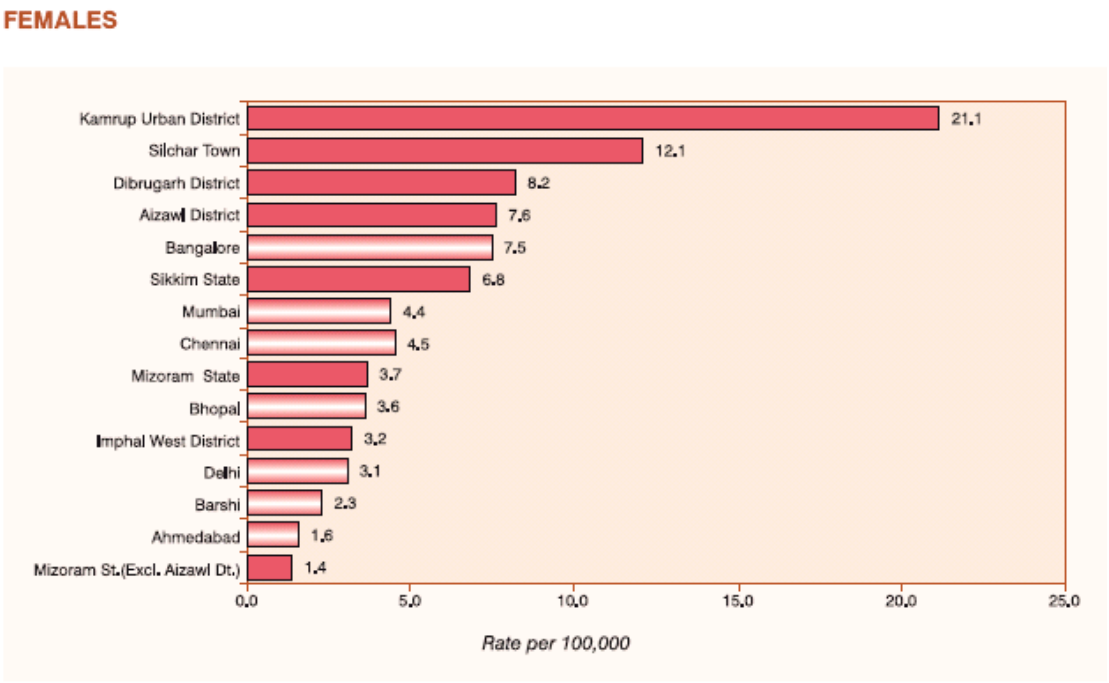


Figure 2: Comparison of age adjusted incidence rates of esophageal cancer in males and females across all PBCRs. (NCRP report: 2001-2004)

D. Risk Factors of esophageal cancer:

Among human cancer, esophageal carcinogenesis appears to be a complex multi-step process with a multi-factorial etiology, where environmental, geographical, and genetic factors appear to play major roles [1]. The large variation in the incidence of esophageal cancer in different geographic regions has often been thought to be due to variation in exposure to environmental factors; however, other studies suggest that hereditary factors may also contribute to the variation in rates [3]. The etiology of esophageal cancer in North-east Indian population is different from western population and other Indian population due to wide variations in dietary habits or nutritional factors, tobacco and alcohol habits.

Excessive use of tobacco has been implicated as a principal factor in the etiology of esophageal SCC. Several tobacco constituents, including nitrosamines, polycyclic aromatic hydrocarbons, aromatic amines, various aldehydes and phenols, may be causally related to esophageal cancer [5]. There are more than 60 carcinogens in cigarette smoke and at least 16 in unburned tobacco. Among these, tobacco specific nitrosamines (such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and *N*′-nitrosonornicotine (NNN)), polycyclic aromatic hydrocarbons (such as benzo[*a*]pyrene) and aromatic amines (such as 4-aminobiphenyl) seem to have an important role as causes of cancer [52]. Alcohol consumption has been shown to further increase the risk of SCC in the esophagus of tobacco smokers. Extensive research in China and South Africa has suggested that *N*-nitroso compounds and their precursors are probable etiological factors for esophageal cancer in these high incidence areas [5]. Reports from Southern India suggest that ESCC occurs in more than 80% of cases in chronic tobacco smokers that is further potentiated by heavy use of alcohol and additional prevailing risk factors, including nutritional factors and vitamin deficiencies. Tobacco and alcohol, known to be involved in the carcinogenesis of esophageal cancer, have been reported to interact in a multiplicative way in the etiology of this neoplasm [1]. Tobacco smoking and alcohol drinking are strongly associated with the risk of esophageal SCC and to a lesser degree with the risk of esophageal adenocarcinoma. The average relative risk of tobacco smoking is 2.0–5.0 for esophageal SCC and 1.5–2.5 for esophageal adenocarcinoma [53].

Areca nut chewing, or betel quid chewing, a common habit in Southeast Asia, has been found recently by epidemiological studies from India, Taiwan and Thailand, to increase the risk of developing esophageal cancer by 4.7–13.3-fold, although other

exogenous risk factors may be also have contributed to the development of the disease in these studies [8]. More than 80% of the people of Assam chew typical raw (green) ripe betel nuts after being specially fermented underground for more than a month. These specially processed nuts (Area catechu) that are at different stages of decay and give pungent smell are chewed with or without betel leaf slaked lime and tobacco. Betel nut chewing with or without tobacco has been shown to be independently associated with the development of esophageal cancer in Assam. The betel nut (*Areca catechu*) has been shown to have carcinogenic potential and 3-methyl nitrosamine propionitrile (MNPN), a potent carcinogen and safrole-like DNA adducts have been detected in the saliva of betel chewers. Arecoline, present in the nut have been shown to be genotoxic and mutagenic and it can produce 2 N-nitrosamine carcinogens [9].

Consumption of salt-pickled, salt-cured and moldy foods has also been implicated in the pathogenesis of this disease. Some of these products are frequently contaminated with *N*-nitrosamine carcinogens and/or fungal toxins. Other factors associated with an increased risk of esophageal SCC include vitamin and trace mineral deficiencies. Plasma levels of vitamins A, C and E, among others, tend to be lower in patients with esophageal cancer. An inverse relationship has been noted between mortality caused by esophageal cancer and levels of zinc, selenium and other trace elements in foods from high risk areas for this disease. Consumption of hot beverages, such as tea, and fungal invasion in esophageal tissues leading to localized inflammation and irritation has been suggested as additional promoting factors for cancers of the esophagus in South America, Central and Southeast Asia [5]. Case control studies from hospital have demonstrated increased

incidence of esophageal cancer in people with poor nutritional status and low socioeconomic groups [4].

Finally, a role for human papilloma virus (HPV) has also been suggested in the etiology of SCC of the esophagus. A low frequency of HPV-16 or HPV-18 positivity has been reported in esophageal tumor samples [5].

Extended epidemiological studies in high incidence areas, such as Northern Iran, Northern China and South Africa, provide evidence that exposure to specific diet-related nitroso compounds in parallel with nutritional deficiencies and consuming food contaminated by mycotoxins are the most important determinants of the disease. The high incidence of esophageal cancer has been attributed to a probable exposure to nitroso compounds, amines and nitrates to be present in local food stuff [1].

E. Genetic factors (Family history) responsible for esophageal cancer:

Although epidemiological studies indicate that tobacco smoking and alcohol consumption are the major risk factors for ESCC in low-risk regions of Europe and North America, the etiological agents in high-risk regions have yet to be convincingly identified. Within these high-risk regions, studies have shown a strong tendency toward familial aggregation, suggesting that genetic susceptibility, in conjunction with potential environmental exposures, may be involved in the etiology of this cancer [10]. Genetic predisposition is based on individual types of carcinogen activation or detoxification and DNA repair in a population at risk [6]. The aggregation of esophageal cancer in families is well-documented; however the predisposing factors have not been so far illucidated.

A positive family history was found to be associated with an increase risk of esophageal cancer in several case-control and cohort studies in China, Iran and Japan

[11]. The influence of environmental factors operating on genetically predisposing individuals leads to familial aggregation of esophageal cancer among the Turkmen population in northern Iran [12]. Environmental factors are probably the main contributor to the familial aggregation of cervical, lung and upper aerodigestive tract cancers, and minor contributor to familial risks for melanoma and squamous cell skin cancer [13]. Molecular profiles in esophageal squamous cell carcinoma are highly consistent and that expression patterns in familial cases differ from those in sporadic cases of high-risk area of China [10].

Familial clustering of cancer has been one of the main avenues to the understanding of cancer etiology and the signal to the involvement of heritable genes. Familial clustering of cancer may be due to environmental factors shared by family members or due to shared genes [16]. Esophageal cancer was significantly more common in the first degree relatives of the cases than the relatives of unaffected controls in Turkmen population of Iran. This suggests that the genetic factors may contribute to the high incidence rate of esophageal cancer among Turkmen [15].

The impact of these differences in risk factors for ESCC in Indian population, in comparison with Chinese and Western population, on the molecular pathogenesis of the disease remains to be determined.

F. Molecular alterations in human esophageal SCC

Esophageal mucosa is lined by a stratified squamous epithelium that functions in epithelial proliferation, differentiation and turnover, and serves as a barrier. Cellular proliferation is frequent in the epibasal layers, and loss of contact with the basement membrane eventually triggers differentiation and the sequential expression of

differentiation markers. Squamous cell differentiation requires the coordinated activation and repression of genes specific to the differentiation process, and disruption of this program accompanies neoplasia. Several studies have made efforts to understand the role of genetics in the etiology of SCC of the esophagus by attempting to identify molecular events associated with the development of precursor and invasive lesions [54].

Molecular studies of human esophageal tumors have revealed frequent genetic abnormalities. Regardless of patient origin and suspected etiological factors, genetic changes that are consistently observed in esophageal SCC are: (i) alterations in tumor suppressor genes, specifically *p53*, leading to altered DNA replication and repair, cell proliferation and apoptosis; (ii) disruption of the G_1/S cell cycle checkpoint and loss of cell cycle control; (iii) alterations in oncogene function leading to deregulation of cell signaling cascades [5].

Altered Ras function can contribute to human esophageal SCC development. Some of the other genetic alterations that are commonly associated with clinical tumors include *p53* mutations, loss of p16MST1 and/or p15 and/or RARß expression, amplification of cyclin D1, HST-1, EGFR and INT-2, elevations in iNOS, hTERT, BMP-6, COX-2 and c-Myc expression and cytoplasmic ß-catenin levels. In addition, loss of heterozygosity on chromosomes 1p, 3p, 4, 5q, 9, 11q 13q, 17q and 18q have also been frequently observed in tumors, supporting a loss of putative tumor suppression function [5]. Preliminary studies have shown high frequencies of loss of heterozygosity (LOH), characteristic patterns of gene expression and significant differences in both LOH and gene expression by family history in ESCC tumors from a high-risk population in neighboring Shanxi provience of China [14] . ESCC develops as the result of a sequence of histopathological

changes that typically involves esophagitis, atrophy, mild to severe dysplasia, carcinoma in situ and finally, invasive cancer. Genetic changes associated with the development of ESCC include mutation of the *p53* gene, disruption of cell-cycle control in G1 by several mechanisms inactivation of p16MTS1, amplification of Cyclin D1, alterations of RB, activation of oncogenes e.g., *EGFR*, *c-MYC* and inactivation of several tumor suppressor genes [55].

Over expression of Cyclin B1 in ESCC have been reported to induce tumor cell invasive growth and metastasis [19]. *PTEN* may play a role in carcinogenesis and progression of ESCC in a high incidence area of northern China [20]. Most commonly used serum markers of ESCC include cytokeratin 19-fragment, squamous cell carcinoma antigen, and carcinoembryonic antigen, which show poor sensitivity of 43.9%, 26.8% and 17% respectively. Additional markers that are known to be over expressed in early stages of ESCC include ALCAM, BPA-2, MMP13, MMP7 and MMP9. Co-expression of *MMP7, MMP9, MMP13*, has also been associated with poor outcome in ESCC [7]. It has been previously reported that E-cadherin is commonly down regulated by CpG island hypermethylation in ESCC that predicts the poor prognosis of patients [21, 22]. Cytokeratin expression is a reliable biochemical indicator of the epithelial differentiation process. Three different cytokeratins, KRT4, KRT13 and KRT15, were found to be down regulated, whereas KRT16 and KRT17 were up regulated in ESCC. It was revealed that the expression of cytokeratins (KRT4 and KRT13) and the cross-linked envelope were aberrant during the different stages of terminal differentiation in ESCC. Altered expression of several cytokeratins might be associated with the abnormal differentiation and proliferation of ESCC [54, 55]. Reduced expression of TGM3 may play an important

an important role in esophageal carcinogenesis [23, 24]. Up-regulation of Aurora-A expression may reflect the malignant behavior of ESCC and may prove useful information as a prognostic factor for ESCC patients [25, 26]. Esophagin and PCNA are reported to be biomarkers of esophageal cellular differentiation and proliferation [56].

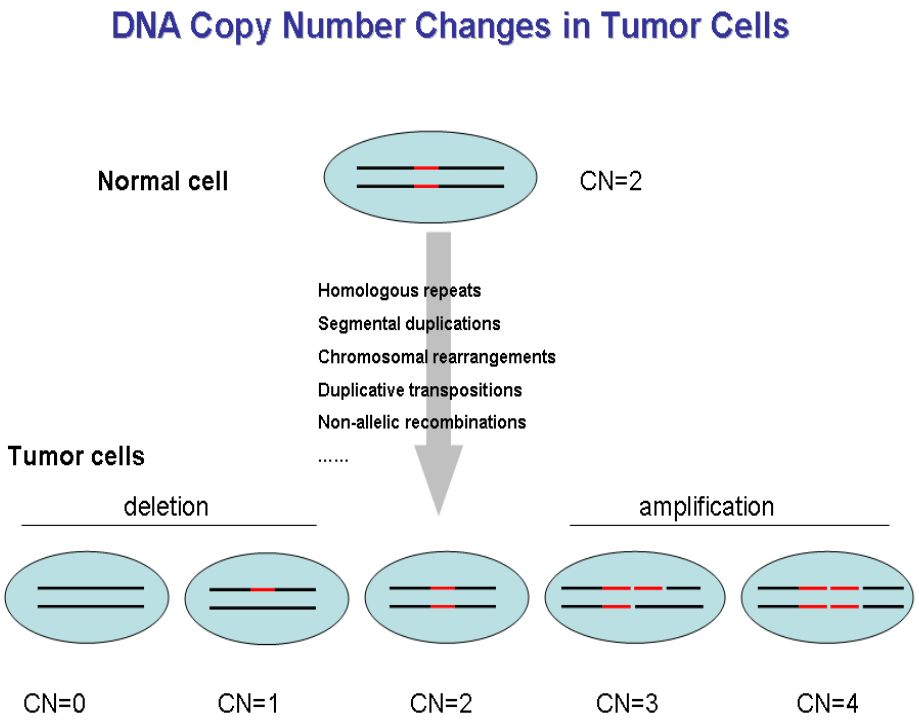
Over expression of oncogenes *Fra-1* and Neogenin and cell cycle-related genes *Id-1* and *CDC25B* were reported in ESCC [27]. *VEGF-C* may play a role in tumor progression via lymphangiogenesis and angiogenesis in human esophageal carcinoma [28, 29, 30]. It has been demonstrated that COX-2 is expressed in the majority of esophageal SCCs and COX-2 derived prostaglandins play an important role in the regulation of proliferation and apoptosis of esophageal tumor cells [31, 32]. *EMP3* may be a tumor suppressor gene at the last step of esophageal carcinogenesis [33]. Down-regulation of RUNX3 may play a role in disease progression of ESCC [34]. mRNA and protein level of CDC25B and LAMC2 were over expressed in esophageal squamous cell carcinomas and premalignant lesions in subjects from a high-risk population in China [35]. A loss of FHIT expression is associated not only with alcohol-induced esophageal carcinogenesis, but also with multicentric carcinogenesis [36]. Stefin A plays an important role in the growth, angiogenesis, invasion, and metastasis of human esophageal squamous cell carcinoma cells and suggests that stefin A may be useful in cancer therapy [37]. Genes including annexin I, small proline-rich proteins, calcium- binding S100 proteins, transglutaminase 3, *KRT4, KRT13* and cystatin A involved in squamous cell differentiation were coordinately down-regulated in ESCC. Moreover, genes associated with invasion or proliferation were up regulated, including genes such as fibronectin, secreted protein acidic and rich in cystein (*SPARC*), cathepsin B and *KRT17*. Multiple

genes are involved in ESCC, including cyclin D1, *EGFR*, annexin I, cystatin B, int-2, *CASK* and keratin 4 [54]. Over expression of MET protein was more often seen in well/moderately differentiated than in poorly differentiated ESCC [57]. Previous study reported that Arachidonic acid metabolism pathway and its altered expression may contribute to esophageal squamous cell carcinogenesis [38]. Detection of DDH over expression in ESCC cells would then provide an alternative link between chronic inflammation and carcinogenesis of esophagus, and, possibly, the disease progression of ESCC [39]. Survivin mRNA expression in ESCC has been reported as a good biomarker for identifying patients with high risk of cancer recurrence [40]. It has been reported earlier that OPN is associated with esophageal tumorigenesis and progression, but not patients' survival [41]. Up-regulation of *MAP3K3, AKAP13, ZnT7* and *TG2* and deregulation of genes associated with zinc homeostasis were reported in ESCC of low-risk region of India [4].

Studies on differential global gene expression profiling in ESCCs using cDNA and oligonucleotid arrays have been performed in Chinese, Japanese and Western population. These studies revealed the complex nature of molecular changes associated with development of ESCC and suggested deregulation of differentiation associated genes and arachiodonic acid metabolism in ESCC [4]. Several studies have made efforts to understand the role of genetics in the etiology of SCC of the esophagus by attempting to identify molecular events associated with the development of precursor and invasive lesions.

Chromosomal or gene copy number alterations are one of the main important mechanisms that perturb normal gene function by inducing changes reflected in gene

expression [42]. Gene amplifications and deletions frequently contribute to tumorigenesis [43]. Earlier studies have shown a very high level of chromosomal instability in ESCC from this high-risk area of China. The molecular events associated with the initiation and progression of esophageal squamous cell carcinoma remains poorly understood, although frequent allelic deletions and other genetic abnormalities affecting individual tumor suppressor genes have been detected in these tumors. Chromosomal regions with frequent allelic loss may point to major susceptibility genes that will assist in understanding molecular events involved in esophageal carcinogenesis and serve as the basis for the development of markers for genetic susceptibility testing and screening for the early detection of this cancer [44]. Genomic alterations are believed to be the major underlying cause of cancer. These alterations include various types of mutations, translocations, and copy number alterations. The last category involves chromosomal regions with either more than two copies (amplifications), one copy (heterozygous deletions), or zero copies (homozygous deletions) in the cell (Fig.3) [45].



Copy number (CN) alterations:

- Amplifications
- Deletions
- Aneuploidy (abnormal number of chromosomes)

Figure 3: DNA copy number alterations in cancer cells.

Genes contained in amplified regions are natural candidates for cancer-causing oncogenes, while those in regions of deletion are potential tumor-suppressor genes. Thus, the localization of these alterations in cell lines and tumor samples is a central aim of cancer research [45]. Comparative Genome Hybridization (CGH) and loss of heterozygosity (LOH) analysis are standard approaches used to characterize copy number changes on a whole-genome or candidate gene level. Traditional techniques, such as

microsatellite analysis and BAC-CGH, are laborious, have limited whole-genome resolution, and represent challenges to data standardization initiatives. LOH studies and conventional comparative genomic hybridization (CGH) analyses have demonstrated genetic complexity in ESCC and have identified multiple recurrent copy number alterations, namely gains of 1q, 2q, 3q, 5p, 7p, 7q, 8q, 11q, 12p, 12q, 14q, 17q, 20p and 20q. Amplifications of regions harboring oncogenes e.g. 7p12 (*EGFR*), 8q24 (*MYC*), 17q21 (*FGFR*) and 11q13 (*CCND1, FGF4/3, and EMS1*) have consistently been observed. Losses, albeit at a lower frequency than gains, have recurrently involved 3p, 5q, 9p, 13q, 18q and 21q and include target genes such as *FHIT, APC, RB1* and *CDKN2A*. Moreover, some of the changes identified, e.g. gain of 8q24, 11q13, 12p and 20q12 and loss of 3p have been associated with poor prognosis, but genetic alterations and biological characteristics have so far had a limited impact on clinical prognostication and treatment [46]. Frequent deletions were found on chromosome arms 1p, 3p, 4p, 5q, 8p, 9p, 9q, 11q, 13q, 16p, 17p, 18q, 19p, and 19q in ESCC [47].

Gene Chip oligonucleotide arrays provide a standard platform to enable data standardization, as well as higher resolution, to detect smaller changes and map genetic boundaries. Furthermore, single nucleotide polymorphism (SNP) arrays offer significant advantages over traditional methods by combining copy number, LOH, and SNP genotyping analysis into a single assay, thereby enabling the detection of copy neutral LOH events (Fig.4 & Fig.5).

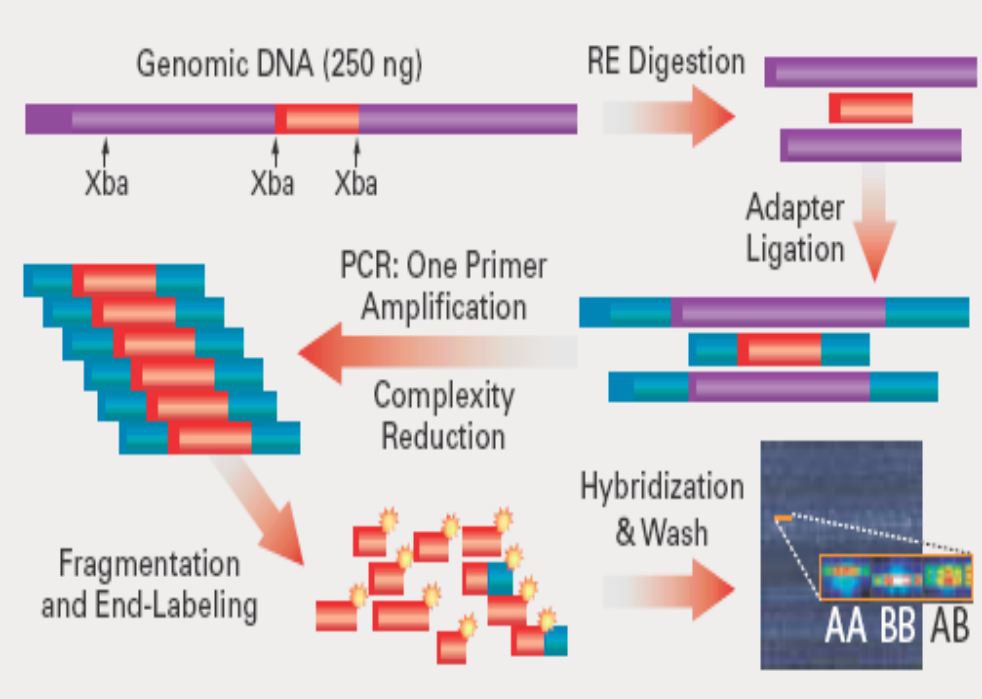


Figure 4: GeneChip® Mapping Assay Overview. (Affymetrix 10K SNP Array Manual: www.affymetrix.com)

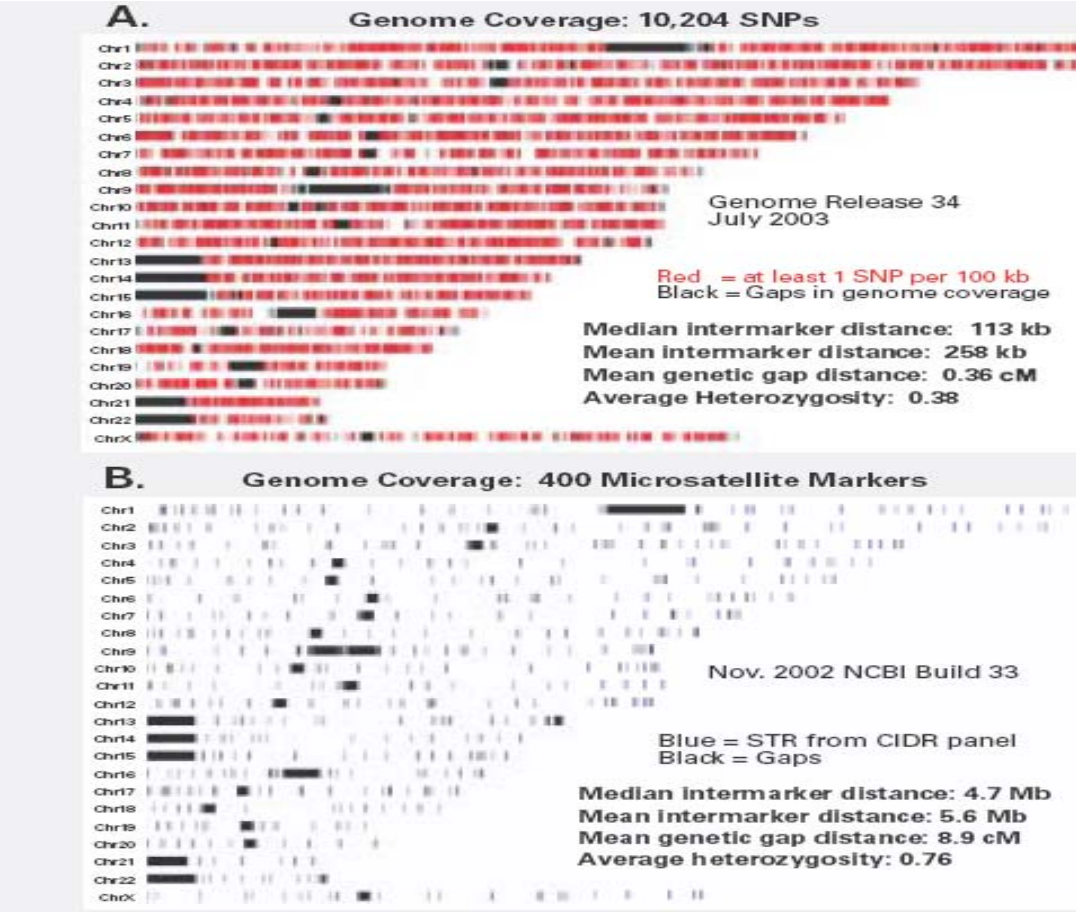


Figure 5. A Genome Coverage of Mapping 10K 2.0 SNPs by chromosome. Black areas represent gaps in the human genome sequence, primarily centromeres and telomeres.

B. Genome coverage of 400 microsatellite markers from the CIDR panel, by chromosome

(Affymetrix 10K SNP Array Manual: www.affymetrix.com)

G. Role of *BRCA2* gene in the genetic susceptibility of familial esophageal cancer

Previous studies of esophageal squamous cell carcinoma (ESCC) have shown a high frequency of allelic loss on chromosome 13q, infrequent somatic mutations in *BRCA2*, and a suggested association between a positive family history (FH+) of upper gastrointestinal cancer and germline *BRCA2* mutations. Germline mutations in *BRCA2* cause increased susceptibility to breast, ovarian, and other cancer types and have been identified in individuals of different races and ethnic groups with varying frequencies. Most of the deleterious alterations described in *BRCA2* are frameshift mutations that result in a truncated protein; however, in many cases of hereditary breast and ovarian cancer, amino-acid changes of unknown significance are seen. A previous study from China and Iran suggested that *BRCA2* might play a role in the etiology of ESCC. Germ line mutations in *BRCA2* in ESCC patients from this high-risk area of China are more frequent in FH+ than FH- cases, suggesting that *BRCA2* may play a role in genetic susceptibility to familial ESCC. Five *BRCA2* germ line mutations (N1600del, A2054P, Q2580H, V2109I, and C315S) were identified in six familial ESCC patients of China [17]. Deleterious *BRCA2* variant was identified in 15 of 197 ESCC cases in Turkmen population of Iran [15].

Recent studies reveal that the BRCA2 protein is required for maintenance of chromosomal stability in mammalian cells and functions in the biological response to DNA damage, as evidenced by the finding that mutations in *BRCA2* lead to chromosomal instability due to defects in the repair of double-strand and single-strand DNA breaks. Earlier studies have shown a very high level of chromosomal instability in ESCC from this high-risk area of China [17]. BRCA2 is mainly involved in homologous

recombination repair (HR) through control of RAD51 recombinase and it interacts with many other proteins involved in various cellular functions, including cell cycle regulation, transcription regulation, cytokinesis and control of cell proliferation [15]. Rad51 is a key protein in HR and BRCA2 may serve as a scaffold for regulation of Rad51 induced nucleoprotein filaments, as well as of its nuclear localization. BRCA2 was found to interact with a DNA binding protein, BRCA2-associated factor 35 (BRAF35), in close association with condensed chromatin. A role of the BRCA2–BRAF35 complex in resolving and packaging of entangled chromatin fibers, or maintenance of chromosomes throughout segregation at mitosis was suggested. BRCA2 becomes phosphorylated by the mitotic checkpoint protein hBUBR1 in cells with microtubule disruption, and BRCA2 (or hBUBR1) deficiency could result in genomic instability. BRCA2 might activate transcription by modulating histone acetylation. BRCA2 interacts with the transcriptional co-activator protein P/CAF (p300/CBP-associated factor) and its associated p300/CBP, both of which possess histone acetylase activity. BRCA2 might recruit these histone modifiers to the transcription complex to induce transcriptional activity (Fig.6) [58].

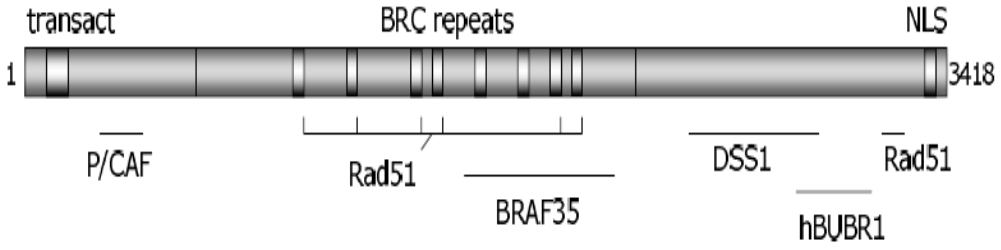


Figure 6. Schematic diagram of *BRCA2* gene.

H. Justification of advanced genomic technology for molecular signature of esophageal cancer:

Cancer is a highly complex disease which can encompass multiple genomic alterations, including point mutations, translocations, gene amplifications, epigenetic modifications, deletions, aberrant splicing, and altered gene expression (Fig. 7). These changes may be inherited or somatically acquired during progression from a normal to a cancerous cell. Progress of phenotypes from normal to advanced carcinoma is controlled by a transcriptional hierarchy that coordinates the action of hundreds of genes (Fig.8). Development of better preventive and diagnostic approaches as well as more effective treatment modalities requires in-depth understanding of molecular mechanisms implicated in the complex process of esophageal carcinogenesis.

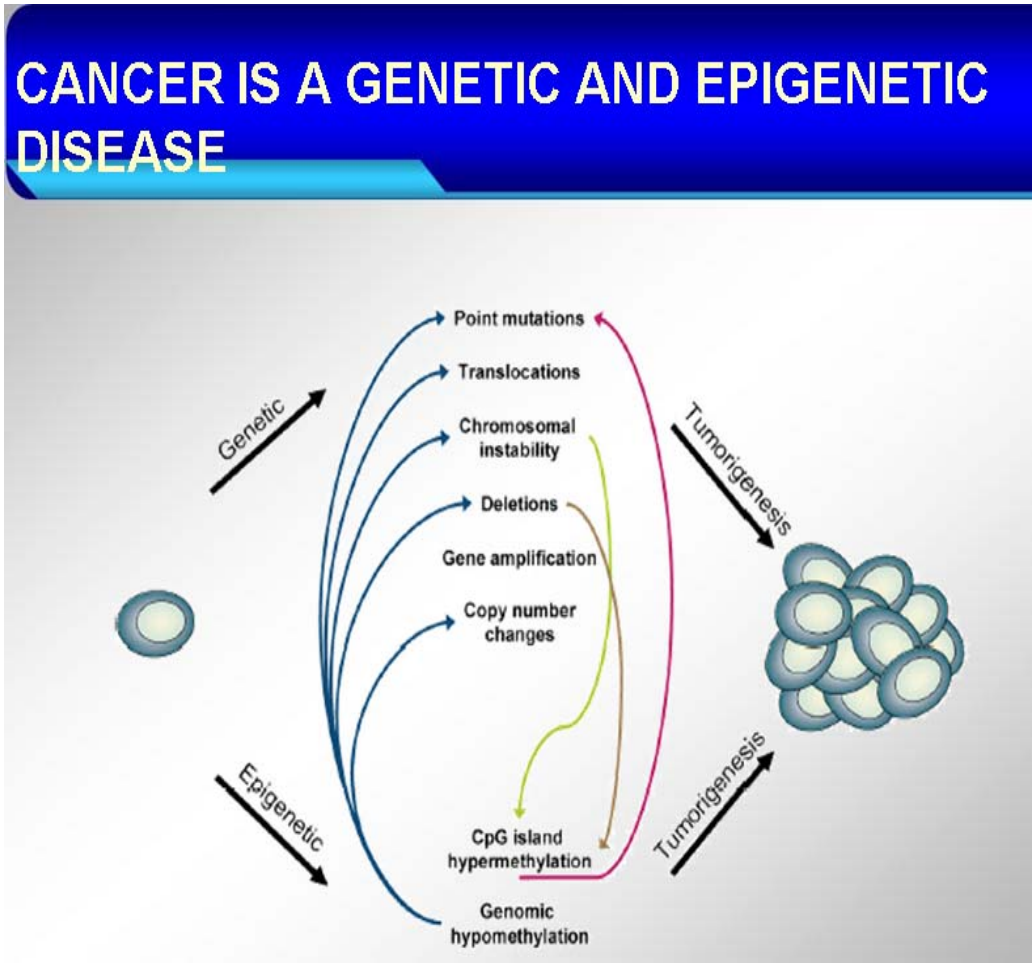


Figure 7: Genetic and epigenetic changes leading to tumorigenesis. (Human Molecular Genetics, 2007, Vol.16, R96-R105)

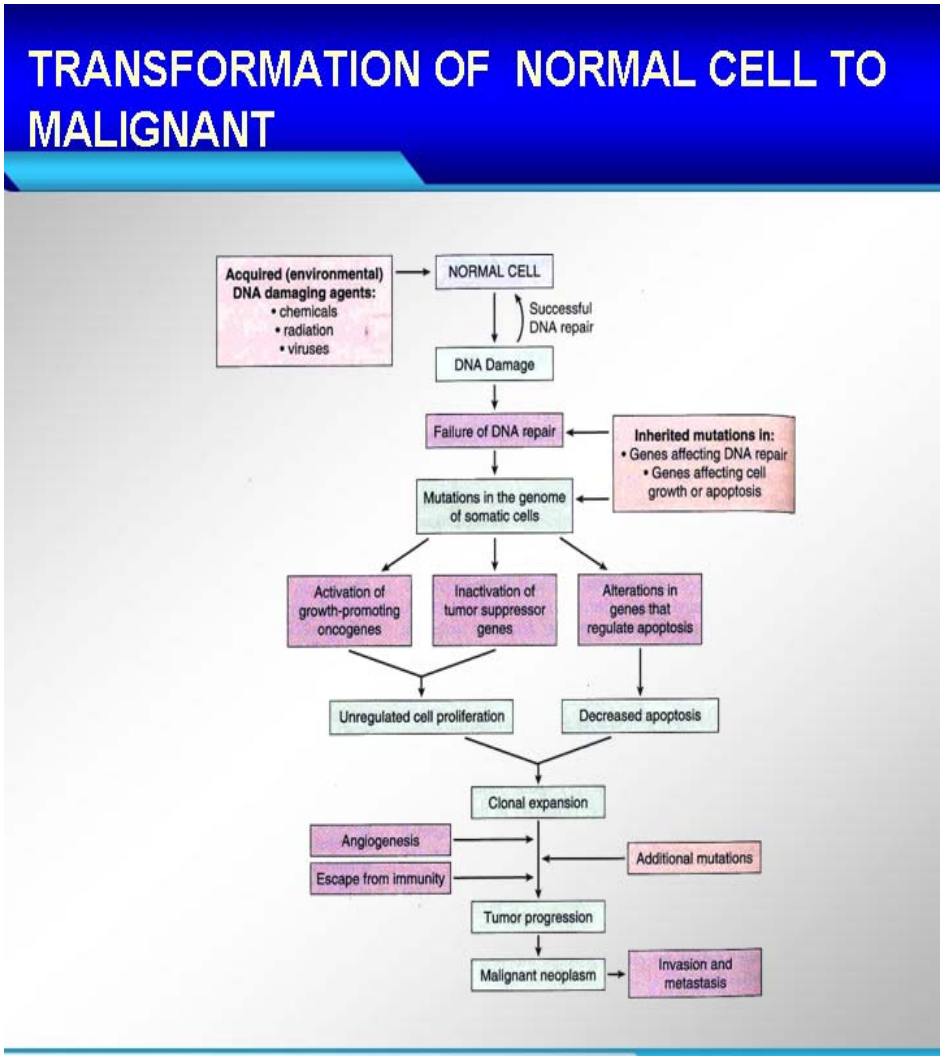


Figure 8: Molecular mechanisms leading to transformation of normal cell to malignant cell.

Conventional approaches investigating one or several candidate genes at a time can not show the whole story of carcinogenesis. The generation of vast amounts of DNA sequence information, coupled with advances in technologies developed for the experimental use of such information, allows the description of biological processes from a view of global genetic perspective. However, little is known about the exact expression

changes in each stage of tumorigenesis, which will help us identify the exact series of events that leads to the initiation and progression of cancer development. As a step toward understanding the complicated changes between normal and malignant cells, this report focused on gene expression profile variations among normal and abnormal esophageal epithelium tissues. Analyzing alterations of gene expression profiles in different stages of neoplasia is necessary for establishing the preventive, diagnostic, therapeutic, and prognostic potential of each related gene. To illustrate the mechanisms controlling malignant changes at molecular level may provide a further understanding of tumorigenesis, as well as new approaches in early detection and treatment of esophageal cancer. One such technology, DNA microarray, permits simultaneous monitoring of thousands of genes [48]. 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 in patient samples [59].

Scope of microarrays (Fig.9):

Gene expression profiling: comparison of the expression level of genes under different conditions such as in health, disease, after therapeutic interventions and following exposure to drugs and radiation.

Gene expression localization: identification of unique genes related to different subcellular organelles and tissues.

Gene function: analysis of gene behavior in relation to specific metabolic pathways such as cell signaling and apoptosis.

Gene characterization: define genes within an organism and compare it to a reference organism.

Single nucleotide polymorphism: detect differences in single nucleotides between genomic samples obtained from similar organisms.

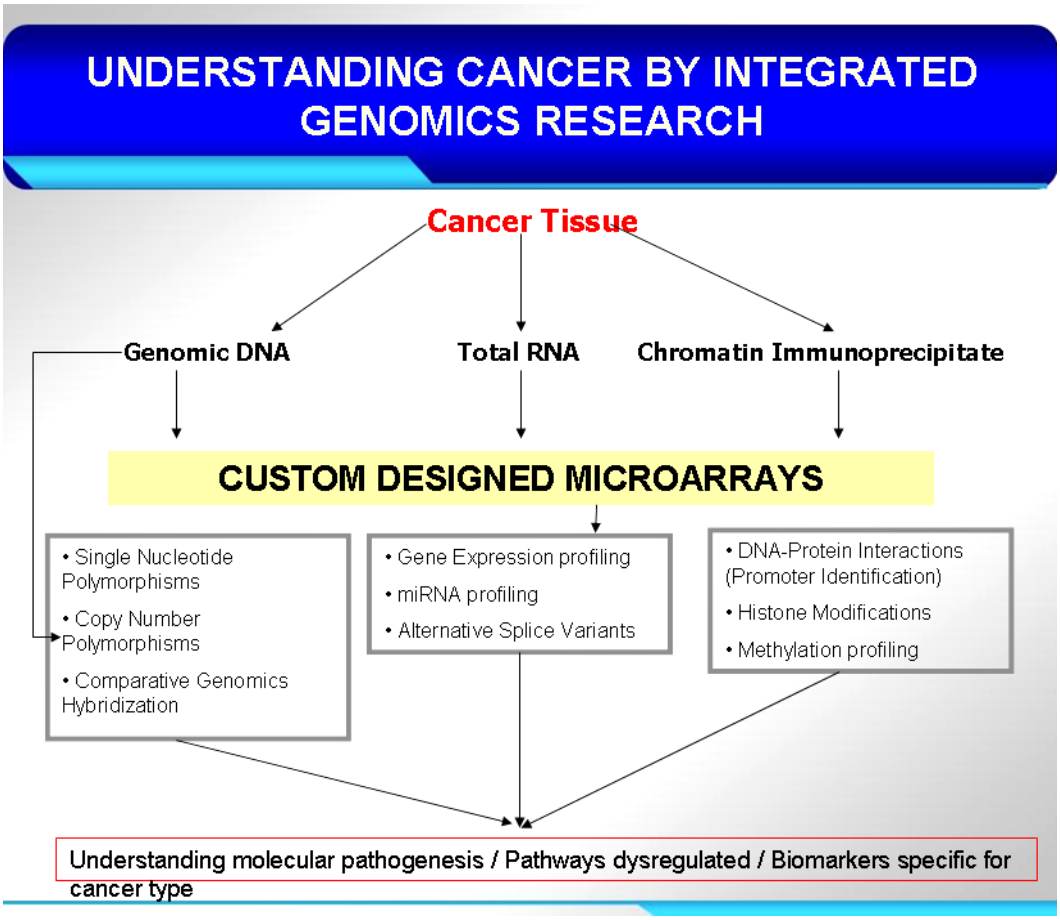


Figure 9: Application of microarrays to identify biomarkers in cancer research

1) Microarray-

In the past several years, numerous systems were developed for the construction of large-scale DNA arrays. All of these platforms are based on cDNAs or oligonucleotides immobilized to a solid support. In the cDNA approach, cDNA (or genomic) clones of interest are arrayed in a multi-well format and amplified by polymerase chain reaction. The products of this amplification, which are usually 500- to 2000-bp clones from the 3' regions of the genes of interest, are then spotted onto solid support by using high-speed robotics. By using this method, microarrays of up to 10 000 clones can be generated by spotting onto a glass substrate. Sample detection for microarrays on glass involves the use of probes labeled with fluorescent or radioactive nucleotides. Fluorescent cDNA probes are generated from control and test RNA samples in single-round reverse-transcription reactions in the presence of fluorescently tagged dUTP (e.g., Cy3-dUTP and Cy5-dUTP), which produces control and test products labeled with different fluors. The cDNAs generated from these two populations, collectively termed the "probe," are then mixed and hybridized to the array under a glass cover slip. The fluorescent signal is detected by using a custom-designed scanning confocal microscope equipped with a motorized stage and lasers for fluor excitation. The data are analyzed with custom digital image analysis software that determines for each DNA feature the ratio of fluor 1 to fluor 2, corrected for local background. The strength of this approach lies in the ability to label RNAs from control and treated samples with different fluorescent nucleotides, allowing for the simultaneous hybridization and detection of both populations on one microarray (Fig.10).

In an alternative approach, large numbers of cDNA clones can be spotted onto a membrane support, albeit at a lower density. This method is useful for expression profiling and large-scale screening and mapping of genomic or cDNA clones. In expression profiling on filter membranes, two different membranes are used simultaneously for control and test RNA hybridizations, or a single membrane is stripped and reprobed. The signal is detected by using radioactive nucleotides and visualized by phosphorimager analysis or autoradiography.

Oligonucleotide microarrays are constructed either by spotting prefabricated oligos on a glass support or by the more elegant method of direct in situ oligo synthesis on the glass surface by photolithography. The strength of this approach lies in its ability to discriminate DNA molecules based on single base-pair difference. This allows the application of this method to the fields of medical diagnostics, pharmacogenetics, and sequencing by hybridization as well as gene-expression analysis. Fabrication of oligonucleotide chips by photolithography is theoretically simple but technically complex. The light from a high-intensity mercury lamp is directed through a photolithographic mask onto the silica surface, resulting in deprotection of the terminal nucleotides in the illuminated regions. The entire chip is then reacted with the desired free nucleotide, resulting in selected chain elongation. This process requires only 4n cycles (where n = oligonucleotide length in bases) to synthesize a vast number of unique oligos, the total number of which is limited only by the complexity of the photolithographic mask and the chip size. Sample preparation involves the generation of double-stranded cDNA from cellular poly (A) + RNA followed by antisense RNA synthesis in an in vitro transcription reaction with biotinylated or fluortagged nucleotides.

The RNA probe is then fragmented to facilitate hybridization. If the indirect visualization method is used, the chips are incubated with fluor-linked streptavidin (e.g., phycoerythrin) after hybridization. The signal is detected with a custom confocal scanner [60].

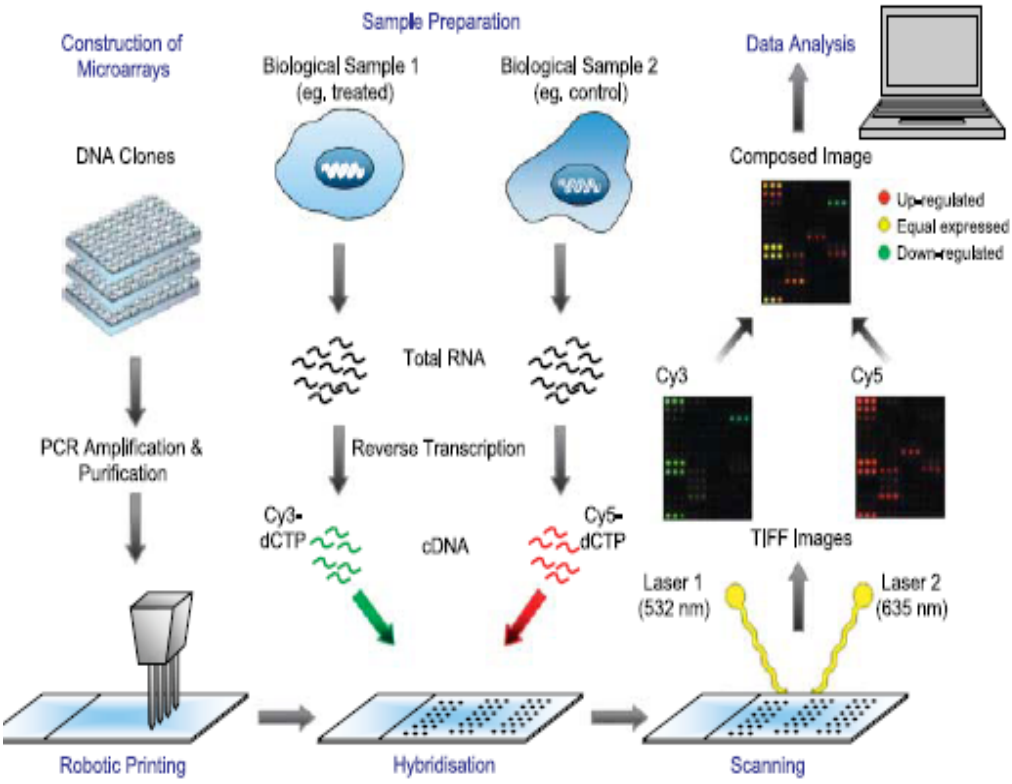


Figure 10: Dual color fluorescence cDNA microarray analysis. (Ref: 60)

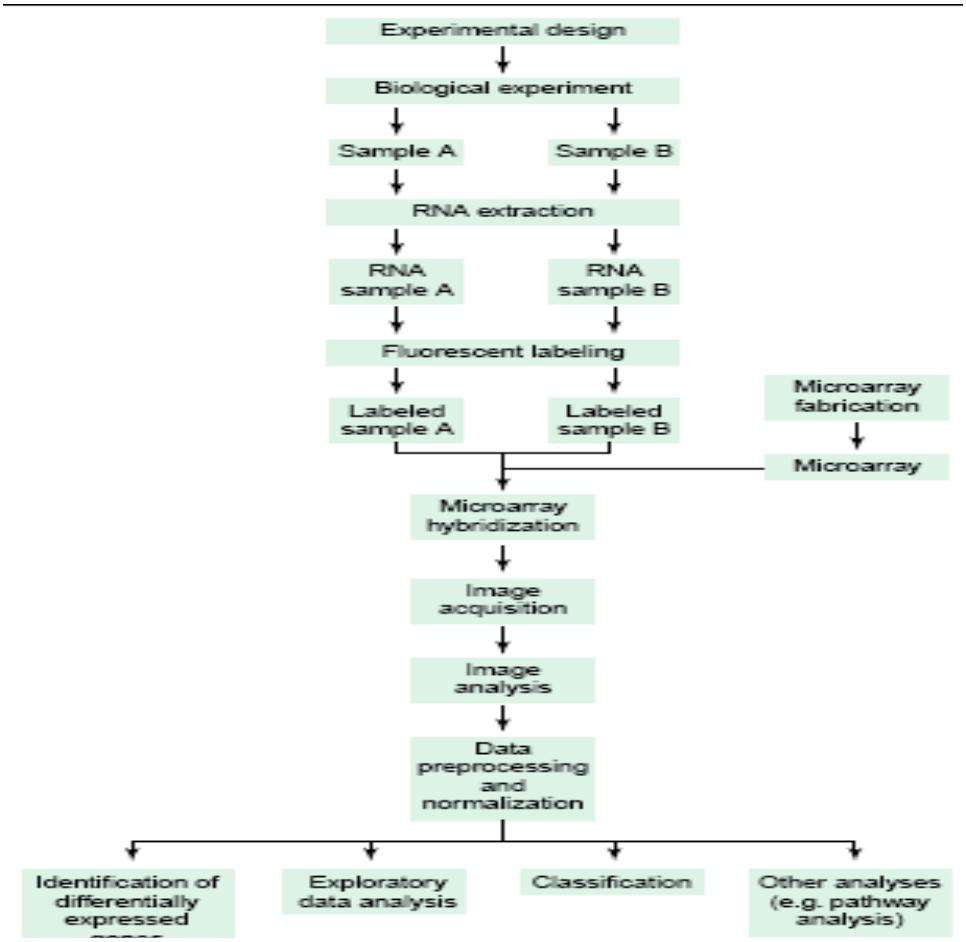


Figure 11: Flow of a typical microarray experiment.

A typical microarray experiment begins with good experimental design. After carrying out the biological experiment, the samples, either tissues from patient or animal model, or cells from in vitro cultures, are collected. Their RNAs are then extracted and labeled with different fluorescent dyes, and co-hybridized to a microarray. The hybridized microarray is scanned to acquire the fluorescent images. Image analysis is performed to obtain the raw signal data for every spot. Poor quality data are filtered out and the remaining high quality data are normalized. Finally depending on the aim of the study, one can infer statistical significance of differential expression, perform various

exploratory data analyses, classify samples according to their disease subtypes and carry out pathway analysis. Note that data from all the steps should be collected according to certain standards, minimum information about a microarray experiment (e.g. MIAME), and archived properly (Fig.11) [61, 62, 63, 64].

2) Real-Time PCR:

Because of the power of real-time PCR applications, it is already used in many different fields within biomedical research and molecular diagnostics. Because of the reliability of real-time PCR, many researchers use the relative or absolute quantitation of gene expression to validate and corroborate the results of printed DNA microarrays or oligonucleotide arrays (e.g., Affymetrix GeneChip). Arrays are used because they allow a researcher to look in an "unbiased" fashion at how experimental manipulation might affect any of the thousands of genes present on the array. Some arrays purport to contain the entire "genome" of a model organism and thus can theoretically be probed to comprehensively determine changes in expression within the entire "transcriptome." The problem is that there can be artifacts, and it is often difficult to get reliable quantitative data or adequate statistical power with present array technology. Thus many researchers choose real-time PCR as a supporting technique to validate and better quantitate the most interesting candidate genes from their arrays. The key to real-time PCR is the ability to monitor the progress of DNA amplification in real time.

This is accomplished by specific chemistries and instrumentation. Generally, chemistries consist of special fluorescent probes in the PCR. Several types of probes exist, including DNA-binding dyes like EtBr or SYBR green I, hydrolysis probes (5'-nuclease probes), and 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.

SYBR green I binds to the minor groove of dsDNA, emitting 1,000-fold greater fluorescence than when it is free in solution. Therefore, the greater the amounts 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. 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.

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. These are likely the most widely used fluorogenic probe format 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 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. Hydrolysis probes afford similar precision as SYBR green I, 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. There are several other variations on the reporter-quencher theme, including molecular beacons, sunrise primers, and scorpion primers. They each seek to keep the reporter and quencher together before amplification while separating them and generating the fluorescence signal during amplification. Another class, called hybridization probes, uses donor and acceptor fluorophores, whereas PNAs containing thiazole orange fluorophores (called light-up probes) also emit greater signal upon binding of DNA. These do not represent an exhaustive list, as much other specific and nonspecific chemistry exist.

To evaluate gene expression, RNA must first be isolated from the samples to be studied. After isolation, RNA is linearly converted to cDNA, which is used for real-time PCR. Amplification curves are graphed by the software to help determine the "cycle time" at which fluorescence reaches a threshold level (CT). This CT value is inversely proportional to the amount of specific nucleic acid sequence in the original sample. Both relative and absolute quantitations of gene expression utilize the CT value to quantitate cDNA and thereby determine gene expression. In a perfectly efficient PCR, the amount of amplified product doubles each cycle. Therefore, a difference of 1 between sample CTs means that the sample with the lower CT value had double the target sequence of the

other sample; a change in CT of 2 means a fourfold difference; a change in CT of 3 means an eightfold difference, and so on. Relative quantitation measures changes in the steady-state levels of a gene of interest relative to an invariate control gene Housekeeping genes (e.g., cyclophilin, glyceraldehyde-3-phosphate dehydrogenase, ribosomal protein, β-actin, 18S rRNA, transferrin receptor, etc.) that are not expected to change under the experimental conditions serve as a convenient internal standard. Because the absolute quantity of the internal standard is not known, only relative changes can be determined by this method. This may not pose a problem for most research projects because the fold change may be informative irrespective of the absolute value. The limitations of this approach include a lack of absolute quantitation and the necessity for unchanging housekeeping genes as internal standards. Absolute quantitation attempts a more ambitious task to measure the actual nucleic acid copy number in a given sample. This requires a sample of known quantity (copy number) of the gene of interest that can be diluted to generate a standard curve. This is an external "absolute" standard. Unknown samples are compared with the standard curve for absolute quantitation. The primary limitation to this approach is the necessity of obtaining an independent reliable standard for each gene to be analyzed and then running concurrent standard curves during each assay [65, 66, 67, 68].

3) Tissue microarray technology (TMA):

Tissue microarray technology allows rapid visualization of molecular targets in thousands of tissue specimens at a time either at the DNA, RNA or protein level. The technique facilitates rapid translation of molecular discoveries to clinical applications. By revealing the cellular localization, prevalence and clinical significance of candidate

genes, TMAs are ideally suitable for genomics-based diagnostic and drug target discovery. The speed of molecular analysis is increased by more than 100 fold, precious tissues are not destroyed and a very large number of molecular targets can be analyzed from consecutive TMA sections. Construction of TMAs is achieved by acquiring cylindrical core specimens from up to 1000 fixed and paraffin-embedded tissue specimens and arraying them at high density into a recipient TMA block. Tissue microarrays (TMA) represent one of the most useful approaches in the microarray field. It consists of dozens to hundreds of predefined microscopic elements of tissue arrayed on a glass slide, and permits the simultaneous analysis of protein, RNA or DNA expression in multiple tissue samples. A particular advantage of this method is that same-sample and new-sample analysis can be carried out using the same TMA [69].

4) Application of single nucleotide polymorphism array in cancer

Cancer arises from the accumulation of inherited polymorphism (i.e. SNPs) and mutation and/or sporadic somatic polymorphism (i.e. non-germline polymorphism) in cell cycle, DNA repair, and growth signaling genes. Neoplastic progression is generally characterized by the accumulation of multiple genetic alterations including loss of tumor suppression gene function. Loss of heterozygosity (LOH) has been used to identify genomic regions that harbor tumor suppressor genes and to characterize different tumor types, pathological stages and progression. Loss of heterozygosity (LOH) refers to change from a state of heterozygosity in a normal genome to a homozygous state in a paired tumor genome. LOH is most often regarded as a mechanism for disabling tumor suppressor genes (TSGs) during the course of oncogenesis. Although LOH is often thought to result from copy-loss events such as hemizygous deletions, a large proportion

of LOH results from copy neutral events such as chromosomal duplications. LOH pattern has been detected by allelotyping using restriction fragment length polymorphism, and later by simple sequence length polymorphisms (SSLPs or microsatellite) for 10 years.

LOH is caused by a variety of genetic mechanisms, including physical deletion of chromosome nondisjunction, mitotic nondisjunction followed by reduplication of the remaining chromosomes, mitotic recombination, and gene conversion. The mechanisms of LOH are remarkably chromosome-specific. Some chromosomes display complete loss. However, more than half of the losses are associated with the loss of only a part of the chromosome rather than the whole chromosome. LOH is also a common form of allelic imbalance and the detection of LOH has been used to identify genomic regions that harbor tumor suppressor genes and to characterize different tumor types, pathological stages, and progression. Global patterns of LOH can be analyzed through allelotyping of tumors with polymorphic genetic markers from each chromosomal arm.

Two allele RFLPs and Southern analysis give way to simple sequence length polymorphisms such as PCR-based microsatellite, and both have been proved to be reliable genetic markers for studying LOH. RFLP markers have low heterozygosity rates and are available in small number, gel-based microsatellite assay is difficult to automate and not readily scalable. Microsatellite markers are reliable genetic markers for studying LOH, but only a modest number of SSLPs are used in LOH studies because the genotyping procedure is rather tedious and difficult to automate and are not readily scalable.

As a result, most genome-wide scans for LOH have been conducted at low resolution with a relatively small number of polymorphic markers. Previous allelotyping

analysis of cancer by many groups was restricted to particular chromosomal regions or arms, or else used a relatively low density of markers. For example, an average of 120 microsatellites has been used to determine the allelotype of multiple different human neoplasms in a series of studies since 1995, and the highest density microsatellite allelotype is ~280 polymorphic markers before the year 2000. SNPs which are less polymorphic than microsatellite markers are the most common form of sequence variation in human genome, occurring approximately in every 1 in 200 bp. SNPs may occur at more than 2 million sites in the genome, thus making it possible to place SNPs at high density along the genome. High-throughput polymorphism detection technologies hold great promise for the characterization of complex diseases including cancer. High-density mapping of genetic losses reveals potential tumor suppressor loci and might be useful in the clinical classification of individual tumors. SNP array has been introduced recently for genome-wide screening of chromosome imbalance. Higher density SNP array can be used effectively to detect small regions of chromosomal changes and provide more information regarding the boundaries of loss regions. In addition, more markers increase confidence in a detected event. If multiple adjacent SNPs show a consistent change, the confidence in the call is much higher than when it is based on a single SNP. The density, distribution, and allele specificity of SNPs makes them attractive for high-resolution analyses of LOH and copy number alterations in cancer genomes.

The Affymetrix 10K SNP array (the second generation) contains 11 560 SNP alleles with high frequency of heterozygosity (average 36% based on Affymetrix in-house data). The Affymetrix 100K SNP array, a new SNP array platform, provides a high accuracy (99.5%), reproducibility (91.1%) and a high call (heterozygous or homozygous)

rate (95%). The average accuracy is calculated as 81% at 95% significance with a median inter-SNP distance of 105 kb in osteosarcoma using 10K array. In summary, with the increasing number of SNPs available and technical progression, it is possible to probe the entire genome, and specific regions at much higher resolution. SNP array hybridization is an accurate and efficient method for evaluating genome-wide tumor LOH at present. Genome wide detection of loss of heterozygosity (LOH), as well as copy-number (CN) alterations in cancer genomes, has drawn recent attention in the field of cancer genetics, because LOH has been closely related to the pathogenesis of cancers, in that it is a common mechanism for inactivation of tumor suppressor genes in Knudson's paradigm. The simultaneous measurement of DNA copy number changes and loss of heterozygosity events by SNP arrays should strengthen our ability to discover cancer-causing genes and to refine cancer diagnosis. DNA copy number changes, such as amplifications and deletions, frequently cause oncogene activation and tumor suppressor gene inactivation in cancer. The high density of SNP arrays may also make possible the characterization of haplotype structures to analyze cancer predisposition. Furthermore, the detection of single-copy changes with SNP arrays suggest that these arrays could be used to study other genetic diseases in addition to cancers, such as Down, Prader Willi, Angelman, and cri du chat syndromes. The SNP arrays may find application as diagnostic as well as research reagents in this area. In conclusion, SNP array hybridization is a highly efficient method for evaluating genome-wide copy number changes. Application of the SNP array approach to large cancer data sets should prove highly fruitful in discovering cancer-specific genomic alterations. Chromosomal or gene copy number alterations are one of

the main important mechanisms that perturb normal gene function by inducing changes reflected in gene expression [70, 71].

5) Automated DNA Sequencing

DNA polymerases copy single-stranded DNA templates, by adding nucleotides to a growing chain (extension product). Chain elongation occurs at the 3′ end of a primer, an oligonucleotide that anneals to the template. The deoxynucleotide added to the extension product is selected by base-pair matching to the template. The extension product grows by the formation of a phosphodiester bridge between the 3′-hydroxyl group at the growing end of the primer and the 5′-phosphate group of the incoming deoxynucleotide. The growth is in the 5′→3′ direction. DNA polymerases can also incorporate analogues of nucleotide bases. The dideoxy method of DNA sequencing developed by Sanger *et al.* (1977) takes advantage of this ability by using 2′, 3′-dideoxynucleotides as substrates. When a dideoxynucleotide is incorporated at the 3′ end of the growing chain, chain elongation is terminated selectively at A, C, G, or T because the chain lacks a 3′-hydroxyl group.

In the Applied Biosystems strategy for automated fluorescent sequencing, fluorescent dye labels are incorporated into DNA extension products using 5′-dye labeled primers (dye primers) or 3′-dye labeled dideoxynucleotide triphosphates (dye terminators). The most appropriate labeling method to use depends on your sequencing objectives, the performance characteristics of each method, and on personal preference. Applied Biosystems DNA sequencers detect fluorescence from four different dyes that are used to identify the A, C, G, and T extension reactions. Each dye emits light at a

different wavelength when excited by an argon ion laser. All four colors and therefore all four bases can be detected and distinguished in a single gel lane or capillary injection.

AmpliTaq® DNA polymerase, FS is the sequencing enzyme used in ABI PRISM cycle sequencing kits. It is a mutant form of *Thermus aquaticus* (Taq) DNA polymerase and contains a point mutation in the active site, replacing phenylalanine with tyrosine at residue 667 (F667Y). This mutation results in less discrimination against dideoxynucleotides, and leads to a much more even peak intensity pattern. AmpliTaq DNA Polymerase, FS also contains a point mutation in the amino terminal domain, replacing glycine with aspartate at residue 46 (G46D), which removes almost all of the 5´ →3´ nuclease activity. This eliminates artifacts that arise from the exonuclease activity. The enzyme has been formulated with a thermally stable inorganic pyrophosphatase that cleaves the inorganic pyrophosphate (PPi) byproduct of the extension reaction and prevents its accumulation in the sequencing reaction. In the presence of high concentrations of PPi, the polymerization reaction can be reversed, a reaction called pyrophosphorolysis. In this reaction, a nucleoside monophosphate is removed from the extension product with the addition of PPi to form the nucleoside triphsphate.

The ABI PRISM™ Dye Terminator Cycle Sequencing Kits combine AmpliTaq® DNAPolymerase, FS, rhodamine dye terminators, and all the required components for the sequencing reaction. The concentrations of the dye-labeled dideoxynucleotides and deoxynucleotides in the dNTP mix have been optimized to give a balanced distribution of signal above 700 bases. The dNTP mix includes dITP in place of dGTP to minimize band compressions. In the Ready Reaction format, the dye terminators, deoxynucleoside triphosphates, AmpliTaq DNA Polymerase, FS, r*Tth* pyrophosphatase, magnesium

chloride, and buffer are premixed into a single tube of Ready Reaction Mix and are ready to use.These reagents are suitable for performing fluorescence-based cycle sequencing reactions on single-stranded or double-stranded DNA templates, or on polymerase chain reaction (PCR) fragments. Applied Biosystems has developed a set of dye terminators labeled with novel, high-sensitivity dyes. The new dye structures contain a fluorescein donor dye, *e.g.*, 6-carboxyfluorescein (6-FAM), linked to one of four dichlororhodamine (dRhodamine) acceptor dyes. The excitation maximum of each dye label is that of the fluorescein donor, and the emission spectrum is that of the dRhodamine acceptor. The donor dye is optimized to absorb the excitation energy of the argon ion laser in the Applied Biosystems DNA sequencing instruments. The linker affords extremely efficient energy transfer (quantum efficiency nearly 1.0, *i.e.*, 100%) between the donor and acceptor dyes. The BigDye™ terminators are 2–3 times brighter than the rhodamine dye terminators when incorporated into cycle sequencing products.

The BigDye terminators are labeled with the following dRhodamine acceptor dyes:

Terminator	**Acceptor Dye**
A	dichloro[R6G]
C	dichloro[ROX]
G	dichloro[R110]
T	dichloro[TAMRA]

The BigDye terminators also have narrower emission spectra than the rhodamine dye terminators, giving less spectral overlap and therefore less noise. The ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kits combine AmpliTaq DNA Polymerase, FS, the new BigDye terminators, and all the required components for the sequencing reaction. In the Ready Reaction format, the dye terminators, deoxynucleoside triphosphates, AmpliTaq DNA Polymerase, FS, r*Tth* pyrophosphatase, magnesium chloride, and buffer are premixed into a single tube of Ready Reaction Mix and are ready to use. These reagents are suitable for performing fluorescence-based cycle sequencing reactions on single-stranded or double-stranded DNA templates, on polymerase chain reaction (PCR) fragments, and on large templates, *e.g.*, BAC clones. The dNTP mix includes dITP in place of dGTP to minimize band compressions. The dNTP mix also uses dUTP in place of dTTP. dUTP improves the incorporation of the T terminator and results in a better T pattern (www.appliedbiosystems.com).

Chapter 4

Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma of high-risk area in India by SNP array

Introduction:

Cancer is the result of a series of genetic or epigenetic changes, including aneuploidy, multiple gene amplifications, deletions and translocations. These genetic instabilities are caused by either inherited mutations in genes that monitor genome integrity or mutations that are acquired in somatic cells during tumor development. Environmental risk factors and individual cancer genetic susceptibilities could contribute to tumor development and progression by facilitating the inactivation or loss of tumor suppressor genes and by favoring the activation or amplification of oncogenes [72]. Apart from numeric aberrations, such as gene amplification or deletion, which are frequently detected in carcinomas, other genomic changes in carcinogenesis, such as loss of heterozygosity (LOH), do not necessarily lead to DNA copy-number gain or loss. Deletion, mitotic non-disjunction, reduplication, mitotic recombination and gene conversion may lead to LOH. The two-hit model describes a tumor suppressor gene inactivation due to a loss of one allele and mutation of the other allele, resulting in an LOH [73]. Comprehensive analysis of genetic alterations (loss of heterozygosity and copy number alterations, either gain or loss) in tumors and identification of genes involved in tumorigenesis has been a major focus of cancer research [72]. LOH is frequently observed in a variety of human cancers, and regions with frequent LOH may contain tumor suppressor genes. Thus, detection of LOH will likely remain a cornerstone for predicting tumor aggressiveness for many human tumors [74].

The development of esophageal cancer is a leading example in which environmental carcinogens in addition to geographic and genetic factors appear to play major etiologic roles [1]. Esophageal cancer occurs at very high frequencies in certain

parts of China, Iran, South Africa, Uruguay, France, Italy and in some regions of India [1, 5]. The highest incidence of this cancer in India has been reported from Assam (AAR of 33/100,000) in the North-east region where it is the second leading cancer in men and third leading cancer in women [9]. Tobacco smoking, betel quid chewing, and alcohol consumption are the major known risk factors for esophageal cancer [8]. Chewing of fermented areca nut with or without tobacco has been shown to be independently associated with the development of esophageal cancer in Assam region of NE India [9].

Esophageal squamous cell carcinoma (ESCC) develops through a multistep process from dysplasia, through carcinoma in situ to invasive carcinoma, and the acquisition of genetic alterations is tightly related to the dysplasia-carcinoma sequence. The characterization of genetic alterations inherently linked to ESCC development and an in-depth under standing of the molecular mechanisms underlying carcinogenesis and growth control may therefore provide information relevant for early tumor detection, refined prognosis and development of novel targeted therapeutics [46].

LOH studies and conventional comparative genomic hybridization (CGH) analyses have demonstrated genetic complexity in ESCC and have identified multiple recurrent copy number alterations, including gains of 1q, 2q, 3q, 5p, 7p, 7q, 8q, 11q, 12p, 12q, 14q, 17q, 20p and 20q. Amplifications of regions harboring oncogenes such as 7p12 (*EGFR*), 8q24 (MYC), 17q21 (*FGFR*) and 11q13 (*CCND1, FGF4/3, and EMS1*) have consistently been observed. Losses, albeit at a lower frequency than gains, have recurrently involved 3p, 5q, 9p, 13q, 18q and 21q and include target genes such as *FHIT, APC, RB1* and *CDKN2A*. Some of the changes identified, such as gain of 8q24, 11q13, 12p and 20q12 and loss of 3p have been associated with poor prognosis. However, conventional genetic

alterations and biological characteristics have so far had a limited impact on clinical prognostication and treatment [46]. In another study, frequent deletions were found on chromosome arms 1p, 3p, 4p, 5q, 8p, 9p, 9q, 11q, 13q, 16p, 17p, 18q, 19p, and 19q in ESCC [47]. Deletion of chromosome 3p is one of the most frequent allelic imbalances in ESCC detected by comparative genomic hybridization (CGH), loss of heterozygosity (LOH) and genome-wide genotyping [75].

Recently, the discovery of large-scale genome-wide copy number variation has stimulated interest in elucidating the role of copy number alterations (CNA) in the development of malignancy. With the development of comparative genomic hybridization (CGH) arrays using more than 30,000 BAC clones spanning the human genome, and high-density single-nucleotide polymorphism (SNP) microarrays designed to genotype more than 100,000 SNPs in the human genome DNA, the resolution of the whole genome scanning technique has increased considerably and allowed accurate and reproducible determination of copy number changes in the cancer genome [76]. CGH array has a limited resolution within a range of 1-20Mb. The SNP array technology, which was originally developed for allelotyping and linkage analyses, allows a genome-wide fine mapping of copy-number changes within a range of 30-900kb. In addition, the genotypes of the SNPs provide information about LOH throughout the genome [73]. The 10 K SNP array (GeneChip Mapping 10 K array, Affymetrix) offers a high-resolution genomic approach to screen chromosomal alterations systematically. Several studies on allelic imbalance or loss in cancers and cancer cell lines using the 10 K SNP array have been published [74]. These arrays provide clear technical advantages including robust single-primer assay methodology, accurate and reproducible genotyping, and copy-

neutral LOH detection compared with array-CGH, karyotyping, and other oligonucleotide CGH arrays. Gene copy number changes are one of the main mechanisms causing cancerous alterations in gene expression [42].

Genome-wide analysis of genetic alterations that may occur in esophageal cancer in a high incidence region of India where tobacco use and alcohol consumption are widespread and the users of these two substances are also betel quid chewers, has so far not been investigated. In the present study, GeneChip® Human Mapping 10K Array Xba 142 2.0 was applied to determine common aberrations (deletion or amplification) and copy number alterations in patients with esophageal cancer from a high-risk region of India.

Materials and Methods:

Selection of Patients and Collection of Samples:

Endoscopic biopsy specimens from tumor and matched normal tissue distant to tumor were collected from twenty patients during diagnostic endoscopy at Dr. Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam. Part of tumor and normal tissue was preserved in formalin for histopathologic diagnosis/confirmation and remaining parts were immediately immersed in RNA later solution (Ambion, Austin, USA) and stored at -70ºC until processed. 5 ml of blood was also collected into EDTA vials from all 20 patients and frozen as a source of normal germ line DNA. Informed consent was obtained from all the patients to use their surgical specimens and clinicopathologic data for research purposes. Institutional Human Ethics Committee had approved the study. The twenty patients' demographic and lifestyle cancer risk factors (e.g., smoking, chewing, and alcohol drinking) and clinical data were collected (Table 1).

Table 1: Demographic and lifestyle cancer risk factors of 20 patients:

Sample ID	Age (years)	Sex	Histological Grade	Tobacco Chewing habit	Smoking habit	Alcohol habit	Betel quid chewing habit
134T	55	F	G1	No	Yes	No	Yes
147T	55	M	G3	Yes	Yes	No	Yes
211T	35	M	G2	No	Yes	No	Yes
213T	52	M	G3	Yes	No	Yes	Yes
215T	70	M	G1	No	Yes	Yes	Yes
259T	70	M	G1	No	Yes	No	Yes
261T	32	M	G1	Yes	Yes	Yes	Yes
270T	65	F	G1	Yes	No	No	Yes
272T	60	M	G3	No	Yes	No	Yes
274T	55	M	G2	No	Yes	No	Yes
275T	54	M	G2	No	No	Yes	Yes
276T	42	M	G2	Yes	No	No	Yes
478T	61	F	G3	No	No	No	Yes
278T	50	M	G1	No	No	No	Yes
289T	55	M	G2	No	No	No	Yes
139T	70	M	G1	Yes	Yes	No	Yes
140T	53	M	G2	Yes	Yes	Yes	Yes
277T	60	M	G3	Yes	No	No	Yes
302T	75	F	G2	No	No	No	Yes
318T	50	M	G3	Yes	No	No	Yes

G1=well differentiated, **G2** = moderately differentiated, **G3**= poorly differentiated.

Genomic DNA Extraction:

Germ-line DNA and tumor tissue DNA was extracted using the Qiagen QIamp DNA Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instruction. Quantification was assessed from 1µl of sample using ND1000 spectrophotometer (NanoDrop, Wilmington, DE) (Table 2). DNA quality of all samples was assessed by the ratios of absorbance at 260 nm and 280 nm and by agarose gel electrophoresis (0.7%). DNA samples were normalized at 50ng/µl concentration using reduced TE buffer (10 mmol/L Tris-HCl, pH8 and 0.1mmol/L EDTA).

Table 2: ESTIMATION of DNA concentration by NANODROP SPECTROPHOTOMETER after DNA Cleanup:-

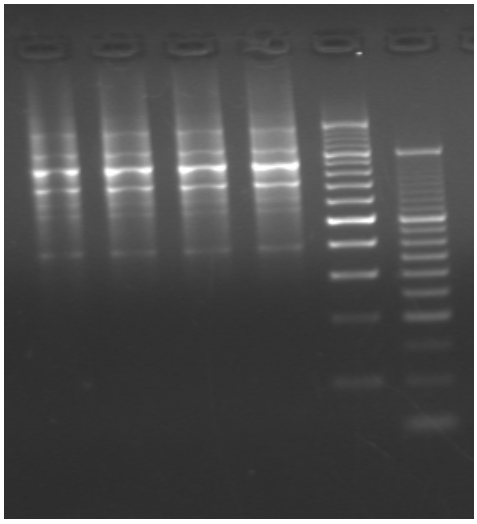
Sample ID	ng/µl	260/280	260/230
261B	176.67	1.88	2.25
213B	87.81	1.87	2.20
270B	97.58	1.86	2.17
274B	111.76	1.89	2.07
201B	194.94	1.88	1.90
269B	104.85	1.90	2.11
161B	87.66	1.83	1.88
211-B-C	128.76	1.9	2.25
215-B-C	124.58	1.88	2.21

Sample ID	**ng/µl**	**260/280**	**260/230**
211-T	139.24	1.94	1.99
215-T	119.09	1.94	2.19
259-T	108.59	2.07	2.03
261-T	135.83	2.05	2.1
270-T	143.76	1.91	2.01
272-T	122.94	1.96	2.22
274-T	217.69	1.89	1.63
275-T	166.58	1.96	1.61
276-T	110.18	1.93	1.68
134-T	159.37	1.88	2.07
147-T	254.78	1.89	1.44
213T	116.93	1.82	2.57

Affymetrix GeneChip® Human Mapping 10K Array Xba 142 2.0:

Further sample processing, including digestion, adaptor ligation, amplification, fragmentation, labeling, hybridization, washing and scanning was assayed according to the standard protocol (Affymetrix GeneChip Mapping 10K 2.0 Assay Manual). Briefly, 250 ng of germ-line and tumor DNA was digested with 20 units of XbaI restriction enzyme and maintained at 4°C. The 4-bp XbaI overhangs were ligated to an Adaptor Xba fragment by T4DNA Ligase and stored at -20˚C until PCR amplification.

For amplification, diluted adaptor-ligated DNA was used as template and universal PCR primer (PCR primer Xba 10μM, Affymetrix) complementary to the Adaptor Xba fragment was used along with dNTPs (2.5mM each) and 5U/μl of AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA). For each sample, four to five reactions were needed. PCR reaction was conducted using a Gene Amp PCR System 9700 Program (Applied Biosystem) using denaturation at 95˚C for 3 minutes, followed by 35 cycles at 95˚C for 30 seconds, 59˚C for 30 seconds, and 72˚C for 30 seconds, followed by final extension at 72˚C for 7 minutes. The PCR products were purified according to the Qiagen manual (QIAquick PCR Purification Kit Protocol, Qiagen, Darmstadt, Germany) except that all DNA elutes from the four to five PCR reactions were collected in one tube. After purification from free primers and nucleotides, PCR products were checked by agarose gel (2%) electrophoresis (Fig.12) and quantified using the ND1000 spectrophotometer (Table 3).



DNA Ladder: 100bp & 50bp (MBI Fermentas)

PCR product range should be 1000bp-400bp

Figure 12: **2% Agarose gel showing the PCR product of tumor samples**

Table 3: Quantification of Purified PCR Product Using Nanodrop Spectrophotometer Analysis:-

A. Controls

Sample ID	Concentration (ng/ul)	260/280	260/230	20µg =µl of PCR Product	EB Buffer (µl)
C2-213	823.27	1.9	1.88	24.29	20.71
C3-270	896.9	1.93	1.98	22.29	22.71
C4-274	786.72	1.93	2.22	25.42	19.58
C5-201	558.99	1.93	2.6	35.78	9.22
C6-269	907.62	1.93	2.22	22.04	22.96
C7-161	771.44	1.92	2.34	25.93	19.07
C1-261	646.46	1.92	2.45	30.94	14.06
211-B-C	474.02	1.91	2.44	42.19	2.81
215-B-C	461.85	1.92	2.44	43.30	1.7

B. Tumor samples:

Sample ID	Concentration (ng/ul)	260/280	260/230	20µg =µl of PCR Product	EB Buffer (µl)
211-T	720.98	1.93	2.46	27.74	17.26
215-T	714.13	1.93	2.42	28	17
259-T	731.19	1.95	2.45	27.33	17.67
261-T	578.83	1.90	2.44	34.55	10.45
270-T	798.61	1.93	2.40	25.04	19.96
272-T	709.25	1.94	2.44	28.19	16.81
274-T	805.51	1.93	2.44	24.83	20.17
275-T	692.47	1.94	2.47	28.88	16.12
276-T	540.06	1.92	2.40	37.03	7.97
134-T	763.32	1.95	2.45	26.20	18.8
147-T	801.18	1.95	2.43	24.96	20.04
213T	915.16	1.93	2.46	21.85	23.15

Purified PCR product (20 µg in 45µl volume) was fragmentized with DNase. The fragmented PCR products were labeled with GeneChip DNA labeling reagent (Affymetrix) by adding 19.4 µl of labeling master mix including 30 units of Terminal Deoxynucleotidyl Transferase (TdT) with 50.6µl of fragmented DNA to make 70µl reaction volume in 0.2 ml PCR tube. The sample was incubated at 37˚C for 2 hours, followed by heat deactivation at 95˚C for 15 minutes and maintained at 4°C. This 70µl of labeled DNA was mixed with 190µl of hybridization cocktail in a 1.5 ml eppendrof tube. The samples were denatured at 95˚C for 10 minutes, then the tubes were transferred to ice for exactly 10 seconds. The target DNA was placed on heating block at 48˚C for 2 minutes, after which 80µl of the denatured hybridization mix was injected into the 10K 2.0 Mapping Array (Affymetrix, Santa Clara, CA, USA). This array contains complementary probes of 10,204 biallelic SNPs, which are located within the amplified 250-1000 base XbaI fragments. For each probe, both the sense and antisense strand in both a perfect match and a mismatch sequence are synthesized on the chip. Hybridization

was done at 48˚C for 16-18 hours at 60 rpm in the hybridization oven. The arrays were washed and stained by incubation with streptavidin, then biotinylated antistreptavidin, followed by streptavidin-R-phycoerythrin conjugates using the Affymetrix Fluidics Station 450. Finally, the microarrays were scanned in the Affymetrix GeneChip Scanner 3000 using GeneChip Operating System 1.1.1 with Patch 5 or higher (GCOS, Affymetrix).

Data Analysis:

CEL files, containing intensity value and standard deviation for each probe on the chip were generated for each array using the GeneChip Operating Software (Affymetrix). Perfect match-mismatch average difference intensities as a signal to noise value and the genotypes (AA, AB and BB) were calculated for each biallelic SNP using the GeneChip DNA Analysis Software (GDAS) and GTYPE v4.0 (Affymetrix). "Signal Detection Rate" is the percentage of SNPs that passed the discrimination filter. "Call Rate" is the percentage of SNPs called on the array. Genotype calls were defined as AA, AB, or BB; "no call" means the SNP for that sample did not pass the discrimination filter and was excluded from further evaluation in the present study. LOH was defined in a traditional manner as a change in genotyping call from heterozygosity (AB) in the germ-line DNA, to homozygosity (AA or BB) in the matched tumor DNA (all calls from GDAS). The identification of genomic areas showing copy loss was dependent on sampling of pure tumor DNA without contamination of normal epithelium, stroma and/or inflammatory cells. Each array allowed analysis of 10,204 SNPs, distributed evenly across the genome with a mean interval of 258 kb and median interval of 113 kb. Each SNP on the array is

represented by 40 different 25-bp oligonucleotides, each with slight variations that allow accurate genotyping.

For further data analysis and illustration of the results, the CEL files were imported into the Copy Number Analysis Tool 4.0 software (CNAT) from Affymetrix. Quantile normalization was performed for all the arrays. Log2 ratio values were imported from CNAT to DNA Analytics 4.0 CGH Module (Agilent Technologies) to find out common aberration. Common aberration analysis was used for all the samples to identify genomic intervals that had statistically significant common aberrations. The two algorithms used in DNA Analytics 4.0 were ADM-1 and Fuzzy Zero.

Aberration Detection Method 1 (ADM-1 or "adam-one") was used as a detection algorithm to identify all aberrant intervals in the samples with consistently high or low log ratios based on the statistical score. The ADM-1 algorithm was used to searche for intervals in which the statistical score exceeded a user specified threshold. The statistical score was calculated based on the average log ratios of the probes and the number of probes in the interval. ADM-1 reported contiguous genomic regions of arbitrary size as aberrant regions. The ADM-1 statistical score was computed as the average normalized log ratios of all probes in the genomic interval multiplied by the square root of the number of these probes. It represented the deviation of the average of the normalized log ratios from its expected value of zero.

The Fuzzy Zero algorithm is an improved error model that explicitly includes the effects of long-range log ratio correlations. In this model, we assumed that there were two independent sources of noise contributing to the total log ratio variation: the local uncorrelated probe-to-probe noise and a global noise describing baseline variation. The

ADM-1 algorithm estimated the standard deviation of the mean log ratio of an interval using a statistical error model that treated probe to probe errors as independent. The errors of the probes were often correlated over wide genomic intervals, and the ADM algorithms therefore underestimated the error for long intervals. Long aberrations with low average log ratios were thus often incorrectly deemed significant. Fuzzy zero correction applied a "Global error model" to all aberrant intervals identified in ADM-analysis. The global error model used a more realistic error model to avoid erroneous aberration calls when the errors were correlated. For the global error model, we assumed that there were two independent sources of noise contributing to the total noise of the intervals. The global noise component is calculated as the variation of the average log ratios in large genomic intervals.

The following parameters were selected for common aberration analysis: statistical algorithm= ADM-1, sensitivity threshold = 6.0, minimum number of probes = 3, the moving average window= 2Mb, p value threshold =0.05, minimum ±log2 ratio=0.25. National Center for Biotechnology Information March 2006 Genome Build 36.1(hg 18) (the International Human Genome Sequencing Consortium) was used for genome locations.

Copy number information (gain and loss) was derived after first normalizing each tumor to its match normal in CNAGv3.0.

Results and Discussion:

10K array performance:

Twenty tissue samples of ESCC and matched germline DNA from same patients with ESCC were analyzed for allelic imbalance and copy number alterations using the GeneChip® Human Mapping 10K Array Xba 142 2.0 from Affymetrix in the current study. The average genotyping call rates for the tumor and matched normal samples were 95.54 % and 95.63% respectively (Table 4).

Table 4: SNP call rate and summary of amplified and deleted probes in tumor samples:

Sample ID	No. of Probes Amplification	No. of Probes Deletion	SNP Call rate XbaI (%)
134T	202	59	92.53
147T	179	38	91.91
211T	2022	1262	92.90
213T	669	342	98.91
215T	1119	210	97.87
259T	141	204	99.24
261T	562	244	95.82
270T	334	172	99.12
272T	456	37	91.11
274T	448	480	94.56
275T	462	210	99.54
276T	148	203	96.14
478T	379	49	90
278T	236	104	97.94
289T	359	39	98.94
139T	349	375	95.43
140T	612	16	92.11
277T	393	378	96.16
302T	367	266	95.85
318T	314	304	94.78

Note: Call rate: The percentage of SNPs called on the array.

Criteria for assessing genome-wide common aberration:

Several variables may compromise the accuracy of copy number estimations, including germ-line copy number variation, germ-line polymorphisms that interfere with the restriction digestion, and random noise. To counter the first two sources of potential error, we analyzed matching normal DNA for all cases and normalized all tumors against their matching normal DNA. The data point for each SNP was a mean –normalized hybridization intensity value represented as a $\log_2$ ratio relative to normal. This strategy effectively masked any germ-line copy number variations and emphasized only somatically acquired copy number changes. To compensate for random noise, a specific SNP was considered as showing gain or loss if the average $\log_2$ ratio of the region containing that SNP was a least ±0.25.

We have taken a conservative stance and designated a region as showing copy number gain or loss if it was represented by at least five consecutive SNPs with average copy number values of ±0.25 ($\log_2$) and encompassed regions of >25-kb in size. It was suggested that a 25 –kb cutoff would exclude most false-positive copy number changes [77].

Common regions of amplification:

DNA amplifications were detected at 1p36.13-p36.12 (*PLA2G5*, 85%), 1p21.1 (*COL11A1*, 70%), 1q21.1-q44 (*OBSCN, PTPRC, KCNK2, RGS1, KCNH1, S100A3*, *ENAH*, *RGS1*, 55%), 2p25.3-p25.2 (*TPO*, 90%), 2p14 (*ARHGAP25*, 95%), 2q24.3-q31.1 (*STK39*, 75%), 3q26.33-q27.1 (*ABCC5*, 60%), 3q28 (*FGF12*, 50%), 4q21.23-q21.3 (*ARHGAP24*, *MAPK10*,65%), 5p15.2-p12 (*SLC1A3* ,*OXCT1* ,*TRIO* ,*NIPBL, RNASEN*,85%), 5q11.2-q12.1 (*PLK2, RAB3C,KIF2A*, 65%), 5q13.1-q13.3 (*ENC1*, 80%),

6p25.3-q11.1 (*CDKAL1, CD83, TUBB2B, TXNDC5,* 50%), 7p21.1-p15.3 (*TWIST1*, 50%), 7p12.3 (*ABCA13*, 50%), 7q11.21-q11.22 (*CALN1*,50%), 9q13-q34.13 (*TSC1, TLR4, TNC, NTRK2,* 45%), 10q21.3 (*CTNNA3*, 75%), 11p13-p11.2 (*CD44, SLC1A2, RAG2*,90%), 11q22.1-q25 (*OPCML, NCAM1, TRIM29, GRIA4, ZBTB16*,50%), 12p13.1-p12.3 (*GRIN2B*,95%) , 13q12.12-q12.13 (*CDK8*, 45%), 13q21.33-q22.1 (*DACH1*, 50%), 17q21.2-q21.31 (*KRT13*,50%), and18q21.31-q21.33 (*TXNL1, CCBE1*, 65%) as shown in Table 5.

Table 5: Genomic regions of amplification

Chromosomal position	Start (bp)	End (bp)	Region Size(bp)	No of SNPs in region	% cases	P-value	Candidate genes
chr1 p36.13-p36.12	20281941	20682391	400451	7	85	0.0001179	*PLA2G5*
chr1 p22.3-p22.2	88038572	91331285	3292714	10	45	6.57E-05	*GBP7, PKN2*
chr1 p21.1	102333724	105069559	2735836	11	70	1.46E-06	*COL11A1* *AMY2B*
chr1 q21.1-q44	145093109	246923597	101830489	367	55	0.01832	*OBSCN* *PTPRC* *TNR* *RXRG* *KCNK2* *S100A3* *ESRRG* *ENAH* *RGS1* *PBX1* *KCNH1*
chr2 p25.3-p25.2	1482433	4591670	3109238	14	90	2.00E-05	*TPO* *TSSC1*
chr2 p16.2	53391905	54561969	1170065	8	45	2.82E-07	*PSME4* *SPTBN1*
chr2 p14	67807953	68907273	1099321	5	95	5.63E-05	*ARHGAP25*

chr2 q14.1-q14.2	118304850	120653278	2348429	5	35	0.001213	*DDX18* *INSIG2* *DBI*
chr2 q24.3-q31.1	168315163	169551130	1235968	10	75	6.19E-07	*STK39*
chr2 q31.1-q31.3	171751689	182713942	10962254	46	35	0.03297	*ITGA4* *CHN1*
chr2 q36.3-q37.1	228762237	230927706	2165470	13	55	1.27E-06	*SP140* *PID1*
chr3 p21.31	46218411	46931606	713196	5	35	0.007192	*CCR1* *LTF* *PTHR1*
chr3 q22.1	133202773	134013218	810446	5	45	0.003484	ACPP *CPNE4*
chr3 q26.33-q27.1	181750414	185121227	3370814	13	60	0.00021	*ABCC5*
chr3 q28	191368617	193791202	2422586	7	50	0.02495	*FGF12*
chr3 q29	195214362	197794932	2580571	8	40	0.04729	*HES1*
chr4 p15.2-p15.1	25223427	28894874	3671448	19	55	0.001708	*RBPJ* *CCKAR*
chr4 q21.23-q21.3	86135637	87471473	1335837	7	65	0.0002125	*ARHGAP24 MAPK10*
chr5 p15.32-p15.31	5869796	8193732	2323937	16	40	0.00302	*ADCY2*
chr5 p15.2-p12	13981677	44856159	30874483	103	85	0.004212	*SLC1A3* *OXCT1* *TRIO* *NIPBL* *RNASEN*
chr5 q11.2-q12.1	57543289	62783493	5240205	23	65	0.008236	*PLK2* *RAB3C* *KIF2A*
chr5 q13.1-q13.3	67902695	75316213	7413519	16	80	7.94E-05	*ENC1*
chr5 q14.1-q14.3	79825615	91157510	11331896	51	35	0.01033	*XRCC4* *RASGRF2* *SSBP2*
chr6 p25.3-q11.1	1428122	62400294	60972173	243	50	0.01162	*TINAG* *IL17F* *RIPK1* *CDKAL1* *CD83* *TUBB2B* *TXNDC5*

chr6 q25.2	152772950	153265455	492506	8	50	0.002916	*VIP* *SYNE1* *MYCT1*
chr6 q14.2-q14.3	84442199	86260641	1818443	8	40	9.97E-05	*SNAP91* *TBX18*
chr7 p21.1-p15.3	19114646	19946346	831701	6	50	3.88E-05	*TWIST1*
chr7 p12.3	47646647	48532909	886263	5	50	3.08E-07	ABCA13
chr7 q11.21-q11.22	61606041	71074933	9468893	10	50	1.93E-05	*CALN1*
chr7 q33	133701183	135428843	1727661	9	40	0.01142	*CALD1*
chr8 q11.22-q24.3	51088197	142773655	91685459	310	25	0.001507	*DDEF1* *ENPP2* *KCNK9* *RB1CC1* *ANGPT1* *EXT1* *TRPS1* *ST18* *RPL7*
chr9 p22.3-p22.2	15367858	16665012	1297155	9	35	0.00327	*BNC2*
chr9 q13-q34.13	70181245	134794803	64613559	255	45	0.005858	*TSC1* *FPGS* *TLR4* *TNC* *PRG-3* *NTRK2* *GSN* *DAPK1*
chr10 q11.23	50392956	51772407	1379452	5	40	3.46E-05	*ERCC6* *CHAT*
chr10 q21.3	68534465	69351872	817408	8	75	0.0003406	*CTNNA3*
chr10 q23.2-q23.31	88322941	91679417	3356477	12	25	1.71E-05	*FAS*
chr11 p15.3	10774980	11193239	418260	8	30	0.0001239	*EIF4G2*
chr11 p13-p11.2	32432150	44178222	11746073	56	90	0.02409	*CD44* *SLC1A2* *RAG2* *EHF* *CHST1*
chr11 q12.3-q14.1	63056827	78502105	15445279	20	30	0.03371	*GAB2* *NEU3* *PAK1* *FGF19*
chr11 q22.1-q25	100278627	134132834	33854208	164	50	0.004959	*OPCML* *NCAM1*

							TRIM29 *GRIA4* *ZBTB16* *MMP13* *CASP4* *GRIK4*
chr12 p13.1-p12.3	14020148	17547095	3526948	16	95	2.48E-06	*PTPRO* *GRIN2B* *MGST1* *RERGL*
chr12 q21.2-q21.31	77984002	79532118	1548117	7	25	5.59E-05	*PPP1R12A*
chr15 q13.3-q14	29278408	33058096	3779689	14	40		*RYR3*
chr15 q22.2-q22.31	59177132	61603631	2426500	11	30	7.36E-05	*RORA* *VPS13C*
chr15 q26.1	89146848	89685282	538435	6	45	3.30E-06	*BLM*
chr16 p12.1-p11.2	27424589	33506241	6081653	6	35	0.0006352	*XPO6* *GSG1L*
chr17 p13.1	8557946	9971130	1413185	7	40	4.61E-05	*CCDC42*
chr17 q11.1-q12	22844802	28976425	6131624	15	35	0.008428	*NOS2A* *KSR1*
chr17 q21.2-q21.31	36526037	40675706	4149670	6	50	1.83E-05	*KRT13*
chr18 q11.2	18685750	20646425	1960676	7	45	0.0000795 1	*RBBP8*
chr18 q21.1-q21.2	46257312	47389543	1132232	7	35	4.84E-06	*MAPK4*
chr18 q21.31-q21.33	52143401	57287738	5144338	20	65	0.0005508	*TXNL1* *CCBE1*
chr20 p13-p11.21	132803	24124359	23991557	112	40	0.03024	*CST3* *BTBD3* *PRNP* *CD93* *PLCB1* *FKBP1A* *NSFL1C* *THBD* *SLC24A3* *TASP1*

1p36.13-p36.12 that contains the gene *PLA2G5* was amplified in our study and has earlier been reported in breast and colon cancer progression [78, 79]. *PLA2G5* is involved in MAPK signaling pathway.1p21.1 contains the gene *COL11A1,* which is associated with malignancy in colon cancer [80] and involved in Focal Adhesion signaling pathway. 1q21.1-q44 contains *OBSCN*, *PTPRC*, *TNR*, *KCNK2*; *S100A3, ENAH, RGS1 and KCNH1* genes which are involved in progression of cancer. Mutations affecting *OBSCN* may be involved in cancer predisposition [81]. *KCNK2* expression is associated with abnormal cell proliferation and may be a novel marker and a molecular target in prostate cancer [82]**.** *S100A3* may play an important role in differentiation and progression of gastric cancer [83]**.** *ENAH,* a cytoskeleton regulatory protein involved in the regulation of cell motility and adhesion, is over expressed in breast cancer [84]. *RGS1* was found to be up regulated in late stage cancer compared to early stage cancer**,** suggesting the possible involvement in cancer progression [85]. *KCNH1* can mediate cancer progression [86]. KCNH1 protein and mRNA are expressed aberrantly in colorectal cancer and occasionally expressed in colorectal adenoma [87]. *PTPRC* and *TNR* gene are involved in Cell Adhesion and Cell Communication pathway respectively. 2q14.1-q14.2 contains *INSIG2,* which is over expressed in colon cancer cells resulting in increased cellular proliferation, invasion, anchorage independent growth and inhibition of apoptosis [88]. *TPO* gene at 2p25.3-p25.2 is involved in cytokine-cytokine receptor interaction pathway.

The most frequently amplified genomic regions were chromosome 3q26.33-q27.1 (*ABCC5*), 3q28 (*FGF12*), 3q29 (*HES1*). *HES1* has earlier been reported to be over-expressed in gastric, pancreatic and colorectal cancer [89]. *FGF12* and *HES1* are involved in MAPK signaling pathway and Notch Signaling pathway respectively. *FGF12*

was significantly up-regulated at mRNA level in ESCC in our previously reported gene expression study [].

4p15.2-p15.1 contains *CCKAR,* which is expressed in gastric and pancreatic cancer. *CCKAR* gene is involved in Neuroactive Ligand Receptor Interaction Pathway. *MAPK10* gene located at amplified region 4q21.23-q21.3 is involved in MAPK signaling pathway. 5p15.2-p12 contains *SLC1A3, TRIO, NIPBL, and RNASEN. SLC1A3* shows significantly over expression in prostrate cancer as compared to benign prostatic hyperplasia and is candidate marker of hypoxia in prostate cancer [90]. *TRIO* amplification and abundant mRNA expression are associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer [91]. *RNASEN* regulates cell proliferation and affects survival in esophageal cancer patients [92]. 5q11.2-q12.1 contains *PLK2,* which is over expressed in pancreatic cancer cell lines and pancreatic tumor tissues, is involved in a wide variety of cell cycle processes [93]. *CDKAL1* candidate gene from 6p22 region is frequently amplified in bladder cancer [94]. 6p25.3-q11.1 contains other candidate genes such as *CD83*, *TUBB2B*, and *TXNDC5*. *TXNDC5* was significantly up-regulated in colorectal adenoma and malignant tissues as compared to normal mucosa [95]. *CD83* gene polymorphisms have been reported to increase susceptibility to human invasive cervical cancer. CD83 cells were identified in the lamina propria surrounding intestinal type glands in Barrett s IM (intestinal metaplasia), dysplasia, and tumor tissues [96].

7p21.1-p15.3 contains *TWIST1,* which is found to be significantly over expressed in colorectal cancer samples compared to nontumorous colon mucosa. *TWIST1,* a basic helix loop helix (bHLH) transcription factor, has been reported to be involved in tumor progression and metastasis in several cancers. *TWIST1* has been shown to favor the

metastatic dissemination of cancer cells through its ability to induce an epithelial mesenchymal transition (EMT) [97]. 9q13-q34.13 contains *TSC1, TLR4, FPGS, TNC, NTRK2, and GSN* which are candidate genes for tumor progression. *TSC1* is a tumor suppressor gene involved in the development of various malignancies, including bladder cancer [98]**.** *TLR4* which is involved in Toll-like receptor signaling pathway was found to be aberrantly expressed in bladder cancer**,** inducing some genes expression and facilitating tumor progression. *TNC,* a new marker of cancer stroma, is elevated in sera of colon and breast cancer patients. TNC expression in the activated tumor stroma facilitates tumorigenesis by supporting the migratory behavior of breast cancer cells [99]. *TNC* gene is involved in Focal Adhesion pathway. *NTRK2* which is involved in MAPK signaling pathway, enhances ovarian cancer cell migration and proliferation [100]. *GSN* gene is involved in Regulation of Actin Cytoskeleton pathway.

11p13-p11.2 contains *CD44* and *SLC1A2* candidate genes which are involved in cancer progression. CD44 is implicated in various adhesion- dependent cellular processes, including cell migration, tumor cell metastasis and invasion [101]. *SLC1A2* gene shows significant over expression in prostrate cancer compared to BPH tissue and is candidate marker of hypoxia in prostate cancer [90]. *FGF19* locus at human chromosome 11q13.3 is amplified in head and neck tumors, esophageal cancer**,** Kaposi s sarcoma, bladder tumors, breast and liver cancer [102]**.** *FGF19* and *PAK1* gene at 11q12.3-q14.1 are involved in MAPK signaling pathway. 11q22.1-q25 contains candidate genes for tumor progression such as *NCAM1* and *TRIM29*. *NCAM1* up-regulation induces the formation of novel signaling complexes that correlate with NCAM1 dependent focal adhesion assembly, migration, and cancer cell invasion [103]. The over expression of

TRIM29 in gastric cancer tumor tissue has been reported as a candidate marker of lymph node metastasis [104]. 13q12.12-q12.13 contains *CDK8* which is a colorectal cancer oncogene regulating beta catenin activity [105]. 17q21.2-q21.31 contains *KRT13* which encodes a cytoskeletal protein and is considered to be important roles in breast cancer growth and metastasis [106]. *KRT13* gene is involved in cell communication. 18q11.2 contains *RBBP8* gene which may be useful biomarkers for breast cancer prognosis and clinical management [107]. 20p13-p11.21 contains candidate gene, *PRNP,* which is over expressed in gastric cancer [108].

Common regions of deletion:

Highest frequency of deletions were detected in chromosomal regions 1p36.32-p36.31 (*NPHP4, PRDM16*, 45%), 1p36.21-p36.13 (*PAX7*, 45%), 1p32.3-p33 (*SCP2,* 50%), 2p16.1 (*PNPT1*, 45%), 3p26.3-p14.3 (*WNT7A, KCNH8, FBLN2, TGFBR2, SYN2*, 30%), 4p13 (*KCTD8*, *GNPDA2*, 55%), 5q32 *(TCERG1, POU4F3, KCTD16*, 45%), 5q35.1-q35.2 (*RANBP17, FGF18, DOCK2, KCNIP1*,55%), 6p22.1-p21.33 (*TRIM26, TRIM31*), 6p21.32-p21.2 (*HLA-DRA*,40%), 6q13-q14.1 (*COL12A1*), 6q23.3-q24.2 (*HECA, NMBR, GPR126,* 40%) 6q25.3-q26 (*AGPAT4, MAP3K4,PLG*, 35%), 8p23.2-p21.3 (*DLC1, LZTS1, MTUS1, CSMD1, PPP3CC, MCPH1, PSD3*, 50%), 8p21.1-p12 (*WRN, PPP2CB*,40%), 8q13.3-q21.1 (*RPL7*,30%), 10p15.3-p11.21 (*OPTN, AKR1C3, ZEB1, CACNB2, WAC*, *CUBN*,50%), 11p15.1 (*CSRP3*, 35%), 11p11.2 (*TSPAN18, CRY2, PRDM11*, 45%), 11q23.2-q23.3 (*CEP164*,45%), 11q24.2 (45%), 12q15 (*IFNG,IL26,IL22,MDM1*,45%), 12q22 (*PLXNC1, CCDC41, 50*%), 13q12.11-q34 (*CLYBL, FARP1, KCTD12, KLHL1, PCDH9, COL4A1*, 40%), 14q23.1, 14q23.2-q23.3 (55%), 16q12.1, 16q13, 17q21.31-q21.33 (*MAPT, CDC27*, 55%), 18q12.3-q21.1

(*SLC14A2, SLC14A1*, 50%), 18q21.2 (*DCC*, 60%), 18q22.1 (*CDH19*, 55%), 18q22.3-q23 (*FBXO15, CYB5A*, 55%), 19p13.2-p12, and 19q13.32-q13.33 (CARD8,45%) as shown in Table 6.

Table 6: Genomic regions of deletion:

Chromosomal position	Start (bp)	End (bp)	Region Size(bp)	No of SNPs in region	% cases	P-value	Candidate genes
chr1 p36.32-p36.31	3095215	5845656	2750442	6	45	0.000143	*NPHP4* *PRDM16*
chr1 p36.21-p36.13	14901106	18940996	4039891	7	45	1.42E-05	*PAX7*
chr1 p32.3-p33	51202917	53393081	2190165	5	50	2.08E-06	*NRD1* *SCP2*
chr1 p12	117966066	119479925	1513860	9	55	4.81E-05	*FAM46C* *WARS2* *TBX15*
chr2 p16.1	55672398	57157076	1484679	6	45	8.43E-05	*PNPT1*
chr2 q21.2-q22.1	133576384	139639475	6063092	23	35	1.48E-05	*CXCR4* *NXPH2* *MGAT5* *R3HDM1* *TMEM163*
chr3 p26.3-p14.3	816534	55517580	54701047	226	30	0.001811	*TGFBR2* *LMCD1* *ULK4* *SRGAP3* *VIPR1* *SLC6A20* *VGLL4* *FBLN2* *UBE2E1* *UBE2E2* *WNT7A* *KCNH8* *DOCK3* *GRM7* *TATDN2* *CACNA1D* *CNTN4* *RBMS3* *ERC2* *SYN2*

chr3 q13.13-q13.2	111105232	112928439	1823208	7	35	0.008792	*PLCXD2* *CD96* *PVRL3* *ZBED2*
chr3 q25.33-q26.1	161100763	162595748	1494986	7	35	0.001509	*PPM1L* *IL12A*
chr4 p13	43527831	44635753	1107923	6	55	2.28E-07	*GNPDA2* *KCTD8*
chr5 q31.3-q32	142322646	144021014	1698369	11	30	0.000335	*ARHGAP26*
chr5 q32	143335099	146281601	2946503	19	45	0.000128	*TCERG1* *POU4F3* *KCTD16* *PPP2R2B* *PRELID2*
chr5 q35.1-q35.2	169421367	174105870	4684504	17	55	5.40E-06	*RANBP17* *FGF18* *DOCK2* *KCNIP1*
chr6 p22.1-p21.33	29420126	30279766	859641	6	65	0.03349	*TRIM26* *TRIM31*
chr6 p21.32-p21.2	32505276	37806755	5301480	9	40	0.001323	*HLA-DRA*
chr6 q13-q14.1	75165874	75976829	810956	5	45	0.01808	*COL12A1*
chr6 q23.3-q24.2	138011488	143113369	5101882	15	40	4.68E-07	*HECA* *NMBR* *GPR126* *CITED2*
chr6 q25.3-q26	160804768	161505967	701200	6	35	1.35E-05	*AGPAT4* *MAP3K4* *PLG*
chr7 q21.11	78447713	79453421	1005709	5	40	7.93E-06	*MAGI2*
chr7 q21.2-q21.3	91065927	94664563	3598637	6	50	0.000104	*COL1A2* *PPP1R9A* *MTERF*
chr8 p23.2-p21.3	3284123	22576721	19292599	78	50	0.001538	*DLC1* *LZTS1* *MTUS1* *CSMD1* *PPP3CC* *MCPH1* *ANGPT2* *PSD3* *NAT2* *TUSC3*
chr8 p21.1-p12	28162657	32261478	4098822	20	40	0.00076	*INTS9* *WRN* *PPP2CB* *NRG1* *TEX15* *PNOC* *RBPMS*

chr8 q13.3-q21.11	73656766	75995821	2339056	5	30	0.008016	*KCNB2* *RPL7* *JPH1*
chr8 q23.1-q23.2	110496031	110645915	149885	5	35	0.008568	*PKHD1L1*
chr10 p15.3-p11.21	2232977	38385694	36152718	159	50	0.03882	*OPTN* *AKR1C3* *ZEB1* *CACNB2* *WAC* *CUBN* *FRMD4A* *GAD2* *PARD3* *NEBL* *RBM17* *KIAA1217* *GPR158* *PLXDC2* *NRP1*
chr11 p11.2	44745944	45853528	1107585	7	45	1.21E-05	*TSPAN18* *CRY2* *PRDM11*
chr11 q23.2-q23.3	114229200	118943200	4714001	13	45	4.39E-05	*CEP164*
chr11 q24.2	125475083	126563346	1088264	11	45	0.000532	*FAM118B* *KIRREL3* *ST3GAL4* *RPUSD4*
chr12 q15	66786868	67054801	267934	6	45	0.000352	*IFNG* *IL26* *IL22* *MDM1*
chr12 q22	93062205	93485924	423720	5	50	0.04572	*PLXNC1* *CCDC41*
chr12 q24.32-q24.33	127328943	130687507	3358565	15	40	0.000263	*TMEM132D*
chr13 q12.11-q34	18582246	111843582	93261337	434	40	0.004564	*CLYBL* *FARP1* *KCTD12* *KLHL1* *POLR1D* *PCDH9* *LCP1* *DIAPH3* *USP12* *FREM2* *COL4A1* *SUCLA2* *CLDN10* *GPC5* *ESD* *HTR2A* *TRPC4* *TBC1D4*

							TPTE2 *PIBF1* *FGF14* *FLT1* *DGKH* *GPC6* *MYO16* *DCLK1* *PCDH20* *TDRD3* *ARHGEF7* *NBEA* *DOCK9* *SPATA13* *SOX1* *TNFSF11* *WASF3* *CDK8* *ABCC4*
chr14 q23.1	59754188	61019845	1265658	7	45	1.67E-06	*PRKCH*
chr14 q23.2-q23.3	62891871	64678800	1786930	5	55	0.000238	*PPP2R5E*
chr16 q12.1	49017772	49928092	910321	6	50	0.000435	*SALL1*
chr16 q13	55534503	56472990	938488	7	45	7.67E-07	*GPR56* *HERPUD1*
chr17 q21.31-q21.33	41458104	47581513	6123410	8	55	9.34E-08	*TTLL6* *MAPT* *CDC27*
chr18 p11.21	11836012	14920951	3084940	6	30	0.001404	*SPIRE1*
chr18 q12.3-q21.1	40925672	42130277	1204606	7	50	0.000429	*SLC14A2* *SLC14A1*
chr18 q21.2-q23	47389143	75328829	27939687	109	25	0.00811	*NFATC1* *CDH19* *DOK6* *DCC* *MBD2* *FBXO15* *TCF4* *CYB5A*
chr18 q21.2	48085370	48773244	687875	5	60	0.000149	*DCC*
chr18 q22.1	61438527	62328959	890433	5	55	4.62E-07	*CDH19*
chr18 q22.3-q23	69454857	73772793	4317937	14	55	0.000283	*FBXO15* *CYB5A*
chr19 p13.2-p12	9590984	22158041	12567058	27	45	0.001975	GATAD2A
chr19 q13.32-q13.33	52837585	55076946	2239362	10	45	2.35E-06	*GLTSCR1* *CARD8*
chr20 q11.21-q13.33	29531586	61447029	31915444	88	40	0.001317	*BCAS1* *EYA2*

							KCNB1 PTPRT AHCY
chr22 q11.21	16971116	19782953	2811838	5	40	4.98E-07	*CLDN5*

Deletion of chromosome band 1p36 is frequently found in many malignancies, including gastric cardia carcinoma, colon cancer and ESCC. 3p26.3-p14.3 contains candidate genes such as *TGFBR2, FBLN2, and WNT7A*. The carcinogenesis process in ulcerative colitis associated colorectal cancer is closely related to the microsatellite instability pathway through *TGFBR2* mutation by a dysfunction of the mismatch repair system [109]. *TGFBR2* gene is involved in MAPK signaling pathway. Loss of Fibulin 2 [*FBLN2*] expression may facilitate migration and invasion in breast cancer [110]. Loss of *WNT7A* expression may be important in lung cancer development or progression due to its influence on E cadherin [111]. *WNT7A* gene is involved in Wnt signaling pathway. Minimal deleted regions at 3p26.3, 3p22, 3p21.3, and 3p14.2 containing potential tumor suppressor genes may contribute to the pathogenesis of esophageal cancer in a high-risk Chinese population [75]. *PPP2R2B* gene at 5q32 is involved in Wnt signaling pathway. *HLA-DRA* gene at 6p21.32-p21.2 is involved in cell-adhesion. *COL1A2* gene at 7q21.2-q21.3 is involved in Focal adhesion.

8p23.2-p21.3 contains candidate genes such as *DLC1, LZTS1, MTUS1, MCPH1 and PSD3*. *DLC1* (deleted in liver cancer 1) is a potential tumor suppressor gene, which is inactive in liver carcinogenesis [112]. *DLC1* encodes a RhoGTPase activating protein, is recurrently down regulated or silenced in various solid tumors and hematological malignancies because of epigenetic modifications or genomic deletion [113]. *DLC1* suppresses non small cell lung cancer growth and invasion by RhoGAP dependent and independent mechanisms [114]. *LZTS1* inhibits cancer cell growth through regulation of

mitosis [115]. *LZTS1* is down regulated in high grade bladder cancer**,** and its restoration suppresses tumorigenicity in transitional cell carcinoma cells [116]. *LZTS1* is inactivated in many cancers with 8p deletions, including prostate, esophageal, gastric, bladder, and breast cancer [117]. *MTUS1* maps to chromosome 8p, a region frequently deleted and associated with disease progression in human cancers, including breast cancer [118]. Microcephalin [*MCPH1*] expression is reduced in breast cancer cell lines and in tumors of the ovary and prostate [119]. *MCPH1* has a crucial role in the DNA damage response by promoting the expression of Checkpoint kinase 1 (CHK1) and Breast cancer susceptibility gene 1 (BRCA1). Aberrantly reduced expression of *MCPH1* in human carcinomas implicates this protein in cancer initiation and progression [120]. *PSD3* shows a complete loss of expression in several ovarian cancer cell lines [121]. 8p21.1-p12 contains *WRN* and loss of the RecQ helicase WRN protein causes the cancer prone progeroid disorder Werner syndrome (WS) [122]. *PPP3CC* gene at 8p23.2-p21.3 and *PPP2CB* gene at 8p21.1-p12 are involved in Wnt signaling pathway. Deletion of 8p22 has been found in patients with SCC of the lung and esophagus [123]. *RPL7* gene at 8q13.3-q21.11 is involved in Ribosomal biosynthesis pathway. *AKR1C3* gene at 10p15.3-p11.21 is involved in metabolism of xenobiotics by cytochrome P450. *IFNG, IL26* and *IL22* genes at 12q15 are involved in Natural Killer Cell mediated cytotoxicity and Cytokine-cytokine receptor interaction pathway.

13q12.11-q34 contains candidate genes such as *COL4A1,* which can inhibit hepatic metastasis from colorectal cancer**,** as a result of its inhibition on tumor angiogenesis. *COL4A1* gene is involved in cell communication and focal adhesion pathway. *CLDN10, FLT1, WASF3, and TNFSF11* genes at 13q12.11-q34 are involved in cell adhesion, focal

adhesion, adherens junction and cytokine-cytokine receptor interaction pathway respectively. 16q13 contains *GPR56* which is suppressed in human pancreatic cancer cells [124]. *CDC27* gene at 17q21.31-q21.33 is involved in cell cycle. 18q21.2 contains candidate gene *DCC* which shows loss or decreased expression in esophageal cancer [125]. Deleted in Colorectal Cancer (*DCC*) is a putative tumor suppressor gene, whose loss has been implicated in colorectal tumorigenesis and whose tumor suppressor activity appears to be dependent on its ability to trigger apoptosis when disengaged by its ligand netrin 1 [126]. *NFATC1* gene at 18q21.2-q23 is involved in Wnt signaling pathway and Natural Killer cell mediated cytotoxicity. *CLDN5* gene at 22q11.21 is involved in cell-adhesion.

Sample specific regions of loss in ESCC:

261T tumor showed homozygous loss (average $\log_2$ copy number ratio $\leq$-0.7) on 3q26.1 (*PPM1L*) and 13q31.3 (*GPC5, GPC6*). Sample specific hemizygous loss (average $\log_2$ copy number ratio between -0.3 and -0.7) was seen on 3p14.2-q13.31 (*FHIT*), 3p26.3-p11.2 (*CNTN6, CNTN4*), 4p16.3-p16.1 (*SLC2A9*), 4p15.33-p12 (*KCNIP4, CCKAR, PCDH7, and UCHL1*), 4q22.1-q35.2 (*CASP3, PDE5A, and PCDH18)*, 6q11.1-q16.1 (*MAP3K7, COL12A1*), 7p21.3-p14.3 (*PDE1C, HDAC9, ETV1, and DGKB*), 8p23.2-p12 (*ANGPT2, DLC1, NAT2, NRG1, PTK2B, EPHX2*), 9p21.3 (*CDKN2A,CDKN2B*), 10p13-p12.33 (*CACNB2*), 11q13.4-q25 (*CASP1, GRIA4, CASP4, NOX4, NCAM1, MMP13, GRM5*), 11p11.2-p11.12 *(FOLH1*), 11q13.3-q13.5 *(PDE2A*), 11q21-q22.1 (*JRKL*, *MAML2*), 12p11.22-p11.21 (*OVOS2, CAPRIN2*), 13q12.11-q21.33 (*PCDH9, PCDH17*), 13q31.1 (*RBM26*), 13q31.1-q31.3 (*GPC5*), 13q31.3-q32.1(*GPC6, CLDN10, ABCC4*), 13q12.11-q32.1 (*PCDH9,PCDH17,PCDH8, GPC5, GPC6,TDRD3,*

DACH1,*FOXO1*), 13q32.1-q34 (*COL4A1, CLDN10, GPR18, LIG4*), 14q32.2 (*CCNK*,*BCL11B*), 17p13.3-p11.2 (*TP53, GAS7, NLRP1*), 18p11.32 (THOC1), 18p11.31-p11.21 (*RAB12*, *LAMA1*,*PTPRM*), 18q11.2-q23 (*NFATC1, NEDD4L, BCL2*,*DTNA, LAMA3, TXNL1*) and 18q21.2-q21.33 (*TCF4, DCC, NEDD4L, BCL2, TXNL1*) as shown in Table7.

Table 7: Sample Specific Homozygous Loss (< -0.7) and hemizygous loss (-0.3 to -0.7) in ESCC

Sample ID	Chromosomal position	Start (bp)	End (bp)	Size (bp)	Mean log2	No of SNPs	Genes of interest
261T	3q26.1	165428475	165727121	298647	-1.4777	5	*PPM1L*
261T	13q31.3	91870277	93533153	1662877	-1.1333	8	*GPC5* *GPC6*
261T	3p14.2-q13.31	60291457	115340891	55049435	-0.5092	166	*FHIT*
272T	3p26.3-p11.2	653347	88845422	88192076	-0.3587	349	*CNTN6* *CNTN4*
261T	4p16.3-p16.1	398952	10760950	10361999	-0.4702	27	*SLC2A9*
261T 272T 211T	4p15.33-p12	11587429	46766741	35179313	-0.4315	143	*KCNIP4*, *CCKAR*, *PCDH7*, *UCHL1*
272T 274T 211T	4q22.1-q35.2	89933759	191091333	101157575	-0.3592	372	*CASP3*, *PDE5A*, *PCDH18*
261T	6q11.1-q16.1	62394790	95125887	32731098	-0.4709	143	*MAP3K7* *COL12A1*
272T	7p21.3-p14.3	11033479	32639963	21606485	-0.3663	92	*PDE1C*, *HDAC9*, *ETV1*, *DGKB*,
272T 261T 274T	8p23.2-p12	4473854	36843477	32369624	-0.3789	139	*ANGPT2*, *DLC1*, *NAT2*, *NRG1*, *PTK2B*, *EPHX2*,
147T	9p21.3	21525576	23996322	2470747	-0.6812	16	*CDKN2A*, *CDKN2B*
	10p13-p12.33	17051181	19165187	2114007	-0.6026	10	*CACNB2* *CUBN*
261T	11q13.4-11q25	71886708	134402514	62515807	-0.4835	292	*CASP1* *GRIA4* *CASP4* *NOX4* *NCAM1* *MMP13* *GRM5*
134T	11p11.2-p11.12	48332180	51247890	2915711	-0.3745	11	*FOLH1*
134T	11q13.3-q13.5	69291704	75130040	5838337	-0.4582	8	*PDE2A*
276T	11q21-q22.1	95601158	97426454	1825297	-0.3308	13	*JRKL* *MAML2*
272T 261T 274T	13q12.11-q32.1	18478972	94172452	75693481	-0.3713	357	*PCDH9* *PCDH17* *GPC5* *GPC6*

							PCDH8 *TDRD3* *DACH1*
272T 261T	13q32.1-q34	94627589	114040116	19412528	-0.3818	77	*COL4A1* *CLDN10* *LIG4* *GPR18* *PCCA*
134T	14q32.2	98329121	99862235	1533115	-0.5095	5	*CCNK*
272T	17p13.3-p11.2	3565047	20147237	16582191	-0.3631	51	*TP53* *GAS7* *NLRP1*
261T 134T	18q11.2-18q23	19749481	75267441	55517961	-0.4307	232	*NFATC1* *NEDD4L* *TCF4* *DCC*

Loss of 3p14.2-q13.31 in ESCC may indicate a specific tumor suppressor gene (*FHIT*) at this locus that may be involved in ESCC of high-risk areas. Inactivation of tumor suppressor gene *FHIT* is a frequent genetic change in lung cancer [127]. 4p15.33-p12 contains candidate gene *UCHL1* which showed methylation in primary ovarian cancer, primary colon cancer and diffuse type gastric cancer. A high frequency of promoter methylation of *UCHL1* gene was observed in head and neck squamous cell carcinoma, ESCC, gastric, lung, prostrate and hepatocellular carcinoma. *UCHL1* is a tumor suppressor gene that is inactivated by promoter methylation or gene deletion in several cancers [128]. *CCKAR* gene at 4p15.33-p12 is involved in Neuroactive ligand receptor interaction pathway. *CASP3* gene at 4q22.1-q35.2 is involved in apoptosis. *MAP3K7* gene at 6q11.1-q16.1 is involved in both MAPK signaling and Wnt signaling pathway. *EPHX2*, *NRG1* and *PTK2B* genes at 8p23.2-p12 are involved in metabolisms of xenobiotics by cytochrome P450 /Arachidonic acid metabolism pathway, ErbB signaling pathway and Natural Killer Cell mediated cytotoxicity respectively.

The loss of 9p21.3 suggests that inactivation of the *CDKN2A/2B* genes is of pathogenetic importance. *CDKN2A* codes for two key tumor suppressor proteins, $p16^{INK4a}$ and $p14^{ARF}$ involved in the regulation of cell cycle G1 and G2/M control. *CDKN2B* codes for a p16-related cyclin-dependent kinase inhibitor $p15^{INK4b}$. Together these three tumor suppressors regulate two important cell cycle checkpoints and the loss of these genes can lead to replicative senescence, cell immortalization, and tumor generation. Deletion of 9p was related with metastasis in patients with solid tumor types. *CDKN2A* on 9p21 was identified as an important tumor suppressor gene in patients with different tumors and its association with metastatic and invasive phenotypes of ESCC has been reported earlier [47].

13q12.11-q32.1 which showed copy number loss contains a candidate tumor suppressor gene *PCDH8* of breast cancer and *DACH1*. *DACH1* inhibited prostrate cancer cellular DNA synthesis, oncogene mediated breast oncogenesis and breast cancer epithelial cell DNA synthesis [129]. An association of 13q12-q14 deletion with lymph node metastasis was found in an earlier report [47]. *GPC6* on 13q31.3 –q32 is markedly decreased in human gastric cancer. *GPC* is frequently silenced in ovarian cancer and breast cancer. Ectopic expression of *GPC* inhibited growth of tumor cells, suggesting that *GPC* plays a negative role in cell proliferation [130,131].

Loss of 17p13.3 has been reported in ESCC. One potentially relevant gene at 17p is TP53, whose product contributes to the control of cell proliferation and malignant transformation. In cancer patients, *TP53* is frequently inactivated through mutations [123]. Copy number loss was also observed at 18q21.2-q21.33 containing *TCF4* and *NEDD4L*. *NEDD4L* expression is down regulated in prostrate cancer and its expression

may be involved in prostrate cancer development [132] while down regulation of *TCF4* has been reported to increase cell migration [133].

Sample specific regions of low level gain in ESCC:

Low-level gains (average $\log_2$ copy number ratio between 0.3 and 0.99) were detected in chromosomal regions 1q23.1-q23.3 (*CD84, CRP*), 3q11.2-q29 (*EPHA6*), 3q22.1-q29 (*CLDN18, FXR1, BCL6*, *FGF12, IL12A, DLG1*), 3q26.2, 5p15.33-p12 (*CDH12, CDH18, CDH6, FGF10*), 6p24.1-p22.3 (*PHACTR1, EDN1, ATXN1, CDKAL1*), 6p22.3-p21.1 (*SOX4, HLA-DRA, HLA-DQA1, and HLA-DQB1*), 6q16.1-q27 (*GRIK2, PTPRK, and PDE10A*), 6p12.3 (*RUNX2*), 6p21.1-p12.3 (*RUNX2*, *CDC5L, and SPATS1*), 6q16.1 (*EPHA7*), 7p22.3-p21.3 (*MAD1L1*), 8p12-q24.3(*PDE7A,HNF4G*), 8q13.2-q13.3 (*SULF1*), 8q23.3(*CSMD3*), 8q24.11-q24.3(*ZFAT*), 11p13-p11.2(*CD44,RAG2*), 11q14.1(*PAK1,DLG2,USP35*), 12p13.33-q12, 12q12-q24.33, 13q22.1-q31.1(*KLF12, LMO7*, *SPRY2*), 15q21.2-q21.3 (*TMOD3, NEDD4*, *TMOD2, RAB27A*), 15q22.2 (*VPS13C*, *RORA*), 18p11.32-p11.31 (*DLGAP1*), 20p13-p12.3 (*SNPH, PRNP*, *BMP2*) and 22q13.31 (*SMC1B*) as shown in Table 8.

Table 8: Sample Specific Low level gain (0.3 to 1) in ESCC

Sample ID	Chromosomal position	Start	End	Size (bp)	Mean log2	No of SNPs	Genes of interest
289T	1q23.1-q23.3	156376556	158900901	2524346	0.4147	14	*CD84 CRP*
272T 211T	3q22.1-q29	133361727	196303395	62941669	0.6332	224	*CLDN18 FXR1 BCL6 FGF12 IL12A*
134T 274T	5p15.33-p12	1677351	45757044	44079694	0.3004	173	*CDH12 CDH18 CDH6 FGF10 TRIO RNASEN*
274T	6p24.1-p22.3	12372985	22502049	10129065	0.4661	50	*PHACTR1 EDN1 ATXN1 CDKAL1*
259T 215T 213T	6p21.1-p12.3	44348171	45529608	1181438	0.6628	9	*RUNX2 CDC5L SPATS1*
261T	6p22.3-p21.1	18515722	41551593	23035872	0.3195	86	*SOX4 HLA-DRA HLA-DQA1 HLA-DQB1*
261T	6q16.1-q27	95212976	170506286	75293311	0.4315	298	*GRIK2 PTPRK PDE10A*
147T	7p22.3-p21.3	1838367	13201036	11362670	0.4032	51	*MAD1L1*
261T	8p12-q24.3	34747946	142072236	107324291	0.5466	331	*PDE7A HNF4G*
139T	8q13.2-q13.3	68812139	70795945	1983807	0.3	8	*SULF1*
276T	11p13-p11.2	32009583	44160735	12151153	0.4124	57	*CD44 RAG2*
272T	11q14.1	76774826	84897809	8122984	0.3297	45	*PAK1 DLG2 USP35*
261T	13q22.1-q31.1	73251899	78642856	5390958	0.8312	30	*KLF12 LMO7*
261T	13q31.1	79781647	83938003	4156357	0.7278	15	*SPRY2*
147T	15q21.2-q21.3	49815069	54816207	5001139	0.3589	32	*TMOD3 NEDD4 RAB27A TMOD2*
270T	15q22.2	59210481	60246906	1036426	0.9665	7	VPS13C RORA
261T	18p11.32-	2512159	3921934	1409776	0.6645	8	*DLGAP1*

	p11.31						
277T	20p13-p12.3	4260153	6755276	2495124	0.7329	5	*PRNP* *BMP2*
147T	22q13.31	44038578	44361533	322956	0.3145	5	*SMC1B*

1q23.1-q23.3 which showed copy number gain in our study contains candidate gene *CRP*. Elevated CRP levels have been reported to be associated with early death in patients with cancer having localized disease [134]. *CLDN18* and *IL12A* genes at 3q22.1-q29 are involved in cell adhesion and Toll-like receptor signaling pathway respectively. Copy number gain was also observed in 5p15.33-p12 containg *CDH6* that showed elevated expression in kidney cancer tissues and *FGF10* which involved in cell migration and invasion in pancreatic cancer cells through interaction with FGFR2, resulting in a poor prognosis [135]. *FGF10* gene is involved in MAPK signaling pathway.

6p24.1-p22.3 containg candidate genes *EDN1* and *CDKAL1* showed copy number gain. *CDKAL1*, candidate gene from 6p22, is frequently amplified in bladder cancer [94]. EDN1 is produced by several types of cancer cells and has been proposed to participate in tumor development or progression by exerting autocrine or paracrine actions on neoplastic cells and their surrounding stromal cells [136]. *EDN1* enhances the expression of the androgen receptor via activation of the c-myc pathway in prostrate cancer cells [137] and proliferation of lung cancer cells by increasing intracellular free calcium [138]. *SOX4* is a target of gene amplification at chromosome 6p in lung cancer [139]. Gain of 6p12.3 which harbors runt-related transcription factor (*RUNX*) gene was detected in ESCC. *RUNX* is associated with cell migration and invasion. This implies that the over expression of an oncogene(s) at 6p12.3 confers a selective advantage in ESCC. In cancer cells, *RUNX2* activates expression of bone matrix and adhesion proteins, matrix

metalloproteinases and angiogenic factors that have long been associated with metastasis. Gains of 6p and 20p may contribute to the lymph node metastasis of ESCC [123].

Gain of a region on 7p22.3, containing *MAD1L1*, is the most frequent event in small cell lung cancer cell lines [140]. Gain of 7p22.3 predicted nodal metastasis [46]. 8q13.2-q13.3 which showed copy number gain contains candidate gene *SULF1*. High expression of *SULF1* occurs in the stromal elements as well as in the tumor cells in pancreatic cancer [141]. *PAK1*, candidate gene from 11q14.1, is over expressed during the progression of breast cancer [142]. *PAK1* gene is involved in MAPK signaling pathway. Gain of a region on 15q21.2-q21.3 contains *RAB27A* whose over expression was found to increase the invasive and metastatic abilities in breast cancer cells in vitro and in vivo [143].

Comparison of chromosomal aberrations with earlier studies in patients with ESCC:

Frequent amplifications were found on chromosomes arms 1p36.13, 1q21.1, 2p14, 3q28, 3q27, 3q26.1, 5p15.2, 5q11.2, 6p25.3, 7q11.21, 9q31.3, and17p13.1 which were previously reported in Chinese population by Nan Hu et al. 9p21.3 showed copy number loss which was reported in Chinese population by Nan Hu et al [74]. Gains of 8q, 3q, 5p and losses of 3p, 8p and 13q were also reported in ESCC from Linzhou population [123].

Frequent gain abnormalities found on chromosome arms 1q, 3q, 5p, 7p, 8q, 11q, 12p, 12q, and frequent deletions found on chromosome arms 1p, 3p, 4p, 5q, 8p, 9p, 11q, 13q, 17p, 18q were also reported from ESCC in Taiwan population [47] and black male population of South Africa .

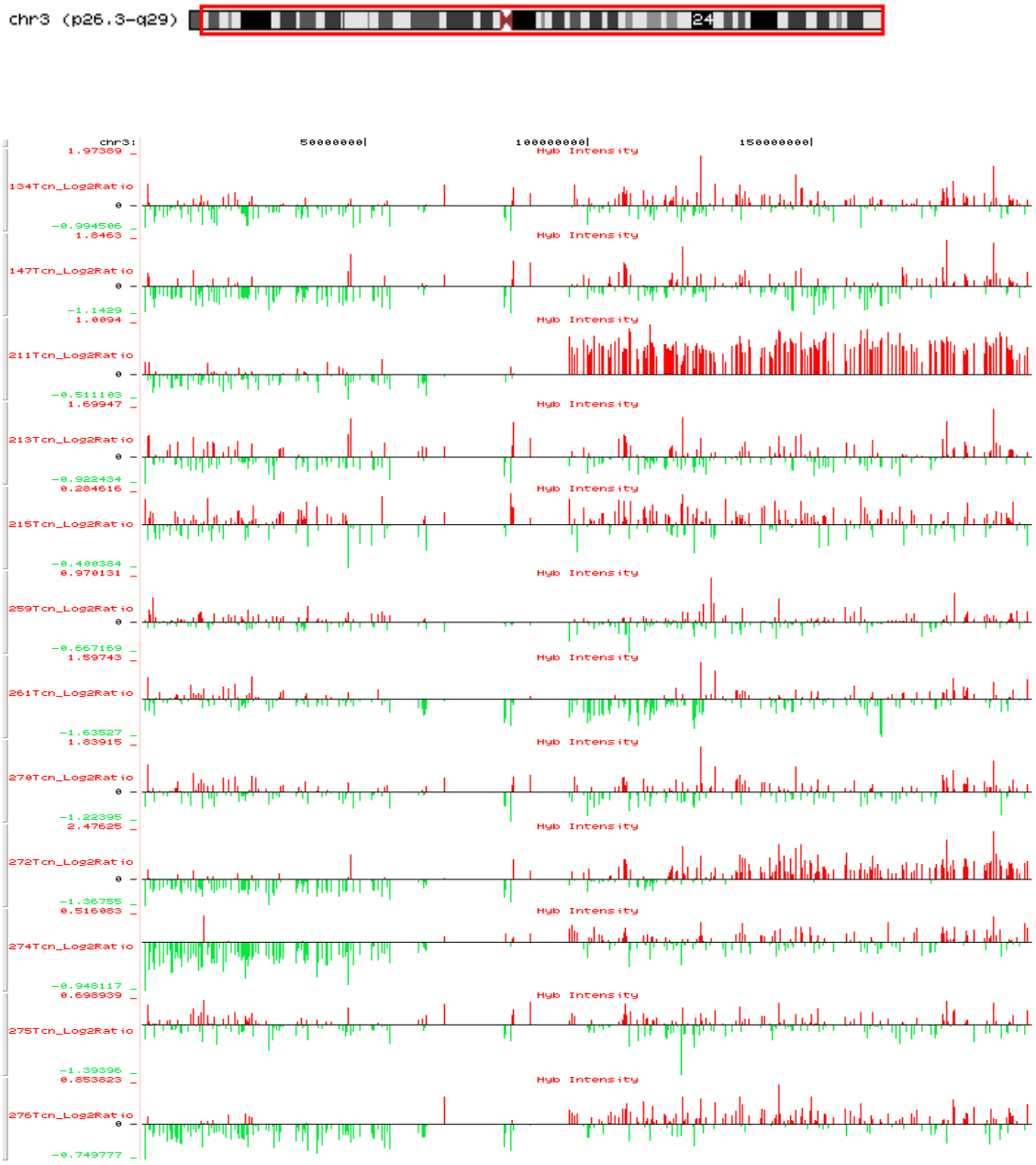
There was a clear reciprocal correlation between gene deletions on chromosomes 3p and gene amplification /duplications on chromosome 3q in ESCC, suggesting a molecular

mechanism for the generation of somatic chromosomal aberrations (Fig.13). Deletion of chromosome 3p or duplication of chromosome 3q, or coupled 3p deletion and 3q duplication have also been observed in lung cancer, suggesting that genetic alterations of genes on chromosome 3 might be a common molecular mechanism for human epithelial cell carcinogenesis [72].

High level amplification detected at 2q24, 7p12, and 7q11.2 was previously reported in the black male population in South Africa [47]. Frequent amplification detected at 2p25.3-p25.2, 4p15.2-p15.1, 4q21.23-q21.3, 5q13.1-q13.3, 7p21.1-p15.3, 11p13-p11.2, 11q22.1-q25, 12p13.1-p12.3, 17q21.2-q21.31 and 18q21.31-q21.33 and frequent deletion detected at 2p16.1, 6p22.1-p21.33, 6p21.32-p21.2, 6q13-q14.1, 6q23.3-q24.2, 7q21.2-q21.3, 10p15.3-p11.21, 12q15, 17q21.31-q21.33 and 22q11.21 were reported in ESCC for the first time in our study.

However, different chromosomal abnormalities have been reported by different authors. One possible explanation for such differences is that there may be some ethnic or geographic factors that contribute to the various patterns of chromosomal changes in patients with ESCC from different reports. Another explanation is the *mutator phenotype* theory. Tumor cells initially gain their phenotype as a result of mutations in genes that function in the maintenance of genomic stability. The genetic instability of tumor cells then increases with tumor progression, which allows a rapid accumulation of errors that favor their survival. Therefore, the different chromosomal changes in ESCC seen in different reports may be the results of repetitive rounds of selective expansion from different clones of tumor cells [47].

Figure 13: Copy number alteration in Chromosome 3 of ESCC cases.



Conclusion:

This is the first study to provide genome wide chromosomal alterations of esophageal cancer in a high-risk area of India. Our analysis revealed a characteristic pattern of genomic imbalances in squamous cell carcinoma tissue from patients with esophageal cancer and identified genomic regions frequently associated with tumor initiation, metastasis and high-grade disease. Advancing array technology can now provide a common platform for identifying specific chromosomal regions that undergo frequent copy number alterations as a biomarker for cancer risk assessment that can be further evaluated in larger studies. Our results of DNA copy number changes and common aberrations in ESCC are relatively concordant with published data using other techniques such as CGH, PCR, and FISH.

The most common sites for gain in ESCC are 3q and 5p. Regions on 3p, 8p, 13q and 18q are the most common sites for loss in ESCC. The target genes located at amplified regions of chromosomes such as *PLA2G5* (1p36-p34), *COL11A1* (1p21), *KCNK2* (1q41), *S100A3* (1q21), *ENAH* (1q42.12), *RGS1* (1q31), *KCNH1* (1q32-q41), *INSIG2* (2q14.1), *FGF12* (3q28), *TRIO* (5p15.2), *RNASEN* (5p15.2), *FGF10* (5p13-p12), *TLR4* (9q32-q33), *TNC* (9q33), *NTRK2* (9q22.1), *CD44* (11p13), *NCAM1* (11q23.1), *TRIM29* (11q22-q23), and *PAK1* (11q13-q14) are found to be associated with cellular migration and proliferation, tumor cell metastasis and invasion, anchorage independent growth and inhibition of apoptosis. The target genes located at deleted regions of chromosomes include *FBLN2* (3p25.1), *WNT7A* (3p25), *DLC1* (8p22), *LZTS1* (8p22), *CDKN2A* (9p21), *COL4A1* (13q34), *DACH1* (13q22) and *DCC* (18q21.3) which are found to be associated with the suppression of tumor.

The genes located at amplified regions of chromosomes were mostly involved in MAPK signaling pathway and inflammation. Genotoxic chemicals in tobacco and fermented betel quid may induce the release of inflammatory mediators via MAPK activation. Genotoxic stress as well as tissue inflammation and release of inflammatory mediators have been suggested to be key factors in carcinogenesis of gastrointestinal system. The genes located at deleted regions of chromosomes were mostly involved in Wnt signaling pathway, Focal/cell adhesion, and metabolism of xenobiotics by cytochrome P450 pathways which were reported to be associated with tobacco. Since there is wide spread use of tobacco and fermented betel quid among patients with ESCC in high risk area of India, exposure to tobacco and betel quid constituents may contribute to the development and progression of ESCC in this area by facilitating the deregulation of genes involved in these pathways.

These findings suggest that the gains and losses of chromosomal regions may contain ESCC-related oncogenes and tumor suppressor genes and provide important theoretic information for identifying and cloning novel ESCC-related oncogenes and tumor suppressor genes. Some of these regions contain cancer genes known to be involved in ESCC and others hold genes that have a known role in other cancers but which have yet to be established as esophageal cancer genes. The comprehensive knowledge gained from this analysis will enable improved strategies to prevent, diagnose and treat ESCC.

Chapter 5

Gene expression profile of esophageal squamous cell carcinoma and its association with tobacco and betel quid consumption in high-risk area of India by cDNA microarray analysis

INTRODUCTION: -

Environmental carcinogens have repeatedly been shown to affect the genetic material of host cells, leading to uncontrolled growth and ultimately malignant tumors [5]. The development of esophageal cancer is a leading example in which environmental carcinogens in addition to geographic and genetic factors appear to play major etiologic roles [1]. Esophageal cancer occurs at very high frequencies in certain parts of China, Iran, South Africa, Uruguay, France, Italy and in some regions of India [5]. The highest incidence of this cancer in India has been reported from Assam in the North-east region where it is the second leading cancer in men and third leading cancer in women [9].

Tobacco smoking, betel quid chewing, and alcohol consumption are the major known risk factors for esophageal cancer [8]. Tobacco smoke contains over 60 established carcinogens including polycyclic aromatic hydrocarbons and nitrosamines. Tobacco specific nitrosamines such as 4-(methyl nitrosamino)-1-butanone (NNK) and N'-nitrosonornicotine (NNN) that are carcinogens in smokeless tobacco, have been shown to enhance the risk of cancer development by forming adducts with DNA [144]. Betel quid chewing, a common habit in Southeast Asia has been found to increase the risk of developing esophageal cancer by 4.7-13.3 fold, although other exogenous risk factors may also be involved [8]. Betel quid usually comprises a piece of areca nut, which contains many polyphenols and several alkaloids, *Piper betle*, and lime with or without *Piper betle* leaves [145]. Arecoline, a major component of areca nut can produce 3-methyl nitrosamine propionitrile (MNPN), a potent carcinogen and safrole-like DNA adducts that have been shown to be genotoxic and mutagenic. Furthermore,

contamination of areca nuts by fungi has been reported to produce carcinogenic aflatoxins. This assumes importance since using fermented areca nut with any form of tobacco is a common habit of people in Assam and has been reported to be a potential risk factor of esophageal cancer in this region [9].

The molecular mechanisms that may lead to the development of esophageal cancer in betel quid chewers and tobacco users are unknown. Recent studies are focusing on mechanisms that can explain the carcinogenic effects of tobacco and areca nut on epithelial cell lines. Incubation of areca nut extract or arecoline with primary oral keratinocytes has been reported to promote cell survival and an inflammatory response by induction of prostaglandin E_2, interleukin-6 (IL-6) and cyclooxygenase-2 (COX-2) production via activation of MEK1/ERK/c-Fos pathway [145]. Genotoxic stress as well as tissue inflammation and release of inflammatory mediators have been suggested to be key factors in carcinogenesis of gastrointestinal systems. Genotoxic chemicals may induce the release of inflammatory mediators via Mitogen Activated Protein Kinase (MAPK) activation. Phosphorylated ERK1/2, JNK, p38 and ERK5 are reported to be significantly increased by exposure to tobacco smoke, indicating the activation of MAPK pathways [146]. NNK has recently been identified as a ligand of neuronal nicotinic acetylcholine receptors, which belong to G-protein-coupled receptors (GPCR). GPCR induces proliferation through activation of members of the family of MAPKs [147, 148].

The gene expression profile of esophageal cancer in a high incidence region of Assam where tobacco use and alcohol consumption are widespread and the users of these two substances are also betel quid chewers, has so far not been investigated. In the current study, cDNA microarray gene expression analysis was done to identify the genes

differentially expressed in esophageal cancer associated with prevalent risk factors such as tobacco use and betel quid chewing in a high-risk Indian population.

MATERIALS and METHODS: -

Collection of tumor samples: -

Endoscopic tissue biopsy specimens were taken from 16 patients at Dr. Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam. Routine histopathologic analysis was done to confirm the diagnosis. Tumor tissue and matched normal tissue distant to the tumor were collected during endoscopy in RNA later (Ambion, Austin, USA), snap-frozen in liquid nitrogen and stored at -70ºC until processed. Informed consent was obtained from all patients. Data of clinicopathologic parameters were obtained from patient's clinical records, operative notes and pathologic reports. Institutional Human Ethics Committee approved the study.

Sample preparation and chip hybridization: -

Total RNA isolation: -

Tissues were ground into powder in -196ºC liquid nitrogen and homogenized using Trizol reagent (Invitrogen Life Technologies, CA) for extraction of total RNA following the protocol of the manufacturer. Briefly, the tumor and normal tissue biopsies (<20mg) were freezed in liquid nitrogen and grind to a fine powder under liquid nitrogen by using a motor and pestle. After evaporation of liquid nitrogen, 1 ml of Trizol reagent was added to the powder without allowing the sample to thaw and tissue lysate was homogenized by Polytron Homogenizer (Kinematica, Switzerland). The lysate was centrifuged at 12,000 x g for 10 min at 4°C to remove insoluble material and the supernatant was processed. The supernatant in TRIzol was incubated for 5 min at room

temperature. 200µl of Chloroform (Sigma, Molecular Biology Grade) was added to supernatant, shake vigorously for 15 sec and incubated at room temperature for 2-3 min. The sample was centrifuged at 12,000 x g for 15 min at 4°C. The aqueous phase was transferred to a fresh tube and 0.5 ml of Isopropanol (Sigma, Molecular Biology Grade) was added to aqueous phase for precipitating the RNA. The mixture was incubated at room temperature for 10 min and centrifuged at 12,000 x g for 10 min at 4°C. The RNA should be visible on the side of the tube. The supernatant was removed and the pellet was washed by 1ml of 75% ice cold ethanol through mixing by vortexing and centrifuging at 7,500 x g for 5 min at 4°C. Air dried RNA pellet and RNA pellet was dissolved in RNA Storage Solution (1mM Sodium Citrate, pH6.4, Ambion). Sample was incubated at 55-60°C for 10-15 min to completely redissolve. RNA was stored at -70°C for long-term storage.

Qualitative and Quantitative estimation of RNA

The integrity of total RNA was checked on 1.2% Formaldehyde agarose gel electrophoresis (visual presence of 28S [5 kb] and 18S [1.9 kb] rRNA bands).

Preparation of 1.2% Formaldehyde Agarose Gel

0.6 gm of agarose (USB) was dissolved in 37 ml of 0.1% of DEPC (Sigma) water by boiling at microwave oven for 3-5 min and placed for cooling at room temperature for 3-5min. 5 ml of 10X MOPS (0.2M MOPS pH7.0, 20 mM sodium-acetate, 10 mM EDTA pH 8.0) and 8 ml of 37 % Formaldehyde (Sigma, Molecular Biology Grade) were added to dissolved agarose by shaking vigorously and kept at room temperature for 5-10 min. The gel was poured at the gel caster. 1 X MOPS was used as a gel running buffer. The gel was run at 100 volt for 20 min.

Preparation of RNA sample for 1.2% Formaldehyde Agarose Gel

2 µl of RNA sample (1µg) was incubated with 10µl of deionized Formamide (Sigma, Molecular Biology Grade), 4µl of 37% Formaldehyde, 2µl of 10 X MOPS, 2µl of DEPC water, 10X RNA dye (50% Glycerol, 10mM EDTA pH8.0, 0.25% Bromophenol blue and 0.25% Xylene cyanol FF) and 0.5µl of Ethidium bromide (10mg/ml) at 65°C for 10 min and snap-chilled on ice. RNA sample was ready to load into the gel.

RNA quantity was determined by the NanoDrop® ND-1000 UV-Vis Spectrophotometer (Nanodrop technologies, Rockland, USA). Total RNA with $OD_{260}/OD_{280}>1.8$ was used for microarray experiments.

Concentration of RNA= (O.D. at 260 nm x dilution factor x 40) µg/ml.

Experimental design: -

Total RNA was isolated from normal tissue of esophagus from all the patients involved in this study and combined to make one common control. We pooled total RNA from biopsy samples of three to four patients in all five experiments on the basis of matching the histological grade to get sufficient amount of total RNA for direct labeling. Pooled tumor RNA were labeled with Cy-3 and hybridized against the Cy-5 labeled pooled samples of normal esophageal RNA, which generated a constant control to be used on chip analyzed. **Figure 14** shows the study design. Sample pooling was done to rapidly identify tumor markers that were expressed by the majority of tumors in a population. Pooling of RNA samples isolated from tissue is a strategy that can be implemented in microarray experiments when the amount of sample RNA is limited or

when variations across the sample must be reduced. The reason behind this approach is that the concentration of an mRNA molecule in a pooled sample is likely to be closer to the average concentration for the class than the concentration in a sample from a single individual. Pooled samples have been shown to accurately reflect gene expression in individual samples and yield reproducible data [149, 150, &151].

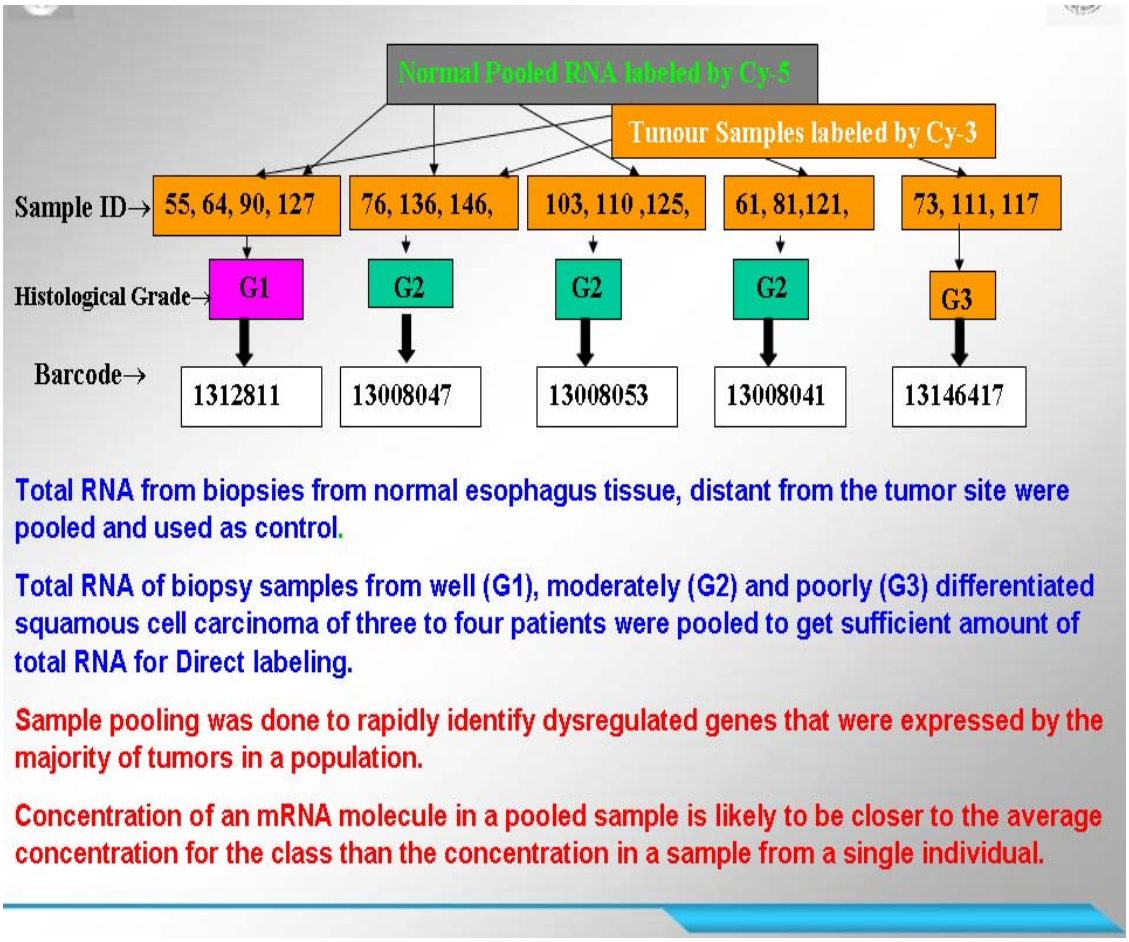


Figure 14: Experimental Design: G1, G2 and G3 indicate well-differentiated, moderately and poorly differentiated squamous cell carcinoma respectively. 13008041-53, 1312811 and 13146417 indicated barcode of microarray chips.

Labeling and hybridization: -

Twenty µg of total RNA from the tumor and matched normal tissue were labeled with Cyanine 3-dUTP and Cyanine 5-dUTP by Direct Labeling Method (PerkinElmer Life Sciences, USA: Micromax Direct labeling kit). Briefly, cDNA were synthesized by incubating 20µg of total RNA with 2µl of Primer Mix Concentrate II (FF) at 65°C for 10 minutes to denature any secondary structure in the RNA followed by cooling the reactions at 25°C for 5minutes to anneal the primers to the RNA template and incubation with 2µl of Cyanine3-dUTP or 2µl of Cyanine5-dUTP (PerkinElmer), 2.5µl of 10XRT Reaction Buffer, and 2µl of AMV RT/RNase Inhibitor Mix, at 42°C for 1 hour. Immediately following cDNA synthesis, the reaction mixture was incubated with 2.5 µl of 0.5M EDTA, pH8.0 (Ambion) and 2.5 µl of 1.0 N NaOH at 65°C for 30 minutes to hydrolyze the RNA template. The labeled cDNA was purified by Isopropyl Alcohol precipitation method. The labeled probes were denatured at 95ºC for 5 min and hybridized with Human 10K cDNA chip (University Health Network, Microarray Centre, Toronto, Canada), which contains 9,914 well-characterized human clones, in a hybridization chamber (Corning Life Sciences, USA) at 65ºC water bath for 18 hrs. Before hybridization, slides were pre-hybridized in 5XSSC, 0.1% SDS and 1% BSA solution at 65ºC for 45 minutes to prevent nonspecific hybridization. After hybridization, the slides were washed in 2XSSC with 0.1% SDS, 0.1X SSC with 0.05% SDS and 0.1XSSC sequentially for 20 min each and then spin-dried.

Microarray image analysis: -

Hybridized arrays were scanned at 5µm resolution on a Gene Pix 4200A scanner (Axon Instruments Inc. Foster City, CA) at various PMT voltage settings to obtain maximal signal intensities with <0.1% probe saturation. The Cy5-labelled cDNA were scanned at 635 nm and the Cy3-labeled cDNA samples were scanned at 532nm. The resulting TIFF images were analyzed by Gene Pix Pro 6.0.1.27 software (Axon Instrument). Both digital images were overlaid to form a pseudo colored image and a detection method was then used to determine the actual target region based on the information from both red (Cy5) and green (Cy3) pixel values. The ratios of the sample intensity to the reference intensity (green: red) for all of the targets were determined and ratio normalization was performed to normalize the center of the ratio distribution to 1.0. Image processing analysis was used for estimation of spot quality by assigning a quality score to each ratio measurement [10].

Data Analysis: -

Quality assurance: -

The data sets were imported into Microsoft Excel spreadsheets. Four parameters were used to assess quality of spots with the following features excluded: diameter < 50µm; >50% saturated pixels in both channels; < 54% of the pixels with an intensity greater than the median background intensity plus one standard deviation in either channel; flagged by Gene Pix as “not found” or “absent” or manually flagged as “bad” due to high background, misshapen features, scratches or debris on the slide undetected by Gene Pix [152].

Normalization: -

The results were normalized for labeling and detection efficiencies of the two fluorescence dyes, prior to determining differential gene expression between tumor and normal tissue samples. Intensities of selected spots were transformed into $\log_2$ (Cy3/Cy5) and data were normalized by locally weighted linear regression (LOWESS) method. LOWESS has mainly been used for correcting intensity-dependent dye-label effects between two samples on the same array. Genespring software version GX7.3.1 (Agilent Technologies, Palo Alto, CA) was used to normalize values for each gene for data analysis.

Ranking of genes: -

Data analysis was performed using Genespring Software GXV 7.3 (Agilent Technologies, Palo Alto, CA). Differentially regulated genes were ranked on the basis of signal intensity, normalized ratio, flag value and variance across replicate experiments. Top ranked genes had a higher intensity, high-normalized ratio for up and low for down, unflagged, very low variance or standard deviation. Genes were considered to be up-regulated when the median of the normalized ratio was ≥2. Genes were considered to be down regulated when the median of the normalized ratio was ≤ 0.5. Filtered genes identified to be differentially expressed by 1.5 fold or greater in three of five chips were analyzed for functional gene clusters using Genespring software GXV 7.3. Genespring used data found publicly in genomic databases to build gene ontology based on annotation information. *T-test* was performed at the 0.05% significance level to find genes that vary significantly across samples. *P*-Value or probability value is the chance of

set of genes involved in a particular function to be present in any given gene list with reference to the number of genes known to be involved in the function.

Hierarchical clustering: -

Average linkage hierarchical clustering was done using the Cluster Software version 3.0 that was written by Michael Eisen [153]. The Euclidean distance metric was used as a measure of similarity between the gene expression patterns for each pair of samples based on log-transformed ratios across all genes. The results were analyzed and visualized with the Tree View Program Version 1.50 that was also written by Michael Eisen. Those genes showing progressive fold increase or decreases in gene expression relative to normal mucosa were shown proportionally in red and green respectively. Hierarchical clustering analysis was also performed with Genespring software GXV 7.3.1. in which the average linkage and Pearson correlation (centered correlation) clustering algorithm was used. Pearson's correlation coefficient metric was used as a measure of similarity between the gene expression patterns for each pair of samples based on log-transformed ratios across all genes. Euclidean distance is best when the magnitude of the expression level is important, whereas Pearson's correlation coefficients are useful when the pattern of expression in the genes or samples is more important.

Pathway prediction analysis: -

We obtained annotations of the bioprocesses, molecular function and cellular localization using the freely available Gene Ontology and Source database [154]. The significant gene clusters were queried with known components of the biological pathways on the freely available KEGG database [155]. We also used the Biointerpreter software (http://www.genotypic.co.in/biointerpreter) for gene ontology.

Tissue Microarray (TMA) based Immunohistochemical analysis for validation of microarray data:

A TMA was constructed from formalin-fixed, paraffin-embedded blocks of 100 tissue samples that included 20 controls and 80 ESCC from high-risk population. Sampling sites were marked on the donor blocks and the tissue cylinders precisely arrayed into two recipient blocks each with a core size of 1.5 mm using a manual tissue microarrayer (Beecher Instruments, Silver Spring, MD, USA). The recipient blocks were tempered at 37^0C overnight. Multiple 5-µm sections were cut. After dewaxed in xylene and rehydrated in graded alcohol and distilled water, antigen retrieval was performed by microwave oven heating for 10 minutes at middle power in 0.01 M sodium citrate buffer (pH 6.0). Then, sections were incubated with 3% hydrogen peroxide for 5 minutes to block endogenous peroxidase activity. Nonspecific staining was blocked by 10% normal goat serum (Vector Laboratories Inc, Burlingame, CA) for 10 minutes. TMA sections were incubated overnight at 4˚C with primary antibodies of CK4 (1:100, clone 6B10, Novocastra laboratories, Norwell, MA), NPY (1:1000, clone CPON, Abcam, Cambridge, UK) and FGF12 (1:100, clone 1D9, Abnova corporation, Taiwan). The standard Streptavidin peroxidase method was employed for immunostaining. The bound antibody was detected with biotinylated anti-mouse/rabbit IgG (H+L) (Vector Laboratories Inc, Burlingame, CA) and horseradish peroxidase streptavidin (Beijing Zhongshan Biotech, Beijing, China). 3, 3'-diaminobenzidine (Maixin Biotech, Fuzhou, Fujian, China) was used as the chromogen. Slides were lightly counterstained with Mayer's hematoxylin. Non-neoplastic esophageal squamous epithelium was used as a control. The cytoplasmic staining was considered positive staining for CK4, NPY and FGF12. An adequate core

was defined as one in which the tumor occupied more than 50% of the core area. The immunohistochemical staining was assesses semi-quantitatively using a 3 point scale as no staining=0, weak/focal=1, and strong/diffuse=2. Pearson's Chi-square tests with SPSS 13.0 for Windows (SPSS Chicago, IL, USA) were performed to compare the expression of these proteins among the control and high-risk population samples. *P* value of less than 0.05 was considered statistically significant.

RESULTS: -

Clinical and epidemiological information of enrolled patients:

317 cases of esophageal cancer were registered at Dr. Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam during the period of 2004-2006. The univariate analysis revealed that tobacco chewers (OR=3.1; 95% CI: 1.8-4.2) and betel quid users (OR=2.9; 95% CI: 1.3-4.1) had higher risk than non-users (Fig.15). Principal component analysis revealed that betel nut and tobacco chewing habit contributed significantly to develop esophageal cancer (Fig.16). Sixteen esophageal biopsy samples were compared with normal pooled esophageal tissue. All patients were male and gave a history of tobacco consumption and betel nut chewing (**Table 9)**.

Figure 15

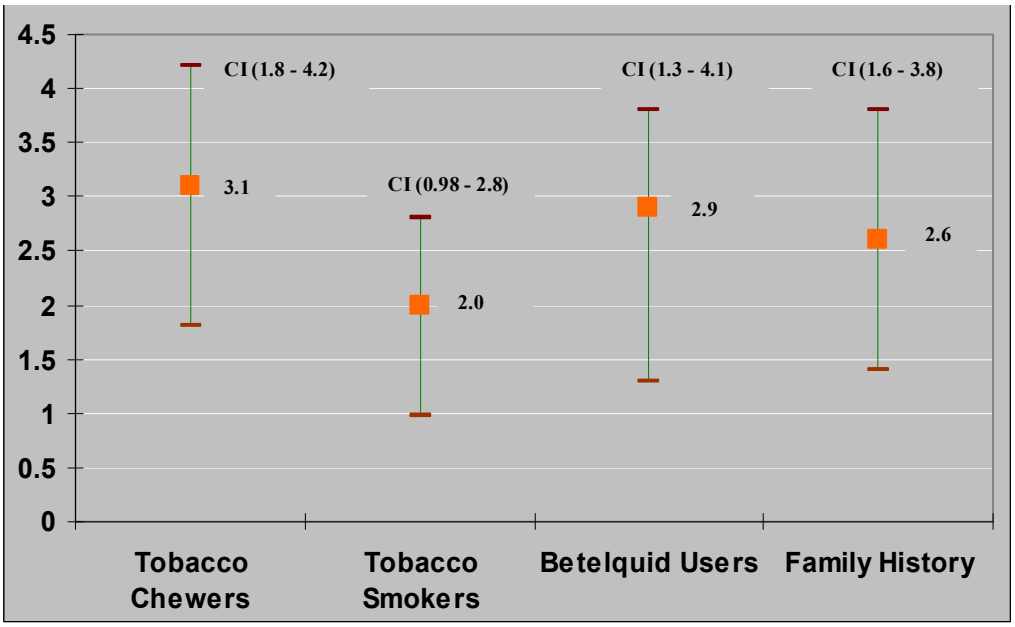


Figure15: Risk estimates of tobacco chewers, tobacco smokers, betel quid user and family history in all esophageal cancer patients

Figure 16

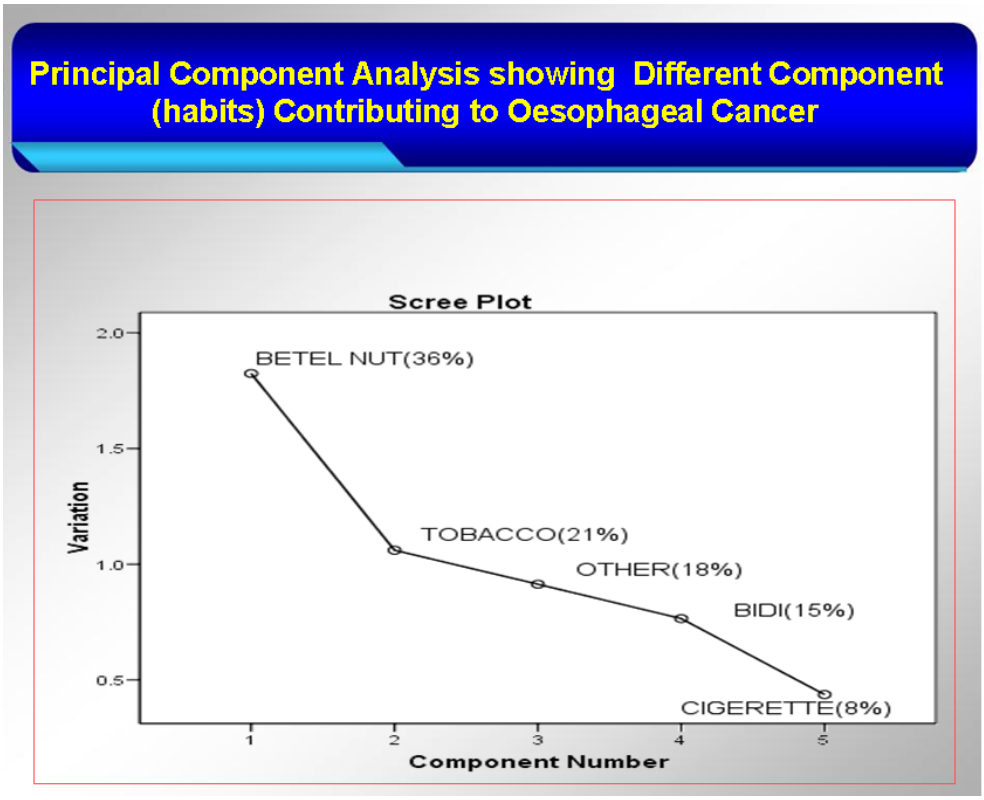


Figure 16: Principal component analysis showing different habits (components) contributing to develop esophageal cancer in North-east region of India.

Table 9: - Demographic and Clinical Characteristics of esophageal squamous cell carcinoma cases

Patient ID	Age	Sex	Tobacco/smoking habit	Alcohol use	Betel nut chewing	Pathological grade
EC-61	55	M	Yes	Yes	Yes	G2
EC-64	45	M	Yes	Yes	Yes	G1
EC-55	52	M	Yes	Yes	Yes	G1
EC-103	59	M	Yes	Yes	Yes	G2
EC-111	55	M	Yes	Yes	Yes	G3
EC-90	50	M	Yes	Yes	Yes	G1
EC-117	60	M	Yes	No	Yes	G3
EC-73	50	M	Yes	No	Yes	G3
EC-125	71	M	Yes	No	Yes	G2
EC-110	70	M	Yes	Yes	Yes	G2
EC-146	45	M	Yes	No	Yes	G2
EC-136	48	M	Yes	Yes	Yes	G2
EC-127	58	M	Yes	Yes	Yes	G1
EC-121	55	M	Yes	Yes	Yes	G2
EC-76	50	M	Yes	Yes	Yes	G2

EC-81	55	M	Yes	Yes	Yes	G1

G1= Well differentiated squamous cell carcinoma

G2= Moderately differentiated squamous cell carcinoma

G3= Poorly differentiated squamous cell carcinoma

Gene expression profiling by cDNA microarray:

An O.D.260/O.D.280 ratio > 1.8 was achieved in all samples (Table 10) and the presence of 28S and 18S rRNA bands were clearly visible in all RNA samples (Figure 17). Total RNA was also used for RT-PCR of Human *GAPDH*, *DDH* and *GSTP* gene to assess RNA quality. All the genes were robustly amplified in all samples (Figure 18). The fluorescent scanning profile of gene expression was showed in Figure 19. The scatter plot that was plotted with Cy3 and Cy5 fluorescent signal values displayed a quite disperses pattern in distribution (Figure 20).

Table 10: Quantitative and qualitative estimation of pooled RNA from tumor and normal tissue biopsies

Sample ID	Concentration µg/µl	O.D.260/O.D.280
127T/90T	3.94	1.87
117T/73T	5.82	1.95
(117+127+121+73)N	3.3	1.96
125T/110T	3.92	1.84
125N/110N	5.98	1.91
61T/121T	2.54	1.92
103T	3.96	1.80
64T/55T	4.2	1.85
61N/64N	2.7	1.8

Figure 17

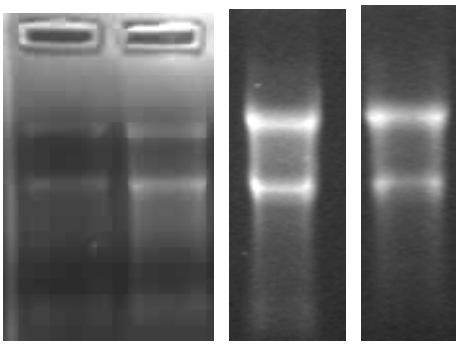


Fig.17: 1.2% Formaldehyde- Agarose Gel showing quality of RNA with presence of 28S and 18S rRNA band

Figure 18

Fig.18A (*GSTP*)

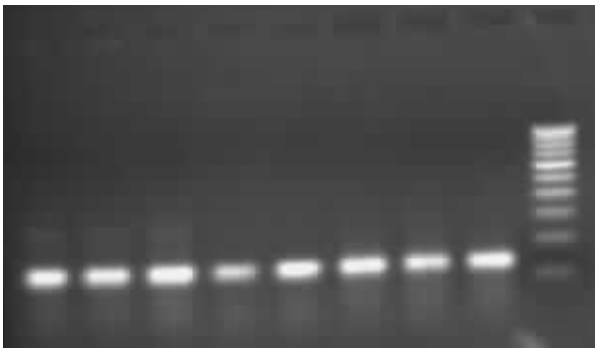


Fig.18B (*GAPDH*)

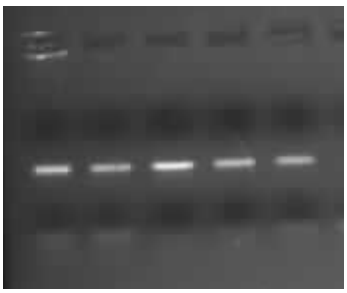


Fig.18C (*DDH*)

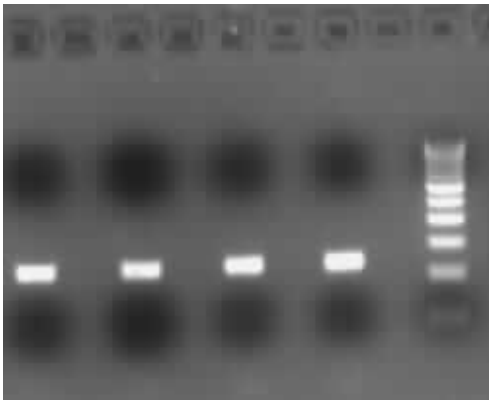


Fig.18: RT-PCR of Human *GSTP, GAPDH* and *DDH* gene to assess RNA quality

Figure 19

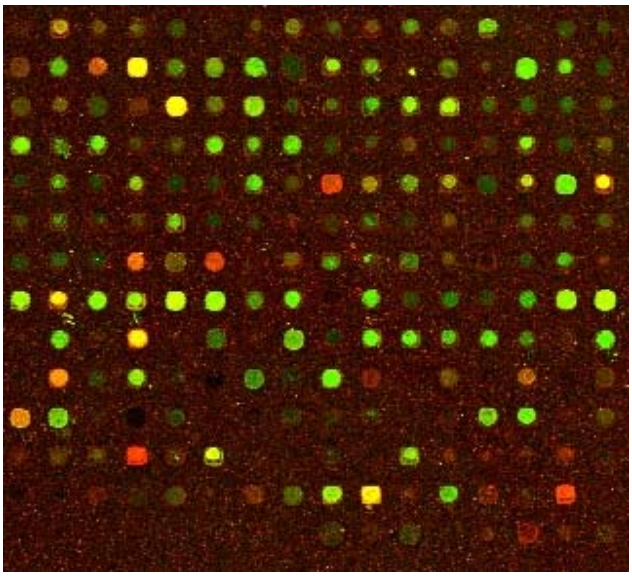


Figure 19: Fluorescent scanning profile of gene expression

Figure 20

Fig.20A

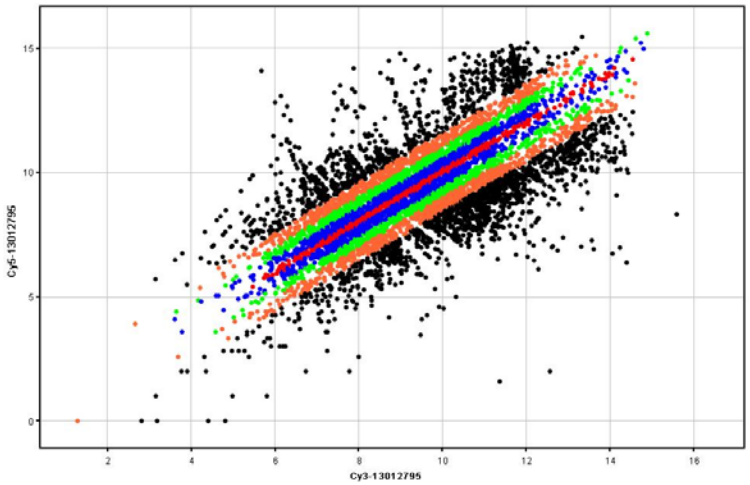


Fig.20B

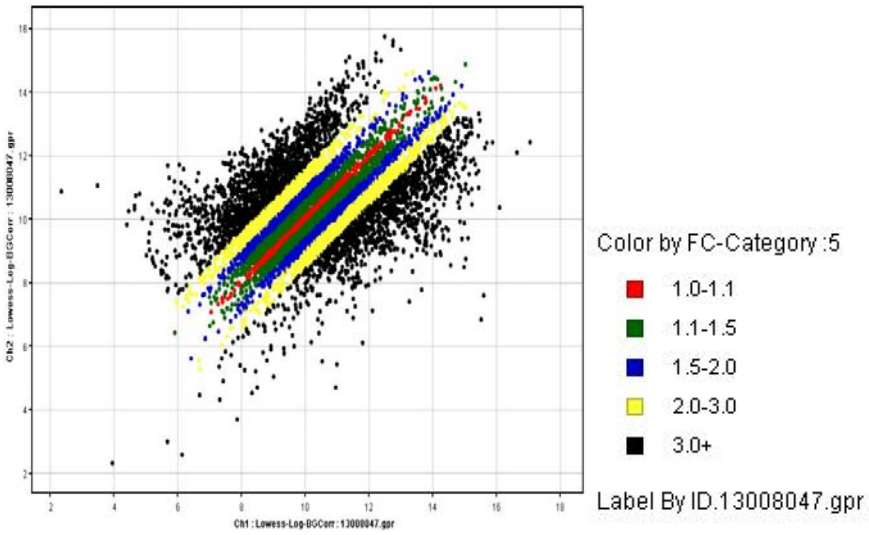


Figure 20: Scatter plots of gene expression pattern

Gene expression was measured using microarrays to detect changes in tumor sample compared to normal. Nine hundred and twenty three genes were differentially expressed at least 1 folds in 3 out of 5 experiments. Of these, 611 genes were up regulated and 312 genes were down regulated. The scaled data generated from Gene Pix Pro 6.0.1.27 software were imported into GeneSpring for fold change analysis, filtering, and cluster analysis. Hierarchical Clustering analysis of 923 genes selected from the 10,000-gene set was performed (**Figure 21**). The two dimensional hierarchical clustering showed that the 66% (611 genes) of the differentially expressed genes were significantly up regulated, whereas 34% of genes (312 genes) showed down-regulation **(Figure 22**).

Figure 21

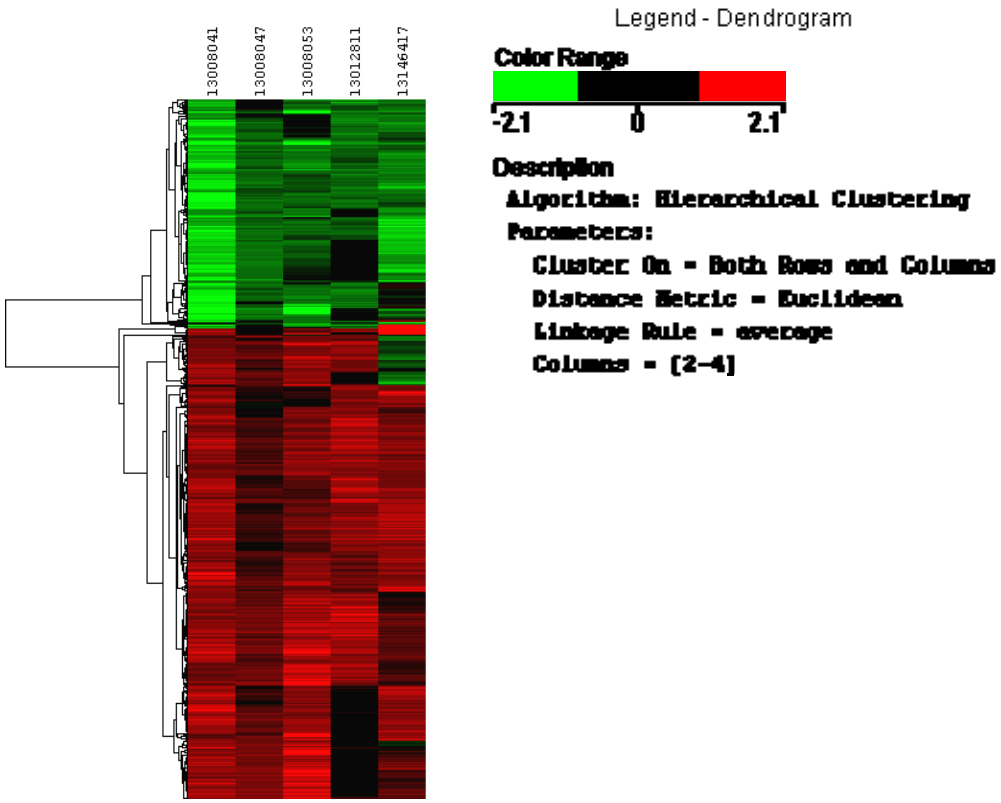


Figure 21: - Hierarchical Clustering (Average Linkage Clustering) of the genes that were over or under expressed in tumor versus normal tissue in the 16 ESCC patients

Figure 22

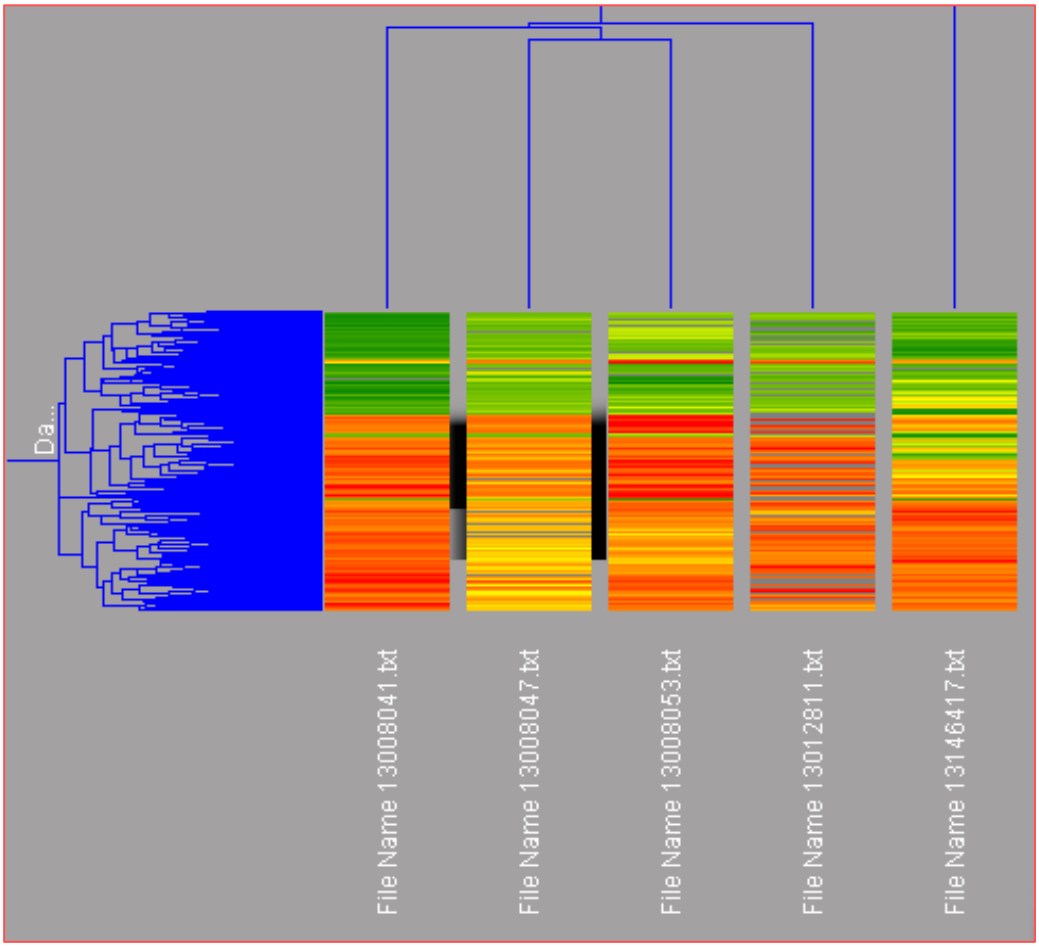


Figure 22: Two-way unsupervised hierarchical clustering (average linkage clustering) of the 923 differentially expressed genes that were over or under expressed in tumor versus normal tissue of 16 ESCC patients. Red and green colors indicate up-regulated and down-regulated gene expression respectively.

Condition tree (Fig.23) is a statistical tree that groups similar experiments under one branch. This tree can be used for quality control of experiments.

Figure 23

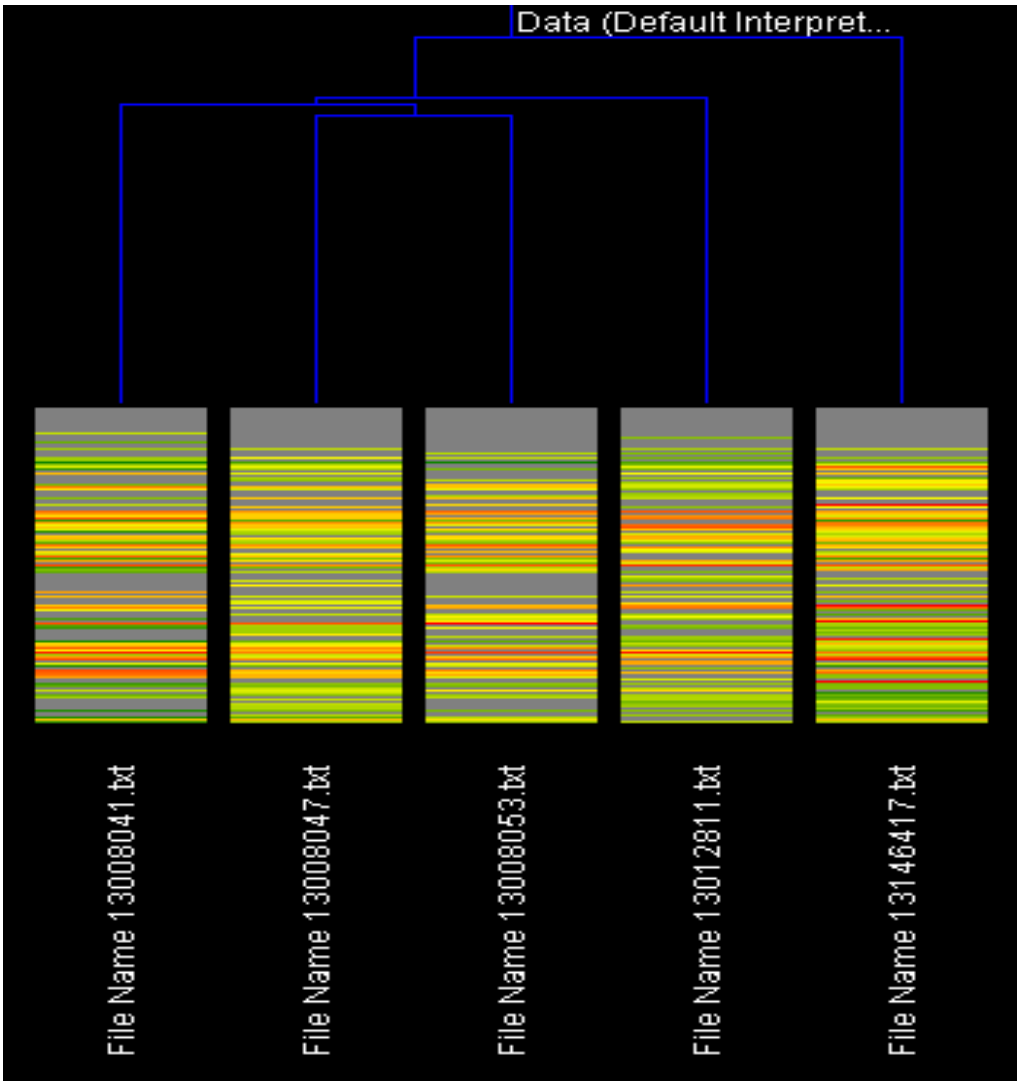


Figure 23: Conditional Tree of all the Experiments. Overview of all 5 experiments including flags.

Box Plot (Figure 24) displayed log ratio distribution in the four experiments investigating differentially expressed genes in esophageal cancer. Spots with extreme log ratios represent genes that appear to be strongly differentially expressed.

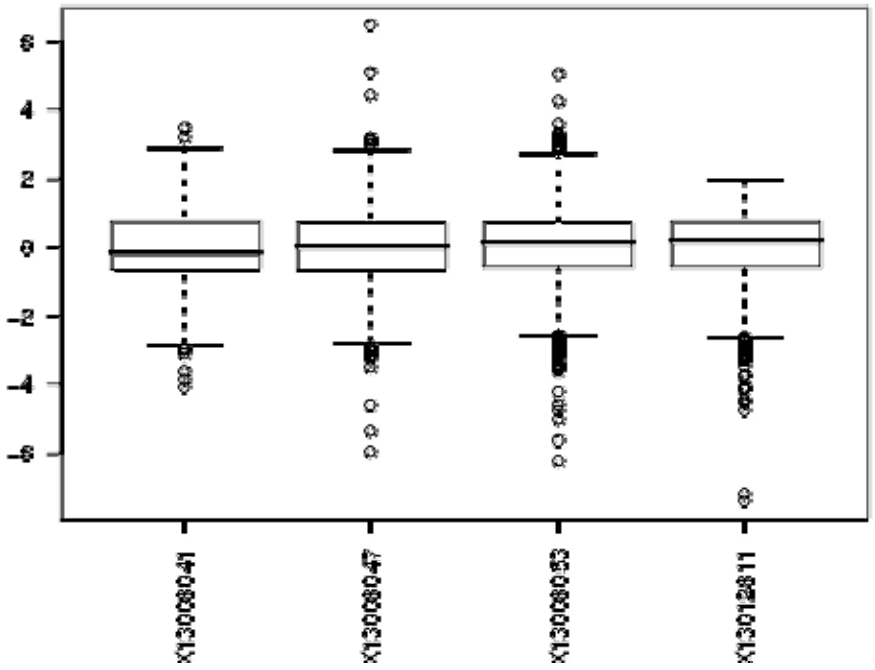


Figure 24: Density Distributions of selected gene expression experiments using Box Plot.

To identify highly reproducible changes, data were filtered based on select criteria. Transcripts modulated by minimum 1.5 fold in at least three of five chips were used for further analyses. However, genes that had ≥1.5 fold cutoff or had a *P* value of ≤ 0.05 were also included for analysis to look for subtle changes in gene expression. Using stringent criteria ($P \leq 0.05$ and ≥1.5 fold change), 127 differentially expressed genes (87 up regulated and 40 down regulated) were identified in tumor tissue. Using the Gene

Ontology database, we categorized the 923 differentially expressed genes into known or probable functional categories. Genes involved in dimethylallyltransferase activity and farnesyltransferase activity (*CTLA4*), cation antiporter activity (*SLC9A2*) and cation transporter activity (*KCNN2*, *SLC30A4, KCNJ15, CACNA2D3*), G-protein coupled receptor (GPCR) activity (*GPR87, NPY*), MAPK signaling pathway (*FGF12*), and protein serine or threonine kinase activity (*GRK4*) were significantly up regulated (**Table 11**). Out of 87 up regulated genes, two genes involved in GPCR group, five genes involved in cation transporter activity, one gene each involved in protein serine or threonine kinase activity and MAPK activity. Gene involved in anti-apoptosis activity (*BIRC1*), and cellular proliferation (*EGR2*) were also significantly up-regulated. Genes involved in structural constituent of the ribosome (*RPL32*, *RPS4X),* structural constituent of cytoskeleton (*KRT17*, *KRT4*, *PLA2G1B*), cysteine protease inhibitor activity (*CSTB*, *CSTA*), anti-oxidant activity (*PRDX6*), acyl groups transferase activity (*TGM3*), and translation elongation (*EEF1A1*) were significantly down-regulated (**Table 12**). Gene involved in humoral immune response (*CD24*) and base-excision repair (*MPG*) were significantly down regulated. Out of 40 down regulated genes, five genes were involved in structural constituent of the ribosome, four genes were involved in structural constituent of cytoskeleton, two genes were involved in cysteine protease inhibitor activity and one gene each was involved in anti-oxidant activity, acyl group transferase activity and translation elongation.

Table 11: -List of Significantly Up-Regulated Genes in Esophageal Cancer Patients

Category	Genes in Category	% Genes in Category	Genes in List in Category	% Genes in List Category	*P*-value
GO: 4930: G-protein coupled receptor activity	150	2.148	23	3.433	**0.0156**
GO: 8324: cation transporter activity	307	4.396	39	5.821	**0.0404**
GO: 4674: protein serine/threonine kinase activity	285	4.081	38	5.672	**0.022**
GO: 4707: MAP kinase activity	20	0.286	5	0.746	**0.0368**
GO: 4161: dimethylallyltranstransferase activity	4	0.0573	3	0.448	**0.00327**

Table 12: - List of Significantly Down-Regulated Genes in Esophageal Cancer Patients

Category	Genes in Category	% Genes in Category	Genes in List in Category	% Genes in List Category	*P*-value
GO: 3735: structural constituent of ribosome	115	1.647	32	4.992	**5.54E-09**
GO: 4869: cysteine protease inhibitor activity	27	0.387	11	1.716	**1.19E-05**
GO: 5200: structural constituent of cytoskeleton	63	0.902	12	1.872	**0.011**
GO: 16746: transferase activity,	90	1.289	19	2.964	**0.000422**

transferring acyl groups					
GO: 16209: antioxidant activity	25	0.358	9	1.404	**0.000232**
GO: 3746: translation elongation factor activity	14	0.2	5	0.78	**0.00636**

Functional category of differentially regulated genes revealed that genes involved in immune response category were significantly down-regulated (Fig.25).

Figure 25

Fig.25A

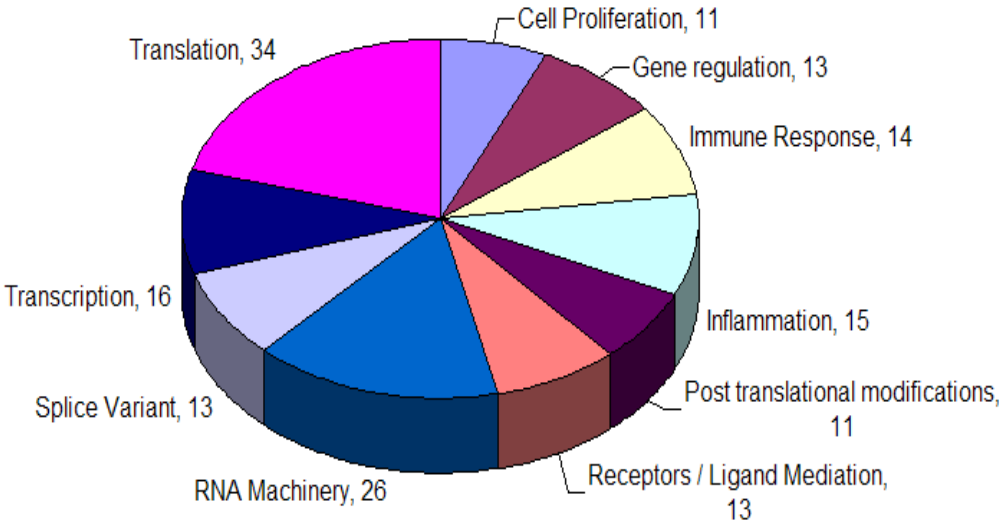


Figure 25A: Significant functional classes of down regulated genes

Fig.25B

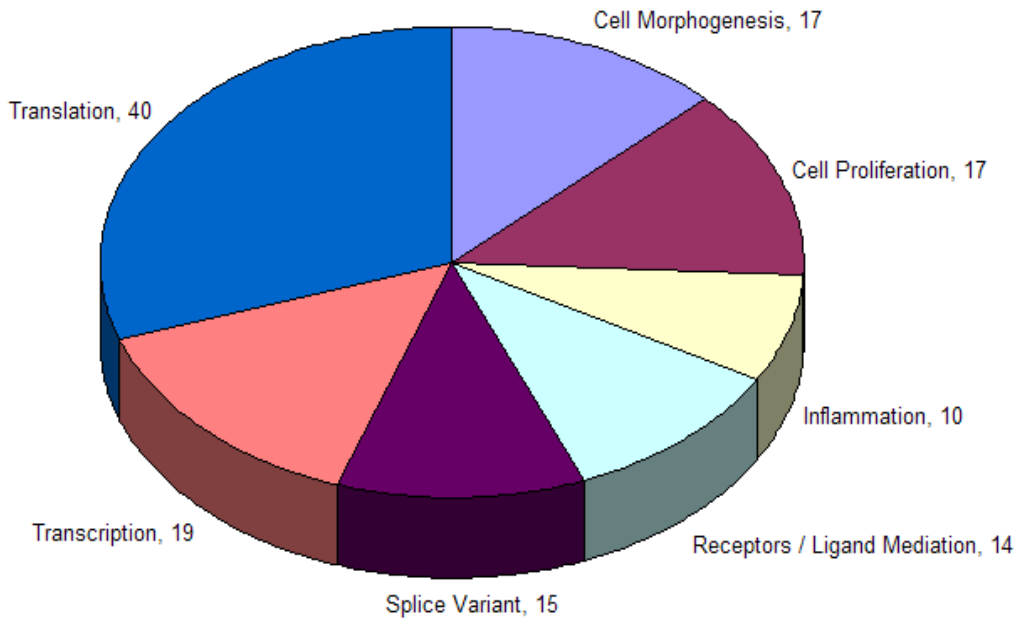


Fig.25B: Significant functional classes of up regulated genes

TMA based Immunohistochemical analysis:

Differential expressions of *KRT4, NPY* and *FGF12* genes identified in cDNA microarray analysis were validated at protein level using immunohistochemistry on TMA. The immunohistochemical staining was assessed semi-quantitatively using a 3 point scale as no staining=0, weak/focal=1, and strong/diffuse=2. Expression of KRT4 ($P<0.001$) was found only in overlying non-neoplastic epithelium and absent in tumor cells (52 out of 80 showing a score of 0 while remaining 28 score of 1). NPY ($P<0.001$) showed expression in tumor cells with score of 0 in 2 cases, with score of 1 in 19 (24%) cases and with score of 2 in 59 (74%) cases. NPY showed expression in normal epithelium with score of 1 (Fig.26). FGF12 ($P=0.001$) showed expression in tumor cells with score of 0 in 2 cases, with a score of 1 in 35 cases (44%) and with a score of 2 in 43 cases (53%). FGF12 showed expression in normal epithelium with a score of 0 in 3 controls and with a score of 1 in 17 controls (Fig.27).

Figure26

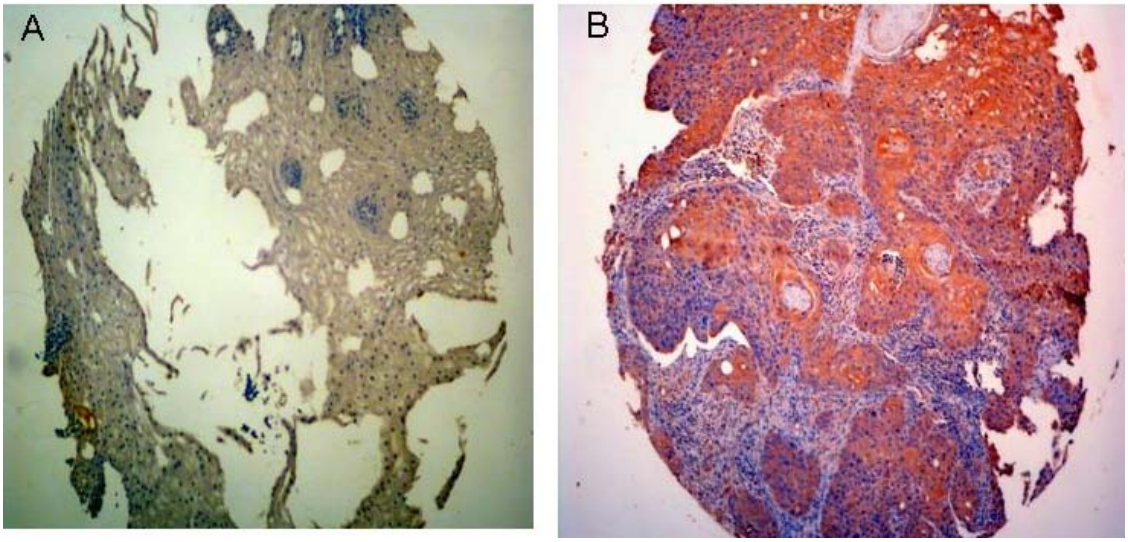


Figure 26: TMA cores showing weak immunostaining for NPY in control (A) and strong staining in ESCC (B).

Figure 27

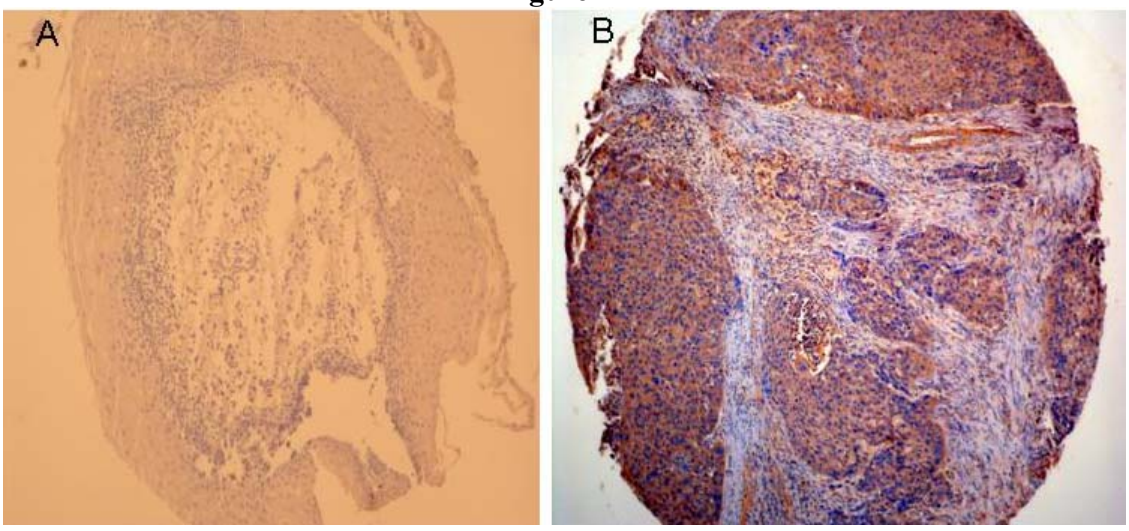


Figure 27: TMA cores showing no immunostaining for FGF12 in control (A) and strong staining in ESCC (B).

DISCUSSION: -

Several tobacco constituents, including nitrosamines, polycyclic aromatic hydrocarbons, aromatic amines, various aldehydes and phenols, may be causally related to esophageal cancer [5]. A previous report has shown that betel quid chewing with or without tobacco consumption is associated with the development of esophageal cancer in Assam [9]. However there are very few studies of gene expression profiles of tumors that may be associated with betel quid chewing as well as tobacco consumption. The current study was aimed to analyze genes and pathways that may be involved in tobacco and betel quid chewing related esophageal malignancies in this high incidence region of India.

Gene expression levels showed that there were five different molecular functional pathways that were most significantly up regulated and four different molecular functional pathways that were most significantly down regulated (**Table 13**). Some of the significantly over expressed molecular functional pathways like MAPK signaling pathway, G-protein coupled receptor and cation transporter activity have earlier been reported in esophageal and other cancers. However, genes such as *CTLA4* (involved in dimethylallyltransferase activity and farnesyltransferase activity) and *NPY*, *FGF12, KCNN2,* and *KCNJ15* were found to be significantly up regulated in our study and have not been reported earlier. Genes involved in structural constituent of ribosome (*RPL32, RPS4X, RPL7A,)* and anti-oxidant activity (*PRDX6*) that were found to be significantly down regulated in our study, have also not been reported earlier. Some of the significantly down-regulated molecular functional pathways like structural constituent of cytoskeleton (*KRT17*, *KRT4),* acyl groups transferase activity (*TGM3)* and cysteine

protease inhibitor activity (*CSTB*, *CSTA*), have already been reported previously in esophageal carcinoma. The data may be used for selection of a limited number of markers that can be screened in large populations by RT-PCR. All the genes identified here are of interest because of their potential role in the natural history of esophageal squamous cell carcinoma.

Table 13: Functionally category of biologically relevant genes in ESCC

Functional category of Up regulated Genes:	**Functional category of Down regulated Genes:**
GPCR activity (*GPR87,NPY*)	Structural Constituent of Ribosome (*RPS4X*)
Cation transporter activity (*SLC9A2, SLC30A4*)	Structural Constituent of Cytoskeleton (*KRT4*)
Serine or threonine kinase activity (*GRK4*)	Acyl groups transferase activity (*TGM3*)
MAPK signaling pathway (*MAPK8, FGF12*)	Cysteine protease inhibitor activity (*CSTB*)
Anti-apoptosis activity (*BIRC1*)	

Mitogen Activated signaling cascade (Fig.28) showed significant up-regulation in our study. Activated MAPK pathway has been detected in many human tumors including carcinomas of the breast, colon, kidney, and lung suggesting the possibility that MAPK may play a major role in tumor progression and metastasis. MAPK induces proteolytic enzymes that degrade the basement membrane, enhances cell migration, initiates several

pro-survival genes and maintains growth [156]. Oxidants in cigarette smoke have previously been reported to activate MAPK signaling cascades in lung epithelial cells in vitro and in vivo. These signaling pathways lead to the enhanced ability of Jun and Fos family members to activate transcription of a number of AP-1 dependent target genes involved in cell proliferation, differentiation, and inflammation [157]. Phosphorylated ERK 1/2, JNK, p38 and ERK5 genes were significantly increased upon exposure to tobacco smoke, indicating the activation of MAPK pathways [146]. Benzo (a) pyrene quinines, which are the non-volatile component of cigarette smoke, have also been reported to increase cell proliferation, generate reactive oxygen species, and transactivate the epidermal growth factor receptor in breast epithelial cells [158].

Figure 28

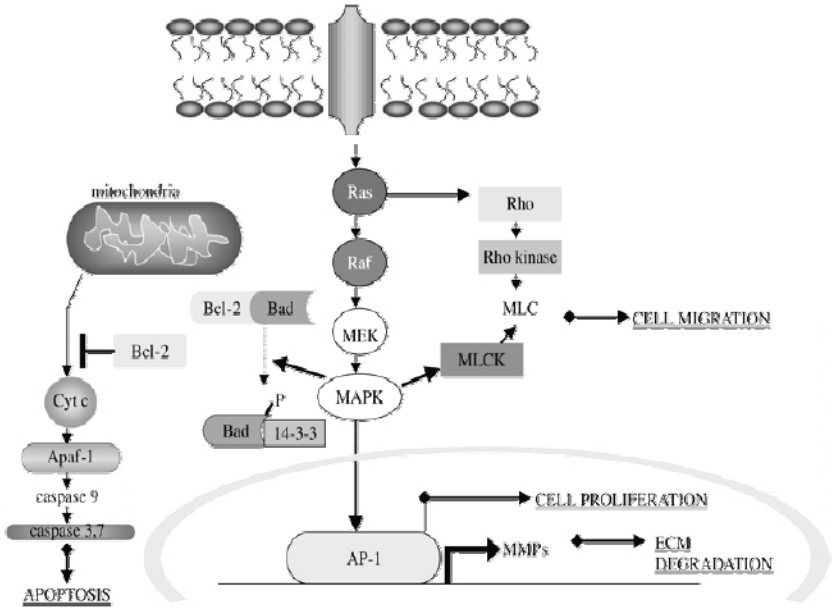


Figure 28: MAPK Pathway

TNF- alpha was found to be up regulated in our study. Cigarette smoke is known to enhance the induction of TNF- alpha by differentiated macrophage that is regulated

primarily via the ERK ½ pathway [159]. Tumor necrosis factor receptor super family, member 17 has been shown to specifically bind to the tumor necrosis factor (ligand) super family, member 13b (TNFSF13B/TALL-1/BAFF) that leads to activation of NF-kappaB and MAPK8/JNK.

NeuropeptideY (NPY), which belongs to G-protein coupled receptor family, was found to be significantly up regulated in our study. Aberrant NPY expressions are early events in prostrate cancer development and are associated with poor prognosis [160]. Activation of the Y1 receptor by NPY regulates the growth of prostate cancer cells. Estrogen up regulates NPY receptor expression in a human breast cancer cell line [161].

Ion channel regulating genes showed significant up-regulation in our study. Several tumors such as prostrate, uterus, glial cells, stomach, pancreas, breast and colorectum are known to express Ca^{2+}activated K^+ channels. The complex process of carcinogenesis is triggered by oncogenic pathways via activation of K^+ channels. For example, p21 ras and the Raf kinase are known to induce oncogenic transformation via activation of Ca^{2+}dependent K^+ channels. Enhanced cell migration and tumor metastasis are associated with fluctuations in the activity of membrane transporters and ion channels that require K^+ channel activity [162]. Genes involved in calcium regulation and calcium signaling also seemed to be important in the progression of esophageal squamous dysplasia [163].

Transglutaminase-3 (TGase-3) that showed significant down regulation in our study has been reported in earlier studies. *TGM3* is a Ca^{2+}dependent enzyme that catalyzes covalent cross-linking reactions between proteins or peptides by glutamyl lysine isopeptide bonds and is important for effective epithelial barrier formation [7].

TGase-3 has been implicated in the formation and assembly of the cornified cell envelope of the epidermis, hair follicle and perhaps other stratified squamous epithelia. Alterated TGase-3 expression is a common event in human esophageal cancer [164]. The lowest levels of TGases 3 and 7 have been reported in patients with metastatic disease [165]. Tissue transglutaminase (tTG) is a high level phenotypic biomarker that is down regulated in prostate cancer [166]. Down regulation of *TGM3* might lead to poor epithelial barrier formation, which may promote tumor invasion easily [7].

Intermediate filaments form the cytoskeleton of cells and maintain the integrity of cell. Keratins (*KRT4* and *KRT8*) showed significant down regulation in our study. Over expression of epithelial cell intermediate filaments and their isoforms (*KRT8)* have been reported earlier in colorectal polyp and cancer [167].

Reduced expression of cystatin B as found in our study, has earlier been reported to be associated with lymph node metastasis and may therefore prove to be a useful marker for predicting the biologic aggressiveness of esophageal cancer [168]. Over expression of cystatin A has been shown to inhibit tumor cell growth, angiogenesis, invasion, and metastasis in esophageal cancer [37]. High levels of cystatin A and cystatin B have been reported to correlate with more favorable patient outcome in breast, lung and head and neck tumors [37, 168].

A group of ribosomal proteins (Figure 29) may function as cell cycle checkpoints and compose a new family of cell proliferation regulators. They play an important role in translational regulation and control of cellular transformation, tumor growth, aggressiveness and metastasis. Thus in addition to protein synthesis, they are involved in neoplastic transformation of cells. Ribosomal proteins were found to be significantly

down regulated in our study. Expression of human ribosomal protein L13a has been shown to induce apoptosis, presumably by arresting cell growth in the G2/M phase of the cell cycle. In addition, a closely related ribosomal protein, L7, arrests cells in G1 and also induces apoptosis [154]. Mitochondrial ribosomal protein L41 suppresses cell growth in association with p53 and p27Kip1. *MRPL41* is reported to be either expressed at reduced levels or absent in most tumor types and cell lines [169]. Human apurinic apyrimidinic endonuclease (RPLP0) and its N terminal truncated form (AN34) are involved in DNA fragmentation during apoptosis**.** Down regulation of apurinic apyrimidinic endonuclease (RPLP0) expression is associated with the induction of apoptosis in differentiating myeloid leukemia cells [170]. Simultaneously, three ribosomal proteins (RPL10, RPL32, and RPS16) also showed up-regulation in C81 cells. Over expression of several ribosomal proteins including RPS16 has been reported in colon, breast, liver, and pancreatic tumors [171].

Fig.29

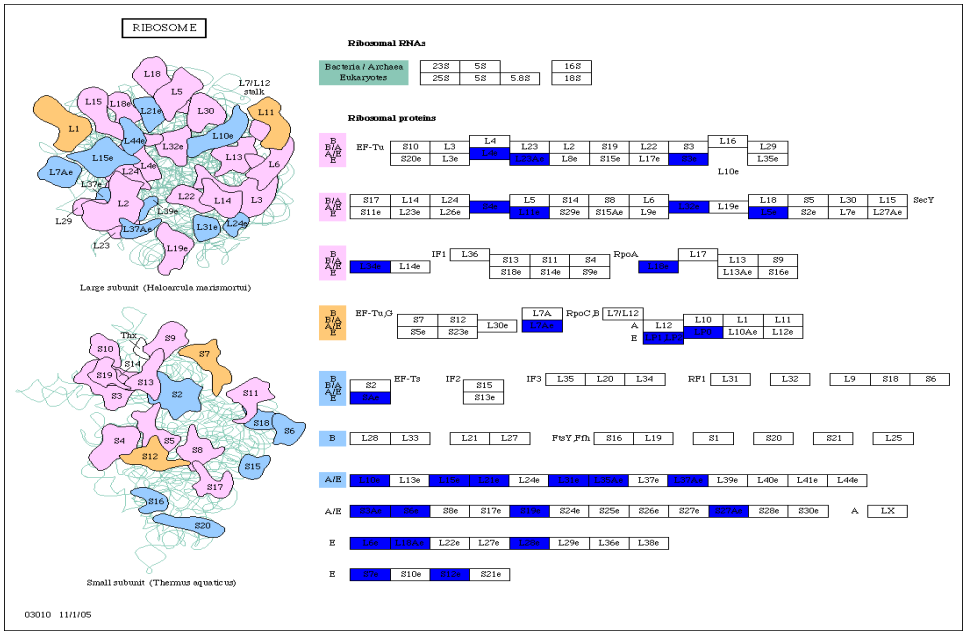
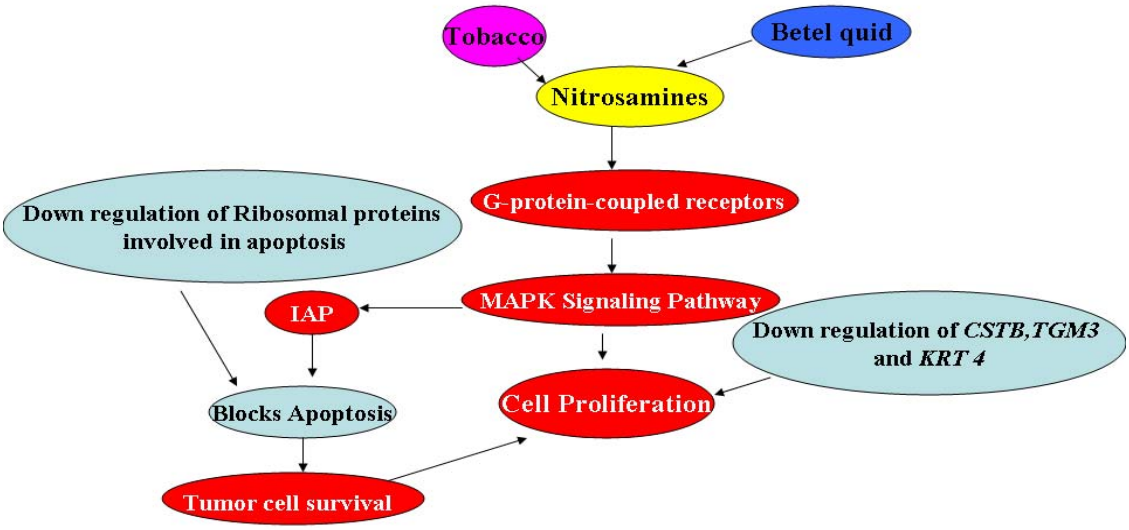


Figure 29. Ribosomal biosynthesis pathway.

This is the first study to provide gene expression profile of esophageal cancer in a high-risk region of India. The most salient finding was identification of down regulated genes involved in structural constituents of ribosome and up regulated genes involved in cation transporter activity (Fig.30). In a similar study of gene expression profile in a high-risk area of China, Taylor et al and Liu et al have reported down regulation of *CSTB*, *CSTA*, *KRT4*, and *TGM3* and up-regulation of G-coupled signaling, ion transport activity and MAPK activity [6,10,54,163]. Genotoxic chemicals in tobacco and fermented betel quid may induce the release of inflammatory mediators via MAPK activation. Genotoxic stress as well as tissue inflammation and release of inflammatory mediators have been suggested to be key factors in carcinogenesis of gastrointestinal system [146]. Since there is wide spread use of tobacco and fermented betel quid among patients with ESCC in

high risk area of India, exposure to tobacco and betel quid constituents may contribute to the development and progression of ESCC in this area by facilitating the deregulation of genes involved in these pathways (Fig.30). In a recent study from low risk area of India, deregulation of genes associated with zinc homeostasis in esophageal squamous cell carcinoma (ESCC) has been reported [4]. These data indicate the consistency of molecular profiles of esophageal cancer in two different geographic locations that have a high incidence of ESCC but different food habits and customs.

Fig.30



Schematic illustration of functional pathway for tobacco and betel quid associated esophageal tumorigenesis.

Figure 30: Possible molecular pathogenesis involved in tobacco and betel quid associated esophageal tumorigenesis in high-risk area of India.

Chapter 6

Molecular profiling to identify molecular mechanism in esophageal cancer with familial clustering

Introduction:

Esophageal cancer is among the ten most common malignancies worldwide and ranks as the sixth leading cause of death from cancer [50]. Esophageal cancer occurs at very high frequencies in certain parts of China, Iran, South Africa, Uruguay, France, and Italy [5]. Association of family history with an increased risk of esophageal cancer has been reported in several case-control and cohort studies from China, Iran and Japan suggesting possible role of environmental as well as genetic factors [10, 11]. A high prevalence of esophageal cancer with familial aggregation has also been reported from Assam in the Northeast (NE) region of India with an Age-adjusted rate (AAR) of 33 /100,000 males [172, 173].

Familial clustering of cancer may be due to shared environmental factors or shared genes by family members [16]. Hua Su et al reported that molecular profiles in esophageal squamous cell carcinoma (ESCC) were highly consistent and expression patterns in familial cases were different from those in sporadic cases [10]. We have earlier reported that gene expression profile of non-familial ESCC in Assam, a high risk zone for esophageal cancer in India, are highly consistent with ESCC in China [173].

In the current study, molecular signature of ESCC from high risk area of India has been studied by gene expression profiling in esophageal cancer patients with family history of esophageal cancer with the aim to elucidate molecular pathogenesis of esophageal cancer in these patients.

Materials and methods

Selection of patients and collection of samples:

Among 317 cases of esophageal cancer registered at Dr. Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam, 92 (29%) cases had family history of esophageal and other cancers beside habit of tobacco and betel quid chewing. Among 92 patients, 45 patients (49%) had family history of esophageal cancer. Patients diagnosed with metastases and at advanced stage of the disease were excluded from the study. Endoscopic biopsy specimens from tumor and matched normal tissue distant to tumor were collected during diagnostic endoscopy. Part of tumor and normal tissues was preserved in formalin for histopathologic diagnosis/confirmation, and remaining tissue was immediately immersed in RNA later solution (Ambion, Austin, USA) and stored at -70ºC until processed. Out of 45 patients, 20 patients with tumor biopsies containing more than 80% tumor were included for gene expression study. Demographic and lifestyle cancer risk factors and clinical data of all twenty patients were collected (Table 14). Informed consent was obtained from all the patients to use their surgical specimens and clinicopathologic data for the study. Institutional Human Ethics Committee had approved the study.

Table 14: - Demographic and clinical characteristics of esophageal squamous cell carcinoma cases with family history of esophageal cancer

Patient ID	Age	Sex	Tobacco chewing habit	Smoking habit	Alcohol use	Betel quid use	Family history of esophageal cancer	Pathological grade†	Experiments carried out‡
EC-116	40	F	Yes	No	No	Yes	Father	G2	MA, RT
EC-88	50	M	No	Yes	Yes	Yes	Cousin brother	G3	MA, RT
EC-99	50	M	No	Yes	No	Yes	Mother	G2	MA, RT
EC-84	50	M	No	Yes	Yes	Yes	Father	G1	MA, RT
EC-53	54	M	Yes	No	Yes	Yes	Elder Brother	G3	MA, RT
EC-69	70	M	Yes	Yes	No	Yes	Mother	G1	MA, RT
EC-124	69	M	No	Yes	Yes	Yes	Brother	G2	MA, RT
EC-129	65	M	Yes	Yes	No	Yes	Sister	G1	MA, RT,
EC-83	52	F	No	No	No	Yes	Father	G3	MA, RT
EC-54	45	M	No	Yes	Yes	Yes	Mother	G2	RT

EC-72	32	M	No	Yes	No	Yes	Father	G2	RT
EC-247	45	F	No	No	No	Yes	Mother	G2	RT
EC-248	56	M	Yes	Yes	Yes	Yes	Brother	G2	RT
EC-283	55	F	Yes	No	No	Yes	Father	G2	RT
EC-291	55	M	No	Yes	No	Yes	Brother	G2	RT
EC-65	50	F	No	No	No	Yes	Parental Uncle	G1	RT
EC-217	56	M	Yes	Yes	Yes	Yes	Elder Brother	G1	RT
EC-244	45	M	Yes	Yes	Yes	Yes	Elder brother/ Father	G2	RT
EC-85	85	M	No	Yes	No	Yes	Son	G3	RT
EC-187	58	M	Yes	Yes	Yes	Yes	Father	G2	RT

Footnotes† G1= Well differentiated squamous cell carcinoma

G2= moderately differentiated squamous cell carcinoma

G3= poorly differentiated squamous cell carcinoma

‡MA= cDNA microarry

‡ RT= Real-Time PCR for validation of microarray data.

Microarray experiments:

Isolation of RNA

Total RNA was isolated from snap-frozen biopsies using the Qiagen (Valencia, CA) RNeasy Mini kit. The tumor and normal tissue biopsies (<20mg) were freezed in liquid nitrogen and grind to a fine powder under liquid nitrogen by using a motor and pestle. After evaporation of liquid nitrogen, 600µl of RLT Buffer with 5% (V/V) of β-Mercaptoethanol was added to the powder without allowing the sample to thaw and tissue lysate was homogenized by Polytron Homogenizer (Kinematica, Switzerland). The lysate was centrifuged for 3min at 14,000 rpm. The supernatant was carefully removed by pipetting and transferred it to a new microcentrifuge tube. 600µl of 70% ethanol was added to the supernatant and mixed immediately by pipetting. The mixture was transferred to an RNeasy spin column that was placed in a 2ml collection tube and was centrifuged for 30 second at 10,000 rpm. The flow-through was discarded. 700µl of RW1 Buffer was added to the RNeasy spin column and the column was centrifuged for 30sec at 10,000 rpm to wash the spin column membrane. The flow-through was discarded. 500µl of Buffer RPE was added to the RNeasy spin column and the column was centrifuged for 30sec at 10,000 rpm to wash the spin column membrane. The RNeasy spin column membrane was again washed by 500µl of Buffer RPE for 2min. The RNeasy spin column was placed in a new 2ml collection tube and the column was centrifuged at 14,000 rpm for 5min. The RNeasy spin column was placed in a new 1.5ml collection tube and 30µl of RNase-free water was directly added to the spin column membrane. The column was centrifuged for 1min at 10,000 rpm to elute the RNA.

Qualitative and Quantitative estimation of RNA

RNA integrity was examined using the RNA 6000 Nano LabChip on the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). The Agilent 2100 bioanalyzer is a small bench top system that uses LabChip technology from Caliper Technologies Corp. to integrate sample separation, detection, quantification, and data analysis. Each disposable RNA chip is used to determine the concentration and purity/integrity of 12 RNA samples with a total analysis time of about 25 minutes. The 18S and 28S ribosomal RNA peaks are identified by the Agilent 2100 bioanalyzer software and dominate the electropherogram. The Agilent 2100 biosizing software includes data collection, presentation and interpretation functions. Data can be displayed as a gel-like image and/or as electropherogram(s). For each sample, the ratio of the two major ribosomal RNA bands is automatically determined and displayed with the RNA quantitation data on the electropherogram. For the RNA applications, the instrument uses fluorescence detection, monitoring the fluorescence between 670 nm and 700 nm. In order to standardize the process of RNA integrity interpretation, Agilent Technologies has introduced a new tool for RNA quality assessment. The RNA Integrity Number (RIN) was developed to remove individual interpretation in RNA quality control. It takes the entire electrophoretic trace into account. The RIN software algorithm allows for the classification of eukaryotic total RNA, based on a numbering system from 1 to 10, with 1 being the most degraded profile and 10 being the most intact. The RIN software algorithm was developed for samples acquired with the Eukaryote Total RNA Nano assay on the Agilent 2100 bioanalyzer.

All chips were prepared according to the chip preparation protocol provided with the RNA 6000 LabChip kit. The kit includes 25 chips, syringe, spin filters and the following reagents: sample buffer, gel matrix, and dye concentrate. In addition, the custom-made RNA 6000 ladder for use with the instrument/assay was purchased from Ambion, Inc. The gel-dye mix was prepared by mixing 400 µl of the gel matrix with 4 µl of the dye concentrate and filtering through a spin filter. The chip was filled with the gel-dye mixture and 5 µl of the sample buffer was then added to each sample well. 12 samples (1 µl each) in the concentration range from 10 to 500 ng/ µl were loaded into the sample wells of the chip. Finally, 1 µl of the RNA ladder was loaded into the assigned ladder well. The chip was then vortexed, and run on the Agilent 2100 bioanalyzer. The RNA 6000 ladder standard is run on every chip from a specified ladder well and is used as a reference for data analysis. The RNA 6000 ladder contains six RNA fragments ranging in size from 0.2 to 6 kb at a total concentration of 150 ng/ µl. The software automatically compares the unknown samples (1–12) to the ladder fragments to determine the concentration of the unknown samples and to identify the ribosomal RNA peaks. The ladder also serves as a built-in quality control measurement of system performance under standard conditions.

RNA quantity was determined by the NanoDrop® ND-1000 UV-Vis Spectrophotometer (Nanodrop technologies, Rockland, USA).

Labeling and Hybridization

Low RNA Input Fluorescent Linear Amplification Kit (Agilent, Santa Clara, CA) was used for labeling. Briefly, both first and second strand cDNA were synthesized by incubating 500ng of total RNA with 1.2µl of oligo dT-T7 Promoter Primer in nuclease-free water at 65°C for 10 min followed by incubation with 4.0µl of 5X First strand buffer, 2µl of 0.1M DTT, 1µl of 10mM dNTP mix, 1µl of 200U/µl MMLV-RT, and 0.5µl of 40U/µl RNaseOUT, at 40°C for 2 hour. Immediately following cDNA synthesis, the reaction mixture was incubated with 2.4µl of 10mM Cyanine-3-CTP or 2.4µl of 10mM Cyanine-5-CTP (Perkin-Elmer, Boston, MA), 20µl of 4X Transcription buffer, 8µl of NTP mixture, 6µl of 0.1M DTT, 0.5µl of RNaseOUT, 0.6µl of Inorganic pyrophosphate, 6.4µl of 50% PEG, 0.8µl of T7 RNA polymerase, and 15.3µl of nuclease-free water at 40°C for 2 hour (Fig 31).

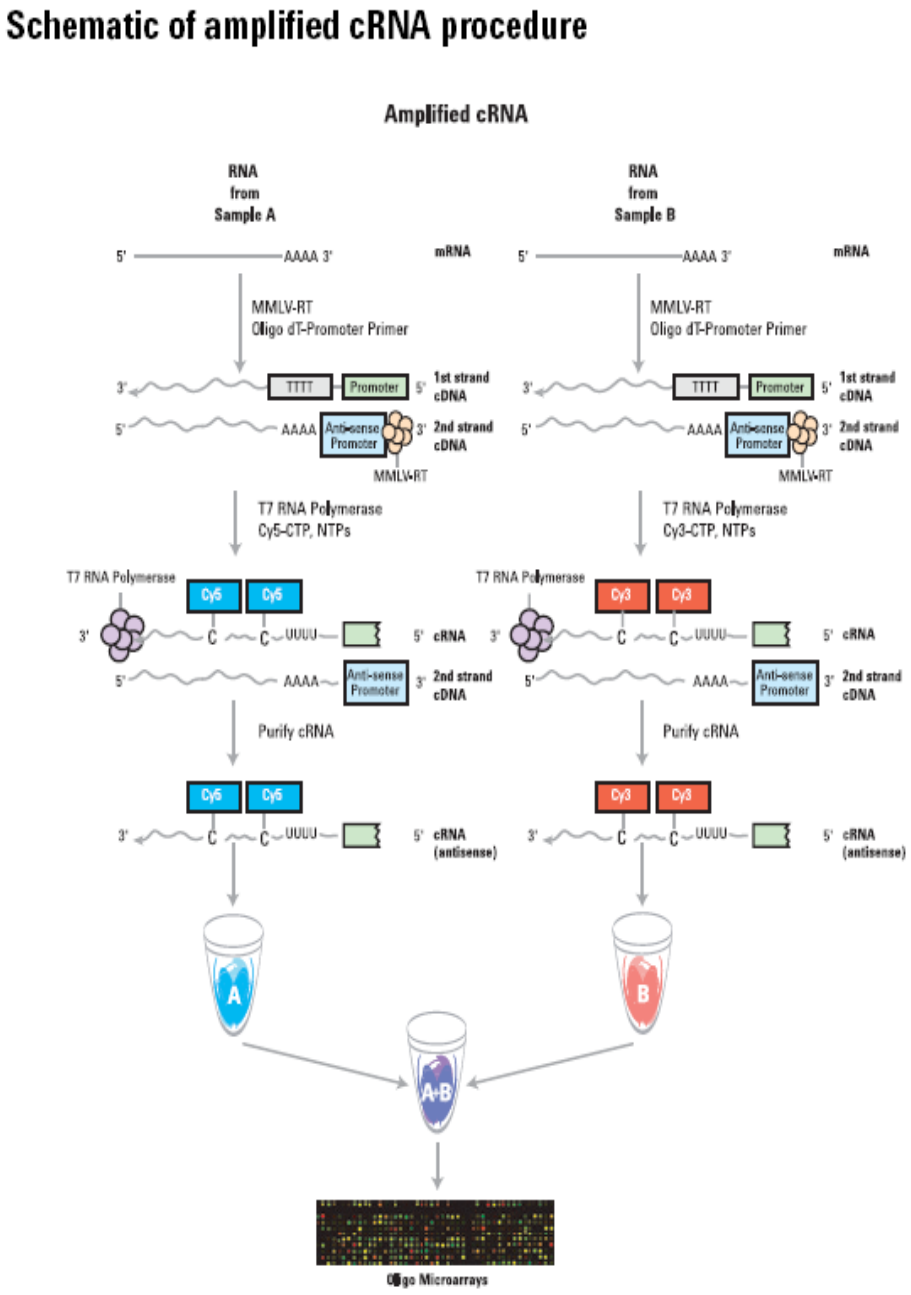


Figure 31: Schematic representation of amplified cRNA procedure.

Qiagen's RNAeasy mini spin columns were used for purifying amplified aRNA samples. The quantity and specific activity of cRNA was determined by using NanoDrop ND-1000UV-VIS Spectrophotometer version 3.2.1. Samples with specific activity >8 were used for hybridization.

Individual tumor cRNA was labeled with Cyanine 5 and hybridized with Cyanine 3- labeled pooled normal esophageal cRNA. 825ng of each Cyanine 3 or Cyanine 5 labeled cRNA in a volume of 41.8µl were combined with 11µl of 10X Blocking agent (Agilent) and 2.2µl of 25X Fragmentation buffer (Agilent), and incubated at 60°C for 30 minutes in dark to fragment RNA. The labeled and fragmented cRNAs were mixed with 55µl of 2X Hybridization Buffer (Agilent) and hybridized at 65 C for 17 hours with Human 10K cDNA array (University Health Network, Microarray Centre, Toronto, Canada), which contained 9,914 well-characterized human clones in an Agilent Microarray Hybridization Chamber (SureHyb:G2534A) with Hybridization Oven. After hybridization, slides were washed with Agilent Gene expression Wash Buffer I for 1 minute at room temperature followed by a 1min wash with Agilent Gene expression Wash Buffer II for 37°C. Slides were finally rinsed with Acetonitrile for cleaning up and drying.

Microarray image acquisition and data analysis:

Hybridized arrays were scanned at 5µm resolution on an Agilent DNA Microarray Scanner; model G2565AA at 100% laser power, and 30% PMT at 635 nm for Cy5-labelled samples and at 532 nm for Cy3-labeled samples. The resulting TIFF images were analyzed by Agilent Feature Extraction Software 9.1.3, which performed spot localization (Find Spot Algorithm), outlier pixels rejection based on the interquartile

range method (Cookie Cutter Algorithm), and flagging of saturated features. The local background value, calculated by the software through the radius method (Cookie Cutter Algorithm), was subtracted from every feature intensity signal and only features positive and significantly above background with a confidence level of 99% were selected for subsequent analysis.

Genespring software version GXV7.3.1 (Agilent Technologies, Palo Alto, CA) was used to normalize values for each gene and for further data analysis. Intensities of selected spots were transformed into $\log_2$ (Cy5/Cy3). Data was normalized for labeling and detection efficiencies of the two fluorescence dyes by locally weighted linear regression (LOWESS) method. LOWESS has mainly been used for correcting intensity-dependent dye-label effects between two samples on the same array. Briefly, the normalization process is as follows, *M–A* plots are constructed for each slide, where log-intensity ratios *M* (normalized log ratio of the two dyes) = log (Cy5/Cy3) = [log Cy5 - log Cy3] are plotted against the mean log-intensity *A* (average logarithmic signal intensity) = [(log Cy5 ± log Cy3)/2] as described by Yang et al. (2002). Positive *M* values indicate higher normalized signal intensity in the tumor RNA sample, negative *M* values indicate higher signal intensity in the control RNA sample, and *M* values of zero indicate equal intensity in the two samples. Higher *A* values indicate brighter signal, lower *A* values indicate dimmer signals. In Agilent Feature Extraction software, the log (red signal/ green signal) was pegged between –2 and 2; therefore, in the *M-A* plots all the spots distribute in the range of *M* from –2 to 2. Efficiency of LOESS normalization is assessed by monitoring *M–A* plots before and after normalization.

Differentially regulated genes were ranked on the basis of signal intensity, normalized ratio, flag value and variance across replicate experiments. Genes were considered to be up-regulated when the median of the normalized ratio was ≥2. Genes were considered to be down regulated when the median of the normalized ratio was ≤ 0.5. The observed number of differentially expressed genes in each GO category was compared to the corresponding number estimated from a random model (hypergeometric distribution); significance was assessed by a *P*-value. Hierarchical clustering analysis was also performed with Genespring software GXV 7.3.1. in which the average linkage and Pearson correlation (centered correlation) clustering algorithm was used. The microarray data set was submitted to the GEO repository (GSE 10127) at http://www.ncbi.nlm.nih.gov/geo. Annotations of the bioprocesses, molecular function and cellular localization were obtained using the freely available Gene Ontology [Source database (http://source.stanford.edu) and Biointerpreter software (http://www.genotypic.co.in/biointerpreter)].

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 PCR reactions were performed using an ABI Prism 7000 Sequence Detection System (Applied Biosystems, Foster City, CA). Primers and Taqman probes of five target genes and an internal control gene *18S rRNA* were purchased as Assays-on-Demand from Applied Biosystems (Table 15). Because18S rRNA resulted in the least variation throughout the tumor samples among a total of 11 housekeeping genes using Taqman Human Endogenous Control Plate (Applied Biosystems), this gene was

used as the endogenous control. The thermal cycling conditions included an initial denaturation step at 95˚C for 10 min, 40 cycles at 95˚C for 15s, and 60˚C for one min. All PCR reactions were done in duplicate. The $2^{-\Delta\Delta CT}$ method was used to calculate relative changes in gene expression determined from real-time quantitative PCR experiments. In the present study, the data are presented as the fold change in target gene expression in tumors normalized to the internal control gene (*18S rRNA*) and relative to the normal tissue control (matched normal as calibrator). Results of the real-time PCR data were represented as C_T values, where C_T was defined as the threshold cycle number of PCRs at which amplified product was first detected. There is an inverse correlation between C_T and amount of target: lower amounts of target correspond to a higher C_T value, and higher amounts of target have lower C_T values. The average C_T was calculated for both the target genes and *18S rRNA* and the ΔC_T was determined as the mean of the duplicate C_T values for the target gene minus the mean of the duplicate C_T values for *18S rRNA*. The $\Delta\Delta C_T$ represented the difference between the paired tissue samples, as calculated by the formula $\Delta\Delta C_T$= (ΔC_T of tumor- ΔC_T of normal). The N-fold differential expression in the target gene of a tumor sample compared to the normal sample counterpart was expressed as $2^{-\Delta\Delta CT}$ [67].

Validation of microarray results was done in 20 familial ESCC and 10 non-familial ESCC cases. Wilcoxon signed rank tests were used to determine the statistical significance of expression difference for each test gene in 20 familial ESCC cases (Table 15). The expression profiles of target genes by QRT-PCR in two groups (familial and non-familial ESCC cases) were then compared using Mann-Whitney *U*-test. A *P* –value $\leq$0.05 was considered significant.

Table 15: Information on the five genes examined by Real-Time PCR: location, function, primers and probes

Gene	Gene Bank ID	Location	Gene Expression status in our study	Putative function	P[a]	Assay ID[b]	Amplicon Size (bp)
CD14	W56632	5q22-q32	Up regulated	Inflammatory response	0.0002	Hs02621496_s1	140
ARG1	AA149501	6q23	Up regulated	Arginine catabolism	0.0002	Hs00163660_m1	86
PF4	AA024929	4q12-q21	Down regulated	Negative regulation of Angiogenesis	0.0028	Hs00236998_m1	86
EPHX1	AA838691	1q42.1	Down regulated	Xenobiotic metabolism	0.0002	Hs01116802_m1	89
MAPK7	H39192	17p11.2	Down regulated	MAP kinase activity	0.0012	Hs00964720_g1	97

a= Wilcoxon signed rank tests were used to determine the statistical significance of expression difference for each test gene in 20 samples. Statistical significance was defined as $P<0.05$

b= 'm1' denotes that assay's probe span an exon junction and will not detect genomic DNA. 's1' denotes that assay's primers and probes are designed within a single exon and will detect genomic DNA. 'g' denotes that assay may detect genomic DNA.

Tissue Microarray (TMA) based Immunohistochemical analysis:

A TMA was constructed from 120 formalin-fixed, paraffin-embedded blocks of esophageal biopsy samples. These included 20 controls (non-neoplastic esophageal squamous epithelium) and 100-ESCCs, of which 20 biopsies were obtained from familial cases. The tissue cylinders were precisely arrayed into the recipient block with core size of 1.5mm using a manual tissue microarrayer (Beecher Instruments, Silver Spring, MD, USA). After dewaxed in xylene and rehydrated in graded alcohol and distilled water, antigen retrieval was performed by microwave oven heating for 10 minutes at middle power in 0.01 M sodium citrate buffer (pH 6.0) or incubated with protease XXIV (Biogenex, San Ramon, CA) for 20 minutes at room temperature. Then, sections were incubated with 3% hydrogen peroxide for 5 minutes to block endogenous peroxidase activity. Nonspecific staining was blocked by 10% normal goat serum (Vector Laboratories Inc, Burlingame, CA) for 10 minutes. TMA sections were incubated overnight at 4˚C with primary antibodies of KRT4 (1:100, Clone 6B10, Novacastra, Newcastle upon Tyne, UK), VEGF (1:50, clone G153-694, BD Biosciences), NFkB/p65 (1:100, NeoMarkers, Fremont, USA), and anti-collagen IV (1:50, Clone COL-94, Biogenex, USA). The standard Streptavidin peroxidase method was employed for immunostaining [174]. The bound antibody was detected then with biotinylated anti-mouse/rabbit IgG (H+L) (Vector Laboratories Inc, Burlingame, CA) and horseradish peroxidase streptavidin (Beijing Zhongshan Biotech, Beijing, China). 3, 3'-diaminobenzidine (Maixin Biotech, Fuzhou, Fujian, China) was used as the chromogen. Slides were lightly counterstained with hematoxylin. In control experiments, the primary

antibody was replaced by PBS. Internal positive controls were available on the TMA itself, for example, normal sqaumous cell epithelia expressed CK4.

Results:

Clinical and epidemiological information of enrolled patients:

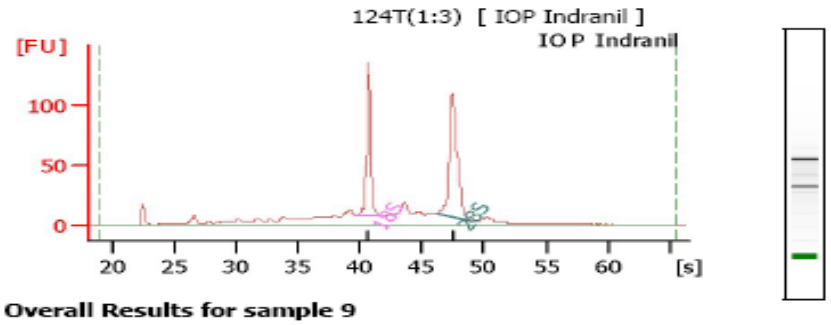
In 64% (59/92) cases, cancer occurred in the first-degree relatives whereas in 11% (10/92) cases, cancer occurred in the second-degree relatives. In about 10% (9/92) cases, the cancer occurred in both the first-degree and second-degree relatives. The univariate analysis revealed that the risk of developing esophageal cancer was more among subjects whose family history showed occurrence of cancer among the first-degree relatives (OR: 3.1; CI: 1.9-5.3) than the second-degree relatives (OR: 1.3, CI: 0.25-3.2). The estimates also revealed that the risk of developing esophageal cancer was more in subjects whose pee-degree suffered from esophageal cancer (OR: 2.4; CI: 1.1-4.1) than any other cancer (OR: 1.1; CI: 0.32-3.3).

Gene expression profiling by cDNA microarray:

As a reference RNA, we used pool of total RNA from normal esophageal tissue of all patients. Total RNA with $OD_{260}/OD_{280}>1.8$ and OD_{260}/OD_{270} 1.3 was used for microarray experiments. We considered RNA to be good quality when the rRNA 28S/18S ratios were greater than or equal to 1.2. Out of 20 patients, RNA samples from nine patients' tumor biopsies with RNA Integrity Number (RIN) of ≥7 were used for gene expression study by cDNA microarray (Fig.32 & Fig.33, Table 16).

Fig.32

32A



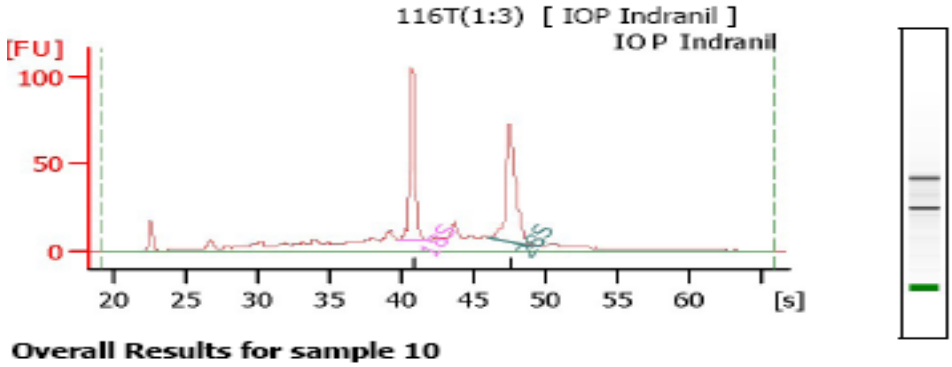
Overall Results for sample 9

RNA Area:	701.4
RNA Concentration:	331 ng/µl
rRNA Ratio [28s / 18s]:	1.5
RNA Integrity Number (RIN):	7.9 (B.02.02)

Fragment table for sample 9

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	39.64	41.68	111.1	15.8
28S	46.14	49.08	161.5	23.0

32B



Overall Results for sample 10

RNA Area:	604.2
RNA Concentration:	285 ng/µl
rRNA Ratio [28s / 18s]:	1.3
RNA Integrity Number (RIN):	7.5 (B.02.02)

Fragment table for sample 10

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	39.71	41.85	86.5	14.3
28S	45.91	49.05	113.6	18.8

Fig 32C

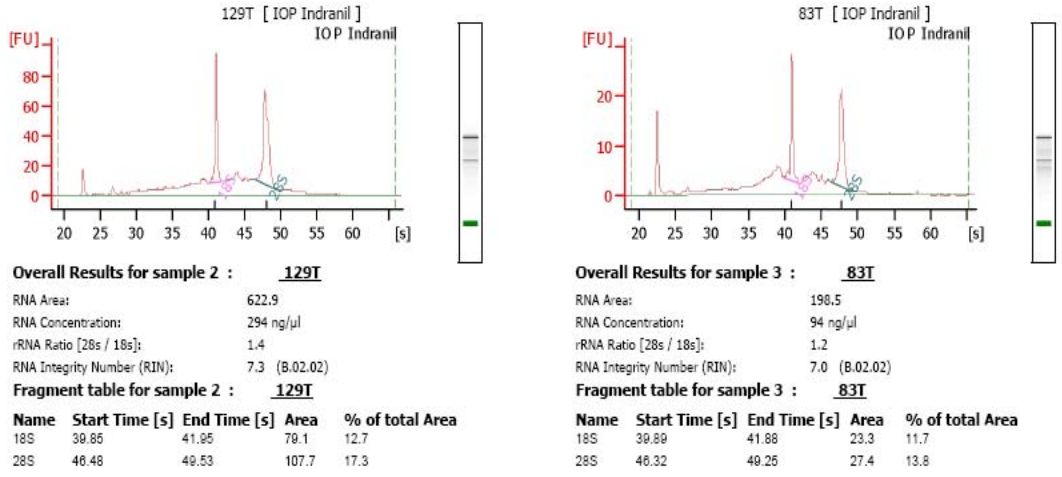


Fig 32D

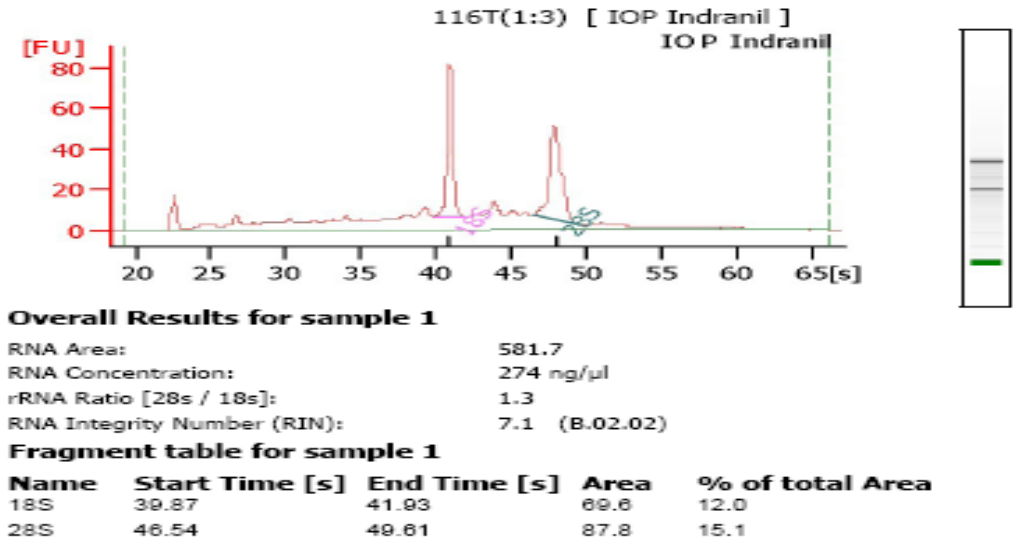
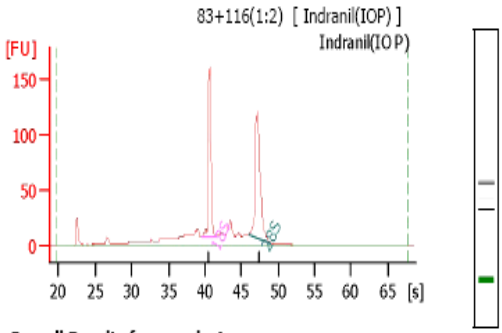


Figure 32: Elelectropherograms of tumor tissue RNA from Agilent 2100 bioanalyzer. Total fluorescence area, sample concentration, and the ratio of the ribosomal bands are reported with the electropherogram.

Fig.33

Fig.33A

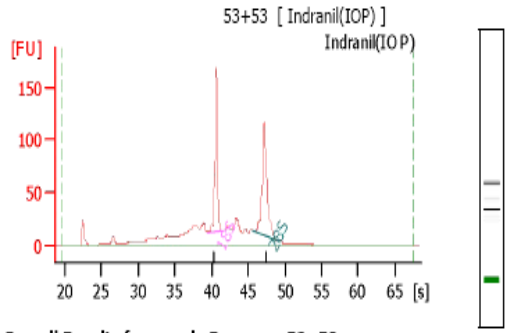


Overall Results for sample 4

RNA Area:	737.5
RNA Concentration:	271 ng/µl
rRNA Ratio [28s / 18s]:	1.3
RNA Integrity Number (RIN):	7.8 (B.02.02)

Fragment table for sample 4

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	39.60	41.51	133.5	18.1
28S	46.03	49.03	174.2	23.6



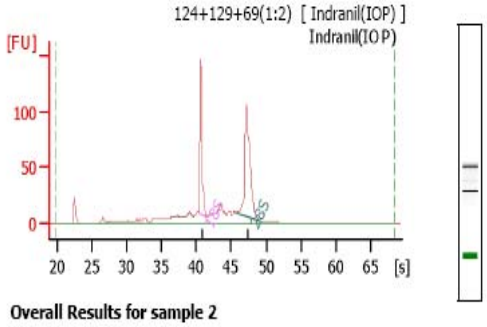
Overall Results for sample 5 : 53+53

RNA Area:	949.9
RNA Concentration:	349 ng/µl
rRNA Ratio [28s / 18s]:	1.3
RNA Integrity Number (RIN):	7.2 (B.02.02)

Fragment table for sample 5 : 53+53

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	39.52	41.48	136.7	14.4
28S	45.73	49.11	172.0	18.1

Fig .33B

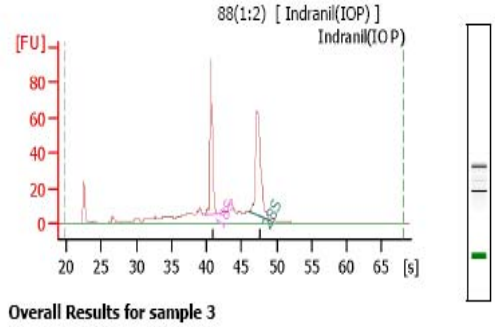


Overall Results for sample 2

RNA Area:	645.4
RNA Concentration:	237 ng/µl
rRNA Ratio [28s / 18s]:	1.5
RNA Integrity Number (RIN):	8.1 (B.02.02)

Fragment table for sample 2

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	40.34	41.53	112.1	17.4
28S	46.79	49.11	172.4	26.7



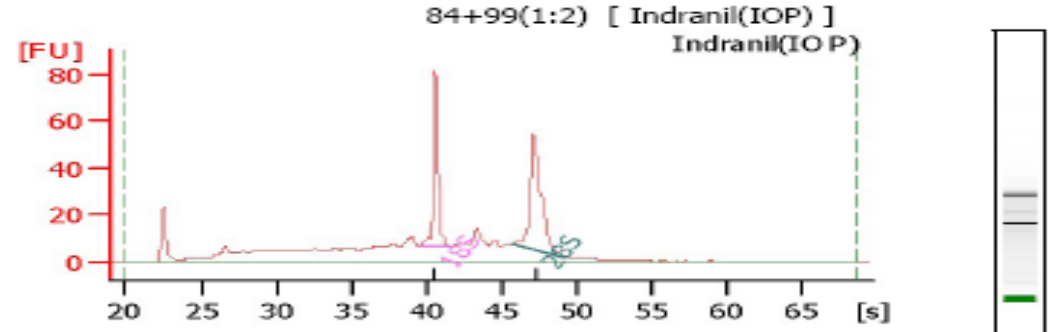
Overall Results for sample 3

RNA Area:	454.7
RNA Concentration:	167 ng/µl
rRNA Ratio [28s / 18s]:	1.3
RNA Integrity Number (RIN):	7.9 (B.02.02)

Fragment table for sample 3

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	39.67	41.94	77.6	17.1
28S	46.04	49.24	102.2	22.5

Fig. 33C



Overall Results for sample 1

RNA Area:	561.5
RNA Concentration:	206 ng/µl
rRNA Ratio [28s / 18s]:	1.3
RNA Integrity Number (RIN):	7.0 (B.02.02)

Fragment table for sample 1

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	39.49	41.47	69.9	12.4
28S	45.75	49.07	88.7	15.8

Fig.33: Elelectropherograms of pooled normal tissue RNA from Agilent 2100 bioanalyzer. Total fluorescence area, sample concentration, and the ratio of the ribosomal bands are reported with the electropherogram.

Table 16. Qualitative and quantitative estimation of RNA

Sample ID	RNA Conc. (ng/µl)§	O.D.260/ O.D280§	O.D.260/ O.D230§	rRNA (28S/18S)*	RNA Integrity Number (RIN)*
129T	185.68	2.06	2.12	1.4	7.3
116T	513.97	2.06	2.13	1.3	7.1
83T	35.02	2.1	2.03	1.2	7.0
88T	347.4	2.09	2.25	1.2	7.4
84T	344.61	2.05	2.07	1.2	7.2
99T	164.87	2.08	2.03	1.2	7.2
53T	251.4	2.05	2.12	1.2	7.3
124T	793.89	2.06	2.18	1.5	7.9
69T	248.14	2.05	2.08	1.2	7.3
(84+99)N	403.31	2.09	2.06	1.3	7.0
(124+129 +69)N	504.39	2.08	2.12	1.5	8.1
88N	300.75	2.11	2.1	1.3	7.9
(83+116) N	512.91	2.08	1.5	1.3	7.8
53N	293.7	2.09	2.08	1.3	7.2

Note: §: This information was obtained from NanoDrop® ND-1000 UV-Vis Spectrophotometer

*: This information was obtained from Agilent 2100 Bioanalyzer

Microarray images using amplified aRNA were shown in Fig.34.

Fig.34 Control (cy3) vs. 88T (cy5)

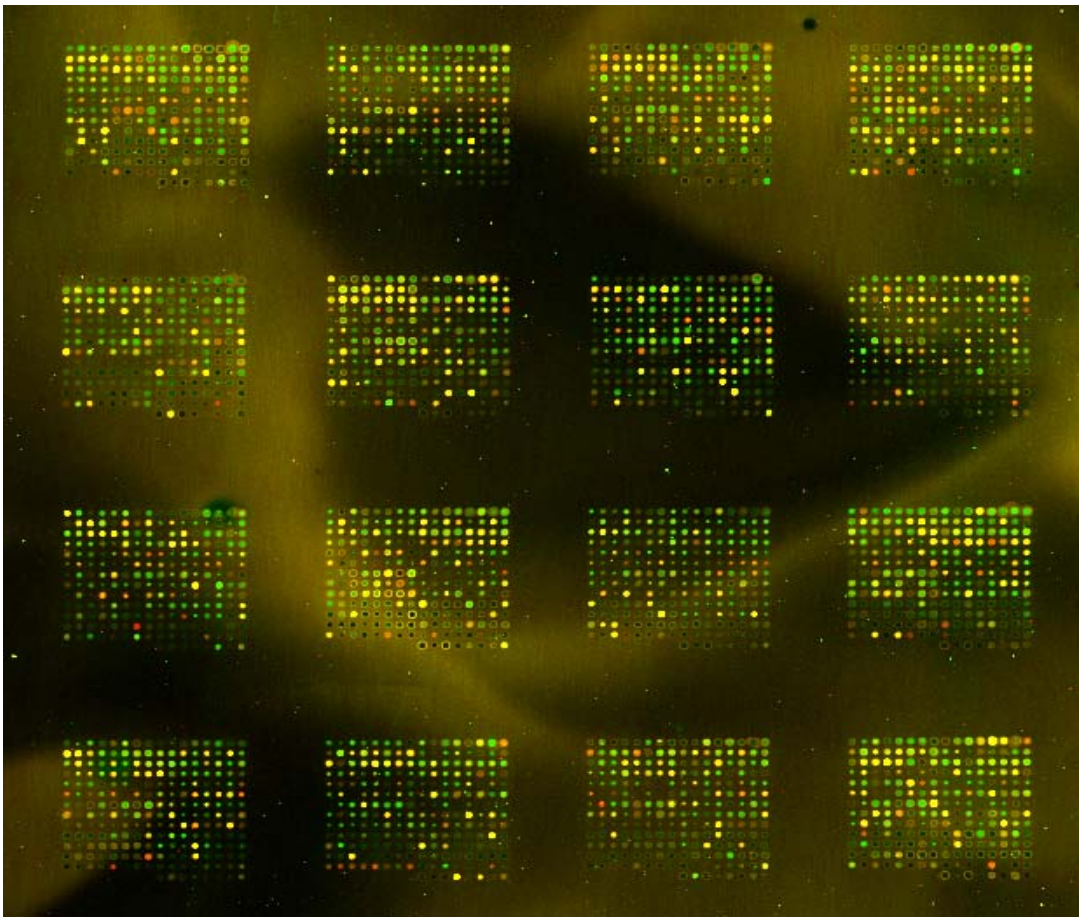


Figure 34: Example of array image generated from amplified cRNA of control and tumor tissue.

The Scatter plots that were plotted with Cy3 and Cy5 fluorescent signal values displayed a quite disperses pattern in distribution. Scatter plot (Fig.35) is of the log of the red background-corrected signal versus the log of the green background-corrected signal for non-control inliers features. The linearity or curvature of this plot can indicate the appropriateness of background method choices. The plot should be linear. The values below the plot indicate the number of non-control features that have a background-corrected signal less than zero.

Fig.35

Fig.35A Control (cy3) vs. 129T (cy5)

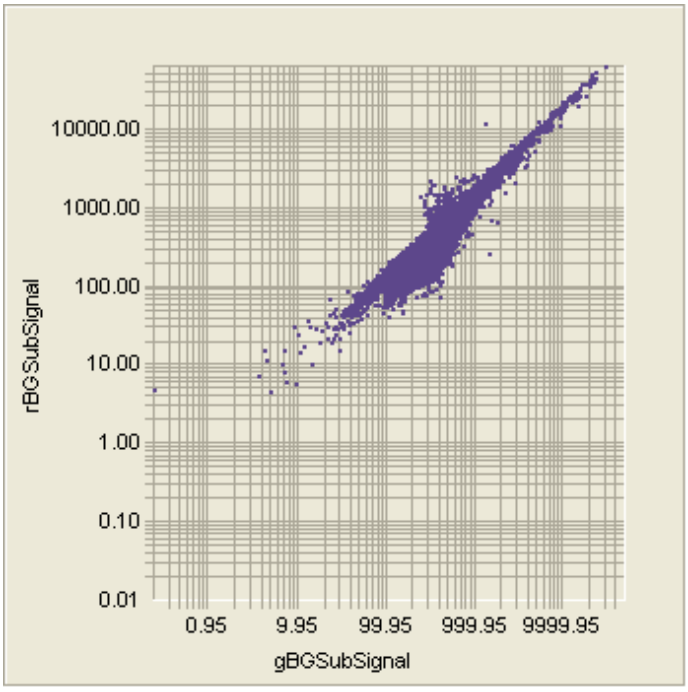


Fig.35B Control (cy3) vs. 83T (cy5)

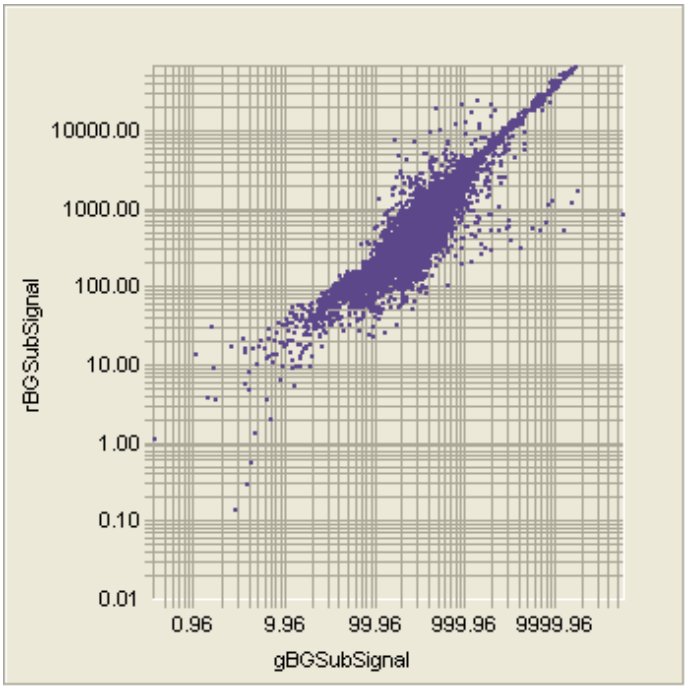


Figure 35: Scatter plot of hybridization signals on gene chip.

Four hundred and thirty eight genes were differentially expressed at least 1 fold in all experiments. The two dimensional hierarchical clustering showed that the majority of the differentially expressed genes were significantly down regulated (84%, 367 genes), whereas 16% genes (71 genes) showed up-regulation (Fig. 36).

Figure 36

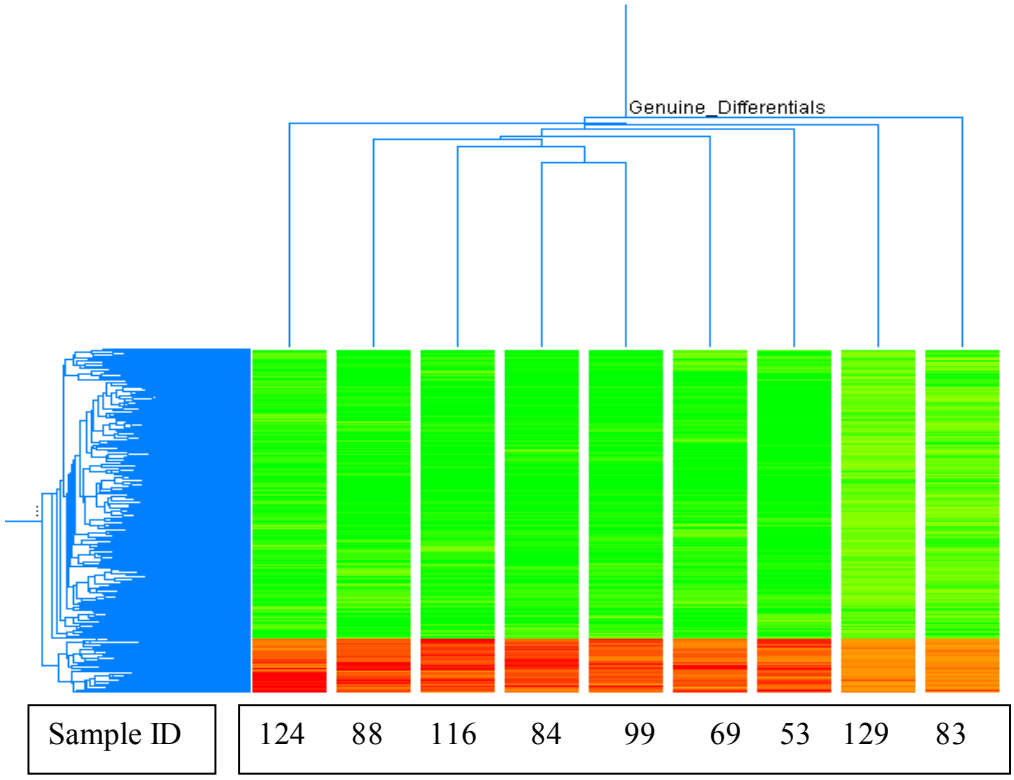


Figure 36: Two-way unsupervised hierarchical clustering (average linkage clustering) of the 438 differentially expressed genes that were over or under expressed in tumor versus normal tissue of nine familial ESCC patients. Red and green colors indicate up-regulated and down-regulated gene expression respectively.

Using stringent criteria ($P \leq 0.05$ and ≥ 1.2 fold change), 350 differentially expressed genes (26 up-regulated and 324 down-regulated) were identified and categorized using the Gene Ontology database into known or probable functional categories on the basis of biological processes and molecular function.

Out of 26 significantly up-regulated genes, genes involved in inflammatory response (*CD14*), immune response (*IFITM1, VDR, CD24*), cell motility (*WASL*), anti-apoptosis (*FOXO1A*), glucocorticoid receptor activity (*NR3C1*), steroid hormone receptor activity (*VDR*), arginase activity (*ARG1*) and metabotropic glutamate, GABA-B-like receptor activity (*GRM3*) were found to be biologically relevant in tumorigenesis (Table 17A). Out of 324 significantly down-regulated genes, genes involved in extracellular matrix organization (*KRT4, COL4A4, and COL14A1*), BMP signaling pathway (*SMAD1*), epoxide hydrolase activity (*EPHX1*), apoptogenic cytochrome c release channel activity (*VDAC1, TXNL1*), DNA damage response (*SMC1A*), humoral immune response (*POU2AF1, PF4, LY9 ,NFAT5,KLRC1*), ion transport (*SLC22A4, SLC23A1*), and MAP kinase activity (*MAPK7, SHC1*) were selected on the basis of biological relevance in tumorigenesis (Table 17B).

Table 17: Biologically relevant and statistically significant up-regulated (Table 17A) and down-regulated genes (Table 17B) in esophageal cancer patients with family history of esophageal cancer.

Table 17A

Genes	Gene symbol	Gene bank ID	Chromosomal location	Median fold change	GO category	*P*-value*	Pathway **
CD14	*CD14*	**W56632**	5q22-q32	2.05	GO: 6954: inflammatory response	0.00	Toll-like receptor signalin-g
Arginase	*ARG1*	**AA149501**	6q23	1.55	GO: 4053: arginase activity	0.01	Urea cycle and metabol-ism of amino groups
Glutamate receptor, metabotropi-c 3	*GRM3*	**N98673**	7q21.1-q21.2	1.54	GO: 8067: metabotropic glutamate, GABA-B-like receptor activity	0.04	Neuroac-tive ligand receptor interacti-on

Forkhead box O1A (rhabdomyosarcoma)	*FOXO1A*	**W32908**	13q14.1	1.43	GO: 6916: anti-apoptosis	0.03	Insulin Signalin-g
Interferon induced transmembrane protein 1 (9-27)	*IFITM1*	**H49853**	11p15.5	1.41	GO: 6955: immune response	0.00	B-cell receptor signalin-g
Wiskott-Aldrich syndrome-like	*WASL*	**AI271884**	7q31.3	1.41	GO: 6928: cell motility	0.01	Regulati-on of Actin Cytoske-leton
Nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)	*NR3C1*	**AA053901**	5q31.3	1.32	GO: 4883: glucocorticoid receptor activity	0.00	Neuroac-tive ligand receptor interacti-on
Vitamin D (1,25-dihydroxyvitamin D3) receptor	*VDR*	**BG149860**	12q13.11	1.20	GO: 3707: steroid hormone receptor activity	0.00	NA[c]
Gardner-	*FGR*	**W81591**	1p36.2-p36.1	1.42	GO: 6928: cell	0.00	Focal

Rasheed feline sarcoma viral (v-fgr) oncogene homolog					motility		Adhesio-n

Table 17B

Genes	Gene Symbol	Gene bank ID	Chromosomal location	Median fold change	GO category	*P*-value*	Pathway**
Keratin 4	*KRT4*	**AA629189**	12q12-q13	-5.20	GO: 30198: extracellu-lar matrix organizati on and biogenesi s-s	0.00	Cell communicatio-n
SMAD, mothers against DPP homolog 1 (Drosoph-ila)	*SMAD1*	**R83757**	4q31	-2.48	GO: 30509: BMP signaling pathway	0.00	TGF-beta signaling

Epoxide hydrolase 1, microsomal (xenobiotic)	*EPHX1*	**AA838691**	1q42.1	-2.76	GO: 4301: epoxide hydrolase activity	0.01	Metabolism of xenobiotics by cytochrome p450
Voltage-dependent anion channel 1	*VDAC1*	**AA025089**	5q31	-2.31	GO: 15283: apoptogenic cytochrome c release channel activity	0.00	Calcium Signaling
Platelet factor 4 (chemokine (C-X-C) motif) ligand 4	*PF4*	**AA024929**	4q12-q21	-4.11	GO: 6959: humoral immune response	0.00	Leukocyte transendothelial migration
Solute carrier family 22 (organic cation transport-	*SLC22A4*	**N26836**	5q31.1	-2.94	GO: 6811: ion transport	0.00	NA[c]

er), member 4							
Mitogen activated protein kinase 7	*MAPK7*	**H39192**	17p11.2	-2.89	GO: 4707: MAP kinase activity	0.00	GnRH signaling /Gap Junction
Killer cell lectin-like receptor subfamil-y C, member 1	*KLRC1*	**AA913480**	12p13	-2..39	GO: 6959: humoral immune response	0.00	Natural killer cell mediated cytotoxicity
Nuclear factor of activated T-cells 5	*NFAT5*	**H60999**	16q22.1	-2.23	GO: 6959: humoral immune response	0.00	Natural killer cell mediated cytotoxicity
Collagen, type XIV, alpha 1 (undulin)	*COL14A1*	**AA167222**	8q23	-2.01	GO: 30198: extracellu -lar matrix organizati	0.00	Cell communicatio -n

					-on and biogenesi -s		
Collagen, type IV, alpha 4	*COL4A4*	**H67349**	2q35-q37	-2.18	GO: 30198: extracellu -lar matrix organizati -on and biogenesi -s	0.00	Cell communicatio -n
SHC (Src homolog y 2 domain containin g) transform ing protein 1	*SHC1*	**R52961**	1q21	-3.28	GO: 4707: MAP kinase activity	0.00	Natural killer cell mediated cytotoxicity
Solute carrier family 23	*SLC23A1*	**AI334656**	5q31.2-q31.3	-3.03	GO: 6811: ion transport	0.00	NA[c]
Lymphoc yte antigen 9	*LY9*	**AI056539**	1q21.3-q22	-2.43	GO: 6959: humoral immune	0.00	NA[c]

					response		
POU domain, class 2, associating factor 1	*POU2AF1*	**AI028546**	11q23.1	-2.36	GO: 6959: humoral immune response	0.00	NA[c]
SMC1 structural maintenance of chromosomes 1-like 1 (yeast)	*SMC1A*	**AA598887**	Xp11.22-p11.21	-2.43	GO:42770: DNA damage response	0.01	Cell Cycle
Thioredoxin-like 1	*TXNL1*	**AA078976**	18q21.31	-2.19	GO: 15283: apoptogenic cytochrome c release channel activity	0.00	NA[c]

Footnotes

*= Biological significance of differentials was computed and functionally classified using the GeneSpring GX software (Agilent Technologies, USA) on the basis of gene ontology. GeneSpring GX software depicts the biologically significant ontology for any given gene list as follows: number of genes in the array known to be present in the ontology category vs. number of genes in the array that are differentially regulated.

[c] NA= No Information Available.

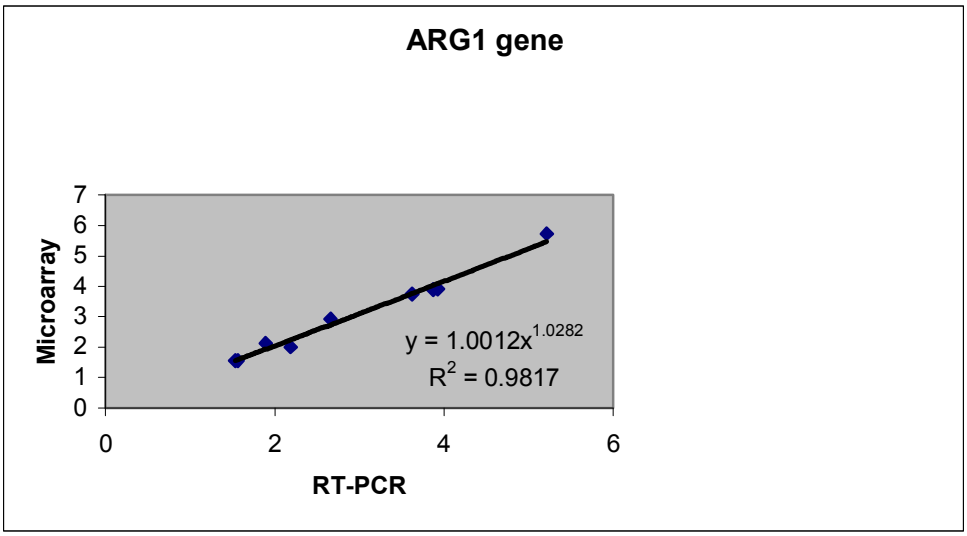
**= Pathways were obtained using enrichment analysis based on gene ontology categories using the Biointerpreter software and DAVID knowledgebase (http://david.abcc.ncifcrf.gov/knowledgebase).

Validation of microarray results with quantitative real -time PCR:

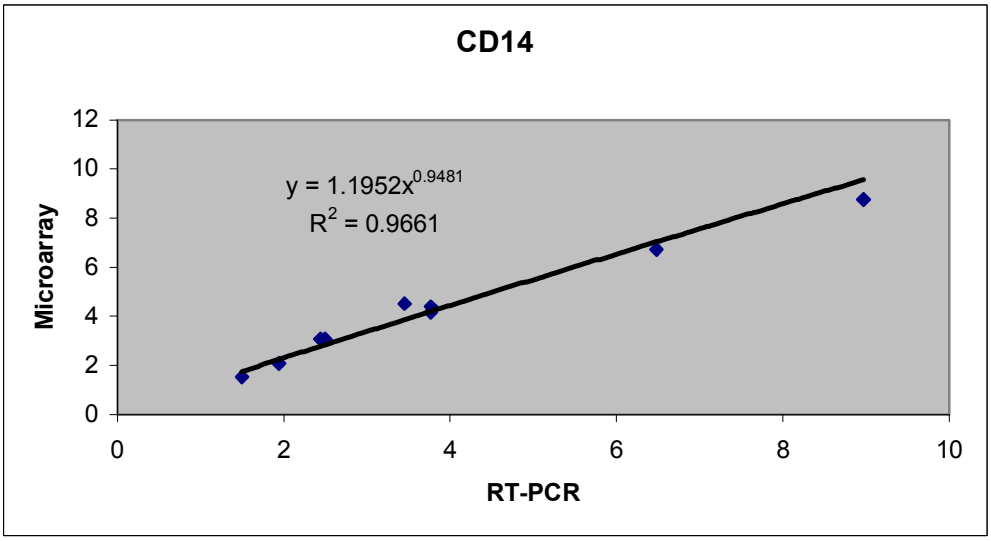
Five top ranked genes viz *ARG1*, *CD14*, *PF4*, *MAPK7* and *EPHX1* were selected to confirm the microarray results with real-time RT-PCR. Regression plot analyses for the five genes showed positive correlation between the gene expression measured by cDNA microarray and real-time RT-PCR (Fig. 37). Pearson correlation coefficient of each gene was *ARG1* 0.99, *CD14* 0.98, *PF4* 0.99, *MAPK7* 0.98 and *EPHX1* 0.414 (Table 18). Quantitative real-time PCR performed on all nine patients' specimens previously arrayed, 11 other familial ESCC patients' specimens and 10 non-familial ESCC, indicated that *ARG1* and *CD14* were consistently up-regulated and *PF4, MAPK7*and *EPHX1* were consistently down-regulated in all patients' specimens (Table 15). No statistical significant difference (P = 0.1508-0.5358) was observed in relative gene expression for five target genes in between familial and non-familial ESCC groups (Table 19).

Figure 37

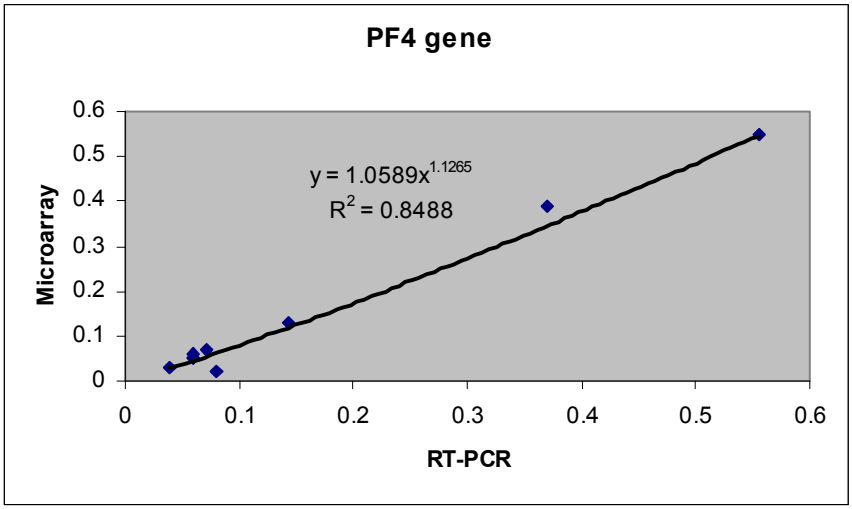
A


ARG1 gene
Microarray
RT-PCR
y = 1.0012x^1.0282
R^2 = 0.9817
0
1
2
3
4
5
6
7
0
2
4
6

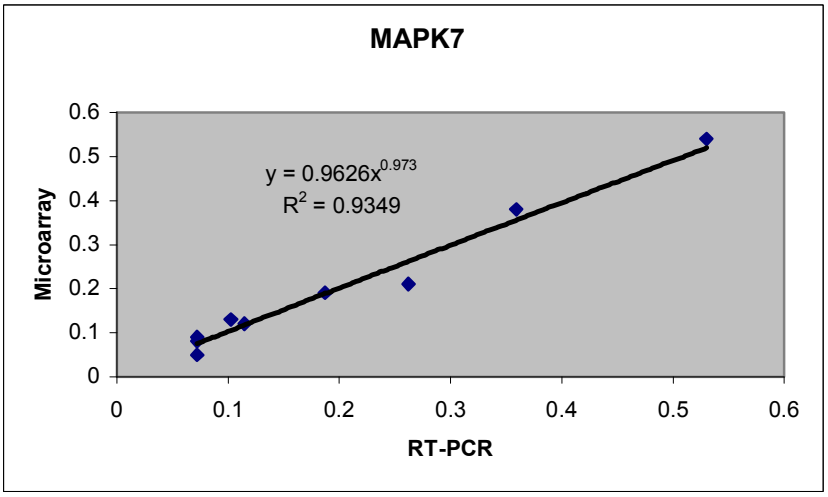
B


CD14
Microarray
RT-PCR
y = 1.1952x^0.9481
R^2 = 0.9661
0
2
4
6
8
10
12
0
2
4
6
8
10

C


PF4 gene
y = 1.0589x^1.1265
R^2 = 0.8488
Microarray
RT-PCR
0
0.1
0.2
0.3
0.4
0.5
0.6

D


MAPK7
y = 0.9626x^0.973
R^2 = 0.9349
Microarray
RT-PCR
0
0.1
0.2
0.3
0.4
0.5
0.6

E

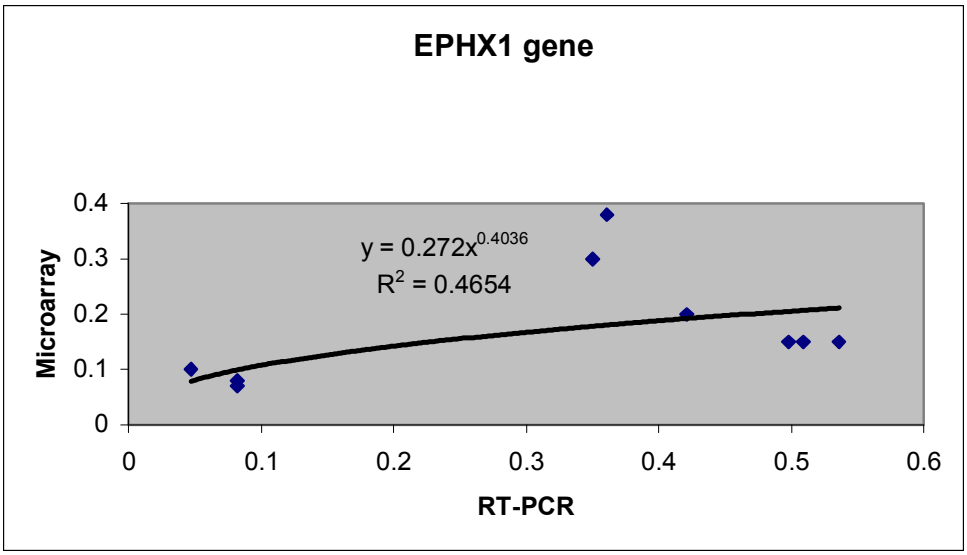


Figure 37. Regression plots for fold change by microarray (Y-axis) and quantitative real-time PCR assay (X-axis) for *ARG1* (A), *CD14* (B), *PF4* (C), *MAPK7* (D) and *EPHX1* (E).

Table 18: - Fold change comparison in tumor tissue of ESCC patients with family history of esophageal cancer by microarray and real-time RT- PCR.

Table 18A

Sample ID	Fold Change of ARG1 gene in Microarray	Fold Change of ARG1 gene in RT-PCR	Pearson Correlation coefficient
116	3.74	3.625	0.99196541
129	1.55	1.533	
99	2.01	2.185	
84	3.91	3.9217	
69	5.73	5.212	
124	2.13	1.895	
53	2.93	2.66	
88	3.89	3.871	
83	1.55	1.56	

Table 18B

Sample ID	Fold Change of CD14 gene in Microarray	Fold Change of CD14 gene in RT-PCR	Pearson Correlation coefficient
116	4.52	3.453	0.9876678
129	1.51	1.5	
99	3.07	2.44	
84	3.08	2.492	
69	4.15	3.767	
124	4.38	3.767	
53	8.74	8.963	
88	6.73	6.482	
83	2.07	1.939	

Table 18C

Sample ID	Fold Change of PF4 gene in Microarray	Fold Change of PF4 gene in RT-PCR	Pearson Correlation coefficient
116	0.05	0.059	0.99428726
129	0.55	0.555	
99	0.06	0.059	
84	0.13	0.144	
69	0.05	0.059	
124	0.07	0.071322446	
53	0.03	0.04	
88	0.02	0.08	
83	0.39	0.37	

Table 18D

Sample ID	Fold Change of MAPK7 gene in Microarray	Fold Change of MAPK7 gene in RT-PCR	Pearson Correlation coefficient
116	0.12	0.1146	0.98827255
129	0.54	0.53	
99	0.13	0.1027	
84	0.19	0.187	
69	0.08	0.072	
124	0.21	0.262	
53	0.05	0.072	
88	0.09	0.072	
83	0.38	0.359	

Table 18E

Sample ID	Fold Change of EPHX1 gene in Microarray	Fold Change of EPHX1 gene in RT-PCR	Pearson Correlation coefficient
116	0.07	0.082	0.41487558
129	0.38	0.361	
99	0.2	0.421	
84	0.15	0.509	
69	0.15	0.498	
124	0.3	0.35	
53	0.1	0.047	
88	0.08	0.082	
83	0.15	0.536	

Table 19: - *P*- value by Mann-Whitney *U*-test

Gene	*P* value-determined by Mann-Whitney U-test in between Non-familial ESCC cases Vs Control	*P value* -determined by Mann-Whitney U-test in between familial Vs Non-familial ESCC cases
CD14	0.0005	0.3424
ARG1	0.0004	0.3021
PF4	0.0031	0.1508
EPHX1	<0.0001	0.2532
MAPK7	0.0056	0.5358

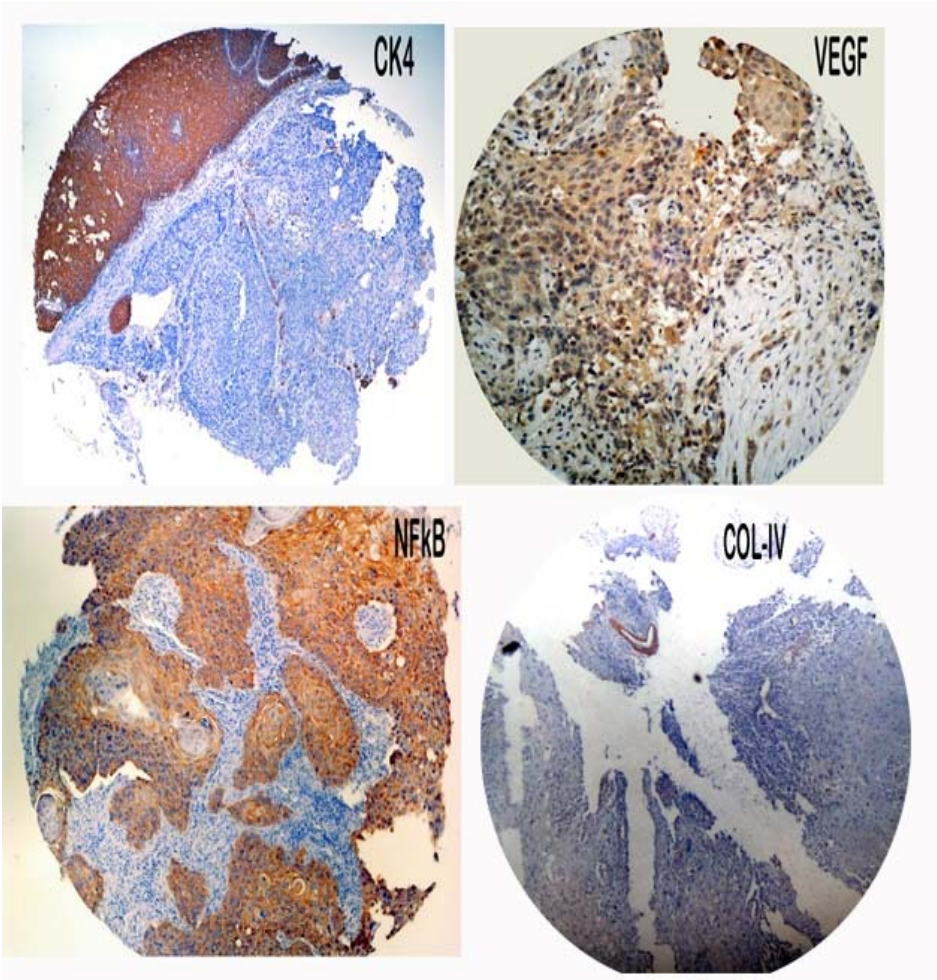
TMA based Immunohistochemical analysis:

Differential expression of *KRT4* and *COL4* genes identified in cDNA microarray analysis were validated at protein level using immunohistochemistry on TMA. We also studied expression of *VEGF* (which is down stream target gene of *PF4* and *MAPK7*) and *NF-kB* (which is down stream target gene of *CD14*) at protein level. The cytoplasmic staining was considered positive for CK4 (KRT4), VEGF, NFkB and collagen4 (COL4) and was scored as <5% or no staining= 0, 5-25%=1, 26-50%= 2, 51-100%= 3.

Expression of KRT4 was found only in overlying non-neoplastic epithelium and absent in tumor cells (14 out of 20 showing a score of 0 while remaining 6 score of 1). VEGF showed expression in tumor cells with score of 2 in 8 (40%) cases and with score of 3 in 12 (60%) cases. NFkB showed a diffuse and strong expression in tumor cells with score of 3 in all familial cases (100%). The COL4A4 was not expressed in the tumor cells. The staining intensity was not graded as all the cores were more or less uniformly stained (Fig. 38). Staining pattern in tumor cores from non-familial ESCC cases showed similar staining pattern as in familial cases.

Figure 38

A **B**



C **D**

Figure 38 Photomicrograph of tissue microarray cores from esophageal tumor biopsies obtained from familial ESCC patients showing negative immunostaining for KRT4 (A) and Collagen IV (D) and positive immunostaining for VEGF (B) and NF-kB (C).

Figure 39

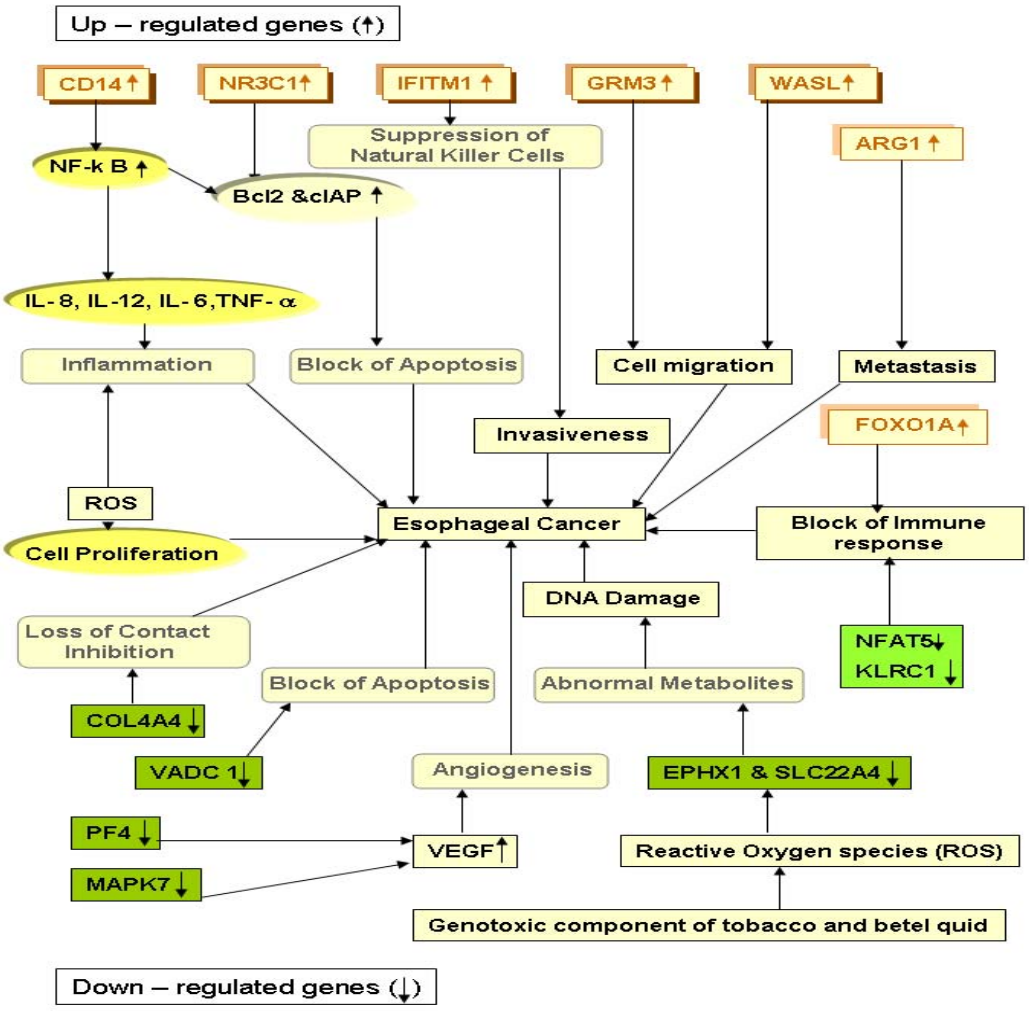


Figure 39: Schematic illustration of differentially expressed genes involved in molecular mechanism of esophageal tumorigenesis. ↑ indicates up-regulated genes and ↓indicates down regulated genes. The key differentially expressed genes (*CD14, ARG1, EPHX1, MAPK7, PF4, COL4A4,* and *CK4*) were validated by RT-PCR/ tissue microarray.

Discussion:

To the best of our knowledge, this is the first study that gives an insight into the genes and molecular pathways that may be playing an important role in the familial aggregation of esophageal cancer in high-risk area of India. According to Hannahan and Weinberg [175], tumorigenesis requires six essential alterations to normal cell physiology: self-sufficiency in growth signals, insensitivity to growth inhibition, evasion of apoptosis, immortalization, sustained angiogenesis and tissue invasion and metastasis. In addition, another component of cancer progression is the failure of the host immune response to recognize tumor cells [176, 177]. Molecular profiling of esophageal cancer with familial clustering showed deregulation of most of these physiological mechanisms in the current study (Fig.39). On the basis of functional annotation, genes responsible for inflammatory response, immune response, angiogenesis, cell migration and cell proliferation were found significantly deregulated in these cases.

Genes (*KRT4, COL4A4*) involved in extracellular matrix organization and cell communication pathway showed significant down regulation in present study. *CK4* influences the formation of cytoskeletal cells, and its low expression has been reported earlier in upper aero-digestive tract tumors [174]. Loss of expression of type IV collagen alpha5 and alpha6 chains, associated with the hypermethylation of their promoter region, has also been reported in colorectal cancer [178].

Platelet factor 4 (*PF-4*)-a CXC-chemokine-involved in humoral immune response, leukocyte transendothelial migration pathway and inflammatory processes, showed significant down regulation in our study. *PF4* inhibits T cell function by down-modulating cell proliferation and cytokine release. In addition, *PF-4* has strong anti-

angiogenic properties that inhibit endothelial cell proliferation and migration, in vitro and in vivo angiogenesis, tumor-associated neovascularization and tumor growth [179]. *PF-4* inhibits VEGF-induced mitogen-activated protein kinase (MAPK) signaling pathways comprising Raf1, MEK1/2 and ERK1/2 genes [180]. *MAPK7* (*ERK5*) was found to be down regulated in our study. *MAPK7* deficiency leads to an increased expression of VEGF, deregulation of which has been shown to impede angiogenic remodeling and vascular stabilization. Increased VEGF expression in a hypoxic environment promotes vessel growth, angiogenesis and tumor growth [181].

Microsomal epoxide hydrolase 1 (*EPHX1),* involved in metabolism of xenobiotics by CYP450, plays an important role in both the activation and detoxification of tobacco-derived carcinogens [182]. In addition, *SLC22A4,* a novel proton antiporter gene that plays a role in the renal excretion of xenobiotics and their metabolites also showed down regulation in our study. Voltage-dependent anion channel (*VDAC1)* gene, which is involved in calcium signaling and apoptosis inducing pathway, showed down regulation in present study. *VDAC1* controls pro- and anti-apoptotic Bcl2-family proteins by regulating the release of cytochrome c and apoptotic proteins in the inter-membrane space [183]. *SMAD1* that is involved in TGF-β or BMP (bone morphogenetic proteins) signaling pathway and helps in tumor progression, showed significant down regulation in the present study. BMPs are involved in wide range of biological activities including cell growth, apoptosis, morphogenesis, development and immune responses. An earlier study has reported that *SMAD1* signaling is low in androgen-regulated growth of prostate cancer, is activated after castration, and again decreases in hormone independent tumors [184].

CD14, which is involved in Toll-like receptor (TLR) signaling pathway (Fig.40) and inflammatory response, showed significant up-regulation in our study. CD14 induces inflammatory response via MyD88, TIRAP and TRAF6 leading to NF-kB activation and cytokine secretion. The activation of TLR signaling in tumor cells induces the synthesis of proinflammatory factors and immunosuppressive molecules, which enhance the resistance of tumor cells to cytotoxic lymphocyte attack and lead to immune evasion [185]. *WASL* gene, which is involved in cell motility and regulation of actin cytoskeleton, showed up regulation in present study. Bourguignon et al [186] had earlier reported that *N WASP (WASL)* played a pivotal role in regulating *CD44-ErbB2* interaction, ss catenin signaling and actin cytoskeleton functions that were required for tumor specific behavior such as transcriptional up regulation and tumor cell migration. Over expression of interferon inducible 9-27 [*IFITM1*] gene, which is involved in B-cell receptor signaling pathway and immune response, has been reported to play a role in malignant progression by suppressing natural killer cells and by increasing the invasive potential of gastric cancer cells [187]. Gene involved in arginine metabolism (*ARG1*) has been found to be up-regulated in these ESCC patients. Myeloid suppressor cells (MSCs) producing high levels of arginase I block T cell function by depleting l-arginine in cancer, chronic infections, and trauma patients. In cancer, infiltration of MSCs in circulation is an important mechanism for tumor evasion and impairs the therapeutic potential of cancer immunotherapies [188]. Over expression of arginase in colorectal carcinoma is associated with metastasis [189].

FOXO1A, a transcription factor that is involved in anti-apoptosis and insulin signaling pathway (Fig.41), showed up-regulation in our study. It can promote tumor

growth and tissue invasion while inhibiting local inflammatory and immune responses. *FOXO1A* is the pathogenetic marker for alveolar rhabdomyosarcoma, an aggressive form of childhood cancer [190]**.** Constitutive phosphorylation of the *FOXO1A* transcription factor has also been reported as a prognostic variable in gastric cancer [191].

NR3C1 (the glucocorticoid receptor family) and *GRM3* (GABA-B-like receptor activity) genes, showing up-regulation in present study, are associated with enhanced anti-apoptotic effect and tumor cell migration, respectively. Both are involved in the neuroactive ligand receptor interaction pathway. The ligand-activated glucocorticoid receptor activates the anti-apoptotic Bcl-2 family protein Bcl-x (L) that inhibits apoptosis and caspase-3 activity in fibrosarcoma cells [192]. Activation of the glucocorticoid receptor in epithelial ovarian cancer cells has earlier been reported to have an enhanced cellular expression level of cIAP2 and anti-apoptotic effect [193]. The multifunctional G-protein-coupled metabotropic glutamate receptor (mGluR) family contributes to tumor cell migration and invasion in oral cancer [194].

Several differentially regulated genes in familial ESCC are functionally annotated to immune response category. Up regulation of *CD14, WASL, IFITM1, FOXO1A, GRM3, ARG1* and *NR3C1* genes is found to be associated with suppression of NK cells, inhibition of immune response, immune evasion, tumor cell migration, invasion, metastasis and anti-apoptosis respectively. Down regulation of *PF4, SMAD1, SLC22A4*, *MAPK7*, *KLRC1, NFAT5, SHC1, LY9*, *POU2AF1* and *VDAC1* genes may be involved in invasion, inhibition of humoral immune response, angiogenesis and anti-apoptosis respectively in these familial ESCC cases. Furthermore, the data presented here will not only provide important information about tumorigenesis of this tumor, but also facilitate

the identification of candidate genes that could be used as therapeutic targets for the treatment of patients with this tumor.

Validation of *CD14, ARG1, PF4, MAPK7* and *EPHX1* genes at the mRNA level by real-time PCR and *KRT4, COL4, NF-kB* and *VEGF* genes at the protein level by tissue microarray did not show any difference in familial and non-familial ESCC cases from the same high-risk area of India, suggesting that familial clustering of cancer in these patients is more due to shared environmental factors rather than shared genes by family members.

In this study, the use of high throughput genomic technology in clinical specimens from well characterized populations that have familial clustering of cancer may lead to identification of molecular mechanism associated with progression of esophageal cancer. Functional analyses of these genes will lead to better understanding of the development and progression of ESCC.

Figure 40

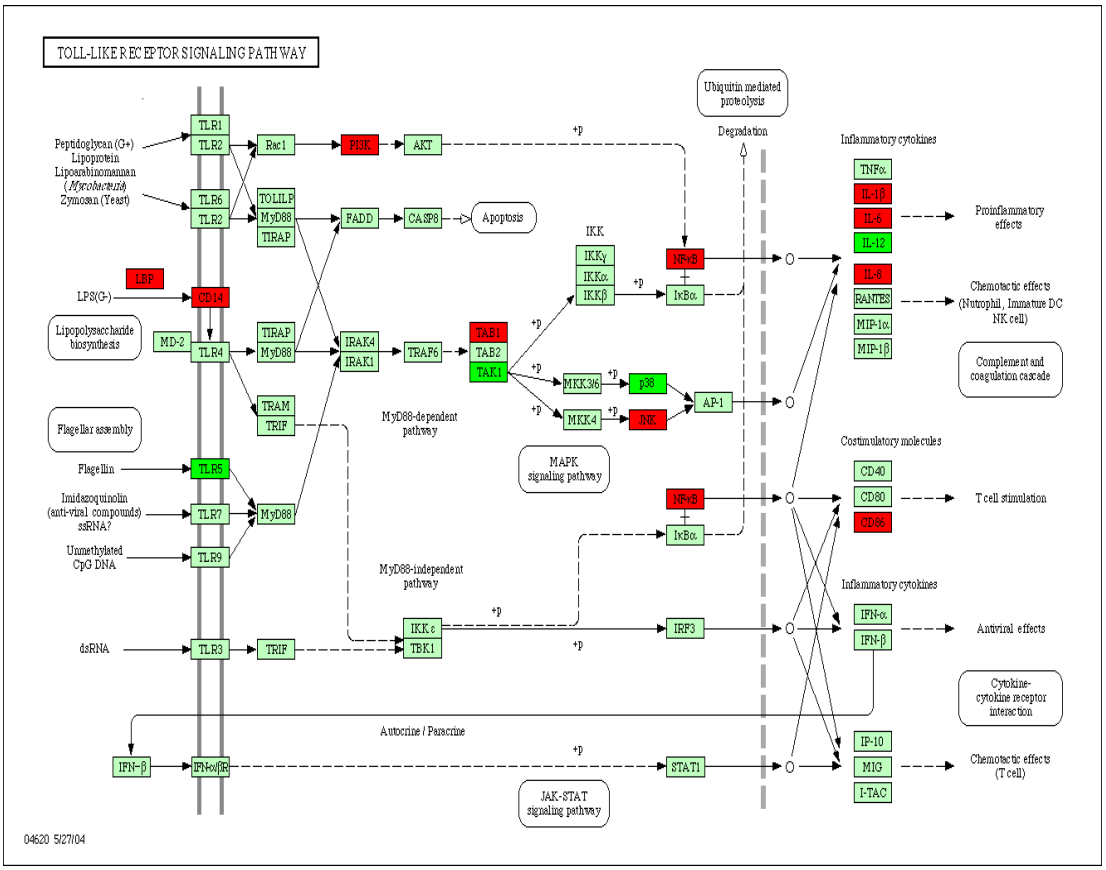


Figure.40: *CD14* which is involved in Toll-Like Receptor Signaling Pathway significantly up-regulated in ESCC.

Figure 41

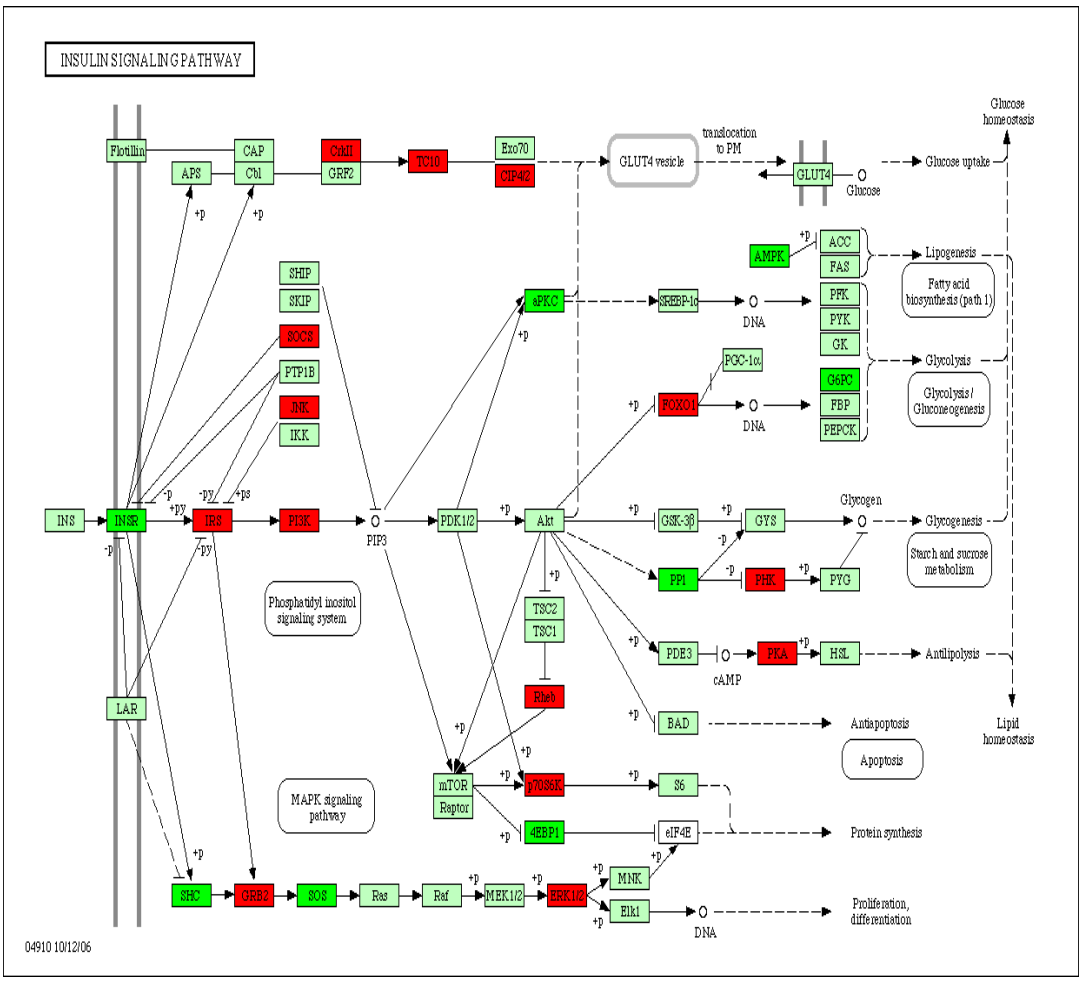


Figure 41: *FOXO1* which is involved in Insulin Signaling Pathway significantly up-regulated in our study.

Chapter 7

Contribution of germ line *BRCA2* sequence alterations to risk of familial aggregation of esophageal cancer in high-risk area of India.

Introduction:

Esophageal cancer is among the ten most common malignancies worldwide and ranks as the 6th leading cause of death from cancer [50]. The incidence of esophageal cancer varies greatly between developed and developing countries and a 50-fold difference has been observed between high and low risk populations [3]. The esophageal cancer belt is a geographic area of high incidence that stretches from north-central China westward through central Asia to northern Iran [50]. Association of family history with an increased risk of esophageal cancer has been reported in several case-control and cohort studies from China, Iran and Japan suggesting possible role of environmental as well as genetic factors [10, 11]. Esophageal cancer is reported to be significantly more common in the first-degree relatives of the cases than the relatives of unaffected controls in Turkmen population of Iran [15]. A high prevalence of esophageal cancer has also been reported from Assam in the northeast (NE) region of India with an age-adjusted rate (AAR) of 33 /100,000 males [9, 172].

Contribution of *BRCA2* mutations for the development of esophageal squamous cell carcinoma (ESCC) has been reported in high-risk Chinese populations and in Turkmen population of Iran. Five germ-line mutations (N1600 del, A2054P, V2109I, Q2580H, and C315S) have been reported in *BRCA2* gene in 14% ESCC (six of 44) patients with a family history of esophageal cancer in a high-risk area of China [17]. In Turkmen population of Iran, K3326X nonsense variant has been reported in eight ESCC cases with family history of esophageal cancer [15].

BRCA2 was identified as a breast cancer susceptibility gene in 1995 by Wooster et al [195]. Germ line mutations of *BRCA2* gene contribute to the development of breast, ovarian, prostrate and pancreatic cancers [15, 17]. *BRCA2* is mainly involved in homologous recombination repair (HRR) through control of *RAD51* recombinase and it interacts with many other proteins involved in various cellular functions, including cell cycle regulation, transcription regulation, cytokinesis and control of cell proliferation [196].

In the current study, screening for germline mutations in BRCA2 gene has been carried out in 20 ESCC cases having family history of esophageal cancer and 80 non-familial ESCC cases from high-risk area of India. Genomic DNA from 100 age and sex matched controls from the same population was also screened to confirm whether variants identified in cases are associated with elevated risk of familial ESCC in this population of India.

MATERIALS and METHODS

Selection of Patients and Collection of Samples:

Of 317 cases of esophageal cancer registered at Dr. Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam during the year of 2005-2006, 92 (29%) cases had family history of esophageal and/or other cancers beside habit of tobacco and betel quid chewing. Of 92 patients, 45 patients (49%) had family history of esophageal cancer. Out of 45 patients with familial ESCC, 26 patients had esophageal cancer in 1st degree relatives, 2 patients had esophageal cancer in 2nd degree relatives, 8 patients had esophageal cancer in spouse and 9 patients had esophageal cancer in family with no

blood relation. Of 45 patients with familial ESCC, 20 patients who had esophageal cancer in either 1st or 2nd degree relatives gave consent to donate blood for this study. An amount of 5 ml of blood was collected into EDTA tube from these 20 patients having family history of esophageal cancer and 80 non-familial ESCC cases. Blood was also collected from 100 age and sex matched healthy controls from the same ethnic population. The age group of both control and patient is 40 -70 years with mean age of 52.9 years in patients and 58.2 years in controls. Informed consent was obtained from all the patients and controls to use their specimens and clinicopathologic data for this study. Institutional Human Ethics Committee had approved the study.

Isolation of blood genomic DNA

Genomic DNA was extracted by QIAamp DNA Blood Mini Kit (Qiagen, GmbH, Hilden, Germany). 20 µl of QIAGEN Protease (or Proteinase K) was pipette into the bottom of a 1.5 ml microcentrifuge tube. 200 µl of blood sample was added to the microcentrifuge tube and mixed by pulse-vortexing for 15 s. In order to ensure efficient lysis, it is essential that the sample and Buffer AL were mixed thoroughly to yield a homogeneous solution. The mixture was incubated at 56°C for 60 min with mild shaking. 200 µl of ethanol (96–100%) was added to the sample, and mixed again by pulse-vortexing for 15 s. After mixing, briefly centrifuged the 1.5 ml microcentrifuge tube to remove drops from the inside of the lid. The mixture was carefully applied to the QIAamp Spin Column (in a 2 ml collection tube) without wetting the rim, closed the cap, and centrifuged at 6000 x *g* (8000 rpm) for 1 min. QIAamp Spin Column was placed in a clean 2 ml collection tube (provided), and discarded the tube containing the filtrate. Carefully opened the QIAamp Spin Column and 500 µl of Buffer AW1 was added into

the QIAamp Spin Column without wetting the rim. Closed the cap and centrifuged at 6000 x *g* (8000 rpm) for 1 min. QIAamp Spin Column was placed in a clean 2 ml collection tube (provided), and discarded the collection tube containing the filtrate. Carefully opened the QIAamp Spin Column and 500 µl of AW2 Buffer was added into the column without wetting the rim. Closed the cap and centrifuged at full speed (20,000 x *g*; 14,000 rpm) for 3 min. QIAamp Spin Column was placed into a new 2 ml collection tube and discarded the collection tube with the filtrate. QIAamp Spin Column was centrifuged at 20,000 x *g* (14,000 rpm) for 1 min. QIAamp Spin Column was placed in a clean 1.5 ml microcentrifuge tube, and discarded the collection tube containing the filtrate. Carefully opened the QIAamp Spin Column and 50 µl of Buffer AE was added. Incubate at room temperature (15–25°C) for 5 min, and then centrifuged at 6000 x *g* (8000 rpm) for 1 min. For long-term storage of DNA, DNA was stored in eluting Buffer AE at –20°C.

Quantitative estimation of DNA

Isolated DNA was checked for purity and concentration was taken by using NanoDrop® ND-1000 UV-Vis Spectrophotometer (Nanodrop technologies, Rockland, USA).The ratio of absorbance at 260 and 280 nm is used to assess the purity of DNA. A ratio of ~1.8 is generally accepted as "pure" for DNA. If the ratio is appreciably lower, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm.

Concentration of DNA = O.D. (Optical density) at 260 nm x 50x dilution factor (µg/ml).

Qualitative estimation of DNA by agarose gel electrophoresis

0.8% agarose gel in 0.5X TBE buffer was prepared. The solution was kept for cooling at room temperature and then ethidium bromide was added to a final concentration of 0.5µg/ml. Comb was fixed properly at correct position in the gel tank so that complete wells were formed when agarose was added. Agarose gel was mixed properly and poured immediately into the gel tank. After the gel was completely set, comb was carefully removed. Gel was placed in the electrophoresis apparatus and sufficient amount of electrophoresis buffer was added to cover the gel to a depth of about 1mm. 4µl of isolated DNA was mixed with 1µl of 6X gel loading dye and this mixture was loaded carefully into the submerged wells using a micropipette. Gel was run at 70-80 volts for 20- 25 mins. Gel was then removed and visualized under UV light.

Mutation detection:

Amplification of *BRCA2* exons by radioactive PCR

The complete coding regions (27 exons) and exon-intron boundaries for *BRCA2* gene were screened for DNA sequence variants by Heteroduplex analysis (HDX) of PCR amplicons using exon specific primers (Table 20) [197,198]**.** PCR reactions were carried out in a volume of 15 µl with 70–100 ng genomic DNA, 1× PCR buffer (20 mM Tris-Hcl pH 8.4, 50 mM KCl), 1.5 mM $MgCl_2$, 5 mM dNTP mix, 10 µM of both forward and reverse primer, 0.2 U platinum Taq (Invitrogen) and 0.4 µCi [α-P33] dATP (BRIT, Department of Atomic Energy, India). An initial de-naturation of 94°C for 3 min was followed by 40 cycles of amplification (30 s/94°C, 30 s/primer specific annealing temperature, and 30 s/72°C) and final elongation of 3 min/74°C.

Screening for DNA sequence variants in *BRCA2* gene by Heteroduplex (HDX) analysis

Samples were diluted 1:1 in formamide dye (98% formamide, 10 mM NaOH, 0.05% bromophenol blue and 0.05% xylene cyanol) and 5 µl of each was loaded onto a HDX gel (40 × 40 cm; containing 0.5× MDE, 0.6× TBE, 4% glycerol, 400 µl 10%APS, 40 µl TEMED) and run at 8–10 mA for 16–20 hrs in 0.6× TBE at room temperature. Gels were dried under vacuum at 80°C for 2 hrs and exposed to film (KODAK BioMax-MR Amersham, USA) for 10–12 hrs with an intensifying screen. For the possibility of PCR fidelity artifacts to be ruled out, both PCR amplification and gel based heteroduplex analysis were carried out twice for samples that showed altered mobility on HDX gels.

Purification of PCR product for sequencing

PCR products showing an aberrant banding pattern were re-amplified by cold PCR and PCR products were purified by QIAquick Gel Extraction Kit (Qiagen). PCR product (DNA fragment) was excised from the agarose gel with clean, sharp scalpel. 3 volumes of Buffer QG were added to 1 volume of gel and incubated at 50°C for 10 min. After the complete dissolving of gel slice, 1 gel volume of isopropanol (Sigma, Molecular Biology Grade) was added to the sample and mixed properly. QIAquick spin column was placed in a provided 2ml collection tube. To bind DNA, the sample was applied to the QIAquick spin column and centrifuged at 13,000 rpm for 1min. Flow-through was discarded and QIAquick column was placed back in the same collection tube. 0.5ml of Buffer QG was added to QIAquick column and centrifuged for 1min. 0.75ml of Buffer PE was added to QIAquick column for washing the column, kept the column stand for 2-5min after

addition of Buffer PE and centrifuged for 1min. The flow-through was discarded and the column was centrifuged for an additional 1min at 13,000rpm. QIAquick column was placed into a clean 1.5ml microcentrifuge tube. To elute the DNA, 30µl of buffer EB (10mM Tris-Cl, pH8.5) was added to the centre of the column, kept the column stand for 1min and then centrifuged the QIAquick column for 1min. Purified PCR products were kept at -20°C. DNA template quantitation is critical for successful sequencing reactions.

Confirmation of sequence variants by Automated Direct Sequencing

The purified PCR Product was directly sequenced on 3130XL Genetic Analyzer (Applied Biosystems Inc., Foster City, CA, USA) by using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) (Table 21). Data analysis was done by Sequencing Analysis Software V5.2 Patch2 and ABI PRISM SeqScape Software Version 2.X (Applied Biosystems, CA).

Table 20: Primers used for amplification of exons in *BRCA2* gene

Exon	Forward Primer Sequence 5'------3'	Reverse Primer Sequence 5'------3'	Tm	Fragment length (bp)
2	CTCAGTCACATAATAAGGAAT	ACACTGTGACGTACTGGGTTTT	52	256
3	TCTGGGTCACAAATTTGTCTGTCA	TGATTTGCCCAGCATGACAC	55	418
4	AGAATGCAAATTTATAATCCAGAGTA	AAATCAGATTCATCTTTATAGAACAAA	50	249
5	AACAATTTATATGAATGAGAATC	AATTGTTAAGTTTTATTTTTATTA	50	220
6	CCACAAAGAGATAAGTCAGGTA	TGTAAATCTCAGGGCAAAGGTA	55	234
7	TAAGTGAAATAAAGAGTGAA	AACAGAAGTATTAGAGATGAC	50	275
8	TGCTTTTTGATGTCTGACAA	ACATATAGGACCAGGTTTAGAGAC	60	274
9	CTAGTGATTTTAAACTATAATTTTG	GTTCAACTAAACAGAGGACT	50	264
10A	TGCCAAGTACTCAGAATAACCC	CTTTTTGATACCCTGAAATGAAGAAG	60	864
10B	TTTCAGAAAAAGACCTATTAGACA	AAACACAGAAGGAATCGTCATC	60	710
11A	ATTTAGTGAATGTGATTGATGG	TCATTGTCTGAGAAAAGTTC	52	869
11B	TCTAGAGGCAAAGAATCATA	CCTGCTTGGAAAATAACATCTG	52	933
11C	ACAAATGGGCAGGACTCTTAGG	TATCAGTTGGCATTTATTATTTTT	58	908
11D	CTTCAAGTAAATGTCATGATTCTGTT	CATTGATGGCTAAAACTGGTG	58	907
11E	TCATACAGCTAGCGGGAAAAA	TCCTCAACGCAAATATCTTCAT	60	864
11F	TTTCCAAGTAATAATATCCAATGTA	TTGGGATATTAAATGTTCTGGAGTA	55	767
11G	AAAGTAACGAACATTCAGACCAG	AGCATACCAAGTCTACTGAATAAAC	55	866
12	AGAGTCAATACTTTAGCTTTA	AGTGGCTCATGTCTGTAAT	54	318
13	TAAAGCCTATAATTGTCTCA	CTTCTTAACGTTAGTGTCATT	50	270
14	TGCAACAAGGCATATTCCT	CAAAGGGGGAAAACCATCAG	55	609
15	GGCCAGGGGTTGTGCTTTTT	AGGATACTAGTTAATGAAATA	50	314
16	TTTGGTAAATTCAGTTTTGGTTT	GCCAACTTTTTAGTTCGAGA	55	395
17	CAGAGAATAGTTGTA	AGAAACCTTAACCCA	55	306

	GTTGTTGAA	TACTGC		
18	GTGACTTGTTTAAAC AGTGGAA	ATTGAGCATCCTTAG TAAGCA	48	524
19	AAGTGAATATTTTTA AGGCAGTT	TATATGGTAAGTTTC AAGAAT	50	342
20	CACTGTGCCTGGCCT GATAC	TGTCCCTTGTTGCTAT TCTTT	55	401
21	AATCTCCCTTCTTTGG GTGT	CATTTCAACATATTCC TTCCTG	60	318
22	TTTTGTTCTGATTGCT TTTTATTC	AATCATTTTGTTAGTA AGGTCAT	50	314
23	CCACTACTAATGCCC ACAAA	AAAACAAAACAAAA ATTCAACATA	55	367
24	CAGTTTTGATAAGTG CTTGTT	AGCTGGAACTAATCA TAAGA	50	290
25	TTAGAGTTTCCTTTCT TGCATC	AAGCTATTTCCTTATA CTGGA	55	399
26	AAGGAAATACTTTTG GAAACATAA	TTTACTAGGTATACA ACAGAA	50	299
27A	TAGGAGTTAGGGGAG GGAGACTGTGT	CAAGGCTCTTCTCTTT TTGC	55	294
27B	CTGTCTCAGCCCAGA TGACT	TGTTGAACCAGACAA AAGAGC	58	344
27C	TCAATGAAATTTCTCT TTTGGA	TGTGTGGTTTGAAAT TATATTC	50	345

Table 21A: The amount of template to use in a cycle sequencing reaction.

Template	**Quantity**
PCR product:	
100–200 bp	1–3 ng
200–500 bp	3–10 ng
500–1000 bp	5–20 ng
1000–2000 bp	10–40 ng
>2000 bp	20–50 ng

Table 21B: Reaction mixtures for cycle sequencing reaction

Reagent	**Concentration**	**Volume**
Ready Reaction Premix	2.5X	4 µl
BigDye Sequencing Buffer	5X	2 µl
Primer	-	3.2 pmol
Template	-	According to Table2A
Water	-	to 20 µl
Final Volume	1X	20 µl

Table 21C: Cycle sequencing condition on the System 9700

Perform an initial denaturation. a. Rapid thermal ramp to 96 °C b. 96 °C for 1 min
Repeat the following for 25 cycles: • Rapid thermal ramp* to 96 °C • 96 °C for 10 sec • Rapid thermal ramp to 50 °C • 50 °C for 5 sec • Rapid thermal ramp to 60 °C • 60 °C for 4 min *Rapid thermal ramp is 1 °C/second

Rapid thermal ramp to 4 °C and hold until ready to purify.

Table 21D: Purifying Extension Products (Big Dye Terminator v 3.1 Clean up)

The best results are obtained when unincorporated dye terminators are completely removed prior to electrophoresis. Excess dye terminators in sequencing reactions obscure data in the early part of the sequence and can interfere with basecalling. Ethanol/EDTA/sodium acetate precipitation is recommended when good signal from base 1 is required.

1	Add 2 µl of 125 mM EDTA per 20 µl reaction.
2	Add 2 µl of 3 M sodium acetate per 20 µl reaction
3	Add 50 µl of 100% ethanol per reaction
4	Seal the plate with aluminum tape and mix by inverting 4 times
5	Incubate at room temperature for 15 min.
6	Spin at a speed of 3000g for 30 minutes at room temperature.
7	Decant the supernatant by inverting the plate on the paper towels.
8	Spin the inverted plate up to 180g to remove residual supernatant.
9	Add 100 µl of 70% ethanol, seal the plate and spin at 3000g for 5 minutes at room temperature.
10	Repeat steps 7-9 and then 7 again.
11	Add 10 µl of Hi-Di formamide, spin the plate, denature, snapchill and proceed for electrophoresis.

Results:

Epidemiological information of enrolled patients and estimation of risk factors:

Of 92 patients, 45 patients (49%) had family history of esophageal cancer. In 64% (59/92) cases, cancer occurred in the first-degree relatives whereas in 11% (10/92) cases, cancer occurred in the second-degree relatives. In 8.69% (8/92) of the cases, the cancer occurred in spouse. In 2 (2.2%) cases, esophageal cancer involved siblings, and 13 (14.1%) cases had cancer in family with no blood relation (Fig.42). The odds ratio (OR) and 95% confidence interval (CI) were calculated by using a logistic regression model and adjusted for age and gender. The multiple model revealed that the risk were more for the esophageal cancer cases (OR=2.6; 95% CI: 1.6 -3.8) whose pre-degree has positive family history of cancers (Fig 43). The univariate analysis revealed that the risk of developing esophageal cancer was higher in cases where family history showed occurrence of cancers among first-degree relatives i.e., parents, brother & sister, (OR: 3.1; CI: 1.9-5.3) than among second-degree relatives i.e., paternal & maternal grand parents (OR: 1.3; CI: 0.25-3.2). Moreover, the risk of developing esophageal cancer was higher in subjects whose pre-degree suffered from esophageal cancer (OR: 2.4; CI: 1.1-4.1) than from other cancers (OR: 1.1; CI: 0.32-3.3). The subjects with family history of cancer were more likely to develop ESCC if they were tobacco chewers (OR: 4.2; CI: 2.1-5.8) and betel quid users (OR: 3.6; CI: 1.8-4.6) (Fig 44). Demographic, lifestyle cancer risk factors such as smoking, chewing, and alcohol drinking and family history of cancer in twenty cases are shown in Table 22. No associated breast or ovarian cancers were reported in the family members of these patients.

Table 22: - Demographic characteristics of esophageal squamous cell carcinoma cases with family history of esophageal cancer

Patient ID	Age (Years)	Sex	Tobacco chewing habit	Smoking habit	Alcohol use	Betel quid use	Family history of esophageal cancer
EC-116	40	F	Yes	No	No	Yes	Father
EC-88	50	M	No	Yes	Yes	Yes	Cousin brother
EC-99	50	M	No	Yes	No	Yes	Mother
EC-84	50	M	No	Yes	Yes	Yes	Father
EC-53	54	M	Yes	No	Yes	Yes	Elder Brother
EC-69	70	M	Yes	Yes	No	Yes	Mother
EC-124	69	M	No	Yes	Yes	Yes	Brother
EC-129	65	M	Yes	Yes	No	Yes	Sister

EC-83	52	F	No	No	No	Yes	Father
EC-54	45	M	No	Yes	Yes	Yes	Mother
EC-72	32	M	No	Yes	No	Yes	Father
EC-247	45	F	No	No	No	Yes	Mother
EC-248	56	M	Yes	Yes	Yes	Yes	Brother
EC-283	55	F	Yes	No	No	Yes	Father
EC-291	55	M	No	Yes	No	Yes	Brother
EC-65	50	F	No	No	No	Yes	Parental Uncle
EC-217	56	M	Yes	Yes	Yes	Yes	Elder Brother
EC-244	45	M	Yes	Yes	Yes	Yes	Elder brother/ Father
EC-85	85	M	No	Yes	No	Yes	Maternal Grand Father
EC-187	58	M	Yes	Yes	Yes	Yes	Father

Footnotes ‡ Sex M= Male, F= Female

Fig 42:

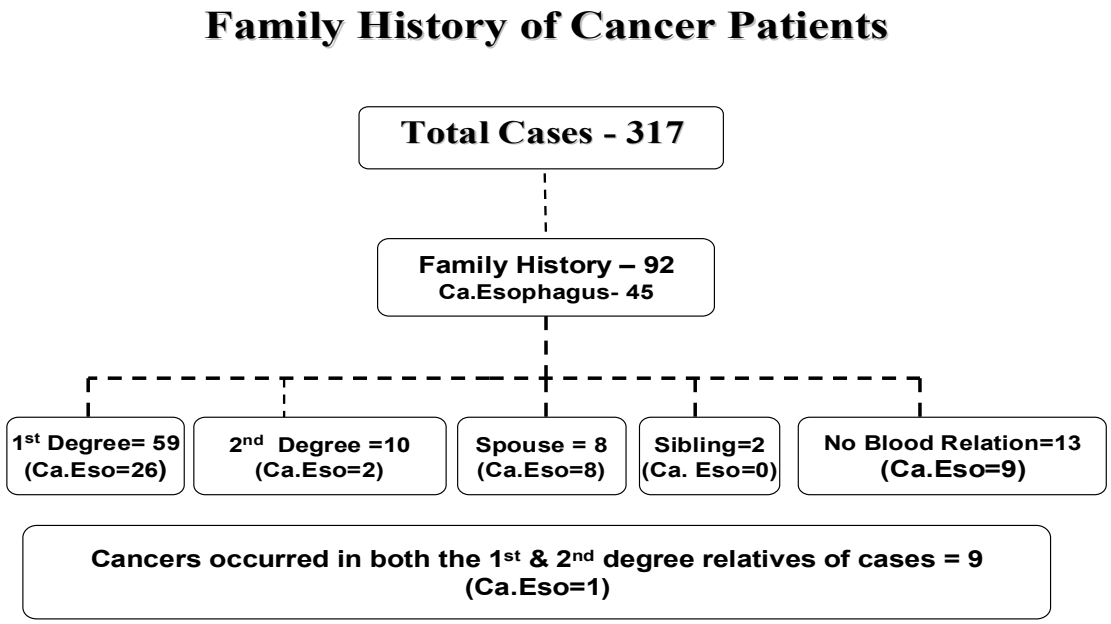


Figure 42. Status of family history of cancer among esophageal squamous cell carcinoma patients from high-risk area of India registered at BBCI during the year of 2005-2006.

Fig 43:

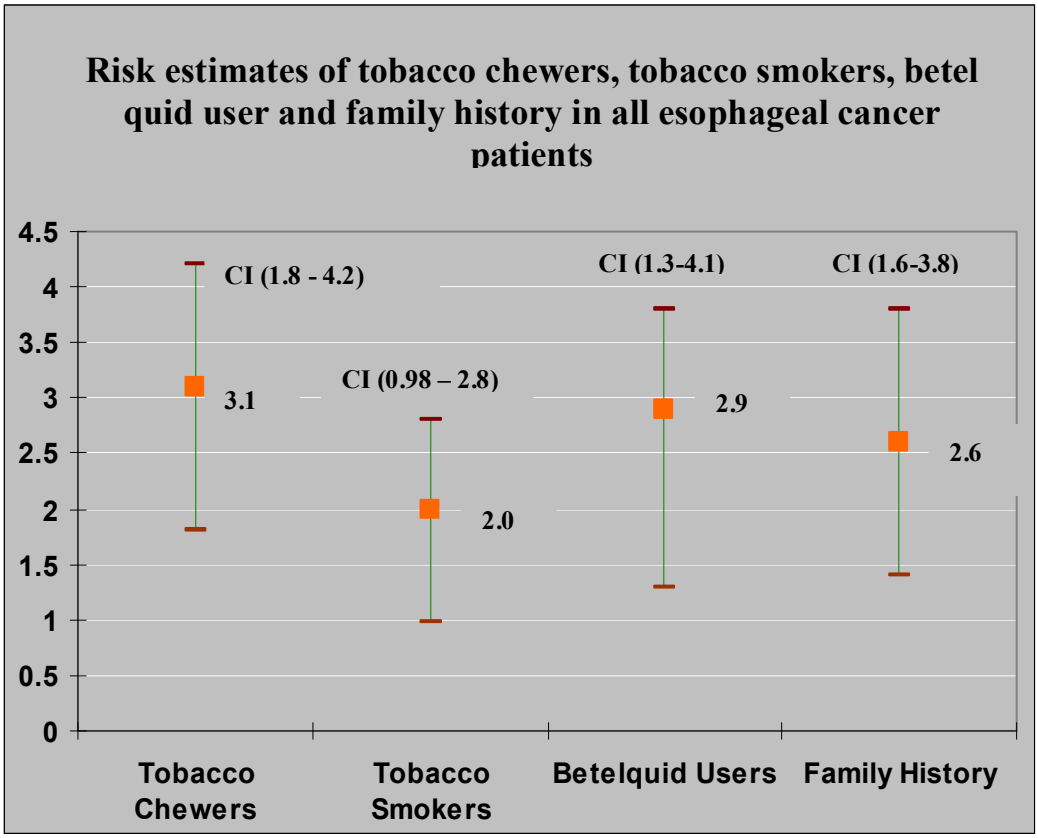


Figure 43. Risk estimates of lifestyle cancer risk factors (e.g., tobacco chewing, smoking, betel quid chewing and family history) in all esophageal cancer patients by Logistic regression model using SPSS 15.0 version software (SPSS, Chicago, IL, USA) package. CI indicates confidence intervals. X- Axis denotes lifestyle cancer risk factors and Y-axis denotes Odd ratio. The odds ratio (OR) and 95% confidence interval (CI) were calculated using a logistic regression model and adjusted for age and gender.

Fig 44

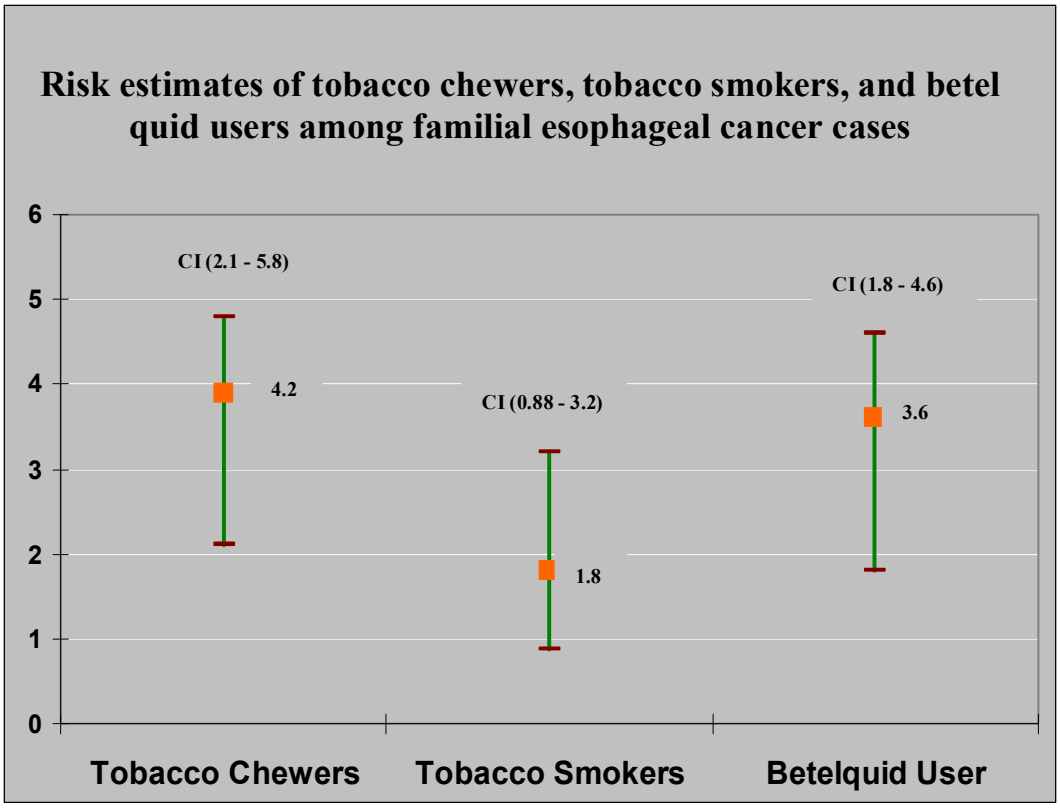


Figure 44. Risk estimates of lifestyle cancer risk factors (e.g., tobacco chewing, smoking, and betel quid chewing) among familial esophageal cancer patients by Logistic regression model using SPSS 15.0 version software (SPSS, Chicago, IL, USA) package. CI indicates confidence intervals. X- Axis denotes lifestyle cancer risk factors and Y-axis denotes Odd ratio. The odds ratio (OR) and 95% confidence interval (CI) were calculated using a logistic regression model and adjusted for age and gender.

***BRCA2* gene sequence variation in germ-line DNA:**

Two germline variants in exons 27 and 18 of *BRCA2* gene were found in three patients with familial ESCC (Fig.45 & Fig.46). A missense variant (10462A>G: I3412V in exon 27) was found in two patients with ESCC with a family history of esophageal cancer in elder brother (EC-217) and father (EC-187), respectively. Another missense variant (8415G>T: K2729N) was found in exon 18 of *BRCA2* gene in one ESCC patient with a family history of esophageal cancer in maternal grand father (Table 23). No sequence alterations were observed in control group (15 vs. 0%, *P*= 0.0037) and 80 non-familial ESCC cases (15 vs.0%, *P* = 0.01).

Fig 45A

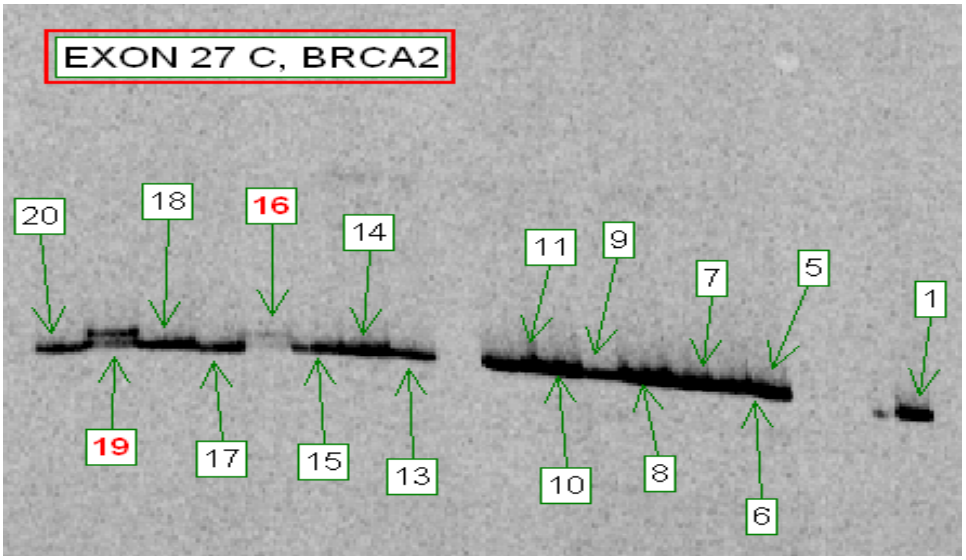


Figure 45A: Heteroduplex analysis showed alteration in exon 27C of *BRCA2* gene at 2 (16: Sample ID 217 & 19: Sample ID 187) out of 20 samples.

Fig 45B

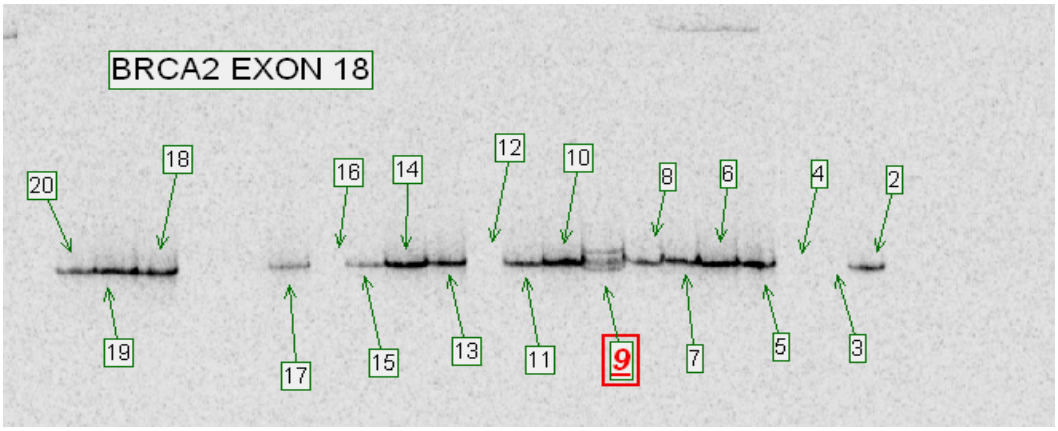


Fig 45B: Heteroduplex analysis showed alteration in exon 18 of *BRCA2* gene at 1 (9: Sample ID 85) out of 20 samples.

Fig 46

Fig 46A

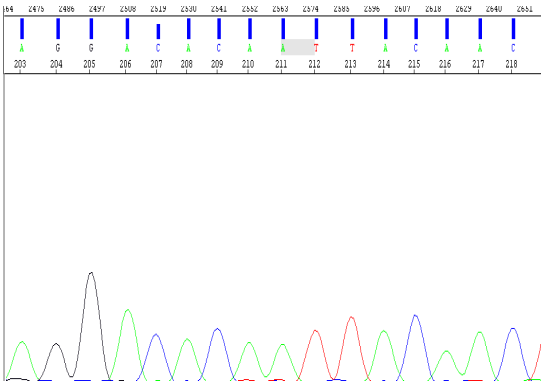


Figure: a

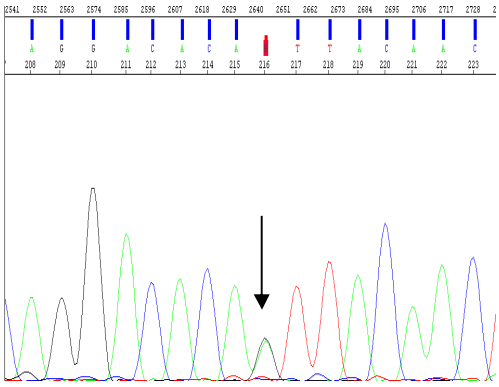


Figure: b (A/G)

Wild type (a: ATT) and mutant (b: GTT) forward sequence of exon 27 of BRCA2 gene

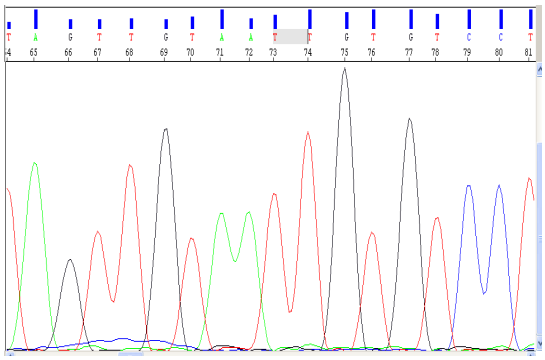


Figure: c

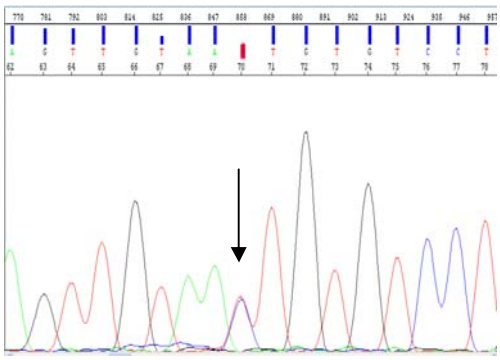


Figure: d (T/C)

Wild type (c: TAA) and mutant (d: CAA) reverse sequence of exon 27 of BRCA2 gene

Fig 46B:

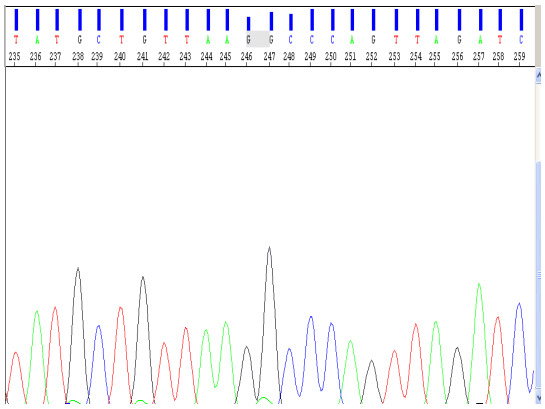


Figure: e

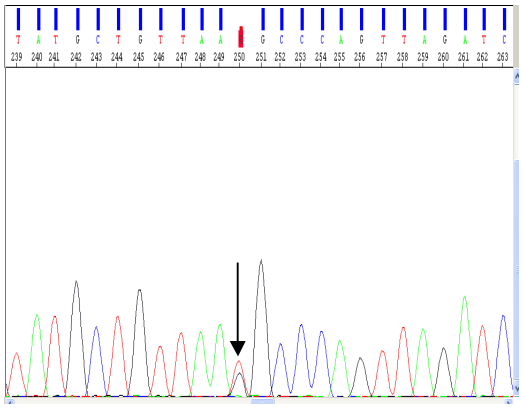


Figure: f (G/T)

Wild type (e: AAG) and mutant (f: AAT) forward sequence of exon 18 of BRCA2 gene

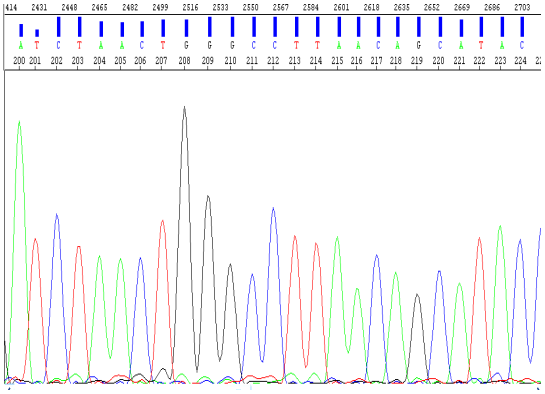


Figure: g

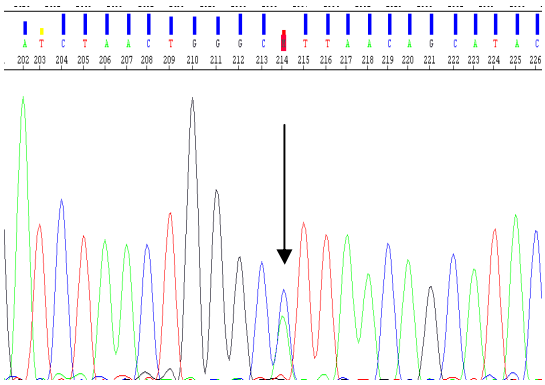


Figure: h (C/A)

Wild type (g: TTC) and mutant (h: TTA) reverse sequence of exon 18 of BRCA2 gene

Figure 46. Sequence chromatograms of missense mutation as determined by automated sequence analysis. Missense mutation in exon 27 (46A) and exon 18 (46B) in *BRCA2* gene. Arrow indicates the position of changes in nucleotide sequences of *BRCA2* gene.

Table 23: Germ-line *BRCA2* variants identified in ESCC patients with family history of esophageal cancer

Patient ID	Exon	Nucleotide change	Amioacid change	Mutation type	Reported in BIC§	Alteration reported by Taylor et.al in Chinese population (2004)†	Alteration reported by MR Akbari in Turkmen population (2008)††
EC-217	27	10462A>G (ATT→GTT)	I3412V (Ile→Val)	Missense mutation	Yes	9 of 70 (13%) ESCC patients	12 out of 197 (6.1%) cases and 20 out of 245 (8.2%) controls.
EC-187	27	10462A>G (ATT→GTT)	I3412V (Ile→Val)	Missense mutation	Yes	9 of 70 (13%) ESCC patients	12 out of 197 (6.1%) cases and 20 out of

							245 (8.2%) controls.
EC-85	18	8415G>T (AAG→AAT)	K2729N (Lys→Asn)	Missense mutation	Yes	2 of 70(3%) ESCC patients and 7 of 232 (3%) healthy controls.	One familial ESCC patient who had a family history of esophageal cancer in both father and mother.

Footnotes §= Breast Cancer Mutation Data Base

†= Oncogene 2004; 23:852-858. ††Oncogene 2008; 27, 1290-1296.

Discussion

Germline mutations in *BRCA2* cause increased susceptibility to breast, ovarian and other cancer types, and have been identified with varying frequencies in individuals of different races and ethnic groups. Most of the deleterious alterations described in *BRCA2* are frameshift mutations that result in a truncated protein; however, in many cases of hereditary breast and ovarian cancer, amino-acid changes of unknown significance are seen. Recent studies reveal that the BRCA2 protein is required for the maintenance of chromosomal stability in mammalian cells and functions in the biological response to DNA damage, as evidenced by the finding that mutations in *BRCA2* lead to chromosomal instability because of defects in repairing of double-strand and single-strand DNA breaks [17].This suggests that genetic changes may result in chromosomal instability and increased genetic susceptibility to cancer.

We screened the entire coding region of *BRCA2* in the germline DNA of 20 ESCC patients with family history of esophageal cancer. Current study showed germ-line variations in *BRCA2* gene in 15% of familial ESCC cases. I3412V is an interesting variant identified in 10% of ESCC patients in our study and this variant result in the conservative substitution of valine for isoleucine at amino-acid position 3412. The terminal region of the BRCA2 protein, where this variant is found, can be entirely truncated. I3421V has earlier been reported in 13% (9/70) of ESCC patients in Chinese population [17]. This variant has also earlier been reported from both cases (12/197; 6.1%) as well as controls (20 /245; 8.2%) in Turkmen population of Iran [15].Thus this variant is common in ESCC cases from high risk population of China, Iran and Indian.

This variant had also been reported in breast Cancer in Asian, African American, Latin American, and Caucasian population (BIC mutation Data Base).

K2729N variant, reported in our study, is located in the conserved *BRCA2* COOH-terminal domain bound to deleted in split-hand/split-foot 1 region (*DSS1*), which can be associated with *BRCA2* in the region of amino acids 2472-2957 [17]. This variant has been reported in 3% (2/70) of ESCC cases and 3% (7/232) of controls in Chinese population and one familial ESCC case in Turkmen population of Iran [15, 17]. This variant which seems to be pathogenic has also been reported in Breast cancer in Asian population [BIC mutation Data Base]. Missense mutations may be pathogenic, depending upon the nature of the amino acid substitution and its effect on protein structure or function. In general, missense alterations in conserved protein motifs are more likely to be deleterious [198].Structural crystallography, of BRCA2 and DSS1 shows that the lysine at codon 2729 of BRCA2 is involved in α-helix and β-sheet structures of oligosaccharide-binding (OB) fold 1 [199]. OB1 is also a site of interaction with FANCG protein – a proposed regulator for unfolding of RAD51 from BRCA2 to the damaged DNA [200]. This amino acid is also located in the binding domain of BRCA2 to MAGE-D1 protein (residues 2393-2952) a synergistic suppressor of cell proliferation [201]. This variant has been reported in Fanconi anemia (FA) patients with biallelic BRCA2 mutations. Therefore, it is tempting to speculate that the *BRCA2* mutation (K2729N) increases the risk for development of ESCC by a mechanism related to FA pathway interruption. However, there may be different mechanism involved for predisposing carriers of BRCA2 mutations to breast cancer or ovarian cancer [15].

Esophageal epithelial cells are exposed to exogenous carcinogens, some of which produce inter-strand DNA links that cannot be repaired in cells with defective HRR. Some other carcinogens cause double-strand DNA breaks repaired by error-prone non-homologous end joining pathway in the absence of intact HRR. Defective HRR in esophageal epithelial cells can result from the mutation of the wild-type copy of BRCA2 in cells that already have a germline BRCA2 mutation, thus leaving no copy of BRCA2 protein capable of binding to FANCD2 and FNACG proteins. Interruption of BRCA2 interaction with these FA pathway genes results in disruption of the FA pathway and consequently causes ineffective HRR. Inappropriate repair of damaged DNA results in loss of genomic integrity and chromosomal instability that eventually leads to cancer development in esophageal epithelial tissue. If HRR is defective in ESCC cells of patients with BRCA2 mutations, it is expected that ESCC tumors in these patients might be more sensitive to chemotherapeutic agents like cis-platin (which makes inter-strand DNA links which require HRR for their repair) than that of radiation which produces double strand breaks repaired by non-homologous end joining pathway [15].

Familial clustering of cancer may be a result of shared environmental factors or shared genes by family members [16]. Environmental factors, which are probably the major contributor for the familial aggregation of upper aero-digestive tract cancers, can significantly modify cancer risk in the presence of an inherited cancer-susceptibility gene [13].

Our results suggest that germ line variants (K2729N and I3412V) of *BRCA2* gene that have earlier been reported from high risk area of China and Iran in ESCC may play a role in familial aggregation of ESCC in high-risk region of India.

Conclusions & Specific Contributions

This is the first study to provide genome-wide analysis of chromosomal alterations and gene expression profile of esophageal cancer in a high-risk region of India.

These findings suggest that the gains and losses of chromosomal regions may contain ESCC-related oncogenes and tumor suppressor genes and provide important theoretic information for identifying and cloning novel ESCC-related oncogenes and tumor suppressor genes.

Genetic alterations of genes on chromosome 3 may be a common molecular mechanism for human epithelial cell carcinogenesis.

Our analysis revealed a characteristic pattern of genomic imbalances in squamous cell carcinoma tissue from patients with esophageal cancer and identified genomic regions associated with tumor initiation, metastasis and high-grade disease.

The target genes located at amplified regions of chromosomes or low-level gain regions such as *PLA2G5* (1p36-p34), *COL11A1* (1p21), *KCNK2* (1q41), *S100A3* (1q21), *ENAH* (1q42.12), *RGS1* (1q31), *KCNH1* (1q32-q41), *INSIG2* (2q14.1), *FGF12* (3q28), *TRIO* (5p15.2), *RNASEN* (5p15.2), *FGF10* (5p13-p12), *EDN1*(6p24.1-p22.3), *SULF1*(8q13.2-13.3), *TLR4* (9q32-q33), *TNC* (9q33), *NTRK2* (9q22.1), *CD44* (11p13), *NCAM1* (11q23.1), *TRIM29* (11q22-q23), *PAK1* (11q13-q14) and *RAB27A* (15q15-q21.1) are found to be associated with cellular migration and proliferation, tumor cell metastasis and invasion,

anchorage independent growth and inhibition of apoptosis. The genes located at amplified regions of chromosomes were mostly involved in MAPK signaling pathway and inflammation.

The target genes located at deleted regions of chromosomes seem to include *FBLN2* (3p25.1), *WNT7A* (3p25), *DLC1* (8p22), *LZTS1* (8p22), *CDKN2A* (9p21), *COL4A1* (13q34), *CDK8* (13q12) and *DCC* (18q21.3) which are found to be associated with the suppression of tumor. The genes located at deleted regions of chromosomes were mostly involved in Wnt signaling pathway, Focal/cell adhesion, and metabolism of xenobiotics by cytochrome P450 pathways which were reported to be associated with tobacco associated carcinogenesis.

The most salient finding was identification of down regulated genes involved in structural constituents of ribosome and up regulated genes involved in cation transporter activity, GPCR activity and MAPK activity in non-familial ESCC cases with history of tobacco and betel quid consumption.

Several deregulated genes identified in the current study are of interest because of their potential role in the natural history of esophageal and other cancers.

Since there is wide spread use of tobacco and fermented betel quid among patients with ESCC in high risk area of India, exposure to tobacco and betel quid constituents may contribute to the development and progression of ESCC in this area by facilitating the deregulation of genes involved in these pathways.

Several genes that showed alteration in our study have also been reported from a high incidence area of esophageal cancer in China**.** These data indicate the consistency of molecular profiles of esophageal cancer in two different geographic locations (China and North-east region of India) that have a high incidence of ESCC but different food habits and customs.

On the basis of functional annotation, genes responsible for inflammatory response, immune response, angiogenesis, cell migration and cell proliferation were found significantly deregulated in patients with ESCC.

Differential expression of cytokeratins played significant contribution in both familial and non-familial ESCC.

Neuoactive ligand receptor interaction pathway showed significant up-regulation in both familial and non-familial ESCC.

Up-regulation of gene involved in B-cell receptor signaling pathway and down regulation of genes involved in Natural killer cell mediated cytotoxicity in familial ESCC suggests that immune response may influence the natural history of ESCC in high-risk area of India.

Validation of *CD14, ARG1, PF4, MAPK7* and *EPHX1* genes at the mRNA level by real-time PCR and *KRT4, COL4, NF-kB* and *VEGF* genes at the protein level by tissue microarray did not show any difference in familial and non-familial ESCC cases, suggesting that familial clustering of cancer in these patients is more due to shared environmental factors rather than shared genes by family members.

The data presented here will not only provide important information about the pathogenesis of esophageal cancer, but also facilitate the identification of candidate genes that could be used as therapeutic targets for the treatment of patients with this tumor.

The use of high throughput genomic technology in clinical specimens from well characterized populations that have familial clustering of cancer may lead to identification of molecular mechanism associated with progression of esophageal cancer.

Germ line variants (K2729N and I3412V) of *BRCA2* gene that have earlier been reported in ESCC from high risk area of China and Iran may play a role in familial aggregation of ESCC in high-risk area of India. Germ-line sequence alterations in *BRCA2* gene in familial ESCC patients from this high-risk area of India suggested that *BRCA2* may play a role in genetic susceptibility to familial ESCC.

Future Scope

Microarrays are used in cancer biology for several applications, such as identification of target genes of tumor suppressors, identification of cancer biomarkers, and identification of genes associated with chemoresistance and drug discovery. Our data provides a rich source of information for further studies into molecular basis of tumorigenesis of the esophagus, as well as identification of biomarkers for early detection of progression. The identification of cancer biomarkers related to the oncogenic process could increase the sensitivity and specificity of conventional histopathologic evaluation by targeting genes whose expression is critical for invasion, metastasis, and cell survival. Gene expression profiles may help to unlock the molecular basis of phenotype, response to treatment, and heterogeneity of disease. They may also help to define patterns of expression that will aid in diagnosis as well as define susceptibility loci that may lead to the identification of individuals at risk. Finally, as specific genes are identified and their functional roles in the development and course of disease are characterized, new targets for therapy would be identified.

References:

1] Murtaza I, Mushtaq D, Margoob MA, Dutt A, Wani NA, Ahmad I, Bhat ML. A study on p53 gene alterations in esophageal squamous cell carcinoma and their correlation to common dietary risk factors among population of the Kashmir valley. World J Gastroenterol. 2006 Jul 7; 12(25):4033-4037.

2] Katiyar S, Hedau S, Jain N, Kar P, Khuroo MS, Mohanta J, Kumar S, Gopalkrishna V, Kumar N, Das BC. p53 gene mutation and human papillomavirus (HPV) infection in esophageal carcinoma from three different endemic geographic regions of India. Cancer Lett. 2005 Jan 31; 218(1):69-79

3] Akbari MR, Malekzadeh R, Nasrollahzadeh D, Amanian D, Sun P, Islami F, Sotoudeh M, Semnani S, Boffeta P, Dawsey SM, Ghadirian P, Narod SA. Familial risks of esophageal cancer among the Turkmen population of the Caspian littoral of Iran. Int J Cancer. 2006 Sep 1; 119(5):1047-1051.

4] Kumar A, Chatopadhyay T, Raziuddin M, Ralhan R. Discovery of deregulation of zinc homeostasis and its associated genes in esophageal squamous cell carcinoma using cDNA microarray. Int J Cancer. 2007 Jan 15; 120(2):230-242.

5] Stoner GD, Gupta A. Etiology and chemoprevention of esophageal squamous cell carcinoma. Carcinogenesis. 2001 Nov; 22(11):1737-1746.

6] Mir MM, Dar NA, Gochhait S, Zargar SA, Ahangar AG, Bamezai RN. p53 mutation profile of squamous cell carcinomas of the esophagus in Kashmir (India): a high incidence area. Int J Cancer. 2005 Aug 10; 116(1):62-68.

7] Kashyap MK, Marimuthu A, Kishore CJ, Peri S, Keerthikumar S, Prasad TS, Mahmood R, Rao S, Ranganathan P, Vijayakumar M, Kumar KV, Montgomery E,

Kumar RV, Pandey A. Genome wide mRNA profiling of esophageal squamous cell carcinoma for identification of cancer biomarkers. Cancer Biol Ther. 2009 Jan 1; 8(1), 1-11.

8] Lee CH, Lee JM, Wu DC, Hsu HK, Kao EL, Huang HL, Wang TN, Huang MC, Wu MT. Independent and combined effects of alcohol intake, tobacco smoking and betel quid chewing on the risk of esophageal cancer in Taiwan. Int J Cancer. 2005 Jan 20; 113(3):475-482.

9] Phukan RK, Ali MS, Chetia CK, Mahanta J. Betel nut and tobacco chewing; potential risk factors of cancer of oesophagus in Assam, India. Br J Cancer. 2001 Sep 1; 85(5):661-667.

10] Su H, Hu N, Shih J, Hu Y, Wang QH, Chuang EY, Roth MJ, Wang C, Goldstein AM, Ding T, Dawsey SM, Giffen C, Emmert-Buck MR, Taylor PR. Gene expression analysis of esophageal squamous cell carcinoma reveals consistent molecular profiles related to a family history of upper gastrointestinal cancer. Cancer Res. 2003 Jul 15; 63(14):3872-3876.

11] Olof Nyren, Hans-Olov Adami, Esophageal Cancer, in: Hans-Olov Adami, David Hunter, Dimitrios Trichopoulos (Eds.), Text Book of Cancer Epidemiology, Oxford University Press, pp.137-155,2002.

12] Ghadirian P. Familial history of esophageal cancer. Cancer. 1985 Oct 15; 56(8):2112-2116.

13] Hemminki K, Rawal R, Chen B, Bermejo JL. Genetic epidemiology of cancer: from families to heritable genes. Int J Cancer. 2004 Oct 10; 111(6):944-950.

14] Tran GD, Sun XD, Abnet CC, Fan JH, Dawsey SM, Dong ZW, Mark SD, Qiao YL,Taylor PR. Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. Int J Cancer. 2005 Jan 20; 113(3):456-463.

15] Akbari MR, Malekzadeh R, Nasrollahzadeh D, Amanian D, Islami F, Li S, Zandvakili I, Shakeri R, Sotoudeh M, Aghcheli K, Salahi R, Pourshams A, Semnani S, Boffetta P, Dawsey SM, Ghadirian P, Narod SA. Germline BRCA2 mutations and the risk of esophageal squamous cell carcinoma. Oncogene. 2008 Feb 21; 27(9):1290-1296.

16] Hemminki K, Li X. Familial risks of cancer as a guide to gene identification and mode of inheritance. Int J Cancer. 2004 Jun 10; 110(2):291-294.

17] Hu N, Wang C, Han XY, He LJ, Tang ZZ, Giffen C, Emmert-Buck MR, Goldstein AM, Taylor PR. Evaluation of BRCA2 in the genetic susceptibility of familial esophageal cancer. Oncogene. 2004 Jan 22; 23(3):852-8.

18] Yang H, Jeffrey PD, Miller J, Kinnucan E, Sun Y, Thoma NH, Zheng N, Chen PL, Lee WH, Pavletich NP. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science. 2002 Sep 13; 297(5588):1837-1848.

19] Song Y, Zhao C, Dong L, Fu M, Xue L, Huang Z, Tong T, Zhou Z, Chen A, Yang Z, Lu N, Zhan Q. Overexpression of cyclin B1 in human esophageal squamous cell carcinoma cells induces tumor cell invasive growth and metastasis. Carcinogenesis. 2008 Feb; 29(2):307-315.

20] Chang D, Wang TY, Li HC, Wei JC, Song JX. Prognostic significance of PTEN expression in esophageal squamous cell carcinoma from Linzhou City, a high incidence area of northern China. Dis Esophagus. 2007; 20(6):491-496.

21] Si HX, Tsao SW, Lam KY, Srivastava G, Liu Y, Wong YC, Shen ZY, Cheung AL. E-cadherin expression is commonly downregulated by CpG island hypermethylation in esophageal carcinoma cells. Cancer Lett. 2001 Nov 8; 173(1):71-78.

22] Lin YC, Wu MY, Li DR, Wu XY, Zheng RM. Prognostic and clinicopathological features of E-cadherin, alpha-catenin, beta-catenin, gamma-catenin and cyclin D1 expression in human esophageal squamous cell carcinoma. World J Gastroenterol. 2004 Nov 15; 10(22):3235-3239.

23] Liu W, Yu ZC, Cao WF, Ding F, Liu ZH. Functional studies of a novel oncogene TGM3 in human esophageal squamous cell carcinoma. World J Gastroenterol. 2006 Jun 28; 12(24):3929-3932.

24] Uemura N, Nakanishi Y, Kato H, Saito S, Nagino M, Hirohashi S, Kondo T.Transglutaminase 3 as a prognostic biomarker in esophageal cancer revealed by proteomics. Int J Cancer. 2009 May 1; 124(9):2106-15.

25] Tanaka E, Hashimoto Y, Ito T, Okumura T, Kan T, Watanabe G, Imamura M, Inazawa J, Shimada Y. The clinical significance of Aurora A/STK15/BTAK expression in human esophageal squamous cell carcinoma. Clin Cancer Res. 2005 Mar 1; 11(5):1827-1834.

26] Tong T, Zhong Y, Kong J, Dong L, Song Y, Fu M, Liu Z, Wang M, Guo L, Lu S, Wu M, Zhan Q. Overexpression of Aurora-A contributes to malignant development of human esophageal squamous cell carcinoma. Clin Cancer Res. 2004 Nov 1; 10(21):7304-7310.

27] Hu YC, Lam KY, Law S, Wong J, Srivastava G. Identification of differentially expressed genes in esophageal squamous cell carcinoma (ESCC) by cDNA expression

array: overexpression of Fra-1, Neogenin, Id-1, and CDC25B genes in ESCC. Clin Cancer Res. 2001 Aug; 7(8):2213-2221.

28] Ding MX, Lin XQ, Fu XY, Zhang N, Li JC. Expression of vascular endothelial growth factor-C and angiogenesis in esophageal squamous cell carcinoma. World J Gastroenterol. 2006 Jul 28; 12(28):4582-4585.

29] Krzystek-Korpacka M, Matusiewicz M, Diakowska D, Grabowski K, Blachut K, Banas T. Up-regulation of VEGF-C secreted by cancer cells and not VEGF-A correlates with clinical evaluation of lymph node metastasis in esophageal squamous cell carcinoma (ESCC). Cancer Lett. 2007 May 8; 249(2):171-177.

30] Matsumoto M, Natsugoe S, Okumura H, Arima H, Yanagita S, Uchikado Y, Yokomakura N, Setoyama T, Ishigami S, Takao S, Aikou T. Overexpression of vascular endothelial growth factor-C correlates with lymph node micrometastasis in submucosal esophageal cancer. J Gastrointest Surg. 2006 Jul-Aug; 10(7):1016-1022.

31] Nozoe T, Ezaki T, Kabashima A, Baba H, Maehara Y. Significance of immunohistochemical expression of cyclooxygenase-2 in squamous cell carcinoma of the esophagus. Am J Surg. 2005 Jan; 189(1):110-115.

32] Zhi H, Wang L, Zhang J, Zhou C, Ding F, Luo A, Wu M, Zhan Q, Liu Z. Significance of COX-2 expression in human esophageal squamous cell carcinoma. Carcinogenesis. 2006 Jun; 27(6):1214-1221.

33] Fumoto S, Hiyama K, Tanimoto K, Noguchi T, Hihara J, Hiyama E, Noguchi T, Nishiyama M. EMP3 as a tumor suppressor gene for esophageal squamous cell carcinoma. Cancer Lett. 2009 Feb 8; 274(1):25-32.

34] Sugiura H, Ishiguro H, Kuwabara Y, Kimura M, Mitsui A, Mori Y, Ogawa R, Katada T, Harata K, Fujii Y. Decreased expression of RUNX3 is correlated with tumor progression and poor prognosis in patients with esophageal squamous cell carcinoma. Oncol Rep. 2008 Mar;19(3):713-719.

35] Shou JZ, Hu N, Takikita M, Roth MJ, Johnson LL, Giffen C, Wang QH, Wang C, Wang Y, Su H, Kong LH, Emmert-Buck MR, Goldstein AM, Hewitt SM, Taylor PR. Overexpression of CDC25B and LAMC2 mRNA and protein in esophageal squamous cell carcinomas and premalignant lesions in subjects from a high-risk population in China. Cancer Epidemiol Biomarkers Prev. 2008 Jun; 17(6):1424-1435.

36] Morita M, Oyama T, Nakata S, Ono K, Sugaya M, Uramoto H, Yoshimatsu T, Hanagiri T, Sugio K, Yasumoto K. Expression of FHIT in esophageal epithelium and carcinoma: reference to drinking, smoking and multicentric carcinogenesis. Anticancer Res. 2006 May-Jun; 26(3B):2243-2248.

37] Li W, Ding F, Zhang L, Liu Z, Wu Y, Luo A, Wu M, Wang M, Zhan Q, Liu Z. Overexpression of stefin A in human esophageal squamous cell carcinoma cells inhibits tumor cell growth, angiogenesis, invasion, and metastasis. Clin Cancer Res. 2005 Dec 15; 11(24 Pt 1):8753-8762.

38] Zhi H, Zhang J, Hu G, Lu J, Wang X, Zhou C, Wu M, Liu Z. The deregulation of arachidonic acid metabolism-related genes in human esophageal squamous cell carcinoma. Int J Cancer. 2003 Sep 1; 106(3):327-333.

39] Wang LS, Chow KC, Wu YC, Lin TY, Li WY. Inverse expression of dihydrodiol dehydrogenase and glutathione-S-transferase in patients with esophageal squamous cell carcinoma. Int J Cancer. 2004 Aug 20; 111(2):246-251.

40] Ikeguchi M, Kaibara N. survivin messenger RNA expression is a good prognostic biomarker for oesophageal carcinoma. Br J Cancer. 2002 Oct 7; 87(8):883-887.

41] Wu IC, Wu MT, Chou SH, Yang SF, Goan YG, Lee JM, Chou YP, Bair MJ, Wang TE, Chen A, Chang WH, Kuo FC, Wu DC. Osteopontin expression in squamous cell cancer of the esophagus. World J Surg. 2008 Sep; 32(9):1989-1895.

42] Albertson DG. Gene amplification in cancer. Trends Genet. 2006 Aug; 22(8):447-455.

43] Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet. 1999 Sep; 23(1):41-46.

44] Hu N, Roth MJ, Emmert-Buck MR, Tang ZZ, Polymeropolous M, Wang QH, Goldstein AM, Han XY, Dawsey SM, Ding T, Giffen C, Taylor PR. Allelic loss in esophageal squamous cell carcinoma patients with and without family history of upper gastrointestinal tract cancer. Clin Cancer Res. 1999 Nov; 5(11):3476-3482.

45] LaFramboise T, Weir BA, Zhao X, Beroukhim R, Li C, Harrington D, Sellers WR, Meyerson M. Allele-specific amplification in cancer revealed by SNP array analysis. PLoS Comput Biol. 2005 Nov; 1(6):e65. Epub 2005 Nov 25.

46] Carneiro A, Isinger A, Karlsson A, Johansson J, Jönsson G, Bendahl PO, Falkenback D, Halvarsson B, Nilbert M. Prognostic impact of array-based genomic profiles in esophageal squamous cell cancer. BMC Cancer. 2008 Apr 11; 8:98.

47] Yen CC, Chen YJ, Chen JT, Hsia JY, Chen PM, Liu JH, Fan FS, Chiou TJ, Wang WS, Lin CH. Comparative genomic hybridization of esophageal squamous cell

carcinoma: correlations between chromosomal aberrations and disease progression/prognosis. Cancer. 2001 Dec 1; 92(11):2769-2777.

48] Zhou J, Zhao LQ, Xiong MM, Wang XQ, Yang GR, Qiu ZL, Wu M, Liu ZH. Gene expression profiles at different stages of human esophageal squamous cell carcinoma. World J Gastroenterol. 2003 Jan; 9(1):9-15.

49] Lu J, Liu Z, Xiong M, Wang Q, Wang X, Yang G, Zhao L, Qiu Z, Zhou C, Wu M. Gene expression profile changes in initiation and progression of squamous cell carcinoma of esophagus. Int J Cancer. 2001 Feb 1; 91(3):288-294.

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

51] Greenawalt DM, Duong C, Smyth GK, Ciavarella ML, Thompson NJ, Tiang T, Murray WK, Thomas RJ, Phillips WA. Gene expression profiling of esophageal cancer: comparative analysis of Barrett's esophagus, adenocarcinoma, and squamous cell carcinoma. Int J Cancer. 2007 May 1; 120(9):1914-1921.

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

53] Hashibe M, Boffetta P, Janout V, Zaridze D, Shangina O, Mates D, Szeszenia-Dabrowska N, Bencko V, Brennan P. Esophageal cancer in Central and Eastern Europe: tobacco and alcohol. Int J Cancer. 2007 Apr 1; 120(7):1518-1522.

54] Luo A, Kong J, Hu G, Liew CC, Xiong M, Wang X, Ji J, Wang T, Zhi H, Wu M, Liu Z. Discovery of Ca2+-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. Oncogene. 2004 Feb 12; 23(6):1291-9.

55] Mandard AM, Hainaut P, Hollstein M. Genetic steps in the development of squamous cell carcinoma of the esophagus. Mutat Res. 2000 Apr; 462(2-3):335-342.

56] Kimos MC, Wang S, Borkowski A, Yang GY, Yang CS, Perry K, Olaru A, Deacu E, Sterian A, Cottrell J, Papadimitriou J, Sisodia L, Selaru FM, Mori Y, Xu Y, Yin J, Abraham JM, Meltzer SJ. Esophagin and proliferating cell nuclear antigen (PCNA) are biomarkers of human esophageal neoplastic progression. Int J Cancer. 2004 Sep 1; 111(3):415-417.

57] Hu YC, Lam KY, Law S, Wong J, Srivastava G. Profiling of differentially expressed cancer-related genes in esophageal squamous cell carcinoma (ESCC) using human cancer cDNA arrays: overexpression of oncogene MET correlates with tumor differentiation in ESCC. Clin Cancer Res. 2001 Nov; 7(11):3519-3525.

58] Borg A. Molecular and pathological characterization of inherited breast cancer. Semin Cancer Biol. 2001 Oct; 11(5):375-385.

59] Mohr S, Leikauf GD, Keith G, Rihn BH. Microarrays as cancer keys: an array of possibilities. J Clin Oncol. 2002 Jul 15; 20(14):3165-3175.

60] Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog. 1999 Mar; 24(3):153-159.

61] Yang YH, Speed T. Design issues for cDNA microarray experiments. Nat Rev Genet. 2002 Aug; 3(8):579-588.

62] Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. Nat Rev Genet. 2006 Jan; 7(1):55-65.

63] Butte A. The use and analysis of microarray data. Nat Rev Drug Discov. 2002 Dec; 1(12):951-960.

64] Fan J, Ren Y. Statistical analysis of DNA microarray data in cancer research. Clin Cancer Res. 2006 Aug 1; 12(15):4469-4473.

65] Valasek MA, Repa JJ. The power of real-time PCR. Adv Physiol Educ. 2005 Sep; 29(3):151-159.

66] Bustin SA, Benes V, Nolan T, Pfaffl MW. Quantitative real-time RT-PCR--a perspective. J Mol Endocrinol. 2005 Jun;34(3):597-601.

67] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) Method. Methods. 2001 Dec; 25(4):402-408.

68] Yuan JS, Reed A, Chen F, Stewart CN Jr. Statistical analysis of real-time PCR data. BMC Bioinformatics. 2006 Feb 22; 7:85.

69] Hewitt SM. Tissue microarrays as a tool in the discovery and validation of tumor markers. Methods Mol Biol. 2009; 520:151-161.

70] Zheng HT, Peng ZH, Li S, He L. Loss of heterozygosity analyzed by single nucleotide polymorphism array in cancer. World J Gastroenterol. 2005 Nov 21; 11(43):6740-6744.

71] Zhao X, Li C, Paez JG, Chin K, Jänne PA, Chen TH, Girard L, Minna J, Christiani D, Leo C, Gray JW, Sellers WR, Meyerson M. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res. 2004 May 1; 64(9):3060-3071.

72] Chen J, Guo L, Peiffer DA, Zhou L, Chan OT, Bibikova M, Wickham-Garcia E, Lu SH, Zhan Q, Wang-Rodriguez J, Jiang W, Fan JB. Genomic profiling of 766 cancer-

related genes in archived esophageal normal and carcinoma tissues. Int J Cancer. 2008 May 15; 122(10):2249-2254.

73] Wiech T, Nikolopoulos E, Weis R, Langer R, Bartholomé K, Timmer J, Walch AK, Höfler H, Werner M. Genome-wide analysis of genetic alterations in Barrett's adenocarcinoma using single nucleotide polymorphism arrays. Lab Invest. 2009 Apr; 89(4):385-397.

74] Hu N, Wang C, Hu Y, Yang HH, Kong LH, Lu N, Su H, Wang QH, Goldstein AM, Buetow KH, Emmert-Buck MR, Taylor PR, Lee MP. Genome-wide loss of heterozygosity and copy number alteration in esophageal squamous cell carcinoma using the Affymetrix GeneChip Mapping 10 K array. BMC Genomics. 2006 Nov 29; 7:299.

75] Qin YR, Fu L, Sham PC, Kwong DL, Zhu CL, Chu KK, Li Y, Guan XY. Single-nucleotide polymorphism-mass array reveals commonly deleted regions at 3p22 and 3p14.2 associate with poor clinical outcome in esophageal squamous cell carcinoma. Int J Cancer. 2008 Aug 15; 123(4):826-830.

76] Tørring N, Borre M, Sørensen KD, Andersen CL, Wiuf C, Ørntoft TF. Genome-wide analysis of allelic imbalance in prostate cancer using the Affymetrix 50K SNP mapping array. Br J Cancer. 2007 Feb 12; 96(3):499-506.

77] Gorringe KL, Jacobs S, Thompson ER, Sridhar A, Qiu W, Choong DY, Campbell IG. High-resolution single nucleotide polymorphism array analysis of epithelial ovarian cancer reveals numerous microdeletions and amplifications. Clin Cancer Res. 2007 Aug 15; 13(16):4731-4739.

78] Mannello F, Qin W, Zhu W, Fabbri L, Tonti GA, Sauter ER. Nipple aspirate fluids from women with breast cancer contain increased levels of group IIa secretory phospholipase A2. Breast Cancer Res Treat. 2008 Sep; 111(2):209-218.

79] Tribler L, Jensen LT, Jørgensen K, Brünner N, Gelb MH, Nielsen HJ, Jensen SS. Increased expression and activity of group IIA and X secretory phospholipase A2 in peritumoral versus central colon carcinoma tissue. Anticancer Res. 2007 Sep-Oct; 27(5A):3179-3185.

80] Fischer H, Stenling R, Rubio C, Lindblom A. Colorectal carcinogenesis is associated with stromal expression of COL11A1 and COL5A2. Carcinogenesis. 2001 Jun; 22(6):875-878.

81] Balakrishnan A, Bleeker FE, Lamba S, Rodolfo M, Daniotti M, Scarpa A, van Tilborg AA, Leenstra S, Zanon C, Bardelli A. Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma. Cancer Res. 2007 Apr 15; 67(8):3545-3550.

82] Voloshyna I, Besana A, Castillo M, Matos T, Weinstein IB, Mansukhani M, Robinson RB, Cordon-Cardo C, Feinmark SJ. TREK-1 is a novel molecular target in prostate cancer. Cancer Res. 2008 Feb 15; 68(4):1197-1203.

83] Liu J, Li X, Dong GL, Zhang HW, Chen DL, Du JJ, Zheng JY, Li JP, Wang WZ.In silico analysis and verification of S100 gene expression in gastric cancer. BMC Cancer. 2008 Sep 16; 8:261.

84] Di Modugno F, Mottolese M, Di Benedetto A, Conidi A, Novelli F, Perracchio L, Venturo I, Botti C, Jager E, Santoni A, Natali PG, Nisticò P. The cytoskeleton regulatory protein hMena (ENAH) is overexpressed in human benign breast lesions with high risk of transformation and human epidermal growth factor receptor-2-positive/hormonal receptor-negative tumors. Clin Cancer Res. 2006 Mar 1; 12(5):1470-1478.

85] Wong YF, Cheung TH, Tsao GS, Lo KW, Yim SF, Wang VW, Heung MM, Chan SC, Chan LK, Ho TW, Wong KW, Li C, Guo Y, Chung TK, Smith DI. Genome-wide gene expression profiling of cervical cancer in Hong Kong women by oligonucleotide microarray. Int J Cancer. 2006 May 15; 118(10):2461-2469.

86] Pardo LA, Stühmer W. Eag1: an emerging oncological target. Cancer Res. 2008 Mar 15; 68(6):1611-1613.

87] Ding XW, Yan JJ, An P, Lü P, Luo HS. Aberrant expression of ether à go-go potassium channel in colorectal cancer patients and cell lines. World J Gastroenterol. 2007 Feb 28; 13(8):1257-1261.

88] Li CG, Gruidl M, Eschrich S, McCarthy S, Wang HG, Alexandrow MG, Yeatman TJ. Insig2 is associated with colon tumorigenesis and inhibits Bax-mediated apoptosis. Int J Cancer. 2008 Jul 15; 123(2):273-282.

89] Katoh M, Katoh M. Integrative genomic analyses on HES/HEY family: Notch-independent HES1, HES3 transcription in undifferentiated ES cells, and Notch-dependent HES1, HES5, HEY1, HEY2, HEYL transcription in fetal tissues, adult tissues, or cancer. Int J Oncol. 2007 Aug; 31(2):461-466.

90] Stewart GD, Gray K, Pennington CJ, Edwards DR, Riddick AC, Ross JA, Habib FK. Analysis of hypoxia-associated gene expression in prostate cancer: lysyl oxidase and glucose transporter-1 expression correlate with Gleason score. Oncol Rep. 2008 Dec; 20(6):1561-1567.

91] Zheng M, Simon R, Mirlacher M, Maurer R, Gasser T, Forster T, Diener PA, Mihatsch MJ, Sauter G, Schraml P. TRIO amplification and abundant mRNA expression is associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer. Am J Pathol. 2004 Jul; 165(1):63-69.

92] Sugito N, Ishiguro H, Kuwabara Y, Kimura M, Mitsui A, Kurehara H, Ando T, Mori R,Takashima N, Ogawa R, Fujii Y. RNASEN regulates cell proliferation and affects survival in esophageal cancer patients. Clin Cancer Res. 2006 Dec 15; 12(24):7322-7328.

93] Yu C, Zhang X, Sun G, Guo X, Li H, You Y, Jacobs JL, Gardner K, Yuan D, Xu Z, Du Q, Dai C, Qian Z, Jiang K, Zhu Y, Li QQ, Miao Y. RNA interference-mediated silencing of the polo-like kinase 1 gene enhances chemosensitivity to gemcitabine in pancreatic adenocarcinoma cells. J Cell Mol Med. 2008 Dec; 12(6A):2334-2349.

94] Hurst CD, Tomlinson DC, Williams SV, Platt FM, Knowles MA. Inactivation of the Rb pathway and overexpression of both isoforms of E2F3 are obligate events in bladder tumours with 6p22 amplification. Oncogene. 2008 Apr 24; 27(19):2716-2727

95] Wang Y, Ma Y, Lü B, Xu E, Huang Q, Lai M. Differential expression of mimecan and thioredoxin domain-containing protein 5 in colorectal adenoma and cancer: a proteomic study. Exp Biol Med (Maywood). 2007 Oct; 232(9):1152-1159.

96] Bobryshev YV, Tran D, Killingsworth MC, Buckland M, Lord RV. Dendritic cells in Barrett's esophagus and esophageal adenocarcinoma. J Gastrointest Surg. 2009 Jan; 13(1):44-53.

97] Valdés-Mora F, Gómez del Pulgar T, Bandrés E, Cejas P, Ramírez de Molina A, Pérez-Palacios R, Gallego-Ortega D, García-Cabezas MA, Casado E, Larrauri J, Nistal M, González-Barón M, García-Foncillas J, Lacal JC. TWIST1 overexpression is associated with nodal invasion and male sex in primary colorectal cancer. Ann Surg Oncol. 2009 Jan; 16(1):78-87.

98] Mhawech-Fauceglia P, Alvarez V, Fischer G, Beck A, Herrmann FR. Association of TSC1/hamartin, 14-3-3sigma, and p27 expression with tumor outcomes in patients with pTa/pT1 urothelial bladder carcinoma. Am J Clin Pathol. 2008 Jun; 129(6):918-923.

99] Degen M, Brellier F, Schenk S, Driscoll R, Zaman K, Stupp R, Tornillo L, Terracciano L, Chiquet-Ehrismann R, Rüegg C, Seelentag W. Tenascin-W, a new marker of cancer stroma, is elevated in sera of colon and breast cancer patients. Int J Cancer. 2008 Jun 1; 122(11):2454-2461.

100] Qiu L, Zhou C, Sun Y, Di W, Scheffler E, Healey S, Kouttab N, Chu W, Wan Y. Crosstalk between EGFR and TrkB enhances ovarian cancer cell migration and proliferation. Int J Oncol. 2006 Oct; 29(4):1003-1011.

101] Kuo YC, Su CH, Liu CY, Chen TH, Chen CP, Wang HS. Transforming growth factor-beta induces CD44 cleavage that promotes migration of MDA-MB-435s cells through the up-regulation of membrane type 1-matrix metalloproteinase. Int J Cancer. 2009 Jun 1; 124(11):2568-2576.

102] Katoh M, Katoh M. Comparative genomics on mammalian Fgf3-Fgf4 locus. Int J Oncol. 2005 Jul; 27(1):281-285.

103] Frame MC, Inman GJ. NCAM is at the heart of reciprocal regulation of E-cadherin- and integrin-mediated adhesions via signaling modulation. Dev Cell. 2008 Oct; 15(4):494-496.

104] Kosaka Y, Inoue H, Ohmachi T, Yokoe T, Matsumoto T, Mimori K, Tanaka F, Watanabe M, Mori M. Tripartite motif-containing 29 (TRIM29) is a novel marker for lymph node metastasis in gastric cancer. Ann Surg Oncol. 2007 Sep; 14(9):2543-2549.

105] Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I, Freed E, Ligon AH, Vena N, Ogino S, Chheda MG, Tamayo P, Finn S, Shrestha Y, Boehm JS, Jain S, Bojarski E, Mermel C, Barretina J, Chan JA, Baselga J, Tabernero J, Root DE, Fuchs CS, Loda M, Shivdasani RA, Meyerson M, Hahn WC. CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. Nature. 2008 Sep 25; 455(7212):547-551.

106] Sheng S, Barnett DH, Katzenellenbogen BS. Differential estradiol and selective estrogen receptor modulator (SERM) regulation of Keratin 13 gene expression and its underlying mechanism in breast cancer cells. Mol Cell Endocrinol. 2008 Dec 16; 296(1-2):1-9.

107] Wu M, Soler DR, Abba MC, Nunez MI, Baer R, Hatzis C, Llombart-Cussac A, Llombart-Bosch A, Aldaz CM. CtIP silencing as a novel mechanism of tamoxifen resistance in breast cancer. Mol Cancer Res. 2007 Dec; 5(12):1285-1295.

108] Liang J, Wang JB, Pan YL, Wang J, Liu LL, Guo XY, Sun L, Lin T, Han S, Xie HH, Yin F, Guo XG, Fan D. High frequency occurrence of 1-OPRD variant of PRNP gene in gastric cancer cell lines and Chinese population with gastric cancer. Cell Biol Int. 2006 Nov; 30(11):920-923.

109] Fujiwara I, Yashiro M, Kubo N, Maeda K, Hirakawa K. Ulcerative colitis-associated colorectal cancer is frequently associated with the microsatellite instability pathway. Dis Colon Rectum. 2008 Sep; 51(9):1387-1394

110] Yi CH, Smith DJ, West WW, Hollingsworth MA. Loss of fibulin-2 expression is associated with breast cancer progression. Am J Pathol. 2007 May; 170(5):1535-1545.

111] Ohira T, Gemmill RM, Ferguson K, Kusy S, Roche J, Brambilla E, Zeng C, Baron A,Bemis L, Erickson P, Wilder E, Rustgi A, Kitajewski J, Gabrielson E, Bremnes R,Franklin W, Drabkin HA. WNT7a induces E-cadherin in lung cancer cells. Proc Natl Acad Sci U S A. 2003 Sep 2; 100(18):10429-10434.

112] Jin Y, Tian X, Shang Y, Huang P. Inhibition of DLC-1 gene expression by RNA interference in the colon cancer LoVo cell line. Oncol Rep. 2008 Mar; 19(3):669-674.

113] Ullmannova-Benson V, Guan M, Zhou X, Tripathi V, Yang XY, Zimonjic DB, Popescu NC. DLC1 tumor suppressor gene inhibits migration and invasion of multiple myeloma cells through RhoA GTPase pathway. Leukemia. 2009 Feb; 23(2):383-90.

114] Lahoz A, Hall A. DLC1: a significant GAP in the cancer genome. Genes Dev. 2008 Jul 1; 22(13):1724-1730.

115] Ishii H, Vecchione A, Murakumo Y, Baldassarre G, Numata S, Trapasso F, Alder H, Baffa R, Croce CM. FEZ1/LZTS1 gene at 8p22 suppresses cancer cell growth and regulates mitosis. Proc Natl Acad Sci U S A. 2001 Aug 28; 98(18):10374-10379.

116] Vecchione A, Ishii H, Baldassarre G, Bassi P, Trapasso F, Alder H, Pagano F, Gomella LG, Croce CM, Baffa R. FEZ1/LZTS1 is down-regulated in high-grade bladder cancer, and its restoration suppresses tumorigenicity in transitional cell carcinoma cells. Am J Pathol. 2002 Apr; 160(4):1345-1352.

117] Nonaka D, Fabbri A, Roz L, Mariani L, Vecchione A, Moore GW, Tavecchio L, Croce CM, Sozzi G.Reduced FEZ1/LZTS1 expression and outcome prediction in lung cancer. Cancer Res. 2005 Feb 15; 65(4):1207-1212.

118] Frank B, Bermejo JL, Hemminki K, Sutter C, Wappenschmidt B, Meindl A,Kiechle-Bahat M, Bugert P, Schmutzler RK, Bartram CR, Burwinkel B. Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. Carcinogenesis. 2007 Jul; 28(7):1442-1445.

119] Hagemann C, Anacker J, Gerngras S, Kühnel S, Said HM, Patel R, Kämmerer U, Vordermark D, Roosen K, Vince GH. Expression analysis of the autosomal recessive primary microcephaly genes MCPH1 (microcephalin) and MCPH5 (ASPM, abnormal spindle-like, microcephaly associated) in human malignant gliomas. Oncol Rep. 2008 Aug; 20(2):301-308.

120] Chaplet M, Rai R, Jackson-Bernitsas D, Li K, Lin SY. BRIT1/MCPH1: a guardian of genome and an enemy of tumors. Cell Cycle. 2006 Nov; 5(22):2579-2583.

121] Pils D, Horak P, Gleiss A, Sax C, Fabjani G, Moebus VJ, Zielinski C, Reinthaller A, Zeillinger R, Krainer M. Five genes from chromosomal band 8p22 are significantly

down-regulated in ovarian carcinoma: N33 and EFA6R have a potential impact on overall survival. Cancer. 2005 Dec 1; 104(11):2417-2429.

122] Sowd G, Lei M, Opresko PL. Mechanism and substrate specificity of telomeric protein POT1 stimulation of the Werner syndrome helicase. Nucleic Acids Res. 2008 Aug; 36(13):4242-4256.

123] Qin YR, Wang LD, Fan ZM, Kwong D, Guan XY. Comparative genomic hybridization analysis of genetic aberrations associated with development of esophageal squamous cell carcinoma in Henan, China. World J Gastroenterol. 2008 Mar 28; 14(12):1828-1835.

124] Huang Y, Fan J, Yang J, Zhu GZ. Characterization of GPR56 protein and its suppressed expression in human pancreatic cancer cells. Mol Cell Biochem. 2008 Jan; 308(1-2):133-139.

125] Park HL, Kim MS, Yamashita K, Westra W, Carvalho AL, Lee J, Jiang WW, Baek JH, Liu J, Osada M, Moon CS, Califano JA, Mori M, Sidransky D. DCC promoter hypermethylation in esophageal squamous cell carcinoma. Int J Cancer. 2008 Jun 1; 122(11):2498-2502.

126] Furne C, Rama N, Corset V, Chédotal A, Mehlen P. Netrin-1 is a survival factor during commissural neuron navigation. Proc Natl Acad Sci U S A. 2008 Sep 23; 105(38):14465-14470.

127] Cecener G, Tunca B, Egeli U, Karadag M, Vatan O, Uzaslan E, Tolunay S. Mutation analysis of the FHIT gene in bronchoscopic specimens from patients with suspected lung cancer. Tumori. 2008 Nov-Dec; 94(6):845-848.

128] Tokumaru Y, Yamashita K, Kim MS, Park HL, Osada M, Mori M, Sidransky D. The role of PGP9.5 as a tumor suppressor gene in human cancer. Int J Cancer. 2008 Aug 15; 123(4):753-759.

129] Wu K, Katiyar S, Witkiewicz A, Li A, McCue P, Song LN, Tian L, Jin M, Pestell RG. The cell fate determination factor dachshund inhibits androgen receptor signaling and prostate cancer cellular growth. Cancer Res. 2009 Apr 15; 69(8):3347-3355.

130] Zhu Z, Friess H, Kleeff J, Wang L, Wirtz M, Zimmermann A, Kore M, Buchler MW. Glypican-3 expression is markedly decreased in human gastric cancer but not in esophageal cancer. Am J Surg. 2002 Jul; 184(1): 78-83.

131] Sung YK, Hwang SY, Farooq M, Kim JC, Kim HK. Growth promotion of HepG2 hepatoma cells by antisense-mediated knockdown of glypican-3 is independent of insulin-like growth factor2 signaling. Exp Mol Med. 2003 Aug 31; 35 (4): 257-262.

132] Hu XY, Xu YM, Fu Q, Yu JJ, Huang J. Nedd4L expression is downregulated in prostate cancer compared to benign prostatic hyperplasia. Eur J Surg Oncol. 2009 May; 35(5):527-531.

133] Kim SK, Jang HR, Kim JH, Kim M, Noh SM, Song KS, Kang GH, Kim HJ, Kim SY, Yoo HS, Kim YS. CpG methylation in exon 1 of transcription factor 4 increases with age in normal gastric mucosa and is associated with gene silencing in intestinal-type gastric cancers. Carcinogenesis. 2008 Aug; 29(8):1623-1631.

134] Allin KH, Bojesen SE, Nordestgaard BG. Baseline C-Reactive Protein Is Associated With Incident Cancer and Survival in Patients With Cancer. J Clin Oncol. 2009 Mar 16.

135] Nomura S, Yoshitomi H, Takano S, Shida T, Kobayashi S, Ohtsuka M, Kimura F,Shimizu H, Yoshidome H, Kato A, Miyazaki M. FGF10/FGFR2 signal induces cell migration and invasion in pancreatic cancer. Br J Cancer. 2008 Jul 22; 99(2):305-313.

136] Bagnato A, Catt KJ. Endothelins as autocrine regulators of tumor cell growth. Trends Endocrinol Metab. 1998 Nov; 9(9):378-383.

137] Lee JG, Zheng R, McCafferty-Cepero JM, Burnstein KL, Nanus DM, Shen R. Endothelin-1 enhances the expression of the androgen receptor via activation of the c-myc pathway in prostate cancer cells. Mol Carcinog. 2009 Feb; 48(2):141-149.

138] Zhang WM, Zhou J, Ye QJ. Endothelin-1 enhances proliferation of lung cancer cells by increasing intracellular free Ca2+.Life Sci. 2008 Mar 26; 82(13-14):764-771.

139] Medina PP, Castillo SD, Blanco S, Sanz-Garcia M, Largo C, Alvarez S, Yokota J, Gonzalez-Neira A, Benitez J, Clevers HC, Cigudosa JC, Lazo PA, Sanchez-Cespedes M. The SRY-HMG box gene, SOX4, is a target of gene amplification at chromosome 6p in lung cancer. Hum Mol Genet. 2009 Apr 1; 18(7):1343-1352.

140] Coe BP, Lee EH, Chi B, Girard L, Minna JD, Gazdar AF, Lam S, MacAulay C, Lam WL. Gain of a region on 7p22.3, containing MAD1L1, is the most frequent event in small-cell lung cancer cell lines. Genes Chromosomes Cancer. 2006 Jan; 45(1):11-19.

141] Li J, Kleeff J, Abiatari I, Kayed H, Giese NA, Felix K, Giese T, Büchler MW, Friess H. Enhanced levels of Hsulf-1 interfere with heparin-binding growth factor signaling in pancreatic cancer. Mol Cancer. 2005 Apr 7; 4(1):14.

142] Li Q, Mullins SR, Sloane BF, Mattingly RR. p21-Activated kinase 1 coordinates aberrant cell survival and pericellular proteolysis in a three-dimensional culture model for premalignant progression of human breast cancer. Neoplasia. 2008 Apr; 10(4):314-329.

143] Wang JS, Wang FB, Zhang QG, Shen ZZ, Shao ZM. Enhanced expression of Rab27A gene by breast cancer cells promoting invasiveness and the metastasis potential by secretion of insulin-like growth factor-II. Mol Cancer Res. 2008 Mar; 6(3):372-382.

144] Rohatgi **N**, Matta A, Kaur J, Srivastava A, Ralhan R. Novel molecular targets of smokeless tobacco (khaini) in cell culture from oral hyperplasia. Toxicology 2006 Jul 5; 224(1-2): 1-13.

145] Chang MC, Wu HL, Lee JJ, Lee PH, Chang HH, Hahn LJ, Lin BR, Chen YJ, Jeng JH. The induction of prostaglandin E2 production, interleukin-6 production, cell cycle arrest, and cytotoxicity in primary oral keratinocytes and KB cancer cells by areca nut ingredients is differentially regulated by MEK/ERK activation. J Biol Chem. 2004 Dec 3; 279(49): 50676-50683.

146] Zhong CY, Zhou YM, Douglas GC, Witschi H, Pinkerton KE. MAPK/AP-1 signal pathway in tobacco smoke-induced cell proliferation and squamous metaplasia in the lungs of rats. Carcinogenesis. 2005 Dec; 26(12): 2187-2195.

147] Arredondo J, Chernyavsky AI, Jolkovsky DL, Pinkerton KE, Grando SA. Receptor-mediated tobacco toxicity: cooperation of the Ras/Raf-1/MEK1/ERK and JAK-2/STAT-3 pathways downstream of alpha7 nicotinic receptor in oral keratinocytes. FASEB J. 2006 Oct; 20(12): 2093-2101.

148] Sud **N**, Sharma R, Ray R, Chattopadhyay TK, Ralhan R. Differential expression of G-protein coupled receptor 56 in human esophageal squamous cell carcinoma. Cancer Lett. 2006 Feb 28; 233(2): 265-270.

149] Helm J, Enkemann SA, Coppola D, Barthel JS, Kelley ST, Yeatman TJ. Dedifferentiation precedes invasion in the progression from Barrett's metaplasia to esophageal adenocarcinoma. Clin Cancer Res. 2005 Apr 1; 11(7): 2478-2485.

150] Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. J Natl Cancer Inst. 2002 Apr 3; 94(7): 513-521.

151] Xuejun Peng, Constance L Wood, Eric M Blalock, Kuey Chu Chen, Philip W Landfield and Arnold J Stromberg. Statistical implications of pooling RNA samples for microarray experiments. BMC Bioinformatics 2003, 4:26.

152] Samimi G, Manorek G, Castel R, Breaux JK, Cheng TC, Berry CC, Los G, Howell SB. cDNA microarray-based identification of genes and pathways associated with oxaliplatin resistance. Cancer Chemother Pharmacol. 2005 Jan; 55(1): 1-11.

153] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998 Dec 8; 95(25): 14863-14868.

154] Diehn M, Sherlock G, Binkley G, Jin H, Matese JC, Hernandez-Boussard T, ReesCA, Cherry JM, Botstein D, Brown PO, Alizadeh AA. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Res. 2003 Jan 1; 31(1): 219-223.

155] Altermann E, Klaenhammer TR. Pathway Voyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC Genomics. 2005 May 3; 6(1): 60.

156] Reddy KB, Nabha SM, Atanaskova N. Role of MAP kinase in tumor progression and invasion. Cancer Metastasis Rev. 2003 Dec; 22(4): 395-403.

157] Mossman BT, Lounsbury KM, Reddy SP. Oxidants and signaling by mitogen-activated protein kinases in lung epithelium. Am J Respir Cell Mol Biol. 2006 Jun; 34(6): 666-669.

158] Burdick AD, Davis JW 2nd, Liu KJ, Hudson LG, Shi H, Monske ML, Burchiel SW. Benzo(a)pyrene quinones increase cell proliferation, generate reactive oxygen species, and transactivate the epidermal growth factor receptor in breast epithelial cells. Cancer Res. 2003 Nov 15; 63(22): 7825-7833.

159] Demirjian L, Abboud RT, Li H, Duronio V. Acute effect of cigarette smoke on TNF-alpha release by macrophages mediated through the erk1/2 pathway. Biochim Biophys Acta. 2006 Jun; 1762(6): 592-597.

160] Rasiah KK, Kench JG, Gardiner-Garden M, Biankin AV, Golovsky D, Brenner PC, Kooner R, O'neill GF, Turner JJ, Delprado W, Lee CS, Brown DA, Breit SN, Grygiel JJ, Horvath LG, Stricker PD, Sutherland RL, Henshall SM. Aberrant neuropeptide Y and macrophage inhibitory cytokine-1 expression are early events in prostate cancer development and are associated with poor prognosis. Cancer Epidemiol Biomarkers Prev. 2006 Apr; 15(4): 711-716.

161] Amlal H, Faroqui S, Balasubramaniam A, Sheriff S. Estrogen up-regulates neuropeptide Y Y1 receptor expression in a human breast cancer cell line. Cancer Res. 2006 Apr 1; 66(7): 3706-3714.

162] Karl Kunzelmann. Ion Channels and Cancer. J. Membrane Biol. 2005; 205**:** 159–173.

163] Joshi N, Johnson LL, Wei WQ, Abnet CC, Dong ZW, Taylor PR, Limburg PJ, Dawsey SM, Hawk ET, Qiao YL, Kirsch IR. Gene expression differences in normal esophageal mucosa associated with regression and progression of mild and moderate squamous dysplasia in a high-risk Chinese population. Cancer Res. 2006 Jul 1; 66(13): 6851-6860.

164] Chen BS, Wang MR, Xu X, Cai Y, Xu ZX, Han YL, Wu M. Transglutaminase-3, an esophageal cancer-related gene. Int J Cancer. 2000 Dec 15; 88(6): 862-865.

165] Jiang WG, Ablin R, Douglas-Jones A, Mansel RE. Expression of transglutaminases in human breast cancer and their possible clinical significance. Oncol Rep. 2003 Nov-Dec; 10(6): 2039-2044.

166] Lewis TE, Milam TD, Klingler DW, Rao PS, Jaggi M, Smith DJ, Hemstreet GP, Balaji KC. Tissue transglutaminase interacts with protein kinase A anchor protein 13 in prostate cancer. Urol Oncol. 2005 Nov-Dec; 23(6): 407-412.

167] Polley AC, Mulholland F, Pin C, Williams EA, Bradburn DM, Mills SJ, Mathers JC, Johnson IT. Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia. Cancer Res. 2006 Jul 1; 66(13): 6553-6562.

168] Shiraishi T, Mori M, Tanaka S, Sugimachi K, Akiyoshi T. Identification of cystatin B in human esophageal carcinoma, using differential displays in which the gene expression is related to lymph-node metastasis. Int J Cancer. 1998 Apr 17; 79(2): 175-178.

169] Yoo YA, Kim MJ, Park JK, Chung YM, Lee JH, Chi SG, Kim JS, Yoo YD. Mitochondrial ribosomal protein L41 suppresses cell growth in association with p53 and p27Kip1. Mol Cell Biol. 2005 Aug; 25(15): 6603-6616.

170] Yoshida A, Urasaki Y, Waltham M, Bergman AC, Pourquier P, Rothwell DG, Inuzuka M, Weinstein JN, Ueda T, Appella E, Hickson ID, Pommier Y. Human apurinic/apyrimidinic endonuclease (Ape1) and its N-terminal truncated form (AN34) are involved in DNA fragmentation during apoptosis. J Biol Chem. 2003 Sep 26; 278(39): 37768-37776.

171] Zhu Y, Lin H, Li Z, Wang M, Luo J. Modulation of expression of ribosomal protein L7a (rpL7a) by ethanol in human breast cancer cells. Breast Cancer Res Treat. 2001 Sep; 69(1): 29-38.

172] Phukan R.K, Mahanta J.; Hazarika N.C: Annual Report, Regional Medical Research Center, Dibrugarh, Assam, India. pp 21-24:2005-2006.

173] Chattopadhyay I, Kapur S, Purkayastha J, Phukan R, Kataki A, Mahanta J, Saxena S. Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis. World J Gastroenterol. 2007 Mar 7; 13 (9): 1438-1444.

174] Xue LY, Hu N, Song YM, Zou SM, Shou JZ, Qian LX, Ren LQ, Lin DM, Tong T, He ZG, Zhan QM, Taylor PR, Lu N. Tissue microarray analysis reveals a tight

correlation between protein expression pattern and progression of esophageal squamous cell carcinoma. BMC Cancer. 2006 Dec 22; 6:296.

175] Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 Jan7; 100 (1): 57-70.

176] S.Perwez Hussain and Curtis C. Harris. Inflammation and cancer: An ancient link with novel potentials. Int.J.Cancer 2007; 121: 2373-2380.

177] Federico A, Morgillo F, Tuccillo C, Ciardiello F, Loguercio C. Chronic inflammation and oxidative stress in human carcinogenesis. Int J Cancer. 2007; 121: 2381-2386.

178] Ikeda K, Iyama K, Ishikawa N, Egami H, Nakao M, Sado Y, Ninomiya Y, Baba H. Loss of expression of type IV collagen alpha5 and alpha6 chains in colorectal cancer associated with the hypermethylation of their promoter region. Am. J. Pathol. 2006 Mar; 168 (3): 856-865.

179] Yamaguchi K, Ogawa K, Katsube T, Shimao K, Konno S, Shimakawa T, Yoshimatsu K, Naritaka Y, Yagawa H, Hirose K. Platelet factor 4-gene transfection into tumor cells inhibits angiogenesis, tumor growth and metastasis. Anticancer Res. 2005 Mar-April; 25 (2A): 847-851.

180] Sulpice E, Contreres JO, Lacour J, Bryckaert M, Tobelem G. Platelet factor 4 disrupts the intracellular signalling cascade induced by vascular endothelial growth factor by both KDR dependent and independent mechanisms. Eur J Biochem. 2004; 271: 3310-3318.

181] Sohn SJ, Sarvis BK, Cado D, Winoto A. ERK5 MAPK regulates embryonic angiogenesis and acts as a hypoxia-sensitive repressor of vascular endothelial growth factor expression. J Biol Chem. 2002; 277: 43344-43351.

182] Kiyohara C, Yoshimasu K, Takayama K, Nakanishi Y: EPHX1 polymorphisms and the risk of lung cancer: a HuGE review. Epidemiology. 2006 Jan; 17 (1): 89-99.

183] Shoshan-Barmatz V, Israelson A, Brdiczka D, Sheu SS. The voltage-dependent anion channel (VDAC): function in intracellular signalling, cell life and cell death. Curr Pharm Des. 2006; 12 (18): 2249-2270.

184] Qiu T, Grizzle WE, Oelschlager DK, Shen X, Cao X: Control of prostate cell growth. BMP antagonizes androgen mitogenic activity with incorporation of MAPK signals in Smad1. EMBO J. 2007 Jan 24; 26 (2): 346-357.

185] Huang B, Zhao J, Unkeless JC, Feng ZH, Xiong H. TLR signaling by tumor and immune cells: a double-edged sword. Oncogene. 2008; 27: 218-224.

186] Bourguignon LY, Peyrollier K, Gilad E, Brightman A. Hyaluronan-CD44 interaction with neural Wiskott-Aldrich syndrome protein (N-WASP) promotes actin polymerization and ErbB2 activation leading to beta-catenin nuclear translocation, transcriptional up-regulation, and cell migration in ovarian tumor cells. J Biol Chem. 2007 Jan 12; 282 (2): 1265-1280.

187] Yang Y, Lee JH, Kim KY. The interferon-inducible 9-27 gene modulates the susceptibility to natural killer cells and the invasiveness of gastric cancer cells. Cancer Letters. 2005 Apr 28; 221(2): 191-200.

188] Rodriguez PC, Hernandez CP, Quiceno D. Arginase I in myeloid suppressor cells is induced by COX-2 in lung carcinoma. J Exp Med. 2005 Oct 3; 202 (7): 931-939.

189] Mielczarek M, Chrzanowska A, Scibior D, Skwarek A, Ashamiss F, Lewandowska K, Barańczyk-Kuźma A. Arginase as a useful factor for the diagnosis of colorectal cancer liver metastases. Int J Biol Markers. 2006; 21: 40-44.

190] Nabarro S, Himoudi N, Papanastasiou A. Coordinated oncogenic transformation and inhibition of host immune responses by the PAX3-FKHR fusion oncoprotein. J Exp Med. 2005; 202: 1399-1410.

191] Kim JH, Kim MK, Lee HE. Constitutive phosphorylation of the FOXO1A transcription factor as a prognostic variable in gastric cancer. Mod Pathol. 2007 May 11; 20: 835-842.

192] Gascoyne DM, Kypta RM, Vivanco MM. Glucocorticoids inhibit apoptosis during fibrosarcoma development by transcriptionally activating Bcl-xL. J Biol Chem. 2003; 278: 18022-18029.

193] Runnebaum IB, Bruning A. Glucocorticoids inhibit cell death in ovarian cancer and up-regulate caspase inhibitor cIAP2. Clin Cancer Res. 2005 Sep1; 11 (17): 6325-6332.

194] Park SY, Lee SA, Han IH. Clinical significance of metabotropic glutamate receptor 5 expression in oral squamous cell carcinoma. Oncol Rep. 2007 Jan; 17 (1): 81-87.

195] Wooster R, Bignell G, Lancaster J. Identification of the breast cancer susceptibility gene BRCA2. Nature 1995; 378: 789-792.

196] Pellegrini L, Yu Ds, Lo T, Anand S, Lee M, Blundell TL, Venkitaraman AR. Insights into DNA recombination from the structure of RAD51-BRCA2 complex. Nature: 2002; 420:287-293.

197] Saxena S, Szabo CI, Chopin S, Barjhoux L, Sinilnikova O, Lenoir G, Goldgar DE, Bhatanager D. BRCA1 and BRCA2 in Indian breast cancer patients. Hum Mutat. 2002 Dec; 20 (6): 473-474.

198] Saxena S, Chakraborty A, Kaushal M, Kaushal M, Kotwal S, Bhatanager D, Mohil RS, Chintamani C, Aggarwal AK, Sharma VK, Sharma PC, Lenoir G, Goldgar DE, Szabo CI. Contribution of germline BRCA1 and BRCA2 sequence alterations to breast cancer in Northern India. BMC Med Genet. 2006 Oct 4; 7:75.

199] Yang H, Jeffrey PD, Miller J. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science 2002; 297:1837-1848.

200] Hussain S, Witt E, Huber PA, Medhurst AL, Ashworth A, Mathew CG. Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1. Hum Mol Genet 2003; 12: 2503-2510.

201] Tian XX, BRCA2 suppress cell proliferation via stabilizing MAGE-D1. Cancer Res 2005; 65:4747-4753.

Appendices

A. 10 X Tris EDTA (TE), pH8.0

100 mM Tris-Cl & 10mM EDTA. Sterilize solutions by autoclaving. Store the buffer at room temperature.

B. 1M Tris-Cl, pH 8.0

Dissolve 121.1 gm of Tris base in 800 ml of autoclaved distilled water. Adjust the pH 8.0 by adding 42ml of concentrated HCl. Sterilize solutions by autoclaving. Store the buffer at room temperature.

C. 0.5M EDTA, pH 8.0

Add 186.1gm of disodium EDTA.$2H_2O$ to 800 ml of autoclaved distilled water. Stir vigorously on a magnetic stirrer. Adjust the pH to 8.0 with NaOH. Sterilize solutions by autoclaving. Store the buffer at room temperature.

D. Ethidium Bromide (10mg/ml)

Add 1gm of ethidium bromide to 100 ml of autoclaved distilled water. Stir on a magnetic stirrer for several hours to ensure that the dye has dissolved. Transfer the solution to a dark bottle. Store at room temperature.

E. 20XSSC

Dissolve 175.3 gm of NaCl (3M) and 88.2 gm of sodium citrate (0.3M) in 800 ml of autoclaved distilled water. Adjust the pH to 7.0 with a few drops of a 14N solution of HCl. Adjust the volume to 1L with autoclaved distilled water. Sterilize solutions by autoclaving. Store the buffer at room temperature.

F. 5X TBE

Add the following items:
54 gm of Tris base.
27.5 gm of boric acid.
20 ml of 0.5 M EDTA, pH 8.0

G. 10X MOPS Electrophoresis Buffer

Dissolve 41.8 gm of MOPS (3'N Morpholino Propanesulfonic acid) in 700 ml of sterile 0.1% DEPC (Diethyl pyrocarbonate) water. Adjust the pH to 7.0 with 2N NaOH. Add 20 ml of DEPC-treated 1M sodium acetate and 20 ml of DEPC treated 0.5M EDTA, pH 8.0. Adjust the volume of the solution to 1L with DEPC treated water. Sterilize solutions by autoclaving. Store the buffer at room temperature in dark condition.

H. 6X Gel-loading Buffer

Add the following items:
0.25% (w/v) bromophenol blue
0.25% (w/v) xylene cyanol FF
30% (v/v) glycerol in autoclaved distilled water. Store at 4°C.

I. 10X Formaldehyde Gel-loading Buffer

50% (v/v) glycerol diluted in DEPC treated water.
10mM EDTA, pH8.0
0.25% (w/v) bromophenol blue
0.25% (w/v) xylene cyanol FF

J. Prepare the following Hybridization Cocktail Master Mix (for multiple samples make 5% excess):

Reagent	1X	Final Conc. In Sample	For 8 Samples
MES (12X; 1.22 M)	12 µL	0.056 M	96µl
DMSO (100%)	13 µL	5.0%	104 µl
Denhardt's Solution (50X)	13 µL	2.50	104 µl
EDTA (0.5 M)	3 µL	5.77 mM	24 µl
HSDNA (10 mg/mL)	3 µL	0.115 mg/mL	24 µl
Oligonucleotide Control	2 µL	1X	16 µl
Human Cot-1 (1 mg/mL)	3 µL	11.5 µg/mL	24 µl
Tween-20 (3%)	1 µL	0.0115%	8 µl
TMACL (5M)	140 µL	2.69 M	1120 µl
Total	**190 µL**		1520 µl

K. 12 X MES Stock Buffer
(1.22 M MES, 0.89 M [Na+])
Do not autoclave. Store between 2ºC and 8ºC, and shield from light.
Discard solution if it turns yellow.
For 1000 mL:
70.4 g MES Hydrate
193.3 g MES Sodium Salt
800 mL Molecular Biology Grade water
Mix and adjust volume to 1,000 mL.
The pH should be between 6.5 and 6.7.
Filter through a 0.2 µM filter.

L. Wash A: Non-Stringent Wash Buffer
(6X SSPE, 0.01% Tween 20)
For 1000 mL:
300 mL of 20X SSPE
1.0 mL of 10% Tween-20
699 mL of water
Filter through a 0.2 µm filter.
Store at room temperature.

M. Wash B: Stringent Wash Buffer
(0.6X SSPE, 0.01% Tween 20)
For 1000 mL:
30 mL of 20X SSPE
1.0 mL of 10% Tween-20
969 mL of water
Filter through a 0.2 µm filter.
Store at room temperature.

N. 0.5 mg/mL Anti-Streptavdin Antibody
Resuspend 0.5 mg in 1 mL of water.
Store at 4°C.

O. 1X Array Holding Buffer

(Final 1X concentration is 100 mM MES, 1M [Na+], 0.01% Tween-20)

For 100 mL:

8.3 mL of 12X MES Stock Buffer
18.5 mL of 5M NaCl
0.1 mL of 10% Tween-20
73.1 mL of water
Store at 2°C to 8°C, and shield from light
Do not autoclave. Store at 2°C to 8°C, and shield from light. Discard solution if yellow.

P. Stain Buffer for 8 samples:-

Reagent	1X	Master mix for 8 samples
H_20	666.7µl	5333 µl
SSPE(20X)	300 µl	2400 µl
Tween-20(3%)	3.3 µl	26.4 µl
Denhardt's (50X)	20 µl	160 µl
Total	990 µl	7920 µl

Q. Antibody Stain Buffer/ SAPE Stain Solution:-

Stain Buffer:-5940µl.
Antibody/SAPE:-60µl.
Total Volume=6000µl.

Publications: -

1) Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis. **Indranil Chattopadhyay,** Sujala Kapur, Joydeep Purkayastha, Rupkumar Phukan, Amal Kataki, Jagadish Mahanta, Sunita Saxena**. World Journal of Gastroenterology, 2007 March 7; 13 (9): 1438-1444. PMID: 17457978**

2) **Chattopadhyay I,** Phukan R, Singh A, Vasudevan M, Purkayastha J, Hewitt S, Kataki A, Mahanta J, Kapur S, Saxena S. Molecular profiling to identify molecular mechanism in esophageal cancer with familial clustering. **Oncol Rep. 2009 May; 21(5):1135-46. PubMed PMID: 19360286.**

3) Avninder Singh, MD; Sujala Kapur, MD; **Indranil Chattopadhyay**, MSc; Joydeep Purkayastha, MS; Jagannath Sharma, MD; Ashwani Mishra, PhD; Stephen M. Hewitt, MD, PhD; Sunita Saxena, MD . Cytokeratin immunoexpression in esophageal squamous cell carcinoma of high-risk populationin Northeast India. **Applied Immunohistochemistry & Molecular Morphology (2009, May 4, PMID: 19417629)**

4) Mishi Kaushal, **Indranil Chattopadhyay**, Rupkumar Phukan, Joydeep Purkayastha, Jagadish Mahanta, Sujala Kapur, Sunita Saxena. Contribution of germline *BRCA2* sequence alterations to risk of familial esophageal cancer in high-risk area of India. **Diseases of the Esophagus. (Accepted: 2009 March 31, Published Online: May 15 2009).**

International Conference, Meeting & Symposium Proceedings/ Abstract: -

1) Genetic Profile of Patients with Esophageal Cancer in a High-incidence Region of Northeast India and Its Association with Tobacco Consumption**.** Sujala Kapur**, Indranil Chatterjee,** Afshaan Noor, J. Mahanta, R.K. Phukan, M.N. Barooah, Bedanshu Saikia, A.C. Kataki, Joydeep Purkayastha and Sunita Saxena**. UICC World Cancer Congress, 2006, Published by Medimond: 311- 315**

2) Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis. Sunita Saxena, **Indranil Chatterjee**, Joydeep Purkayastha, Sujala Kapur. **Cancer Conference at National Cancer Research Institute, 30 September-3 October, 2007, Birmingham, UK.**

3) Prevalence of Glutathione-S-transferase (GST) polymorphisms in tobacco-associated malignancies in high risk Northeast Indian population. D.S.Yadav, Th.Regina, R.Ihsan, P.S.Chauhan, **I.Chattopadhyay**, A.C.Kataki, J.Sharma, E.Zomawia, Y.Verma, S.Kapur, S.Saxena. **International Symposium on Cancer Biology, National Institute of Immunology, New Delhi, India, November14-16, 2007.**

4) Alteration of Molecular Functional Pathways in Esophageal Cancer and its association with tobacco and betel quid consumption. **Chattopadhyay I,** Kapur S, Saxena S. **International Symposium on Chromosomes to Genomes, Centre for Cellular and Molecular Biology, Hyderabad, July3-5, 2007.**

5) Differential expression of MAPK and GPCR pathway in esophageal cancer of North-east region of India. (Poster No 388) **Indranil Chattopadhyay**, Avninder Pal Singh, Sujala Kapur, Rupkumar Phukan, Joydeep Purkayastha, Jagadish Mahanta, Hewitt

Stephen, Jaganath Sharma, Sunita Saxena. **Human Genome meeting (HGM2008Workshops), Hyderabad, 27th September-30 September.**

6) Differential gene expression profile of stomach and oral cancer in high risk region of India. (Poster No. 449) Regina D Thoudam, Dhirendra S Yadav, **I Chattopadhyay**, AC Kataki, E Zamoawia, S Kapur, S Saxena. **Human Genome meeting (HGM2008Workshops), Hyderabad, 27th September-30 September.**

National Conference, Meeting & Symposium Proceedings/ Abstract: -

1) Gene expression profile in esophageal cancer patients of Northeast region of India. **Chatterjee I,** Kapur S, Mahanta J, Phukan RK, Barooah MN, Kataki AC, Purkayastha J, Saxena S**. J Cancer Res Ther, 2 (1), S 20, 2006.**

2) Comparison of HPRT Mutant Frequency in Human Peripheral Blood Lymphocytes of Smokers and Non-Smokers. P.R.Vivek Kumar, Mary N. Mohankumar, **Indranil Chatterjee,** R.K.Jeevanram. **Radiation Protection and Environment, Vol 26, No 1-2, 389- 396, 2003.**

3) Interpreting Low-Dose Exposures by the ODDS -RATIO Method – A Case Report. Mary N. Mohankumar**, Indranil Chatterjee,** P.R. Vivek Kumar, and R.K. Jeevanram**. Radiation Protection and Environment, Vol 26, No 1-2, 290-292; 2003.**

4) Assessment of Breast Cancer Risk: Contribution of Genetic Polymorphisms in Estrogen-synthesizing and Metabolizing Genes. Kaushal Mishi, Chakraborty A, Bagadi S.A Raju, Ishan R, Regina T, Yadav D.S. **Chatterjee I**, Zomawia E, Kataki AC, Sharma J, Verma Y, Mishra AK, Kapur Sujala, Saxena Sunita. **27th Annual Convention of Indian Association for Cancer Research Networking Research to Applications &**

International Symposium on Frontiers in Functional Genomics. Gujarat Cancer & Research Institute, Ahmedabad, February 7-9, 2008.

Brief Biography of the Supervisor

Dr. Sujala Kapur is working as a Deputy Director at Institute of Pathology (ICMR), Safdarjung Hospital Campus, New Delhi. She did her MBBS in the year of 1973 from Lardy Hardinge Medical College, New Delhi. She did her MD in Pathology & Microbiology in the year of 1985 from S.N. Medical College, Agra. She joined as a research officer at Institute of Pathology in the year of 1988. Her research area of interest is molecular profiling of gastrointestinal cancer and Flow cytometric assays to analyse chemotherapeutic sensitivity in leukaemic cells. She received Scientific Fellowship from Kiel University, Germany in the year of 1995 & 1998. She has 30 publications in reputed national and international journals. She has 36 proceedings/ abstracts in national and international conferences.

Brief Biography of the Candidate

I, Mr.Indranil Chattopadhyay joined as Senior Research Fellow at Institute of Pathology (ICMR), Safdarjung Hospital Campus, New Delhi on 5th August, 2004. Currently, I am an Off-campus Ph.D. student of BITS, Pilani. The focus of my research work is to study the global gene expression profile in esophageal squamous cell carcinoma (ESCC) from high-risk area of India by cDNA microarray. I obtained Master degree in Zoology with first class from VISVA-BHARATI University in the year of 1997. I worked in the field of reproductive immunology at IICB, Kolkata for a period of one year (April, 1999-June, 2000) and orphan nuclear receptor at Chonnam National

University, South Korea for a period of more than one year (September , 2000-December, 2001). I got UICC International Cancer Technology Transfer Fellowship, 2007 at Peter MacCallum Cancer Institute, Melbourne, Australia. During my research carrier, I gained experience in all sorts of molecular biology techniques related to human genomics. I qualified NET for Lectureship on 25th June, 2000. I have 4 publications in International Journals and 10 proceedings / abstracts in national and international conferences.

PO Box 2345, Beijing 100023, China
www.wjgnet.com
wjg@wjgnet.com

World J Gastroenterol 2007 March 7; 13(9): 1438-1444
World Journal of Gastroenterology ISSN 1007-9327

RAPID COMMUNICATION

Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis

Indranil Chattopadhyay, Sujala Kapur, Joydeep Purkayastha, Rupkumar Phukan, Amal Kataki, Jagadish Mahanta, Sunita Saxena

Indranil Chattopadhyay, Sujala Kapur, Sunita Saxena, Institute of Pathology, Indian Council of Medical Research, Safdarjang Hospital Campus, Post Box No. 4909, New Delhi 110029, India
Joydeep Purkayastha, Amal Kataki, Dr.B.Borooah Cancer Institute, Guwahati-781016, Assam, India
Rupkumar Phukan, Jagadish Mahanta, Regional Medical Research Center, Dibrugarh-786001, Assam, India
Supported by Non Communicable Disease Division, Indian Council of Medical Research
Correspondence to: Dr. Sunita Saxena, Director, Institute of Pathology, Indian Council Of Medical Research, Safdarjang Hospital Campus, Post Box, No. 4909, New Delhi 110029, India. sunita_saxena@yahoo.com
Telephone: +91-11-26165797 Fax: +91-11-26198401
Received: 2006-11-17 Accepted: 2007-01-26

Abstract

AIM: To identify alterations in genes and molecular functional pathways in esophageal cancer in a high incidence region of India where there is a widespread use of tobacco and betel quid with fermented areca nuts.

METHODS: Total RNA was isolated from tumor and matched normal tissue of 16 patients with esophageal squamous cell carcinoma. Pooled tumor tissue RNA was labeled with Cy3-dUTP and pooled normal tissue RNA was labeled with Cy5-dUTP by direct labeling method. The labeled probes were hybridized with human 10K cDNA chip and expression profiles were analyzed by Genespring GX V 7.3 (Silicon Genetics) .

RESULTS: Nine hundred twenty three genes were differentially expressed. Of these, 611 genes were upregulated and 312 genes were downregulated. Using stringent criteria ($P \leqslant 0.05$ and $\geqslant 1.5$ fold change), 127 differentially expressed genes (87 upregulated and 40 downregulated) were identified in tumor tissue. On the basis of Gene Ontology, four different molecular functional pathways (MAPK pathway, G-protein coupled receptor family, ion transport activity, and serine or threonine kinase activity) were most significantly upregulated and six different molecular functional pathways (structural constituent of ribosome, endopeptidase inhibitor activity, structural constituent of cytoskeleton, antioxidant activity, acyl group transferase activity, eukaryotic translation elongation factor activity) were most significantly downregulated.

CONCLUSION: Several genes that showed alterations in our study have also been reported from a high incidence area of esophageal cancer in China. This indicates that molecular profiles of esophageal cancer in these two different geographic locations are highly consistent.

Key words: Esophageal cancer; Gene expression profile; Tobacco consumption; Betel quid chewing; North East India

Chattopadhyay I, Kapur S, Purkayastha J, Phukan R, Kataki A, Mahanta J, Saxena S. Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis. *World J Gastroenterol* 2007; 13(9): 1438-1444

http://www.wjgnet.com/1007-9327/13/1438.asp

INTRODUCTION

Environmental carcinogens have repeatedly been shown to affect the genetic material of host cells, leading to uncontrolled growth and ultimately malignant tumors[1]. The development of esophageal cancer is a leading example in which environmental carcinogens in addition to geographic and genetic factors appear to play major etiologic roles[2]. Esophageal cancer occurs at very high frequencies in certain parts of China, Iran, South Africa, Uruguay, France, Italy and in some regions of India[1]. The highest incidence of this cancer in India has been reported from Assam in the North-east region where it is the second leading cancer in men and third leading cancer in women[3].

Tobacco smoking, betel quid chewing, and alcohol consumption are the major known risk factors for esophageal cancer[4]. Tobacco smoke contains over 60 established carcinogens including polycyclic aromatic hydrocarbons and nitrosamines. Tobacco specific nitrosamines such as 4- (methyl nitrosamino)-1-butanone (NNK) and N'-nitrosonornicotine (NNN) that are carcinogens in smokeless tobacco, have been shown to enhance the risk of cancer development by forming adducts with DNA[5]. Betel quid chewing, a common habit

in Southeast Asia, has been found to increase the risk of developing esophageal cancer by 4.7-13.3 fold, although other exogenous risk factors may also be involved[4]. Betel quid usually comprises a piece of areca nut, which contains many polyphenols and several alkaloids, *Piper betle*, and lime with or without *Piper betle* leaves[6]. Arecoline, a major component of areca nut can produce 3-methyl nitrosamine propionitrile (MNPN), a potent carcinogen and safrole-like DNA adducts that have been shown to be genotoxic and mutagenic. Furthermore, contamination of areca nuts by fungi has been reported to produce carcinogenic aflatoxins. This assumes importance since using fermented areca nut with any form of tobacco is a common habit of people in Assam and has been reported to be a potential risk factor of esophageal cancer in this region[3].

The molecular mechanisms that may lead to the development of esophageal cancer in betel quid chewers and tobacco users are unknown. Recent studies are focusing on mechanisms that can explain the carcinogenic effects of tobacco and areca nut on epithelial cell lines. Incubation of areca nut extract or arecoline with primary oral keratinocytes has been reported to promote cell survival and an inflammatory response by induction of prostaglandin E_2, interleukin-6 (IL-6) and cyclooxygenase-2 (COX-2) production *via* activation of MEK1/ERK/c-Fos pathway[6]. Genotoxic stress as well as tissue inflammation and release of inflammatory mediators have been suggested to be key factors in carcinogenesis of gastrointestinal system. Genotoxic chemicals may induce the release of inflammatory mediators *via* mitogen activated protein kinase (MAPK) activation. Phosphorylated ERK1/2, JNK, p38 and ERK5 are reported to be significantly increased by exposure to tobacco smoke, indicating the activation of MAPK pathways[7]. NNK has recently been identified as a ligand of neuronal nicotinic acetylcholine receptors, which belong to G-protein-coupled receptors (GPCRs). GPCR induces proliferation through activation of members of the family of MAPKs[8,9].

The gene expression profile of esophageal cancer in a high incidence region of Assam where tobacco use and alcohol consumption are widespread and the users of these two substances are also betel quid chewers, has so far not been investigated. In the current study, cDNA microarray gene expression analysis was done to identify the genes differentially expressed in esophageal cancer associated with prevalent risk factors such as tobacco use and betel quid chewing in a high-risk Indian population.

MATERIALS AND METHODS

Collection of tumor samples

Endoscopic tissue biopsy specimens were taken from 16 patients at Dr. Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam. Routine histopathologic analysis was done to confirm the diagnosis. Tumor tissue and matched normal tissue distant to the tumor were collected during endoscopy in RNA later (Ambion, Austin, USA), snap-frozen in liquid nitrogen and stored at -70℃ until processed. Informed consent was obtained from all patients. Data of clinicopathologic parameters were obtained from patients' clinical records, operative notes and pathologic reports. Institutional Human Ethics Committee approved the study.

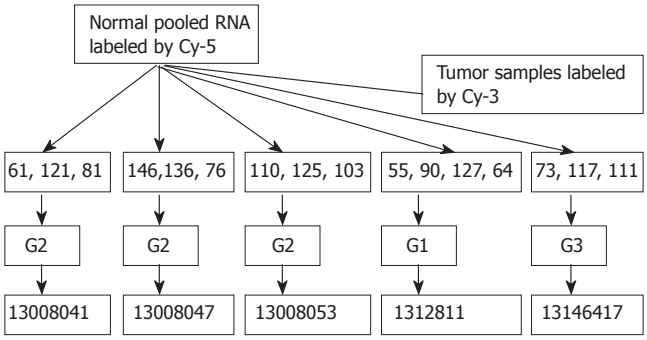


Figure1 Experimental design: G1, G2 and G3 indicate well-differentiated, moderately and poorly differentiated squamous cell carcinoma respectively. 13008041-53, 1312811 and 13146417 indicated barcode of microarray chips.

Sample preparation and chip hybridization

Total RNA isolation: Tissues were ground into powder in -196℃ liquid nitrogen and homogenized using Trizol reagent (Invitrogen Life Technologies, CA) for extraction of total RNA following the instruction of the manufacturer. The integrity of total RNA was checked by 1.2% formaldehyde agarose gel electrophoresis (visual presence of 28S and 18S bands). Total RNA with $OD_{260}/OD_{280} > 1.8$ was used for microarray experiments.

Experimental design: Total RNA was isolated from normal tissue of the esophagus from all the patients involved in this study and combined to make one common control. We pooled total RNA from biopsy samples of three to four patients in all five experiments on the basis of matching the histological grade to get sufficient amounts of total RNA for direct labeling. Pooled tumor RNAs were labeled with Cy-3 and hybridized against the Cy-5 labeled pooled samples of normal esophageal RNA, which generated a constant control to be used on chips analyzed. Figure 1 shows the study design. Sample pooling was done to rapidly identify tumor markers that were expressed by the majority of tumors in a population. Pooling of RNA samples isolated from tissue is a strategy that can be implemented in microarray experiments when the amount of sample RNA is limited or when variations across the samples must be reduced. The reason behind this approach is that the concentration of an mRNA molecule in a pooled sample is likely to be closer to the average concentration for the class than the concentration in a sample from a single individual. Pooled samples have been shown to accurately reflect gene expression in individual samples and yield reproducible data[10,11].

Labeling and hybridization: Twenty μg of total RNA from the tumor and matched normal tissue were labeled with cyanine 3-dUTP and cyanine 5-dUTP by direct labeling method (Perkin Elmer Life Sciences, USA: Micromax Direct labeling kit). The labeled probes were denatured at 95℃ for 5 min and hybridized with a human 10K cDNA chip (University Health Network, Microarray Centre, Toronto, Canada), which contains 9914 well-

characterized human clones, in a hybridization chamber (Corning Life Sciences, USA) at 65℃ water bath for 18 h. Before hybridization, slides were pre-hybridized in 5XSSC, 0.1% SDS and 1% BSA solution at 65℃ for 45 min to prevent nonspecific hybridization. After hybridization, the slides were washed in 2XSSC with 0.1% SDS, 0.1X SSC with 0.05% SDS and 0.1XSSC sequentially for 20 min each and then spin-dried.

Microarray image analysis: Hybridized arrays were scanned at 5 μm resolution on a Gene Pix 4200A scanner (Axon Instruments Inc. Foster City, CA) at various PMT voltage settings to obtain maximal signal intensities with < 0.1% probe saturation. The Cy5-labelled cDNAs were scanned at 635 nm and the Cy3-labeled cDNA samples were scanned at 532 nm. The resulting TIFF images were analyzed by Gene Pix Pro 6.0.1.27 software (Axon Instrument). Both digital images were overlaid to form a pseudo colored image and a detection method was then used to determine the actual target region based on the information from both red (Cy5) and green (Cy3) pixel values. The ratios of the sample intensity to the reference intensity (green: red) for all of the targets were determined and ratio normalization was performed to normalize the center of the ratio distribution to 1.0. Image processing analysis was used for estimation of spot quality by assigning a quality score to each ratio measurement[12].

Data analysis

Quality assurance: The data sets were imported into Microsoft Excel spreadsheets. Four parameters were used to assess quality of spots with the following features excluded: diameter < 50 μm; > 50% saturated pixels in both channels; < 54% of the pixels with an intensity greater than the median background intensity plus one standard deviation in either channel; flagged by Gene Pix as "not found" or "absent" or manually flagged as "bad" due to high background, misshapen features, scratches or debris on the slide undetected by Gene Pix[13].

Normalization: The results were normalized for labeling and detection efficiencies of the two fluorescence dyes, prior to determining differential gene expression between tumor and normal tissue samples. Intensities of selected spots were transformed into $\log_2$ (Cy3/Cy5) and data were normalized by locally weighted linear regression (LOWESS) method. Genespring software version GX7.3.1 (Silicon Genetics, Redwood, CA) was used to normalize values for each gene for data analysis.

Ranking of genes: Data analysis was performed using Genespring Software GXV 7.3 (Silicon Genetics, Redwood City, CA). Differentially regulated genes were ranked on the basis of signal intensity, normalized ratio, flag value and variance across replicate experiments. Top ranked genes had a higher intensity, high-normalized ratio for up and low for down, unflagged, very low variance or standard deviation. Filtered genes identified to be differentially expressed by 1.5 fold or greater in three of five chips were analyzed for functional gene clusters using GeneSpring software GXV 7.3. GeneSpring used data found publicly in genomic databases to build gene ontology based on annotation information. *t*-test was performed at the 0.05% significant level to find genes that vary significantly across samples. *P*-Value or probability value is the chance of set of genes involved in a particular function to be present in any given gene list with reference to the number of genes known to be involved in the function.

Hierarchical clustering: Average linkage hierarchical clustering was done using the Cluster Software version 3.0 written by Michael Eisen[14]. The Euclidean distance metric was used as a measure of similarity between the gene expression patterns for each pair of samples based on log-transformed ratios across all genes. The results were analyzed and visualized with the Tree View Program Version 1.50 also written by Michael Eisen. Those genes showing progressive fold increases or decreases in gene expression relative to normal mucosa were shown proportionally in red and green, respectively.

Pathway prediction analysis: We obtained annotations of the bioprocesses, molecular function and cellular localization using the freely available Gene Ontology and Source database[15]. The significant gene clusters were queried with known components of the biological pathways on the freely available KEGG database[16]. We also used the Biointerpreter software (http://www.genotypic.co.in/biointerpreter) for gene ontology.

Table 1 Demographic and clinical characteristics of esophageal squamous cell carcinoma cases

Patient ID	Age	Sex	Tobacco/ smoking habit	Alcohol use	Betel nut chewing	Pathological grade
EC-61	55	M	Yes	Yes	Yes	G2
EC-64	45	M	Yes	Yes	Yes	G1
EC-55	52	M	Yes	Yes	Yes	G1
EC-103	59	M	Yes	Yes	Yes	G2
EC-111	55	M	Yes	Yes	Yes	G3
EC-90	50	M	Yes	Yes	Yes	G1
EC-117	60	M	Yes	No	Yes	G3
EC-73	50	M	Yes	No	Yes	G3
EC-125	71	M	Yes	No	Yes	G2
EC-110	70	M	Yes	Yes	Yes	G2
EC-146	45	M	Yes	No	Yes	G2
EC-136	48	M	Yes	Yes	Yes	G2
EC-127	58	M	Yes	Yes	Yes	G1
EC-121	55	M	Yes	Yes	Yes	G2
EC-76	50	M	Yes	Yes	Yes	G2
EC-81	55	M	Yes	Yes	Yes	G1

G1 = Well differentiated squamous cell carcinoma; G2= Moderately differentiated squamous cell carcinoma; G3= Poorly differentiated squamous cell carcinoma.

RESULTS

Sixteen esophageal biopsy samples were compared with normal pooled esophageal tissue. All patients were male and gave a history of tobacco consumption and betel nut chewing (Table 1). Gene expression was measured using microarrays to detect changes in tumor samples compared to normal. Nine hundred and twenty three genes were differentially expressed at least 1 fold in 3 out of 5 experiments. Of these, 611 genes were upregulated and 312 genes were down regulated. The scaled data generated from Gene Pix Pro 6.0.1.27 software were imported into GeneSpring

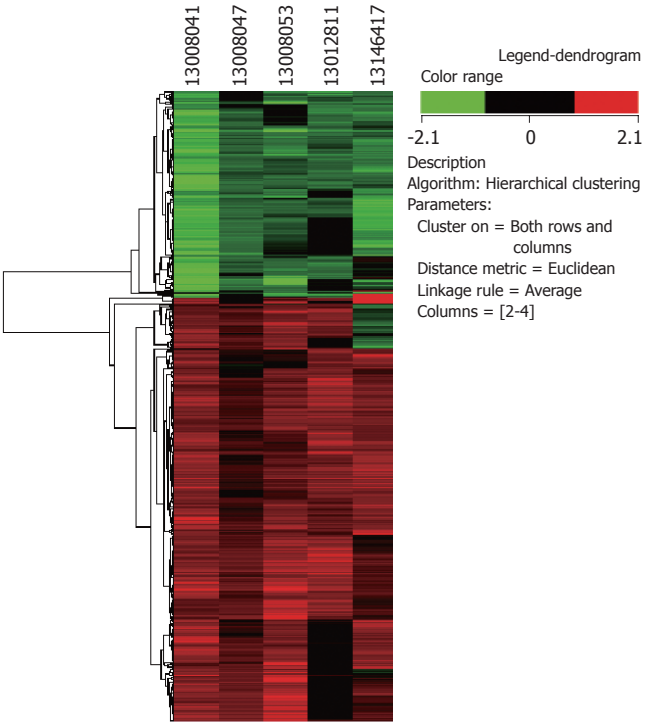


Figure 2 Hierarchical Clustering (Average Linkage Clustering) of the genes that were over or under expressed in tumor versus normal tissue in the 16 ESCC patients.

for fold change analysis, filtering, and cluster analysis. Hierarchical Clustering analysis of 923 genes selected from the 10 000-gene sets was performed (Figure 2). To identify highly reproducible changes, data were filtered based on select criteria. Transcripts modulated by a minimum 1.5 fold in at least three of five chips were used for further analyses. However, genes that had a $\geqslant$ 1.5 fold cutoff value or had a *P* value of $\leqslant$ 0.05 were also included for analysis to seek for subtle changes in gene expression. Using stringent criteria ($P \leqslant 0.05$ and $\geqslant$ 1.5 fold change), 127 differentially expressed genes (87 upregulated and 40 down regulated) were identified in tumor tissue. Using the Gene Ontology database, we categorized the 923 differentially expressed genes into known or probable functional categories. Genes involved in dimethylallyltransferase activity and farnesyltransferase activity (*CTLA4*), cation antiporter activity (*SLC9A2*) and cation transporter activity (*KCNN2*, *SLC30A4, KCNJ15, CACNA2D3*), G-protein coupled receptor (GPCR) activity (*GPR87, NPY*), MAPK signaling pathway (*FGF12*), and protein serine or threonine kinase activity (*GRK4*) were significantly upregulated (Table 2). Out of 87 upregulated genes, two genes were involved in GPCR group, five genes were involved in cation transporter activity, one gene each was involved in protein serine or threonine kinase activity and MAPK activity. Genes involved in anti-apoptosis activity (*BIRC1*), omega peptidase activity (*UCHL1*) and cellular proliferation (*EGR2*) were also significantly activated. Genes involved in structural constituent of the ribosome (*RPL32*, *RPS4X),* structural constituent of cytoskeleton (*KRT17*, *KRT8*, *PLA2G1B*), cysteine protease inhibitor activity (*CSTB*, *CSTA*), anti-oxidant activity (*PRDX6*), acyl groups transferase activity (*TGM3*), and

Table 2 List of significantly upregulated genes in esophageal cancer patients

Category	Genes in category	% Genes in category	Genes in list in category	% Genes in list category	*P*-value
GO: 4930: G-protein coupled receptor activity	150	2.148	23	3.433	0.0156
GO: 8324: cation transporter activity	307	4.396	39	5.821	0.0404
GO: 4674: protein serine/threonine kinase activity	285	4.081	38	5.672	0.022
GO: 4707: MAP kinase activity	20	0.286	5	0.746	0.0368
GO: 4161: dimethylallyltransferase activity	4	0.0573	3	0.448	0.00327

Table 3 List of significantly down-regulated genes in esophageal cancer patients

Category	Genes in category	% Genes in category	Genes in list in category	% Genes in list category	*P*-value
GO: 3735: structural constituent of ribosome	115	1.647	32	4.992	5.54E-09
GO: 4869: cysteine protease inhibitor activity	27	0.387	11	1.716	1.19E-05
GO: 5200: structural constituent of cytoskeleton	63	0.902	12	1.872	0.011
GO: 16746: transferase activity, transferring acyl groups	90	1.289	19	2.964	0.000422
GO: 16209: antioxidant activity	25	0.358	9	1.404	0.000232
GO: 3746: translation elongation factor activity	14	0.2	5	0.78	0.00636

translation elongation (*EEF1A1*) were significantly down-regulated (Table 3). Genes involved in humoral immune response (*CD24*) and base-excision repair (*MPG*) were significantly down regulated. Out of 40 down regulated genes, five genes were involved in structural constituent of the ribosome, four genes were involved in structural constituent of cytoskeleton, two genes were involved in cysteine protease inhibitor activity and one gene each was involved in anti-oxidant activity, acyl group transferase activity and translation elongation.

DISCUSSION

Several tobacco constituents, including nitrosamines, polycyclic aromatic hydrocarbons, aromatic amines, various aldehydes and phenols, may be causally related to esophageal cancer[1]. A previous report has shown that betel quid chewing with or without tobacco consumption

is associated with the development of esophageal cancer in Assam[3]. However, there are very few studies of gene expression profiles of tumors that may be associated with betel quid chewing as well as tobacco consumption. The current study was aimed to analyze genes and pathways that may be involved in tobacco and betel quid chewing related esophageal malignancies in this high incidence region of India.

Gene expression levels showed that there were four different molecular functional pathways that were most significantly upregulated and six different molecular functional pathways that were most significantly down regulated. Some of the significantly overexpressed molecular functional pathways like MAPK signaling pathway, G-protein coupled receptor and cation transporter activity have earlier been reported in esophageal and other cancers. However, genes such as *CTLA4* (involved in dimethylallyltransferase activity and farnesyltransferase activity) and *UCHL1, NPY*, *FGF12, KCNN2,* and *KCNJ15* were found to be significantly upregulated in our study and have not been reported earlier. Genes involved in structural constituent of ribosome (*RPL32, RPS4X, RPL7A)* and anti-oxidant activity (*PRDX6*) that were found to be significantly down regulated in our study, have also not been reported earlier. Some of the significantly down regulated molecular functional pathways like structural constituent of cytoskeleton (*KRT17*, *KRT8, KRT4),* acyl groups transferase activity (*TGM3)* and cysteine protease inhibitor activity (*CSTB*, *CSTA*), have already been reported previously in esophageal carcinoma. The data may be used for selection of a limited number of markers that can be screened in large populations by RT-PCR. All the genes identified here are of interest because of their potential roles in the natural history of esophageal squamous cell carcinoma.

The mitogen activated signaling cascade showed significant up-regulation in our study. Activated MAPK pathway has been detected in many human tumors including carcinomas of the breast, colon, kidney, and lung suggesting the possibility that MAPK may play a major role in tumor progression and metastasis. MAPK induces proteolytic enzymes that degrade the basement membrane, enhance cell migration, initiate several pro-survival genes and maintain growth[17]. Oxidants in cigarette smoke have previously been reported to activate MAPK signaling cascades in lung epithelial cells *in vitro* and *in vivo*. These signaling pathways lead to the enhanced ability of Jun and Fos family members to activate transcription of a number of AP-1 dependent target genes involved in cell proliferation, differentiation, and inflammation[18]. Phosphorylated ERK 1/2, JNK, p38 and ERK5 genes were significantly increased upon exposure to tobacco smoke, indicating the activation of MAPK pathways[7]. Benzo (a) pyrene quinines, which is the non-volatile component of cigarette smoke, have also been reported to increase cell proliferation, generate reactive oxygen species, and transactivate the epidermal growth factor receptor in breast epithelial cells[19].

TNF- alpha was found to be upregulated in our study. Cigarette smoke is known to enhance the induction of TNF-alpha by differentiated macrophage that is regulated primarily *via* the ERK ½ pathway[20]. TNF receptor super family member 17 has been shown to specifically bind to the TNF (ligand) super family member 13b (TNFSF13B/ TALL-1/BAFF) , which leads to activation of NF-kappaB and MAPK8/JNK.

Neuropeptide Y (NPY), which belongs to G-protein coupled receptor family, was found to be significantly upregulated in our study. Aberrant NPY expressions are early events in prostrate cancer development and are associated with a poor prognosis[21]. Activation of the Y1 receptor by NPY regulates the growth of prostate cancer cells. Estrogen upregulates NPY receptor expression in a human breast cancer cell line[22].

Ion channel regulating genes showed significant upregulation in our study. Several tumors such as prostrate, uterus, glial cells, stomach, pancreas, breast and colorectum are known to express Ca^{2+}activated K^+ channels. The complex process of carcinogenesis is triggered by oncogenic pathways *via* activation of K^+ channels. For example, p21 ras and the Raf kinase are known to induce oncogenic transformation *via* activation of Ca^{2+} dependent K^+ channels. Enhanced cell migration and tumor metastasis are associated with fluctuations in the activity of membrane transporters and ion channels that require K^+ channel activity[23]. Genes involved in calcium regulation and calcium signaling also seemed to be important in the progression of esophageal squamous dysplasia[24].

The deleted in colorectal cancer (DCC) gene, which was found to be upregulated in our study, is a candidate tumor suppressor gene which may be associated with differentiation and proliferation of normal cells. DCC protein expression seems to be a significant prognostic factor in high-risk resected gastric cancer. This gene may play a role in the metastatic potential of these tumors[15].

Transglutaminase-3 (TGase-3) that showed significant down regulation in our study has been reported in earlier studies. TGase-3 has been implicated in the formation and assembly of the cornified cell envelope of the epidermis, hair follicle and perhaps other stratified squamous epithelia. Alterated TGase-3 expression is a common event in human esophageal cancer[25]. The lowest levels of TGases 3 and 7 have been reported in patients with metastatic disease[26]. Tissue transglutaminase (tTG) is a high level phenotypic biomarker down regulated in prostate cancer[27].

Intermediate filaments form the cytoskeleton of cells and maintain the integrity of cells. Keratins (*KRT4* and *KRT8*) showed significant down regulation in our study. Overexpression of epithelial cell intermediate filaments and their isoforms (*KRT8)* have been reported earlier in colorectal polyp and cancer[28].

Reduced expression of cystatin B as found in our study, has earlier been reported to be associated with lymph node metastasis and may therefore prove to be a useful marker for predicting the biologic aggressiveness of esophageal cancer[29]. Overexpression of cystatin A has been shown to inhibit tumor cell growth, angiogenesis, invasion, and metastasis in esophageal cancer[30]. High levels of cystatin A and cystatin B have been reported to correlate with more favorable patient outcome in breast, lung and head and neck tumors.

A group of ribosomal proteins may function as cell cycle checkpoints and compose a new family of cell proliferation regulators. They play an important role in translational regulation and control of cellular transformation, tumor growth, aggressiveness and metastasis. Thus in addition to protein synthesis, they are involved in neoplastic transformation of cells. Ribosomal proteins were found to be significantly down regulated in our study. Expression of human ribosomal protein L13a has been shown to induce apoptosis, presumably by arresting cell growth in the G2/M phase of the cell cycle. In addition, a closely related ribosomal protein, L7, arrests cells in G1 and also induces apoptosis[15]. Mitochondrial ribosomal protein L41 suppresses cell growth in association with p53 and p27Kip1. *MRPL41* is reported to be either expressed at reduced levels or absent in most tumor types and cell lines[31]. Human apurinic apyrimidinic endonuclease (RPLP0) and its N terminal truncated form (AN34) are involved in DNA fragmentation during apoptosis. Down regulation of RPLP0 expression is associated with the induction of apoptosis in differentiating myeloid leukemia cells[32]. Simultaneously, three ribosomal proteins (RPL10, RPL32, and RPS16) also showed up-regulation in C81 cells. Overexpression of several ribosomal proteins including RPS16 has been reported in colon, breast, liver, and pancreatic tumors[33].

This is the first study to provide gene expression profiles of esophageal cancer in a high-risk region of Assam in India. The most salient finding was identification of down regulated genes involved in structural constituents of ribosome and upregulated genes involved in cation transporter activity. In a similar study of gene expression profiles in a high-risk area of China, Taylor *et al* and Liu *et al* have reported down regulation of *CSTB*, *CSTA*, *KRT4*, and *TGM3* and upregulation of G-coupled signaling, ion transport activity and MAPK activity[12,24,35,36]. In a recent study from a low risk area of India, deregulation of genes associated with zinc homeostasis in esophageal squamous cell carcinoma (ESCC) has been reported[34]. These data indicate the consistency of molecular profiles of esophageal cancer in two different geographic locations that have a high incidence of ESCC but different food habits and customs.

ACKNOWLEDGMENTS

The authors are thankful to Madavan Vasudevan, Dr. Sudha Rao and Dr. Raja Maugasimangalam of Genotypic Technology (P) Ltd. Bangalore, India for their valuable support in Microarray Data Analysis. The authors are thankful to Indian Council of Medical Research for their financial support. The authors are thankful to Jagannath Sharma of BBCI, Guwahati for histopathological analysis.

REFERENCES

1 **Stoner GD**, Gupta A. Etiology and chemoprevention of esophageal squamous cell carcinoma. *Carcinogenesis* 2001; **22**: 1737-1746
2 **Murtaza I**, Mushtaq D, Margoob MA, Dutt A, Wani NA, Ahmad I, Bhat ML. A study on p53 gene alterations in esophageal squamous cell carcinoma and their correlation to common dietary risk factors among population of the Kashmir valley. *World J Gastroenterol* 2006; **12**: 4033-4037
3 **Phukan RK**, Ali MS, Chetia CK, Mahanta J. Betel nut and tobacco chewing; potential risk factors of cancer of oesophagus in Assam, India. *Br J Cancer* 2001; **85**: 661-667
4 **Lee CH**, Lee JM, Wu DC, Hsu HK, Kao EL, Huang HL, Wang TN, Huang MC, Wu MT. Independent and combined effects of alcohol intake, tobacco smoking and betel quid chewing on the risk of esophageal cancer in Taiwan. *Int J Cancer* 2005; **113**: 475-482
5 **Rohatgi N**, Matta A, Kaur J, Srivastava A, Ralhan R. Novel molecular targets of smokeless tobacco (khaini) in cell culture from oral hyperplasia. *Toxicology* 2006; **224**: 1-13
6 **Chang MC**, Wu HL, Lee JJ, Lee PH, Chang HH, Hahn LJ, Lin BR, Chen YJ, Jeng JH. The induction of prostaglandin E2 production, interleukin-6 production, cell cycle arrest, and cytotoxicity in primary oral keratinocytes and KB cancer cells by areca nut ingredients is differentially regulated by MEK/ERK activation. *J Biol Chem* 2004; **279**: 50676-50683
7 **Zhong CY**, Zhou YM, Douglas GC, Witschi H, Pinkerton KE. MAPK/AP-1 signal pathway in tobacco smoke-induced cell proliferation and squamous metaplasia in the lungs of rats. *Carcinogenesis* 2005; **26**: 2187-2195
8 **Arredondo J**, Chernyavsky AI, Jolkovsky DL, Pinkerton KE, Grando SA. Receptor-mediated tobacco toxicity: cooperation of the Ras/Raf-1/MEK1/ERK and JAK-2/STAT-3 pathways downstream of alpha7 nicotinic receptor in oral keratinocytes. *FASEB J* 2006; **20**: 2093-2101
9 **Sud N**, Sharma R, Ray R, Chattopadhyay TK, Ralhan R. Differential expression of G-protein coupled receptor 56 in human esophageal squamous cell carcinoma. *Cancer Lett* 2006; **233**: 265-270
10 **Helm J**, Enkemann SA, Coppola D, Barthel JS, Kelley ST, Yeatman TJ. Dedifferentiation precedes invasion in the progression from Barrett's metaplasia to esophageal adenocarcinoma. *Clin Cancer Res* 2005; **11**: 2478-2485
11 **Agrawal D**, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J Natl Cancer Inst* 2002; **94**: 513-521
12 **Su H**, Hu N, Shih J, Hu Y, Wang QH, Chuang EY, Roth MJ, Wang C, Goldstein AM, Ding T, Dawsey SM, Giffen C, Emmert-Buck MR, Taylor PR. Gene expression analysis of esophageal squamous cell carcinoma reveals consistent molecular profiles related to a family history of upper gastrointestinal cancer. *Cancer Res* 2003; **63**: 3872-3876
13 **Samimi G**, Manorek G, Castel R, Breaux JK, Cheng TC, Berry CC, Los G, Howell SB. cDNA microarray-based identification of genes and pathways associated with oxaliplatin resistance. *Cancer Chemother Pharmacol* 2005; **55**: 1-11
14 **Eisen MB**, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95**: 14863-14868
15 **Diehn M**, Sherlock G, Binkley G, Jin H, Matese JC, Hernandez-Boussard T, Rees CA, Cherry JM, Botstein D, Brown PO, Alizadeh AA. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res* 2003; **31**: 219-223
16 **Altermann E**, Klaenhammer TR. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* 2005; **6**: 60
17 **Reddy KB**, Nabha SM, Atanaskova N. Role of MAP kinase in tumor progression and invasion. *Cancer Metastasis Rev* 2003; **22**: 395-403
18 **Mossman BT**, Lounsbury KM, Reddy SP. Oxidants and signaling by mitogen-activated protein kinases in lung epithelium. *Am J Respir Cell Mol Biol* 2006; **34**: 666-669
19 **Burdick AD**, Davis JW 2nd, Liu KJ, Hudson LG, Shi H, Monske ML, Burchiel SW. Benzo(a)pyrene quinones increase cell proliferation, generate reactive oxygen species, and transactivate the epidermal growth factor receptor in breast epithelial cells. *Cancer Res* 2003; **63**: 7825-7833
20 **Demirjian L**, Abboud RT, Li H, Duronio V. Acute effect

of cigarette smoke on TNF-alpha release by macrophages mediated through the erk1/2 pathway. *Biochim Biophys Acta* 2006; **1762**: 592-597

21 **Rasiah KK**, Kench JG, Gardiner-Garden M, Biankin AV, Golovsky D, Brenner PC, Kooner R, O'neill GF, Turner JJ, Delprado W, Lee CS, Brown DA, Breit SN, Grygiel JJ, Horvath LG, Stricker PD, Sutherland RL, Henshall SM. Aberrant neuropeptide Y and macrophage inhibitory cytokine-1 expression are early events in prostate cancer development and are associated with poor prognosis. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 711-716

22 **Amlal H**, Faroqui S, Balasubramaniam A, Sheriff S. Estrogen up-regulates neuropeptide Y Y1 receptor expression in a human breast cancer cell line. *Cancer Res* 2006; **66**: 3706-3714

23 **Kunzelmann K**. Ion channels and cancer. *J Membr Biol* 2005; **205**: 159-173

24 **Joshi N**, Johnson LL, Wei WQ, Abnet CC, Dong ZW, Taylor PR, Limburg PJ, Dawsey SM, Hawk ET, Qiao YL, Kirsch IR. Gene expression differences in normal esophageal mucosa associated with regression and progression of mild and moderate squamous dysplasia in a high-risk Chinese population. *Cancer Res* 2006; **66**: 6851-6860

25 **Chen BS**, Wang MR, Xu X, Cai Y, Xu ZX, Han YL, Wu M. Transglutaminase-3, an esophageal cancer-related gene. *Int J Cancer* 2000; **88**: 862-865

26 **Jiang WG**, Ablin R, Douglas-Jones A, Mansel RE. Expression of transglutaminases in human breast cancer and their possible clinical significance. *Oncol Rep* 2003; **10**: 2039-2044

27 **Lewis TE**, Milam TD, Klingler DW, Rao PS, Jaggi M, Smith DJ, Hemstreet GP, Balaji KC. Tissue transglutaminase interacts with protein kinase A anchor protein 13 in prostate cancer. *Urol Oncol* 2005; **23**: 407-412

28 **Polley AC**, Mulholland F, Pin C, Williams EA, Bradburn DM, Mills SJ, Mathers JC, Johnson IT. Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia. *Cancer Res* 2006; **66**: 6553-6562

29 **Shiraishi T**, Mori M, Tanaka S, Sugimachi K, Akiyoshi T. Identification of cystatin B in human esophageal carcinoma, using differential displays in which the gene expression is related to lymph-node metastasis. *Int J Cancer* 1998; **79**: 175-178

30 **Li W**, Ding F, Zhang L, Liu Z, Wu Y, Luo A, Wu M, Wang M, Zhan Q, Liu Z. Overexpression of stefin A in human esophageal squamous cell carcinoma cells inhibits tumor cell growth, angiogenesis, invasion, and metastasis. *Clin Cancer Res* 2005; **11**: 8753-8762

31 **Yoo YA**, Kim MJ, Park JK, Chung YM, Lee JH, Chi SG, Kim JS, Yoo YD. Mitochondrial ribosomal protein L41 suppresses cell growth in association with p53 and p27Kip1. *Mol Cell Biol* 2005; **25**: 6603-6616

32 **Yoshida A**, Urasaki Y, Waltham M, Bergman AC, Pourquier P, Rothwell DG, Inuzuka M, Weinstein JN, Ueda T, Appella E, Hickson ID, Pommier Y. Human apurinic/apyrimidinic endonuclease (Ape1) and its N-terminal truncated form (AN34) are involved in DNA fragmentation during apoptosis. *J Biol Chem* 2003; **278**: 37768-37776

33 **Zhu Y**, Lin H, Li Z, Wang M, Luo J. Modulation of expression of ribosomal protein L7a (rpL7a) by ethanol in human breast cancer cells. *Breast Cancer Res Treat* 2001; **69**: 29-38

34 **Kumar A**, Chatopadhyay T, Raziuddin M, Ralhan R. Discovery of deregulation of zinc homeostasis and its associated genes in esophageal squamous cell carcinoma using cDNA microarray. *Int J Cancer* 2007; **120**: 230-242

35 **Luo A**, Kong J, Hu G, Liew CC, Xiong M, Wang X, Ji J, Wang T, Zhi H, Wu M, Liu Z. Discovery of Ca2+-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. *Oncogene* 2004; **23**: 1291-1299

36 **Zhou J**, Zhao LQ, Xiong MM, Wang XQ, Yang GR, Qiu ZL, Wu M, Liu ZH. Gene expression profiles at different stages of human esophageal squamous cell carcinoma. *World J Gastroenterol* 2003; **9**: 9-15

S- Editor Liu Y **L- Editor** Zhu LH **E- Editor** Ma WH

Molecular profiling to identify molecular mechanism in esophageal cancer with familial clustering

INDRANIL CHATTOPADHYAY[1], RUPKUMAR PHUKAN[3], AVNINDER SINGH[1], MADAVAN VASUDEVAN[4], JOYDEEP PURKAYASTHA[2], STEPHEN HEWITT[5], AMAL KATAKI[2], JAGADISH MAHANTA[3], SUJALA KAPUR[1] and SUNITA SAXENA[1]

[1]Institute of Pathology, New Delhi-110029; [2]Dr B. Borooah Cancer Institute, Guwahati-781016, Assam; [3]Regional Medical Research Center, Dibrugarh-786001, Assam; [4]Genotypic Technology [P] Ltd, 112, Bangalore-560001, India; [5]Laboratory of Pathology, National Cancer Institute, Bethesda, MD 20892, USA

Received November 5, 2008; Accepted February 6, 2009

DOI: 10.3892/or_00000333

Abstract. To identify the genes and molecular functional pathways involved in esophageal cancer, we analyzed the gene expression profile of esophageal tumor tissue from patients having family history of esophageal cancer by cDNA microarray. Three hundred and fifty differentially expressed genes (26 up-regulated and 324 down-regulated) were identified. Genes involved in humoral immune response (*PF4*), extracellular matrix organization (*COL4A4*), metabolism of xenobiotics (*EPHX1*), TGF-ß signaling (*SMAD1*) and calcium signaling pathways (*VDAC1*) were down-regulated and genes involved in regulation of actin cytoskeleton (*WASL*), neuroactive ligand receptor interaction (*GRM3*), Toll-like receptor (*CD14*), B-cell receptor (*IFITM1*) and insulin signaling pathways (*FOXO1A*) were up-regulated. Validation of differential expression of subset of genes by QRT-PCR and tissue microarray in familial and non-familial cases showed no significant difference in expression of these genes in two groups suggesting familial clustering occurs as result of sharing of common environmental factors. Gene expression profiling of clinical specimens from well characterized populations that have familial clustering of cancer identified molecular mechanism associated with progression of esophageal cancer.

Introduction

Esophageal cancer is among the ten most common malignancies worldwide and ranks as the sixth leading cause of death from cancer (1). Esophageal cancer occurs at very high frequencies in certain parts of China, Iran, South Africa, Uruguay, France and Italy (2). Association of family history with an increased risk of esophageal cancer has been reported in several case control and cohort studies from China, Iran and Japan suggesting possible role of environmental as well as genetic factors (3,4). A high prevalence of esophageal cancer with familial aggregation has also been reported from Assam in the Northeast (NE) region of India with an age-adjusted rate (AAR) of 33/100,000 males (5,6).

Familial clustering of cancer may be due to shared environmental factors or shared genes by family members (7). Su *et al* reported that molecular profiles in esophageal squamous cell carcinoma (ESCC) were highly consistent and expression patterns in familial cases were different from those in sporadic cases (3). We have earlier reported that gene expression profile of non-familial ESCC in Assam, a high-risk zone for esophageal cancer in India, are highly consistent with ESCC in China (6).

In the current study, molecular signature of ESCC from high-risk area of India has been studied by gene expression profiling in esophageal cancer patients with family history of esophageal cancer with the aim to elucidate molecular pathogenesis of esophageal cancer in these patients.

Materials and methods

Selection of patients and collection of samples. Among 317 cases of esophageal cancer registered at Dr Bhubaneshwar Borooah Cancer Institute (BBCI), Guwahati, Assam, 92 (29%) had family history of esophageal and other cancers besides habit of tobacco and betel quid chewing. Among 92 patients, 45 patients (49%) had family history of esophageal cancer. Patients diagnosed with metastases and at advanced stage of the disease were excluded from the study. Endoscopic biopsy specimens from tumor and matched normal tissue distant to tumor were collected during diagnostic endoscopy. Part of tumor and normal tissues was preserved in formalin for histopathological diagnosis/confirmation and remaining tissue was immediately immersed in RNA later solution (Ambion, Austin, TX) and stored at -70˚C until processed. Out of 45 patients, 20

Correspondence to: Dr Sunita Saxena, Institute of Pathology, Safdarjung Hospital Campus, New Delhi-110029, India
E-mail: sunita_saxena@yahoo.com

Key words: expression profile, familial esophageal cancer, microarray

patients with tumor biopsies containing >80% tumor were included for gene expression study. Demographic and lifestyle cancer risk factors and clinical data of all 20 patients were collected (Table I). Informed consent was obtained from all the patients to use their surgical specimens and clinicopathological data for the study. Institutional Human Ethics Committee approved this study.

Microarray experiments. Total RNA was isolated from snap-frozen biopsies using the Qiagen (Valencia, CA) RNeasy mini kit and its integrity was examined using the RNA 6000 Nano LabChip on the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). RNA quantity was determined by the NanoDrop® ND-1000 UV-Vis spectrophotometer (Nanodrop technologies, Rockland, MD). As a reference RNA, we used pool of total RNA from normal esophageal tissue of all patients. Out of 20 patients, RNA samples from nine tumor biopsies with RNA Integrity Number (RIN) of 8 were used for gene expression study by cDNA microarray. Low RNA input fluorescent linear amplification kit (Agilent, Santa Clara, CA) was used for labeling. Individual tumor cRNA was labeled with cyanine 5 and hybridized with cyanine 3-labeled pooled normal esophageal cRNA. The labeled and fragmented cRNAs were hybridized at 65˚C for 17 h with human 10K cDNA array (University Health Network, Microarray Centre, Toronto, Canada), which contained 9,914 well-characterized human clones.

Microarray image acquisition and data analysis. Hybridized arrays were scanned at 5 μm resolution on an Agilent DNA microarray scanner; model G2565AA at 100% laser power and 30% PMT at 635 nm for Cy5-labelled samples and at 532 nm for Cy3-labeled samples. The resulting TIFF images were analyzed by Agilent Feature Extraction Software 9.1.3, which performed spot localization (Find Spot Algorithm), outlier pixel rejection based on the interquartile range method (Cookie Cutter Algorithm) and flagging of saturated features.

Genespring software version GXV7.3.1 (Agilent Technologies) was used to normalize values for each gene and for further data analysis. Differentially regulated genes were ranked on the basis of signal intensity, normalized ratio, flag value and variance across replicate experiments. Genes were considered to be up-regulated when the median of the normalized ratio was ≥2. Genes were considered to be down-regulated when the median of the normalized ratio was ≤0.5. The observed number of differentially expressed genes in each GO category was compared to the corresponding number estimated from a random model (hypergeometric distribution); significance was assessed by a p-value. Hierarchical clustering analysis was also performed with Genespring software GXV 7.3.1. in which the average linkage and Pearson correlation (centered correlation) clustering algorithm was used. The microarray data set was submitted to the GEO repository (GSE 10127) at http://www.ncbi.nlm.nih.gov/geo. Annotations of the bioprocesses, molecular function and cellular localization were obtained using the freely available Gene Ontology [Source database (http://source.stanford.edu) and Biointerpreter software (http://www. genotypic. co.in/biointerpreter)].

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 PCR reactions were performed using an ABI Prism 7000 sequence detection system (Applied Biosystems). Primers and TaqMan probes of five target genes and an internal control gene *18S rRNA* were purchased as assays-on-demand from Applied Biosystems (Table II). 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 20 familial ESCC and 10 non-familial ESCC cases. Wilcoxon signed rank tests were used to determine the statistical significance of expression difference for each test gene in 20 familial ESCC cases (Table II). The expression profiles of target genes by QRT-PCR in two groups (familial and non-familial ESCC cases) were then compared using Mann-Whitney U test. A p-value ≤0.05 was considered significant.

Tissue microarray (TMA)-based immunohistochemical analysis. A TMA was constructed from 120 formalin-fixed, paraffin-embedded blocks of esophageal biopsy samples. These included 20 controls (non-neoplastic esophageal squamous epithelium) and 100-ESCCs, of which 20 biopsies were obtained from familial cases. The tissue cylinders were precisely arrayed into the recipient block with core size of 1.5 mm using a manual tissue microarrayer (Beecher Instruments, Silver Spring, MD). TMA sections were incubated overnight at 4˚C with primary antibodies of KRT4 (1:100, Clone 6B10, Novacastra, Newcastle upon Tyne, UK), VEGF (1:50, clone G153-694, BD Biosciences), NF-κB/p65 (1:100, NeoMarkers, Fremont, USA), and anti-collagen IV (1:50, Clone COL-94, Biogenex, USA). The standard streptavidin peroxidase method was employed for immuno-staining (8).

Results

Clinical and epidemiological information of enrolled patients. In 64% (59/92) of cases, cancer occurred in the first-degree relatives whereas in 11% (10/92), cancer occurred in the second-degree relatives. In ~10% (9/92) of cases, the cancer occurred in both the first-degree and second-degree relatives. The univariate analysis revealed that the risk of developing esophageal cancer was more among subjects whose family history showed occurrence of cancer among the first-degree relatives (OR: 3.1; CI: 1.9-5.3) than the second-degree relatives (OR: 1.3, CI: 0.25-3.2). The estimates also revealed that the risk of developing esophageal cancer was more in subjects whose pee-degree suffered from esophageal cancer (OR: 2.4; CI: 1.1-4.1) than any other cancer (OR: 1.1; CI: 0.32-3.3).

Gene expression profiling by cDNA microarray. Four hundred and thirty-eight genes were differentially expressed at least 1-fold in all experiments. The two dimensional hierarchical

Table I. Demographic and clinical characteristics of esophageal squamous cell carcinoma cases with family history of esophageal cancer.

Patient ID	Age	Gender	Tobacco chewing habit	Smoking habit	Alcohol use	Betel quid use	Family history of cancer	Pathological grade[a]	Experiments carried out[b]
EC-116	40	F	Yes	No	No	Yes	Esophageal cancer (father)	G2	MA, RT
EC-88	50	M	No	Yes	Yes	Yes	Esophageal cancer (cousin brother)	G3	MA, RT
EC-99	50	M	No	Yes	No	Yes	Esophageal cancer (mother)	G2	MA, RT
EC-84	50	M	No	Yes	Yes	Yes	Esophageal cancer (father)	G1	MA, RT
EC-53	54	M	Yes	No	Yes	Yes	Esophageal cancer (elder brother)	G3	MA, RT
EC-69	70	M	Yes	Yes	No	Yes	Esophageal cancer (mother)	G1	MA, RT
EC-124	69	M	No	Yes	Yes	Yes	Esophageal cancer (brother)	G2	MA, RT
EC-129	65	M	Yes	Yes	No	Yes	Esophageal cancer (sister)	G1	MA, RT,
EC-83	52	F	No	No	No	Yes	Esophageal cancer (father)	G3	MA, RT
EC-54	45	M	No	Yes	Yes	Yes	Esophageal cancer (mother)	G2	RT
EC-72	32	M	No	Yes	No	Yes	Esophageal cancer (father)	G2	RT
EC-247	45	F	No	No	No	Yes	Esophageal cancer (mother)	G2	RT
EC-248	56	M	Yes	Yes	Yes	Yes	Esophageal cancer (brother)	G2	RT
EC-283	55	F	Yes	No	No	Yes	Esophageal cancer (father)	G2	RT
EC-291	55	M	No	Yes	No	Yes	Esophageal cancer (brother)	G2	RT
EC-65	50	F	No	No	No	Yes	Esophageal cancer (parental uncle)	G1	RT
EC-217	56	M	Yes	Yes	Yes	Yes	Esophageal cancer (elder brother)	G1	RT
EC-244	45	M	Yes	Yes	Yes	Yes	Esophageal cancer (elder brother/father)	G2	RT

Table I. Continued.

Patient ID	Age	Gender	Tobacco chewing habit	Smoking habit	Alcohol use	Betel quid use	Family history of cancer	Pathological grade[a]	Experiments carried out[b]
EC-85	85	M	No	Yes	No	Yes	Esophageal cancer (son)	G3	RT
EC-187	58	M	Yes	Yes	Yes	Yes	Esophageal cancer (father)	G2	RT

[a]G1, Well differentiated squamous cell carcinoma; G2, moderately differentiated squamous cell carcinoma; G3, poorly differentiated squamous cell carcinoma; [b]MA, cDNA microarry; RT, real-time PCR for validation of microarray data.

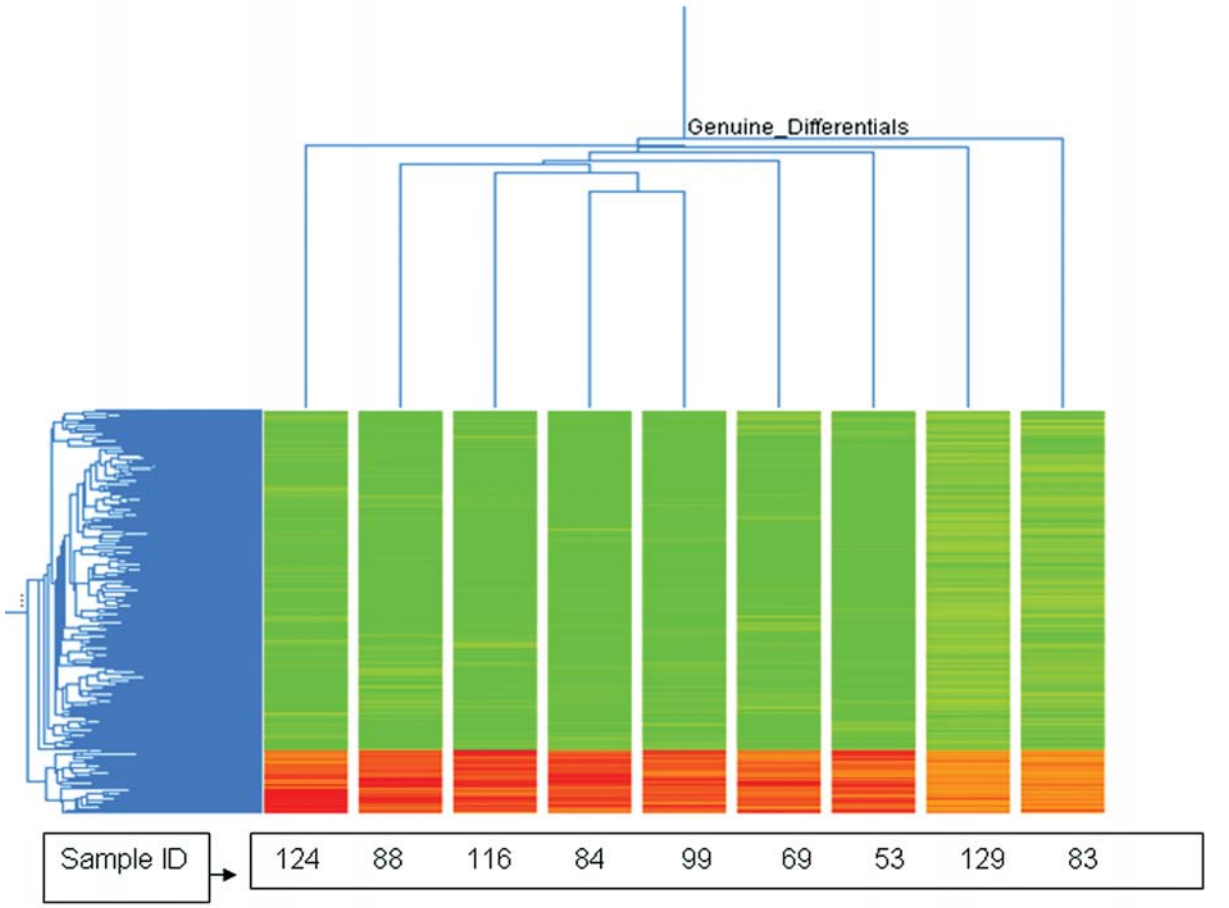


Figure 1. Two-way unsupervised hierarchical clustering (average linkage clustering) of the 438 differentially expressed genes that were over- or underexpressed in tumor vs. normal tissue of nine familial ESCC patients. Red and green colors indicate up-regulated and down-regulated gene expression respectively.

clustering showed that the majority of the differentially expressed genes were significantly down-regulated (84%, 367 genes), whereas 16% genes (71 genes) showed up-regulation (Fig. 1). Using stringent criteria ($P \leq 0.05$ and ≥1.2-fold change), 350 differentially expressed genes (26 up-regulated and 324 down-regulated) were identified and categorized using the Gene Ontology database into known or probable functional categories on the basis of biological processes and molecular function.

Out of 26 significantly up-regulated genes, genes involved in inflammatory response (*CD14*), immune response (*IFITM1*, *VDR*, *CD24*), cell motility (*WASL*), anti-apoptosis (*FOXO1A*), glucocorticoid receptor activity (*NR3C1*), steroid hormone receptor activity (*VDR*), arginase activity (*ARG1*) and metabotropic glutamate, GABA-B-like receptor activity (*GRM3*) were found to be biologically relevant in tumorigenesis (Table IIIA). Out of 324 significantly down-regulated genes, genes involved in extracellular matrix organization (*KRT4*, *COL4A4* and *COL14A1*), BMP signaling pathway (*SMAD1*), epoxide hydrolase activity (*EPHX1*), apoptogenic cytochrome c release channel activity (*VDAC1*, *TXNL1*), DNA damage response (*SMC1A*), humoral immune response (*POU2AF1*, *PF4*, *LY9*, *NFAT5*, *KLRC1*), ion transport (*SLC22A4*, *SLC23A1*) and MAP kinase activity (*MAPK7*,

Table II. Information on the five genes examined by real-time PCR: location, function, primers and probes.

Gene	Gene Bank ID	Location	Gene Expression status in our study	Putative function	P[a]	Assay ID[b]	Amplicon size (bp)
CD14	W56632	5q22-q32	Up-regulated	Inflammatory response	0.0002	Hs02621496_s1	140
ARG1	AA149501	6q23	Up-regulated	Arginine catabolism	0.0002	Hs00163660_m1	86
PF4	AA024929	4q12-q21	Down-regulated	Negative regulation of Angiogenesis	0.0028	Hs00236998_m1	86
EPHX1	AA838691	1q42.1	Down-regulated	Xenobiotic metabolism	0.0002	Hs01116802_m1	89
MAPK7	H39192	17p11.2	Down-regulated	MAP kinase activity	0.0012	Hs00964720_g1	97

[a]Wilcoxon signed rank tests were used to determine the statistical significance of expression difference for each test gene in 20 samples. Statistical significance was defined as $P<0.05$. [b]m1 denotes that assay's probe spans an exon junction and will not detect genomic DNA. s1 denotes that assay's primers and probes are designed within a single exon and will detect genomic DNA. g denotes that assay may detect genomic DNA.

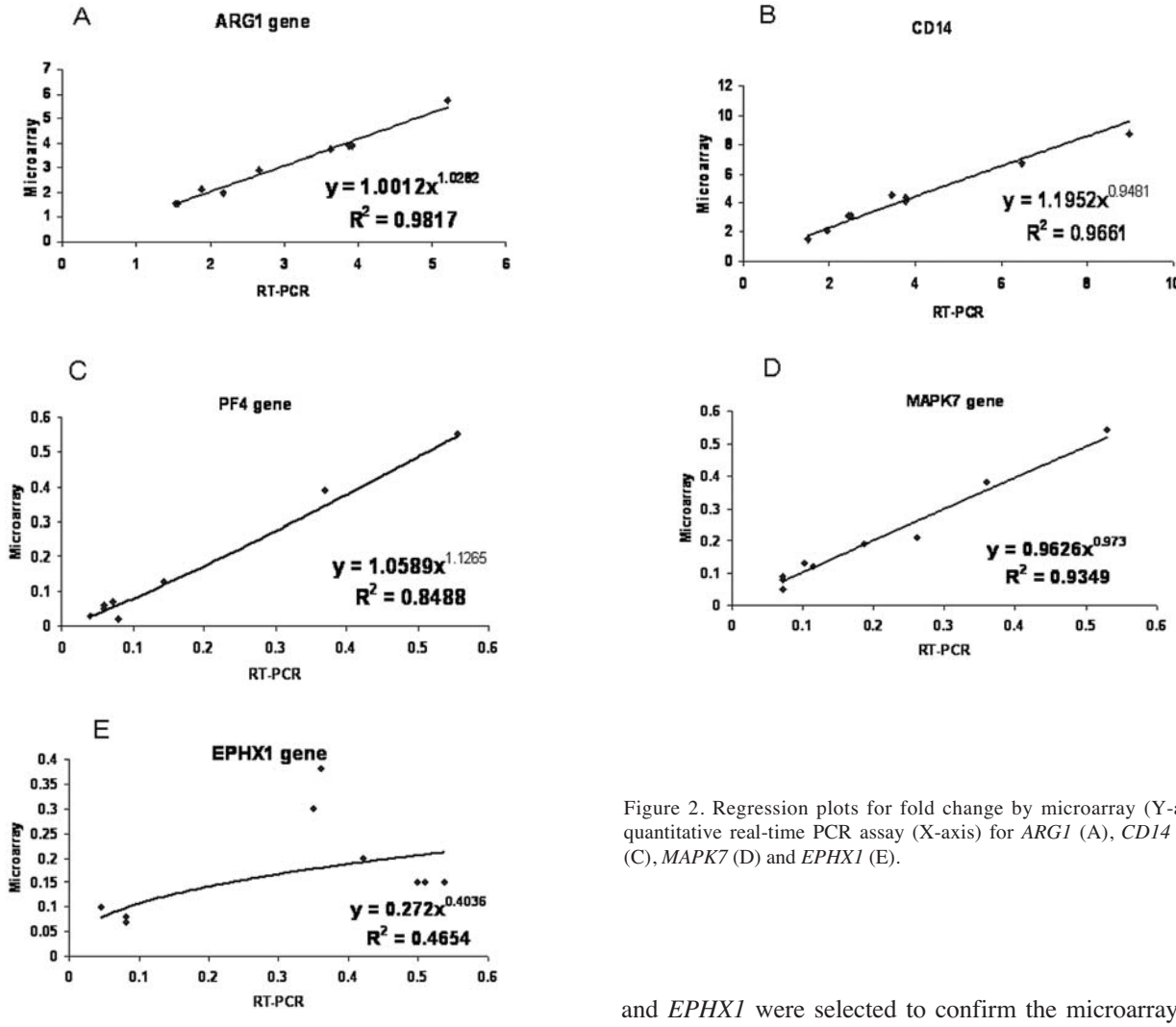


Figure 2. Regression plots for fold change by microarray (Y-axis) and quantitative real-time PCR assay (X-axis) for *ARG1* (A), *CD14* (B), *PF4* (C), *MAPK7* (D) and *EPHX1* (E).

SHC1) were selected on the basis of biological relevance in tumorigenesis (Table IIIB).

Validation of microarray results with quantitative real-time PCR. Five top ranked genes viz *ARG1*, *CD14*, *PF4*, *MAPK7* and *EPHX1* were selected to confirm the microarray results with real-time RT-PCR. Regression plot analyses for the five genes showed positive correlation between the gene expression measured by cDNA microarray and real-time RT-PCR (Fig. 2). Pearson correlation coefficient of each gene was *ARG1* 0.99, *CD14* 0.98, *PF4* 0.99, *MAPK7* 0.98 and *EPHX1* 0.414. Quantitative real-time PCR performed on all the nine patient specimens previously arrayed, 11 other familial ESCC

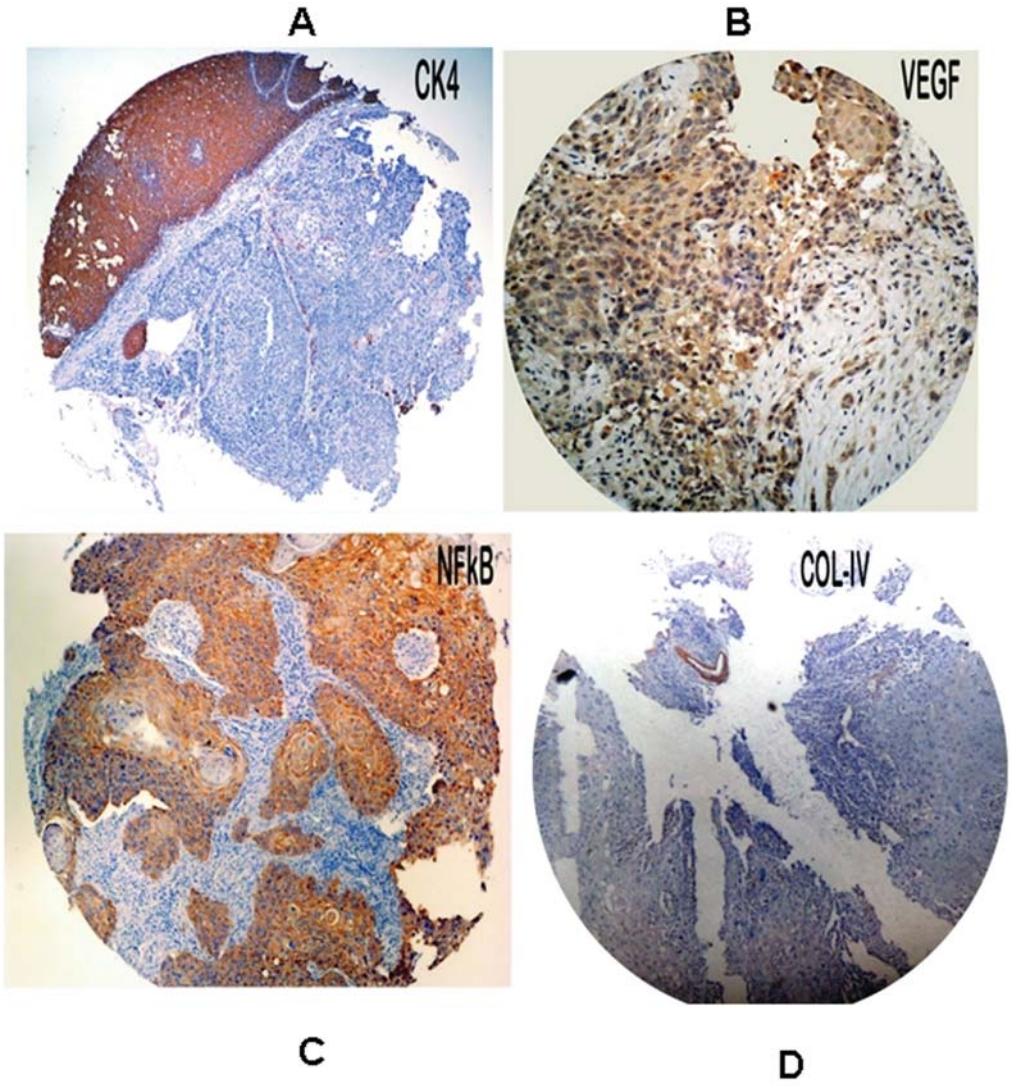


Figure 3. Photomicrograph of tissue microarray cores from esophageal tumor biopsies obtained from familial ESCC patients showing negative immunostaining for KRT4 (A) and Collagen IV (D) and positive immunostaining for VEGF (B) and NF-κB (C).

patient specimens and 10 non-familial ESCC, indicated that *ARG1* and *CD14* were consistently up-regulated and *PF4*, *MAPK7* and *EPHX1* were consistently down-regulated in all patient specimens (Table II). No statistical significant difference (P=0.1508-0.5358) was observed in relative gene expression for five target genes in between familial and non-familial ESCC groups.

TMA-based immunohistochemical analysis. Differential expression of *KRT4* and *COL4* genes identified in cDNA microarray analysis were validated at protein level using immunohistochemistry on TMA. We also studied expression of *VEGF* (which is downstream target gene of *PF4* and *MAPK7*) and *NF-κB* (which is down-stream target gene of *CD14*) at protein level. The cytoplasmic staining was considered positive for CK4 (KRT4), VEGF, NF-κB and collagen 4 (COL4) and was scored as <5% or no staining = 0, 5-25%=1, 26-50%=2, 51-100%=3. Expression of KRT4 was found only in overlying non-neoplastic epithelium and absent in tumor cells (14 out of 20 showing a score of 0 while remaining 6 score of 1). VEGF showed expression in tumor cells with score of 2 in 8 (40%) cases and with score of 3 in 12 (60%) cases. NF-κB showed a diffuse and strong expression in tumor cells with score of 3 in all familial cases (100%). The COL4A4 was not expressed in the tumor cells. The staining intensity was not graded as all the cores were more or less uniformly stained (Fig. 3). Staining pattern in tumor cores from non-familial ESCC cases showed similar staining pattern as in familial cases.

Discussion

To the best of our knowledge, this is the first study that gives an insight into the genes and molecular pathways that may be playing an important role in the familial aggregation of esophageal cancer in high-risk area of India. According to Hanahan and Weinberg (9), tumorigenesis requires six essential alterations to normal cell physiology: self-sufficiency in growth signals, insensitivity to growth inhibition, evasion of apoptosis, immortalization, sustained angiogenesis and tissue invasion and metastasis. In addition, another component of cancer progression is the failure of the host immune response

Table III. Biologically relevant and statistically significant up-regulated and down-regulated genes in esophageal cancer patients with family history of esophageal cancer.

A, Up-regulated genes

Genes	Gene symbol	Gene bank ID	Chromosomal location	Median fold change	GO category	P-value[a]	Pathway[b]
CD14	*CD14*	W56632	5q22-q32	2.05	GO: 6954: inflammatory response	0.00	Toll-like receptor signaling
Arginase	*ARG1*	AA149501	6q23	1.55	GO: 4053: arginase activity	0.01	Urea cycle and metabolism of amino groups
Glutamate receptor, metabotropic 3	*GRM3*	N98673	7q21.1-q21.2	1.54	GO: 8067: metabotropic glutamate, GABA-B-like receptor activity	0.04	Neuroactive ligand receptor interaction
Forkhead box O1A (rhabdomyosarcoma)	*FOXO1A*	W32908	13q14.1	1.43	GO: 6916: anti-apoptosis	0.03	Insulin signaling
Interferon-induced transmembrane protein 1 (9-27)	*IFITM1*	H49853	11p15.5	1.41	GO: 6955: immune response	0.00	B-cell receptor signaling
Wiskott-Aldrich syndrome-like	*WASL*	AI271884	7q31.3	1.41	GO: 6928: cell motility	0.01	Regulation of actin cytoskeleton
Nuclear receptor subfamily 3, group C, member 1 glucocorticoid receptor	*NR3C1*	AA053901	5q31.3	1.32	GO: 4883: glucocorticoid receptor activity	0.00	Neuroactive ligand receptor interaction
Vitamin D (1,25- dihydroxyvitamin D3) receptor	*VDR*	BG149860	12q13.11	1.20	GO: 3707: steroid hormone receptor activity	0.00	NA[c]
Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog	*FGR*	W81591	1p36.2-p36.1	1.42	GO: 6928: cell motility	0.00	Focal adhesion

B, Down-regulated genes

Keratin 4	*KRT4*	AA629189	12q12-q13	-5.20	GO: 30198: extracellular matrix organization and biogenesis	0.00	Cell communication

Table IIIB. Continued.

Genes	Gene symbol	Gene bank ID	Chromosomal location	Median fold change	GO category	P-value[a]	Pathway[b]
SMAD, mothers against DPP homolog 1 (*Drosophila*)	*SMAD1*	R83757	4q31	-2.48	GO: 30509: BMP signaling pathway	0.00	TGF-ß signaling
Epoxide hydrolase 1, microsomal (xenobiotic)	*EPHX1*	AA838691	1q42.1	-2.76	GO: 4301: epoxide hydrolase activity	0.01	Metabolism of xenobiotics by cytochrome p450
Voltage-dependent anion channel 1	*VDAC1*	AA025089	5q31	-2.31	GO: 15283: apoptogenic cytochrome c release channel activity	0.00	Calcium signaling
Platelet factor 4 [chemokine (C-X-C) motif] ligand 4	*PF4*	AA024929	4q12-q21	-4.11	GO: 6959: humoral immune response	0.00	Leukocyte transendo-thelial migration
Solute carrier family 22 (organic cation transporter), member 4	*SLC22A4*	N26836	5q31.1	-2.94	GO: 6811: ion transport	0.00	NA[c]
Mitogen activated protein kinase 7	*MAPK7*	H39192	17p11.2	-2.89	GO: 4707: MAP kinase activity	0.00	GnRH signaling/gap junction
Killer cell lectin-like receptor subfamily C, member 1	*KLRC1*	AA913480	12p13	-2.39	GO: 6959: humoral immune response	0.00	Natural killer cell-mediated cytotoxicity
Nuclear factor of activated T-cells 5	*NFAT5*	H60999	16q22.1	-2.23	GO: 6959: humoral immune response	0.00	Natural killer cell-mediated cytotoxicity
Collagen, type XIV, α 1 (undulin)	*COL14A1*	AA167222	8q23	-2.01	GO: 30198: extracellular matrix organization and biogenesis	0.00	Cell communication
Collagen, type IV, α 4	*COL4A4*	H67349	2q35-q37	-2.18	GO: 30198: extracellular matrix organization and biogenesis	0.00	Cell communication
SHC (Src homology 2 domain containing) transforming protein 1	*SHC1*	R52961	1q21	-3.28	GO: 4707: MAP kinase activity	0.00	Natural killer cell-mediated cytotoxicity

Table IIIB. Continued.

Genes	Gene symbol	Gene bank ID	Chromosomal location	Median fold change	GO category	P-value[a]	Pathway[b]
Solute carrier family 23	*SLC23A1*	AI334656	5q31.2-q31.3	-3.03	GO: 6811: ion transport	0.00	NA[c]
Lymphocyte antigen 9	*LY9*	AI056539	1q21.3-q22	-2.43	GO: 6959: humoral immune response	0.00	NA[c]
POU domain, class 2, associating factor 1	*POU2AF1*	AI028546	11q23.1	-2.36	GO: 6959: humoral immune response	0.00	NA[c]
SMC1 structural maintenance of chromosomes 1-like 1 (yeast)	*SMC1A*	AA598887	Xp11.22-p11.21	-2.43	GO: 42770: DNA damage response	0.01	Cell cycle
Thioredoxin-like 1	*TXNL1*	AA078976	18q21.31	-2.19	GO: 15283: apoptogenic cytochrome c release channel activity	0.00	NA[c]

[a]Biological significance of differentials was computed and functionally classified using the GeneSpring GX software on the basis of gene ontology. GeneSpring GX software depicts the biologically significant ontology for any given gene list as follows: number of genes in the array known to be present in the ontology category vs. number of genes in the array that are differentially regulated. [b]Pathways were obtained using enrichment analysis based on gene ontology categories using the Biointerpreter software and DAVID knowledgebase (http://david.abcc.ncifcrf.gov/knowledgebase). [c]NA, No information available.

to recognize tumor cells (10,11). Molecular profiling of esophageal cancer with familial clustering showed deregulation of most of these physiological mechanisms in the current study (Fig. 4). On the basis of functional annotation, genes responsible for inflammatory response, immune response, angiogenesis, cell migration and cell proliferation were found significantly de-regulated in these cases.

Genes (*KRT4*, *COL4A4*) involved in extracellular matrix organization and cell communication pathway showed significant down-regulation in present study. *CK4* influences the formation of cytoskeletal cells and its low expression has been reported earlier in upper aero-digestive tract tumors (8). Loss of expression of type IV collagen α5 and α6 chains, associated with the hypermethylation of their promoter region, has also been reported in colorectal cancer (12).

Platelet factor 4 (*PF-4*)-a CXC-chemokine-involved in humoral immune response, leukocyte transendothelial migration pathway and inflammatory processes, showed significant down-regulation in our study. *PF4* inhibits T cell function by down-modulating cell proliferation and cytokine release. In addition, *PF-4* has strong anti-angiogenic properties that inhibit endothelial cell proliferation and migration, *in vitro* and *in vivo* angiogenesis, tumor-associated neovascularization and tumor growth (13). *PF-4* inhibits VEGF-induced mitogen-activated protein kinase (MAPK) signaling pathways comprising Raf1, MEK1/2 and ERK1/2 genes (14). *MAPK7* (*ERK5*) was found to be down-regulated in our study. *MAPK7* deficiency leads to an increased expression of VEGF, deregulation of which has been shown to impede angiogenic remodeling and vascular stabilization. Increased VEGF expression in a hypoxic environment promotes vessel growth, angiogenesis and tumor growth (15).

Microsomal epoxide hydrolase 1 (*EPHX1*), involved in metabolism of xenobiotics by CYP450, plays an important role in both the activation and detoxification of tobacco-derived carcinogens (16). In addition, *SLC22A4*, a novel proton antiporter gene that plays a role in the renal excretion of xenobiotics and their metabolites also showed down-regulation in our study. Voltage-dependent anion channel (*VDAC1*) gene, which is involved in calcium signaling and apoptosis inducing pathway, showed down-regulation in present study. *VDAC1* controls pro- and anti-apoptotic Bcl2-family proteins by regulating the release of cytochrome c and apoptotic proteins in the inter-membrane space (17). *SMAD1* that is involved in TGF-ß or BMP (bone morphogenetic proteins) signaling pathway and helps in tumor progression, showed significant down-regulation in the present study. BMPs are involved in wide range of biological activities including cell growth, apoptosis, morphogenesis, development and immune responses. An earlier study has reported that

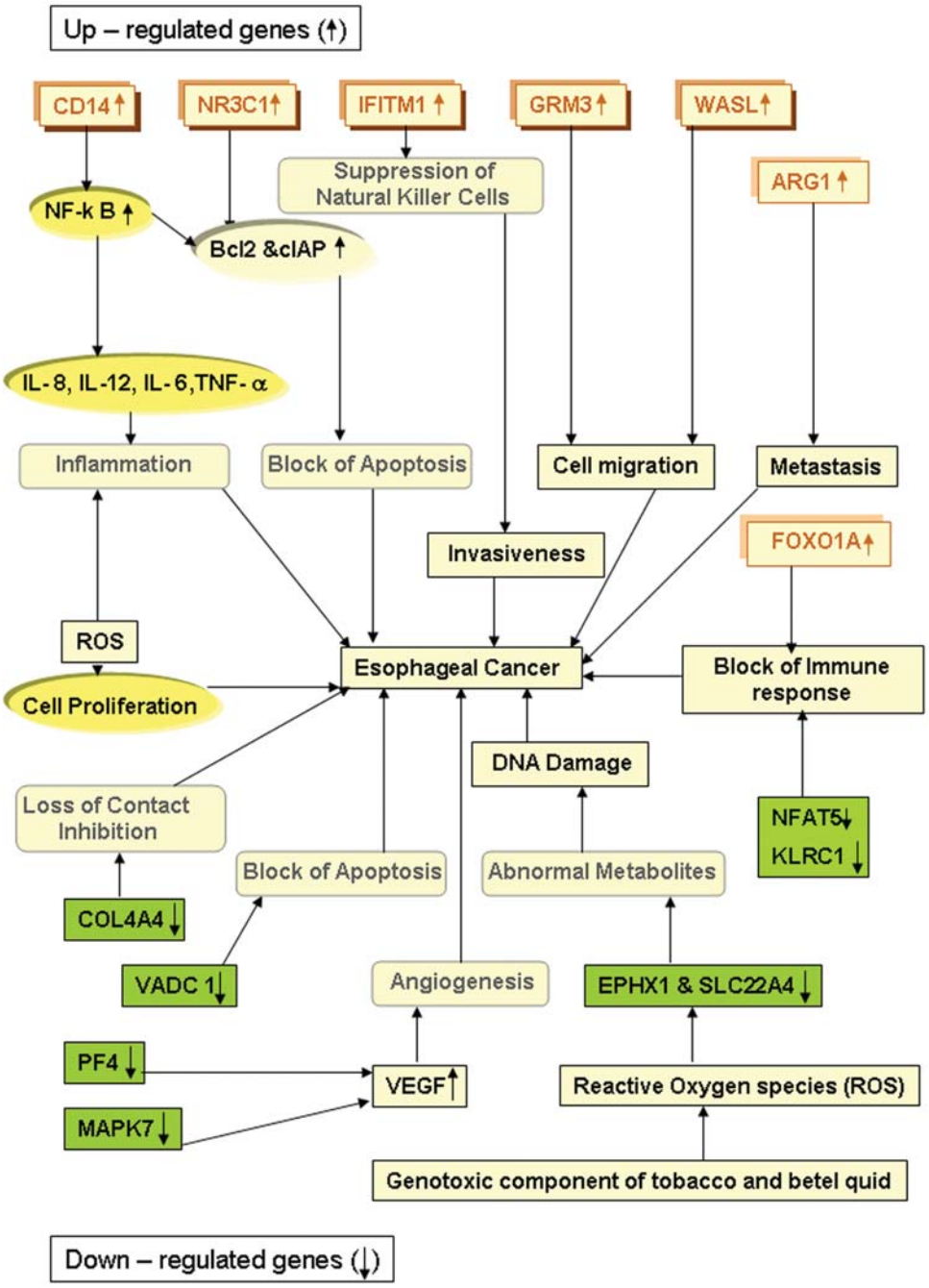


Figure 4. Schematic illustration of differentially expressed genes involved in molecular mechanism of esophageal tumorigenesis. ↑ Indicates up-regulated genes and ↓ indicates down-regulated genes. The key differentially expressed genes (*CD14*, *ARG1*, *EPHX1*, *MAPK7, PF4*, *COL4A4* and *CK4*) were validated by RT-PCR/tissue microarray.

SMAD1 signaling is low in androgen-regulated growth of prostate cancer, is activated after castration and again decreases in hormone-independent tumors (18).

CD14, which is involved in Toll-like receptor (TLR) signaling pathway and inflammatory response, showed significant up-regulation in our study. CD14 induces inflammatory response via MyD88, TIRAP and TRAF6 leading to NF-κB activation and cytokine secretion. The activation of TLR signaling in tumor cells induces the synthesis of proinflammatory factors and immunosuppressive molecules, which enhance the resistance of tumor cells to cytotoxic lymphocyte attack and lead to immune evasion (19).

WASL gene, which is involved in cell motility and regulation of actin cytoskeleton, showed up-regulation in present study. Bourguignon *et al* (20) earlier reported that *N WASP* (*WASL*) played a pivotal role in regulating *CD44-ErbB2* interaction, ss-catenin signaling and actin cytoskeleton functions that were required for tumor-specific behavior such as transcriptional up-regulation and tumor cell migration. Overexpression of interferon inducible 9-27 (*IFITM1*) gene, which is involved in B-cell receptor signaling pathway and immune response, has been reported to play a role in malignant progression by suppressing natural killer cells and by increasing the invasive potential of gastric cancer cells (21).

Gene involved in arginine metabolism (*ARG1*) has been found to be up-regulated in these ESCC patients. Myeloid suppressor cells (MSCs) producing high levels of arginase I block T cell function by depleting l-arginine in cancer, chronic infections, and trauma patients. In cancer, infiltration of MSCs in circulation is an important mechanism for tumor evasion and impairs the therapeutic potential of cancer immunotherapies (22). Overexpression of arginase in colorectal carcinoma is associated with metastasis (23).

FOXO1A, a transcription factor that is involved in anti-apoptosis and insulin signaling pathway, showed up-regulation in our study. It can promote tumor growth and tissue invasion while inhibiting local inflammatory and immune responses. *FOXO1A* is the pathogenetic marker for alveolar rhabdomyosarcoma, an aggressive form of childhood cancer (24). Constitutive phosphorylation of the *FOXO1A* transcription factor has also been reported as a prognostic variable in gastric cancer (25).

NR3C1 (the glucocorticoid receptor family) and *GRM3* (GABA-B-like receptor activity) genes, showing up-regulation in present study, are associated with enhanced anti-apoptotic effect and tumor cell migration, respectively. Both are involved in the neuroactive ligand receptor interaction pathway. The ligand-activated glucocorticoid receptor activates the anti-apoptotic Bcl-2 family protein Bcl-x (L) that inhibits apoptosis and caspase-3 activity in fibrosarcoma cells (26). Activation of the glucocorticoid receptor in epithelial ovarian cancer cells has earlier been reported to have an enhanced cellular expression level of cIAP2 and anti-apoptotic effect (27). The multifunctional G-protein-coupled metabotropic glutamate receptor (mGluR) family contributes to tumor cell migration and invasion in oral cancer (28).

Several differentially regulated genes in familial ESCC are functionally annotated to immune response category. Up-regulation of *CD14*, *WASL*, *IFITM1*, *FOXO1A*, *GRM3*, *ARG1* and *NR3C1* genes is found to be associated with suppression of NK cells, inhibition of immune response, immune evasion, tumor cell migration, invasion, metastasis and anti-apoptosis, respectively. Down-regulation of *PF4*, *SMAD1*, *SLC22A4*, *MAPK7*, *KLRC1*, *NFAT5*, *SHC1*, *LY9*, *POU2AF1* and *VDAC1* genes may be involved in invasion, inhibition of humoral immune response, angiogenesis and anti-apoptosis respectively in these familial ESCC cases. Furthermore, the data presented here will not only provide important information on tumorigenesis of this tumor, but also facilitate the identification of candidate genes that could be used as therapeutic targets for the treatment of patients with this tumor.

Validation of *CD14*, *ARG1*, *PF4*, *MAPK7* and *EPHX1* genes at the mRNA level by real-time PCR and *KRT4*, *COL4*, *NF-κB* and *VEGF* genes at the protein level by tissue microarray did not show any difference in familial and non-familial ESCC cases from the same high-risk area of India, suggesting that familial clustering of cancer in these patients is more due to shared environmental factors rather than shared genes by family members.

In this study, the use of high throughput genomic technology in clinical specimens from well characterized populations that have familial clustering of cancer may lead to identification of molecular mechanism associated with progression of esophageal cancer. Functional analyses of these genes will lead to better understanding of the development and progression of ESCC.

Acknowledgements

The authors are thankful to Jagannath Sharma of BBCI for histopathological analysis, Manisha Sangma, L.C. Singh and Shweta Agarwal of Institute of Pathology, for immunohistochemistry and real-time PCR.

References

1. Parkin DM, Bray F, Ferlay J and Pisani P: Global cancer statistics, 2002. CA Cancer J Clin 55: 74-108, 2005.
2. Stoner GD and Gupta A: Etiology and chemoprevention of esophageal squamous cell carcinoma. Carcinogenesis 22: 1737-1746, 2001.
3. Su H, Hu N, Shih J, Hu Y, Wang QH, Chuang EY, Roth MJ, Wang C, Goldstein AM, Ding T, Dawsey SM, Giffen C, Emmert-Buck MR and Taylor PR: Gene expression analysis of esophageal squamous cell carcinoma reveals consistent molecular profiles related to a family history of upper gastrointestinal cancer. Cancer Res 63: 3872-3876, 2003.
4. Nyren O and Adami H-O: Esophageal Cancer. In: Text Book of Cancer Epidemiology, H-O Adami, D Hunter, D Trichopoulos (eds). Oxford University Press, pp137-155, 2002.
5. Phukan RK, Mahanta J and Hazarika NC: Annual Report, Regional Medical Research Center, Dibrugarh, Assam, India. pp21-24: 2005-2006.
6. Chattopadhyay I, Kapur S, Purkayastha J, Phukan R, Kataki A, Mahanta J and Saxena S: Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis. World J Gastroenterol 13: 1438-1444, 2007.
7. Hemminki K and Li X: Familial risks of cancer as a guide to gene identification and mode of inheritance. Int J Cancer 110: 291-294, 2004.
8. Xue LY, Hu N, Song YM, Zou SM, Shou JZ, Qian LX, Ren LQ, Lin DM, Tong T, He ZG, Zhan QM, Taylor PR and Lu N: Tissue microarray analysis reveals a tight correlation between protein expression pattern and progression of esophageal squamous cell carcinoma. BMC Cancer: 6296, 2006.
9. Hanahan D and Weinberg RA: The hallmarks of cancer. Cell 100: 57-70, 2000.
10. Perwez HS and Curtis CH: Inflammation and cancer: An ancient link with novel potentials. Int J Cancer 121: 2373-2380, 2007.
11. Federico A, Morgillo F, Tuccillo C, Ciardiello F and Loguercio C: Chronic inflammation and oxidative stress in human carcinogenesis. Int J Cancer 121: 2381-2386, 2007.
12. Ikeda K, Iyama K, Ishikawa N, Egami H, Nakao M, Sado Y, Ninomiya Y and Baba H: Loss of expression of type IV collagen alpha5 and alpha6 chains in colorectal cancer associated with the hypermethylation of their promoter region. Am J Pathol 168: 856-865, 2006.
13. Yamaguchi K, Ogawa K, Katsube T, Shimao K, Konno S, Shimakawa T, Yoshimatsu K, Naritaka Y, Yagawa H and Hirose K: Platelet factor 4-gene transfection into tumor cells inhibits angiogenesis, tumor growth and metastasis. Anticancer Res 25: 847-851, 2005.
14. Sulpice E, Contreres JO, Lacour J, Bryckaert M and Tobelem G: Platelet factor 4 disrupts the intracellular signalling cascade induced by vascular endothelial growth factor by both KDR dependent and independent mechanisms. Eur J Biochem 271: 3310-3318, 2004.
15. Sohn SJ, Sarvis BK, Cado D and Winoto A: ERK5 MAPK regulates embryonic angiogenesis and acts as a hypoxia-sensitive repressor of vascular endothelial growth factor expression. J Biol Chem 277: 43344-43351, 2002.
16. Kiyohara C, Yoshimasu K, Takayama K and Nakanishi Y: EPHX1 polymorphisms and the risk of lung cancer: a HuGE review. Epidemiology 17: 89-99, 2006.
17. Shoshan-Barmatz V, Israelson A, Brdiczka D and Sheu SS: The voltage-dependent anion channel (VDAC): function in intracellular signalling, cell life and cell death. Curr Pharm Des 12: 2249-2270, 2006.

18. Qiu T, Grizzle WE, Oelschlager DK, Shen X and Cao X: Control of prostate cell growth: BMP antagonizes androgen mitogenic activity with incorporation of MAPK signals in Smad1. EMBO J 26: 346-357, 2007.
19. Huang B, Zhao J, Unkeless JC, Feng ZH and Xiong H: TLR signaling by tumor and immune cells: a double-edged sword. Oncogene 27: 218-224, 2008.
20. Bourguignon LY, Peyrollier K, Gilad E and Brightman A: Hyaluronan-CD44 interaction with neural Wiskott-Aldrich syndrome protein (N-WASP) promotes actin polymerization and ErbB2 activation leading to beta-catenin nuclear translocation, transcriptional up-regulation, and cell migration in ovarian tumor cells. J Biol Chem 282: 1265-1280, 2007.
21. Yang Y, Lee JH and Kim KY: The interferon-inducible 9-27 gene modulates the susceptibility to natural killer cells and the invasiveness of gastric cancer cells. Cancer Lett 221: 191-200, 2005.
22. Rodriguez PC, Hernandez CP and Quiceno D: Arginase I in myeloid suppressor cells is induced by COX-2 in lung carcinoma. J Exp Med 202: 931-939, 2005.
23. Mielczarek M, Chrzanowska A, Scibior D, Skwarek A, Ashamiss F, Lewandowska K and Baranczyk-Kuzma A: Arginase as a useful factor for the diagnosis of colorectal cancer liver metastases. Int J Biol Markers 21: 40-44, 2006.
24. Nabarro S, Himoudi N and Papanastasiou A: Coordinated oncogenic transformation and inhibition of host immune responses by the PAX3-FKHR fusion oncoprotein. J Exp Med 202: 1399-1410, 2005.
25. Kim JH, Kim MK and Lee HE: Constitutive phosphorylation of the FOXO1A transcription factor as a prognostic variable in gastric cancer. Mod Pathol 20: 835-842, 2007.
26. Gascoyne DM, Kypta RM and Vivanco MM: Glucocorticoids inhibit apoptosis during fibrosarcoma development by transcriptionally activating Bcl-xL. J Biol Chem 278: 18022-18029, 2003.
27. Runnebaum IB and Bruning A: Glucocorticoids inhibit cell death in ovarian cancer and up-regulate caspase inhibitor cIAP2. Clin Cancer Res 11: 6325-6332, 2005.
28. Park SY, Lee SA and Han IH: Clinical significance of metabotropic glutamate receptor 5 expression in oral squamous cell carcinoma. Oncol Rep 17: 81-87, 2007.

Cytokeratin Immunoexpression in Esophageal Squamous Cell Carcinoma of High-risk Population in Northeast India

Avninder Singh, MD, Sujala Kapur, MD,* Indranil Chattopadhyay, MSc,* Joydeep Purkayastha, MS,† Jagannath Sharma, MD,† Ashwani Mishra, PhD,* Stephen M. Hewitt, MD, PhD,‡ and Sunita Saxena, MD**

Abstract: Esophageal cancer is a frequently fatal malignancy, and is described in certain regions in Northeast India with an incidence of esophageal squamous cell carcinoma many fold higher than the rest of the population. The population in Northeast India is at higher risk due to poor nutritional status, consumption of fermented betel quid and other oral tobacco products besides smoking and alcohol intake. Cytokeratins (CKs) are the major constituents of the esophageal epithelium and may show gain or loss of CKs as the cancer progresses from normal epithelium to invasive phenotype. In this study, we studied the immunohistochemical expression of 5 CKs (CK4, CK5, CK8, CK14, and CK17) in the normal esophageal epithelium and esophageal squamous cell carcinoma from both the general population and the high-risk population of Assam in Northeast India. The CK expression profile was similar to other published data in general. Further analysis demonstrated differences in CK expression between the general and the high-risk tumor samples. CK5 and CK8 expression was altered in the high-risk population. The significance of these differences is unclear, but suggests a connection to the etiologic factors.

Key Words: cytokeratins, esophageal cancer, high risk, immunohistochemistry, squamous cell carcinoma, tissue microarray

(*Appl Immunohistochem Mol Morphol* 2009;00:000–000)

Esophageal cancer is a frequently fatal malignancy known to occur in very high frequencies in certain parts of the world such as Northern China, Iran, South Africa, and India with up to 50-fold difference has been observed between high-risk and low-risk populations.[1,2] The highest incidence of cancer of the esophagus in India has been reported from Assam in the Northeast region, where it is the second leading cancer in men and third leading cancer in women.[3] The age adjusted incidence rates per 100,000 persons for males and females from the high-risk region of Assam in Northeast India have been reported to be 32.6 and 21.1 in the district of Kamrup and 15.7 and 8.1 in the district of Dibrugarh as compared with adjusted incidence rates of 4.7 and 3.1 in Delhi.[4] Besides smoking and alcohol consumption that are the major risk factors for developing esophageal cancer, association of poor nutritional status, and dietary habits such as consumption of unconventional smokeless forms of tobacco, chewing betel quid, and consuming fermented preserved food predispose the people from this region with higher mortality rates.[5]

As cytokeratin (CK) are the major constituent of the esophageal epithelium and esophageal squamous cell carcinoma (ESCC) is the predominant histologic subtype of esophageal cancer in India, we chose to study the immunohistochemical expression of those keratins that had shown differential gene expression on microarray (unpublished data) in the high-risk population of Northeast India. The objective of this study was to ascertain whether there was any interrelationship of CK expression between the normal epithelium and ESCC between high-risk population and the general (low-risk) population of Delhi which may have potential for classifying ESCC in these population groups.

MATERIALS AND METHODS

Patients and Samples

A written informed consent was obtained from the patients to use these samples for research and ethical clearance was obtained for this study from the institutional ethical committee of Institute of Pathology, Indian Council of Medical Research.

This study was performed on 125 formalin-fixed, paraffin-embedded tissue specimens obtained by endoscopic biopsy of normal esophageal epithelium and ESCC cases received during the period between January 2006 and July 2007. Of these, 85 comprised ESCC obtained from B. Borooah Cancer Institute, Guwahati belonging to the high-risk population of Northeast India, 25

Received for publication October 29, 2008; accepted January 19, 2009.
From the *Institute of Pathology, Indian Council of Medical Research, New Delhi; †B. Borooah Cancer Institute, Guwahati, Assam, India; and ‡Tissue Array Research Laboratory, National Cancer Institute, National Institutes of Health, Bethesda.
Source of support: Financial support to Avninder Singh for the ICRETT technology transfer fellowship UICC (International Union against Cancer). This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.
Reprints: Stephen M. Hewitt, MD, PhD, Tissue Array Research Laboratory, National Cancer Institute, National Institutes of Health, Bethesda 20892-4605 (e-mail: hewitts@mail.nih.gov).

biopsies of ESCC from low-risk general population of Delhi, and 15 endoscopic biopsies from normal esophageal mucosa taken from a distant site of ESCC cases from high-risk patients as control. Hematoxylin and eosin-stained slides of all these samples were reviewed by 2 pathologists (A.S. and S.M.H.) and graded into well, moderately, and poorly differentiated ESCC.

Tissue Microarray Construction

A tissue microarray (TMA) was constructed from formalin-fixed, paraffin-embedded blocks of 125 tissue samples that included 15 controls, 25 ESCC from low-risk population of Delhi, and 85 ESCC from high-risk population. Sampling sites were marked on the donor blocks and the tissue cylinders precisely arrayed into 2 recipient blocks each with a core size of 1.5-mm using a manual tissue microarrayer (Beecher Instruments, Silver Spring, MD). The recipient blocks were tempered at 37°C overnight. A TMA section was stained with hematoxylin and eosin for quality assurance and confirmation of diagnostic elements on the TMA.

Immunohistochemistry

The LSAB$^+$ detection system (Dako, Carpinteria, CA) was used for immunohistochemical staining. Briefly, 5-μm sections were cut from TMA blocks, deparaffinized in xylene and rehydrated in graded alcohol. Heating the sections immersed in 0.01 M sodium citrate buffer in a steamer for 20 minutes performed antigen retrieval. After bringing the sections to room temperature the sections were incubated with 3% hydrogen peroxide for 10 minutes to block endogenous peroxidase activity. The primary antibodies used were CK4 (1:100, Clone 6B10, Novocastra laboratories, Norwell, MA), CK5 (1:100, Clone EP1601Y, Abcam Inc), CK8 (prediluted, Clone 35 βH11, Dako,), CK14 (1:100, Clone LL002, NeoMarkers, Fremont, CA) and CK17 (1:20, Clone E3, Dako). 3, 3′-Diaminobenzidine was used as the chromogen and sections were counterstained with hematoxylin.

Evaluation of Immunohistochemistry

The TMA sections were screened and scored by 2 pathologists (A.S., S.M.H.) who independently evaluated the immunohistochemical staining and reached a consensus. An adequate core was defined as one in which the tumor occupied more than 50% of the core area.

Out of 125 cores, 114 cores were found suitable for evaluation due to loss of cores, inadequate tumor or insufficient immunohistochemical staining. The cytoplasmic staining was considered positive staining for all the CKs. The distribution of positive staining (X) was scored as percentage of cells labeled by counting individual tumor cells under ×40 objective using an eyepiece pinhole and then graded as 0 = no staining, 1 = upto 30%, 2 = 31% to 60%, and 3 = more than 60%. As immunohistochemical staining intensity in the cores was not uniform, the intensity of staining (Y) was graded as 0 = no staining, 1 = mild, 2 = moderate, and 3 = strong. The final score (X × Y) was the multiplication of distribution score by intensity score, generating scores over a range of 0 to 9. The immunoexpression was divided into 2 groups based on the final score as underexpressed (score below 6) and as overexpressed (score of 6 and above).

Statistical Analysis

Pearson χ^2 tests with SPSS 13.0 for Windows (SPSS Chicago, IL) were performed to compare the expression of the CK proteins among the control, low-risk and high-risk population samples. P value of less than 0.05 was considered statistically significant.

RESULTS

Of the 100 ESCC patients from both groups, 86% were male and 14% were female with a median age of 56.4 years (range: 39 to 85). ESCC cases were further subdivided into well differentiated (n = 22), moderately differentiated (n = 66), and poorly differentiated (n = 12). Immunohistochemical staining for CKs showed cytoplasmic as demonstrated in Figure 1. The normal esophageal epithelium was taken as baseline in which CK4, CK5, CK8, and CK17 stained the stratified squamous epithelium in different patterns and intensities, whereas CK14 showed staining of only the basal cells. Detailed distributions of staining for the individual CKs are summarized in Table 1.

Utilizing a dichotomization, based on a scoring index of 6, CK4 expression was identical in tumors from the low and high-risk groups, staining 0/19 and 0/81, respectively, compared with the normal epithelium of the controls (n = 14) which were all positive (Table 2). CK5 expression was also present in the normal epithelium of all of the controls, but only 7/19 (36.8%) of the tumors from the low-risk population compared with 65/81 (80.3%) of the tumors n the high-risk population. The difference in expression in tumors between low and high-risk populations was statistically significant, $P < 0.001$.

Staining for CK8 showed a similar pattern, present in the normal epithelium of all control patients, but 7/19 (36.0%) of tumors from patients in the low-risk population and 68/81 (84%) of tumors from the high-risk population, resulting in a P value of less than 0.001 between low and high-risk populations. CK14 expression was limited to the normal epithelium of 1/14 (7.2%) patients in the control population, but was detectable in 12/19 (63.%) and 65/81 (80.3%) of tumors from low and high-risk populations, respectively.

CK17 expression was found in the normal epithelium of all the normal controls, but was limited to 15/19 (78.9%) and 69/81 (85.2%) of tumors from low and high-risk populations, respectively. Detailed analysis failed to demonstrate a statistically significant correlation between age, sex, and histologic grade of ESCC between high and low-risk groups. Further analysis failed to demonstrate statistical significance of differentiation and staining for the individual CKs (Table 3).

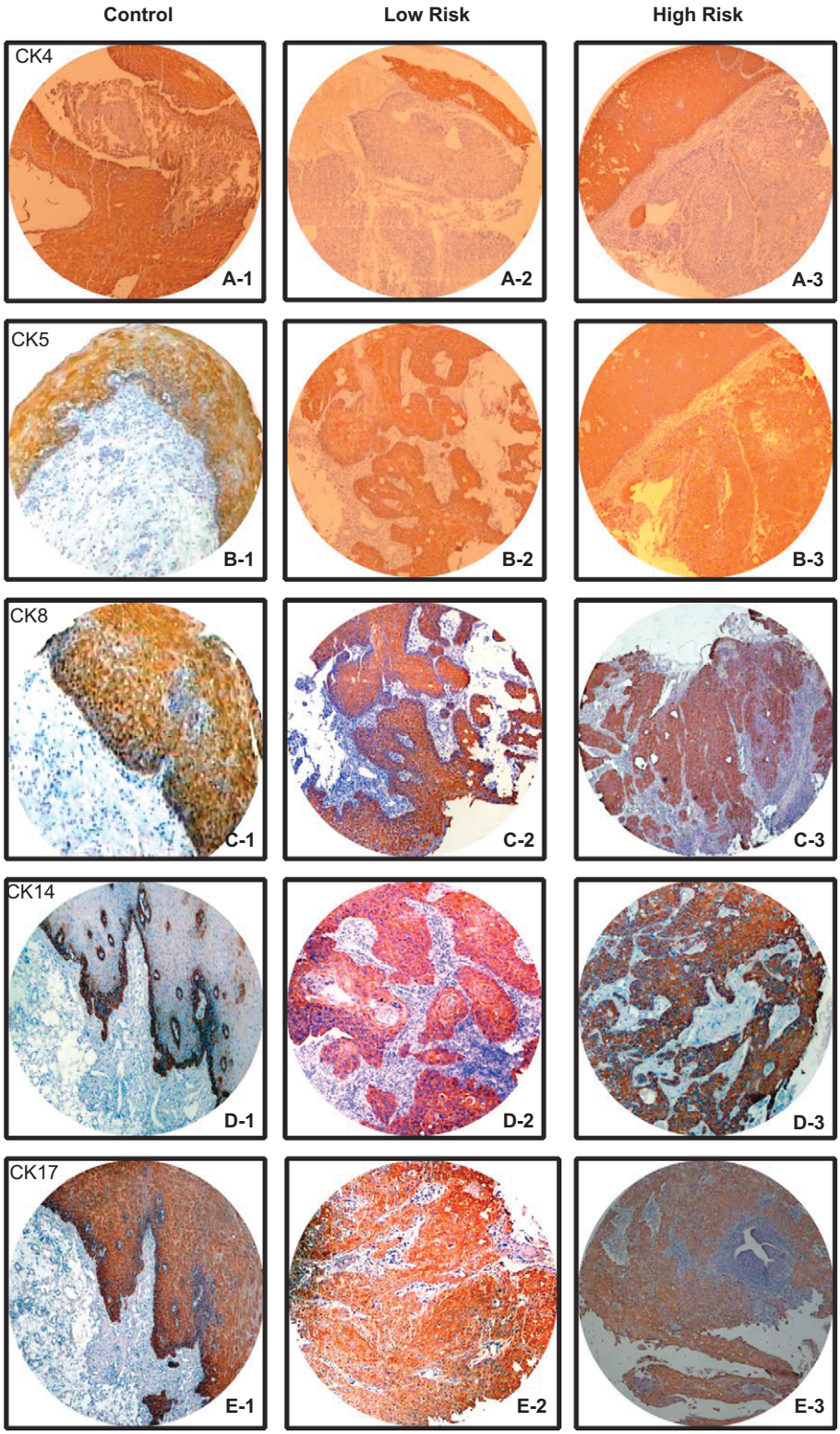


FIGURE 1. Tissue microarray cores showing cytokeratin (CK) staining pattern. CK4: A-1, normal; A-2, low risk; A-3, high risk. CK5: B-1, normal; B-2, low risk; B-3, high risk. CK8: C-1, normal; C-2, low risk; C-3, high risk. CK14: D-1, normal; D-2, low risk; D-3, high risk. CK17: E-1, normal; E-2, low risk; E-3, high risk ($\times$ 100).

TABLE 1. Immunohistochemical Staining Pattern of Cytokeratins

	Distribution (%)				Intensity (%)			
	0: No Stain, 1: Up to 35%, 2: 36-65, 3: >65				0: No Stain, 1: Mild, 2: Moderate, 3: Strong			
Cytokeratins	Grade (X)	Control, n = 14	Low Risk, n = 19	High Risk, n = 81	Grade (Y)	Control, n = 14	Low Risk, n = 19	High Risk, n = 81
CK4	0	0/14 (0.0)	14/19 (73.7)	52/81 (64.2)	0	0/14 (0.0)	0/19 (0.0)	54/81 (66.7)
	1	0/14 (0.0)	1/19 (5.3)	28/81 (34.6)	1	0/14 (0.0)	4/19 (21)	27/81 (33.3)
	2	0/14 (0.0)	1/19 (5.3)	1/81 (1.2)	2	1/14 (7.1)	8/19 (42.1)	0/81 (0.0)
	3	14/14 (100)	3/19 (15.8)	0/81 (0.0)	3	13/14 (92.9)	7/19 (36.8)	0/81 (75.3)
CK5	0	0/14 (0.0)	0/19 (0.0)	2/81 (2.5)	0	0/14 (0.0)	0/19 (0.0)	2/81 (2.5)
	1	0/14 (0.0)	4/19 (21.1)	6/81 (7.4)	1	0/14 (0.0)	5/19 (26.3)	5/81 (6.2)
	2	1/14 (7.1)	8/19 (42.1)	12/81 (14.8)	2	3/14 (21.4)	9/19 (47.4)	20/81 (24.7)
	3	13/14 (92.9)	7/19 (36.8)	61/81 (75.3)	3	11/14 (78.6)	5/19 (26.3)	54/81 (66.7)
CK8	0	0/14 (0.0)	4/19 (21.1)	3/81 (3.7)	0	0/14 (0.0)	4/19 (21.1)	3/81 (3.7)
	1	0/14 (0.0)	6/19 (31.6)	6/81 (7.4)	1	0/14 (0.0)	3/19 (15.8)	8/81 (9.9)
	2	0/14 (0.0)	4/19 (21.1)	13/81 (16)	2	0/14 (0.0)	7/19 (36.8)	7/81 (8.6)
	3	14/14 (100)	5/19 (26.3)	59/81 (72.8)	3	14/14 (100)	5/19 (26.3)	63/81 (77.8)
CK14	0	0/14 (0.0)	0/19 (0.0)	3/81 (3.7)	0	0/14 (0.0)	0/19 (0.0)	3/81 (3.7)
	1	12/14 (85.7)	3/19 (15.8)	11/81 (13.6)	1	0/14 (0.0)	2/19 (10.5)	2/81 (2.5)
	2	1/14 (7.1)	7/19 (36.8)	33/81 (40.7)	2	3/14 (21.4)	7/19 (36.8)	6/81 (7.4)
	3	1/14 (7.1)	9/19 (47.4)	34/81 (42)	3	11/14 (78.6)	10/19 (52.6)	70/81 (86.4)
CK17	0	0/14 (0.0)	1/19 (5.3)	3/81 (3.7)	0	0/14 (0.0)	1/19 (5.3)	3/81 (3.7)
	1	0/14 (0.0)	1/19 (5.3)	8/81 (9.9)	1	0/14 (0.0)	2/19 (10.5)	2/81 (2.5)
	2	0/14 (0.0)	4/19 (35.7)	19/81 (23.5)	2	5/14 (35.7)	5/19 (26.3)	16/81 (19.7)
	3	14/14 (100)	13/19 (68.4)	51/81 (63)	3	9/14 (64.3)	11/19 (57.9)	60/81 (74.1)

DISCUSSION

CKs are a family of about 30 different related cytoplasmic proteins encoded by multiple genes and constitute the intermediate filaments that are incorporated into the cytoskeleton of almost all epithelial cells. The epithelial or soft keratins (CK1 to CK20) are divided into 2 subfamilies, type 1 (acidic, CK9 to CK20) and type 2 (basic, CK1 to CK8) depending on their molecular weights and isoelectric point.[6] These keratin polypeptides are the product of different genes and are expressed in different cells at different stages of development and differentiation.[7] Carcinoma cells retain an ability to produce the CKs of their progenitor cells and may also gain an ability to develop new types of intermediate filaments.[8,9] Therefore, CK subtypes have potential to be useful biomarkers for studying the evolution of cancer in different stages and differentiation states.

CKs are the major constituents of the normal esophageal epithelium and epithelial cancer. In past, several reports[10–13] have shown alterations in the CK protein expression in the normal esophageal epithelium in comparison with ESCC, however, these studies were based on western blot analyses and were not possible to relate keratin expression to specified cell populations or different states of cell differentiation. Chung et al[14] were able to demonstrate reduction of CK4 and CK14 in the transition from normal epithelium to invasive tumor in a small number of cases. Vianne and Baert[15] studied the relationship between morphologic characteristics and in situ hybridization for CK-mRNAs and demonstrated that changes in CK expression occur with differences in malignant potential in the esophageal squamous epithelium. In their CK-mRNA expression they showed that in well-differentiated ESCC, CK6 and CK14 were

TABLE 2. Immunohistochemical Expression Pattern of Cytokeratins (CKs)

					P		
Cytokeratins	Score	Control (%) (n = 14)	Low Risk (%) (n = 19)	High Risk (%) (n = 81)	P_a	P_b	P_c
CK4	< 6	0	19/19 (100)	81/81 (100)	< 0.001	< 0.001	Not computable
	> 6	14/14 (100)	0	0			
CK5	< 6	0	12/19 (63.2)	16/81 (19.7)	< 0.001	0.06	< 0.001
	> 6	14/14 (100)	7/19 (36.8)	65/81 (80.3)			
CK8	< 6	0	12/19 (63.2)	13/81 (16)	< 0.001	0.107	< 0.001
	> 6	14/14 (100)	7/19 (36.8)	68/81 (84%)			
CK14	< 6	13/14 (92.8)	7/19 (36.8)	16/81 (19.7)	0.001	< 0.001	0.197
	> 6	1/14 (7.2)	12/19 (63.2)	65/81 (80.3)			
CK17	< 6	0	4/19 (21.1)	12/81 (14.8)	0.067	0.123	0.5
	> 6	14/14 (100)	15/19 (78.9)	69/81 (85.2)			

P_a indicates control versus low risk; P_b, control versus high risk; P_c, low risk versus high risk.

TABLE 3. Cytokeratin Expression in Relation to Histologic Grade

Score	Grade	Cytokeratins															
		CK4			CK5			CK8			CK14			CK17		
		Low Risk (%)	High Risk (%)	P	Low Risk (%)	High Risk (%)	P	Low Risk (%)	High Risk (%)	P	Low Risk (%)	High Risk (%)	P	Low Risk (%)	High Risk (%)	P
< 6	Well differentiated	3/22 (13.6)	19/22 (86.4)	Not comp	2/7 (28.5)	5/7 (71.4)	0.163	2/5 (40)	3/5 (60)	0.051	0	3/3 (100)	0.459	0	3/3 (100)	0.459
	Moderately differentiated	14/66 (13.6)	52/66 (78.8)		8/13 (61.5)	5/13 (38.4)	< 0.001	9/17 (53)	8/17 (47)	< 0.001	5/16 (31.2)	11/16 (68.8)	0.259	3/10 (30)	7/10 (70)	0.461
	Poor differentiated	2/12 (16.7)	10/12 (83.3)		2/8 (25.0)	6/8 (75.0)	0.273	1/3 (33.3)	2/3 (66.6)	0.371	2/4 (50)	2/4 (50)	0.028	1/3 (33.3)	2/3 (66.6)	0.371
≥ 6	Well differentiated	0	0		1/15 (6.7)	14/15 (93.3)		1/17 (5.9)	16/17 (94.1)		3/19 (15.8)	16/19 (84.2)		3/19 (15.8)	16/19 (84.2)	
	Moderately differentiated	0	0		6/53 (11.4)	47/53 (88.6)		5/49 (10.2)	44/49 (89.8)		9/50 (18)	41/50 (82)		11/56 (19.6)	45/56 (80.4)	
	Poor differentiated	0	0		0	4/4 (100)		1/9 (11.2)	8/9 (88.8)		0	8/8 (100)		1/9 (11.2)	8/9 (88.8)	

abundantly expressed whereas CK4 and CK13 were only focally expressed and were in contrast to the normal epithelium. In the moderately differentiated ESCC, CK6, CK14, and CK19 were detected whereas in the poorly differentiated ESCC none of these were detected.

With the availability of monoclonal antibodies for small subsets of keratin polypeptides, it was important to know what components of esophageal cells were related to cell differentiation and the need for indicators other than morphologic parameters in evaluating precursor lesions, prognosis, and therapeutic intervention and it was a challenge for researchers. Yang and Lipkin[16] were the first to study CK immunoexpression using AE1 antibody and concluded that there were 3 patterns of CK expression that paralleled the degree of differentiation. Type 1 pattern was present in well-differentiated ESCC whereas types 2 and 3 were seen in poorly differentiated and undifferentiated carcinomas. In dysplasia and carcinoma in situ abnormal cells had reaction patterns in which they lost or increased AE1 expression suggesting that AE1 could be used as a biomarker for identifying early aberrations in the esophageal epithelium.

In another study, Takahashi et al[17] reported that neoplastic epithelial cells showed different keratin reactivity and distribution compared with normal esophageal epithelium. CK17 was limited only to the keratinized cells whereas CK14 selectively labeled the basal cells of the normal esophageal epithelium and all the ESCC cells and adjacent carcinoma in situ. Itakura et al[18] studied esophageal dysplasia both from cancer and cancer-free patients using antibodies to CK10/13 and CK14 and concluded that abnormal patterns of CK expression occur not only in atypical cells but also in nonatypical cells in esophageal squamous dysplasia. Lam et al[19] studied 35 cases of ESCC and reported that CK18 and CK19 are overexpressed in ESCC whereas expression of CK10 is seen in well-differentiated ESCC and not in normal epithelium. Cintorino et al[20] suggested from their study on ESCC in high-risk population of China that significant differences in expression of CK8, CK18, and CK19 were seen between lymph node positive and lymph node negative ESCC and the pattern of distribution of these CKs could be of predictive value. More recently, Xue et al[21] studied 205 ESCC and 173 precursor lesions and found that CK14 were positive only in basal layer of normal whereas diffusely positive in ESCC and this overexpression occurred in intermediate stage of carcinogenesis between dysplasia to carcinoma in situ and carcinoma whereas the underexpression of CK4 in the ESCC was an early event occurring in the mild-to-moderate dysplasia stage and suggests malignant transformation in this cancer.

In this study, we compared the CK expression of CK4, CK5, CK8, CK14, and CK17 between normal esophageal epithelium and ESCC and then between the low-risk and high-risk populations. We found that CK4 underexpression ($P < 0.001$), and the overexpression of CK5 ($P < 0.001$), CK8 ($P < 0.001$), and CK14 ($P = 0.001$) were useful in differentiating normal

epithelium from cancer in the general population whereas CK4 underexpression ($P < 0.001$) and CK14 overexpression ($P < 0.001$) were useful to differentiate normal epithelium from ESCC in the high-risk population. This finding provides some insights into the alterations in CK expression encountered with malignant transformation. The findings of CK4 underexpression and CK5, CK8, and CK14 overexpression were in concordance with previously published reports.

CK8 is expressed in a wide variety of "simple epithelial cells."[22] CK4 is expressed in stratified and columnar epithelium, and typically absent from the basal cells that give rise to these epithelial layers.[22] In contrast CK5 and CK14 are expressed in the basal layer of squamous epithelium, and not expressed in the stratified elements. With reference to the differential expression between the low-risk and high-risk groups, the relationships of the CKs are less clear. CK8 is described as focally expressed in squamous cell carcinomas, possibly with greater expression in poorly differentiated cases. CK5 is associated with basal or myoepithelial phenotype, and is noted to be expressed in small cell carcinomas of the lung.[22] The greater level of expression of CK5 and CK8 in the high-risk group may suggest a connection with a basal/small cell phenotype of tumor, which would be concordant with a more poorly differentiated phenotype.

More interestingly, our analysis shows that CK5 is more commonly expressed in the esophageal tumors of patients from the high-risk cohort compared with the low-risk cohort. ($P < 0.001$). Additionally, CK8 was also expressed more commonly in the tumors of patients from the high-risk cohort than the low-risk cohort ($P < 0.001$). As a result CK5 and CK8 expression did not decrease with malignant transformation in the high-risk cohort, compared with the low-risk cohort. These changes were independent of differentiation state of the tumor. To our knowledge this is the first example of detection of differences in CK expression, measured by immunohistochemistry that can be correlated to high-risk groups, attributed to tobacco use or diet. Examination of dysplastic or in situ lesions may yield a greater understanding of the significance of these changes in CK expression.

In conclusion, the findings from this study suggest that CK5 and CK8 may be useful markers in separating high-risk from low-risk population groups and detection of changes in the expression of these gene products may be able to shed some light on identification of biomarkers that can help for early diagnosis and may also help to stratify populations at low or high risk.

ACKNOWLEDGMENT

The authors acknowledge the help of Mr Kris Ylaya, for technical support in TMA construction. Financial support to Avninder Singh was in the form of an ICRETT technology transfer fellowship from the UICC (International Union against Cancer).

REFERENCES

1. Parkin DM, Bray F, Ferlay J, et al. Global cancer statistics. *CA Cancer J Clin*. 2005;55:74–108.
2. Akbari MR, Mohammadkhani A, Fakheri H, et al. Familial risks of esophageal cancer among Turkman population of the Caspian littoral of Iran. *Int J Cancer*. 2006;119:1047–1051.
3. Chattopadhyay I, Kapur S, Purkaystha J, et al. Gene expression profile of esophageal cancer in Northeast India by cDNAmicroarray analysis. *World J Gastroenterol*. 2007;13:1438–1444.
4. Consolidated report of population based cancer registry 2001-2004. *http://www.icmr.nic.in/ncrp/report*.
5. Phukan RK, Chetia CK, Ali MS, et al. Role of dietary habits in the development of esophageal cancer in Assam, the North-eastern region of India. *Nutr Cancer*. 2001;39:204–209.
6. Moll R, Franke WW, Schiller DL, et al. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell*. 1982;31:11–24.
7. Cooper D, Schermer A, Sun TT. Classification of human epithelia and their neoplasm using monoclonal antibodies to keratins: strategies, applications and limitations. *Lab Invest*. 1985;52:243–56.
8. Fuchs E, Grace MP, Kim HK, et al. Differential expression of two classes of keratins in normal and malignant epithelial cells and their evolutionary conservation. In: Levine AJ, Topp W, van de Woude G, et al., eds. *Cancer Cells. 1: The Transformed Phenotype*. New York: Cold Spring Harbor Laboratory; 1984:161–167.
9. Gown AM, Gabbiani G. Intermediate-size (10 nm) filaments in human tumors. In: DeLellis RA, ed. *Advances in Immunohistochemistry*. New York: Masson; 1984:89–109.
10. Banks-Schlegel SP, Harris CC. Tissue specific expression of keratin proteins in human esophageal and epidermal epithelium and their cultured keratinocytes. *Exp Cell Res*. 1983;146:271–280.
11. Banks-Schlegel SP, Harris CC. Aberrant expression of keratin proteins and cross-linked envelopes in human esophageal carcinomas. *Cancer Res*. 1984;44:1153–1157.
12. Grace MP, Kim KH, True LD, et al. Keratin expression in normal esophageal epithelium and squamous cell carcinoma of the esophagus. *Cancer Res*. 1985;45:841–846.
13. Moll R, Krepler R, Franke WW. Complex cytokeratin polypeptide pattern observed in certain human carcinomas. *Differentiation*. 1983;23:256–269.
14. Chung JY, Braunshweig T, Hu N, et al. Multiplex tissue immunoblotting for proteomic profiling: a pilot study of the normal to tumor transition of esophageal squamous cell carcinoma. *Cancer Epidemiol Biomarker Prev*. 2006;7:1403–1408.
15. Vianne A, Baert J. Expression of cytokeratin-mRNAs in squamous cell carcinoma and balloon-cell formation of human oesophageal epithelium. *Histochem J*. 1995;27:69–78.
16. Yang K, Lipkin M. AE1 cytokeratin reaction patterns in different differentiation states of squamous cell carcinoma of the esophagus. *Am J Clin Pathol*. 1990;27:261–269.
17. Takahashi H, Shikata N, Senzaki H, et al. Immunohistochemical staining patterns of keratins in normal oesophageal epithelium and carcinoma of the oesophagus. *Histopathology*. 1995;26:45–50.
18. Itakura Y, Sasano H, Abe K, et al. Cytokeratin immunolocalization and lectin binding studies in esophageal squamous dysplasia. *Histopathology*. 1996;29:3–10.
19. Lam KY, Loke SL, Shen XC, et al. Cytokeratin expression in non-neoplastic oesophageal epithelium and squamous cell carcinoma of the oesophagus. *Virchows Arch*. 1995;426:345–349.
20. Cintorino M, Tripod SA, Santopietro R, et al. Cytokeratin expression patterns as an indicator of tumor progression in oesophageal squamous cell carcinoma. *Anticancer Res*. 2001;21:4195–4202.
21. Xue L, Hu N, Song Y, et al. Tissue microarray analysis reveals a tight correlation between protein expression and progression of esophageal squamous cell carcinoma. *BMC Cancer*. 2006;6:296.
22. Moll R, Divo M, Langein L. The human keratins: biology and pathology. *Histochem Cell Biol*. 2008;129:705–733.