“Contribution of Alterations In Putative Susceptibility Genes And Genomic Imbalances In The Occurrence Of Breast Cancer In Northeast Indian Population”

THESIS

Submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

Mishi Kaushal

Under the Supervision of

Dr. Sunita Saxena

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

PILANI (RAJASTHAN) INDIA

2013

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

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

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE

PILANI (RAJASTHAN)

CERTIFICATE

This is to certify that the thesis entitled **"Contribution of Alterations In Putative Susceptibility Genes And Genomic Imbalances In The Occurrence Of Breast Cancer In Northeast Indian Population"** and submitted **by Mishi Kaushal**, ID No. **2007PHXF438P** for award of Ph.D. Degree of the Institute, embodies original work done by her under my supervision.

Saxena 19.8.13

DR. SUNITA SAXENA

Supervisor

निदेशक / Director
राष्ट्रीय विकृति विज्ञान संस्थान - भा.आ.अं.प.
National Institute of Pathology - ICMR
सफदरजंग अस्पताल परिसर
Safdarjung Hospital Campus
नई दिल्ली-110029 / New Delhi-110029

Saxena 19.8.13

DR. SUNITA SAXENA

Director

निदेशक / Director
राष्ट्रीय विकृति विज्ञान संस्थान - भा.आ.अं.प.
National Institute of Pathology - ICMR
सफदरजंग अस्पताल परिसर
Safdarjung Hospital Campus
नई दिल्ली-110029 / New Delhi-110029

Date: **19th August 2013**

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

Dedication

Dedicated to my beloved Parents, Avyaan and Satvik

♣*ACKNOWLEDGEMENT*♣

"What we are is God's gift to us. What we become is our gift to God." Thank you God for the Blessings you bestowed upon me to do this work.

This thesis is the end of my journey in obtaining my Ph.D. I have not traveled in vacuum in this journey. This thesis has been kept on track and been seen through to completion with the support and encouragement of numerous people including my well wishers, my friends, colleagues and various institutions. At the end of my thesis I would like to thank all those people who made this thesis possible and an unforgettable experience for me. At the end of my thesis, it is a pleasant task to express my thanks to all those who contributed in many ways to the success of this study and made it an unforgettable experience for me.

I am immensely pleased to place on record my profound gratitude and heartfelt thanks to my supervisor ***Dr Sunita Saxena*** *whose expertise, understanding, and patience, added considerably to my experience. I appreciate her vast knowledge and skill in many areas. To work with you has been a real pleasure to me, with heaps of fun and excitement. You have been a steady influence throughout my Ph.D. career; you have oriented and supported me with promptness and care, and have always been patient and encouraging in times of new ideas and difficulties; you have listened to my ideas and discussions with you frequently led to key insights. Your ability to select and to approach compelling research problems, your high scientific standards, and your hard work set an example.*

I am very grateful to my Co supervisor, ***Dr Sujala Kapur*** *for insightful comments both in my work and in this thesis, for her support, and*

for many motivating discussions. Dr. Kapur, thank you for your helpful, wise guidance and incessant support throughout the study.

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

I am most grateful to all my collaborators. I wish to thank ***Dr. Amal Kataki****,* ***Dr Jagannath Sarma****,* ***Dr. Eric Zomawia*** *and* ***Dr Yogesh Verma*** *for their valuable contribution in sample collection and for their assistance with the clinical information. I'd also like to acknowledge all the men and women who participated in this study. I would like to thank all participants in this study, who by giving samples and personal information, have made it to perform the analysis and conclusion in this work.*

In addition, I have been very privileged to get to know many other great people who became friends over the last several years. I learned a lot from you about life, research, how to tackle new problems and how to develop techniques to solve them. My special gratitude and regard to ***Dr AK Mishra*** *whose statistical understanding is tremendous and whose scientific outlook inspired me a lot. I enjoyed our discussions and talks on the statistical analysis very much and I will never forget that you always had a few minutes for me even if you were very busy. My work has greatly benefited from suggestions and kind encouragement from* ***Dr Usha***

Agarwal. *Your technical depth and attention to detail and important suggestions while discussing my manuscripts have been a great help. I also thank* ***Dr LC Singh*** *on his help during the DHPLC experiments.*

It's been a rollercoaster ride with ***Rakhshan Ihsan****, and* ***Pradeep Singh Chauhan*** *my mates and dear friends since master days without whom I would not be penning these words. The PhD. Journey would always remain memorable because of the fun times we had in the lab and on our conferences trip. Thanks for the time I spent with you, thanks for all nice walks, chats, tea breaks and for helping me. It has been a joy to face the challenges of both the clinical and scientific work, and to share the various moments of enjoyable free time with you. You made my stay and work so enjoyable.*

My work has greatly benefited from suggestions and kind encouragement from ***Anand Dixit*** *and* ***Dr. Anil Tyagi****. I was very lucky to have crossed paths with* ***Dr Zeeba, Dr Raju, Dr Abhilasha*** *and* ***Indranil****. Thank you for your support and encouragement. I would also like to thank* ***Dr. Anand Verma*** *and* ***Dr Bharat Bhushan****, for their valuable suggestion and inputs.*

I am indebted to my many student colleagues for providing a stimulating and fun filled environment. I would also like to thank my colleagues ***Dhirender Singh Yadav, Abha Soni, Thoudam Regina Devi*** *for their encouragement and instructive suggestions. It was a pleasure and privilege working with them. I also warmly thank to my colleagues* ***Shanti, Asheema, Nitu, Shobhit, Ashish, Shubham, Ira, Meena*** *and* ***Shrestha.***

I also thank to ***Mr. Mohan****. for his assistance and help during my research work. I would also like to thank* ***Mrs Vansamma ji****, for their kind help during these years.*

I would like to say thank you to those people as well who are not members of the Tumour Biology but supported me during my Ph.D. studies: ***Vikas sir, Tanvi maam, Ashok sir, Harsh sir, Rajneesh sir,*** *and* ***Hemchandra jha ji.***

I would also like to thank ***Mrs Sharda*** *(Administrative Officer) for her help in various matters. I am also thankful to other staff of Office (****ken ji****), Store Section (****Subash ji****), Library (****Anita madam****) and establishment section (****Yogender ji****) for their kind help and support.*

I wish to thank all the students and staff that have passed through the Institute and friends from the other walks of life- thanks for laughs and keeping me relatively "sane". I would also like I expand my thanks to my "Hamdard" Friends- ***Dr. Shahnawaz Khan*** *and* ***Dr Abhay Singh***

*Finally, to the ones I hold dearest to my heart- my parents (****mamma, papa****) for everything, for believing in me when I myself could not and for all their sacrifices. I hope I have made them proud. Special thanks to* ***Dr. Urmila Vij (Masi)*** *without whose motivation and encouragement I would not have considered a career in research.* ***Madan Uncle*** *and* ***Priyanka*** *thank you for a huge source of strength. Immense gratitude to my* ***mom****,* ***dad (in-laws)*** *and* ***Ira*** *for giving me support, liberty and encouragement to do my research work. Huge hug and love to my darling husband* ***Satvik Wasson*** *for being there through all my failures, giving me strength and pushing me to excellence. Love and thanks to my sweetheart* ***Avyaan*** *for making my lab and thesis writing days memorable. I held have you in my arms and writen my thesis.* ***Reshmi,*** *thank you for just being there and making sure that I don't go insane.*

Finally, I would like to say to myself! Bravo!!

Mishi Kaushal

Abstract

Abstract

Breast cancer is the most common cancer in women worldwide representing over 20% of all malignancies. Differences in breast cancer incidence rates suggest that environmental factors like tobacco exposure influence breast cancer risk significantly. The association of tobacco exposures to breast cancer remains controversial. Human carcinogens and metabolites of cigarette smoke have been found in the breast fluid of smokers. However, smoking also has anti-estrogenic effects that could lower breast cancer risk. Tobacco consumption in the Northeast region of India is extensive and is reported to be different from rest of the India. The habit of chewing betel quid, containing fresh betel nut, slaked lime and tobacco wrapped in betel leaf is widespread in Mizoram. Manju Rani et al have documented that women from Northeast India have the highest tobacco chewing prevalence as compared to rest of the Indian women.

DNA damage repair and cell cycle check point are two primary defense mechanisms against mutagenic exposure. CYP, GST, BRCA2, RAD51, P53 and CCND1 are among the genes that impact detoxification, DNA adduct formation, DNA repair, while p53 and CCND1 are regulators of cell proliferation. The tobacco carcinogens furthermore may append tumorigenesis by causing genomic DNA copy number alterations initiating breast carcinogenesis. The proposed study has been undertaken, to investigate polymorphisms in genes that impact detoxification, DNA repair, and cell cycle that may modify breast cancer

susceptibility following tobacco exposure along with its effect on genomic copy number alterations in the Northeast population of India.

Polymorphisms in xenobiotic and estrogen synthesizing genes (GSTT1, GSTM1, GSTP1, CYP17 and P53) were analyzed with various environmental factors to identify their association with breast cancer risk. Betel quid chewing was identified as the single main risk factor and women with betelquid chewing history had five times the risk of developing breast cancer. GSTT1 null and GSTM1 null genotypes conferred 41 per cent less and 55 per cent less reduced risk to breast cancer respectively, suggesting their protective role.

Betel quid consumption can cause formation of reactive oxygen species and DNA adducts leading to DNA damage. Therefore individuals with defects in DNA repair and cell cycle checkpoint could be more susceptible to breast cancer. Hence, a multigenic approach was applied to identify the association of environmental risk factors and polymorphisms in DNA repair and cell cycle genes (BRCA2, RAD51, p53 and CCND1) with the risk of breast cancer. Betel quid chewing was identified as the predominant risk factor by three advanced analytical approaches viz logistic regression, MDR and CART as reported earlier. Alcohol consumption was also found significantly associated with increased risk for breast cancer.

Homozygous AA genotype of CCND1 gene conferred significant protection in both NBQC and BQC subsets. Homozygous PP genotype of TP53 gene conferred significant protection in NBQC subset whereas the C allele of

RAD51 gene was significantly over-represented in the BQC subset along with mutations in the BRCA2 gene seen in two samples. NBQC showed a best 4 locus models with TBA 0.6765 (0.005) and CVC of 10/10 in MDR analysis. No interaction models were obtained for the BQC subset. Interaction dendogram showed a large part of interaction between TP53 and RAD51 (1.32%) followed by an independent effect of CCND1 (1.89%) in NBQC. Small percentages of the entropy in case–control status explained by TP53 (0.64%), or EX2BRCA2 (0.11%) considered independently, but a large percentage of entropy explained by the interaction between these two loci (1.02%) were also found in NBQC.

Anticipating BQC and NBQC breast cancer subsets to have distinct carcinogenesis it was imperative to investigate genomic alterations. Therefore, a whole genome approach was used to identify copy number gain and loss regions amongst breast cancer patients subsets (exposed and non exposed). Along with the high throughput scanning of the whole genome data mining and computation analysis was also done to explore possible networks and pathways. BQC tumors showed significantly higher total number of alterations, as compared with NBQC tumors. Incidence of gain in fragile sites in BQC tumors was significantly higher as compared with NBQC tumors. Two chromosomal regions (7q33 and 21q22.13) were significantly associated with BQC tumors while two regions (19p13.3-19p12 and 20q11.22) were significantly associated with NBQC tumors. GO terms oxidoreductase and aldo-keto reductase activity in BQC tumors in contrast to G-protein coupled receptor protein signaling pathway and cell surface

receptor linked signal transduction in NBQC tumors were enriched in DAVID. One network “Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism” including genes AKR1B1, AKR1B10, ETS2 etc in BQC and two networks “Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry” and “Cellular Development, Embryonic Development, Organismal Development” including genes RPN2, EMR3, VAV1, NNAT and MUC16 etc were seen in NBQC. Common alterations (>30%) were seen in 27 regions. Three networks were significant in common regions with key roles of PTK2, RPN2, EMR3, VAV1, NNAT, MUC16, MYC and YWHAZ genes. These data show that breast cancer arising by environmental carcinogens exemplifies genetic alterations differing from those observed in the non exposed ones. A number of genetic changes are shared in both tumor groups considered as crucial in breast cancer progression. Biological information obtained from betel quid exposed breast cancer subset is valuable. This subgroup is frequent in the North East Indian population as most of the women in this area are usually chewers. Given a unique set of underlying genomic changes, distinct approaches to treatment may be appropriate for this patient population and others where this habit is highly.

~TABLE OF CONTENTS~

~ List of Tables ~

~ List of Figures ~

Figure 6.4B: NBQC Networks: Cellular Development, Embryonic Development, Organismal Development (Network 2).

Figure 6.5A: Common Networks: Cellular Movement, Connective Tissue Development and Function, Cellular Assembly and Organization (Network 1)

Figure 6.5B: Common Networks: Cell-To-Cell Signaling and Interaction, Tissue Development, Organismal Injury and Abnormalities (Network 2)

Figure 6.5C: Common Networks: Cell Morphology, Cellular Assembly and Organization, Cellular Compromise (Network 3).

Figure 6.6: Cancer related Canonical pathways enriched from common genes.

Abbreviations

IP	-	Immuno pathway.
θ	-	Theta
π	-	Pi
μ	-	micro
AARs	-	Age adjusted rates
BQ	-	Betel Quid
BMI	-	Body mass index
Bp	-	base pair
BQC	-	Betel quid chewers
CART	-	Classification and Regression Trees
CI	-	Confidence Interval
CONC.	-	Concentration
CVC	-	Cross validation consistency
CYP	-	cytochrome P450
DAVID	-	Database for Annotation, Visualization and Integrated Discovery) Functional Annotation Tool and Database
DHPLC	-	Denaturing high performance liquid chromatography
DNA	-	Deoxyribonucleic acid
(DSB)	-	DNA double-strand breaks
DNA RP	-	DNA repair pathway. pathway.
EDTA	-	Ethylenediaminetetraacetic acid
EHBCCG	-	Endogenous Hormones and Breast Cancer Collaborative Group
EP	-	Estrogen pathway.
EPIC	-	European Prospective Investigation into Cancer and Nutrition
EtBr	-	Ethidium Bromide
FPRP	-	False Positive Report Probability
FDR	-	False discovery rates
FDRs	-	First-degree relatives
γ	-	Gamma
GST-pi	-	Glutathione S-transferase-pi
HWE	-	Hardy–Weinberg equilibrium
HMM	-	Hidden Markov Model
HRR	-	homologous recombination repair
HT	-	hormone therapy

IMP	-	iron metabolism
IPA	-	Ingenuity Pathway Analysis
kg/m	-	Kilogram per
LOH	-	Loss of heterozygosity
MP	-	Metabolic pathway.
MDR	-	Multifactor Dimensionality Reduction
NaCl	-	Sodium Chloride
NBQC	-	Non betel quid chewers
nm	-	Nano meter
O.R	-	Odds Ratio
OD	-	Optical Density
[P_h]	-	*P* value of heterogeneity test
PAHs	-	Polycyclic aromatic hydrocarbons
PBCR	-	Population based cancer registry
PCR	-	Polymerase Chain Reaction
RBC	-	Red Blood Cells
ROS	-	reactive oxygen species
RFLP	-	Restriction fragment length polymorphism
RNA	-	Ribonucleic acid.
RT	-	Room Temperature
SDs	-	Standard Deviation
SDS	-	Sodium dodecyl Sulphate
SNP	-	Single nucleiotide polymorphism
ssDNA	-	Single stranded DNA
TBA	-	Testing Balance Accuracy
TDW	-	Triple distilled water
TEAA	-	triethylammonium acetate
TSG	-	tumour suppressor gene.
UTR	-	Untranslated region
α	-	Alpha
β	-	Beta
µg	-	Microgram
µl	-	Microlitre

Chapter 1

Introduction

Chapter 1

Introduction

Breast cancer is the most common female cancer and one of the leading causes of death among women worldwide. With changing disease patterns, breast cancer is also now one of the most common cancers amongst Asian women in whom low incidence of breast cancer had been reported earlier (Tan, S. M. et al. 2007). The several fold difference in incidence rates between different geographical regions suggest that environmental factors besides genetic factors influence breast cancer risk significantly. Among the identified environmental risk factors in general for cancers, tobacco exposure has been reported as the leading preventable risk factor (Terry, P. D. and T. E. Rohan 2002). Despite considerable research, however, the relationship of tobacco exposures to breast cancer incidence remains controversial. Tobacco contains a number of human carcinogens and metabolites of cigarette smoke have been detected in the breast fluid of smokers also. However, smoking also has anti-estrogenic effects that could, paradoxically, act to lower breast cancer risk. The inconsistencies in the literature may be due to heterogeneity of risk according to timing of exposure, age of diagnosis or genetic susceptibilities (Reynolds, P. et al. 2004).

In the Northeast population of India the mean age for tobacco use initiation is 18.5 years and the prevalence of tobacco use is estimated as 41%

that includes a large number of female chewers too apart from male smokers. Method and form of tobacco consumption in this region is reported to be different from the rest of India (Chaturvedi, H. K. et al. 2003). Tobacco smoking rate in Mizoram is very high among adults. A peculiar habit of using ‘‘tuibur’’ (tobacco smoke–infused water) has also been observed in Mizoram. The habit of chewing betel quid, containing fresh betel nut, slaked lime and tobacco wrapped in betel leaf is also widespread in Mizoram. Dried tobacco mixed with lime processed with tips of thumb on the palm of other hand into a powder that is place near the gum known locally as ‘‘Khaini’’ also chewed in Mizoram (Phukan, R. K., et al. (005). In Assam ‘raw’ (‘green’), ‘ripe’ (‘red’) and ‘fermented’ (‘underground’,‘ processed’) betel nuts are all chewed. The latter, known locally as ‘Bura Tamul’, is prepared in a 4–5 foot hole in the ground where ripe betel nuts are left for 3–4 months covered with bark from the betel tree, cow dung and soil. During the period of fermentation the outer fibrous shell of the nuts decays. Chopped or crushed nuts at the different stages of ripening or decay are wrapped in betel leaf and are chewed with or without tobacco. ‘Dhapat’, dried tobacco leaf that may be treated with lime (calcium oxide), is sometimes added to the betel nut in the quid while a mixture of finely cut and dried, ‘raw’ or ‘ripe’ betel nut (‘Supari’) and finely cut, scented tobacco (‘Zarda’) is also chewed. In Assam a larger proportion of betel nut is included in the quid and fewer leaves than in the ‘pan’ which is chewed in Bombay and which includes only a very small quantity of betel nut that

is always processed ('fermented'). As in Assam, the Bombay quid may also include tobacco. Dried tobacco chewed alone in Assam is known locally as 'Chadha'. Whatever the composition of the quids, they are usually retained in the mouth for about 20 to 25 minutes but occasionally the mixture may be retained in the mandibular groove during sleep (Phukan, R. K. et al. 2001). Manju Rani *et al* have documented that women from Northeast India have the highest tobacco chewing prevalence as compared to rest of the Indian women (Rani, M. et al. 2003). The population based cancer registries (PBCR) from different parts of India have documented (2006-2008) high age adjusted rates (AARs) of all cancers in females in the Northeast Indian states as compared to rest of the India (*figure1.1*).

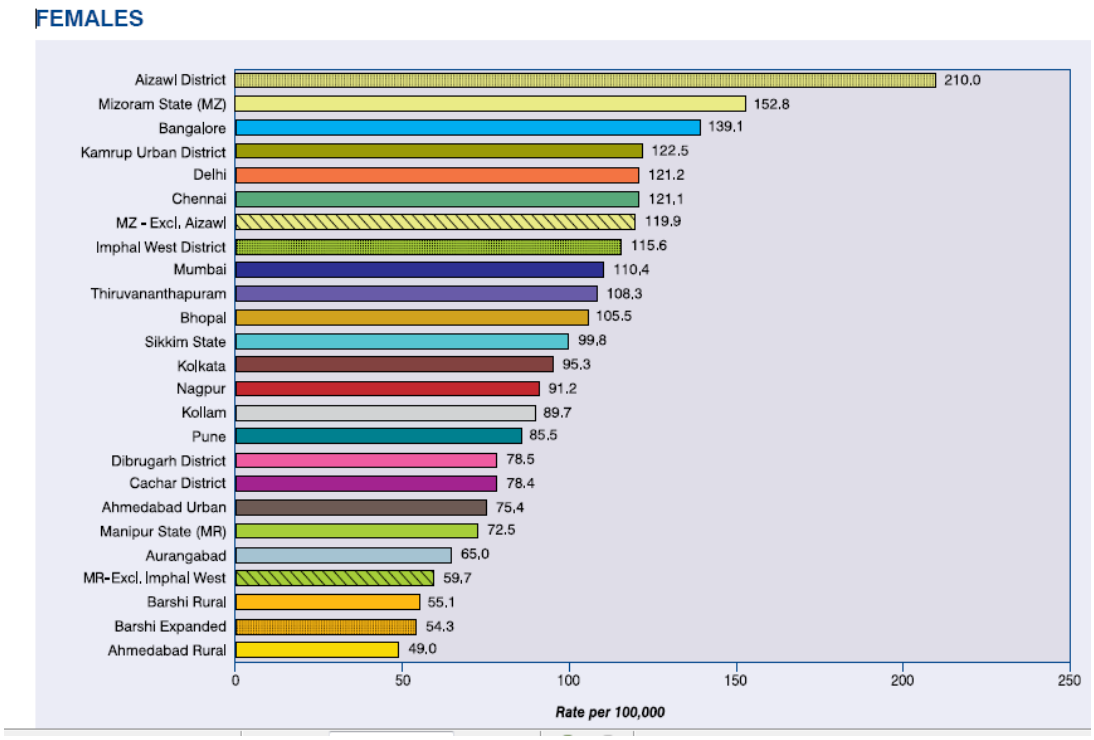


Figure 1.1: Comparison of Age incidence rates (AARs) of all PBCRs in all cancers in females

Large epidemiological studies have shown an association of tobacco with several other cancer sites, either because of direct contact of the tobacco products with the specific organ or because the carcinogenic products reach the organ through the blood stream. The long list of tobacco-associated cancers includes nose, oral cavity, oropharynx, hypopharynx, larynx, esophagus, pancreas, bladder, kidney, stomach, liver, colon, cervix and myeloid leukemia. Genetic pathways of tobacco metabolism polycyclic aromatic hydrocarbons

(PAHs), N-nitrosamines and aromatic amines are metabolized by a two phase process. Phase I involves the activation of the carcinogen by enzymes encoded by the cytochrome P450 (CYP) gene superfamily. These enzymes are involved in the oxidative metabolism of several exogenous compounds along with drugs and endogenous hormones. During the Phase II process, carcinogens are transformed into hydrophilic elements to facilitate excretion. Glutathione S-transferases are mostly responsible for this process. This multigene superfamily includes three classes μ, θ and π and detoxifies carcinogens from cigarette smoke as well as from other sources. DNA damage is usually repaired by a series of genes specialized in this activity. Several polymorphisms in DNA repair genes also have been identified, but their impact on repair phenotype and cancer susceptibility remains uncertain. Since Phase II enzymes conjugate these compounds and make them suitable for excretion, it is reasonable to think that the carcinogenic effect of tobacco compounds should be measured as the result of the phase II enzymes actions (Taioli, E. 2008).

When tobacco carcinogens bind to epithelial cell DNA they form DNA adducts (DNA-xenoligand complexes). Adduct formation initiates tumorigenesis through a series of DNA repair processes that may lead to mutations in genes that initiate or facilitate tumor growth. When adducts are detected they are excised and the excised DNA segment is repaired. DNA repair genes code for proteins that form complexes with DNA to repair gaps left during DNA adduct

excision repair. Therefore, gene polymorphisms can lead to DNA repair deficiency, gene mutations, and ultimately tumorigenesis. Alternatively, the TP53 tumor suppressor gene codes for the nuclear phosphoprotein TP53, a transcription factor regulating cell replication and apoptosis. Therefore, TP53 protein inactivation via nucleotide base transversions, transitions, or deletions may negatively impact appropriate cell replication and death, two processes necessary for tumorigenesis. CYP, GST, BRCA2, RAD51, TP53 and CCND1 are among the genes that impact detoxification, DNA adduct formation, DNA repair, while TP53 and CCND1 are regulators of cell proliferation. Therefore, polymorphisms in any of these genes may modify breast cancer susceptibility following tobacco exposure, but each gene may do so by separate pathways that invokes a myriad of cellular processes (Brownson, R. C. et al. (2002).

Tobacco carcinogens are known to induce anaphase bridges via DNA double stranded breaks causing genomic imbalances in human cells. Regions like 7p11.2 (epidermal growth factor receptor) and 11q13.3 (cyclin D1) playing a role in the pathogenesis of tobacco-related human squamous cell carcinoma has been identified by SNP array. Examination of the copy number changes due to effect of tobacco carcinogens in mouse lung adenocarcinomas revealed gain on chromosomes 6 and 8, and losses on chromosomes 11 and 14. Significant gains of 1p and 3q have been observed in patients with a history of tobacco exposure in head and neck squamous cell carcinomas. In addition, Benzo(*a*)pyrene [B(a)P]

diolepoxide (BPDE), a carcinogen present in cigarette smoke, induces chromosomal 9p21 aberrations seen to be significantly higher in peripheral blood lymphocytes of bladder cancer cases than that of controls. Furthermore, 5q22.2~q22.3 (LOX gene) allelic imbalance has been observed significantly higher among smokers than nonsmokers in clear cell renal carcinomas indicating that tobacco may cause genetic alterations. In vitro and in vivo experiments have shown that betel quid consumption can also cause micronuclei and DNA adducts formation, chromosomal aberrations, allelic imbalances and sister chromatid exchange in oral mucosa cells. Carcinogens in betel quid leads to accumulation of genetic alterations at 3q26.3 locus particularly in recurrent oral tumors along with accelerating tumor migration by stimulating MMP-8 expression through MEK pathway (Kaushal, M. et al. 2012).

High incidence of tobacco associated cancer like esophageal, oral, lung and gastric have been reported in Northeast states along with a high incidence of breast cancer. Hence this study had been undertaken to investigate the genetic and environmental risk factors and molecular mechanism associated with high incidence of breast cancer in this region.

Gaps in Existing Research

Approximately 30% of all cancer deaths in developed countries are caused by tobacco consumption. This includes 87%of lung cancer, 60%of upper aerodigestive cancer, and 8% of other cancer. In the latter group, smoking is a recognized cause of cancers of the pancreas, bladder, kidney, liver, and colon. As these sites are removed from direct contact with tobacco, there are clearly systemic effects that result in cancer. It is possible that tobacco could cause breast cancer as well, but epidemiologic studies are conflicting.

BRCA1 and BRCA2 are tumor suppressor genes, familial mutations in which account for ~5% of breast cancer cases. Therefore, low penetrance genes may be associated with a small increased risk for breast cancer in an individual but the attributable risk in the population as a whole is likely to be higher than for rare, high-penetrance susceptibility genes. There have been numerous studies from India but none of them have been focused on a subset of population where there is an extensive usage of tobacco as an environmental toxicant. The Northeast Indian population is very homogenous with a very extensive and distinct consumption of tobacco. Moreover, apart from males a large part of the female population is also under the influence of tobacco chewing.

The AARs of breast cancer in different states of Northeast region have been constantly increasing over the years and are now almost comparable to the metropolitan cities of India. There is compelling evidence that the characteristic

environmental factors such as tobacco and alcohol use and genetic susceptibility could be most significant driving factors for rise in incidence in these states. This impelled us to investigate the etiology and molecular carcinogenesis of breast cancer in such a homogeneous population.

Genes that impact detoxification, DNA adduct formation, DNA repair, and cell cycle are prime candidates for susceptibility studies and polymorphisms in any of these genes may modify breast cancer susceptibility following tobacco exposure, but each gene may do so by separate pathways that invokes a myriad of cellular processes. Moreover, the effect of environmental toxicant might also cause genomic copy number alterations which need to be further investigated. Thus, in the proposed study polymorphisms in putative xenobiotic metabolizing genes and DNA repair cell cycle genes have been analyzed for assessment of breast cancer risk. The impact of genomic alteration in breast cancer patients exposed to environmental carcinogen (betel quid) had also been investigated.

Scope of the study

Exploring the environmental and genotypic factors in the Northeast population of India will facilitate in identifying the risk factors for breast cancer susceptibility. The study will append new information regarding environmental carcinogen (like tobacco) induced breast carcinogenesis. As this population is ethnically different from rest of the India, the genotypic data of xenobiotic metabolizing genes, DNA damage, repair and cell cycle genes generated will give new insights with underlining principles of breast

tumorgenesis in this population. Analysis of genomic alteration in breast cancer between two groups (exposed and non exposed to environmental carcinogen) will help in understanding the difference in etiologies and help formulating distinct approaches of treatment and prevention.

Chapter 2

Aims and Objectives

Chapter 2

Aim and Objectives

3.1: Aim:

To investigate the role of environmental and genotypic risk factors along with gene-environmental interactions for high incidence of breast cancer in Northeast Indian population.

3.2: Objective

1. **Polymorphism in putative xenobiotic metabolizing GST genes and CYP17 and assessment of breast cancer risk**
   Polymorphisms in xenobiotic and estrogen synthesizing genes (GSTT1, GSTM1, GSTP1, CYP17 and P53) were analyzed with various environmental factors to identify there association with the risk of breast cancer. MDR analysis and Logistic regression method were used to explore high order gene-gene and gene-environment interactions.

2. **Polymorphism in putative DNA repair genes BRCA2, RAD51, P53 and cell cycle gene CCND1 and assessment of breast cancer risk**
   A multigenic approach was applied to identify the association of environmental risk factors and polymorphisms in DNA repair and cell cycle genes (BRCA2, RAD51, p53 and CCND1) with the risk of breast cancer. CART and MDR analysis method were used to explore high order gene-gene interactions.

3. **To screen and differentiate Genomic alteration in breast cancer patients exposed to environmental carcinogen (betel quid)**

   A whole genome approach was used to identify copy number gain and loss regions amongst breast cancer patient. Along with the high throughput scanning of the whole genome data mining and computation analysis was also done to explore possible networks and pathways.

Chapter 3

Review of Literature

Chapter 3

Review of literature

Breast cancer is a genetically and clinically heterogeneous disease arising from epithelium lining of duct, lobules or stroma of the breast. Breast cancer is sub-divided into two major categories, *in situ* disease in the form of ductal carcinoma in situ (DCIS), or invasive cancer. Ductal carcinoma (IDC) is the most common subtype accounting for 70–80% of all invasive lesions (Malhotra, G. K., X et al. 2010) compared to invasive lobular carcinoma which comprises of 5%–15% of the group (Badve, S. et al. 2011). Both are heterogeneous processes with very variable appearances, biology and clinical behaviour. DCIS is predominantly detected by breast screening as micro calcifications on mammography. DCIS grows within a single duct system of the breast but it can vary in size and is sometimes extensive. However, DCIS, by definition, has not spread outside the boundaries of the normal structures of the breast and therefore cannot have metastasized (Yarnold, J. 2009). High grade DCIS is a more inherently high-risk disease in terms of progression into invasive breast cancer and development of local recurrence after surgical excision. DCIS has been classified according to architectural pattern (solid, cribriform, papillary, and micropapillary), tumor grade (high, intermediate, and low grade), and the presence or absence of comedo histology.

Unlike DCIS, invasive breast cancer infiltrates into the breast stroma and thus has the potential to spread to lympho-vascular spaces and to metastasize. Not all invasive breast cancers are the same; some are more aggressive and some may spread earlier to distant sites. There are a variety of methods for classifying invasive breast cancer; most are based on the architectural microscopic pattern and nature of the cancerous cells. The most important of these is histological grading, which identifies tumours as being of histological grade 1 (least aggressive), grade 2 or grade 3 (most aggressive) (Virnig, B. A. et al. 2009).(*figure 2.1*)

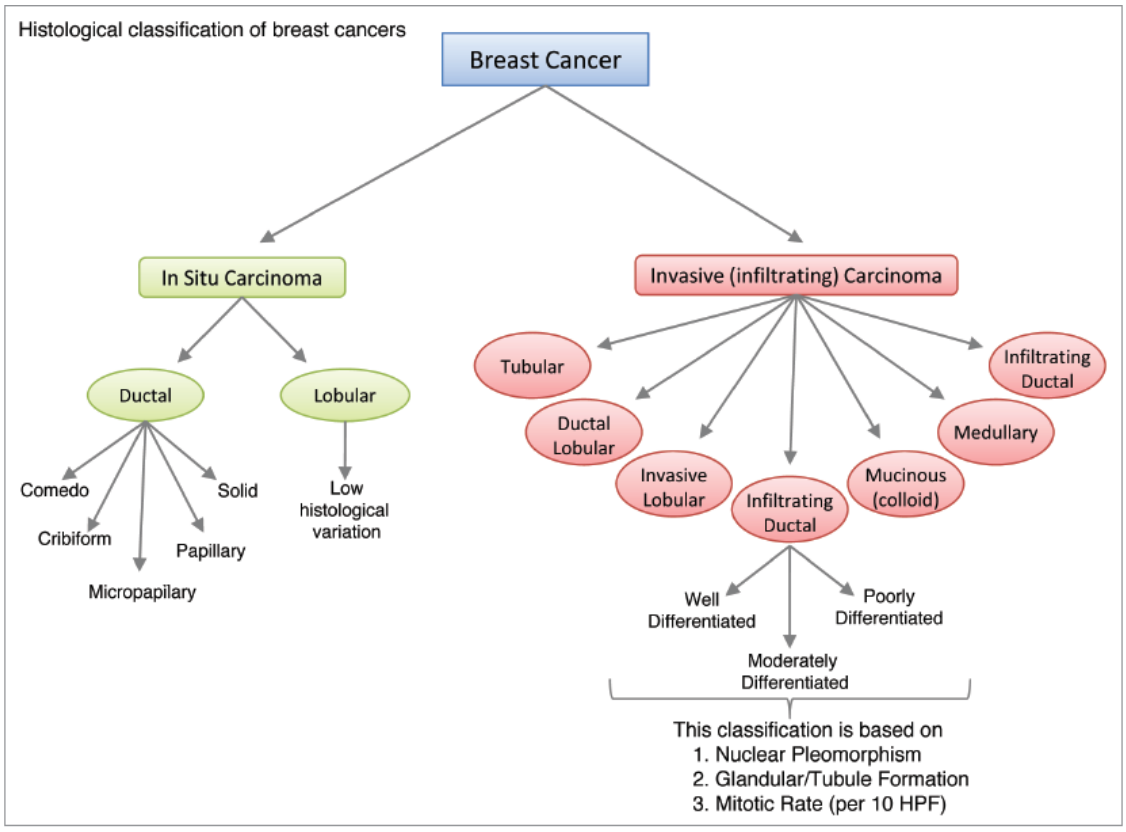


Figure2.1: Histological classification of breast cancer subtypes

2.1. Epidemiology

2.1.1. Worldwide

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in female worldwide, accounting for 23% (1.38 million) of the total new cancer cases and 14% (458,400) of the total cancer deaths in 2008. About half the breast cancer cases and 60% of the deaths are estimated to occur in economically developing countries. In general, incidence rates are high in Western and Northern Europe, Australia/New Zealand, and North America; intermediate in South America, the Caribbean, and Northern Africa; and low in sub-Saharan Africa and Asia (*figure 2.2*).

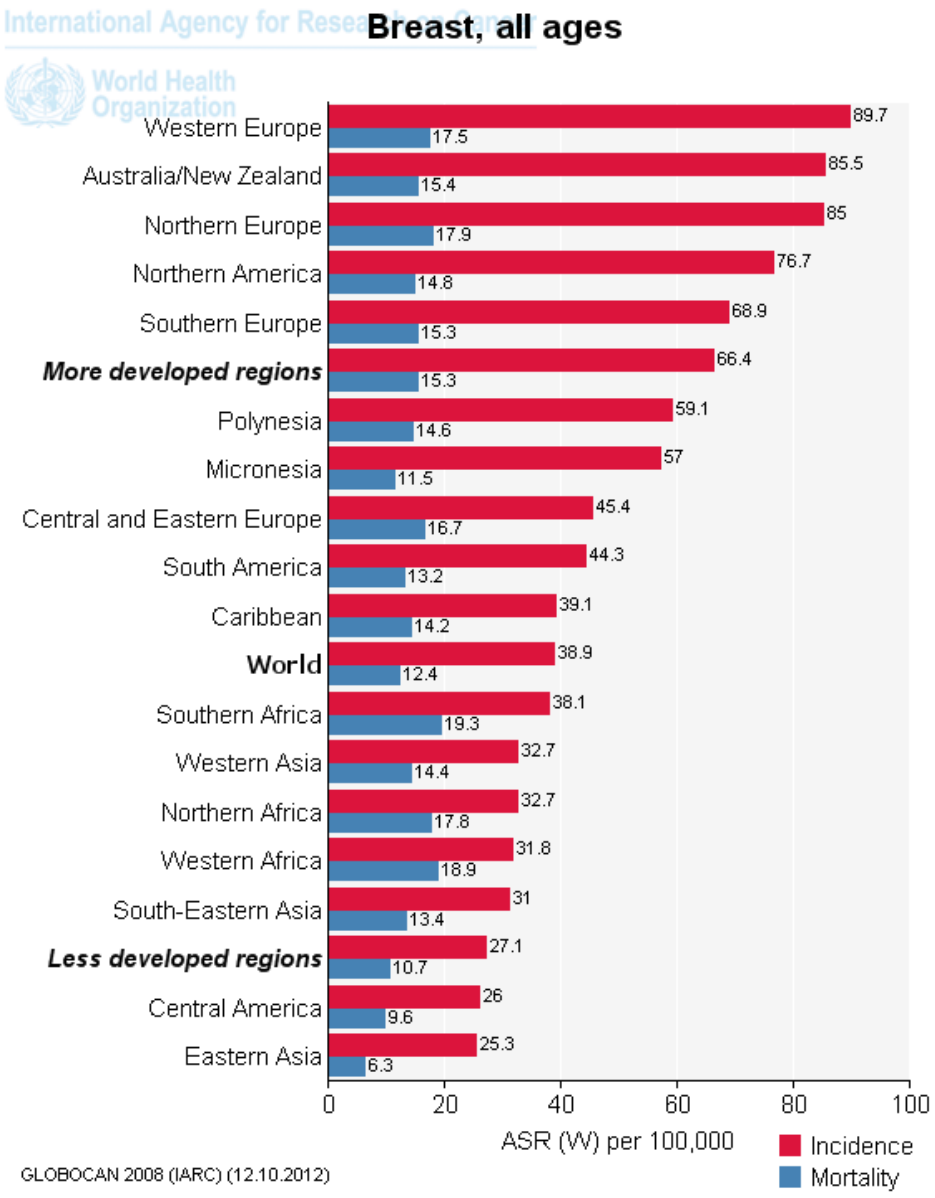


Figure 2.2: Worldwide Breast cancer Incidence and Mortality rates

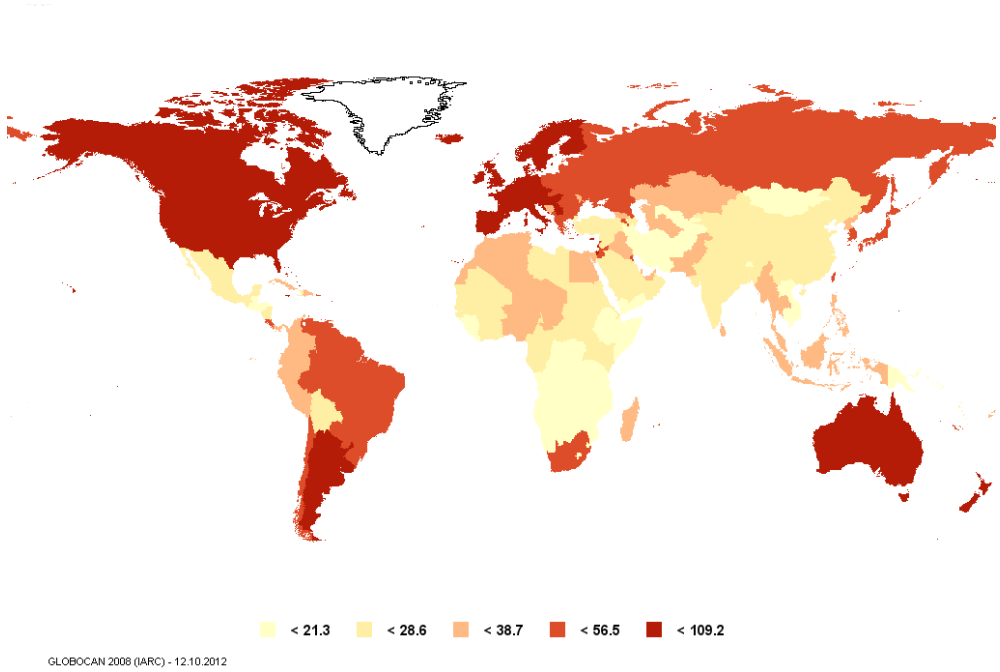


Figure 2.3: Estimated age-standardized incidence rate per 100,000 Breast, all ages

In contrast, breast cancer death rates have been decreasing in North America and several European countries over the past 25 years, largely as a result of early detection through mammography and improved treatment. In many African and Asian countries however, including Uganda, South Korea, and India, incidence and mortality rates have been rising(*figure 2.3 and 2.4* (Jemal, A et al. 2011).

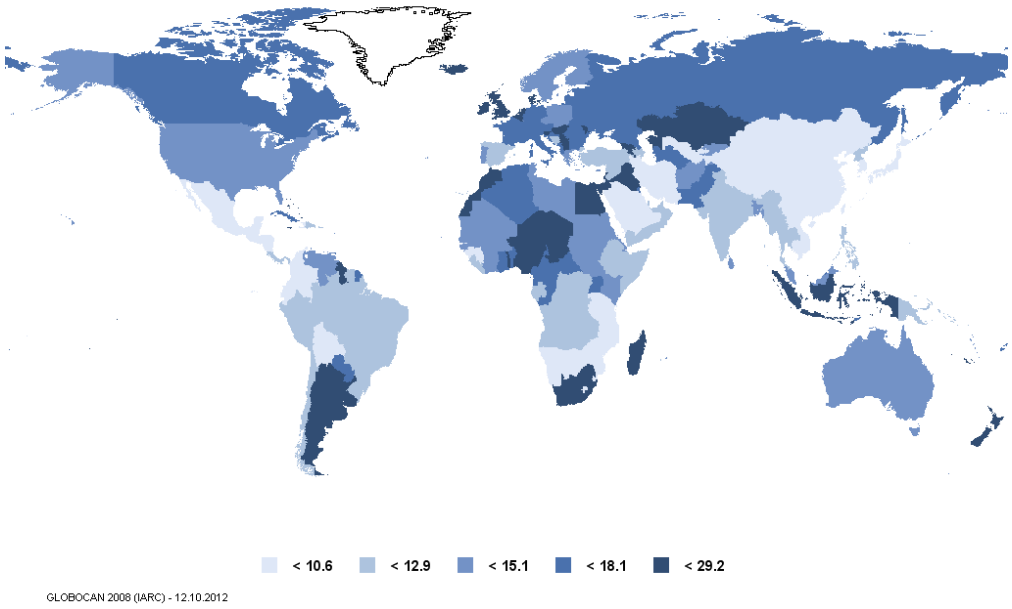


Figure 2.4: Estimated age-standardized mortality rate per 100,000 Breast, all ages

2.1.2. India

Breast cancer accounts for about one-fourth of all cancers in Indian women and about half of all cancer-related deaths. Data from all urban and rural population-based cancer registries in India suggest a rising incidence of breast cancer in India. The age-adjusted rates reported from various urban registries range from 21.6 in Ahmadabad to 36.1 in Bangalore per 100,000, (*figure 2.5*) which is about one-third the incidence reported from Western countries such as the United States (California SF: NH White, 109.6 per 100,000) and Uruguay (Montevideo, 114.9 per 100,000). The minimum age-adjusted rates reported from the rural population-based cancer registry in India are even lower (7.2 per

100,000) and somewhat similar to the incidence reported from other developing countries, such as The Gambia and Jiashan, China (NCRPINDIA 2006). The reason of increasing incidence is India's urbanization and adoption of western lifestyle and food habits. Urban Indian women get married later, have fewer children, breastfeed them less, have a more western diet and higher alcohol intake leading to obesity which increases their lifetime exposure to estrogen and therefore their risk of developing the breast cancer. Although Indian women are still less likely to get breast cancer than those in the west, they are far more likely to die of it. The most overwhelming reason for this is the poor awareness of breast cancer, reticence to see a doctor or the inability to access medical care. Another major reason is the importance still given to alternative medicine in India. Many women turn first to Ayurveda, homoeopathy, or worse, "traditional healers" (Shetty, P. 2012).

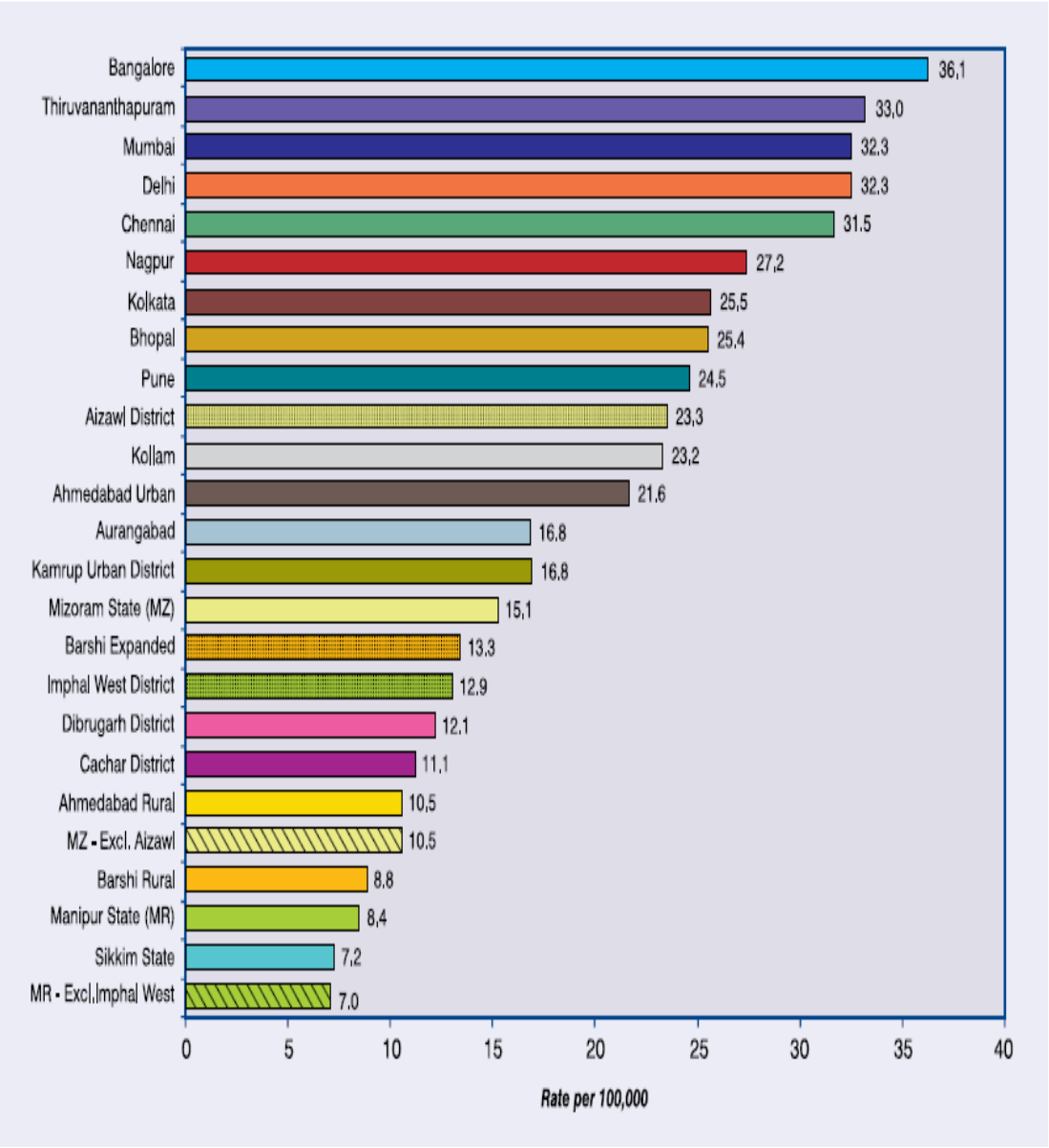


Figure 2.5: Comparison of Age Adjusted Breast cancer incidence rates (AARs) of all PBCRs

2.2. Etiology of Breast Cancer

2.2.1. Age

The incidence of breast cancer increases rapidly with age. The cumulative incidence of breast cancer among women in Europe and North America is about 2.7% by age 55, about 5.0% by age 65, and about 7.7% by age 75 (Key, T. J. et al. 2001).

2.2.2. Reproductive factors

Reproductive factors influence the risk of breast cancer. The following reproductive factors are important: age at menarche; age at first child birth; parity; age at menopause; and duration of breastfeeding.

2.2.2.1. Age at menarche

The older a woman is when she begins menstruating, the lower her risk of breast cancer. For each 1-year delay in menarche, the risk decreases by around 5%. There is also evidence that, although age at menarche is related to breast cancer risk at all ages, the effect may be stronger in younger (premenopausal) women. Other menstrual factors, such as cycle length and regularity, have not been consistently related to risk of breast cancer (Key, T. J. et al. 2001).

2.2.2.2. Age at first child birth

The younger the woman is when she begins childbearing, the lower her risk for breast cancer. The relative risk (RR) of developing breast cancer increases by 3% for each year of delay (Parkin, D. M. 2011).

2.2.2.3. Parity

Increasing parity reduces the risk of breast cancer. The higher the number of full term pregnancies, the greater the protection. Compared with nulliparous women, a woman who has at least one full-term pregnancy reduces her risk of breast cancer by around 25% and women with five or more children experience a 50% reduction in risk (Parkin, D. M. 2011).

2.2.2.4. Age at menopause

Late menopause increases the risk of breast cancer. For breast cancer, risk is doubled for a woman with menopause at 55 years compared with less than 45 years. For each year that the menopause is delayed, there is an approximate 3% increase in breast cancer risk (Parkin, D. M. 2011).

2.2.2.5. Breastfeeding

The role of breastfeeding as a protective factor against the development of breast cancer has been reported. There is a decrease in breast cancer risk of 4.3% for every 12 months of breastfeeding (Parkin, D. M. 2011).

2.2.2.6. Circulating hormone levels

In Postmenopausal women—The body of evidence supporting a link between circulating estrogen levels and breast cancer risk for postmenopausal women are more consistent than for premenopausal women. For example, in the Endogenous Hormones and Breast Cancer Collaborative Group (EHBCCG) analysis of nine prospective studies of endogenous hormone levels and breast cancer risk, levels of multiple sex steroids –including total estradiol, free

estradiol, estrone, and estrone sulfate – were associated with increased breast cancer risk. In general, those in the highest quintile of circulating hormone levels had twice the risk compared to those in the lowest quintile. Similar results were seen within European Prospective Investigation into Cancer and Nutrition (EPIC). The Nurses' Health Study further evaluated the association by hormone receptor status of the tumor, and as expected a stronger association with circulating estrogen levels was found with ER+/PR+ tumors.

Premenopausal women

Stronger evidence exists for associations between breast cancer risk and premenopausal androgen levels. The largest prospective study to date is EPIC, which reported an association between breast cancer and premenopausal levels of testosterone, androstenedione, and dehydroepiandrosterone (DHEAS), but not for estradiol or estrone. However, another large prospective study (the Nurses' Health Study, NHS) presented slightly different results. Similarly to EPIC, the NHS investigators also observed an association with breast cancer risk and premenopausal testosterone levels. In contrast to EPIC, they also observed an increased breast cancer risk with estradiol levels. In summary, for premenopausal women, the associations with circulating androgens appears to be stronger than the data for circulating estrogens, but it is not known how much of the observed differences are due to measurement issues, since androgens can be more reliably measured and have less variation according to the menstrual cycle than estrogens (Chen, W. Y. 2008).

2.2.2.7. Oral Contraceptives

Substantial data show little association between use of oral contraceptives 10 or more years in the past and risk of breast cancer. However, in earlier reports from the prospective Nurses' Health Study and in a pooled analysis of 53,297 cases and 100,239 controls conducted in the 1970s and 1980s, a modest increase in risk was observed among women who were currently using oral contraceptives, or who had stopped using them in the preceding 10 years. However, few studies have examined the relation of newer formulations of oral contraceptives as used in the 1990s with breast cancer risk. A recent large case-control study reported an odds ratio of 0.9 [95% confidence interval (CI), 0.8-1.0] for past use of more recent oral contraceptive preparations, and no elevation in risk for current use (odds ratio, 1.0; 95% CI, 0.8-1.3). A hospital based case-control study conducted between 1993 and 2007 observed an increased odds ratio for one or more years of oral contraceptive use of 1.5 (95% CI, 1.2-1.8) (David J. Hunter et al. 2010). A meta-analysis was done by Zhu H done on 13 prospective cohort studies on oral contraceptives use and breast cancer risk from the period 1960 to 2012 involving 11,722 cases and 859,894 participants. The combined relative risk (RR) of breast cancer for ever- compared with never- oral contraceptives users was 1.08 (95% confidence interval [CI]: 0.99-1.17). Dose-response analysis based on five eligible studies showed that every ten-year' increment of oral contraceptives use was associated with a significant 14% (95% CI: 1.05-1.23) rise in breast cancer risk (Zhu, H., et al. 2012).

2.3. Life style breast cancer risk factors

2.3.1. Body mass index (BMI)

The relationship of BMI with breast cancer risk differs by menopausal status. In premenopausal women, most studies have found either no association or a weak inverse relationship between BMI and breast cancer risk, although positive associations have sometimes been observed in countries with moderate or low rates of breast cancer. In postmenopausal women, however, the risk for breast cancer clearly increases with increasing BMI. A pooled analysis of seven prospective studies has shown that the risk of developing breast cancer is approximately 30% higher among postmenopausal women with a BMI of over 31 kg/m2 compared with women with a BMI of about 20 kg/m. The effect of BMI on breast cancer risk is probably due to the effects of BMI on endogenous estradiol levels. Women with high BMI is associated with an even greater increase in concentration of free estradiol than the two-fold increase in the concentration of total estradiol (Key, T. J., et al. 2003).

2.3.2. Physical activity

In a review of 73 observational epidemiologic studies of physical activity and breast cancer risk, an average decrease in breast cancer risk of 25% when compared the most physically active to the least active women was found. Effect modification was also observed between race, family history of breast cancer, and parity subgroups with a stronger effect of physical activity observed amongst women of non-Caucasian backgrounds, without a family history of breast cancer

and who were parous. One common theme of many hypotheses explaining the relation between physical activity and breast cancer risk is a mediation of the effect through body weight. Adiposity, frequently measured in terms of BMI, is now convincingly associated with increased breast cancer risk in postmenopausal women, and weight gain and abdominal fatness are probably also causally related. In premenopausal women there is no such association; in fact an inverse relation with BMI is probable (Lynch, B. M., et al 2011).

2.3.3. Alcohol and breast cancer

An association between alcohol and breast cancer was first suggested in the early 1980s by case–control studies. More than 100 epidemiological studies on alcohol consumption and female breast cancer were published afterwards, and a positive association is now established. A significant increase of 4% in the risk of breast cancer has been seen at intakes of up to one alcoholic drink/day. Heavy alcohol consumption, defined as three or more drinks/day, is associated with an increased risk by 40–50%. This translates into up to 5% of breast cancers attributable to alcohol in northern Europe and North America for a total of approximately 50 000 alcohol-attributable cases of breast cancer worldwide. Up to 1–2% of breast cancers in Europe and North America are attributable to light drinking alone, given its larger prevalence in most female populations when compared with heavy drinking. Alcohol increases estrogen levels, and estrogens may exert its carcinogenic effect on breast tissue either via the ER or directly. Other mechanisms may include acetaldehyde, oxidative stress, epigenetic

changes due to a disturbed methyl transfer and decreased retinoic acid concentrations associated with an altered cell cycle (Seitz, H. K et al. 2012).

2.3.4. Tobacco smoking

Tobacco smoking is one of the leading preventable risk factors of cancer in respiratory and nonrespiratory sites. Tobacco smoke contains potential carcinogens, including polycyclic aromatic hydrocarbons, aromatic amines, and *N*-nitrosamines. Carcinogens in tobacco pass through the alveolar membrane and enter the bloodstream and are transported to mammary tissue through plasma lipoproteins. Metabolites of cigarette smoke have been detected in nonlactating cigarette smokers in breast fluid obtained through standard nipple aspiration techniques. Furthermore, because these breast carcinogens are lipophilic, they may be stored in breast adipose tissue and metabolized and activated by mammary epithelial cells. Conversely, smoking has been postulated to have an antiestrogenic effect, which may be associated with a lower risk of breast cancer. The antiestrogenic effect of smoking has been supported by an increased risk of osteoporosis, an early age at natural menopause, and attenuated effects of hormone therapy (HT) among smokers.

Numerous epidemiologic studies have been conducted on the association between cigarette smoking and breast cancer risk, and results from these studies have inconsistently suggested positive, inverse, or null associations. The direction and magnitude of the overall association between cigarette smoking and breast cancer may differ according to the hormonal profile and other

characteristics of the study population. Lifetime smoking exposure consists of many facets, including active and passive smoking, as well as quantity, duration, cessation, and age at initiation of smoking, which are difficult to assess accurately (Xue, F., et al. 2011).

2.4. Family History

Breast cancer tends to cluster in families; the disease is approximately twice as common among first-degree relatives of patients as among women in the general population. The higher rate of breast cancer among monozygotic twins of patients than among dizygotic twins or siblings suggests that genetic variation, rather than lifestyle or environmental factors, accounts for most of the familial clustering. Familial clustering of breast cancer occurs in specific inherited breast-cancer syndromes in which single genes confer a high risk. Several such genes, including *BRCA1*, *BRCA2*, *PTEN,* and *TP53* have been identified by means of family based linkage studies. The susceptibility alleles of these genes are rare in the general population and they account for less than 25% of the inherited component of breast cancer. Other genes that confer a risk equivalent to that of *BRCA1* and *BRCA2* are unlikely to exist, since most families with four or more cases of breast cancer can be accounted for by *BRCA1* or *BRCA2*, and extensive attempts to identify similar genes with the use of family-based linkage studies have failed (Pharoah, P. D., et al. 2007). A meta-analysis of 52 epidemiologic studies on familial breast cancer showed risk ratios increase as the number of affected first-degree relatives (FDRs) increases (risk ratios of 1.8,

2.9, and 3.9 respectively for one, two, and three or more affected FDRs compared to women without affected FDRs) (Spector, D., et al. 2009).

Many families affected by breast cancer show an excess of ovarian, colon, prostatic, and other cancers attributable to the same inherited mutation. Patients with bilateral breast cancer, those who develop a combination of breast cancer and another epithelial cancer, and women who get the disease at an early age are most likely to be carrying a genetic mutation that has predisposed them to developing breast cancer. Most breast cancers that are due to a genetic mutation occur before the age of 65, and a woman with a strong family history of breast cancer of early onset who is still unaffected at 65 has probably not inherited the genetic mutation. A woman's risk of breast cancer is two or more times greater if she has a first degree relative (mother, sister, or daughter) who developed the disease before the age of 50, and the younger the relative when she developed breast cancer the greater the risk. For example, a woman whose sister developed breast cancer aged 30-39 has a cumulative risk of 10% of developing the disease herself by age 65, but that risk is only 5% (close to the population risk) if the sister was aged 50-54 at diagnosis. The risk increases by between four and six times if two first degree relatives develop the disease. For example, a woman with two affected relatives, one who was aged under 50 at diagnosis, has a 25% chance of developing breast cancer by the age of 65 (McPherson, K et al. 2000).

2.5. Genetic factors

2.5.1. High Penetrance genes

2.5.1.1. BRCA1 and BRCA2 genes

BRCA1 and BRCA2 are tumor suppressor genes, located on the long arms of chromosomes 17q12-21 and 13q12-13 respectively account for a substantial proportion of very high risk families and familial mutations in which account for ~5% of breast cancer cases. Germ line mutations in BRCA1 that truncate or inactivate the protein lead to 80% cumulative risk of breast cancer by age 70. For germ line BRCA2 mutations, the breast cancer cumulative risk approaches 50%. Both BRCA1 and BRCA2 are involved in maintaining genome integrity at least in part by engaging in DNA repair, cell cycle checkpoint control and even the regulation of key mitotic or cell division steps (O'Donovan, P. J. et al 2010). The BRCA1 gene has been associated with more than 15 different proteins involved in transcription, plays a role in apoptosis and helps in maintaining genomic stability. Mutations of BRCA1 are scattered throughout the gene and consist of insertions, deletions, frameshifts, base substitutions and inferred regulatory mutations. In sporadic breast cancer the gene is rarely mutated, but frequently functionally impaired. The BRCA2 gene codes for proteins involved in DNA repair, cell cycle control and transcription and may have a function in terminal differentiation of breast epithelial cells. In sporadic breast cancer, mutational inactivation of BRCA2 is rare as inactivation requires both gene copies to be mutated or totally lost (Kenemans, P. et al. (2004). Certain

mutations occur at high frequency in defined populations. For instance, some 2% of Ashkenazi Jewish women carry either BRCA1 185 del AG (deletion of two base pairs in position 185), BRCA1 5382 ins C (insertion of an extra base pair at position 5382) or BRCA 6174 del T (deletion of a single base pair at position 6174), while BRCA2 999 del 5 (deletion of five base pairs at position 999) accounts for about half of all familial breast cancer in Iceland (McPherson, K et al. 2000).

Mutation analysis has also been performed in Indian women previously. In a previous study sixteen breast or breast and ovarian cancer families, 20 female patients with sporadic breast cancer regardless of age and family history, and 69 unrelated normal individuals as control of Indian origin were screened for BRCA1 and BRCA2 mutations. Twenty-one sequence variants including fifteen point mutations were identified. Five deleterious pathogenic, protein truncating frameshift and non-sense mutations were detected in exon 2 (c.187_188delAG); and exon 11 (c.3672G>T) [p.Glu1185X] of *BRCA1* and in exon 11 (c.5227dupT, c.5242dupT, c.6180dupA) of *BRCA2* (putative mutations – four novel) as well as fourteen amino acid substitutions were identified. Twelve *BRCA1* and *BRCA2* missense variants were identified as unique and novel. In the cohort of 20 sporadic female patients no mutations were found (Valarmathi, M. T.et al. 2004). In a study from our group series of 20 breast cancer patients were analyzed from North India with either family history of breast and/or ovarian cancer (2 or more affected first degree relatives) or early age of onset (< 35 years) led to

identification of two novel splice variants (331+1G>T; 4476+2T>C) in BRCA1 (10%). In addition, two BRCA2 missense variants were each identified in more than one patient (two unrelated individuals each) (Saxena, S., et al. 2002). In another study from our group, 204 breast cancer cases along with 140 age-matched controls were analyzed for BRCA1 and BRCA2 mutations, 18 genetic alterations were identified. Three deleterious frame-shift mutations (185delAG in exon 2; 4184del4 and 3596del4 in exon 11) were identified in BRCA1, along with one missense mutation (K1667R), one 5'UTR alteration (22C>G), three intronic variants (IVS10-12delG, IVS13+2T>C, IVS7+38T>C) and one silent substitution (5154C>T). Similarly three pathogenic proteintruncating mutations (6376insAA in exon 11, 8576insC in exon19, and 9999delA in exon 27), one missense mutation (A2951T), four intronic alterations (IVS2+90T>A, IVS7+75A>T, IVS8+56C>T, IVS25+58insG) and one silent substitution (1593A>G) were identified in BRCA2. Four previously reported polymorphisms (K1183R, S1613G, and M1652I in BRCA1, and 7470A>G in BRCA2) were detected in both controls and breast cancer patients. Rare BRCA1/2 sequence alterations were observed in 15 out of 105 (14.2%) early-onset cases without family history and 11.7% (4/34) breast cancer cases with family history. Of these, six were pathogenic protein truncating mutations (Saxena, S. et al. 2002).

2.5.1.2. TP53 Gene

The tumour suppressor gene TP53, on chromosome 17p13.1, is one of the most frequently mutated genes in sporadic human cancer. Most mutations

are point mutations leading to proteins defective for sequence-specific DNA binding and activation of TP53-responsive genes. In sporadic breast carcinomas the occurrence of TP53 mutations is a late event. Rarely, a TP53 mutation is associated with hereditary breast cancer, as seen with the Li-Fraumeni Syndrome (Kenemans, P. et al. (2004). Inactivating mutations in the *TP53* gene have been found in many tumour types including breast cancer. The risk of developing breast cancer before the age of 45 is 18-fold higher for affected females as compared to the general population. Germline mutations in the *TP53* gene have been estimated to account for less than 1% of breast cancer cases. However, somatic mutations in the *TP53* gene are reported in 19-57% of human breast cancers and loss of heterozygosity (LOH) is found in 30-42%. Three different *TP53* polymorphisms (in intron 3, exon 4 and intron 6) have been studied in breast cancer patients. All three polymorphisms exhibit strong linkage disequilibrium with each other (de Jong, M. M.et al. 2002).In a meta-analysis with thirty-nine published studies, including 26,041 breast cancer cases and 29,679 controls the overall results suggested that the variant genotypes were associated with a significantly reduced breast cancer risk (GC vs. GG: OR = 0.91, 95% CI: 0.83–1.00; CC/GC vs. GG: OR = 0.90, 95% CI: 0.82–0.99). In the stratified analyses, significantly decreased risks were also found among European populations (GC vs. GG: OR = 0.89, 95% CI: 0.80–0.99; CC/GC vs. GG: OR = 0.88, 95% CI: 0.80–0.98) and studies with population-based controls (GC vs. GG: OR = 0.88, 95% CI: 0.78–0.98; CC/GC vs. GG: OR = 0.87, 95% CI:

0.78–0.97) suggesting that *TP53* codon 72 polymorphism may contribute to susceptibility to breast cancer, especially in Europeans (Zhang, Z.et al. (20101). In another meta-analysis the results showed that codon 72 was not associated with breast cancer risk among 37 combined case–control studies (23,567 cases and 25,995 controls). However, a significant association with decreased risk of breast cancer was found in the Mediterranean studies (PP + PR vs. RR: OR = 0.32, 95% CI = 0.24−0.44, $P < 0.001$; PP vs. RR: OR = 0.35, 95% CI = 0.21−0.60, $P < 0.001$) (Hu, Z. et al. 2010). Another meta-analysis performed to investigate the association between breast cancer and the TP53 polymorphisms codon 72 (27,046 cases and 30,998) controls found no evidence of significant association between breast cancer risk and TP53 codon 72 polymorphism in any genetic model. However, in the stratified analysis for Indian population, significantly association was observed in additive model (OR = 0.62, 95% CI = 0.46–0.82, *P* value of heterogeneity test [P_h] = 0.153) and recessive model (OR = 0.70, 95% CI = 0.50–0.92, P_h = 0.463). indicating that codon 72 homozygous mutants may be associated with decreased breast cancer risk in Indian population (He, X. F. et al.(2011).

2.5.2. Low Penetrance Genes

Low-penetrance susceptibility alleles, sometimes called “modifier genes,” are defined as polymorphic genes with specific alleles that are associated with an altered risk for disease susceptibility. Usually, the variants in these genes are common in the general population. Therefore, although each variant may be

associated with a small increased risk for breast cancer in an individual, the attributed risk in the population as a whole is likely to be higher than for rare, high-penetrance susceptibility genes. Candidate modifier genes are chosen on the basis of biologic plausibility. Modifier genes may be found in a number of pathways, including detoxification of environmental carcinogens, steroid metabolism pathways, DNA damage response pathways, and immunomodulatory pathways (Martin, A. M. et al 2000). There are several classes of potential low penetrance breast cancer susceptibility genes, such as proto-oncogenes, oestrogen pathway genes, immunomodulatory pathway genes and metabolic pathway genes (de Jong, M. M.et al. 2002).

Table 1.1 List of genes implied in breast cancer tumorigenesis.

Gene	Location and function
BRCA1	17q21, DNA RP*, guardian of genome integrity
BRCA2	13q12-13, DNA RP, guardian of genome integrity
TP53	17p13.1, DNA RP, protection against replication of damaged DNA
ATM	11q22-23, DNA RP, sensor in cellular response to DNA double strand breaks
PTEN	10q23.3, TSG†, suppresses cell cycle progression and induction of apoptosis
LKB1	19p13.3, serine/threonine kinase, otherwise unknown function
HRAS1	11p15, proto-oncogene, control of cell growth and differentiation
NAT1	8p22, MP‡, detoxification of arylamines

NAT2	8p22, MP, detoxification of arylamines
GSTM1	1p13.3, MP, detoxification of a wide range of xenobiotics, including environmental carcinogens, chemotherapeutic agents, and reactive oxygen species
GSTP1	11q13, MP, detoxification of numerous chemicals including chemotherapy agents and catechol estrogens
GSTT1	11q, MP, detoxification of a wide range of xenobiotics, including environmental carcinogens, chemotherapeutic agents, and reactive oxygen species
CYP1A1	15q, MP, EP§, metabolism of oestrogens and PAHs
CYP1B1	2p21, MP, metabolism of PAHs
CYP2D6	22q11-ter, MP, metabolism of many commonly prescribed drugs, including debrisoquine and codeine
CYP17	10q24.3, EP, balance of estrogens, progesterones, and androgens
CYP19	11q21.1, EP, catalysing the conversion of androgens into estrogens, determines the local estrogen level
ER	6q25.1, EP, binding and transfer of estrogens to the nuclei, ER modulates transcription of a number of growth factors (IGF-1, TGFα)
PR	11q22-23, EP
AR	Xq11-12, EP
COMT	22q11.2, EP, conjugation and inactivation of catechol estrogens
UGT1A1	2q37, MP, EP, phase II drugs metabolism and maintain intracellular

	steady state levels of estrogen
TNFα	6p21, ¶ IP, central mediator in the inflammatory response and immunological activities to tumour cells
HSP70	6p21, molecular chaperones, regulation of structure, subcellular localisation, and turnover of cell proteins
HFE	6p21, IMP**
TFR	3q, IMP
VDR	12q, cell differentiation
APC	5q22, inhibits the progression of cells from G1 to S phase, apoptosis, cell-cell interactions
APOE	19q13.2, lipid metabolism
CYP2E1	10q24.3-ter, MP, metabolism of acetone, ethyl glycol, and ethanol
EDH17B2	17q12-21, EP, catalyses the reaction between estrone and estradiol
HER2	17q21, proto-oncogene, control of cell growth and proliferation
TβR-I	9q33-34, cell growth

*DNA RP: DNA repair pathway. †TSG: tumour suppressor gene. ‡MP: metabolic pathway. §EP: estrogen pathway. ¶IP: immuno pathway. **IMP: iron metabolism pathway.

2.5.2.1. Proto-oncogene, cell invasion and angiogenesis genes

Many proto-oncogenes, with different functionality and cellular localization, have been reported to play a role in human breast carcinogenesis. In sporadic

breast cancer proto-oncogenes amplification is frequently found, but only a few of these amplified genes are crucial in the development of breast cancer, e.g. MYC, Int2, EMS1, CCND1 and ERBB2. Growth factors like EGF, TGFα and IGF-1 could be also involved in proliferation and growth of breast cancer. Invasion, cell adhesion and 'homing' of tumour cells are essential steps in the metastatic spread of cancer cells. Several genes are involved in this process, e.g. N-CAM, integrins, E-Cadherin, uPA, cathepsinD, B, collagenase I-IV, CD44, NME1 and metalloproteases. Growth and progression of tumours is accompanied by neovascularisation (angiogenesis). Tumour cells in the stroma contribute to an increase of vascular endothelial growth factor (VEGF) and other angiogenic factors, like basic fibroblast growth factor and platelet-derived growth factor (Kenemans, P. et al. 2004).

2.5.2.2. Steroid receptors

The estrogen receptor (ER) α gene located on chromosome 6q25.1 is the most important growth factor receptor involved in hormone-dependent breast carcinogenesis. Whereas the ER β gene is located on 14q22-24 locus (Kenemans, P. et al. 2004). ER-positive breast cancer constitutes about 60% of breast cancers arising in premenopausal women and 80% of those diagnosed after menopause making level of ER expression a strong predictor of response to endocrine manipulation in breast cancer patients. Estrogens promote cell proliferation in both normal and transformed epithelial cells by modifying the expression of hormone-responsive genes involved in the cell cycle and/or

apoptosis. The ER α is shown to be critical for estrogen-induced tumor regression and apoptosis, as blockade of the ER α signaling pathway using the pure anti-estrogen completely inhibits the apoptotic effect of estrogen. It has been reported that estradiol induces regression of tamoxifen-resistant breast cancer tumors by inducing Fas expression and suppressing the anti-apoptotic or pro-survival factors such as nuclear factor-kappa-B (NF-κβ) and HER2/neu. Akt and NF-κβ signaling pathways are involved in cell survival including cell growth, proliferation, motility, and many of breast cancer transforming events are reported to be a result of enhanced signaling of these pathways. There are several reports demonstrating that estradiol can inhibit these signaling pathways, and consequently induce apoptosis of breast cancer cells (Xue, F. et al. 2011).

2.5.2.3. Xenobiotic metabolizing GST and CYP17 and genes

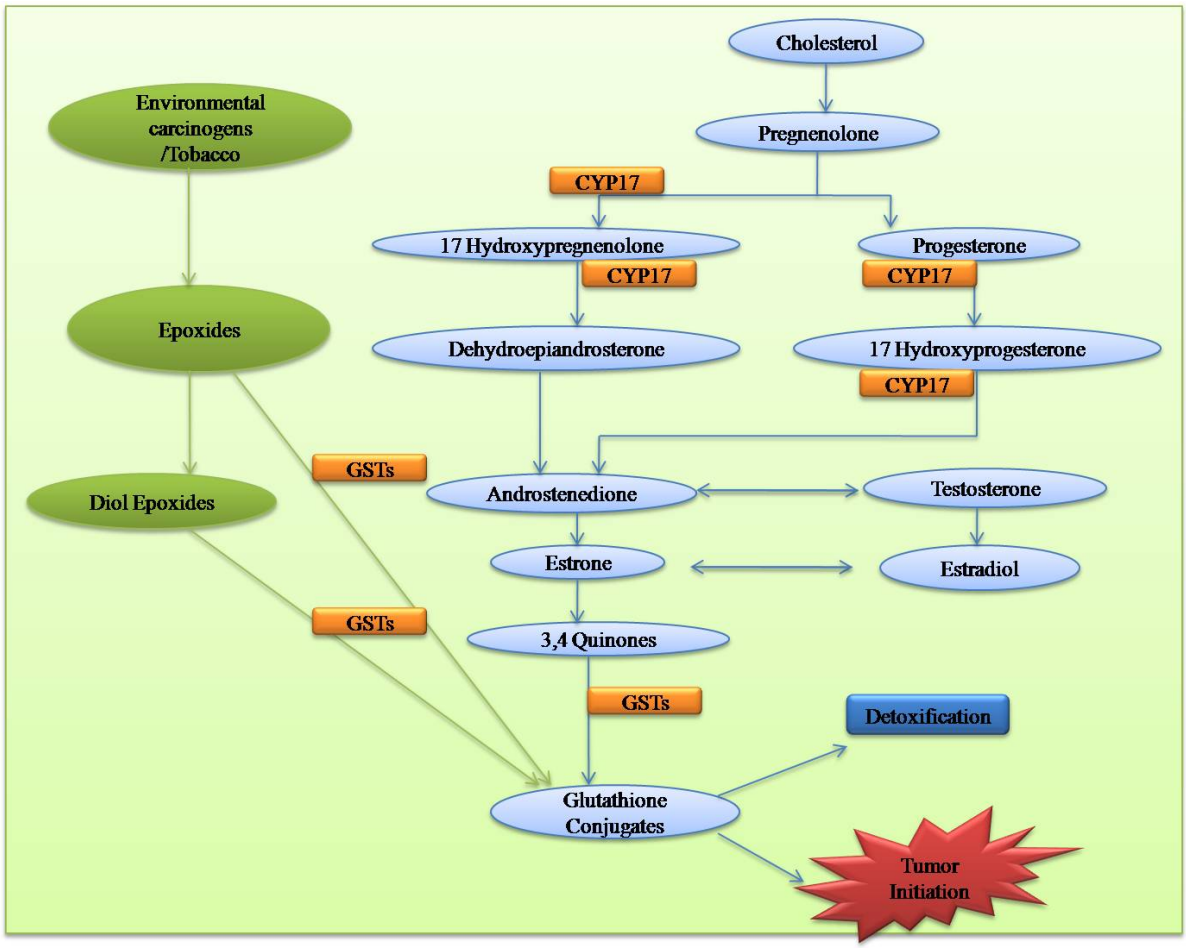


Figure 2.6: Overview of Xenobiotic metabolizing GST genes and CYP17 in breast cancer

An increasing number of epidemiological studies of common genetic polymorphisms having a role in the metabolism of estrogens or in the activation or detoxification of drugs and environmental carcinogens have been reported (Coughlin, S. S. et al 1999). Enzymes involved in metabolic pathways are of

interest because of their possible role in (de)toxification of chemical compounds. A number of metabolic pathway genes, including the *GST* family are thought to have evolved as an adaptive response to environmental exposure to toxins, including some carcinogens. Similarly genes involved in the metabolism of sex hormones are strong candidates for breast cancer susceptibility genes (de Jong, M. M., et al. 2002) Those in the sex hormone biosynthesis pathway may affect production of, and thus exposure to, the most active estrogen; estradiol (Dunning, A. M., et al. 1999). Any alteration in the activity of these enzymes would result in an altered susceptibility to potentially toxic (mutagenic) compounds. This may determine the rate at which somatic mutations occur in genes in response to environmental exposures, resulting in altered cancer susceptibility (de Jong, M. M. et al. 2002) (*figure 2.6*).

The glutathione *S*-transferases (GSTs) constitute a family of genes that encode for enzymes that catalyze the conjunction of reactive chemical intermediates to soluble glutathione conjugates to facilitate clearance. There are four classes (α, μ, π and θ) of cytosolic GSTs, of which at least three are expressed in normal breast tissue *(*Martin, A. M. et al 2000). GSTM1 detoxifies hydrophobic electrophiles derived from the metabolism of xenobiotics, including PAH-derived epoxides. GSTT1 is considered as one of the most ancient enzymes among GSTs and it exhibits a different catalytic activity compared to other GSTs. The deletions of GSTM1 and GSTT1 that lead to loss of enzyme activity are both relatively prevalent in different populations. The prevalence of

GSTM1 null is around 50% and GSTT1 null genotype ranged from 15–25% in Caucasians to 60–80% in Asians. GSTP1 is the only locus of the pi class of GST family to be described so far and it exhibits particularly high and selective activity of detoxifying the carcinogenic epoxide of benzo[a]pyrene (BPDE). Two single nucleotide polymorphisms have been described at the GSTP1 locus. One is an A to G transition resulting in the amino acid change from isoleucine to valine at codon 105; the other is a nucleotide substitution of C to T that results in alanine to valine at codon 114. Some previous studies suggested that 105Val had different enzyme heat stability and affinity, and lower 1-chloro-2,4 dinitrobenzene (CDNB) conjugating activity, whereas other in vitro studies have shown that GSTP1 105Val was more active in conjugation reactions toward carcinogenic diol epoxides of PAHs. The mean GST activity was not significantly lower in individuals with 114Val allele, although GSTP1*C (105Val-114Val) was associated with reduced enzyme activity (Hung, R. J. et al. 2004)

CYP17 is one of the proposed low penetrance susceptibility genes which codes for the enzyme cytochrome P450c17α, which catalyzes both steroid 17α-hydroxlyase and 17,20-lyase activities at key branch points in the oestrogen (oestradiol) biosynthesis pathway. 17α-Hydroxylase activity converts steroids to precursors of the glucocorticoid cortisol and 17,20-lyase activity yields precursors of oestradiol and testosterone. It is conceivable that changes in the expression levels or activities of the cytochrome P450c17α enzyme may have an impact on oestrogen biosynthesis (Ye, Z. and J. M. Parry 2002) Cholesterol may be

converted to progestins, androgens, and estrogens by several pathways, the choice of which is determined by the cytochrome P450C17a enzyme (Coughlin, S. S. and M. Piper (1999). It has been suggested that the *Msp*A1 polymorphism in the cYP17 gene confer susceptibility to breast cancer, since this polymorphism creates an additional Sp-1 type (CCACC box) promoter site 34 bp upstream from the initiation of translation but downstream from the transcription start site. This additional promoter site may increase the rate of transcription of the *CYP17* gene and thus increase enzyme activity. The *Msp*AI polymorphism gives rise to three different genotypes: a homozygous wild-type (A1/A1), a heterozygous variant (A1/A2) and a homozygous variant (A2/A2) (Ye, Z. and J. M. Parry 2002).

2.5.2.4. DNA repair genes and cell cycle genes

DNA in most cells is regularly damaged by endogenous and exogenous mutagens (Goode, E. L. et al. 2002). There are ~50–60 known carcinogens in cigarette smoke. Many of these compounds are converted into reactive metabolites. If not inactivated, the reactive molecules bind to cellular DNA and form adducts (Zienolddiny, S. et al. 2006). The tobacco related metabolites like reactive quinones, are capable of binding and damaging macromolecules including DNA, glutathione, tubulin, histones, topoisomerase II and other DNA-related proteins. Additionally, the metabolites may give rise to reactive oxygen species (ROS). Direct attack of DNA by ROS and reactive quinones or replication of unrepaired DNA damage can result in DNA double-strand breaks (DSB).

DSBs are repaired in vivo by nonhomologous end joining or, after replication when a second identical DNA copy is present, homologous recombination. DSB is especially genotoxic because (i) it affects both DNA strands and no intact template is available for repair; (ii) the repair is intrinsically more difficult than other types of DNA repair mechanisms because erroneous rejoining of broken DNA may occur. DNA DSB are potentially highly cytotoxic and can induce chromosomal aberrations (CA) and disrupt the genomic integrity of a cell. Therefore, the prompt and efficient repair of DSBs is fundamental for genomic stability and cancer prevention in the presence of environmental carciongenes like tobacco (Shen, M. et al. 2006). Due to the importance of maintaining genomic integrity as well as in the prevention of carcinogenesis, genes coding for DNA repair molecules have been proposed as candidate cancer-susceptibility genes (Goode, E. L. et al. 2002).

RAD51, a homologue of the *Escherichia coli* DNA repair protein, RecA is located at chromosome position 15q15.1. RAD51 functions in DNA repair by mediating homologous pairing and strand exchange reactions (Lose, F. et al. 2006). RAD51 binds single-stranded DNA (ssDNA) to form nucleoprotein filaments that are essential for strand transfer during homologous recombination repair (HRR). RAD51 is normally dispersed in the nucleus, but upon DNA damage induction, it redistributes to nuclear foci that are presumed sites of HRR. RAD51 with foci have been shown to be associated ssDNA regions after DNA damage. Several HRR proteins, including XRCC2, XRCC3, RAD51B, RAD51C,

RAD51D, and BRCA2, are important for RAD51 focus formation. BRCA2 has nine RAD51 binding regions, including eight BRC repeats encoded by exon 11 and a distinct RAD51 binding region encoded by exon 27. Expression of individual BRC repeats interferes with RAD51 focus formation and HRR, indicating that RAD51-BRCA2 interactions are important for both processes (Lu, H. et al. 2005) (*figure 2.7*). Besides the interactions of RAD51 with key players in breast tumourigenesis, there is additional evidence to support a role for RAD51 in breast cancer. *RAD51* gene exhibits loss of heterozygosity in a large range of cancers, including those of the lung, the colorectum and the breast. Specifically, 70% of breast tumours (from subjects with an unknown family history) and 32% of sporadic (nonfamilial) breast cancers have been found to exhibit loss of heterozygosity of this region. RAD51 expression has also been found altered in both primary tumours and cancer cell lines. Despite the presented evidence for a role for deregulated expression of RAD51 in DSB repair defects, there have been few studies assessing the effects of *RAD51* gene variation on breast cancer risk. These have largely focused on assessing the risk associated with a polymorphic variation in *RAD51*. Although there is some evidence that the rare -135G>C variant in *RAD51* is involved in modifying the *BRCA1/2*-mutation-positive breast cancer phenotype, studies focusing on the association between the *RAD51* -135G>C and -172G>T variants and breast cancer risk using case-control analysis have shown little support for a significant association with breast cancer (Lose, F. et al. 2006).

TP53, p21, and CCND1 are important genes involved in the G1-S checkpoint (Qiu, Y. L. et al. 2008). Cyclin D1, a protein encoded by the CCND1 gene located on chromosome 11q13, is a key cell-cycle regulatory protein modulating the restriction point early in the G1-phase (Ceschi, M. et al. 2005). Cyclin D1 acts by complexing with the cyclin dependent kinases CDK4 and CDK6, promotes phosphorylation and inactivation of retinoblastoma protein. CCND1 has been identified as an oncogene, and is rearranged, amplified or overexpressed in a variety of tumours. Recent results from several groups suggest that cyclin D1 may also be involved in the activities of transcription factors through CDK independent mechanisms (Bieche, I. et al. 2002). Cyclin D1 gene (CCND1) is amplified or overexpressed in a variety of tumours . In up to 20% of breast cancers, CCND1 is amplified and >50% of mammary tumours overexpress it. CCND1 exhibits a common A/G polymorphism at nucleotide 870, which modulates alternate splicing of CCND1. Both alleles lead to the expression of two different transcripts, but at different proportions. Several studies found the A-allele to be the major source of transcript form b, which encodes a cyclin D1 protein with an altered C-terminus. It lacks a PEST sequence postulated to target protein for rapid degradation. Carriers of one or two A-alleles may thus possess a longer protein half-life (Ceschi, M. et al. (2005).

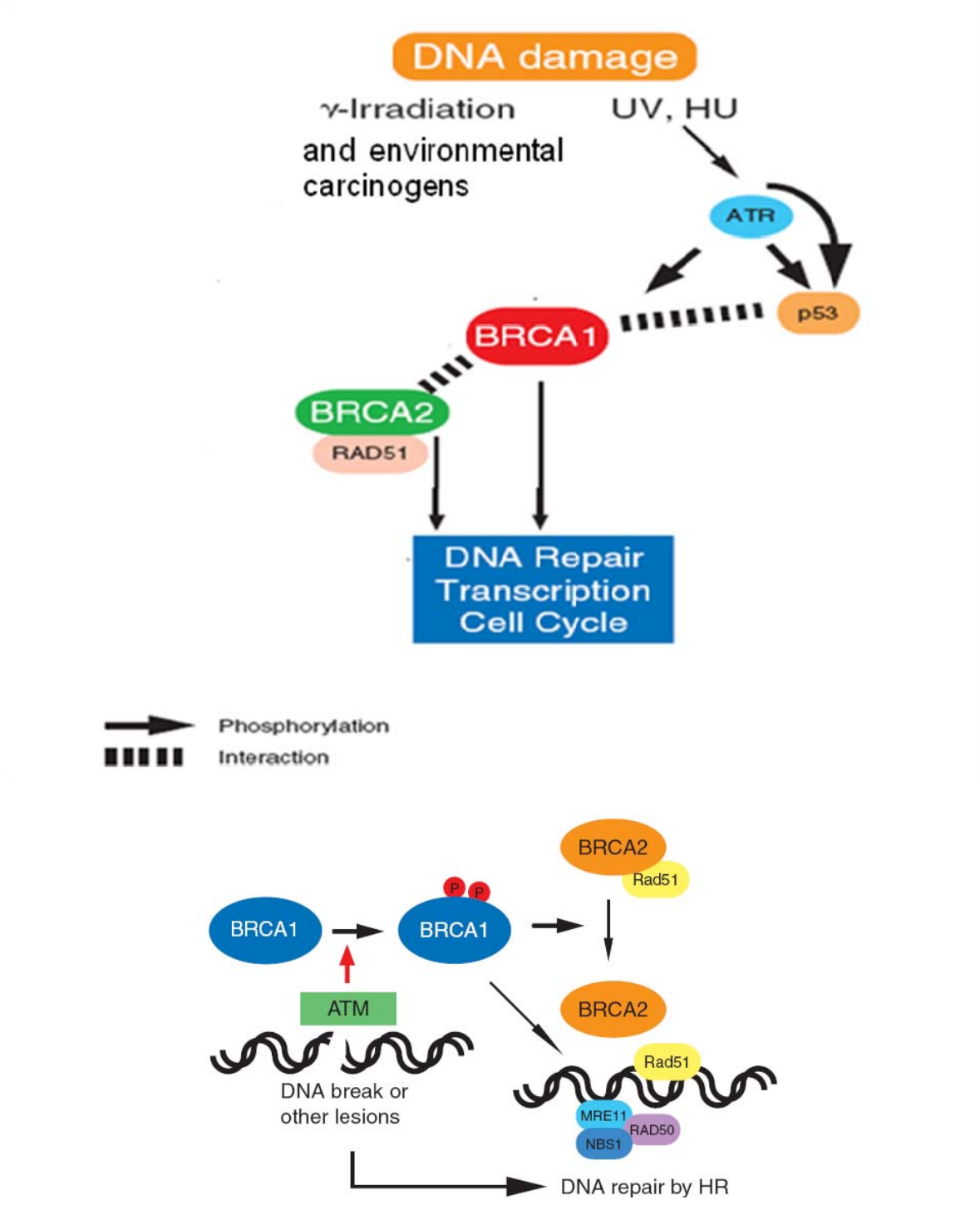


Figure 2.7: RAD51 and BRCA2 interactions necessary for DNA repair

2.5.3. Genomic alteration in breast cancer

Genomic alterations are believed to be the major underlying cause of cancer. Copy number or genomic alterations involves chromosomal regions with either more than two copies (amplifications), one copy (heterozygous deletions), or zero copies (homozygous deletions) in the cell (*figure 2.8*). Genes contained in amplified regions are natural candidates for cancer-causing oncogenes, while those in regions of deletion are potential tumor-suppressor genes (LaFramboise, T. et al. 2005).

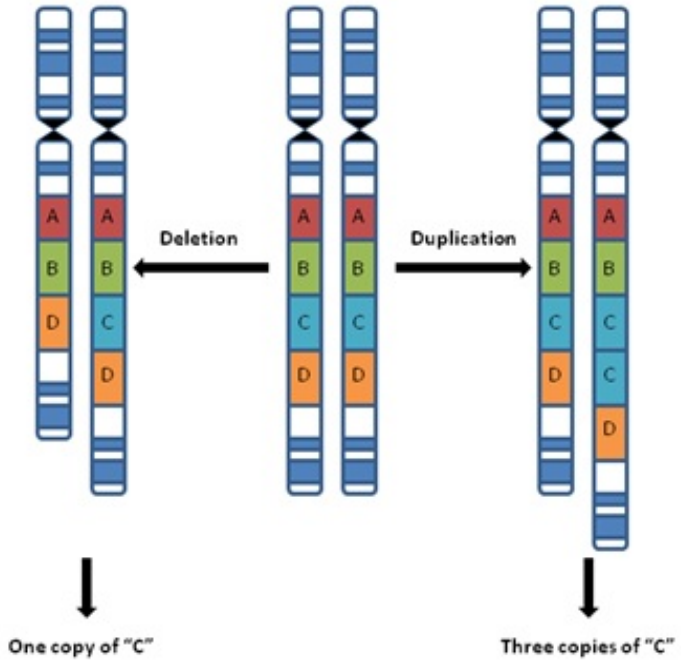


Figure 2.8: Copy number alterations; chromosome depicting gain and loss of gene "C"

Breast cancer is a complex disease in which multiple genetic factors can combine to drive pathogenesis. Changes in copy numbers of genes such as *ERBB2* and c-*MYC* have been extensively documented in breast cancer and are present in model cell lines. Amplified (and overexpressed) genes are prime therapeutic targets as for example, the use of the drug trastuzumab against ERBB2 has been shown to improve breast cancer survival rates alone or in combination with other treatments (Shadeo, A. and W. L. Lam 2006). Copy number alterations may provide potentially useful molecular markers for breast cancer prognostication or prediction of treatment response. Frequently observed CNAs include gain of chromosomal regions 1q, 8q, 17q, and 20q, and loss of 1p, 8p, 13q, and 17p. Sites of localized high-level DNA amplification harboring known oncogenes include 7p12 (EGFR), 8q24 (MYC), 11q13 (CCND1), 12q14 (MDM2), 17q12 (ERBB2), 20q12 (AIB1), and 20q13 (ZNF217). Deletions with known tumor suppressor genes (TSGs) include 13q12 (BRCA2), 17p13 (TP53), and 17q21 (BRCA1). Cytogenetic studies have identified gains on 8q, 17q12, and 20q13 to be associated with poor overall survival (Bergamaschi, A. et al. 2006).

Previously many studies have analyzed effect of environmental or occupational carcinogens on genomic alterations. For instance asbestos has been shown to be a genotoxic and cytotoxic agent that can produce both DNA and chromosomal damage. The mechanisms behind these actions may be multiple. In lung cancer asbestos related areas were detected in 2p21–p16.3,

3p21.31, 5q35.2–q35.3, 16p13.3, 19p13.3–p13.1 and 22q12.3–q13.1, the most prominent of these being 19p13 (Wikman, H. et al. 2007). Additionally, patients with higher levels of arsenic exposure depict higher levels of chromosomal instability in bladder tumors. Chromosomal alterations associated with arsenic exposure were found to be associated with tumor stage and grade, raising the possibility that bladder tumors from arsenic-exposed patients may behave more aggressively than tumors from unexposed patients (Moore, L. E.et al. 2002).

Recent work on breast cancer has been focused on distinguishing copy number alterations between different subtypes of breast cancer where a higher numbers of gains/losses were associated with the \basal-like" tumor subtype, while high-level DNA amplification is more frequent in \luminal-B" subtype tumors (Bergamaschi, A. et al. 2006). A higher frequency of copy number alterations among BRCA1 and BRCA2 tumors compared with sporadic cases have been found. BRCA1 tumors harbored frequent loss at 2q, 4p, 4q, 5q, and 12q, whereas BRCA2 tumors are characterized by a higher frequency of 6q and 13q losses as well as gains on 17q22-24 and 20q13. Use of metaphase CGH analysis to classify BRCA1 tumors based on patterns of genomic alterations have suggested gain of 3q and loss of 3p and 5q to distinguish BRCA1 from sporadic tumors (Jonsson, G. et al. 2005). In another study high-resolution array comparative genomic hybridization with an array of 4153 bacterial artificial chromosome clones had been analyzed to assess copy number changes in 44 archival breast cancer cases. Gene copy number changes were evaluated in the

tumors based on histologic subtype and estrogen receptor (ER) status. There was a consistent association between loss in regions of 5q and ER-negative infiltrating ductal carcinoma, as well as more frequent loss in 4p16, 8p23, 8p21, 10q25, and 17p11.2 in ER-negative infiltrating ductal carcinoma compared with ER-positive infiltrating ductal carcinoma (Loo, L. W.et al. 2004). Also to identify genetic changes involved in the progression of breast carcinoma, cDNA array comparative genomic hybridization (CGH) has been done on a panel of breast tumors and 49 minimal commonly amplified regions (MCRs) were identified that included known (1q, 8q24, 11q13, 17q21-q23, and 20q13) and several uncharacterized (12p13 and 16p13) regional copy number gains (Yao, J. et al. 2006).

2.6. Statistical Analysis

2.6.1. Testing for interaction

Breast cancer is believed to be influenced by several genetic and environmental factors, each factor potentially having a modifying effect on the other. Understanding the interplay between genetic and non-genetic factors is one of the major goals of genetic epidemiology. In genetic association single nucleotide polymorphisms (SNPs) are the most commonly used type of genetic markers (Marnellos, 2003). Looking beyond singular genetic effects and beyond the boundaries of additive inheritance of SNP polymorphisms should reflect biological pathways that are involved in disease etiology (Dixon *et al*., 2000).

Standard methods to analyze the simultaneous evaluation of a large pool of predictors (whether genetic or not) broadly fall into two classes: parametric and non-parametric methods. For instance, in a classic logistic modeling framework, in which case-control status is taken as the outcome variable, the search for functional variants can be carried out by constructing a model for the probability of disease. Quantifying the effects of a single locus is achieved by interpreting the corresponding regression coefficients, conditional on the fixed status at the remaining loci. However, if the single locus is involved in complex multi-co linearity patterns with other loci included in the model, it is questionable how much value can be placed on this interpretation (Van Steen & Molenberghs, 2004). This issue becomes even more relevant as the number of terms increases and interaction terms are considered as well. In addition, traditional parametric approaches have severe limitations when there are too many independent variables in relation to the number of observed outcome events. This is also referred to as the curse of dimensionality (Bellman, 1961). Therefore, alternative methods have been proposed to deal with elevated dimensionality and related problems when investigating interactions, including penalized logistic regression (Park & Hastie, 2008), (bagged) logic regression (Ruczinski *et al*., 2004), and non-parametric multi-locus techniques based on machine learning and data mining. The latter comprise tree-based methods (e.g., Recursive Partitioning and Random Forests), pattern recognition methods (e.g., Symbolic Discriminant Analysis, Mining Association rules, Neural Networks and Support Vector

Machines), and data reduction methods (e.g., Multifactor Dimensionality Reduction) (Cattaert, T. et al. 2011).

2.6.2. Multifactor Dimensionality Reduction (MDR)

MDR seeks to identify combinations of loci that influence a disease outcome. MDR reduces the number of dimensions by converting a high dimensional multilocus model to a one-dimensional model, thus avoiding the issues of sparse data cells and models with too many parameters that can cause problems for traditional regression-based methods. MDR classifies genotypical classes as either high risk or low risk according to the ratio of cases and controls in each class. This approach could be considered overly simplistic, and improves that embed a more traditional regression-based approach into the cell classification step, allowing application of the method to continuous as well as binary traits and adjustment for covariates, have been proposed (Lee, S. Y. et al. 2007 and Lou, X. Y.et al. 2007). Therefore, Multifactor Dimensionality Reduction (MDR) method is a constructive induction algorithm where the observed data is divided into ten equal parts and a model is fit to each nine-tenths of the data (the training data), and the remaining one-tenth (the test data) is used to assess model fit, thus using ten-fold cross-validation. Within each nine-tenths of the data, a set of *n* genetic factors is selected and their possible multifactor classes or cells are represented in *n* dimensional space. For example, for *n* = 2 diallelic loci, there are nine possible genotype classes or cells. The ratio of the number of cases to the number of controls is estimated in each cell and the cell is labelled

as either high risk if the case–control ratio reaches or exceeds a predetermined threshold (for example, ≥1) and low risk if it does not reach this threshold. This reduces the original *n*-dimensional model to a one-dimensional model (that is, one variable with two classes: high risk and low risk). The procedure is repeated for each possible *n*-factor combination and the combination that maximizes the case–control ratio of the high-risk group (that is, the combination that fits the current nine-tenths of the data best, giving minimum classification error among all *n*-locus models) is selected. The testing accuracy (which is equal to 1 – prediction error) of this best *n*-locus model can be estimated using the remaining test data portion of the data. The whole procedure is repeated for each of the nine-tenth-one-tenth partitions of the data, and the final best *n*-locus model is the model that maximizes the testing accuracy or, equivalently, minimizes the prediction error. The cross-validation consistency is defined as the number of cross-validation replicates (partitions) in which that same *n*-locus model was chosen as the best model (that is, the number of replicates in which it minimized classification error). The average prediction error is defined as the average of the prediction errors over the ten cross-validation test data sets. Note that the prediction error of each individual cross-validation replicate refers to the prediction error of the *n*-locus model chosen as the best model in that replicate, which will not always correspond to the final best *n*-locus model. One thus generates a best model within each cross-validation replicate as well as a final best model (with the associated cross-validation consistency and average

prediction error) for each different value of *n*. The cross-validation consistencies and average prediction errors can be used to determine the best value of *n* that gives the highest cross-validation consistency or lowest average prediction error, and thus the resulting overall best model (Cordell, H. J. 2009).

2.6.3. Classification and Regression Trees (CART)

CART is a nonparametric statistical procedure that identifies mutually exclusive and exhaustive subgroups of a population whose members share common characteristics that influence the dependent variable of interest. CART produces a visual output that is a multilevel structure that resembles branches of a tree. A generic illustration of CART output is presented in *figure 2.10*. Classification and regression trees begin with one "node," or group, containing the entire sample, called a parent node, which is illustrated in *figure 2.10* as Node 1. The CART procedure examines all possible independent, or splitting, variables and selects the one that results in binary groups that are most different with respect to the dependent variable, according to a predetermined splitting criterion (described later). The parent node then branches into two descendent, or child, nodes according to the independent variable that was selected. CART only splits each parent node into two child nodes. In Figure 1, Node 1 split into Nodes 2 and 3 according to a level of Independent Variable 1. Within each of these two child nodes, the tree-growing methodology continues by assessing each of the remaining independent variables to determine which variable results in the best split according to the chosen splitting criterion. Thus, each of the two

child nodes becomes a parent node to the two groups into which it splits. In Figure 1, Node 3 splits according to Independent Variable 2 into Nodes 4 and 5. Thus, Node 3 is a child node of Node 1 and a parent node to Nodes 4 and 5. The procedure continues through each branch of the tree until a stopping rule (defined later) is reached. At the point that no further split is made, a terminal node is created. In *figure 2.9*, Nodes 2, 4, and 5 are considered terminal nodes. Terminal nodes are mutually exclusive and exhaustive subgroups of the population. The dependent, or target variable, can be either categorical (i.e., classification tree) or continuous (i.e., regression tree). In a classification tree, this is the type of analysis used in the example presented in this article, the probability of having the dependent measure is estimated among those within each node. In regression trees, the average value of the dependent measure is estimated among those within each node. Independent variables can be any combination of categorical and continuous variables.

However, splits are always binary. In the case of an ordinal or continuous variable, CART searches through the full range of values and finds the best combination of categories or cut-point according to the predetermined splitting criterion (Lemon, S. C. et al. 2003).

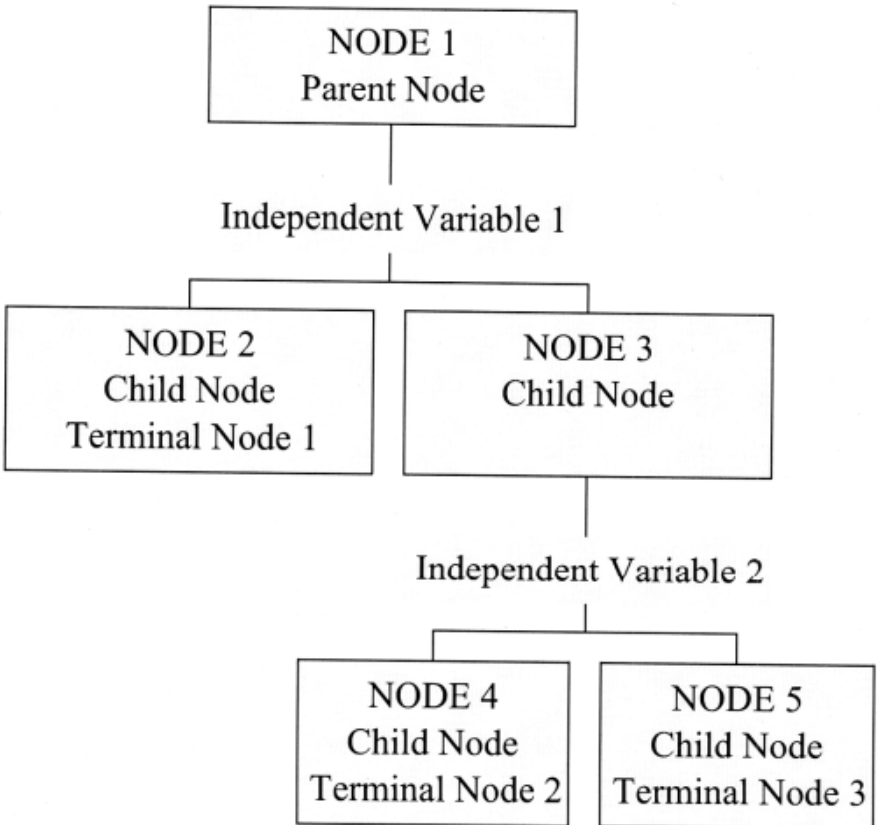


Figure 2.9 Example of classification and regression tree output.

2.6.4. Microarray Analysis using dChip

2.6.4.1. Normalization of arrays based on an 'invariant set'

As array images usually have different overall image brightness, especially when they are generated at different times and places, proper normalization is required before comparing the copy number levels of genes between arrays. Model-based method computation requires normalized probe-level data (from Affymetrix's DAT or CEL files). For a group of arrays, we

normalize all arrays (except the baseline array) to a common baseline array having the median overall brightness (as measured by the median CEL intensity in an array). The normalization is based only on probe values that belong to non-differentially altered SNPs, but generally we do not know which SNPs are non-differentially altered (control or housekeeping genes may also be variable across arrays). Nevertheless, it is expected that a probe of a non-differentially altered SNP in two arrays to have similar intensity ranks (ranks are calculated in two arrays separately). An iterative procedure is used to identify a set of probes (called the invariant set), which presumably consists of points from non-differentially altered SNPs (Li, C. and W. Hung Wong 2001).Therefore, invariant set normalization method adaptively selects probes that have similar ranks (thus more likely to belong to SNPs that have the same copy numbers) between one array and the baseline array to determine the normalization function. After normalization, the two arrays have similar overall brightness (Zhao, X. et al. 2004).

2.6.4.2. Model-Based Signal Values

After normalization, we used a model-based method to obtain the signal values for each SNP in each array. Because the probe response patterns of the three genotypes (AA, BB, and AB) are dissimilar, the new perfect match probe intensity is defined as as pmA + pmB and the new mismatch probe intensity is defined as as mmA + mmB for each probe quartet of a probe set. This transformation makes the probe intensity pattern and magnitude of a probe set

comparable across the genotypes. Then the perfect match/mismatch difference model was applied on the transformed probe-level data to compute mode lbased signal values. The model-based method weighs probes by their sensitivity and consistency when computing signal values, and image artifacts are also identified and eliminated by the outlier detection algorithm in this step.

2.6.4.3. Observed and Inferred DNA Copy Number

For each SNP, the signal values of all of the control are averaged to obtain the mean signal of 2 copy (male X chromosomes are multiplied by 2 before averaging), and the observed copy number is defined as (observed signal/mean signal of two copy) * 2, and visualized either log 2 ratio displayed in blue to white then to red color scale or white (0 copy) to red color scales. In general, a diploid genome is assumed in the absence of specific average DNA content data, but experimental values for mean copy number, derived from flow cytometry, can be substituted for the 2-copy assumption and will give more reliable results. To infer the DNA copy number from the raw signal data, the Hidden Markov Model (HMM) is used. First, it is specified that for each SNP the observed signal values are random values drawn from a t distribution with parameters determined by the underlying real copy number (Fold*2) and the estimated mean signals and their standard deviations (SDs) in the normal samples: (Signal-Mean* Fold / Std *Fold)~ *t*. These distributions give the "emission probabilities" of the HMM. Secondly, it is also assumed that the copy number changes are caused by genetic recombination events: for a particular

sample, the larger the genetic distance between the two markers, the more likely it is that recombination (thus a copy number change) will happen within the interval. The Haldane's map function. $\theta = 1/2\ (1 - e^{-2d})$ is used to convert the genetic distance *d* between two SNP markers to the probability (2θ) that the copy number of the second marker will return to the background distribution of copy numbers in this sample and thus independent from the copy number of the first marker. These probabilities are used as the "transition probabilities" of HMM that determine how *d,* the real copy number of one marker, provides information of the real copy number of the adjacent marker. Thirdly, the background distribution of copy numbers in each sample in two rounds is estimated. The proportion of chromosome regions that have a particular copy number is set to fixed values in the first round [0.9 for 2 copy, 0.1/(N-1) for copy 0 to N except 2, where N is the maximal allowed copy number in inference]. The HMM is run as described below and then the inferred copy numbers are used to re-estimate the sample-specific background distribution of the copy numbers. After this, the HMM model is reruned to obtain the final results. These background distributions are used as the "initial probabilities" of HMM specifying the likelihood of observing a particular copy number at the beginning of the p-arm and also used together with the "transition probabilities" to determine the dependency of the copy number values of two adjacent markers as described above.

A HMM model with these probabilities specifies the joint probability of the unobserved copy number and the observed signal values, and the Viterbi

algorithm was then used to obtain the most likely underlying copy number path of SNPs in a chromosome (in the p-arm to q-arm ordering), given the observed signal values. The algorithm works by analyzing one chromosome at a time. The HMM was applied to all of the chromosomes and all of the samples separately, and the best paths were defined as the inferred DNA copy number values. The inferred copy number is visualized in the same way as the observed copy number. The above described analysis methods are implemented in the dChip software (Zhao, X. et al. 2004).

Chapter 4

Polymorphism in putative xenobiotic metabolizing GST genes and CYP17 and assessment of breast cancer risk

Chapter 4

Polymorphism in putative xenobiotic metabolizing GST genes and CYP17 and assessment of breast cancer risk

4.1. Introduction

The several fold difference in incidence rates between different geographical regions suggest that environmental factors influence breast cancer risk significantly. Among the identified environmental risk factors in general for cancers, tobacco exposure is preventable (Terry, P. D. and T. E. Rohan 2002). The Northeast districts of India have the highest incidence of cancers associated with both smoking and smokeless tobacco (Mudur, G. 2005). Method and form of tobacco consumption in this region is reported to be different from rest of the India. Also, this region reports high risk for developing oesophageal, gastric cancer with betel quid chewing, another form for tobacco consumption (Phukan, R. K., M. S. Ali, et al. 2001). Betelquid is a combination of betel leaf, areca nut slaked lime and tobacco. There is a great spectrum of ingredients and ways of preparing betel quid which differs with different geographical region. Studies from Pakistan and Mainland China also report betelquid chewing as the major aetiological factor for oral leukoplakia and oral submucous fibrosis. It is also commonly used in South and Southeast Asia and Asia Pacific (Gupta, P. C. and C. S. Ray 2004).

Epidemiological perspective suggests an increased risk associated with exposure to genotoxic agents during breast development, as the undifferentiated ductal elements of the breast are more susceptible to the action of genotoxins early in life (Williams, J. A. and D. H. Phillips 2000). Environmental genotoxic stress like tobacco smoke and smokeless tobacco contain polycyclic aromatic hydrocarbons (PAHs), tobacco-specific nitrosamines, nitrosamino acids, aldehydes, metals, aromatic and heterocyclic amines and other genotoxic carcinogens (Boffetta, P. et al. 2008 and Lash, T. L. et al. 2005). The concomitant use of betel quid also lead to a 50-fold increase in reactive oxygen species generated (Anantharaman, D. et al. 2007).

In keeping with the polygene hypothesis of breast cancer (Pharoah, P. D. et al. 2004), the genes responsible for metabolizing the tobacco carcinogens appear to be prime candidates for the investigative search of breast cancer susceptibility genes. As the northeast region has a very high and typical usage of tobacco we selected genes related to catabolism and detoxification of xenobiotics (GSTM1, GSTT1, and GSTP1), tumour suppressor gene (TP53) and oestrogen biosynthesis (CYP17), to explore their contribution for breast.

Null polymorphisms in GSTT1 and GSTM1 genes result in absence of their respective expression of (mu) and (theta) isoenzymes and hence resulting in reduced glutathione binding efficiency of PAH epoxides and reactive conjugates causing DNA damage. The GSTP1 le105Val polymorphism is associated with reduced isoenzyme (pi) expression and a reduced capacity to

inactivate carcinogens including the diol epoxides created during phase I metabolism of PAHs (Phukan, R. K. et al. 2005).

CYP17 codes for the enzyme cytochrome P450c17á, responsible for catalyzing steroid 17á-hydroxylase and 17,20-lyase activities at key branch points in the estrogen biosynthesis pathway (Chaturvedi, H. K. et al. 2003). The 5' UTR T>C change produces A2 allele which has been associated with higher estrogen levels than the wild-type allele (Lash, T. L. et al. 2005). An increase or decrease in activity of cytochrome P450c17á enzyme may alter the level of endogenous estrogen, thereby influencing the estrogen derived quinone formation (Boffetta, P. et al. 2008) and their dextoxification by the GSTs (Gupta, P. C. and C. S. Ray 2004 and Williams, J. A. and D. H. Phillips 2000).

The TP53 tumor suppressor gene encodes a cell-cycle checkpoint protein that functions in the G1 phase of the cell cycle and it has a pivotal role in inducing apoptosis. The wild type TP53 gene polymorphism at codon 72 produces a protein with an arginine (Arg: CGC) or proline (Pro: CCC) genotype. The polymorphism changes the function of the TP53 protein and is strongly associated with the tumor formation process. The wildtype TP53 gene suppresses cellular transformation with activated oncogenes, therefore inhibiting the growth of malignant cells (Anantharaman, D., et al. 2007).

The existing literature in India documents the polymorphisms in GST genes, CYP17 and TP53 genes and has exhibited their role in breast cancer susceptibility. However studies of considerable sample size show conflicting

results on GST polymorphisms. While some studies show lack of association (Pharoah, P. D. et al. 2004, Samson, M. et al. 2007 and Saxena, A. et al. 2009), others depict association of GSTM1 deletion with sporadic (Chacko, P.et al. 2004) and familial (Tiemersma, E. W. et al. 2001) breast cancer risk. Studies on GSTT1 show its association with sporadic breast cancer (Tiemersma, E. W. et al. 2001), particularly in premenopausal women (Pharoah, P. D. et al. 2004) while other studies show insignificant association (Saxena, A. et al. 2009 and Chacko, P.et al. 2004). GSTP1 was found to be significantly associated to breast cancer when in combination with other GST gene genotypes (Chacko, P.et al. 2004). There are few studies reported on CYP17 polymorphism and breast cancer risk, although available results are inconsistent (Zhao, M. et al. 2001 and Chakraborty, A. et al. 2007). Studies on TP53 codon 72 polymorphism illustrate Arg/Arg genotype, alone and in combination with TGF β as a risk factor for breast cancer (Saxena, A. et al. 2009 and Liu, G. et al. 2001). However in some studies Arg/Pro and Pro/Pro genotype exhibited a significant protective association with total and postmenopausal breast cancer risk (Cao, G. et al. 2008). Another study documented lack of association between TP53 codon 72 polymorphism and breast cancer (Samson, M. et al. 2007).

The existing studies in Indian setting on the above genes have adopted a case control study design and applied regression approach to estimate the risk for a particular genotype and environmental factors (Samson, M. et al. 2007, Saxena, A. et al. 2009 and Chacko, P.et al. 2004). However the interaction

between gene-gene and gene-environment factors in such study settings can be enormous and lead to a biased estimate of the regression coefficients. Under such a situation the regression approach are not designed to test the high order interaction and therefore one need to employ advanced methodologies such as multi-factor dimensionality reduction (MDR) method for estimating the risk of cancer accredited to such interactions.

We present single factor and multifactorial analyses of high-order gene–gene and gene–environment interactions, and discuss the findings.

4.2. Experimental methods

4.2.1. Materials

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

4.2.2. Chemicals used

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

PROTEINASE K: 10 mg dissolved in 1 ml of TDW:1% ; AGAROSE: 1gm of agarose dissolved 1% TAE buffer.

4.2.3. Patient recruitment and sample collection

Patients with a confirmed diagnosis of breast cancer admitted in the Dr. Bhubaneswar Borooah Cancer Institute, Guwahati, Civil Hospital, Aizawl, and Sir Thutob Namgyal Memorial Hospital, Gangtok, the collaborating centers in Northeast India from the year 2005-2008 were included in the study. All subjects provided written informed consent for participation done under a protocol approved by the institutional ethics committee of Regional Medical Research Centre, North East Region (Indian Council of Medical Research).

4.2.4. Inclusion criteria

1. Incident cases during the period of December 2005 to 2008 and willing to participate in the study were included.
2. Cases confirmed by microscopy and for whom the breast was the primary site of cancer were included in the study

4.2.5. Exclusion criteria

1. Patients unwilling to give consent.
2. Patients who were too ill to participate in the study were excluded.
3. Patients who had taken any form of treatment earlier (Secondary cases) were also excluded from the study.
4. Patients with any other history of malignancy.

4.2.5. Patient details

Out of all the incident breast cancer cases only 117 cases agreed to participate in the present study and answered the questionnaire. At the same time information was collected from the attendants who accompanied cancer patients and who provided a readily available and cooperative source of controls from the same socio-economic background as the patients. A final group of matched controls were selected by random pairing of the cases with subjects from the pool of controls after matching for sex and age (within ± 5 years). The study included 117 cases and 174 controls between November 2005 and December 2008. All subjects including cases and controls were resident of the North-eastern part of India at the time of recruitment for the past five years and belonged to the same ethnicity. All of the 117 cases were of infiltrating ductal type of breast carcinoma. Details of age and sex and various demographic variables were collected in the course of the interviews as well as details of personal habits that included tobacco smoking and the consumption of alcohol as well as chewing practices.

4.2.6. Collection of blood samples

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

4.2.7. Extraction of Genomic DNA

Genomic DNA from breast cancer patients was extracted by using Himedia kit (Mumbai, India) and stored at -20^{O}C till further analyzed.

4.2.8 Quantification of Genomic DNA

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

4.2.9. Agarose Gel Electrophoresis of Extracted DNA Samples

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

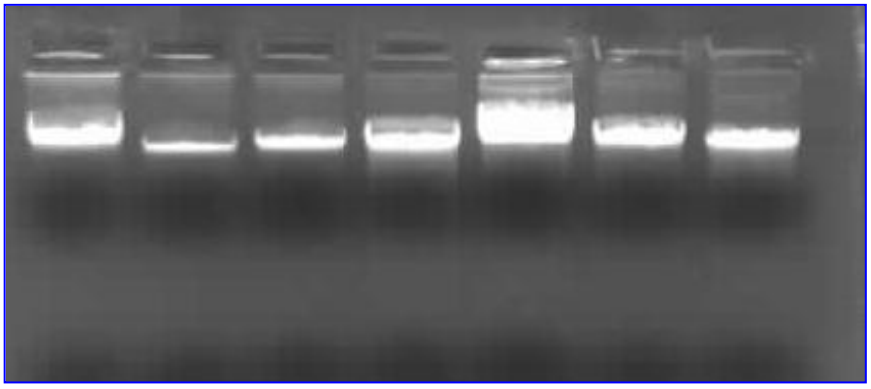


Figure 4.1: Agarose gel picture showing quality of genomic DNA isolated from the subjects

4.2.10. Genotyping of GSTM1 and GSTT1 by Multiplex PCR

A multiplex PCR method was used to detect the presence or absence of the GSTM1 and GSTT1 genes in the genomic DNA samples (*Table 4.1*). This method had both GST primers sets in the same PCR reaction and included a third primer set for β-globin as internal control to ensure proper functioning of PCR. The PCR was carried out for an initial activation step at 94^0C for 4 min, 20 cycles of denaturation at 93^0C for 1 min; annealing at 60^0C for 1 min; 72^0C for 1 min and in addition with these there were 15 cycles of denaturation at 93^0C for 1 min; annealing at 50^0C for 1 min; 72^0C for 1 min and a final extension at 72^0C for 10 min. The PCR products were electrophoresed in 2.5% agarose gels containing ethidium bromide, prepared and run in 0.5X TBE buffer. The absence

of 459 bp band indicates GSTT1 null and the absence of 219bp indicates GSTM1 null genotypes (*figure 4.2*).

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

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

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

Gene	Chr[a]	Loc	Polymorphism	
			Nucleotide	Codon
GSTT1	22q11.23	Gene	Presence>Null	Deletion
GSTM1	1p13.3	Gene	Presence>Null	Deletion
GSTP1	11q13	Exon 5	313A>G	Ile105Val
P53	17p13.1	Exon4	215C>G	Arg72Pro
CYP17	10q24.3	5' UTR	34T>C	5' UTR

[a]Chromosomal position is based on NCBI Build.

4.2.11 Genotyping of GSTP1, CYP17 and p53

Polymorphisms in GSTP1, CYP17and p53 were genotyped using PCR-RFLP (Polymerase chain reaction-Restriction Fragment length polymorphism) method. Standard PCR were performed on PTC-200 (MJ Research, USA). The PCR reaction were performed in a volume of 25µl with a final concentration of 1X PCR Buffer (MBI Fermentas), 1.5mM $Mgcl_{2,}$ 200 µM dNTPs, 2.0 mM , 1 Unit of Taq DNA polymerase and 100-300 ng of DNA Template. Negative controls were included in all PCR-runs to prevent misjudging following contamination of samples. PCR amplification consist of 35 cycles of denaturation at 94^0C for 45s; annealing at depend on gene for 45s; 72^0C for 45s followed by a final extension

at 72^0C for 10 min. PCR products were loaded on 2.5% agarose gel and subjected to gel electrophoresis in 0.5X TBE buffer, stained with ethidium bromide and visualized under UV. Detail of single-nucleotide polymorphisms (SNPs) selected for the study is summarized in *Table 4.2*. Sequence of the primer and their annealing temperatures are given in the *Table 4.3*.

Table 4.3: Sequence of primers used in the study

Gene	Primer sequence	T^0C	PCR (bp)
GSTT1	5'-TTCCTTACTGGTCCTCACATCTC-3' 5'-TCACCGGATCATGGCCAGCA-3'	-	459
GSTM1	5'-GAACTCCCTGAAAAGCTAAAGC-3' 5'- GTTGGGCTCAAATATACGGTGG-3'	-	219
GSTP1	5’-CCAGTGACTGTGTGTTGATC-3’ 5’-CAACCCTGGTGCAGATGCTC-3’	62	189
P53	5’-TTGCCGTCCCAAGCAATGGATGA-3’ 5’-TCTGGGAAGGGACAGAAGATGAC-3’	60	199
CYP17	5’-CATTCGCACTCTGGAGTC- 3’ 5’-AGGCTCTTGGGGTACTTG -3’	55	459

4.2.12 RFLP analysis of GSTP, P53 and CYP17 polymorphism

Restriction digestion of the amplified fragments was carried out for the above polymorphism in a water bath (*Table 4.4*). Heat inactivation of enzyme was done

at 80^0C for 20 minutes after completion of incubation with enzyme. Restriction enzymes that cleave the DNA specifically for different alleles were used (*Table 4.5*) and the alleles of each specific sample can be observed as a specific band pattern on the gel (*figure 4.3-4.5*). The genotyping results were confirmed by repeated analysis of approximately 10% of all samples randomly chosen.

Table 4.4: Standard protocol used for the RFLP experiment

COMPONENTS	STOCK CONC.	WORKING CONC.	1 REACTION (µl)
Water			3
Buffer	10X	1	1.5
Enzyme*	10Units/µl	5Units	0.5
PCR product			10.0
* Enzymes are specific for each polymorphism given in the table 4.5			

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

Gene	Enzyme	Site	Incubation Condition	PCR	RFLP product (bp)	
					Homo wild	Homo variant
GSTP1	BsmA1	5'-GTCTC^-3' 3'-CAGAG^-5'	55^0C for 8 hrs	189	189	148+41
P53	BstUI	5'-CG^CG-3' 3'-GC^GC-5'	37^0C overnight	199	113+86	199
CYP17	MspAI	5'-C^CGC-3' 3'-G^GCC-5'	37^0C overnight	459	459	335+124

4.3 Statistical analysis

A χ^2-test was used to assess whether the genotypes were in Hardy–Weinberg equilibrium (HWE) among case and control subjects. Odds ratios and their corresponding 95% confidence intervals were calculated by both unadjusted and adjusted logistic regression analysis as a measure for association with the risk for the genes and environmental factors considered. A two side $p \leq 0.05$ was considered as statistically significant. Data for family history was missing for 17% of cases and 42% of controls hence family history was not incorporated in the analysis. The gene–gene and gene-environment interaction was examined using the multifactor dimensionality reduction (MDR) method.

With MDR, genotype and environmental factors were pooled into high- and low-risk groups, effectively reducing the multifactor predictors from multi-dimensions to single dimension. The new one-dimensional multifactor variable

was evaluated for its ability to classify and predict disease status through cross-validation and permutation testing (Cao, G., et al. 2008). MDR ultimately selected one genetic model, either single or multilocus, that most successfully predicted phenotype or disease status. The data was then randomly divided into 10 equal parts. A training set of 9/10 of the data was used to search for the best model. The remaining 1/10 of the data is the testing set. Here, we also used 10-fold cross validation, and the analysis was repeated 10 times with different random seeds to reduce the possibility of biased results due to the chance divisions of the data into training and testing sets (Vaarala, M. H. et al. 2008). Finally, all the variables in the best model were combined and dichotomized according to the MDR software and their ORs and 95% CIs in relation to breast cancer risk were calculated in logistic regression models.

4.4 Results

The mean age was 45.5 ±12.86 years for the cases and 45.98 ±14.44 years for the controls. There were no significant differences between case and control subjects in terms of distributions of tobacco smoking (p=0.63), tobacco chewing (p=0.89), alcohol consumption (p=0.10) and they were not found to be associated with breast cancer risk. However, women with a betelquid chewing history had around five times the risk of developing breast cancer [4.78 (2.87 – 8.00), 0.001] (*Table 4.6*).

Table 4.6: Demographic data of patients with breast cancer

Variable	categories	N	Cases (117), (%)	Controls (174), (%)	χ^2 P value	Logistic Regression Analysis O.R(95%CI)
Tobacco Smoking	no	240	95 (81.2)	145 (83.3)	0.63	1
	yes	51	22 (18.8)	29 (16.7)		1.15 (0.62 – 2.13)
Tobacco Chewing	no	168	67 (57.3)	101 (58)	0.89	1
	yes	123	50 (42.7)	73 (42)		1.03 (0.64 – 1.65)
Betel Quid Chewing	no	144	32 (27.4)	112 (64.4)	**0.001**	**1**
	yes	147	85 (72.6)	62 (35.6)		**4.78 (2.87 – 8.00)**
Alcohol consumption	no	276	108 (92.3)	168 (96.6)	0.10	1
	yes	15	9 (7.7)	6 (3.4)		2.33 (0.80 – 6.74)
Distribution of Age	< 30	38	13 (11.1)	25 (14.36)	0.59	
	30-39	58	26 (22.2)	32 (18.3)		
	40-49	80	39 (33.3)	41 (23.5)		
	50-59	63	18 (15.3)	45 (25.8)		
	> 60	52	21 (17.9)	31 (17.8)		

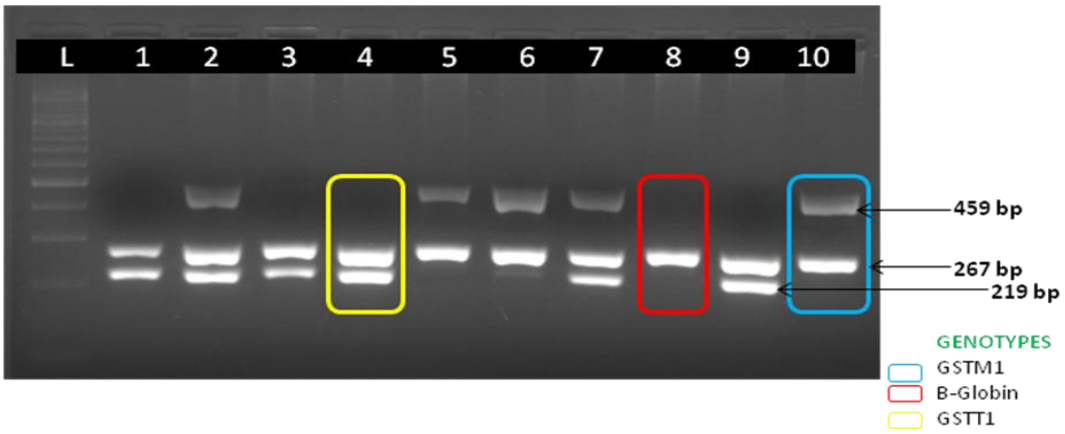


Figure 4.2: Agarose gel picture showing multiplex PCR for GST polymorphism. Lane 1, 2, 3, 4, 7 and 9-samples with 459 bp represent wild type GSTT1; Lane 2,5,6,7 and 9-sample with 219bp represent GSTM1 gene; The presence of 267bp in all lanes represent β-globin gene used as internal control. M-100bp ladder

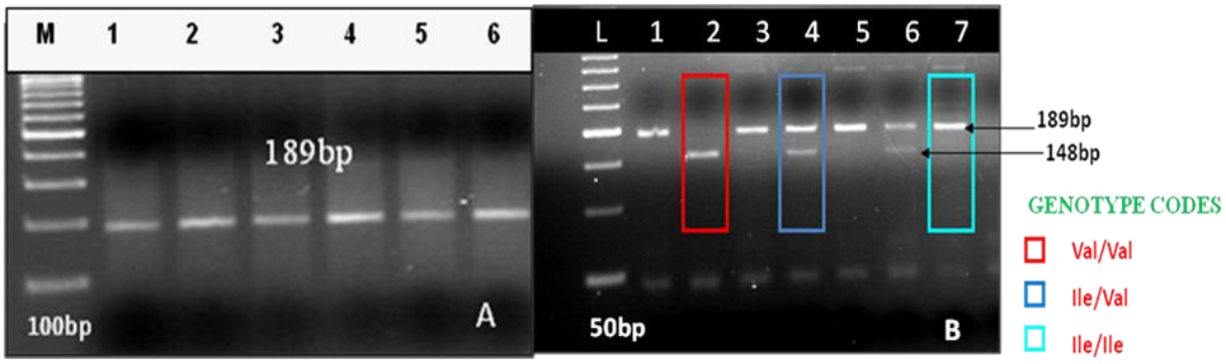


Figure 4.3: Agarose gel picture showing RFLP products of GSTP1 gene: A-showing PCR amplification of GSTP1 gene (189bp). B-RFLP of GSTP1 PCR product; Lane 1,3,5 and 7-samples with 189 bp represent wild type Ile/Ile allele (AA genotype); Lane4 and 6-sample with 189 bp and 148 bp represent heterozygous Ile/Val allele (AG genotype); Lane 2 - sample with 148bp and 41 bp not visible represent homozygous Val/Val allele (GG genotype). M-100 & 50bp ladder.

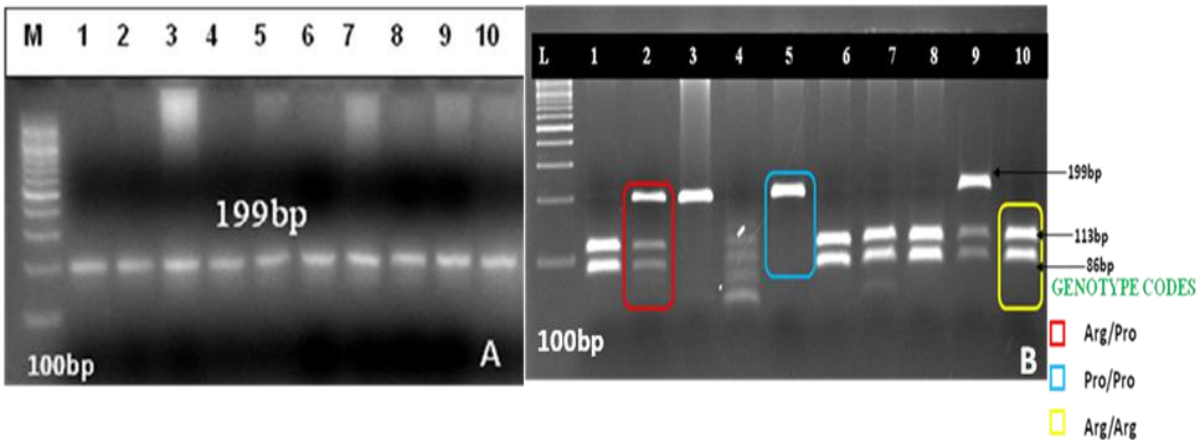


Figure 4.4: Agarose gel picture showing RFLP products of P53 codon 72 polymorphism: A- showing PCR amplification of exon 4 of p53 gene. B-RFLP of p53 PCR product; Lane 3 and 5-sample with 199 bp represent homozygous Pro/Pro allele (GG genotype); Lane 2 and 9 - samples with all three bands (199bp, 113bp, 86bp) represent heterozygous Arg/Pro allele (GC genotype); Lane 1,6,7,8 and 10- samples with two band (113 bp, 86 bp) represent homozygous Arg/Arg allele (CC genotype) . M-100bp ladder.

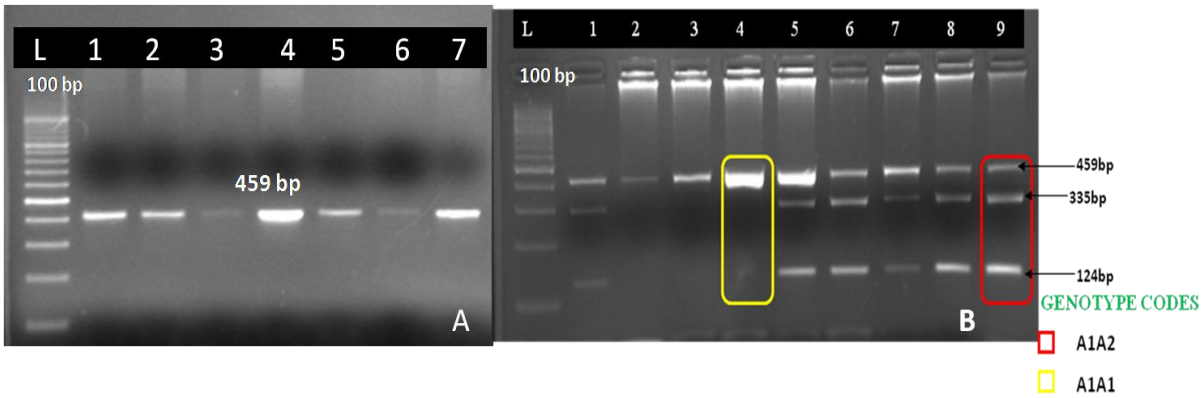


Figure 4.5: Agarose gel picture showing RFLP products of CYP17 T>C polymorphism: A- showing PCR amplification of CYP17 gene. B-RFLP of CYP17 PCR product; Lane 2,3 and 4-sample with 459 bp represent homozygous A1A1 allele; Lane 1,5,6,7,8 and 9 - samples with all three bands (459bp, 335bp, 124bp) represent heterozygous A1A2 allele ; no samples with two band (335 bp, 124 bp) represent homozygous A2A2 allele is seen . M-100bp ladder.

Table 4.7: Association between genotypes, alleles of GSTT1, GSTM1, GSTP1, CYP17 and TP53 polymorphisms and the risk of breast cancer

	Case (117)		Control (174)		Logistic Regression Analysis				
	Count (%)	Allele probabilities (p-value HWE)	Count (%)	Allele probabilities (p-value HWE)	Unadjusted OR (95% CI)	P	Adjusted OR (95% CI)	P	χ^2,df,p
GSTM1									
Present	94 (80.3)		122 (70.1)		1.00		1.00		**3.82, 1, 0.05**
Null	23 (19.7)		52 (29.9)		**0.57 (0.32- 1.00)**	**0.05**	**0.55 (0.30-1.02)**	**0.05**	
GSTT1									
Present	84 (71.8)		105 (60.3)		1.00		1.00		**4.02,1, 0.04**
Null	33 (28.2)		69 (39.7)		**0.59 (0.36-0.99)**	**0.04**	**0.59 (0.34-1.03)**	**0.06**	
GSTP1									
AA	62	A: 0.73, G: 0.26	108	A: 0.79, G: 0.20	1.00		1.00		4.03, 2,

	(53.0)	(0.56)	(62.1)	(0.15)					0.13
AG	48 (41.0)		62 (35.6)		1.34 (0.82 - 2.20)	0.23	1.26 (0.74-2.16)	0.38	
GG	7 (6.0)		4 (2.3)		3.04 (0.85 - 10.82)	0.08	2.01 (0.53 -7.66)	0.30	
AA + AG					0.37 (0.10-1.29)	0.11	0.54 (0.14-2.03)	0.36	
AG + GG					2.70 (0.77 - 9.45)	0.11	1.83 (0.49-6.87)	0.36	
A	172		278		1.000				
G	62		70		**1.43 (0.96 - 2.11)**	**0.07**			
CYP17									3.12, 2, 0.21
A1A1	33 (28.2)	A1: 0.52, A2: 0.47 (0.95)	44 (25.3)	A1: 0.53, A2: 0.46 (0.08)	1.00		1.00		
A1A2	58 (49.6)		98 (56.3)		0.78 (0.45-1.37)	0.40	.072 (.39-1.32)	0.29	
A2A2	26 (22.2)		32 (18.3)		1.08 (0.54 - 2.15)	0.81	1.23 (0.58 - 2.63)	0.57	

A1	124		186		1.				
A2	110		162		1.01 (0.73 - 1.41)	0.91			
A1A1 + A1A2					0.78 (0.44-1.41)	0.42	0.652 (0.34-1.23)	0.19	
A1A2 + A2A2					1.26 (0.70-2.26)	0.42	1.53 (0.80- 2.90)	0.19	
TTP53									
AA	28 (23.9)	A: 0.47, P: 0.52 (0.42)	38 (21.8)	A: 0.47, P: 0.52 (0.84)	1.00		1.00		0.54, 2, 0.76
AP	54 (46.2)		88 (50.6)		0.83 (0.46-1.50)	0.54	0.72 (0.37-1.39)	0.33	
PP	35 (29.9)		48 (27.6)		0.99 (0.51-1.90)	0.97	0.84 (0.41-1.73)	0.65	
AA +AP					0.893 (0.53-1.49)	0.66	0.94 (0.54-1.65)	0.85	
AP + PP					1.12 (0.66-1.87)	0.66	1.05 (0.60-1.84)	0.85	
A	110		164		1.00				
P	124		184		1.00 (0.72- 1.40)	0.97			

The genotypic distribution of the genetic markers under study was found to be in HWE both in cases and controls (*Table 4.7* gives the p values). The GSTP1, CYP17 and TP53 genes were not associated with breast cancer risk when adjusted for age, tobacco smoking, tobacco chewing, betel quid chewing, alcohol consumption and tobacco exposure. However, women with GSTT1 null polymorphism were 41 per cent less susceptible [0.59, (0.34-1.03), 0.06] for having breast cancer. Women with GSTM1 null polymorphism were also 55 per cent less susceptible [0.55 (0.30-1.02), 0.05] for having breast cancer.

A marginally significant risk was observed among women having G/G genotype of *GSTP1* gene [3.04 (0.85 - 0.82), 0.08], but when adjusted for the exposure variables significance was lost. Allele frequencies of different alleles of *GSTP1, CYP17* and *TP53* genes were also compared. The G allele of the *GSTP1* gene was found to be over represented in cases as compared to the controls indicating that G allele might be a risk factor for breast cancer [1.43 (0.96 - 2.11), 0.07]. The frequencies of alleles of other genes did not differ significantly.

Table 4.8: Multifactor dimensionality reduction (MDR) models of selected gene and environmental factors

	Models	Testing Balanced Accuracy	CVC	Odds Ratio (95% CI)[1]	P value of χ^2 test[1]
1 order	Bet	0.6851	10/10	4.7984 (0.9523,24.1769)	**0.05**
2 order	bet alc	0.6808	10/10	4.5982 (0.9214,22.9468)	**0.05**
2 order	GSTP1 bet	0.6851	10/10	4.7984 (0.9523,24.1769)	**0.05**
3 order	smk chw bet	0.7193	10/10	6.822 (1.2482,37.2871)	**0.02**
4 order	GSTP1 smk chw bet	0.7107	9/10	6.1081 (1.1943,31.2402)	**0.02**
5 order	GSTP1 TP53 smk chw bet	0.6625	10/10	3.8925 (0.8129,18.6389	0.08
5 order	M1 p53 smk chw bet	0.6967	10/10	5.2818 (1.0486,26.6041)	**0.03**
5 order	T1 p53 smk chw bet	0.658	10/10	3.7049 (0.7751,17.7084)	0.09

Bet, betel quid chewers; smk, tobacco smoking; chw, tobacco chewing; T1, GSTT1; M1, GSTM1; p53, TP53; CVC, Cross Validation Consistency

Risk estimate and Chi-square test were based on the combination and dichotomization of the distribution of genetic factors according to the MDR software

MDR Analysis revealed *(table 4.8*), betel quid chewing to be the single factor imparting the main effect [testing accuracy of 0.6851 and Cross validation consistency 10/10, p =0.05]. The combination of betel quid chewing *alcohol

consumption [testing accuracy 0.6808 and CVC of 10/10 (p = 0.05)] and GSTP1 * betel quid chewing [testing accuracy of 0.6851 and CVC of 10/10 (p = 0.05)] were seen as the best two factor interaction models. The MDR analysis gave a three factor interaction model which added tobacco smoking and tobacco chewing to betel quid chewing increasing the test accuracy to 0.7193 and CVC of 10/10. Four way (GSTP1* tobacco smoking*tobacco chewing* betel quid chewing) interaction model was also identified which showed a lower cross validation consistency 9/10 but a significant testing accuracy. The addition of TP53 to the four factor interaction model gave a five factor interaction model (GSTP1*TP53* smoking* chewing* betel quid chewing) which decreased its testing accuracy to 0.6625 but increased the CVC to 10/10. Among the remaining two five order interaction models found, GSTM1*TP53*smoking*chewing*betel quid chewing was found to be significant with a CVC of 10/10 and 0.658 training accuracy. CYP17 was not found to interact with the environmental or the genotypic factors in any of the significant interaction models found, suggesting its minor role towards breast cancer development in this population.

Figure 4.6 depicts the interactions between nine attributes from the MDR analysis via a graphical representation of a ‘dendrogram’. It shows betelquid chewing, GSTT1 and GSTM1 on a separate branch imparting there independent effects to breast cancer risk.

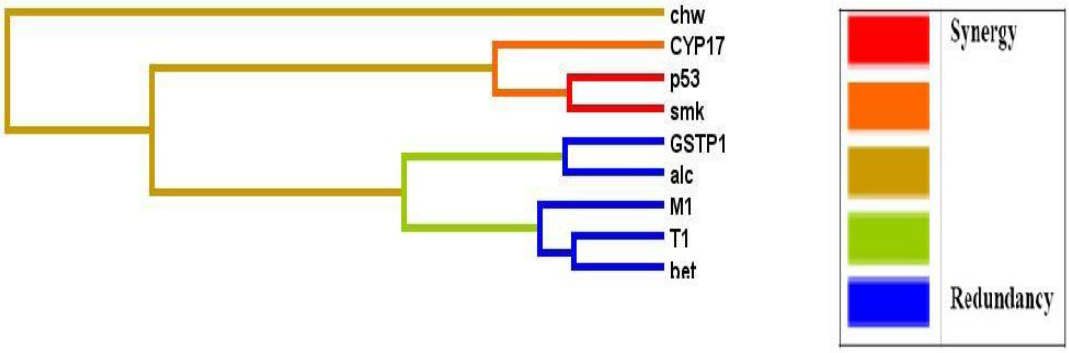


Figure 4.6: Interactions between nine attributes from the MDR analysis via a graphical representation of a 'dendrogram'

4.5 Discussion

There has been increasing interest in the association between tobacco exposure and increased breast cancer risk. The role of smoking in breast cancer aetiology has been extensively studied. Yet, the association remains equivocal and much debated. Smoking has been proposed to increase breast cancer risk, based on studies showing breast epithelial genotoxicity of tobacco-related compounds (Magnusson, C., et al. 2007). Several explanations for the lack of consistency in previous studies have been suggested. Included among these is the possibility that the observed associations are not causal, in which case chance or bias might have driven some of the previous findings in either direction from the null (Alguacil, J. and D. T. Silverman 2004). Tobacco smoke and smokeless tobacco are well known risk factors for pancreatic, bladder and hepatic cancer which occur at sites that are not in direct contact with them (Boffetta, P. et al. 2005 , Brennan, P. et al. 2001 , Tsai, J. F. et al. 2001 and Tsai, J. F. et al. 2001). Smokeless tobacco has been extensively investigated in

both oral and oesophageal cancers. Most of the carcinogenic contents in smokeless and tobacco smoke are similar and it should to be examined in other cancers as well to get a better understanding of the pathogenesis.

Our study provides evidence that **betel quid chewing is a very important independent risk factor for breast cancer.** Since betel quid chewing has not been shown to be risk factors for breast cancer earlier, it is important to validate that our finding is not due to confounding bias. The bias may result from the control selection, information bias, or by un-controlling confounding factor. The estimated prevalence of current betel quid chewers reported in control females in the same population was found to be 38% which is almost similar to prevalence found in our study. Moreover, there was no significant difference in the prevalence rates of habitual alcohol drinking between our controls and those (4% for alcohol) found in another case control study [9]. Based on the information mentioned above, our controls seem be representative for the same population, and make bias unlikely from control selection or under-reporting of life-style habits.

As shown in *Table 4.8*, although **betel quid chewing is identified as the main risk factor, the interactions with other factors (smoking, tobacco chewing, alcohol consumption and GSTM1, GSTT1, GSTP1, CYP17, TP53) only modified the risk.** Interactions conferred insignificant risk to breast cancer on removal of betel quid chewing from MDR analysis. This confirms the major contribution of betel quid to breast cancer risk. Betel quid chewers swallow the

betel quid juice (saliva extract of betel quid produced by chewing) (Xiong, P. et al. 2001) which get absorbed from the intestine and pass through the blood stream to various organs like kidney, pancreas and breast. The carcinogens present in it can be stored in breast adipose tissue and then get metabolized and activated by human mammary epithelial cells (Alguacil, J. and D. T. Silverman 2004). Presence of tobacco-related DNA adducts has been demonstrated in both breast tumour and adjacent normal tissue (Nair, J. et al. 1985). Saliva and urine of betel quid chewers have also shown presence of cancer causing nitrosamines like N'-nitrosonornicotine (1.0 to 51.7ng/ml), N'-nitrosoanatabine, 4-(methyl-nitrosamino)-1-(3-pyridyl)-l-butanone (0 to 2.3 ng/ml), nicotine and cotinine (Lee, C. H. et al. 1996). Animals fed with arecoline, a major constituent of betel quid are shown to develop genotoxcity in the ovary (Rivenson, A. et al. 1988). Animals exposed to nitrosamines, by chronic oral administration and by drinking water, produced during betelquid chewing are shown to develop lung, pancreatic tumours and intestinal metaplasia (Sen, S. 1987) and van Bladeren, P. J. 2000). Moreover carcinogens derived from betel quid chewing are shown to induce p53 mutation and over-expression of c-myc protein with activated ras oncogene and subsequent over-expression of cell cycle regulatory protein, cyclin D1 in oral cancer (Tsai, J. F. et al. 2001). However, similar phenomenon induced by betelquid consumption leading to breast carcinogenesis needs elucidation.

GSTs are predominantly involved in the detoxification of xenobiotics and genetic variations in them have been implicated in the etiology of numerous cancers. However, substrate bioactivation reactions by GSTs are known to occur

through formation of conjugates which are activated through cysteine conjugate beta-lyase, redox cycling and/or release of the original reactive parent compound [36]. Generally, detoxification by GST leads to formation of less reactive products that are readily excreted. However, in specific tissues and with certain exposures, the products formed are found to be more reactive than the parent compound (Kim, W. J. et al. 2002). For example, detoxification of dichloromethane by GSTT1enzyme results in bi-products which are found to be carcinogenic in mouse (Kushman, M. E. et al. 2007).Toxic metabolites formed during CYP activation of reactive diol-epoxides are not detoxified by the GSTM1enzyme (Landi, S. 2000). Glutathione conjugation of halogenated compounds by GST is known to serve as a substrate for renal cysteine conjugate *b*-lyase which forms reactive chlorothioketenes found to directly damage the tissue. Therefore, an active GST enzyme conjugates the substrate and forms more reactive intermediate that directly damages the tissue. Conversely, the deleted variant GST genotype forms an inactive enzyme, metabolizing the compounds through oxidation, without formation of reactive intermediates (Kim, W. J. et al. 2002).

An increase in risk of kidney and liver tumours in humans with the active GSTT1 genotype following exposures to halogenated compounds has been reported (Kelsey, K. T. et al. 1997). In another study increase risk to renal cell carcinoma was reported to be associated with pesticide exposure exclusive to individuals with active GSTM1/T1 genotypes (Kim, W. J. et al. 2002). Previous

studies have also revealed a decrease risk to breast cancer among premenopausal women with the absence of GSTT1 enzyme (null genotype) (Evans, A. J. et al. 2004). Similar findings have been reported in head and neck (Chaudru, V. et al. 2009), bladder (Lemos, M. C. et al. 2008), melanoma (Shimada, T. et al. 1996), and thyroid (Ritchie, M. D. et al. 2001) cancer. Similar mechanisms have been proposed to operate through tobacco carcinogenesis among GSTT1 and GSTM1 positive individuals (Tan, S. M. et al. 2007). In this study a protective role has been observed by the absence of GSTT1 and GSTM1 enzymes for development of breast cancer (Table 2). The presence of the enzymes might have led to the activation of certain known as well as unknown procarcinogens present in the betelquid chewers (Anantharaman, D. et al. 2007) leading to breast carcinogenesis.

GSTP1 catalysis the conjugating reactions of PAHs and their electrophilic compounds to facilitate their excretion. A polymorphic adenine to guanine transition at nucleotide 313 (*A*313*G*) in exon 5 results in an isoleucine to valine substitution in codon 105 (I105V). This codon is located in the substrate-binding site of GSTP1, and the corresponding allozymes exhibited differential catalytic activities toward diverse substrates (Zhao, M., et al. 2001) [19]. In the present study marginal risk conferred by the variant genotype and its overrepresentation in cases suggests that due to the lower activity of this enzyme the detoxification of the carcinogens was hindered leading to breast carcinogenesis.

The most obvious limitation of the present study is its small sample size. However, the sample studied was well characterized and was from a homogenous Northeast Indian population, which reduces the risks of population stratification and false associations. As with all statistical analyses, replication and validity of findings is necessary to separate true relationships from chance findings. One advantage of the MDR method is that false-positive results due to multiple testing are minimized. This is primarily due to the cross-validation strategy used to select optimal models. MDR facilitates the simultaneous detection and characterization of multiple variables associated with a discrete clinical endpoint. This is accomplished by reducing the dimensionality of the multilocus data. Another advantage of MDR is that it can overcome sample size limitation since it is nonparametric and that it assumes no particular genetic model and no mode of inheritance need to be specified. This is important for diseases, such as breast cancer, in which the mode of inheritance is unknown and very complex (Wu, M. T. et al. 2004). However, logistic regression analysis revealed a protective role of GSTT1 and GSTM1 genes which were not observed in the MDR analysis as it only gave the high order risk estimates.

The incidence of breast cancer in Asia has been steadily increasing over the years (Tan, S. M. et al. 2007). The habit of betelquid chewing is known and has been reported from many Asian countries such as Sri Lanka, Bangladesh, Thailand, Cambodia, Malaysia, Indonesia, and China (Gupta, P. C. and C. S. Ray 2004). These populations report betelquid associated increase in risk for

cancers other than breast cancer also (Phukan, R. K. et al. 2001 and Wu, M. T. et al. 2004). This suggests need for investigating the mechanism of betelquid induced carcinogenesis in breast cancer.

In summary, our data provides evidence that **betel quid consumption seems to impose strong environmental effects and appears to be an independent risk factor of breast cancer.** The results also demonstrate the need for more epidemiological and genetic studies demonstrating and confirming the role of betel quid in breast cancer.

Chapter 5

Polymorphism in putative DNA repair genes BRCA2, RAD51, P53 and cell cycle gene CCND1 and assessment of breast cancer risk

Chapter 5

Polymorphism in putative DNA repair genes BRCA2, RAD51, P53 and cell cycle gene CCND1 and assessment of breast cancer risk

Introduction

Among various environmental risk factors we have found betel quid as an important risk factor for breast cancer in the Northeast population. The potential of betel quid carcinogens in causing chromosomal damage and genetic alterations have also been reported (Kaushal, M. et al 2012). Risk factors like betel quid chewing along with chemical carcinogens, alcohol and tobacco exposure produce reactive oxygen species, oxidized bases, bulky DNA adducts, and DNA strand breaks. Due to this cells are under constant mutagenic assault leading to deletions, amplifications, and/or mutations of critical genes that contribute to breast carcinogenesis (Smith, T. R. et al. 2003).

Tobacco exposure results in production of reactive oxygen species, oxidized bases, bulky DNA adducts, and DNA strand breaks. Constant mutagenic assaults from endogenous and exogenous sources lead to deletions, amplifications and mutations of critical genes (Smith, T. R. et al. 2003). DNA damage repair and cell-cycle checkpoints are two primary defense mechanisms against mutagenic exposures. The DNA double strand break repair (DSB) pathway and cell-cycle checkpoints are regulatory pathways that control double-strand breaks repair, order and timing of cell-cycle transitions to ensure the fidelity of DNA replication and chromosome segregation.

Inter individual variations and rare germ line mutations in DNA damage and repair genes have been associated with an increased risk of breast cancer. (Smith, T. R. et al. 2008). DNA damage causes cells to mediate a p53-dependent SOS response that comprises of apoptosis, cell-cycle arrest and DNA-repair(Siddique, M. and K. Sabapathy 2006).It is important to investigate contribution of common genetic variations in DNA repair and cell cycle genes to breast cancer risk in relation to DNA damage caused due to betel quid and tobacco habits. The present study examined DNA repair and cell cycle gene polymorphism (TP53 72Arg>Pro, RAD51 135G>C, BRCA2, and CCND1 G870A) along with betel quid and tobacco exposure in relation to breast cancer risk in Northeast Indian population.

5.2. Experimental methods

5.2.1. Materials

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

5.2.2. Chemicals used

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

5.2.3. Patient recruitment and sample collection

Patients with a confirmed diagnosis of breast cancer admitted in the Dr. Bhubaneswar Borooah Cancer Institute, Guwahati, Civil Hospital, Aizawl, and Sir Thutob Namgyal Memorial Hospital, Gangtok, the collaborating centers in Northeast India from the year 2005-2008 were included in the study. All subjects provided written informed consent for participation done under a protocol approved by the institutional ethics committee of Regional Medical Research Centre, North East Region (Indian Council of Medical Research).

5.2.4. Inclusion criteria

1. Incident cases during the period of December 2005 to 2008 and willing to participate in the study were included.

2. Cases confirmed by microscopy and for whom the breast was the primary site of cancer were included in the study

5.2.5. Exclusion criteria

1. Patients unwilling to give consent.
2. Patients who were too ill to participate in the study were excluded.
3. Patients who had taken any form of treatment earlier (Secondary cases) were also excluded from the study.
4. Patients with any other history of malignancy.

5.2.6. Patient details

This study consisted of 205 histopathologically diagnosed breast cancer cases registered at Dr. Bhubaneswar Borooah Cancer Institute, Guwahati and Civil Hospital,

Aizawl the collaborating centers in north east India. Incident cases during the period of December 2006 to 2009 and willing to participate in the study were included. 217 voluntary, age (±5 years) and sex matched individuals were selected from the unrelated attendants who accompanied cancer patients. This provided a readily available and cooperative source of controls from the same socio-economic background as the cases reducing confounding biases. As our collaborating centers were public hospitals, a large majority of subjects belonged to lower to middle socio-economic background. All subjects including cases and controls were resident of the north-eastern part of India at the time of recruitment for the past 5 years and belonged to the same ethnicity Demographic data and characteristics such as age, sex, smoking habit, usage of tobacco, betel quid and alcohol, were obtained from subjects in a standard questionnaire used for all the centers, in an in-person interview by a trained data collector. A majority of cases and controls were literate with full primary schooling and some up to the college level. The occupational history of the study participants revealed that most of them were farm laborers or engaged in petty jobs and the nature of such jobs did not exposed them to any occupational hazards. Any history of past or present illness was enquired or if undergoing any medication at the time of enrolment. Patients with only breast as their primary site of cancer were included. Final selected controls were included on the basis of no history of any systemic and infectious disease and those not taking any medication at the time of recruitment. Smokers, chewers and drinkers were classified into two categories ever and never.

5.2.7. Collection of blood samples

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

5.2.8. Extraction of Genomic DNA

Genomic DNA from breast cancer patients was extracted by using Himedia kit (Mumbai, India) and stored at -20^{O}C till further analyzed.

5.2.9. Quantification of Genomic DNA

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

5.2.10. Agarose Gel Electrophoresis of Extracted DNA Samples

In order to check the quality of the extracted DNA from blood samples, agarose gel electrophoresis was carried out in a 0.8% agarose gel in TAE buffer. 0.8 gram of agarose was dissolved in 100 ml of 1x TAE buffer and boiled. The solution was cooled to 45-50^0 C and 5-6 ul of EtBr was added in to the solution. Then solution was poured in to the casting tray with a comb. After solidification, gel was placed in electrophoresis tank containing 1% TAE buffer. The DNA samples (5 µl) were mixed with 6X loading dye (1 µl) and loaded into the slot/well of submerged gel. Applying a constant current of

100 mA the gel was run for up to 30 minutes. Gels were visualized under the gel documentation system and images acquired *(figure 5.1*).

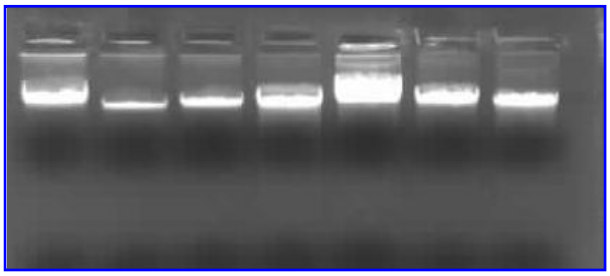


Figure 5.1: Agarose gel picture showing quality of genomic DNA isolated from the subjects

5.2.11. Genotyping of RAD51, CCND1 and p53

Polymorphisms in RAD51, CCND1 and p53 were genotyped using PCR-RFLP (Polymerase chain reaction-Restriction Fragment length polymorphism) method. Standard PCR were performed on PTC-200 (MJ Research, USA). The PCR reaction were performed in a volume of 25µl with a final concentration of 1X PCR Buffer (MBI Fermentas), 1.5mM $Mgcl_2$, 200 µM dNTPs, 2.0 mM , 1 Unit of Taq DNA polymerase and 100-300 ng of DNA Template. Negative controls were included in all PCR-runs to prevent misjudging following contamination of samples. PCR amplification consist of 35 cycles of denaturation at 94^0C for 45s; annealing at depend on gene for 45s; 72^0C for 45s followed by a final extension at 72^0C for 10 min. PCR products were loaded on 2.5% agarose gel and subjected to gel electrophoresis in 0.5X TBE buffer, stained with ethidium bromide and visualized under UV. Detail of single-nucleotide polymorphisms (SNPs) selected for the study is summarized in *Table 5.1*. Sequence of the primer and their annealing temperatures are given in the *Table 5.2*.

Table 5.1: Detail of the single-nucleotide polymorphisms (SNPs) selected for the study

Gene	Chr[a]	Loc	Polymorphism/Mutation	
			Nucleotide	Codon
RAD51	15q15.1	Gene	135 G>C	5'UTR
CCND1	11q13	Gene	A870G	242
P53	17p13.1	Exon4	215C>G	Arg72Pro

[a]Chromosomal position is based on NCBI Build.

Table 5.2: Sequence of primers used in the study

Gene	Primer sequence	T^0C	PCR (bp)
RAD51	5'- TGGGAACTGCAACTCATCTGG -3' 5'- GCGCTCCTCTCTCCAGCAG -3'	53	157
CCND1	5'- GTGAAGTTCATTTCCAATCCGC-3' 5'- GGGACATCACCCTCACTTAC-3'	57	167
P53	5'-TTGCCGTCCCAAGCAATGGATGA-3' 5'-TCTGGGAAGGGACAGAAGATGAC-3'	60	199

5.2.12. RFLP analysis of RAD51, CCND1 and p53polymorphism

Restriction digestion of the amplified fragments was carried out for the above polymorphism in a water bath (*Table 5.3*). Heat inactivation of enzyme was done at

80^0C for 20 minutes after completion of incubation with enzyme. Restriction enzymes that cleave the DNA specifically for different alleles were used (*Table 5.4*) and the alleles of each specific sample can be observed as a specific band pattern on the gel (*figure 5.2-5.4*). The genotyping results were confirmed by repeated analysis of approximately 10% of all samples randomly chosen.

Table 5.3: Standard protocol used for the RFLP experiment

COMPONENTS	**STOCK CONC.**	**WORKING CONC.**	**1 REACTION (µl)**
Water			3
Buffer	10X	1	1.5
Enzyme*	10Units/µl	5Units	0.5
PCR product			10.0
* Enzymes are specific for each polymorphism given in the table 7.5			

Table 5.4: Detail of the RFLP enzymes used for each polymorphism

Gene	**Enzyme**	**Site**	**Incubation Condition**	**PCR**	**RFLP (bp)**	**product**
					Homo wild	**Homo variant**
RAD51	MvaI	5'-GTCTC^-3' 3'-CAGAG^-5'	37^0C for 4 hrs	157	86 + 71	157
P53	BstUI	5'-CG^CG-3' 3'-GC^GC-5'	37^0C overnight	199	113+86	199
CCND1	ScrF1	5'-CC^NGG-3' 3'-GGN^CC-5'	37^0C for 4 hrs	167	145+22	167

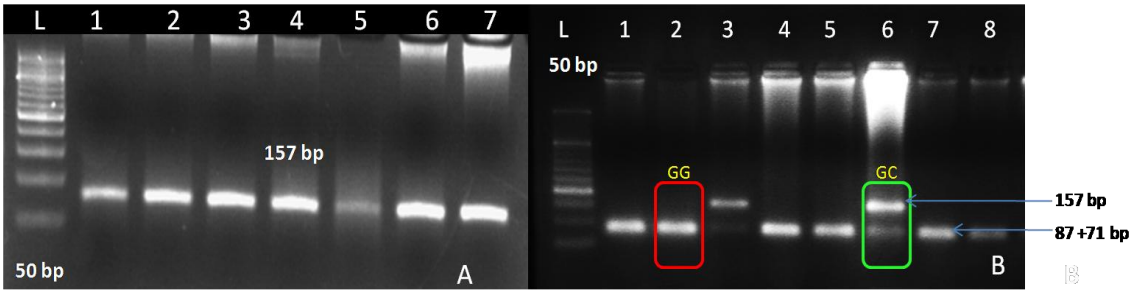


Figure 5.2: Agarose gel picture showing RFLP products of RAD51 5'UTR 135 G>C polymorphism: A- showing PCR amplification of RAD51 gene. B-RFLP of RAD51 PCR product; Lane 1,2,4,5,7 and 8-sample with 157 bp represent homozygous GG genotype; Lane 3 & 6 - samples with two bands (157bp and 87bp, 71bp (one band)) represent heterozygous GC genotype . M-50bp ladder.

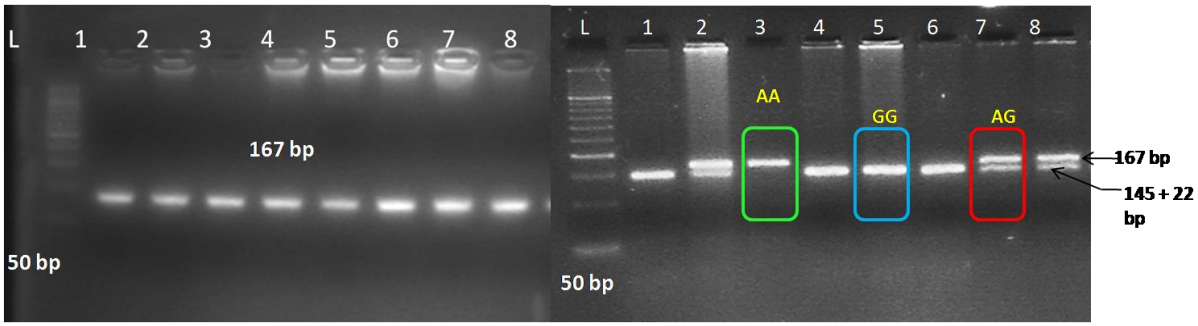


Figure 5.3: Agarose gel picture showing RFLP products of CCND1 A870C polymorphism: A- showing PCR amplification of CCND1 gene. B-RFLP of RCCND1 PCR product; Lane 3-sample with 167 bp represent homozygous AA genotype; Lane 2,7 and 8 - samples with all three bands (167bp and 145bp, 22bp (one band)) represent heterozygous AG genotype; Lane 1,4,5 & 6 - samples with two bands (145bp and 22bp(one band)) represent hhomozygous GG genotype . M-50bp ladder.

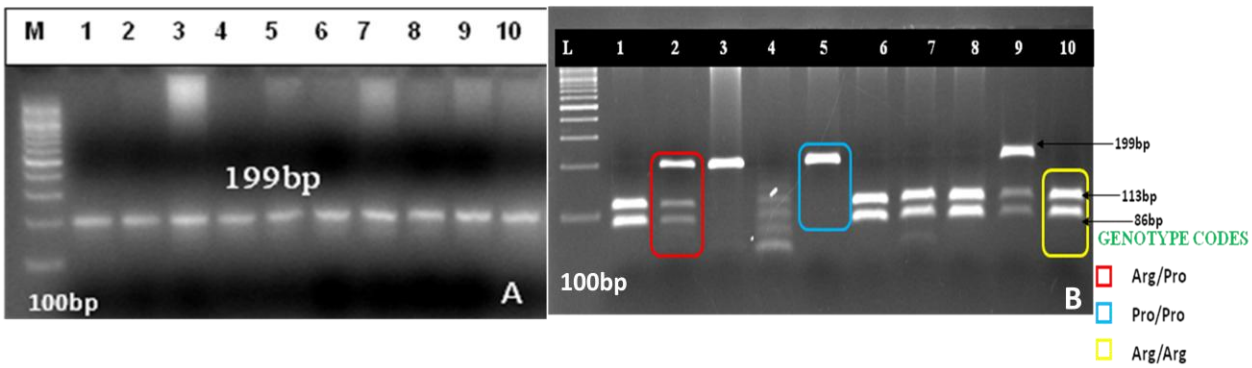


Figure 5.4: Agarose gel picture showing RFLP products of P53 codon 72 polymorphism: A- showing PCR amplification of exon 4 of p53 gene. B-RFLP of p53 PCR product; Lane 3 and 5-sample with 199 bp represent homozygous Pro/Pro allele (GG genotype); Lane 2 and 9 - samples with all three bands (199bp, 113bp, 86bp) represent heterozygous Arg/Pro allele (GC genotype); Lane 1,6,7,8 and 10- samples with two band (113 bp, 86 bp) represent homozygous Arg/Arg allele (CC genotype) . M-100bp ladder.

5.2.13. DHPLC and sequencing Analysis for BRCA2 gene polymorphism and mutation

Polymorphism and mutation in BRCA2 gene were genotyped using PCR-DHPLC-sequencing method. Standard PCR were performed on PTC-200 (MJ Research, USA). The PCR reaction were performed in a volume of 35µl with a final concentration of 1X PCR Buffer (MBI Fermentas), 1.5mM $Mgcl_2$, 200 µM dNTPs, 2.0 mM , 1 Unit of Amplitaq Gold and 100-300 ng of DNA Template. Negative controls were included in all PCR-runs to prevent misjudging following contamination of samples. PCR amplification consist of 35 cycles of initial denaturation at 94^0C for 10 mins; denaturation at 94^0C for 30s; annealing depending on the exon as mentioned in the table for 45s;

72^0C for 45s followed by a final extension at 72^0C for 10 min. PCR products were loaded on 2.5% agarose gel and subjected to gel electrophoresis in 0.5X TBE buffer, stained with ethidium bromide and visualized under UV.

Table 5.5: Detail of the single-nucleotide polymorphisms (SNPs) selected for the study

Gene	**Chr[a]**	**Loc**	**Polymorphism/ Mutation**	**Codon**
BRCA2	13q12.3	Exon2	-26G>A	5'UTR
BRCA2	13q12.3	Exon 27	10462A > G	I3412V
BRCA2	13q12.3	Exon 18	8415G > T	K2729N

Table 5.6: 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	TCTGGGTCACAAATTTGTCTG TCA	TGATTTGCCCAGCATGACAC	55	418
4	AGAATGCAAATTTATAATCCA GAGTA	AAATCAGAT TCATCTTTATAGAACAAA	50	249
5	AACAATTTATATGAATGAGAA TC	AATTGTTAAGTTTTATTTTTATTA	50	220
6	CCACAAAGAGATAAGTCAGG TA	TGTAAATCTCAGGGCAAAGGTA	55	234
7	TAAGTGAAATAAAGAGTGAA	AACAGAAGTATTAGAGATGAC	50	275
8	TGCTTTTTGATGTCTGACAA	ACATATAGGACCAGGTTTAGAGAC	60	274
9	CTAGTGATTTTAAACTATAATT TTG	GTTCAACTAAACAGAGGACT	50	264
10A	TGCCAAGTACTCAGAATAAC	CTTTTTGATACCCTGAAATGAAGAA	60	864

	CC	G		
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	CAGAGAATAGTTGTAGTTGTTGAA	AGAAACCTTAACCCATACTGC	55	306
18	GTGACTTGTTTAAACAGTGGAA	ATTGAGCATCCTTAGTAAGCA	48	524
19	AAGTGAATATTTTTAAGGCAGTT	TATATGGTAAGTTTCAAGAAT	50	342
20	CACTGTGCCTGGCCTGATAC	TGTCCCTTGTTGCTATTCTTT	55	401
21	AATCTCCCTTCTTTGGGTGT	CATTTCAACATATTCCTTCCTG	60	318
22	TTTTGTTCTGATTGCTTTTTATTC	AATCATTTTGTTAGTAAGGTCAT	50	314
23	CCACTACTAATGCCCACAAA	AAAACAAAACAAAAATTCAACATA	55	367

24	CAGTTTTGATAAGTGCTTGTT	AGCTGGAACTAATCATAAGA	50	290
25	TTAGAGTTTCCTTTCTTGCATC	AAGCTATTTCCTTATACTGGA	55	399
26	AAGGAAATACTTTTGGAAACATAA	TTTACTAGGTATACAACAGAA	50	299
27A	TAGGAGTTAGGGGAGGGAGACTGTGT	CAAGGCTCTTCTCTTTTTGC	55	294
27B	CTGTCTCAGCCCAGATGACT	TGTTGAACCAGACAAAAGAGC	58	344
27C	TCAATGAAATTTCTCTTTTGGA	TGTGTGGTTTGAAATTATATTC	50	345

DHPLC analysis was performed on a wave DNA fragment analysis system (Transgenomic, San Jose, CA, USA). Each PCR product was denaturated for 5 min at 95 ^{0}C and then gradually re-annealed by decreasing the sample temperature from 95 to 25^{0}C (with a temperature ramp of 1·75 ^{0}C/min) over a period of 40 min. Two to five µl of PCR product was then applied to a pre-heated C18 reversed-phase column based on non porous poly(styrene-divinylbenzene) particles (DNASep, Transgenomic) and eluted with a mobile phase consisting of a mixture of 0·1 M triethylammonium acetate pH 7·0 (TEAA) (Transgenomic) and 0·1 M TEAA with 25% acetonitrile (Transgenomic) in a linear gradient, at a flow rate of 1·5 ml/min. Gradient parameters were determined by size and GC content of the amplicons.

Heteroduplexes and homoduplexes were detected by monitoring the absorbance at 260 nm. The condition for mutation analysis is based on the melting behaviour of the wild type sequence of the DNA fragment rather than by its length. Temperature for successful resolution of heteroduplex molecules was determined by using the WaveMaker software (Transgenomic) and the Stanford DHPLC program. From the

sequence of each fragment, the algorithm calculates the melting behaviour, the optimal analysis temperature for each domain corresponding to 80–90% of α-helical fraction. During the optimization phase, a few mutated samples for each DNA fragment were additionally run up to 2 ^{0}C above and below of the predicted temperatures, with 0·1 ^{0}C increments, to select the optimal temperatures for detection of sequence variations.

PCR products with heterozygous peaks were sequenced to identify the type of mutation. Amplified fragments were cut and extracted from agarose gels using Qiagen mini elute kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Approximately 100 ng purified PCR product was directly sequenced with 3.3 pmol primers as described above with the Big Dye Terminator Cycle Sequencing Kit v3.1. Sequence analysis was performed on an ABI 3130xl Sequence Detection System (Applied Biosystems, USA). Details of single-nucleotide polymorphisms (SNPs) and mutation detected in the study is summarized in *Table 5.5*. Sequence of the primer and their annealing temperatures are given in the *Table 5.6.* The homoduplex and heteroduplex peaks along with sequencing are depicted in *figures 5.5 – 5.10*

Table 5.7: Standard protocol used for the sequencing experiment

Component	**Volume**
Template DNA (300-600 ng)	x µl
Primer (3.2 uM)	1 µl
ABI BigDye terminator v3.1*	1 µl
5x Sequencing Buffer	3.5 µl
Milli-Q water	y µl
Total volume	20 µl

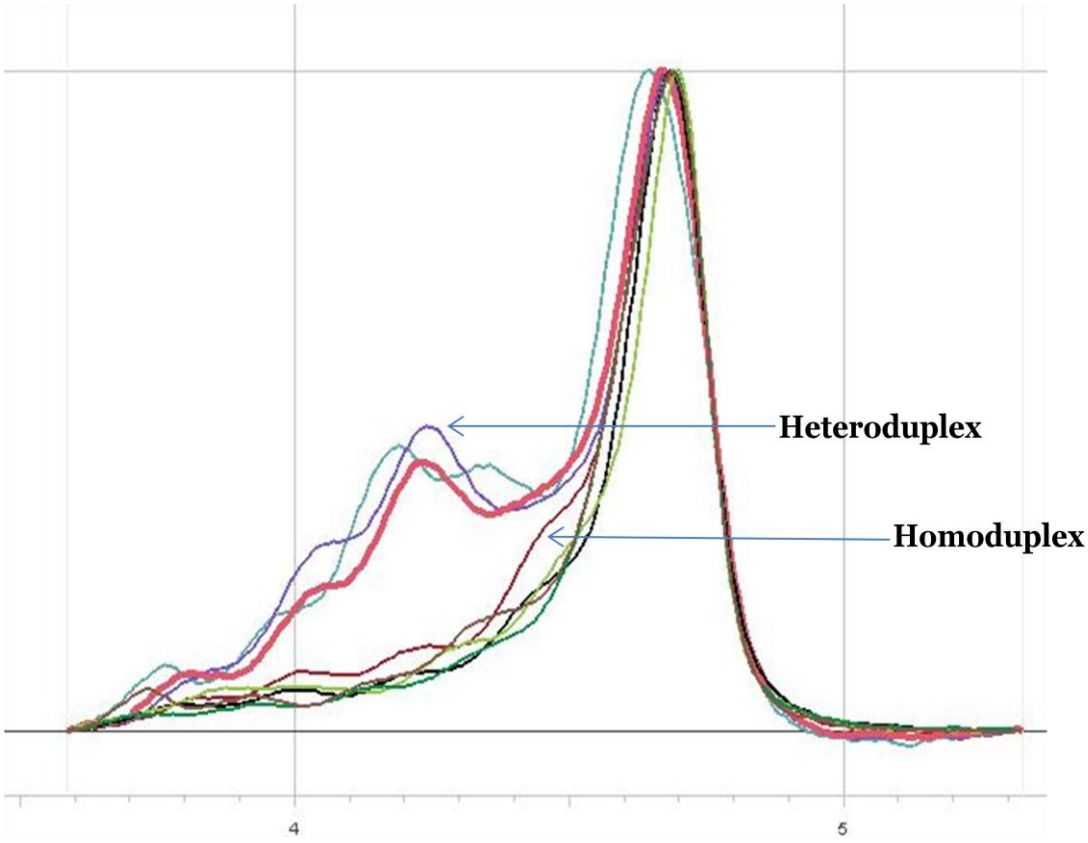


Figure 5.5: DHPLC analysis showing heteroduplex peaks depicting GA and AA genotype and homoduplex peaks depicting GG genotype in Exon2 of BRCA2 gene. The polymorphic variant was identified as a missense variant -26 G > A in Exon2 of BRCA2 gene

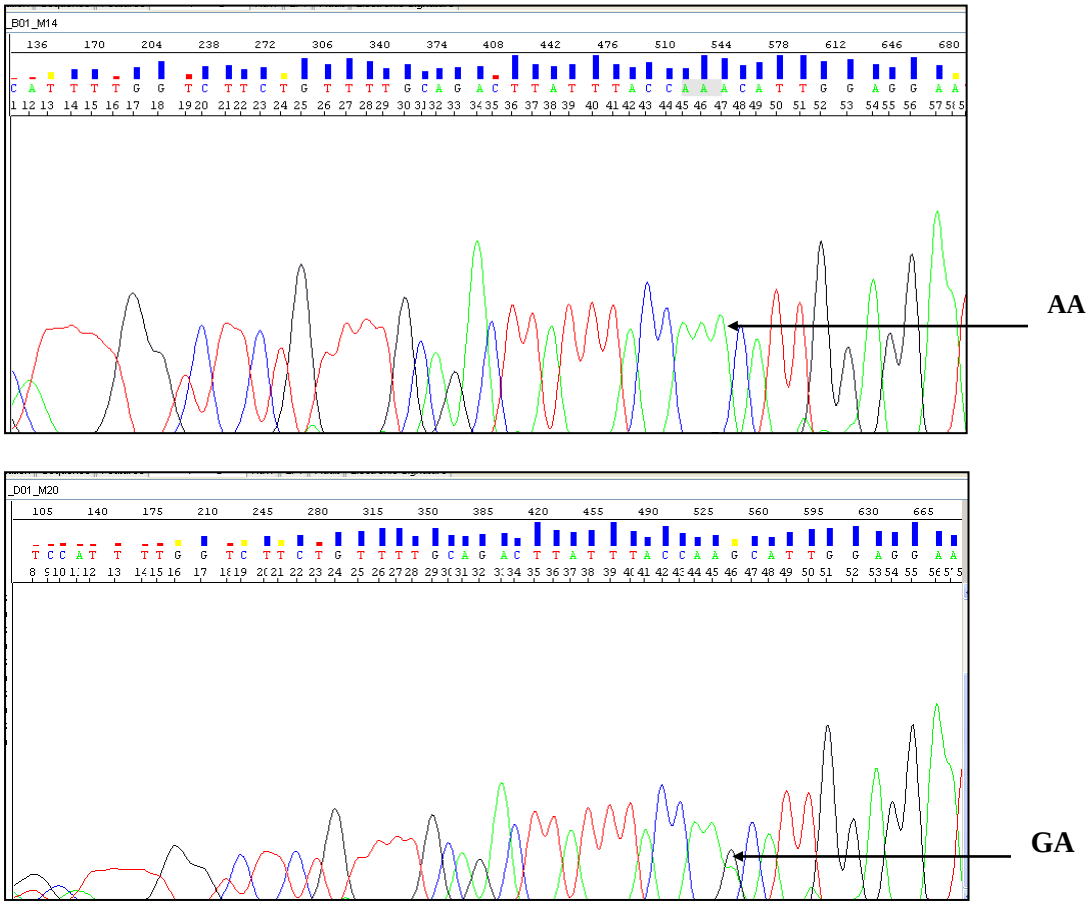


Figure5.6: Sequencing analysis of samples with Exon 2 variants depicting GA heterozygous and AA homozygous genotype of BRCA2 gene.

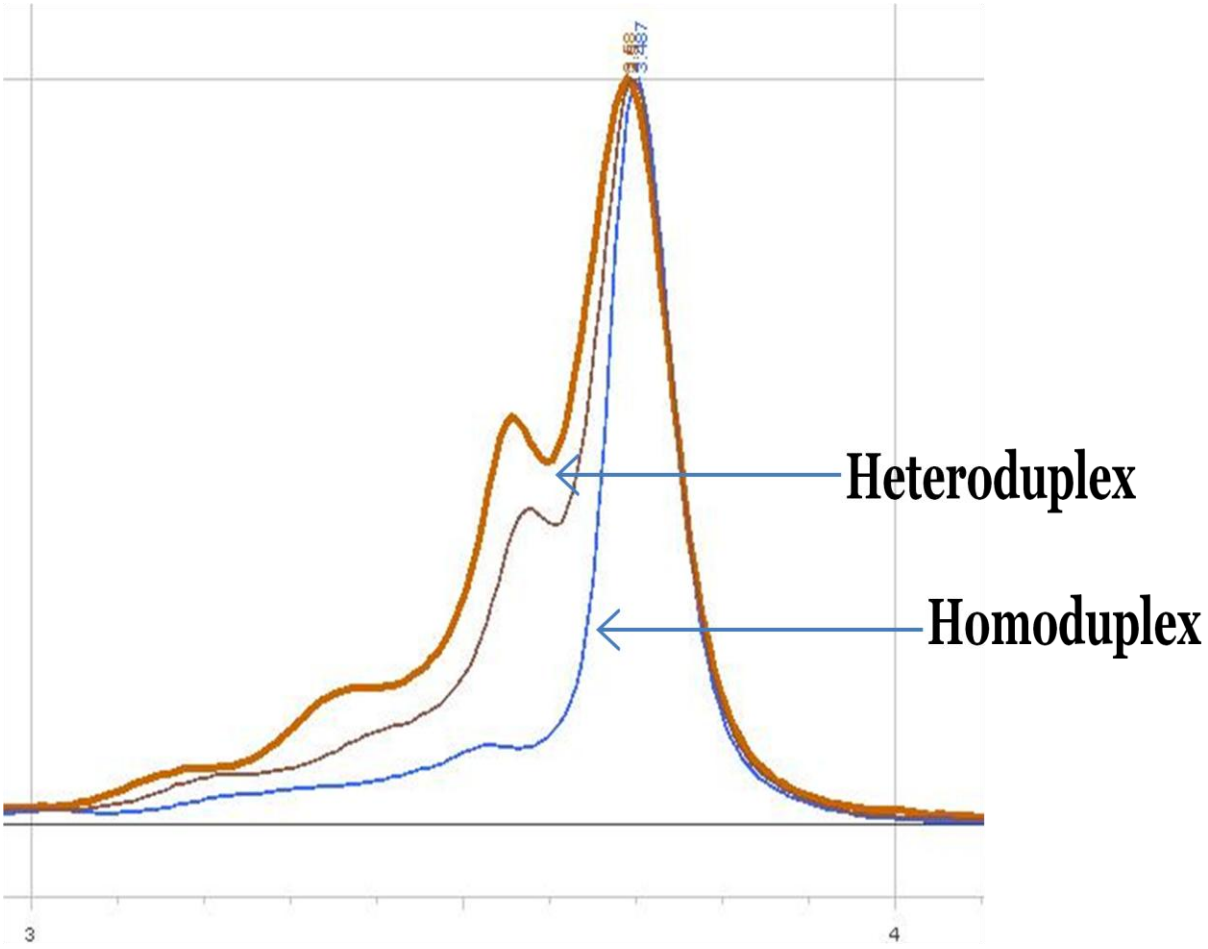


Figure 5.7: DHPLC analysis showing heteroduplex peaks depicting CT and TT genotype and homoduplex peaks depicting CC genotype in Exon18 of BRCA2 gene. The variant was identified as a 8415G > T: K2729N in Exon18 of BRCA2 gene

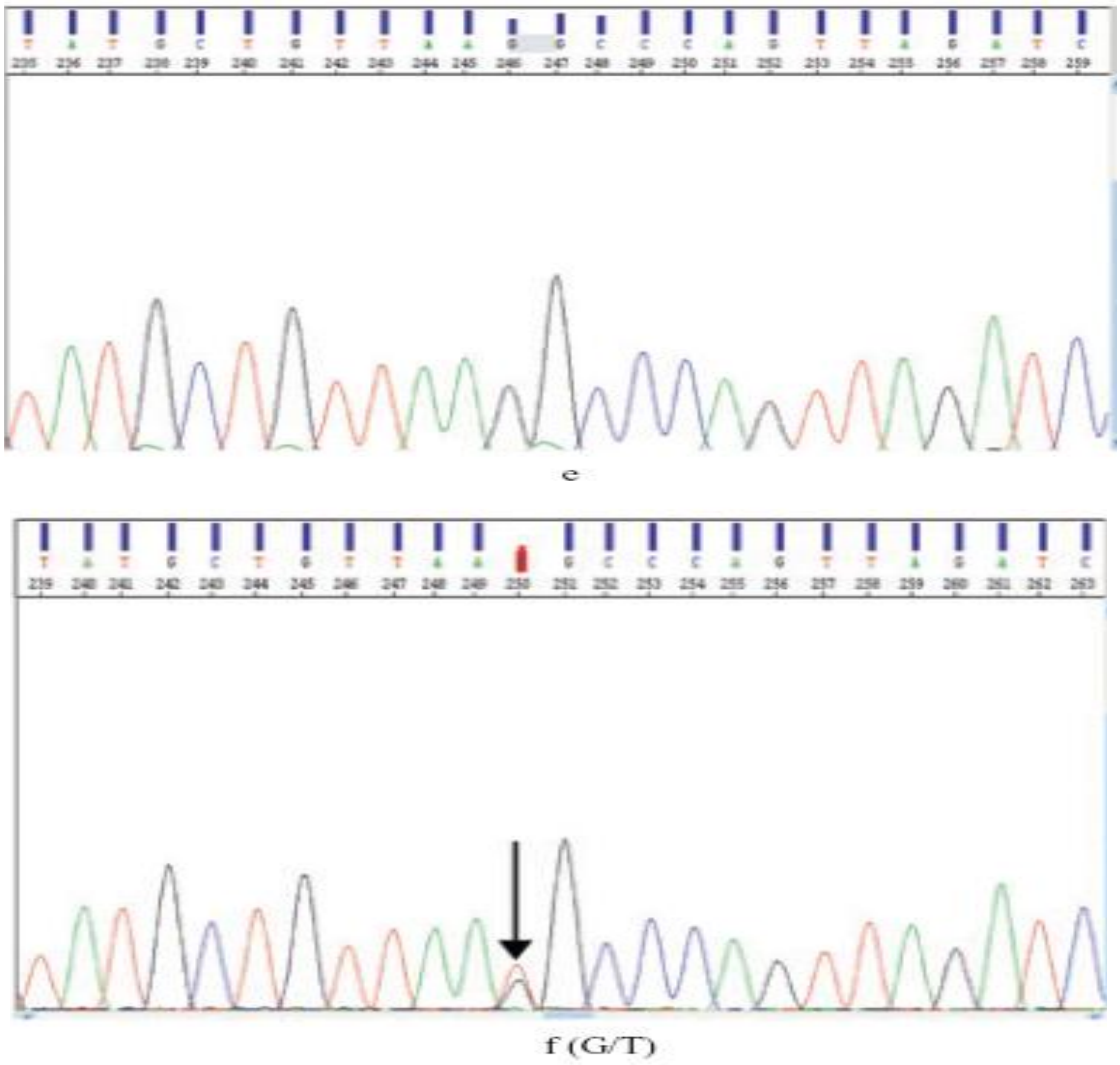


Figure 5.8: Sequencing analysis of samples with Exon 18 variants depicting GT heterozygous and GG homozygous genotype of BRCA2 gene.

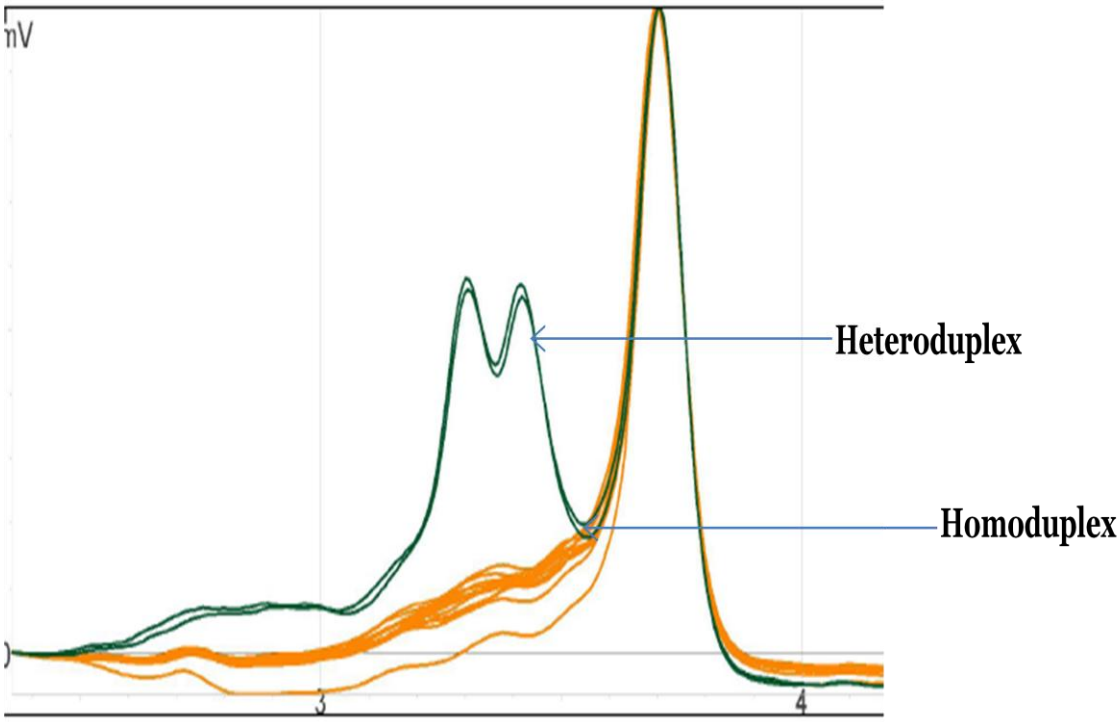


Figure 5.9: DHPLC analysis showing heteroduplex peaks depicting AG and GG genotype and homoduplex peaks depicting AA genotype in Exon27 of BRCA2 gene. The polymorphic variant was identified as 10462A > G: I3412V

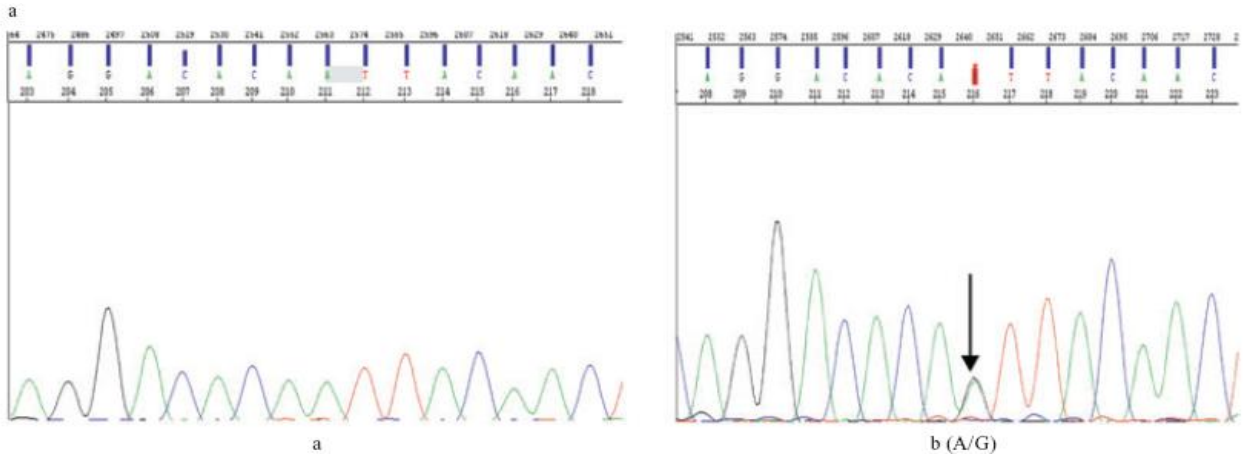


Figure 5.10: Sequencing analysis of samples with Exon 27 variants depicting AG heterozygous and AA homozygous genotype of BRCA2 gene.

5.3. Statistical Analysis

Cases were matched with controls on the basis of age (±5 years), sex and ethnicity. Difference in the distribution of demographic characteristics and genotype frequencies between cases and controls were evaluated using the Chi Square (χ^2) and Fisher's Exact test wherever appropriate. Hardy–Weinberg equilibrium (HWE) was assessed by using the χ^2-test. Estimates of risk to cancer, imparted by genotypes and other covariates as tobacco smoking, tobacco chewing, betel quid chewing and alcohol consumption were determined by deriving the odds ratio (OR) and its corresponding 95% confidence interval (95% CIs) using multivariable conditional logistic regression. For all the tests a two sided $p<0.05$ was considered statistically significant. The data analysis was performed on the SPSS Version 16 software package.

5.3.1 Identification of High Order Interactions

High order interactions were determined using CART and MDR

5.3.1.1. Multifactor dimensionality reduction (MDR)

Multifactor dimensionality reduction (MDR) method as discussed in chapter 4 is non-parametric, genetic model-free method for overcoming some of the limitations of logistic regression (i.e. sample size limitations) for the detection and characterization of gene–gene interactions (Hahn, L. W. et al. 2003).

5.3.1.2. Classification and regression tree (CART) analysis

Classification and regression tree (CART) analysis was performed using the SPSS ver. 16 software to build a decision tree (Srivastava, A.et al 2012) . Decision tree was created by splitting a node into two child nodes repeatedly, beginning with the root node that contains the total sample. Before constructing a tree, we chose measure for goodness of split using Gini criteria, by which splits were found that maximize the homogeneity of child nodes with respect to the value of the target variable. After the tree was grown to its full depth, a pruning procedure was performed to avoid over fitting the model. Finally the risk of various genotypes was evaluated by using the LR analysis. The ORs and 95% CIs were adjusted for age, with treating the <40 case ratio as the reference.

5.3.2. False Positive Report Probability (FPRP)

Reports of gene-environment interaction studies are often challenged by false positive discoveries especially when results are generated by multiple comparisons. To estimate the FPRP and to evaluate robustness of the findings from MDR analysis we used the Bayesian approach described by Wacholder et. al. [Ihsan, R.,et al.]. Considering poor epidemiological data from the study population and inconsistent association of the SNPs with breast cancer risk we set a fairly wider range of prior probabilities (10^{-6} to 10^{-1}) with an estimated statistical power to detect an OR of 0.1 and 0.2 and 0.3 and α level equal to the observed p-value. The FPRP cutoff point was stringently kept to 0.5.

5.4. Results

The mean age was 45.5±12.86 years for the cases and 45.98±14.44 years for the controls. No significant difference in distribution of age (p=0.07), tobacco smoking (p = 0.16) and tobacco chewing (p = 0.78) was seen between cases and controls and were not found associated with breast cancer risk. However, women with a betel quid chewing history and alcohol consumption were found significantly associated with the risk of developing breast cancer (p<0.001 and (p=0.003 respectively) (*Table 5.8*).The distribution of all SNPs in cases and control was in agreement with HWE (p>0.05), except alleles of CCND1 polymorphism in control did not follow HWE (p<0.001).

Table 5.8: Characteristics of case and control

	Case 205 (%)	Control 217 (%)	O.R (95%CI)	P VALUE	χ^2P value
Distribution of Age					
≤29	18 (0.08)	29 (13.3)			0.07
30-39	44 (21.4)	39 (19)			
40-49	72 (35.1)	58 (26.7)			
50-59	33 (16)	53 (24.4)			
≥60	38 (18.5)	38 (17.5)			
Tobacco smoking					
no	179 (87.3)	179 (82.5)			0.167
yes	26 (12.7)	38 (17.5)	0.60 (0.32-1.10)	0.10	
Tobacco chewing					
no	123 (60)	133 (61.3)	1.0		0.786
yes	82 (40)	84 (38.7)	0.78 (0.49-1.22)	0.28	
Betel quid chewing					
no	54 (26.3)	134 (61.8)	1.0		**<0.001**
Yes	151 (73.7)	83 (38.2)	**4.98 (3.15-7.87)**	**<0.001**	
Alcohol consumption					
No	187 (91.2)	212 (97.7)			**0.003**
Yes	18 (8.7)	5 (2.3)	2.59 (0.88-7.55)	0.08	
Family History					
No Family History	158	95			
Family History of any cancer	18	52			
Unknown	29	70			

5.4.1. Association of genetic and environmental factors with breast cancer risk by LR analysis

The distribution and main effects of genetic and environmental factors are summarized in Table I and II. Betel quid chewing was significantly associated with breast cancer risk (OR = 4.98 (3.15-7.87); p<0.001. Genotype distribution of CCND1 and TP53 polymorphism showed significant variation in cases and controls (*Table 5.9).* Main effects of genotypes were evaluated using multivariable LR. Both homozygous and heterozygous genotype (AA, AG) and dominant model (AA+AG) of CCND1 showed a protective trend towards breast cancer risk ((0.28 (0.14-0.57) p=<0.001, 0.37(0.20-0.68) p=0.002 and 0.34(0.18-0.62) p=<0.001respectively). The 'A' allele was also found underrepresented in the breast cancer case population (0.68(0.51-0.89), p=0.006). The Pro/Pro genotype of TP53 also showed a protective trend towards breast cancer risk (0.52(0.28-0.95), p=0.03). The Pro allele of TP53 was also found significantly underrepresented in the case population (0.76(0.58-1.00), p=0.05) (*Table 5.9*). No significant association of breast cancer risk was observed in women with RAD51 polymorphism. Screening for mutation in BRCA2 gene showed presence of 8415G > T: K2729N mutation in Exon 18 in two cases. The variant AA genotype frequency for -26 G>A polymorphism in exon2 was found 17.1% cases and 14.7% controls. The variant GG and AG genotype frequency of 10462A > G: I3412V in exon27 were 4.9% cases and 2.8% controls. The BRCA2 polymorphisms were not found associated with breast cancer risk.

Table 5.9: Distribution of genotypes of DNA repair and cell cycle genes amongst case and control

Gene	Case	Control	$OR_{adjusted}$	P value	χ^{2} p value
RAD51					
GG	117 (57.1)	134 (61.8)	1.0		0.59
GC	76 (37.1)	73 (33.6)	1.29 (0.82-2.04)	0.26	
CC	12 (5.9)	10 (4.6)	1.46 (0.53-3.98)	0.45	
GC+CC			1.31 (0.84-2.04)	0.22	
G	310 (0.75)	341 (0.78)	1.0		
C	100 (0.24)	93 (0.21)	1.18 (0.84-1.65)	0.32	
CCND1					
GG	48 (23.4)	25 (11.5)	1.0		**0.004**
GA	108 (52.7)	125 (57.6)	**0.37 (0.20-0.68)**	**0.002**	
AA	49 (23.9)	67 (30.9)	**0.28 (0.14-0.57)**	**<0.001**	
AG+AA			**0.34 (0.18-0.62)**	**<0.001**	
G	204 (0.50)	175 (0.60)	1.0		
A	206 (0.50)	259 (0.40)	**0.68(0.51-0.89),0.006**		
TP53					
Arg/Arg	53 (25.9)	41 (18.9)	1.0		
Arg/Pro	99 (48.3)	106 (48.8)	0.69 (0.40-1.20)	0.19	0.15
Pro/Pro	53 (25.9)	70 (32.3)	**0.52 (0.28-0.95)**	**0.03**	
Arg/Pro+ Pro/Pro			0.62 (0.37-1.04)	0.72	
Arg	205 (0.50)	188 (0.43)	1.0		
Pro	205 (0.50)	246 (0.57)	**0.76(0.58-1.00),0.05**		
BRCA2 EXON2					
GG	74 (36.1)	76 (35.0)	1.0		0.72
GA	96 (46.8)	109 (50.2)	0.98 (0.61-1.60)	0.96	
AA	35 (17.1)	32 (14.7)	1.09 (0.58-2.08)	0.77	
GA+AA			1.01 (0.64-1.59)	0.94	
G	244 (0.60)	261 (0.60)	1.0		
A	166 (0.40)	173 (0.40)	1.02 (0.77-1.36)	0.85	
BRCA2 EXON27					
TT	195 (95.1)	211 (97.2)	1.0		0.25
TC + CC	10 (4.9)	6 (2.8)	1.89 (0.58-6.17)	0.29	
BRCA2 EXON18					
GG	203 (99)	217 (100)			0.14
GT + TT	2 (1.0)	0			

5.4.2. Risk associated with SNPs stratified by betel quid chewing

Data was further stratified by betel quid chewing as it was the strongest independent risk factor in LR. Stratification of risk associated with genetic factors among betel quid chewers (BQC) and non betel quid chewers (NBQC) is shown in Table III. The AA genotype and dominant model (AA+AG) of CCND1 showed protection towards breast cancer risk in BQC (**0.28(0.10-0.77), 0.01** and **0.32(0.12-0.81), 0.01**) and NBQC (**0.26 (0.09-0.78), 0.01** and **0.37(0.16-0.87), 0.02**). In addition GA genotype of CCND1 was associated with protection towards breast cancer risk in BQC subset (0.34(0.13-0.90), p=0.03). The 'A' allele of CCND1 was also significantly underrepresented in the breast cancer cases in both BQC and NBQC subsets (0.64 (0.43-0.95), p=0.02 and 0.57 (0.36-0.89), p=0.01 respectively). The Pro/Pro genotype and Pro allele of TP53 showed protection towards breast cancer in NBQC (0.29(0.10-0.81), p=0.01) and (0.51(0.32-0.80), p=0.003 respectively). In the BQC group, the C allele of RAD51 was overrepresented in cases and associated with breast cancer risk (2.03 (1.26-3.30) 0.002). Two cases showing BRCA2 8415G > T: K2729N mutation in Exon 18 belonged to BQC (*Table 5.10*).

Table 5.10: Distribution of genotypes of DNA repair and cell cycle genes amongst case and control in two sample subsets (NBQC and BQC)

Gene	BQC Case/Control 151/83 (n,%)	OR (95%CI),p value	NBQC Case/Control 54/134 (n,%)	OR (95%CI),p value
RAD51				
GG	85/55 (56.3/66.3)	1.0	32/79 (59.3/59)	1.0
GC	56/24 (37.1/28.9)	1.50 (0.81-2.78) 0.19	20/49 (37/36.6)	1.07 (0.50-2.26) 0.85
CC	10/4 (6.6/4.8)	1.81 (0.51-6.42) 0.35	2/6 (3.7/4.5)	1.65 (0.27-9.82) 0.58
GC+CC		1.55 (0.86-2.77) 0.14		1.11 (0.53-2.30) 0.77
G	226/194	1.0	84/207	1.0
C	76/32	**2.03 (1.26-3.30) 0.002**	24/61	0.97 (0.54-1.71)
CCND1				
GG	33/7 (21.9/8.4)	1.0	15/18 (27.8/13.4)	1.0
GA	77/47 (51/56.6)	**0.34(0.13-0.90), 0.03**	31/78 (57.4/58.2)	0.43(0.17-1.06), 0.06
AA	41/29 (27.2/34.9)	**0.28(0.10-0.77), 0.01**	8/38 (14.8/28.4)	**0.26 (0.09-0.78), 0.01**
GA+AA		**0.32(0.12-0.81),0.01**		**0.37(0.16-0.87), 0.02**
G	143/61	1.0	61/114	1.0
A	159/105	**0.64 (0.43-0.95) 0.02**	47/154	**0.57 (0.36-0.89),0.01**
TP53				
Arg/Arg	34/17 (22.5/20.5)	1.0	19/24 (35.2/17.9)	1.0
Arg/Pro	74/42 (49/50.6)	0.92(0.44-1.92),0.82	25/64 (46.3/47.8)	0.59 (0.26-1.36),0.22
Pro/Pro	43/24 (28.5/28.9)	0.86(0.38-1.93), 0.72	10/46 (18.5/34.3)	**0.29(0.10-0.81), 0.01**
Arg/Pro+Pro/Pro		0.90(0.45-1.80), 0.76		0.47(0.22-1.04), 0.06
Arg	142/76	1.0	63/112	1.0
Pro	160/90	0.95 (0.65-1.39), 0.79	45/156	**0.51(0.32-0.80),0.003**
BRCA2 EXON2				
GG	55/29	1.0	19/47	1.0

	(36.4/34.9)		(35.2/35.1)	
GA	67/41	0.91 (0.48-1.73) 0.79	29/68	1.13 (0.53-2.41)
	(44.4/49.4)		(53.7/50.7)	0.74
AA	29/13	1.48 (0.63-3.45) 0.36	6/19	0.59 (0.18-1.97)
	(19.2/15.7)		(11.1/14.2)	0.39
GA+AA		1.04 (0.57-1.90) 0.88		1.00 (0.48-2.09)
				0.98
G	177/99	1.0	67/162	1.0
A	125/67	1.05 (0.70-1.57) 0.84	41/106	0.93 (0.57-1.51)
				0.81
BRCA2 EXON27				
AA	144/81	1.0	51/130	1.0
	(95.4/97.6)		(94.4/97)	
AG + GG	7/2 (4.6/2.4)	1.47 (0.26-8.18)	3/4 (5.6/3)	2.46 (0.49-
				12.25) 0.27
BRCA2 EXON18				
GG	149/83		54/134	
	(98.7/100)		(100/100)	
GT + TT	**2/0 (1.3/0)**		NIL	

5.4.3. MDR Analysis

MDR analysis was applied to further explore gene-gene and gene-environment interactions. Best predictive models up to 4 orders of interaction, along with their CVC, TBA were chosen. For total data set, betel quid chewing was the best one locus model with CVC of 10/10 and testing accuracy of 0.6770 which was statistically significant ($p<0.001$) determined by 1000 fold permutation testing. For a 2-locus interaction, combination of betel quid chewing and alcohol consumption was most significant with CVC of 8/10 and TBA of 0.6673 ($p<0.001$). The 3 locus model consisted of tobacco smoking, tobacco chewing and betel quid chewing with CVC of 10/10 and TBA of 0.6952 ($p<0.001$). The best 4 locus model which included genes consisted of RAD51

TP53 tobacco smoking and betel quid chewing with a CVC of 10/ 10 and TBA of 0.6869 (<0.001). MDR analysis performed in NBQC showed a best 4 locus models with TBA 0.6765 (0.005) and CVC of 10/10 (*Table 5.11*). No interaction models were obtained for the BQC subset.

Table 5.11: Multifactor Dimensionality Reduction Analysis (MDR) revealing inteactions

	No. of Locus	Model	TBA	CVC	P for permutation testing
Total Data Set	1st Order	Betchew	0.6770	10/10	<0.001
	2nd Order	Betchw Alc	0.6718	9/10	<0.001
	3rd Order	Tbsmk Tbchew Bqchew	0.6975	10/10	<0.001
	4th Order	**RAD51 TP53 Tbsmk Bqchew**	**0.6869**	**10/10**	**<0.001**
NBQC	1st Order	TP53	0.5112	8/10	0.76
	2nd Order	RAD51 CCND1	0.5837	7/10	0.30
	3rd Order	RAD51 CCND1 EX2	0.6007	6/10	0.18
	4th Order	**RAD51 CCND1 EX2 TP53**	**0.6765**	**10/10**	**0.005**

BQc: Betel quid chewer; Tbsmk: tobacco smokers; Alc: alcoholics

5.4.4. Interpretation of the results using information gain and interaction graphs

As shown in the hierarchical interaction graphs in figure 5.1, for total sample set betel quid chewing large independent effect (9.38%) among environmental factors. A strong interaction (1.32%) was seen between RAD51 and TP53. Similar to total data set, the NBQC (Figure 5.2) depicted a large part of interaction was seen in between TP53 and

RAD51 (1.32%). CCND1 had a large independent effect (1.89%) in NBQC. In addition small percentages of the entropy in case–control status explained by TP53 (0.64%), or EX2BRCA2 (0.11%) considered independently, but a large percentage of entropy explained by the interaction between these two loci (1.02%) were found.

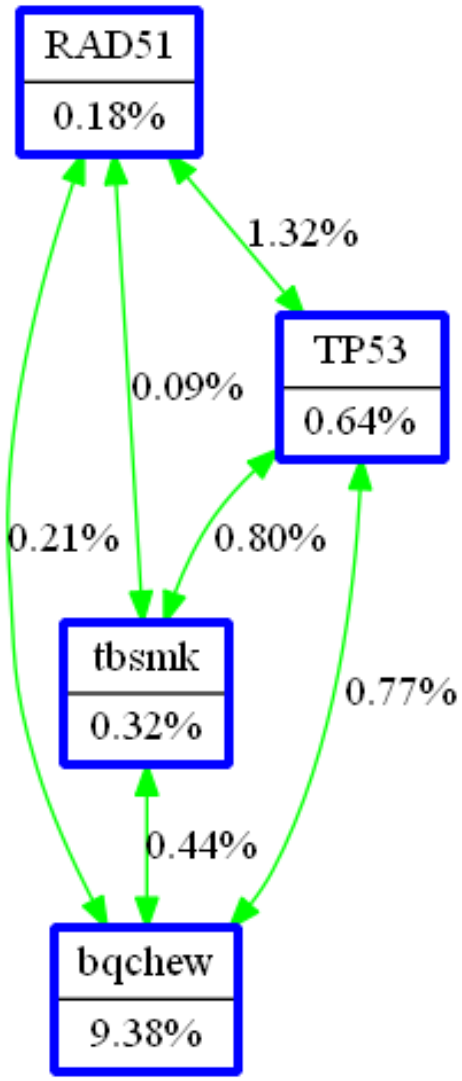


Figure 5.11: Interaction dendogram using orange software for the total data set

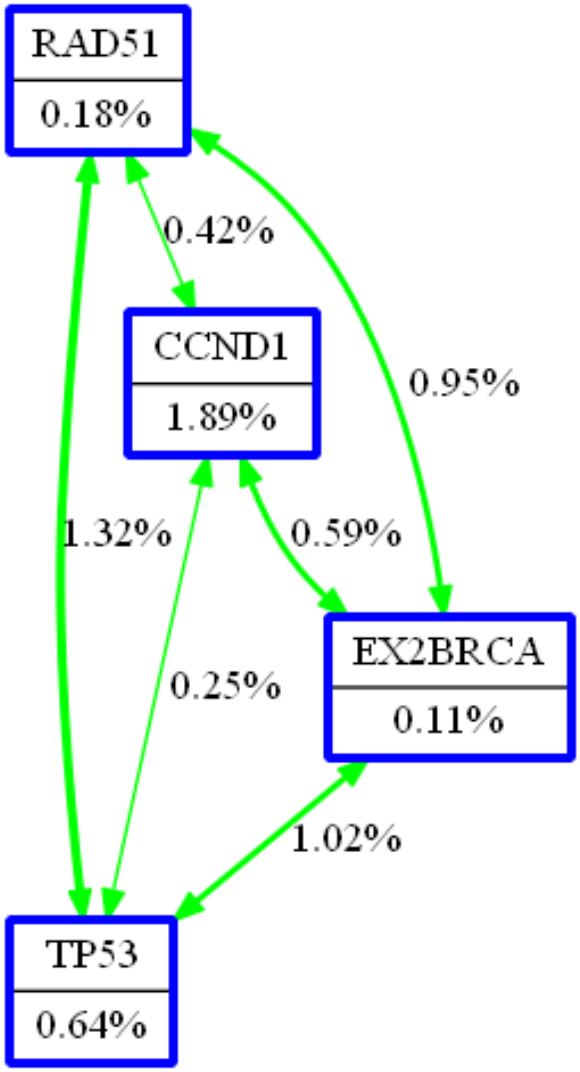


Figure 5.12: Interaction dendogram using orange software for the NBQC data set

5.4.5. CART analysis

Figure 5.3 show the selected CART model constructed on all investigated genetic variants and environmental risk factors. The final tree contained nine terminal nodes. The first split of the root node was on betel quid chewing, indicating that it is the strongest risk factor for breast cancer. Among BQC, the subsequent splits showed

interactions between CCND1, tobacco smoking, alcohol and TP53. In NBQC first split was TP53 which was seen to interact with tobacco chewing and BRCA G>A polymorphism. Terminal node 14 comprising of least percentage of cases was taken as reference to calculate OR for other terminal nodes. Among betel quid chewers significant risk was observed for terminal node 3 consisting of CCND1 GG genotype (OR = 33.0;95%CI = 6.08-179.07), p=<0.001) followed by terminal node 11 (BQC, CCND1 GA,AA, No Smk, Alc) (OR = 42.00;95%CI = 5.11-345.11, p=<0.001). Risk was also observed in Nodes 15 (BQC, CCND1 GA,AA, No Smk, Non Alc, Pro/Pro;Arg/Arg) (OR=14.84;95%CI=3.13-70.34, p= <0.001)and Node 16 (BQC, CCND1 GA,AA, No Smk, Non Alc, Arg/Pro) (OR=9.40;95%CI=1.99-44.34, p= <0.001).In NBQC group risk was seen for terminal node 5 comprising of NBQC and TP53 Arg/Arg (OR = 5.54; 95%CI = 1.11-27.42, p=0.03).

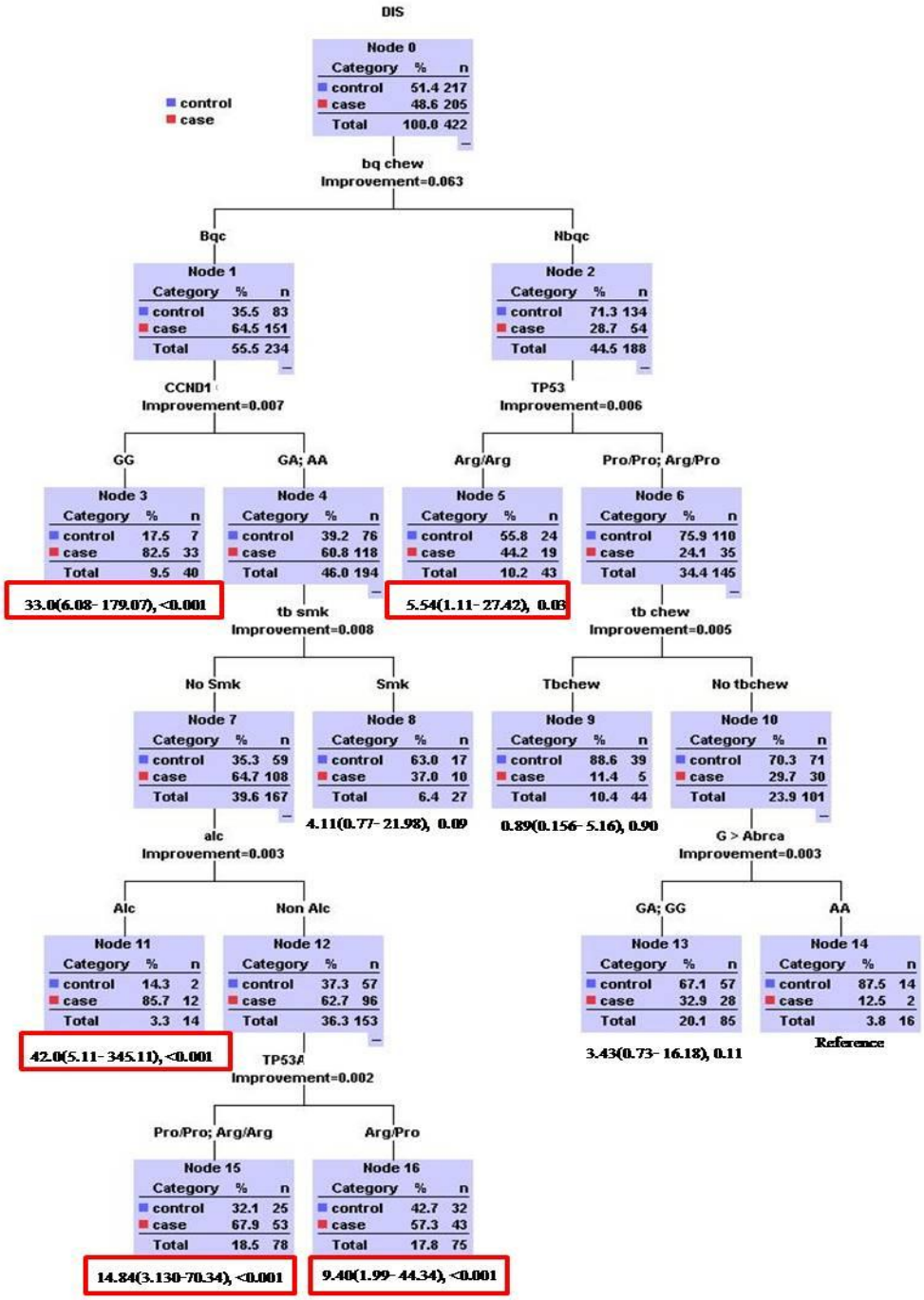


Figure 5.13: CART analysis for the DNA repair and cell cycle genes

5.5.6.False positive report probability (FPRP)

Table 5.14 shows the FPRPs for BQC and NBQC obtained from LR analysis. It reports the FPRP values calculated using the statistical power to detect an OR of 0.1, 0.2 and 0.3 with α level equal to the observed p value. Results show a reliability on CCND1 in the BQC and NBQC with prior probabilities (0.05 and 0.025) for both OR = 0.1 and 0.2. TP53 also showed a good reliability with prior probabilities (0.05 and 0.025) for both OR = 0.1 and 0.2. In addition with a prior probability of 0.05 at OR =0.3, CCND1 for BQC and NBQC and TP53 for NBQC gave reliable results.

Table 5.12: False positive report probability (FPRP) results

Logistic Regression	OR 95% CI	P value	OR = 0.1 power	Prior probability				
				0.05	0.025	0.01	0.001	0.0001
BQC								
CCND1	**0.28(0.10-0.77)**	0.01	0.977	**0.210**	**0.353**	0.580	0.933	0.993
NBQC								
CCND1	**0.26 (0.09-0.78)**	0.01	0.956	**0.244**	**0.399**	0.627	0.944	0.994
TP53	**0.29(0.10-0.81)**	0.01	0.979	**0.261**	**0.420**	0.648	0.949	0.995

Gene	OR = 0.2 power	Prior probability					OR = 0.3 power	Prior probability				
		0.05	0.025	0.01	0.001	0.0001		0.05	0.025	0.01	0.001	0.0001
BQC												
CCND1	0.733	**0.259**	**0.417**	0.645	0.948	0.995	0.447	**0.367**	0.544	0.751	0.968	0.997
NBQC												
CCND1	0.680	**0.312**	**0.482**	0.703	0.960	0.996	0.399	**0.436**	0.613	0.801	0.976	0.998
TP53	0.761	**0.312**	**0.482**	0.703	0.960	0.996	0.474	**0.421**	0.599	0.791	0.975	0.997

5.6. Discussion

Betel quid chewing showed strong association in LR, best one factor model in MDR and formed first split in CART implying its importance as breast cancer risk factor in the northeast population of India. Betel quid is composed of areca nut (Areca catechu), catechu (Acacia catechu), slaked lime (calcium oxide and calcium hydroxide), wrapped in a betel leaf (Piper betel) and tobacco whose each component metabolizes producing potent carcinogens. Tobacco chewing with betel quid results in high exposure to carcinogenic tobacco-specifc nitrosamines (~1000 mg/day, compared with ~20 mg/day in smokers). Many chewers swallow the quid that contains precursors of nitrosamines. The acidic pH of stomach favors nitrosation of secondary and tertiary amines in the quid. Reactive oxygen species (ROS) are generated in the oral cavity during chewing (Nair, U. et al. (2004)). Since metabolic absorption of the ingredients of betel quid directs the cancer-causing principles to other organs/tissues of the body, the evidence is growing to indicate that cancers other than oropharyngeal may also be caused by betel quid chewing [Chatterjee, A. and S. Deb 1999).]. Since betel quid was found as an important risk factor, it was imperative to examine BQC and NBQC breast cancer etiology separately to determine their carcinogenesis mechanisms.

The AA genotype of CCND1 gene was protective against breast cancer in BQC analogous to the risk seen in interaction between BQC and GG genotype in CART. Contrastingly many Caucasian studies report the ‘A’ allele to be associated with breast cancer risk. The G allele has been implicated to increase risk in colorectal cancer in

Singapore and Turkish population, gastric cancer in Chinese population, cervical and nasopharyngeal cancer in Portuguese population (Jia A, et al 2008, Hong Y et al 2005, Catarino R et al 2005, Catarino RJ et al 2006). The GG genotype has previously been associated with worse survival in Squamous Cell Carcinoma of the Head and Neck and von Hippel–Lindau haemangioblastoma (Hong Y et al 2005). Three studies on Southern, Taiwanese and Singapore Chinese have reported similar A and G allelic frequencies to that found in this study (A = 0.60; G = 0.40, mean A allele frequency 0.63). The data thus suggest that the effect of cyclin D1 polymorphism on cancer risk could be population-specific (Hong Y et al 2005). The 'A' allele is predisposed for transcript-b and cyclin D1b production which has tumorigenic effects unlike the transcript-a (Comstock CE et al 2009). Opposed to this GG genotype was seen associated with transcript b in the Singapore Chinese study and both transcripts a and b were expressed in the normal mucosa of all healthy controls irrespective of genotype, suggesting that both transcripts can be spliced from both alleles. Study on leukemia patients has also shown that the predominant transcript in GG patients was transcript b. Another study depicted the GG genotype to be associated with reduced disease free interval (Matthias C et al 1998). In the present study, the AA genotype may have endorsed the transcript a production thus conferring protection. In our study, most likely the distribution of G and A allele would be ethnicity and population associated wherein CCND1 splicing is modified influencing the GG genotype to endorse splicing of the tumorigenic transcript b whilst promoting malignant transformations. CART showed an increased breast cancer risk among alcoholics with CCND1 GA and AA genotypes

which were non smokers was observed. Production of ROS is a possible mechanism of alcohol-related carcinogenesis. Oxidative stress leads to lipid peroxidation, the products of which are reactive electrophilic compounds that react with DNA to form exocyclic DNA adducts and reactive aldehydes (Boffetta P et al 2006). In addition, risk was seen among all three genotypes of TP53 among non alcoholics, non smokers with CCND1 GA/AA genotypes. This could be due to small number of samples in the end of the nodes and therefore it is difficult to interpret the results biologically.

The 'C' allele of RAD51 showed an increased breast cancer risk in BQC. RAD51 polymorphism changing guanine to cytosine at position 135 in the 5' untranslated region is associated with enhanced promoter activity. High level of basal DNA damage in individuals with CC genotype in endometrial cancer has been observed (Krupa R et al 2011) in contrast to reduced risk among heavy smokers in head and neck squamous cell carcinoma (Werbrouck J et al 2008). In *BRCA2* carriers, *RAD51-135C* heterozygote frequency in affected women was significantly higher than in unaffected women (*P* 5 0.07). Increased breast cancer risk in *BRCA2* carriers who were *RAD51-135C* heterozygotes was observed. Therefore *RAD51-135C* could be a clinically significant modifier of *BRCA2* penetrance, specifically in raising breast cancer risk at younger ages (Levy-Lahad E et al 2001). Although, reduced breast cancer risk was also reported with 83 pairs of female BRCA1 5382insC mutation carriers from Poland (Jakubowska A et al 2003). In this study BRCA2 K2729N variant was seen in two BQC cases. This variant has been reported in 3% ESCC cases and controls in one study and in 0.62% in breast cancer cases in the Chinese population and one familial ESCC case

in Turkmen population of Iran. K2729N variant located in the conserved *BRCA2* COOH-terminal domain is involved in α-helix and β-sheet structures of oligosaccharide-binding fold 1. Unfolding of RAD51 from BRCA2 to the damaged DNA by FANCG protein is regulated by Oligosaccharide-binding fold 1 site (Kaushal M et al 2009). which could therefore be effected in these BRCA2 mutated samples. Moreover, as interaction between the BRCA2 and RAD51 is essential for DNA repair, the C allele may act indirectly and disrupt DNA repair allowing the cell to accumulate more mutations (Marx, J. 1997). In addition, K2729 variant is also located in the binding domain of BRCA2 to MAGE-D1 protein, a synergistic suppressor of cell proliferation indicating deregulated cell proliferation in BRCA2 mutated samples due to incorrect/ nonbinding of MAGE-D1 to BRCA2. Overall, breast cancer etiology in BQC was governed by betel quid carcinogens and CCND1 genotypes as seen in LR and CART along with roles of mutations in BRCA2 gene and a minor role of the C allele of RAD51 genes.

LR showed the main effect of CCND1 GA and AA genotypes was protective against breast cancer risk in NBQC. Conversely as opposed to BQC, no interaction with any environmental variable was seen here. It is essential to state that significant association of AA genotype of CCND1 was common is both BQC and NBQC. As both groups were obtain from the same population a similar abovementioned CCND1 spicing mechanism promoting carcinogenesis can be proposed for both groups. LR showed Pro/Pro genotype of TP53 polymorphism as protection against breast cancer in NBQC which was analogous to the risk seen in interaction between NBQC and Arg/Arg genotype in CART. The literature remains highly controversial regarding the role of

TP53 Arg72Pro polymorphism in breast cancer risk. Vannini et al reported that presence of an arginine allele increased drug resistance and metastatic breast cancer patients homozygous for arginine had a significantly shorter time for progression and overall survival than those with heterozygous arginine/proline tumors (Vannini, I. et al. 2008). Higher frequency of *TP53* mutations on the Arg72 compared with the Pro72 allele in different squamous cell cancers have been reported. Preferential selection of the Arg72 allele in cancers with recessive TP53 mutants has also been found. Recessive TP53 mutants achieve a selective growth advantage by an Arg72-dependent inactivation of TP73, whereas the dominant negative TP53 mutants inactivate the remaining wild-type TP53 allele in an Arg72-independent manner (Langerod A et al 2002). Beside apoptosis and cell cycle control, p53 protein seems to be crucial in the regulation of the different DNA repair pathways [43]. A recent study demonstrated that *Pro* variant activates several *TP53* dependent target genes involved in DNA repair and repair DNA damage much more efficiently than the *Arg* variant expressing cells (Costa S et al 2008). These facts indicate towards the protective effects of the Pro/Pro genotype and risk with Arg/Arg genotype as seen in our study.

The MDR analysis did not generate any significant model showing gene-gene or gene environmental interactions in the BQC. However, MDR analysis did generate a 4 order model showing interactions between TP53, BRCA2 -26 G>A, RAD51and CCND1 in NBQC. The -26 G>A polymorphism in the 5' UTR of *BRCA2* has a regulatory role which is further influenced by codon 72 polymorphism in the TP53 gene. The ‘A’ allele is associated with higher expression and increased genomic instability via inhibition of p53

transactivation. The G allele is associated with low expression impairing the DNA repair leading to genomic instability. Either of the homozygous forms is associated with allelic imbalance or *p53* mutations making the heterozygous form protective characterized by intermediate levels of *BRCA2* expression and an optimal balanced DNA damage response (Gochhait S et al 2007). p53 interaction with RAD51 may influence DNA recombination and repair and modifications of p53 by mutation and protein binding may affect this interaction (Buchhop, S et al 1997). The same was seen in the post-hoc analysis done through entropy graph to visualize and interpret interaction model identified by MDR. RAD51 and TP53 had the strongest interaction subsequent to TP53 and BRCA2 EX2. CCND1 had an independent effect. Overall breast cancer etiology in NBQC was governed by interaction between TP53, RAD51, BRCA2 and CCND1 genes along with a major role of TP53 codon 72 polymorphism.

Although LR analysis has the advantage of controlling for confounding variables (Zhai R et al 2010) examining interactions in more than two variables is not feasible as it detects only low-order interactions and the model complexity increases with the order of interactions. This limitation of LR is referred to as the curse of dimensionality. CART and MDR do not assume any specific parametric form for the relation between independent and dependent variables whilst uncovering SNP-SNP interactions that are missed by LR. They can also deal with sparse and high-dimension data and can account for non-linear SNP-SNP interactions (Briollais L et al 2007). Large standard errors (SEs) and an increased type I error can result from high-order interactions involving multidimensional factors. Cross validation and permutation testing procedures

in MDR reduce the chances of making type I errors as a result of multiple testing (Zhai R et al 2010). An important feature of CART is the influence of the first split on the tree structure. In our analyses the main effect, betel quid appears in the first split. As there is a strong main effect the resulting tree is interpreted as very stable (Briollais L et al 2007). Moreover the significance of our results can also be gauged considering the FPRP values we obtained under different scenarios (Table 5.14).

One of the limitations of this study is that we used a candidate polymorphism approach based on potential functional role in genes with higher potential of being associated to cancer risk. A more comprehensive approach including tagging SNPs would present more convincing support for the associations. Second, the sample size of this study is relatively small but the population is ethnically homogeneous with distinct tobacco usage and food intakes. Further studies in ethnically similar populations are required to validate our findings. Third, dietary patterns and other factors like family history neither were unaccounted for nor adjusted in the analyses because of missing or uncollected data. Given the strong interactions detected in this study, these potential confounders would probably have minor influence on the results. Fourth, CCND1 polymorphism in controls showed deviation from HWE. After ruling out false positive associations and genotyping errors perhaps population stratification, could be a reason for this deviation. Also case-control matching was done in reference to age and ethnicity, thereby controlling for any confounding effect accounted by these variables.

In summary multifaceted analytic approach (CART and MDR) along with LR revealed a complex gene-environment interaction with CCND1 in BQC apart from

important roles of RAD51 and BRCA2 in breast pathogenesis. Contrastingly, no environmental interactions were seen in NBQC and imperative and main effect of TP53 and CCND1 along with gene-gene interactions in NBQC.

Chapter 6

To screen and differentiate Genomic alteration in breast cancer patients exposed to betel quid chewing

Chapter 6

To screen and differentiate Genomic alteration in breast cancer patients exposed to betel quid chewing

6.1. Introduction

The case control studies in previous two chapters have highlighted betel quid chewing to be an independent risk factor for breast cancer in the Northeast Indian population. There are large number of studies elucidating molecular mechanisms associated with tobacco associated cancers like oral, esopheal, gastric and lung cancer. Little is known about molecular pathogenesis associated with betel quid chewing and breast cancer.

Direct analysis of the tumor genome provides an alternative and complementary, means of comparing breast tumors by revealing the genetic events accumulated during tumor progression Loo, L. W.et al. 2004). One of the basic goals in the area of genomic changes is to identify genome copy number abnormalities. Genetic alterations are involved in the activation of proto-oncogenes and in the inactivation of tumour suppressor genes. Copy number alterations can cause gene dosage, gene interruption, generation of a fusion gene, position effects, unmasking of recessive coding region mutations (single nucleotide polymorphisms, SNPs, in coding DNA) or other functional SNPs Wong, K. K.et al. 2004).

During the last two decades, technology developments have enabled a higher resolution analysis of the human genome Pollack, J. R. et al. 2002). Array-CGH has been successfully used to identify genomic deletions and duplications. This technology is high throughput and useful in identifying detecting submicroscopic rearrangements not visible by routine chromosome analysis Hu, N., C. Wang, et al. 2006). With more than 1.4 million SNPs, high-density SNP array is a potential platform for high-resolution whole genome allelotyping with accurate copy number measurements. The Affymetrix 10K SNP array contains 11 560 SNP alleles with high frequencies of heterozygosity. This new SNP array platform is shown to have high accuracy (99.5%), reproducibility (99.9%) and call rate (95%). Wong, K. K.et al. 2004).

In vitro and in vivo experiments have shown that BQ consumption can also cause micronuclei and DNA adducts formation, chromosomal aberrations, allelic imbalances and sister chromatid exchange in oral mucosa cells (IARC 2012). Carcinogens in BQ lead to accumulation of genetic alterations at 3q26.3 locus particularly in recurrent oral tumors (Chiang, W. F. et al. 2011) besides accelerating tumor migration by stimulating MMP-8 expression through MEK pathway (Liu, S. Y. et al. 2007).

In addition, calcium hydroxide a major content of slaked lime in the presence of areca nut is responsible for the formation of ROS (reactive oxygen species) known to cause oxidative damage in the DNA of buccal mucosa cells of

BQ chewers. Presence of iron and copper transition metals are also involved in the catalytic process of ROS generation (Nair, U. et al. 2004). This ROS generation leads to structural alterations in DNA, including rearrangements, deletions, insertions and sequence amplification, affect cytoplasmic and nuclear signal transduction pathways, modulate the activity of the proteins and genes that respond to stress and act to regulate genes related to cell proliferation, differentiation and apoptosis (Wiseman, H. and B. Halliwell 1996).

Tobacco chewing with BQ results in increased exposure (~1000 µg/day) to carcinogenic tobacco-specifc nitrosamines (TSNAs). High levels of TSNAs have been found in saliva samples of BQ chewers collected from India. N'-nitrosonornicotine (NNN), 4-(N-methyl-N-nitrosamino)-1-(3-pyridyl)-1-butanone (NNK), N-nitrosoanabasine (NAB), N-nitrosodimethylamine and N-nitrosodiethylamine have been detected in saliva of BQ with tobacco chewers (Nair, U. et al. 2004), breast tissue of women workers and are known to induce mammary tumors in rodents and anaphase bridges via DNA double stranded breaks causing genomic imbalances in human cells (Narayan, S. et al. 2004 and Luo, L. Z. et al. 2004). Regions like 7p11.2 (epidermal growth factor receptor) and 11q13.3 (cyclin D1) playing a role in pathogenesis of tobacco-related human squamous cell carcinoma has been identified by SNP array (Baras, A. et al. 2009). Examination of genomic alteration due to tobacco carcinogens depict gain on chromosomes 6 and 8, and losses on chromosomes 11 and 14 in mouse lung adenocarcinomas (Herzog, C. R. et al. 2006) and gains of 1p and 3q in patients

with tobacco exposure history in head and neck squamous cell carcinomas (Singh, B., et al. 2002). In addition, Benzo(*a*)pyrene [B(a)P] diolepoxide (BPDE), a carcinogen present in cigarette smoke, induces chromosomal 9p21 aberrations seen to be significantly higher in peripheral blood lymphocytes of bladder cancer cases than that of controls (Gu, J. et al. 2008). Allelic imbalance at 5q22.2~q22.3 (LOX gene) is significantly higher among smokers than nonsmokers in clear cell renal carcinomas indicating that tobacco may cause genetic alterations (Korenaga, Y. et al. 2005).

Earlier studies on genomic alterations in breast cancer have investigated copy number changes between different subtypes and BRCA predisposed breast tumors and cell lines (Loo, L. W. et al. 2004 , Fang, M. et al. 2011 and Jonsson, G. et al. 2005). Although, the literature suggests role of BQ carcinogens in mediating genomic alterations, there is dearth of evidence suggesting its role in breast carcinogenesis. The present study has been undertaken to elucidate the genetic alterations induced by BQ chewing leading to breast carcinogenesis utilizing whole genome SNP array and Ingenuity pathway analysis in breast cancer patients with and without BQ chewing history.

6.2. Experimental methods

6.2.1. Materials

Agarose, Tris base, EDTA, NaCl, SDS, Triton X-100 and other fine chemicals were purchased from Sigma Chemicals, USA. RNA later, DNA extraction kit were purchased from Qiagen Sciences, USA, Genechip Mapping 10K early access array analysis The Single Primer Assay Protocol (labeling, hybridization, washing, staining and scanning) from Affymetrix, Santa Clara, CA, USA).

6.2.2. Chemicals used

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

6.2.3. Patient recruitment and sample collection

Patients with a confirmed diagnosis of breast cancer admitted in the Dr. Bhubaneswar Borooah Cancer Institute, Guwahati, Civil Hospital, Aizawl, and Sir Thutob Namgyal Memorial Hospital, Gangtok, the collaborating centers in Northeast India from the year 2005-2008 were included in the study. All subjects provided written informed consent for participation done under a protocol

approved by the institutional ethics committee of Regional Medical Research Centre, North East Region (Indian Council of Medical Research).

6.2.4. Inclusion criteria: Same as chapter 4

6.2.5. Exclusion criteria : Same as chapter 4

6.2.6. Patient details

Ninety two patients with breast tumors histopathologically confirmed as breast cancer at the Dr. B. Borrooah Cancer Institute, Guwahati and Civil Hospital, Aizwal India between November 2005 and December 2008 were registered for this study. Besides collecting tumor tissues in formalin for histopathology, tumor tissue in RNAlater Demographics, including age, sex, menopausal status, BQ history, tobacco history, alcohol drinking, family history and area of residence were obtained for each case. To quantify betel quid chewing we defined a habitual BQ chewer who chewed one betel quid or more daily for no less than ten years. Details of betel quid chewing history for 26 BQC samples are given in supplementary table 6.2. The ingredients of BQ included areca nut (Areca catechu), catechu (Acacia catechu) and slaked lime (calcium oxide and calcium hydroxide) wrapped in a betel leaf (Piper betle) and tobacco. Thirty two patients with locally advanced breast cancer were given neoadjuvant chemotherapy therefore were excluded. DNA was extracted from the fresh frozen tumor tissue and blood. Specimens with lower than 70% cancer cellularity, inadequate DNA

concentration (<50 ng/mL), or a smearing pattern in gel electrophoresis were not included for genotyping. On this basis 43 cases of breast cancer cases were selected and analyzed for copy number assessment which included 26 BQC with only BQ chewing history and 17 NBQC with no history of tobacco chewing, tobacco smoking and alcohol consumption. All 43 cases were morphologically infilterating ductal carcinoma, not otherwise specific. Control germline DNA extracted from blood lymphocytes was used from age matched 14 breast cancer patients.

6.2.7. Collection of blood samples Same as chapter 4

6.2.8. Extraction of Genomic DNA

Tumor and blood DNA from breast cancer patients was extracted by using Himedia kit (Mumbai, India) and stored at -20^{O}C till further analyzed.

6.2.9. Collection of tissue samples

Tumor tissue was collected in formalin for histopathology and in RNAlater. Tissue collected in ENALater vials were stored in -20^{0}C until transported to the laboratory where the study was performed

6.2.10. Quantification of Genomic DNA

Same as chapter 4

6.2.11. Agarose Gel Electrophoresis of Extracted DNA Samples

Same as chapter 4

6.2.12. Single Nucleotide Polymorphism array

Genechip Mapping 10K early access array analysis The Single Primer Assay Protocol (labeling, hybridization, washing, staining and scanning) was performed according to the manufacturer's instructions (Affymetrix, Santa Clara, CA, USA).

6.3.1. Data and Statistical Analysis

The primary experimental data was normalized to a baseline array with median signal intensity by applying invariant set normalization method. Copy number change was measured based on comparing the signal intensities at each probe locus between control and tumor samples by applying the hidden Markov Model using the dChip software, with a sliding window of 3 SNPs. Copy number gain was defined as > 2.8 copies and loss was defined as less than 1.2 copies in at least 3 consecutive SNPs (George, R. E. et al. 2007). Recurrent altered regions were identified as regions with gain or loss in ≥ 3SNPs in not less than 15% of samples (Turner, N. et al. 2010). Mapping information of SNP locations and cytogenetic band were based on curation of Affymetrix and University of California Santa Cruz hg 17 (http://genome.ucsc.edu). To identify exposure-related aberrations, the data from individual patients were analyzed at group

level by comparing gene copy number ratios of the tumors of chewers and nonchewers patients. In each region, we considered a 3 × 2 contingency table, with the rows representing number of patients with copy number gain, copy number loss or normal copy number in that region and the column representing BQC and NBQC breast cancer patients. Significant regions ($p<0.05$) were identified by comparing the copy number changes in the 26 BQC versus 17 NBQC breast cancer patients using a Fisher's Exact Test based on the 3 × 2 table in each region. FDR was calculated using Benjamini and Hoeschbergs using the Q value software in R package (Benjamini, Y. et al. 2001).

6.3.2. Gene Ontology (GO), Pathway and Network Analyses

Functional annotation analysis was performed using the DAVID (Database for Annotation, Visualization and Integrated Discovery) Functional Annotation Tool and Database (Huang da, W. et al. 2009). A modified, more conservative Fisher's exact p-value, or EASE score, is used to determine if there is a significant level of enrichment in the gene set. To determine pathways and networks those were significantly enriched in the two groups we performed pathway analysis using the Ingenuity Pathway Analysis (IPA) program (http://www.ingenuity.com).

6.4. Results

6.4.1. 10K SNP array profiles of overall breast cancer patients

Forty-three tissue samples of breast carcinoma and fourteen matched samples of germline DNA were analyzed for copy number alterations. The mean age of cases was 44.4 ± 9.6 years and maximum cases were between 40-49 years. Twenty six patients were with BQ chewing history (BQC) and seventeen patients were without BQ chewing history (NBQC). Among the total cases, 23 cases were premenopausal and 20 cases were postmenopausal. Stage IV tumors were more, followed by stage III and II tumors whereas stage for six tumors was unknown. The association between above groups of patients with regard to patient age at diagnosis, tumor stage and menopausal status was statistically insignificant and no sample had a previous family history of cancer, alcohol drinking and tobacco smoking (Table 6.1). 110 recurrent altered regions were identified ranging from 0.15Mb to 51Mb in size (Table S2) with more gains than losses. More than 40% alterations were observed in 30 regions which were essentially gains (1q24.1, 1q25.2, 1q31.1, 1q32.1,1q41, 1q42.2-1q42.3, 1q43-44, 2p11.2, 5p13.3, 5p15.2, 7p12.3-7p12.1, 7p14.1-7p12.3, 7p14.3, 7p21.2, 7p21.3-7p21.2, 7q33, 8q12.1-8q12.2, 8q13.2, 8q22.1, 8q22.2, 8q22.3, 8q24.11, 8q24.21, 12q22.3, 16p13.3-16p13.2, 17q23.2, 17q23.3, 20q13.2, 20q13.33, 21q22.1. Most of the recurrent alterations observed were focal amplification (<10Mb). Although, most chromosomes depicted multiple regions of alteration, 10, 22 and X

chromosomes were altered in only single region with no alteration in chromosome 18. Frequent gains were observed in regions of long arm of chromosome 1 and 8 with genes implicated in cancer. 67 and 50 percent samples presented gain at 8q22.1 and 8q24 respectively. 1q43-44 and 1q41 regions presented with gain in 60 and 51 percent samples respectively. The remaining 80 regions were seen to be altered in 39 to 16 percent samples. The key regions comprising tumor associated genes were 6p25.3, 14q21.3, 1q21.1, 15q25.1, 1p13.2,20q13.11 , 15q22.2-15q23, 19q13.11, 9p23,11q13.3, 11p14.3, 17q25.1, 3p24.2-3p23, 9q34.11-9q34.3, 5p15.33-5p12, 5q35.1-5q35.3, 22q12.1. Loss in single regions was more frequent than recurrent loss (losses in ≥15% of samples) .

Table 6.1: Patient and tumor characteristics in relation to betel quid chewing

Variable	BQC 26 (%)	NBQC 17 (%)	P value
Age at diagnosis			
Mean	44.63 ± 5.8	44.3 ± 11.6	0.9
TMN Stage			
II	9 (34.6)	4 (25.5)	0.28
III	4 (15.3)	2 (11.7)	
IV	8 (30.7)	10 (58.8)	
unknown	5 (19.2)	1 (5.8)	
Menopausal status			
Premenopausal	15 (57.6)	8 (47.0)	0.54
Postmenopausal	11 (42.3)	9 (52.9)	
Family history			
Yes	0	0	
no	26	17	

Fisher's exact test was used. A P-value <0.05 was considered to reflect a significant difference.

Table 6.2: Details of betel quid chewing history for 26 BQC samples

Sno	Sample ID	Betel quid chewing (Yrs)	Number of Betel quid chewed per day.
1	13T-371-06-T-GW	40	11
2	14T-143-06-T-GW	30	3
3	17T-372-06-T-GW_rs	30	8
4	18T-530-08-T-GW_rs	10	2
5	20T-531-08-T-GW_rs	22	3
6	86-06-T_RS	29	4
7	IOP_123_1	10	5
8	IOP_143_1	30	6
9	IOP_250_1	12	6
10	IOP_377_1	50	7
11	IOP_411_1	43	8
12	IOP_412_1	25	2
13	IOP_439-G00	10	1
14	IOP_450_1	35	3
15	IOP_493_1	37	12
16	IOP_494T-A2	26	7
17	IOP_529_1	36	12
18	IOP_557	35	5
19	IOP_564_1	34	11
20	IOP_579	25	3
21	IOP_588-G00	20	4
22	IOP_619_1	22	5
23	IOP_644_1	35	6
24	IOP_646_1	45	13
25	IOP_64-G00	12	2
26	IOP_Tumor_1 30 06az	38	6

Table 6.3: Total 110 regions seen to be altered in overall samples

Cytoband	gain	loss	Total samples	%	Start	End	Size (Mb)	Fragile sites	Genes
1p13.2	14	0	14	32.6	111620648	114443955	2.82		Lrig2, Wnt2b, *Nras*
1q21.1	15	0	15	34.9	143780476	144894002	1.11		Bcl9, *Muc1*
1q24.1	22	0	22	51.2	161945480	167344909	5.4		
1q25.2	29	0	29	67.4	176251808	180031538	3.78		*Lamc2, Tnfsf4, Rsg16*
1q31.1	27	0	27	62.8	184769223	187079322	2.31		Fam5c, Tpr
1q32.1	19	0	19	44.2	197931873	201474622	3.54		Kiss1, Pik3c2b, Mdm4
1q41	22	0	22	51.2	216877367	239186501	22.31		Mar1,Dsp1,Dusp10, Ephx1,Tgfb4 ,Laminin B Receptor,Tp53bp2
1q42.2-1q42.3	30	0	30	69.8	230759715	234370852	3.61		Exo84
1q43-44	29	0	29	67.4	237614919	241862421	4.25	FRA1I fragile site, aphidicolin type, common, fra(1)(q44)	Akt3,Fh,
2p25.3-2p23.1	9	0	9	20.9	6837215	8179329	1.34		
2p16.2-2p16.1	7	0	7	16.3	54196092	56418950	2.22	FRA2D fragile site, aphidicolin	

								type, common, fra(2)(p16.2)	
2p11.2	21	0	21	48.8	88241773	105694407	17.45		Fabp1
2q11.2	15	0	15	34.9	97878052	101494278	3.62	FRA2A fragile site, folic acid type, rare, fra(2)(q11.2)	Aff3
2q22.1	8	0	8	18.6	151285083	155788375	4.5		Lrpb1
3p26.3	7	0	7	16.3	653347	2264798	1.61		Ccr2, Cntn4
3p26.1-3p21.31	12	0	12	27.9	6473283	7535811	1.06		Grm7
3p25.1	11	0	11	25.6	15972853	16744642	0.77		
3p24.2-3p23	9	0	9	20.9	27665535	35068906	7.4	FRA3A fragile site, aphidicolin type, common, fra(3)(p24.2)	Tgfbr2
3q13.2-3q13.31	9	0	9	20.9	115340891	117384598	2.04		
3q22.1	14	0	14	32.6	134070697	185118975	51.05	ATR ataxia telangiectasia and Rad3 related	
3q25.2	9	0	9	20.9	155066932	158948974	3.88	FRA3D fragile site, aphidicolin	

								type, common, fra(3)(q25)	
3q26.1-3q27.2	16	0	16	37.2	165409849	167801377	2.39	FRA3C fragile site, aphidicolin type, common, fra(3)(q27)	Pdcd10 *Ect2* Bche
4p16.1	14	0	14	32.6	10760950	11857265	1.1	FRA4A fragile site, aphidicolin type, common, fra(4)(p16.1)	
4p12-4q11	10	0	10	23.3	46766741	53731283	6.96		
4q33-4q33.2	7	1	8	18.6	167150806	167704678	0.55		
5p15.33-5p12	8	0	8	18.6	1677351	8881383	7.2		Tert
5p15.2	19	0	19	44.2	13931843	16045984	2.11		Trio,Dnah5,Fbxl7
5p13.3	17	1	18	41.9	29073506	35879721	6.81	FRA5A fragile site, BrdU type, common, fra(5)(p13)	Cdh6
5q13.3	0	9	9	20.9	77456368	77952104	0.5		Ap3b1
5q11.2-5q12.1	7	0	7	16.3	57466589	58659721	1.19		Plk2,Rab3c,*Map3k1,Actbl2*

5q21.1-5q21.3	10	0	10	23.3	101290873	103669092	2.38	FRA5F fragile site, aphidicolin type, common, fra(5)(q21)	Pam
5q22.1-5q22.3	13	0	13	30.2	110568424	110723657	0.16		Camk4
5q31.3-5q33.1	9	0	9	20.9	143271856	143899067	0.63		Yipf5
5q35.1-5q35.3	8	0	8	18.6	156693377	157479655	0.79	FRA5G fragile site, folic acid type, rare, fra(5)(q35)	Fgfr4, Mapk9
6p25.3	17	0	17	39.5	150610	1416716	1.27		Dusp22
6p25.1	10	0	10	23.3	5966877	8472511	2.51	FRA6B fragile site, aphidicolin type, common, fra(6)(p25.1)	
6p24.3-6p23	8	0	8	18.6	9595662	16186458	6.59	FRA6A fragile site, folic acid type, rare, fra(6)(p23)	
6p22.3	12	0	12	27.9	20345022	21992498	1.65		E2f3
6q13	8	0	8	18.6	75833704	85962027	10.13	FRA6D fragile site, BrdU	

								type, common, fra(6)(q13)	
6q15	14	0	14	32.6	89348510	89867130	0.52	FRA6G fragile site, aphidicolin type, common, fra(6)(q15)	
6q16.1	9	0	9	20.9	93552831	94259637	0.71		
6q21	13	0	13	30.2	110952941	114779647	3.83	FRA6F fragile site, aphidicolin type, common, fra(6)(q21)	
6q23.2	15	0	15	34.9	131432905	135371662	3.94		Ctgf
6q25.3	14	0	14	32.6	155713132	157738990	2.03		*Sod2*, Arid1b, Nox3, Mir1202,
7p21.3-7p21.2	20	0	20	46.5	12577440	14919353	2.34		Etv1
7p21.2	18	0	18	41.9	17333989	17734067	0.4		Snx13
7p21.1	10	0	10	23.3	18938504	21593778	2.66		Macc1
7p15.2	17	0	17	39.5	27835512	29690012	1.85		
7p14.3	21	0	21	48.8	31794828	32639963	0.85		Creb5
7p14.1-7p12.3	18	0	18	41.9	43737457	49336509	5.6		
7p12.3-7p12.1	21	0	21	48.8	49519486	52496765	2.98		Cobl,*Egrf*

7q11.22-7q21.12	12	0	12	27.9	68764919	70839801	2.07	FRA7J fragile site, aphidicolin type, common, fra(7)(q11), FRA7E fragile site, aphidicolin type, common, fra(7)(q21.2)	Caln1,Rabgef1
7q22.1	9	0	9	20.9	102065731	103642482	1.58	FRA7F fragile site, aphidicolin type, common, fra(7)(q22)	
7q33	19	0	19	44.2	133281372	135010987	1.73		Sec8,Akr1b1,Akr1b10
7q35-7q36.1	10	0	10	23.3	133054441	136538478	3.48	FRA7I fragile site, aphidicolin type, common, fra(7)(q36)	Cntnap2
8p23.1	0	7	7	16.3	8256592	11054299	2.8		Cldn23
8p23.1	0	10	10	23.3	11054299	13155237	2.1	*CSMD1 CUB and Sushi multiple domains 1*	Dlc1
8q11.23	14	0	14	32.6	55602883	59758572	4.16		Cyp7a1

8q12.1-8q12.2	18	0	18	41.9	60797111	63244196	2.45		
8q13.2	19	0	19	44.2	70628049	71089425	0.46		Sulf1
8q13.32-8q21.12	12	0	12	27.9	70628049	81210174	10.58		Il7
8q21.2-8q22.2	12	0	12	27.9	87218471	105541089	18.32		
8q22.1	29	0	29	67.4	97006376	99137466	2.13	FRA8B fragile site, aphidicolin type, common, fra(8)(q22.1)	Mtdh,Laptm4b, Cdh17, Angiopoitienin 1
8q22.2	27	0	27	62.8	100574413	103045541	2.47		
8q22.3	22	0	22	51.2	102196496	118844178	16.65	FRA8A fragile site, folic acid type, rare, fra(8)(q22.3)	
8q24.11	22	0	22	51.2	117649151	122675718	5.03	FRA8E fragile site, distamycin A type, rare, fra(8)(q24.1), FRA8C fragile site, aphidicolin type, common,	Smad12, Collectin 10,Tnfrsf11b,Rad21

								fra(8)(q24.1)	
8q24.21	24	0	24	55.8	128051801	132991056	4.94		Myc,Pvt1
8q24.23	16	0	16	37.2	137820222	142072236	4.25		
9p23	11	0	11	25.6	9757713	12742371	2.98		Tyrp1
9q31.2	10	0	10	23.3	107979573	114702228	6.72		Edg2
9q34.11-9q34.3	9	0	9	20.9	127642279	135054595	7.41		Lamc3
10p12.1	8	0	8	18.6	28581402	36424795	7.84		Map3k8,Itgb1
11p14.3	10	0	10	23.3	25138024	27870966	2.73		Muc15
11p14.1	11	0	11	25.6	32009583	35321011	3.31		Wt1,Cd44
11q13.3	11	0	11	25.6	68986287	70642724	1.66	FRA11A fragile site, folic acid type, rare, fra(11)(q13.3) , FRA11H fragile site, aphidicolin type, common, fra(11)(q13)	Ccnd1,Fgf19,Fgf3,Fgf4
12q11-12q12	8	0	8	18.6	36727030	43496926	6.77		Cntn1
12q14.1-12q14.3	9	0	9	20.9	57087294	65279429	8.19		Lrig3
12q23.1	7	0	7	16.3	97151726	100120265	2.97		Tff3,Tff2,Tff1
13q12.3-13q13.1	6	1	7	16.3	29575163	32641347	3.07		

13q21.31-13q31.1	9	1	10	23.3	61969788	63696479	1.73		
13q22.1-13q31.1	9	0	9	20.9	72056009	73093200	1.04	FRA13B fragile site, BrdU type, common, fra(13)(q21	Pcdh9
13q31.2-13q31.3	13	0	13	30.2	87658904	90523458	2.86		Klf5
14q13.3-14q21.1	11	0	11	25.6	37787739	40277626	2.49		
14q21.3	17	0	17	39.5	49633012	51074896	1.44		Foxa1
14q32.13	9	0	9	20.9	94931266	96104558	1.17		Map4k5
14q32.32-14q32.33	7	0	7	16.3	99862235	106312036	6.45	DICER1 dicer 1, ribonuclease type III	Near Dicer
15q15.3-15q26.3	12	0	12	27.9	41404785	47635423	6.23		
15q22.2-15q23	13	0	13	30.2	61445689	68011638	6.57		
15q25.1	15	0	15	34.9	76606208	84172266	7.57	FRA15A fragile site, aphidicolin type, common, fra(15)(q22)	Dapk2,Samd3,Samd6, Mapk25,Itg11
16p13.3-16p13.2	18	0	18	41.9	2747264	7226822	4.48		Il16

16p13.12-16p11.2	8	0	8	18.6	10529386	33498455	22.97		Trap1
17q11.2	8	0	8	18.6	22436842	23092917	0.66	FRA16A fra(16)(p13.11)	
17q23.2	19	0	19	44.2	52197554	55320513	3.12		Wsb1,Ksr
17q23.3	22	0	22	51.2	59136652	64914787	5.78		
17q25.1	10	0	10	23.3	68880877	74687946	5.81		Ern1, Pecam1
19p13.3-19p12	12	0	12	27.9	3542590	17471210	13.93		Birc5(Survivin)
19q13.11	12	0	12	27.9	40000180	43898128	3.9	FRA19B fragile site, folic acid type, rare, fra(19)(p13)	Vav1,Icam3,,Cdkn2d,Mum1
19q13.31	16	0	16	37.2	51160543	63437743	12.28		Etv2,Map4k1
19q13.32-19q13.43	12	0	12	27.9	51160543	63437743	12.28		Bcl3,Sulta1
20p12.1	9	0	9	20.9	16210360	20930386	4.72		
20q11.22	9	0	9	20.9	31982015	35933409	3.95		
20q13.11	14	0	14	32.6	42101802	50441231	8.34		Tgif2
20q12	11	0	11	25.6			0		Ncoa3
20q13.2	20	0	20	46.5	51292709	53751202	2.46		
20q13.33	12	6	18	41.9	59180207	61366354	2.19		
21q22.11	13	0	13	30.2	36049947	38947095	2.9		Birc7
21q22.13	18	0	18	41.9	37974454	40484883	2.51		
12q22.3	18	0	18	41.9	42530317	46782195	4.25		erg,ets2
22q12.1	8	0	8	18.6	26287071	27802395	1.52	*CHEK2 CHK2*	chek2

								checkpoint homolog (S. pombe)	
Xp26.1-X26.2	7	0	7	16.3	130037493	137016423	6.98	fragile sites	

Table 6.4: Chromosomal gains and deletions in breast tumors from 26 BQC and 17 NBQC. High-level amplifications are in boldface.

	SAMPLE	HABIT	GAIN	DELETION	TOTAL
1	**13T-371-06-T-GW**	BQC	11p14.2,**11q13.3**,12q14.1,12q21.31-12q21.32,**,**14q11.2-4q12,14q13.1-14q21.1,14q21.3-14q23.1,**14q24.2**,14q24.3-14q31.1,14q31.2-14q32.33,16p13.3-16q12.1,16q12.2-16q22.1,16q23.1-16q23.2,19p13.3-19p12,1q21.1-1q44,20p12.1-20q13.33,2q21.3-2q22.1,3p22.3,4q21.1,5q14.1,6q21-6q22.1,7p14.3-7p11.1,7p15.3-7p15.2,7p21.3-7p21.1,7q22.2-7q31.1,7q33,7q34-7q35**,**8q11.23,Xp11.4,Xp22.33-Xp22.11,Xq22.3-Xq24,Xq25-Xq27.3	11q14.1-qter, 13q21.2-qter, 4q32.1-4q32.3, ,8p23.1-8p23.2	36
2	**14T-143-06-T-GW**	BQC	10q21.3-10q23.2, 10q26.12-10q26.2, 11q13.3-11q14.1,17q25.3, ,20q13.13-20q13.33,2p25.2-	1q42.2, 4q23-4q24, ,8p22-8p23.2	17

			2p25.1,3q13.12-3q13.13,3q13.2-3q13.32,4q21.1, ,6q22.31,6q22.33-6q25.2**,**8q12.1-8q24.23,**Xq21.32**,Xq22.3-Xq23		
3	**17T-372-06-T-GW_rs**	BQC	**1p36.12,**1q21.1-qter,2p21,**2p25.3-2p25.2,**2q21.2-2q22.1,**2q37.3,**3p21.33-3p21.31,**3p25.1,** 3p25.3,3q11.2-3q13.13,3q21.2-3q24,3q26.1-3q26.2, 5p14.3-5p12,5p15.33-5p15.16p21.1-6p12.3,6p22.3-6p22.2,**6p25.3,6q15,6q16.1,6q21-6q22.1,**7p12.3-7p12.1,7p21.3-7p21.2, 8q12.1-8q12.3,8q13.2-8q21.13,8q21.3-8q22.3,8q24.13-8q24.3, **9q21.31,12q14.1-12q14.3, 14q21.3-14q22.1, 14q32.13-14q32.2, 19q13.2-19q13.41,** 12q12, 12q24.31-12q24.33, 13q12.2-13q13.3, 13q21.2-13q21.31, 13q22.1-13q31.1, 13q31.2-13q31.3,14q11.2-14q12, 15q25.3-15q26.1, 19p13.3-19p12, 20p12.1, 21q21.3-21q22.11, 21q22.11-21q22.3,Xp22.33-Xp22.11	1p32.2, ,4p15.1-4p14,4p15.2,4p15.32-4p15.31,4q31.21,4q32.1-4q32.2, ,5q12.1,5q14.1-5q14.3,5q23.2,5q33.3, 7q31.31-7q32.1,7q36.1-7q36.2,8p22,8p23.1-8p22,9p21.1, 10q21.3, 10q24.1, 10q25.1, 11p15.3, 12p12.3, 15q26.1, 16q23.1	66
4	**18T-530-08-T-GW_rs**	BQC	1p13, **1p22.1,** 1p32.1-1p31.3, 1p36, 1q21.1-1q41, 1q43-1q44, 2p11, 2p12, 2p25, 2q21.2-2q22.3, **2q32.1, 2q33.3, 3p22.3,** 3p25.1-3p24.1, 3p26.1-3p25.3, 3q22, 3q25, 3q26, 4p13-4q12, **4p16.1-4p15.33, 4q13,** 4q21.1, 4q22, 4q28.2-4q31.3, **4q32,** 4q33-4q34.3, **5p13,** 5p15, **5q11.2,** 5q21, **5q22.1,** 5q23.1, 5q23.2-5q31.2, **5q31.3, 6p12.1,** 6p21.1-	1q42.13-1q42.2, 2p24.1-2p23.2, 3p21.31, 5q14.1, 6p22.3, 7q31.32-7q31.33, 8p22-8p21.2, 8p23, 11q22.1, 12q21.33, 14q24.314q31.1-14q31.2, 15q26, 16q12.2, Xq21.33-Xq22.2, Xq28	93

			6p12.3, 6q13-6q14.1, **6q14, 6q15, 6q16, 6q21-6q22.1,** 6q25,7p14.1-7p12.1,**7p21,** 7q22, **7q31.1,** 7q33-7q34, 7q35, 8q11.23-8q24.3, 9q21-9q22.32, 9q31.1-9q33.1, 10p11.21-10q11.21, 10p12.1, **10p15.1-10p14,** 10q21.3, **10q26.12, 11p11.12-11q12.1, 11p15.4,** 12q24, 13q12, **13q22.1,** 13q31.2-13q31.3, **14q11.2-14q12,** 14q13.3-14q21.1, **14q21.3-14q22.1,** 14q24.2, 12q12, 12q13.13-12q14.3, **14q31.1, 14q31.3, 14q32.13-14q32.2,** 15q22.2, 15q26.1, 16p13.12, 17q21.33-17q23.2, 19q13, 20p12.1-20p11.21, 20p13-20p12.3, 20q12-20q13.33, 21q11, **21q22,** 22q12.1-22q13.31, Xp22.11, **Xp22.32-Xp22.31,** Xq25		
5	**20T-531-08-T-GW_rs**	BQC	1p13, 1p22.3-1p21.3, 1q21-qter, 2p22.1-2p21, 2p24, **2p25,** 2q11.2-2q37.3, 3p21.1-3p14.2, 3p25.3-3p22.3, **3p26.3,** 3q11.2-3q12.3, , **3q22.1,**4q13.1-4q13.2, 4q21.1, 4q22, 4q26-4q27,5p15, 5q11.2-5q12.1, 5q21, **5q22.1, 5q31.1,6q15, 6q16.1, 6q21-6q22.1,**7p14.1-7p12.1,7p22.2-7p21.1,7q22.1-7q31.1,7q33-7q34,7q34-7q36.1,8p228q13.2-8q21.11,**8q24.23,**9p21.1,9p24.1-9p23,9q21.13-q21.31,**9q21.32,**9q21.33-9q22.33,9q31.1-9q31.2,9q31.2-9q32,**9q33.1,**10p12.1,**10p15.1-**	3q21.2**,**4p15.2, 4q31.21, 4q32, 5p13.3, 5q14, 5q33.3, 6p21.1, 6p22.3,6q14.1, 6q21,,8p23.1,10q24.2-10q25.1, 15q25.3,16q12.2,17q24.3-17q25.1, Xp11.4-Xp11.3, ,Xq21.33-Xq22.2	90

			10p14,10q21.1-10q21.2,10q26.12,11p11.12-11q11,**11p14.2-11p14.1,11p15.4,**12p11.23,12q12-12q21.31,12q23.1-12q23.2,13q12.11,13q12.13-13q12.3,13q12.3-13q13.3,13q21.1-13q21.32,13q22.1-13q31.1,13q31.1-13q31.3,14q11.2-14q12,**14q13.3-14q21.1,14q21.3-14q22.1,**14q24.1-14q24.2,**14q31.1,**14q31.2-14q31.3,14q32.13-14q32.33,15q21.3-15q23,**18p11.21,19p13.3,**20p12.1-20p11.21,20q12-20q13.33,21q11.2-21q21.1,21q21.3-21q22.11,21q22.11-21q22.2, Xp22.11		
6	**86-06-T_RS**	BQC	1p12-1q24.3, 1p32.3,1p36,1q42.2-1q44,3q12.3-3q21.1,6p21.32-6q25.3,6p24.3-6p22.3,7q34-7q35,8q11.23-8q13.3,8q21.3-8q24.22,**9p24.3-9p22.3,**9q33,11p15.1-11p13,15q11.2-15q13.2,21q22	NIL	15
7	**IOP_123_1**	BQC	1p13.2-1p12,1p22.2-1p22.1,1q32.1-1q42.13,2p23.2-2p22.3, 3p26.3-3p21.33,3q21.2-3q22.2,3q26.33-3q27.2,4p13-q12,**5p15.2,** 6p25.3-6p12.1,7p14.1-7p11.2,7p15.2-7p14.3,7q11.22,7q21.13-7q31.32,7q35-7q36.1,8p11.1-8q24.3, 10p12.31,10q26.12-10q26.3,11q13.3-11q14.1,13q12.11-13q12.12,13q21.2-	2p24.3-2p24.2,3q23-3q24, 5q23.3-5q31.1,5q33.3, 8p23.1-p22,10q21.2-10q21.3,12q21.33-12q23.1,	37

			13q31.1,14q32.2-14q32.33,15q23-15q26.1,18q11.2,19q13.11-19q13.43,20q12-20q13.32,21q22.11-21q22.3,**15q11.2-15q12,16q13-16q21,11p15.2**		
8	**IOP_143_1**	BQC	1p13.2-1q21.1,1p36.21-1p36.11,2p12-2q11.2,2p16.2-2p16.1,2p25.3-2p23.1,3p25.2-3p24.3,3p26.3,3q11.2-3q13.13,3q13.2-3q13.33,**3q22.3,**3q23-3q25.32,3q26.2-3q27.2,4p15.2-4p14,4p16.1-4p15.32,4q21.1-4q21.3,4q27-4q28.1,5q11.2,6q22.31-6q25.3,7p21.2-7p14.3,7q11.22-7q21.12,7q33,8q11.21-8q24.3,9q12-9q21.13,**9q22.33,**10q26.11-10q26.2,11p14.3-11p11.2,11q13.3-11q14.2,12q13.2-12q21.2,12q21.31-12q21.32,12q23.1-12q23.3,13q12.3-13q13.3,14q21.1,15q23-15q26.1,16p12.1-16q12.1,16p13.3-16p13.2,17q22-17q24.3,17q25.3,18p11.32-18q12.2,19q13.11-19q13.43,20p11.23-20p11.21,20q11.23-20q12,21q22.12-21q22.3,Xq13.3-Xq21.32,Xq23-Xq25,10q21.2-10q23.2,**20q13.13-20q13.33**	8p23.1-8p22	47
9	**IOP_250_1**	BQC	**1p13.2,**1p31.2-1p31.1,1q21.1,1q23.1,1q25.2-1q25.3,1q31.1-1q31.3,1q43-1q44,2p12-2q11.2,2p16.2-2p16.1,2q21.3-2q22.3,2q23.3-2q24.1,2q36.1-2q37.1,3p14.1-3p12.1,3p14.3-3p14.2,3p24.1-3p22.3,3p25.1-	7q21.13-7q21.3, 8p22-8p21.2, 8p23.1, 11q12.3-11q13.1, 11q21-11q22.1, 11q24.2,11q25,12p11.22,17p13.2-17p12,Xq11.2-Xq12,	92

			3p24.3,**3p26.1,**3p26.3,3q13.2-3q13.31,3q22.1-3q23,3q25.32-3q27.2,4p12-4q35.2,4p15.2-4p15.1,4p16.1-4p15.32,5p13.3-5p12,5p15.2-5p15.1,**5q11.2,**5q21.1-5q21.3,5q22.1-5q22.3,5q31.3,6p12.1-6q11.2,**6q13,6q15,**6q16.1-6q16.2,6q22.33-6q23.3,6q25.3,6q27,7p22.2-7q11.21,**7q11.22-7q21.12,** 7q32.1-7q35,**8p11.21-8q11.21,** 8q13.2,8q22.1-8q22.2,8q23.3,8q24.21,9p22.2-9p21.3,9q21.13-9q21.31,9q31.1-9q33.1,10p12.1-10p11.21,10q21.1,10q25.1-10q25.2,11p15.2-11p12,12p12.2-12p12.1,12q12-12q21.33,12q23.1-12q23.3,12q24.13-12q24.32,13q13.3,13q21.2-13q21.32,13q21.33-13q22.1,13q22.2-13q31.1,13q31.1-13q31.3,14q13.2-14q21.1,14q21.3-14q22.1,14q31.2-14q31.3,14q32.13-14q32.2,15q11.2-15q12,15q14-15q21.2,15q24.3-15q25.3,16p13.3-,16p13.2,16q13-16q21,18q21.32-18q22.1,20q12-20q13.32,21q11.2-21q21.1,21q21.2-21q21.3,21q22.13-21q22.3,Xp21.3-Xp21.1,Xp22.31-,Xp22.13,Xq26.1-Xq27.1,17p12-17q24.3,**14q24.2,15q22.2,19q13.12**		
10	**IOP_377_1**	BQC	1p13.2,**1p22.1,**1q22-1q23.1,1q25.2-	3p26.2-3p26.1, 6q12-6q13, ,8p23.1,	80

			1q25.3,1q31.1-1q31.2,1q42.2-1q44,2p13.3-2q11.2,**2p16.2-2p16.1,**2p22.1,2p23.2-2p22.3,2p25.3-2p24.3,2q21.3-2q22.3,2q23.3-2q24.1,2q24.1-2q24.3,2q31.3-2q32.1,2q34-2q35,2q36.2-2q37.3,3q13.2-3q13.31,3q22.1-3q26.32,4p12-4q12,4p16.1-4p15.33,4q13.1-4q13.3,4q21.22-4q22.1,4q26-4q28.1,**4q32.3,**5p15.2-5p13.1,5q21.1-5q21.3,5q22.1,5q31.3-5q32,6p22.3-6p22.1,6p25.3-6p24.3,**6q11.1-6q11.2,**6q21-6q22.1,6q25.3,6q27,7p14.1-7p11.2,7p14.3,**7p15.1,**7p21.3-7p21.2,7q21.11,7q22.1-7q31.1,**7q33,**7q35,**7q35-7q36.1,**8p21.1-8q11.21,8q22.1-8q24.3,9q31.1-9q31.2,**9q31.3,**9q34.13-9q34.3	20q13.33,10q23.31-10q23.32,15q26.1,17p12,18q12.1,20p12.2-20p12.1,Xp11.1-Xq12,	
11	**IOP_411_1**	BQC	**1p36.31-1p36.21,**1q23.1,1q25.2-1q25.3,1q41-1q42.12,1q42.2-1q43,2p15-2q11.2,3q26.1-3q27.2,4p12-4q12,**4p16.1-4p15.33,**4q21.22-4q21.3,5p14.1-5p12,5p15.33-5p15.1,6p22.3-6p22.1,6q13-6q14.1,6q15,6q16.1-6q24.3,7q11.22-7q21.11,7q33,7q34-7q35,**8q13.2-8q13.3,8q22.1-8q24.23, 11p14.3-11p12,11q13.4-11q13.5,13q21.33-13q22.1,13q31.2-13q32.1,15q15.3-15q21.2,15q21.3-15q26.1,16p13.3-16p13.2,17q22-17q24.3,19q13.11-**	14q12	34

			19q13.32,21q22.13-21q22.3,11q14.1-11q14.2,21q11.2-21q21.1		
12	**IOP_412_1**	BQC	1p13.2,1q22-1q23.1,1q23.3-1q25.3,1q31.1-1q32.2,**1q41,**1q42.2-1q44,**2p11.2,**2q21.3-2q22.3,3p13-3p12.2,3q13.2-3q13.31,3q22.1,3q26.1-3q27.1,4p15.2-4p15.1,4p16.1-4p15.33,4q24,4q26-4q28.3,4q32.3,5p14.1-5p12,5p15.2-5p14.3,5p15.31-5p15.2,5q22.1,6p25.3-6p25.2,6q14.3-,6q15,6q16.1-6q21,6q25.3,7p14.1-7p11.2,7p21.3-7p14.3,7q11.22-7q21.12,7q21.3-7q31.1,7q31.2-7q35,**7q35-7q36.1,**8p12-8q13.3,8q21.12-8q21.13,8q22.1-8q23.3,8q24.21-8q24.23,9p23,9q21.2-9q21.31,9q31.1-9q33.1,10p12.1-10p11.21,10p13,11p11.12,12q13.2-12q14.1,12q21.1-12q21.2,12q23.1-12q23.2,13q22.3-13q31.3,14q12-14q21.1,17q22-17q24.3,Xq23-Xq24,Xq26.1-Xq27.1,Xq28,**14q21.3-14q22.1,15q22.2,15q25.1**	NIL	53
13	**IOP_439-G00**	BQC	1p13.2,1q25.2-1q25.3,1q31.1-1q31.2,**1q31.3,**1q41,1q42.2-1q43,2p12-2q11.2,2p25.3-2p25.2,2q13-2q14.1,2q24.1-2q24.3,2q36.2-2q37.3,3p14.1-3p12.2,3p14.3-3p14.2,**3p22.1-3p21.33,**3p24.2-	5q14.1,	72

			3p22.3,**3p26.1,**3q26.1-3q26.31,**4p12-4q12,4p15.2,4p16.1-4p15.33,4p16.3,**4q21.23,**4q32.3,**5p14.1-5p13.2, 5q14.3-5q21.3,**5q22.1,5q31.2,5q31.3,**6p25.3-6p25.2,6q13,6q21-6q22.1,6q23.2-6q23.3,6q25.3-6q26,7p13-7p11.2,7p15.1-7p14.3,7p21.3-7p21.1,**7q11.22-7q21.11,**7q22.1-7q22.2,7q31.33,7q32.3-7q33,9p22.2-9p21.3,9p23,**9q31.2, 9q31.3,**10q21.1,11p11.12,**11p13-11p12,**11p14.3-11p14.2,**11q13.3-11q13.4,11q14.1-11q14.2,**12q12,12q14.1,12q23.1-12q23.3,13q21.33-13q22.1,13q31.2-13q31.3,**14q11.2,**14q13.2-14q21.1,14q21.3-14q22.1,14q23.2-14q24.1,**14q32.13,**15q11.2-15q12,15q15.3-15q21.2,15q21.3-15q25.3,**15q26.1,16p13.12,**18q12.1,18q12.1-18q12.2,20q13.13-20q13.32,21q21.3,21q22.13-21q22.3, Xp21.3-Xp21.2	
14	**IOP_450_1**	BQC	1q22-1q23.1,1q25.2-1q25.3,1q31.1-1q31.2,1q43-1q44,2p12-2q13,2q23.3,2q32.2-2q32.3,5q21.1-5q21.3,5q22.1-5q22.3,5q31.3,7p15.2-7p14.3,7p21.3-7p21.1,7q22.1-7q22.2,7q32.1-7q33,**7q35-**	NIL	35

			7q36.1,8q11.23-8q13.3,8q21.12-8q21.13,8q22.1-8q22.2,8q23.3,11p14.3-11p12,**11q12.2,**11q14.1-11q14.3,13q31.2-13q31.3,14q13.2-14q21.1,14q21.3-14q22.1,14q23.1-14q24.3,14q32.13-14q32.33,15q15.3-15q21.2,15q22.2-15q25.3,**15q26.1,**16p13.3-16p13.2,17q22-17q24.3,**20q13.2-20q13.32,**21q22.13-21q22.3,8q24.13-8q24.3		
15	**IOP_493_1**	BQC	1p32.2-1p21.2,1p34.2-1p32.3,**1q21.1-1q21.3,**1q23.3-1q24.1,1q25.2-1q25.3,1q31.1-1q31.2,1q41-1q43,2p12-2q11.2,2q31.3,3p24.3-3p21.31,3p26.1-3p25.3,4p12-4q12,6p25.3-6q14.1,6q15-6q25.3,7p14.1-7p11.2,7p22.2-7p14.2,7q21.11-7q21.13,**8p12-8p11.21,**8q22.1-8q23.1,8q24.21-8q24.3,9p24.3-9p13.2,12p13.33-12p12.3,15q22.2,19q12-19q13.32, 21q11.2-21q21.1,21q22.13-21q22.3,	20q13.33,	27
16	**IOP_494T-A2**	BQC	1p13.2-1p13.1,**1p22.1,**1q23.1,**1q23.3,1q25.2-1q25.3,1q31.1-1q31.2,**1q42.2-1q43,2p12-2q11.2,**2q13,**2q23.3-2q24.1,2q24.1-2q24.3,3p24.2-3p22.3,**3p26.1,**4q13.3-4q21.3,**4q28.2,**5p14.1-5p13.2,5p15.2-5p14.3, **5q31.2,**5q31.3-5q33.1,5q35.2-5q35.3,6p25.3-6p25.2,6q21-6q22.1,6q23.1-6q23.3,7p14.1-	5q14.1, 17p12	46

			7p11.2,7p15.1-7p14.2,7p21.3-7p21.1,7q11.22-7q21.11,7q33,8p23.3-8p23.2,8q23.3,9p23,**9q31.3,** 11p11.12, **11q13.4-11q13.5,11q14.1-11q14.2,**12q12,13q21.31-13q21.32,13q21.33-13q22.1,14q13.3-14q21.1,15q15.3-15q21.2,15q22.2-15q25.3,15q26.1-15q26.2,16p13.3-16p13.2, ,**17q12**		
17	**IOP_529_1**	BQC	1p33-1p31.3,1q22-1q25.1,1q42.2-1q44,3q25.1-3q27.1,**4q27,**6p25.1-6p22.3,6q27,8q22.1-8q24.23,9p22.3,14q21.3-14q22.1,17q22-17q24.3,19q13.11-19q13.2	NIL	12
18	**IOP_557**	BQC	1p31.1-1p21.3,1q31.1-1q44,2p16.3-2p16.1,3p24.3-3p22.3,3p26.3-3p25.3,4p13-4q12,4p15.1-4p14,4p16.1-4p15.33,4q32.2-4q32.3,**5p15.1,**5p15.32-5p15.31,5q35.1-5q35.3,6q22.33-6q24.3,6q25.3-6q27,6q27,7p15.2-7p14.3,7p21.3-7p21.1,7q11.22,8p11.21-8q24.23, 9q31.2-9q34.3, 10p11.21-10q11.21,12q12,15q15.3-15q26.3,16p13.3-16p13.2,18q11.2,20q12-20q13.31,21q22.2-21q22.3,Xq26.1-Xq28,**13q12.11,17q21.33-17q25.3,11p15.4**	9p23-9p22.3, 10q23.31-10q23.32,11q25,19p13.11-19p12,20p13,	36
19	**IOP_564_1**	BQC	1p13.2-1p13.1,1p36.13-1p36.11,1q21.1-1q23.3,1q23.3-1q24.2,1q31.1,1q41-1q44,2p12-2q11.2,2p16.2-2p16.1,2q36.3-2q37.3,3p21.31,3p23-3p22.3,3p25.1-	11q22.1,	61

		3p24.3,3q24-3q27.2,4p16.1-4p15.33,4q32.3-4q34.3,5p14.1-5p12,5p15.2-5p15.1,5q11.2,6p22.3,6p25.3-6p25.2,6q23.1-6q23.3,6q25.3,7p13-7p12.3,7p15.2-7p14.3,7q11.21-7q11.22,7q22.1-7q22.3,7q33,7q34-7q35,**7q35-7q36.1,**8q21.12-8q21.13,8q22.1-8q22.3,9p21.1,**9p23,**9q31.2-9q31.3,9q34.13-9q34.3,10p12.1-10p11.21,10p13,10p15.2-10p14,10q26.11-10q26.12,11p13-11p12,11p14.2,11q13.3-11q13.5,**11q14.1-11q14.2,** 12p13.33-12p12.3,12q12-12q13.12,13q13.3-13q14.11,13q21.33-13q22.1,**14q21.3-14q22.1,**14q23.2-14q24.2,15q22.2-15q22.31,15q23-15q26.1,**16p13.12,**16p13.3-16p13.2,17q23.3-17q25.3,18p11.21-18q11.2,18q12.1,19p13.3-,19p13.11,19q13.11-19q13.43,20q11.22,**22q12.1**		
	BQC	1p12-1q21.1,1p13.2-1p13.1,1q22-1q23.1,1q25.1-1q25.3,1q31.1,**1q41,**1q42.2-1q44,2p12-2q11.2,2q12.3-2q13,2q36.2-2q37.3,3p13-3p12.3,3p25.1-3p24.3,3p26.3,3q22.1-3q22.3,3q26.33-3q27.2,4p16.1-4p15.32,**4q32.3,**5p13.3-5p13.1,5p15.2-5p15.1,**5q11.2,**5q21.1-5q21.3,5q22.1,5q31.3,5q33.2,6q15,6q21-		
20	**IOP_579**		NIL	66

			6q22.1,6q25.3,**6q27,**7p14.1-7p12.3,7p14.3,**7p15.1,7p21.3-7p21.2,**7q22.1-7q22.3,7q32.3-7q33,**7q35-7q36.1,8q22.1-8q22.2,**8q24.23-8q24.3,9p21.3,9p23,9q31.2-9q33.1,9q33.1-9q34.3,10p12.1-10p11.21,10q25.2,11p14.3-11p14.1,**11q14.1,**11q22.3-11q23.1,12q13.2-12q14.3,13q21.2-13q21.31,13q21.33-13q22.1,14q21.1,**14q21.3-14q22.1,14q32.13-14q32.2,**14q32.2-14q32.33,15q15.3-15q21.2,15q22.2-15q26.1,16p13.3-16p13.2,17p12-17p11.2,17q23.3-17q25.3,18q12.1,18q21.32-18q22.1,20p12.1-20p11.21,20q13.13-20q13.32,21q22.13-21q22.3,Xp22.2-Xp22.13,Xq23-Xq24,Xq25-Xq27.1		
21	**IOP_588-G00**	BQC	1q25.2-1q25.3,1q31.1-1q43,3q13.11-3q13.2,3q25.32-3q28,4p12-4q12,6p25.3,8q13.1-8q24.3,11p13-11p12,11q12.1-11q14.1,13q31.3-13q32.1,14q32.13-14q32.33,15q15.3-15q21.2,15q21.3-15q25.1,17q22-17q24.3,21q22.11-21q22.3	NIL	15
22	**IOP_619_1**	BQC	1q22-1q32.2,1q41-1q44,**2p11.2,**2p25.3-2p25.2,2q33.2-2q34,2q36.3-2q37.3,3p14.1-3p12.3,3p26.1-3p25.1,3q21.3-3q27.3,5q21.1-5q21.3,5q22.1,6p21.1-6p12.3,6p22.3-	7q33-7q34,Xq11.2-Xq12,	38

			6p22.1,6p25.3-6p25.1,6q15,6q25.3,7q11.22, ,8q11.21-8q23.1,**8q23.2-8q24.13,8q24.13-8q24.3,**10p15.3-10q21.2,10q21.3-10q23.1,11p14.2,11q13.3-11q14.1,**11q14.1,**12q23.1-12q23.3,12q24.32-12q24.33,**14q21.1,**15q25.1,**15q26.1,**17q21.33-17q25.3,20p11.21,20p12.1-20p11.23,20q13.13-20q13.33,21q22.12-21q22.2 ,Xq23-Xq27.3		
23	**IOP_644_1**	BQC	1p36.13-1p31.3,1q21.1-1q24.3,1q31.1-1q44,2p11.2,3p22.3-3p21.33,3p25.2-3p24.3,3p26.3-3p26.1,5p15.33-5p12, 6p12.3-6p12.1,6p21.33-6p21.1,6p25.3-6p22.3,6q25.1,7p12.3-7q11.21,7q22.3-7q35,**7q35-7q36.1, 8q21.13-8q24.21,**9p21.1-9p13.2,9q21.32-9q22.33,9q33.1-9q34.3,10q22.2-10q22.3,11q11-11q13.1,16p13.3-16p13.12,**17q11.1-17q11.2,**17q22-17q25.3, 21q22.11-21q22.3,22q12.1-22q13.33	5q23.3, 8p12-8p11.1,8p23.2-8p21.1, 18q21.1,19q12-19q13.11,	31
24	**IOP_646_1**	BQC	1p13.2,1q23.1,1q25.2-1q25.3,1q31.1-1q31.2,**1q41,**1q42.2-1q44,2p13.3-2q11.2,2p16.2-2p16.1,2p23.2-2p22.3,2p25.3-2p25.2,2q13,2q21.3-2q22.3,2q36.3-2q37.3,3p12.2-3p12.1,3p14.1,3p26.1-3p21.31,3q13.13-3q13.31,3q26.1-3q27.2,**4p16.1-4p15.33,**5p14.1-	21q21.1	59

			5q11.2,5p15.2-5p15.1,5q22.1,**5q31.2,**5q35.2-5q35.3,6p25.3-6p24.3,6q11.1-6q12,6q13-6q14.1,6q15,6q16.1-6q22.1,6q22.33-6q23.3,6q25.3-6q27,7p14.3,7q11.22-7q21.11,7q33,7q35-7q36.1,8p11.21-8q11.21,8q12.1-8q12.3,8q13.2-8q13.3,8q22.1-8q23.1,8q23.3-8q24.11,8q24.21-q24.3,9q21.13-9q21.31,9q31.2-9q33.1,9q33.1-9q34.3,10q26.11-10q26.2,11p11.12,11p15.1-11p14.1,**11q14.1-11q14.2,**12q23.1-12q23.3,**14q24.1-14q24.2,15q22.2,**15q25.1,**15q26.1,**16p13.3-16p13.2,17q22-17q24.3,19q12-19q13.43, ,**22q12.1,22q12.3-22q13.2**		
25	**IOP_64-G00**	BQC	1q21.1-1q31.3,2p12-2q11.2,6p21.31-6p21.1,6p25.3-6p25.2,6q22.33-6q23.3,7p14.1-7p11.2,7p21.3-7p14.3,7q22.3-7q31.1,7q31.2-7q31.32,7q33,7q35,8p12-8q11.21,8q23.3-8q24.11,9p22.3,17q22-17q24.3,	20q13.33,Xp22.2	17
26	**IOP_Tumor_1**	BQC	1q24.1-1q25.1,1q25.3-1q44,2p24.3-2p24.2,3q24-3q26.1,8q12.1-8q12.3,8q22.2-8q24.13,10p15.3-10p14,16p13.3-16q13,2p22.3-2p21,3q26.31-3q29	NIL	10
27	**1T-7-06-SJH(T)**	NBQC	1p13, **1p22.1-1p21.3,** 1p31.1-1p22.3, 1q21-qter, **2p25.2,** 2q21.2-2q22.1, 3p25.1-3p24.3, 3q22.1-3q23, 3q26-3q27.2, 4p16.1-4p15.33, 4q13.1-4q21.1, 4q28.1-4q31.3, 4q32.3-4q35.2,	2p16.3, 2p24.1-2p23.2, 5q14.1-5q14.2, 5q21.3, 7p22.2, 7q31.32-7q31.33, 8p22, 8p23.1, 9p21.3, 13q12.11-13q12.12, 13q31.1, 16q12.2, 18q22.2,	63

			5p15.31-5p12, 5q22.1, 6p12.1, 6p22.3-6p22.2, 6q, 7p14.1-7p12.1, **7p21.2-7p21.1,**7q33-7q36.1, **7q36.3, 8p11.21-8q11.21,** 8p12, 8q11.23-8q24.3, **9q21.2-9q21.31,** 12q24.13-12q24.33, 10p12.1-10q11.21, 10p15.2-10p13, 10q21.1-10q21.3, 10q22.2-10q23.1, 10q26.12, 11p15.5-11p15.1, 11q14.1-11q14.3, 11q22.3-11q23.1, 11q23.3, 12p12.3, 12q11-12q13.13, 12q13.3-12q14.3, 13q31.2, 14q11.2, 14q31.1, 15q25.3-15q26.1, 16p13.3-16p13.2, 17q21.33-17q22, 20p12.3-20q13.33, 9q31.1-9q31.2, 11p14.3-11q12.2, **14q21.3-14q22.1**	Xq28,	
28	**20T-310-AZ(T)**	NBQC	1p36.12-1p34.2, 1q21.1-1q21.2, 3q29, 5p14.1-5p12, 5p15, 5q33.1-5q35.3, 10p12.1-10p11.23, 10q24, 15q22.2-15q25.2, 17q25, 19p13.3-19p12, 20q11.22-20q13.33,21q22.3, 9q34, **17q11.2**	3p14.1-3p12.1, 3p24.2-3p24.1, 3q11.2-3q12.1, 3q23-3q24, 4q32.3-4q34.3, 7q21, 7q31.2-7q31.32, 8p23, 9p21, 11q22,13q31.1	26
29	**3T-19-07SJH(T)_rs**	NBQC	1p32.3-1p31.3,1p34.3-1p33,1q21.1-qter,7p22,8q,9p13.3,9p22.1-9p21.3,**9p24,**10p13-10q11.21,10q21.3-10q22.2,11q12.2-11q14.1,14q32,17q11,**17q12-17q21.2,**19p13,20p11.21-20q11.23,21q22.3,22q11.1-22q13.1,**Xp11,**Xp22.33	NIL	20
30	**4T-10-06SJH(T)_rs**	NBQC	1p36.11-1p34.2,1q21.1-1q31.1,1q31.2-1q44,3p26.1,5p15.33-5p13.2,6p25.3-	NIL	21

			6p21.31,8p12-8p11.21,8q12.1-8q24.3,9q12-9q21.2,9q22.31-9q34.3,11p15,**11q21-11q22.1,**15q14,17q11,17q21,17q23.2-17q24.3,17q25,18q11.2-18q23,19p13,20q11,21q22		
31	**5T-12-06SJH(T)_rs**	NBQC	1p32.3-1p31.3, 1q21.1-1q44,2q36.2,5p15,**8p12,**8p12,**8p21,**8q21.12-8q24.3,10p13-10p12.1,10q11.21,17q22-17q23.3,20q13	NIL	12
32	**15T-313-AZ(T)**	NBQC	1q21.1-1q44, 16p13.13-16p11.2, 19p13.3-19p12, 19q13.32-19q13.43, 20q11.22-20q11.23, 21q22.3	NIL	6
33	**16T-29-06-T-GW**	NBQC	1q21.1-1q44, 2p11.2-2q12.2, 2p24.3, 2p25.2-2p25.1, 2q13-2q22.1, 2q31.1-2q36.1, 2q36.3-2q37.3, 3q13.12-3q22.1, 3q29, **6q12-6q14.1, 7p12.1-7p11.1,** 7p22.2-7p14.1, 7q21.13-7q21.2, 7q31.1-7q31.2, 13q12.11,13q12.2-13q12.3	3p14.1-3p13, , 4q21.21, 8p23.2-8p22, 9q21.32,	20
34	**73-06-T_RS**	NBQC	1p31.2-1p12,1p32.1-1p31.3,1p34.3,1q22-qter,2p22.1-2p21,**2p25.3-2p24.2,**cen-2q22.1,3p22.3,4q26-4q28.3,5p15.31-5p12, 5q22.1,**5q31.1,**6p12.1-6p11.2,6p22,**6p24.3,**6q12-6q14.1,6q16.3-6q22.1, 7p14.1-7p12.1,**7p21,** 10q26.12,**11p14,**11q14.3-11q21,**12q12,**12q14.1-12q14.3, 12q21.1,-12q24.32,13q,**14q21.3-**	5q14, 8p22-8p21.3,8p23,	36

			14q22.1,17q11.2,17q12,17q21.32-17q21.33,17q25.1,Xq25-Xq26.3		
35	**7T-16-06SJH(T)_rs**	NBQC	1p36.11-1p34.2,1q42.2-1q44,3q21.1-3q22.3,3q25.33-3q29,4q11-4q12,4q13.3-4q22.1,5q31.1-5q31.2, 8q21.3-8q22.3,8q23.3-8q24.13,9q31.3-9q32,9q33.3-9q34.3,10q22.3,14q13.1,16p13.3-16q12.1,16q12.2-16q21,16q22.1-6q24.3,**17q11,17q12,**17q21,17q23.2-17q25.3,19p13,20p12.1-20q13.33,21q22,22q11.1-22q13.1	8p23.1-8p22	25
36	**8T-21-07SJH(T)_rs**	NBQC	1p13.3-1q21.1,1p33,1q24.1-qter,cent-2q14.1,2p,2q32.3-2q33.1,3p25.2-3p24.3,3q12.1,-3q29,**4p12--4q21.21,**4q24-4q28.3,5p15.2-5q11.2,5q34-5q35.3,**6p21.1-6p12.1,**6p25.2-6p22.3,**6q12-6q14.1,**6q14.3-6q16.1,6q16.1-6q23.2,7p11.2-7q21.11,7p14.3,7p21.3-7p15.3,8q11.23-8q12.3,8q22.1-8q24.3,9q21.13-9q21.31,9q31.2-9q32,9q33.2-9q34.3,10p15.3-10q22.3,11p11.2-11q12.1,11p14.1-11p12,12p11.23-12q12,12p13.33-12p13.31,12q14.1-12q14.2,12q21.1-12q21.31,12q21.33-12q23.1,12q24.21-12q24.33, 13q31.1-13q32.3,14q12-14q13.3,15q11.2-15q26.3,16p13.3-16q21,17q11.1-17q12,17q21.32-	1p31.3, ,2q32.2, 2q36.1,3p24.1, **,**4q22.3,8p, 13q12.3-13q13.1, 16q21-16q22.1, ,Xq26.3	57

			,17q25.3,18q12.1-18q12.2,18q21.2-18q21.33,19p13.3-19q13.43,20q12-20q13.2,21q21.1,21q21.3,21q22.2-21q22.3		
37	**9T-MK-SJH(T)**	NBQC	**1p31.3,1p32.3,**1q21.3-1q23.3,1q42.3-1q44,**2q12.1,**2q14.1-2q14.3,3p26.1-3p25.2,3q21.1-3q22.2,**3q26.1-3q26.2,**5p15.33-5p15.31,5q23.3,5q33.2,6q25.1-6q25.3,**7p12.3-7p11.1,**7p21.3-7p21.1, 8q11.23-8q12.1,8q21.3-8q22.3,8q23.3-8q24.3, 9q33.3-9q34.3,10q23.33-10q24.32,10q26.11-10q26.3,11p11.2-11p11.12,15q22.2-15q24.3,17q22-17q25.3,18p11.22-18p11.21,19p13.3-19p12,19q13.32-19q13.43,20q11.22-20q11.23,20q13.12-20q13.33,14q12-14q13.3,**11q13.3-11q13.4,21q22.3**	8p23.1-8p22,8p23.2, ,9p21.3,9p24.1-9p23, ,21q21.2,Xq21.1-Xq21.33,Xq24-Xq25,	39
38	**IOP_28-G00**	NBQC	1q21.1-1q32.1,1q41-1q42.13,1q42.2-1q43,6p25.3-6p25.2,6q16.1-6q22.1,6q22.33-6q23.3,7p21.3-7p11.2,7q21.3-7q22.2,7q33,7q35,8p12-q11.21,8q13.2-8q21.13,8q24.22,10q26.13-10q26.3,17q22-17q24.3,18p11.32-18p11.21,19q13.11-19q13.32,21q21.1	NIL	18
39	**IOP_290-A2**	NBQC	1p34.3-1p34.1,1q23.1-1q24.3,**1q25.2-1q25.3,**1q31.1-1q32.1,1q41-1q43,2p12-2p11.2,2q33.1,3q13.32-3q21.1, 4p12-	3q28-3q29,7q36.3,9p23-9p22.3,10q26.11,12p12.1	51

			4q12,**4p16.1-4p15.33,**4q21.23,5p14.1-5p13.2,5p15.2-5p14.2,5q11.2,5q14.1-5q14.2,5q21.1-5q21.3,5q31.3-5q33.1,5q33.2-5q33.3,5q34,5q34-5q35.3,6p25.3-6p25.2,7p14.1-7p11.2,**7p14.3-7p14.2,**7p15.3-7p14.3,**7q21.11,**8p22-8p21.1,**8q13.2-8q13.3,8q23.3-8q24.11,9p13.2,** 15q14-15q21.2,15q22.2-15q25.3,16p13.3-16p13.2,**10q23.33,10q25.3,13q21.31-13q21.32,13q31.3,14q13.2-14q13.3,14q21.3-14q22.1,14q32.13-14q32.2,17p12,20p11.23-20p11.21,20q13.2-20q13.31,11q14.1-11q14.2,12q24.31,16p13.12,17q22-17q23.2**		
40	**IOP_453-G00**	NBQC	**1p13.2,**1q23.1-1q24.2,**1q25.2-1q25.3,**1q31.1-1q31.2,1q31.3-1q32.1,**1q41,**1q41-1q43,2p12-2q11.2,**2q13-2q14.1,3p22.1-3p21.33,**3p24.2-3p22.3,**3p26.1,3q13.33-3q21.1, 4p15.2,**5p14.1-5p13.2, 5q35.1-5q35.3,6q21-6q22.1,7p14.1-7p11.2,**7p14.3-7p14.2,**7p15.2-7p14.3,**7p21.1,**7p22.2-7p21.3,7q11.22-7q21.11,7q22.3-7q31.1,7q33,8q22.3-8q24.22,**9p13.2,**9p23,**9q31.3,**10p11.21-10q11.21,10q21.1,10q23.33,**10q25.3, 13q12.3-13q13.1,**13q21.31-13q21.32, 14q13.2-14q21.1,14q21.3-14q22.1,**14q32.13,**15q15.3-15q21.2,15q22.2-15q25.3,**15q26.1,**16p13.3-16p11.2,**17p12-**	3q28-3q29, 5q14.1, 12p12.1, 13q31.1, ,20q13.33,Xp21.1-Xp11.4	52

			17q11.2,17q23.2-17q24.3,20p13-20p11.21,20q12-20q13.32		
41	**IOP_645_1**	NBQC	1q22-1q23.1,1q23.3-1q25.3,1q31.1-1q44,2p12-2q11.2,**2q13,**2q21.3-2q22.3,3q25.2-3q26.2,5p15.33-5p12,5q21.1-5q21.2,5q22.1,**5q31.2,**6p25.3-6p25.2,6q15,6q21-6q22.1,6q22.33-6q23.3,6q25.3,7p14.1-7p11.2,7p15.1-7p14.3,7p21.3-7p15.3,8p11.1-8q21.13, ,**8p12,** ,8q22.1-q24.23,9p22.2-9p21.3,**9p23,**11p14.3-11p11.2,12q12,12q13.2-12q14.1,12q23.1-12q23.3,14q13.2-14q21.1,**14q21.3-14q22.1,**14q23.2-14q24.2,**14q32.13,**15q15.3-15q21.2,15q22.2-15q22.31,15q25.1-15q26.1,16p13.3-16p11.2,17q12-17q21.2,17q23.2-17q25.3,19q13.12-19q13.41,20q13.12-20q13.32,**22q12.1-22q12.2**	8p12,8p23.1-8p22	43
42	**12T-18-06SJH(T)_rs**	NBQC	21q22.3, 20q13.31-20q13.33, 10p13, 3q21.1-3q23, 8q22.1-8q22.3, 6q15, 1q32.2-1q41, 3p26.1-3p25.2, 10p12.1-10p11.23, 19q13.32-19q13.43, 6p25.2-6p21.33, 1q42.3-1q44, 19p13.3-19p12, Xp22.33-Xp22.31	NIL	14
43	**6T-15-06SJH(T)_rs**	NBQC	1p32.3-1p32.2,1p34.2-1p33,1p36.21-1p35.3,1q31.1,1q42.13-1q44,2p16.1-2p13.2,2p22.2-2p21,2q35,4q28.3-4q31.22,6p12.2-6q12,**6q21,**7p12.3-	7q22.1-7q22.2,8p23.2, Xp11.4	35

q11.21,7p21.1-7p15.1,7p22.3-7p21.3, ,8q13.1,8q22.2-8q22.3,8q23.3-8q24.12,8q24.13-8q24.21,9q21.33-9q22.33,11p11.2,11p14.2-11p12,11p15.3-11p15.1,,11q12.1-11q13.4,14q24.3-14q31.1,16p13.13-16q12.1,16q12.2-16q21,17q11.1-17q11.2,19p13.3-19p12,19q13.32-19q13.43,20p12.2-20q13.33,21q22.11-21q22.3, ,Xp22.33

6.4.2. Genetic alterations different between BQC and NBQC

Total number of alterations varied considerably between BQC and NBQC tumors from 12 to 93 alterations among BQC tumors, and from 6 to 63 alterations among NBQC tumors. The frequency plot of alterations per chromosome 1–X is shown in Figure 6.1. The BQC tumors showed a significantly (P< 0.01, T test) higher total number of alterations, as compared with NBQC tumors (48 ± 17 % versus 32 ± 25, respectively) (Table 6.4). One of the important finding was significantly high incidence of gain in fragile sites in BQC tumors (P<0.001, T test) as compared in NBQC tumors, 34 versus 23%, respectively (Table 6.3). Significant (P<0.05) differential genetic alterations were found in twelve chromosomal regions among BQC and NBQC tumors (Table 6.5, Figure 6.2). Among the twelve regions seven chromosomal regions (3p26.3, 3q26.1-3q27.2, 4p16.1, 5q11.2-5q12.1, 6q25.3, 7q33 and 21q22.13) presented more gain in BQC tumors while five regions (16p13.12-16p11.2, 17q11.2, 19p13.3-19p12, 19q13.32-19q13.43, 20q11.22) showed more gain in NBQC tumors. The alterations observed were chiefly gains of sizes ranging between 0.65Mb to 22Mb. Multiple testing was controlled using the false discovery rate (FDR) q-value method. The FDR cutoff up to 0.2 has been commonly used in case-control GWAS studies (Saama, P. M. et al. 2006 and Pang, H. et al. 2011). FDR correction is likely to be conservative considering the relatively small number of cases, but four differentially altered regions at various chromosomes remained significant, as indicated by relatively low FDR values. The FDR value of 0.26 as for regions (3p26.3, 3q26.1-3q27.2, 4p16.1, 5q11.2-5q12.1, 6q25.3, 16p13.12-

16p11.2, 17q11.2, 19q13.32-19q13.43) indicates that the relevance of these finding should be interpreted with caution, and we therefore focused particularly on the regions with P-values = 0.005 and low FDR values. More than 50% BQC tumors presented with gain at 7q33 and 21q22.13 in contrast to just 17% gain in NBQC tumors. Among the regions altered more in NBQC tumors, 52% NBQC tumors had gain at 19p13.3-19p12 in comparison to gain in 11% BQC tumors and 47% of NBQC tumors had gain at 20q11.22 in comparison to gain in 3% BQC tumors.

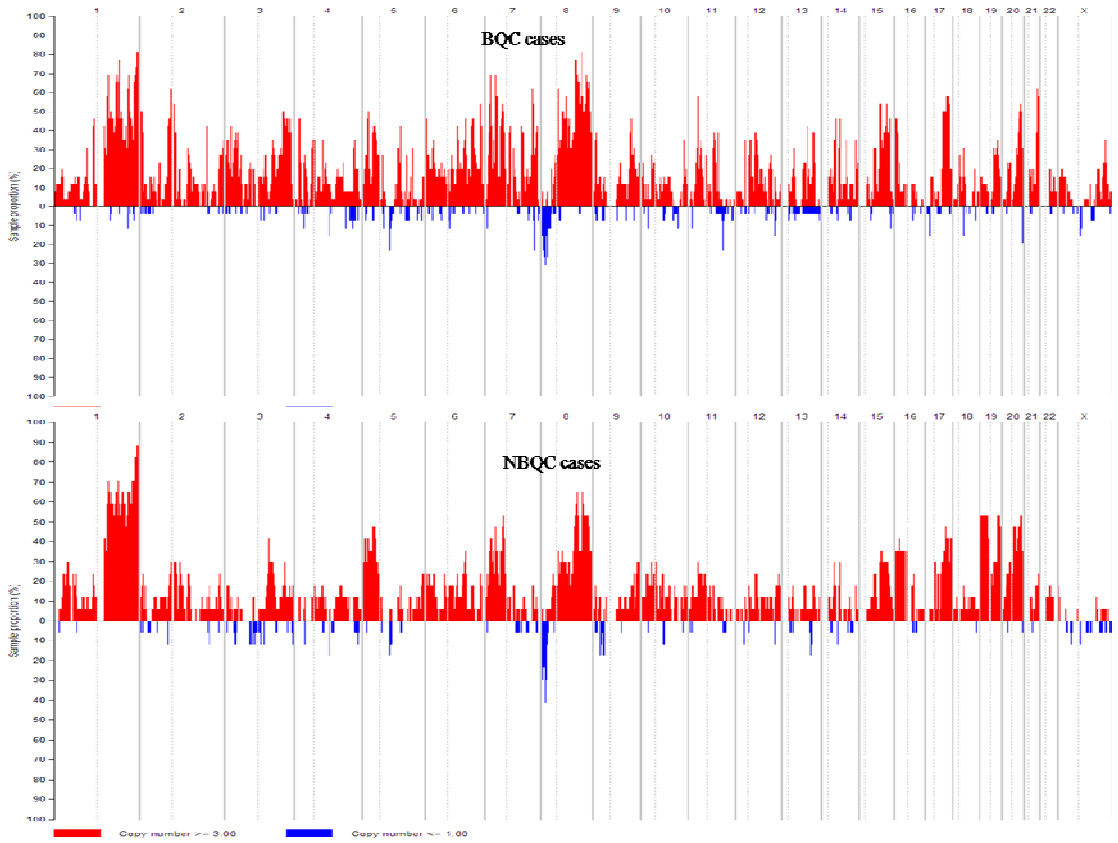
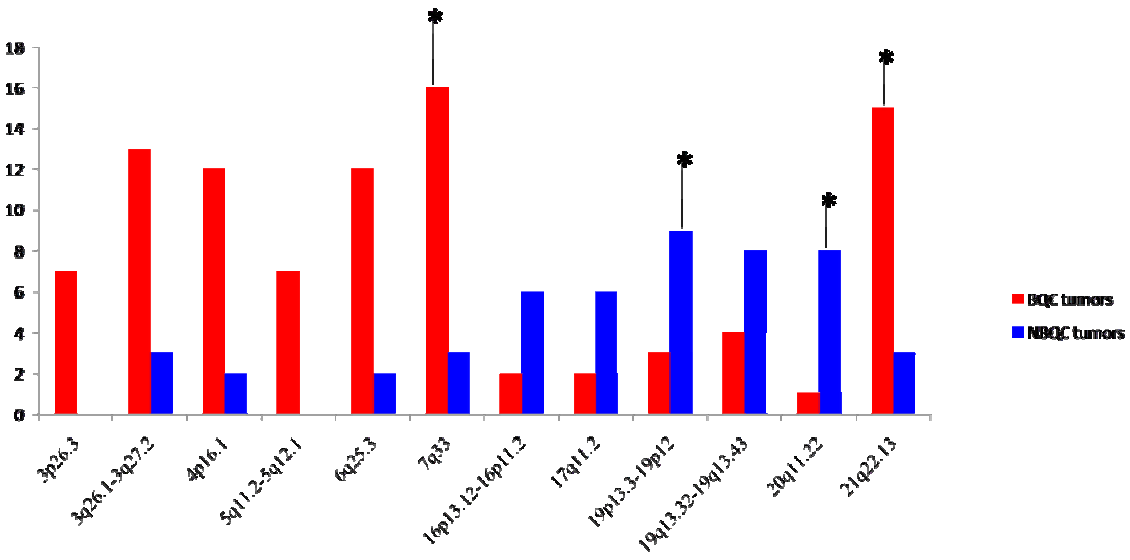


Figure 6.1: Prevalence (%) of patients with ≥3 copies (red) and ≤1 copies (blue) in BQC and NBQC tumors,

Table 6.5: Chromosomal areas with gain those are significantly different between betel quid chewers (BQC) and non betel quid chewers (NBQC) breast cancer patients

Cytoband	BQC (26)	NBQC (17)	p value	Q value (FDR)	Start Site	End Site	Size (Mb)
3p26.3	7	0	0.03	0.26	653347	2264798	1.61
3q26.1-3q27.2	13	3	0.05	0.26	165409849	167801377	2.39
4p16.1	12	2	0.02	0.26	10760950	11857265	1.09
5q11.2-5q12.1	7	0	0.03	0.26	57466589	58659721	1.19
6q25.3	12	2	0.02	0.26	155713132	157738990	2.02
7q33	**16**	**3**	**0.005**	**0.10**	**133281372**	**135010987**	**1.72**
21q22.13	**15**	**3**	**0.01**	**0.10**	**37974454**	**40484883**	**2.51**
16p13.12-16p11.2	2	6	0.04	0.26	10529386	33498455	22.96
17q11.2	2	6	0.04	0.26	22436842	23092917	0.65
19p13.3-19p12	**3**	**9**	**0.005**	**0.10**	**3542590**	**17471210**	**13.92**
19q13.32-19q13.43	4	8	0.03	0.26	51160543	63437743	12.27
20q11.22	**1**	**8**	**0.001**	**0.08**	**31982015**	**35933409**	**3.95**

Frequency of chromosomal regions with significantly different (P<0.05; see Materials and methods for the statistical test) alterations between TBC and NTBC tumors are depicted. Most significant regions, based on the criteria of P<0.05 and a relatively low FDR value, are indicated in bold.. FDR=false discovery rate



*Regions significant after FDR correction.

Figure 6.2: Chromosomal regions altered differently between BQC and NBQC breast tumors.

6.4.3. Gene Ontology (GO) and Network Analyses of associated regions

Genes associated with BQC regions, 7q33 and 21q22.1 were enriched for oxidoreductase ($p<0.001$) and aldo-keto reductase activity ($p= 0.015$) in contrast to G-protein coupled receptor protein signaling pathway ($p= 0.005$) and cell surface receptor linked signal transduction ($p= 0.012$) for 19p13.3-19p12 and 20q11.22 NBQC associated regions. IPA (Ingenuity Pathway Analysis) analysis for BQC associated regions revealed one top network (score=20) "Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism" encompassing

genes like AKR1B1, AKR1B10, AKR1B15, ERG, ETS2 (Figure 6.3). IPA analysis for NBQC genes revealed two top networks (score= 29) “Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry” and “Cellular Development, Embryonic Development, Organismal Development” (Figure 6.4A and B) encompassing genes like RPN2, EMR3, BLCAP and VAV1, NNAT and MUC16 respectively.

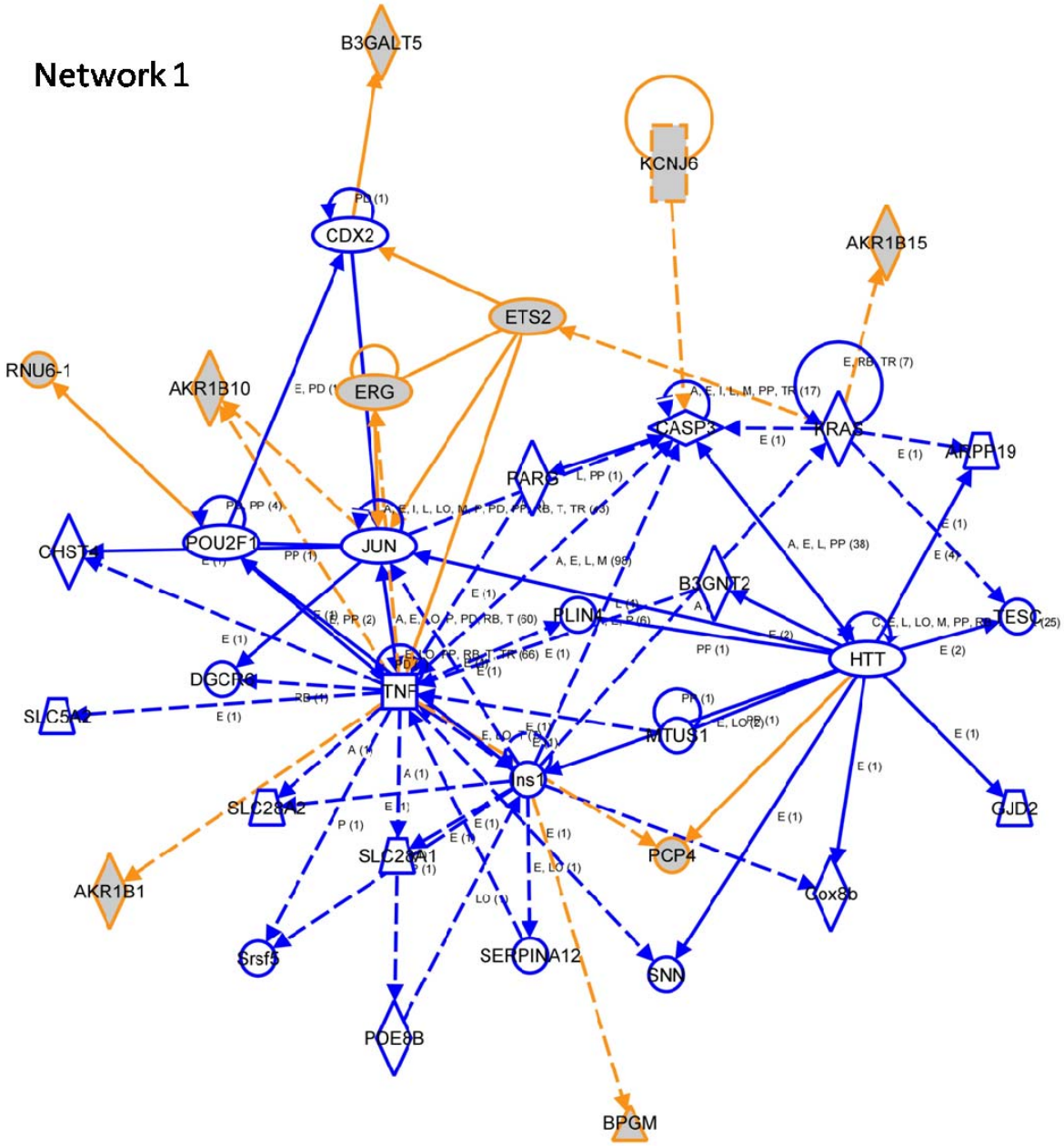


Figure 6.3: BQC Network 1 Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism.

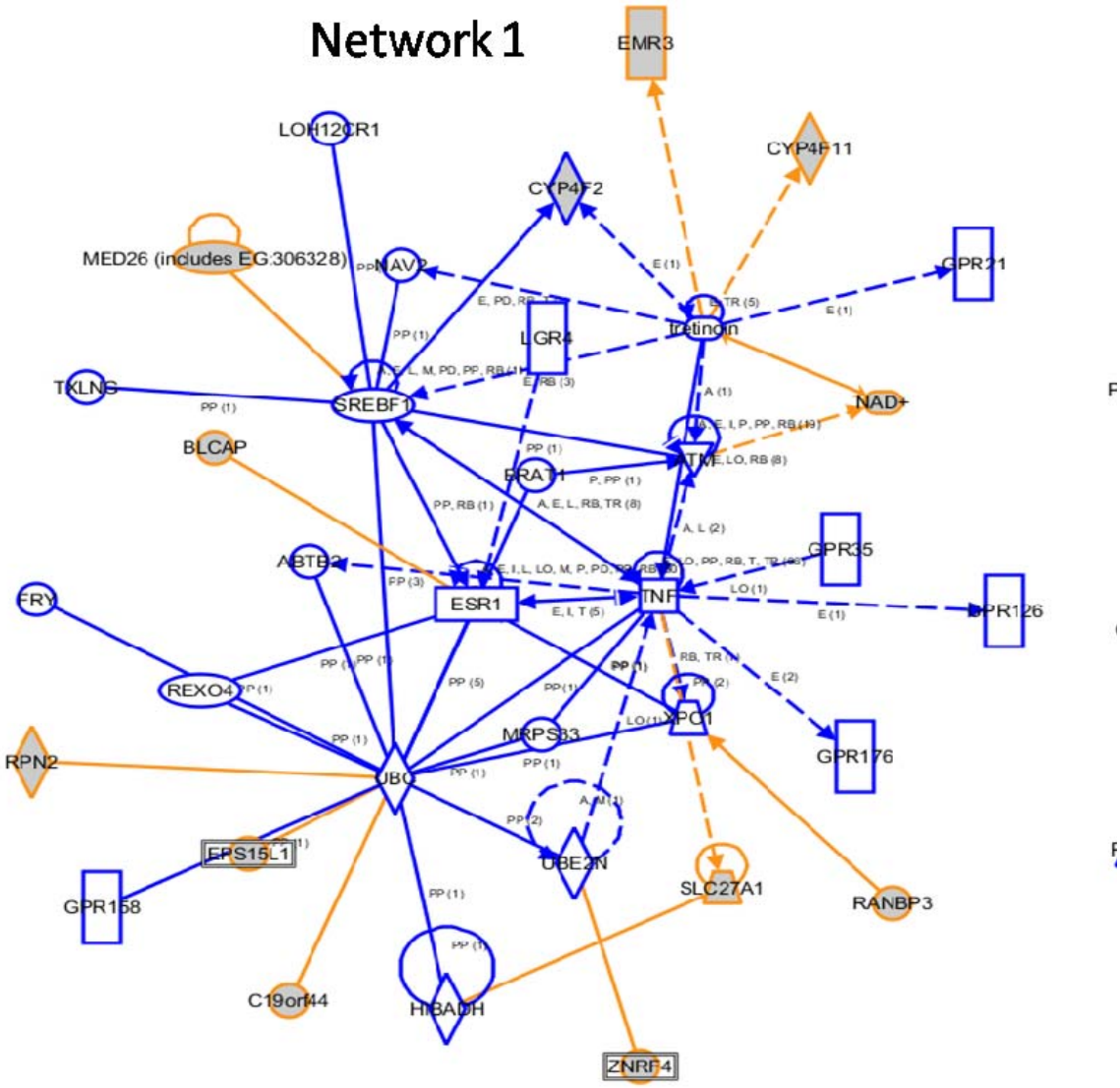


Figure 6.4 A: NBQC Networks: Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry (Network 1)

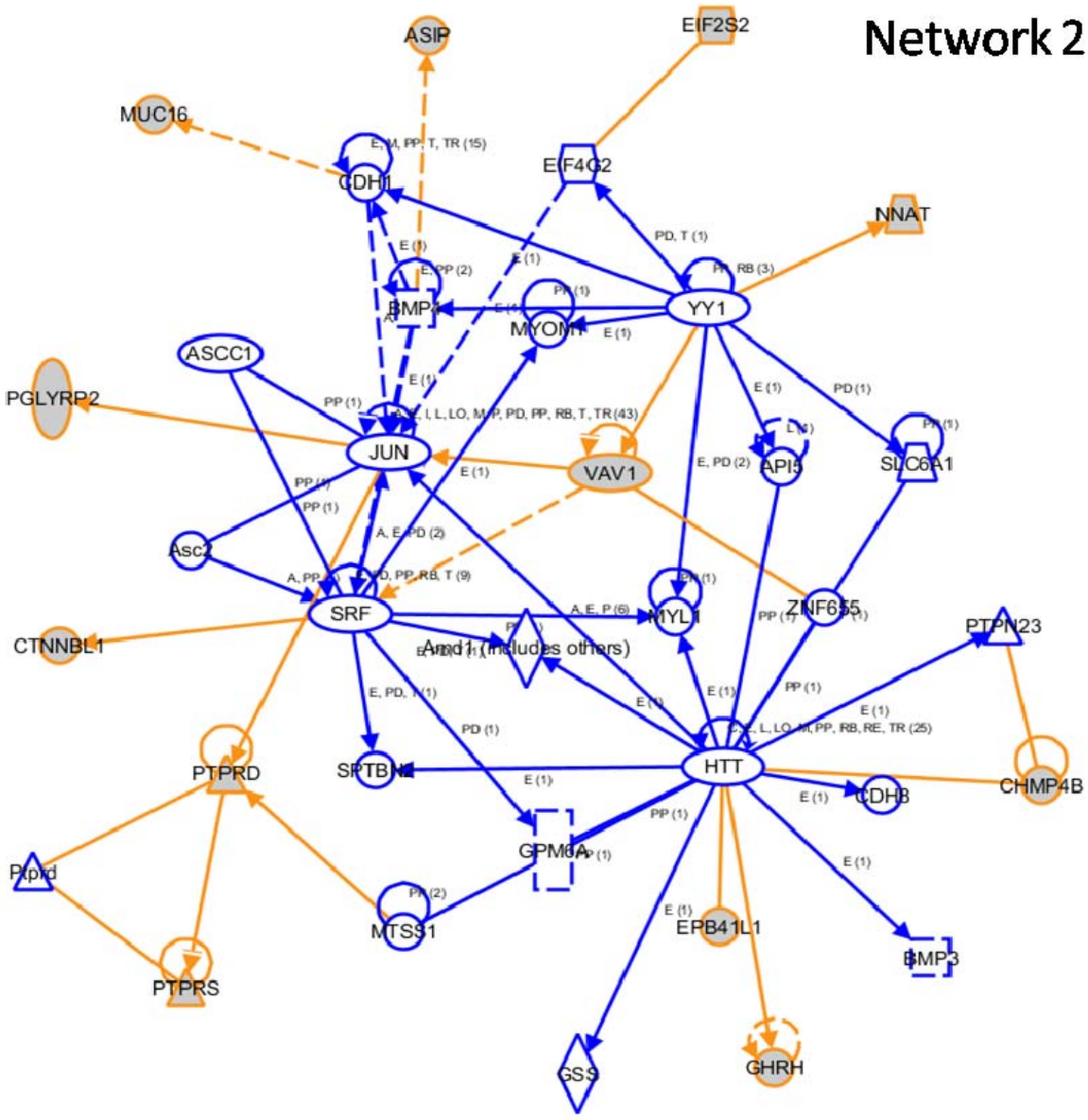


Figure 6.4B: NBQC Networks: Cellular Development, Embryonic Development, Organismal Development (Network 2).

6.4.4. Genetic alterations similar between BQC and NBQC

Twenty seven common regions of gain were illustrated between BQC and NBQC tumors. Regions demonstrating gain in minimum 30% cases from each group were considered as similarly altered (Nowak, N. J. et al. 2005). Both groups exhibited gain on chromosomes 1q, 5p, 7p, 8q, 12q, 16p, 17q and 20q (*Table S4*). Gain in more than 50% samples was seen in six regions (1q31.1, 1q42.2-1q42.3, 1q43-44, 8q22.1, 8q22.2, 8q24.11). Gain in more than 45% samples was seen at 1q24.1, 1q41, 7p12.3-7p12.1, 8q24.21 and 20q13.2 regions. Other regions encompassing probable tumor associated genes were 1q32.1, 1q21.1, 7p21.3-7p21.2, 12q22.3, 16p13.3-16p13.2, 17q23.3.

6.4.5. Gene Ontology (GO) and Network Analyses of similar regions

Enrichment and IPA was performed to investigate the function of genes associated with these regions. Regions were mainly enriched for activation of protein kinase activity (p= 0.009), cell junction (p= 0.01). IPA analysis revealed three top networks (*Table 6.6*, (*figure 6.5A, B and C*). Network 1 functions in Cellular Movement, Connective Tissue Development and Function, Cellular Assembly and Organization (score= 43) with key role played by PTK2. Network 2 functions in Cell-To-Cell Signaling and Interaction, Tissue Development, Organismal Injury and Abnormalities (score= 43) with RPN2, EMR3, VAV1, NNAT and MUC16 important genes. Network 3 functions in Cell Morphology, Cellular Assembly and Organization, Cellular Compromise (score= 32) with key

roles played by MYC and YWHAZ. Among all the tumor associated canonical pathways enriched were GNRH signaling (p= 2.92E-04), cAMP-mediated signaling (p= 3.60E-04), Protein Kinase A signaling (p= 3.77E-04), CXCR4 signaling (p= 4.99E-03), molecular mechanisms of cancer (p= 8.58E-03), phospholipase C Signaling (p= 1.01E-02), RAR Activation (p= 3.16E-02), ILK Signaling (p= 4.21E-02)(Table 6.7, Figure 6.6).

Table 6.6: Significant signaling pathway networks observed in BQC, NBQC and Common altered genomic regions

Networks	Nodes (genes) in Network	Score	Nodes	Identified Nodes	Top Functions
BQC Network 1	AKR1B1,AKR1B10,AKR1B15,ARPP19, B3GALT5,B3GNT2,BPGM,CASP3,CDX2,CHST4, Cox8b,DGCR6,ERG,ETS2,GJD2,HTT,Ins1,JUN, KCNJ6,KRAS,MTUS1,PARG,PCP4,PDE8B,PLIN4, POU2F1,RNU6-1,SERPINA12,SLC28A1,SLC28A2, SLC5A2,SNN,Srsf5,TESC,TNF	28	34	10	Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism
NBQC Network 1	ABTB2,ATM,BLCAP,BRAT1,C19orf44, CYP4F2,CYP4F11,EMR3,EPS15L1, ESR1,FRY,GPR21,GPR35,GPR126, GPR158,GPR176,HIBADH,LGR4, LOH12CR1,MED26 (includes EG:306328),	29	34	12	Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry

	MRPS33,NAD+,NAV2,RANBP3,REXO4, RPN2,SLC27A1,SREBF1,TNF,tretinoin, TXLNG,UBC,UBE2N,XPO1,ZNRF4				
NBQC Network 2	Amd1 (includes others),API5,Asc2,ASCC1, ASIP,BMP3,BMP4,CDH1,CDH8,CHMP4B, CTNNBL1,EIF2S2,EIF4G2,EPB41L1,GHRH,GPM6A, GSS,HTT,JUN,MTSS1,MUC16,MYL1,MYOM1,NNAT, PGLYRP2,PTPN23,PTPRD,Ptprd,PTPRS,SLC6A1, SPTBN2,SRF,VAV1,YY1,ZNF655	29	34	12	Cellular Development, Embryonic Development, Organismal Development
COMMON Network 1	ADCY1,ADCY10,Alpha tubulin,ANGPT1,Arf, ASAP1,ATP2B4,CACNA1E,Calpain,CAPN9, CDC42BPA,CYP24A1,DISC1,EIF3H,ERK1/2, EXT1,GALNT2,Integrin,KIFAP3,NADPH oxidase, NPHS2,Pdgf (complex),PFDN4,Pld,PTK2 (includes EG:14083),Rac,RAD21,Rap1,Rxr,	43	35	25	Cellular Movement, Connective Tissue Development and Function, Cellular Assembly and Organization

	RXRG,RYR2,TNFRSF11B,TNS3,TRIO,TSH				
COMMON Network 2	ABCG1,ADAMTS12,AEBP1,AKAP1, Alpha catenin,APOH,CDH6,CHN2, Collagen type I,Collagen type IV,DOK5, Ecm,EDARADD,Fibrinogen,GRB10, Growth hormone,GTPASE,HAS2,HDL,LDL, MTDH,Mucin,NFkB (complex),NID1,NOV, PKP1,Pro-inflammatory Cytokine,RAB3GAP2, RGS7,SELP,SNX13,SUMO2,SUMO3,TFF3,WIPI1	43	35	24	Cell-To-Cell Signaling and Interaction, Tissue Development, Organismal Injury and Abnormalities
COMMON Network 3	26s Proteasome,Actin,Ck2,DROSHA,ENPP2, ERMAP,FSH,HEATR1,HELZ,HISTONE, Histone h3,Histone h4,IKK (complex),IKZF1, IPO9,Jnk,MARK1,Max-Myc,MYBPH,MYC, NBPF11 (includes others),NCALD,P38 MAPK, PSEN2,RAI14,RNA polymerase II,SRRM2,STRADA,	32	35	20	Cell Morphology, Cellular Assembly and Organization, Cellular Compromise

TARBP1,TBCE,Ube2-ubiquitin,UBE2D4,UBE2G2,
Ubiquitin, YWHAZ

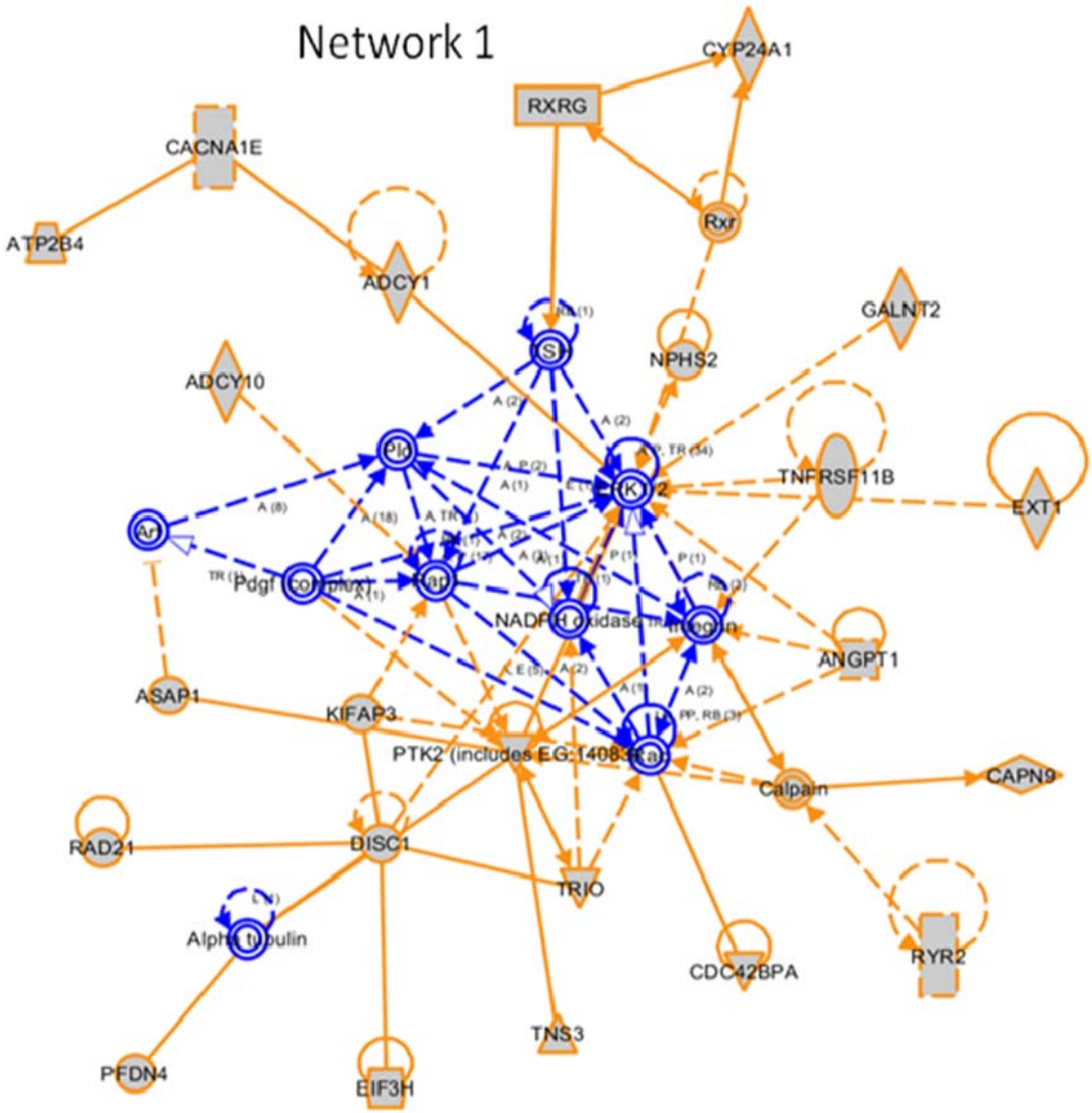


Figure 6.5A: Common Networks: Cellular Movement, Connective Tissue Development and Function, Cellular Assembly and Organization (Network 1)

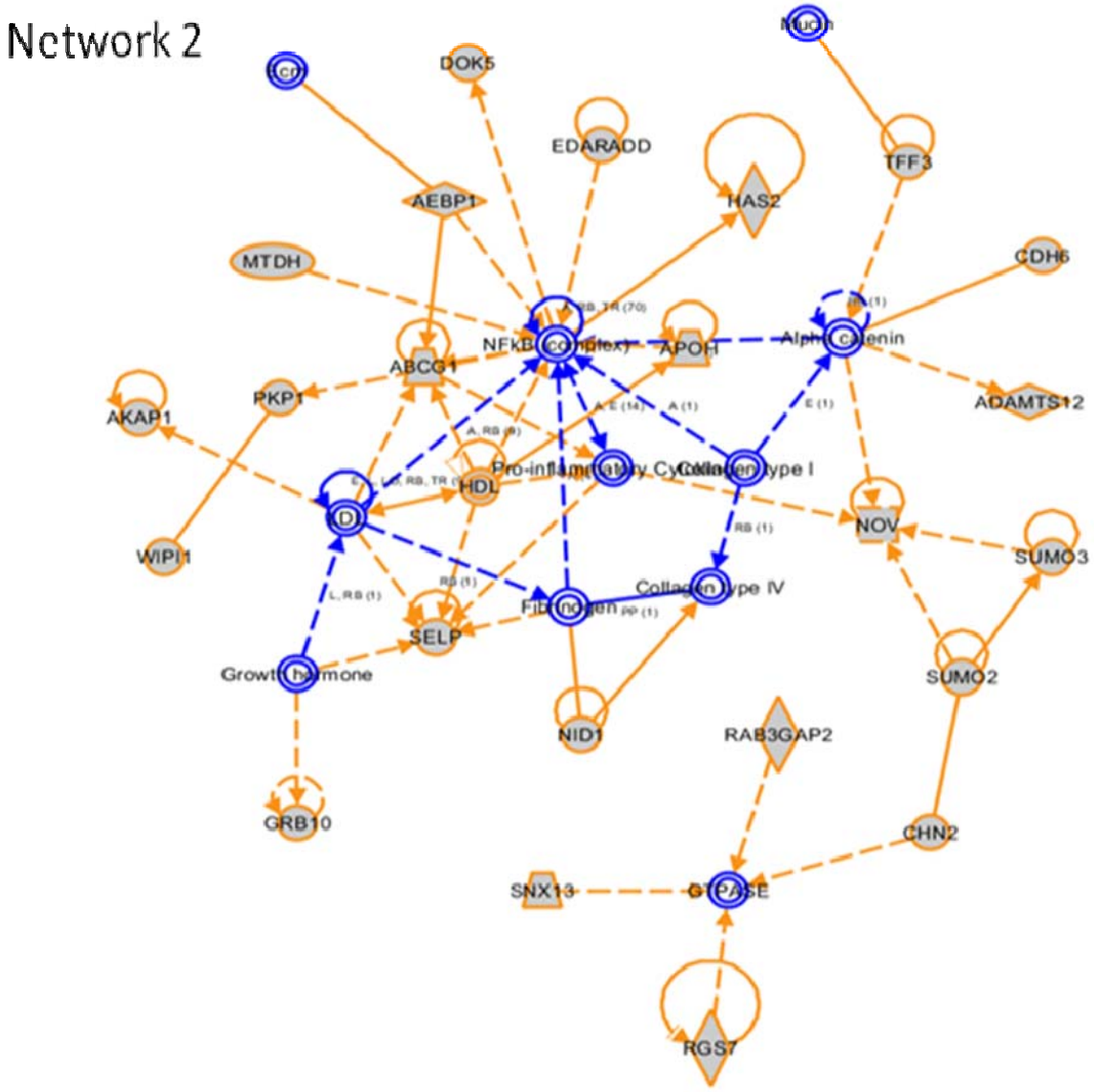


Figure 6.5B: Common Networks: Cell-To-Cell Signaling and Interaction, Tissue Development, Organismal Injury and Abnormalities (Network 2)

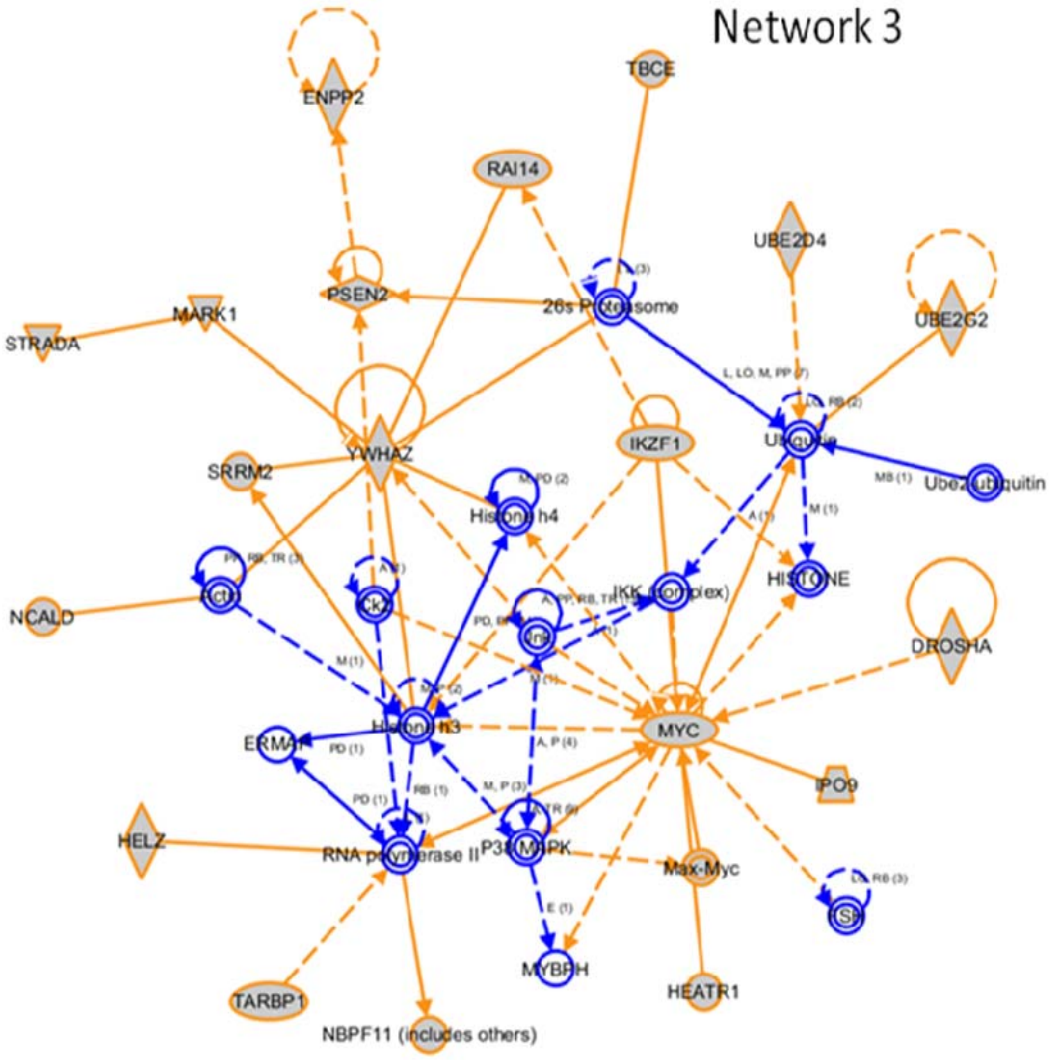


Figure 6.5C: Common Networks: Cell Morphology, Cellular Assembly and Organization, Cellular Compromise (Network 3).

Table 6.7: DAVID analysis of genes in BQC, NBQC and Common regions

Term	Count	%	PValue	Genes	Fold Enrichment	FDR
BQC						
GO:0016616~oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	4	14.814815	1.65E-04	CBR1, AKR1B15, AKR1B10, AKR1B1, CBR3	34.260181	0.1834789
GO:0016614~oxidoreductase activity, acting on CH-OH group of donors	4	14.814815	2.28E-04	CBR1, AKR1B15, AKR1B10, AKR1B1, CBR3	30.71602434	0.2533245
GO:0004090~carbonyl reductase (NADPH) activity	2	7.4074074	0.004220099	CBR1, CBR3	445.3823529	4.5938739
GO:0004033~aldo-keto reductase activity	2	7.4074074	0.015739444	AKR1B15, AKR1B10, AKR1B1	118.7686275	16.173153
GO:0005737~cytoplasm	14	51.851852	0.022484827	B3GALT5, CALD1, HLCS, CBR3, TTC3, CLDN14, LUZP6, CBR1, AKR1B15, KCNJ6, PCP4, AKR1B10,	1.601541769	20.970954

				MTPN, AKR1B1, SLC35B4, DOPEY2, TTC3L		
NBQC						
GO:0004984~olfactory receptor activity	6	14.285714	0.003323315	OR7C1, OR7A5, OR10H5, OR10H4, OR10H3, OR7A10	5.697497962	3.6069446
GO:0007608~sensory perception of smell	6	14.285714	0.003953067	OR7C1, OR7A5, OR10H5, OR10H4, OR10H3, OR7A10	5.458623357	5.4643509
hsa04740:Olfactory transduction	6	14.285714	0.0049168	OR7C1, OR7A5, OR10H5, OR10H4, OR10H3, OR7A10	4.735371721	3.7342917
GO:0007186~G-protein coupled receptor protein signaling pathway	9	21.428571	0.005300215	LOC100133511, LOC653879, C3, GHRH, OR7C1, OR7A5, EMR3, OR10H5, OR10H4, OR10H3, OR7A10	3.142475512	7.2621932

GO:0007606~sensory perception of chemical stimulus	6	14.285714	0.006115541	OR7C1, OR7A5, OR10H5, OR10H4, OR10H3, OR7A10	4.921896792	8.3347998
GO:0004888~transmembrane receptor activity	9	21.428571	0.010161503	PTPRD, OR7C1, PTPRS, OR7A5, EMR3, OR10H5, OR10H4, OR10H3, OR7A10	2.822553588	10.659248
GO:0007166~cell surface receptor linked signal transduction	11	26.190476	0.012301449	PTPRD, LOC653879, C3, OR7A5, OR10H5, VAV1, OR10H4, OR10H3, LOC100133511, GHRH, OR7C1, EMR3, OR7A10	2.323934387	16.104866
GO:0007600~sensory perception	7	16.666667	0.013638582	RAX2, OR7C1, OR7A5, OR10H5, OR10H4, OR10H3, OR7A10	3.38861454	17.701864
GO:0004930~G-protein coupled receptor activity	7	16.666667	0.015935723	OR7C1, OR7A5, EMR3, OR10H5, OR10H4,	3.281663106	16.245259

				OR10H3, OR7A10		
COMMON						
GO:0003012~muscle system process	8	3.539823	0.002927847	TNNT2, CHRM3, PSEN2, SNTB1, TBCE, RYR2, GJA5, MAP2K6	4.201190476	4.7360582
GO:0005488~binding	151	66.814159	0.003959613	RNASEN, CYP24A1, RAB3GAP2, ENAH, ADCY1, NOG, ADCY8, ACBD6, NOV, ERO1LB, RAD21, ATP2B4, KCNK9, SRRM2, KIFAP3, LRRC52, APOH, DISC1, MAP2K6, LOC100132779, EDARADD, NMNAT2, NCALD, SIPA1L2, ZNF648, RXRG, COLEC10, RAI14, NME7, MARK1, TNNT2,	1.092664085	5.4133962

	TNS3, GRB10, NAV1, F5, LRP12, RYR2, FBXL7, PGCP, AKAP1, EXT1, KIF26B, ABCA8, TSHZ2, ENPP2, ASAP1, RIMS2, URGCP, UBE2D4, LOC652798, PTK2, EIF3H, EFR3A, EFCAB2, SLC30A8, TMEM195, OBSCN, IKZF1, TRIO, HEATR1, TRIL, WIPI1, ABCG1, PKP1, TRPS1, TUBD1, WDR4, CDC42BPA, JAZF1, RGS7, AGXT2, CACNA1E, PDZD2, XCL1, SNORA25,

	AEBP1, TAF1A, TAF1D, UBE2G2, PRRX1, HELZ, RHOU, ZNF678, HMOX2, DIP2A, TNFRSF11B, HLX, RSPO2, SNTB1, ANGPT1, ADCY10, MYC, NACAP1, PRKCA, DDC, TTC35, ZP4, SNORA32, TBCE, ADIPOR1, IPO9, MOSC1, DNAJC21, TARBP1, FMN2, CHRM3, DOK5, PSEN2, PFDN4, KCNH6, SPEF2, KLHL12, ZFPM2, PDE9A, SNX13, GALNT2, ZNF554, YWHAZ, MTDH, SUMO2P, MAL2,	

				CD247, DPYS, IL7R, DNAH5, CDH6, SUMO3, FMO5, SUMO2, CHD1L, DGKB, DGKE, PDE1C, ETV1, WIPF3, HSF2BP, ADAMTS12, ABCA13, DPT, SELP, BCAS1, RNF19A, CAPN9, NFASC, STRADA, NID1, CREB5, LMX1A, CPVL, GORAB, KCNV1, MED30, PKNOX1, NPHS2, A2BP1, CHN2, SNX31		
GO:0008289~lipid binding	13	5.7522124	0.004034771	PRKCA, SELP, RXRG, ACBD6, WIPI1, ABCG1, DGKB, DGKE, CDC42BPA, APOH, CHN2, SNX31, SNX13	2.619547572	5.5134646

GO:0019992~diacylglycerol binding	5	2.2123894	0.00658016	PRKCA, DGKB, DGKE, CDC42BPA, CHN2	6.667400493	8.8450341
GO:0004016~adenylate cyclase activity	3	1.3274336	0.007334818	ADCY1, ADCY8, ADCY10	22.66916168	9.8116085
GO:0006936~muscle contraction	7	3.0973451	0.007718719	TNNT2, CHRM3, PSEN2, SNTB1, RYR2, GJA5, MAP2K6	4.036437908	12.033917
GO:0007017~microtubule-based process	9	3.9823009	0.007941366	FMN2, PTK2, RNF19A, KIFAP3, TBCE, TUBD1, MARK1, DNAH5, KIF26B	3.138438735	12.359952
GO:0030054~cell junction	13	5.7522124	0.011183692	ENAH, MTDH, RIMS2, RHOU, GJA5, CBLN4, LOC652798, TNS3, PTK2, PKP1, CHRM3, SNTB1, CDC42BPA, PDZD2	2.294457019	13.632653
GO:0042221~response to chemical stimulus	25	11.061947	0.009020685	CYP24A1, ADCY1, YWHAZ, ADCY8, ENPP2,	1.721799375	13.924437

				HMOX2, TNFRSF11B, TFF3, ANGPT1, SLC30A8, ADCY10, MYC, MAP2K6, OXR1, PRKCA, DDC, SELP, ADIPOR1, ABCG1, TNNT2, GRB10, PSEN2, RYR2, XCL1, PPP1R15B		
GO:0042474~middle ear morphogenesis	3	1.3274336	0.009064738	NOG, PRRX1, MYC	20.35961538	13.987732
GO:0032147~activation of protein kinase activity	6	2.6548673	0.009345971	ADCY1, DGKB, DGKE, ADCY8, STRADA, ADCY10	4.643421053	14.390772
GO:0000287~magnesium ion binding	12	5.3097345	0.011164497	RNASEN, NMNAT2, OBSCN, ADCY1, ATP2B4, ADCY8, CDC42BPA, ADCY10, RHOU, MARK1, NME7, EDARADD	2.407344603	14.571675
GO:0001501~skeletal system	10	4.4247788	0.010166031	PRKCA,	2.765673981	15.555908

development				CYP24A1, AEBP1, TNFRSF11B, NOG, TRPS1, PRRX1, EXT1, GJA5, MYC		
GO:0006171~cAMP biosynthetic process	3	1.3274336	0.010497869	ADCY1, ADCY8, ADCY10	18.90535714	16.023134
GO:0030554~adenyl nucleotide binding	28	12.389381	0.012736735	ABCA8, ADCY1, ADCY8, UBE2G2, HELZ, DNAH5, UBE2D4, FMO5, PTK2, ERO1LB, ATP2B4, DGKB, CHD1L, DGKE, PDE1C, ADCY10, ABCA13, MAP2K6, PRKCA, OBSCN, NMNAT2, STRADA, TRIO, NME7, MARK1, ABCG1, CDC42BPA, KIF26B	1.60998485	16.457444

hsa00230:Purine metabolism	6	2.6548673	0.019249014	ADCY1, PDE1C, ADCY8, PDE9A, ADCY10, NME7	3.762486127	18.98454
GO:0001883~purine nucleoside binding	28	12.389381	0.015297966	ABCA8, ADCY1, ADCY8, UBE2G2, HELZ, DNAH5, UBE2D4, FMO5, PTK2, ERO1LB, ATP2B4, DGKB, CHD1L, DGKE, PDE1C, ADCY10, ABCA13, MAP2K6, PRKCA, OBSCN, NMNAT2, STRADA, TRIO, NME7, MARK1, ABCG1, CDC42BPA, KIF26B	1.585850161	19.446843
GO:0001882~nucleoside binding	28	12.389381	0.016603244	ABCA8, ADCY1, ADCY8, UBE2G2, HELZ, DNAH5, UBE2D4, FMO5,	1.575028603	20.931819

	PTK2, ERO1LB, ATP2B4, DGKB, CHD1L, DGKE, PDE1C, ADCY10, ABCA13, MAP2K6, PRKCA, OBSCN, NMNAT2, STRADA, TRIO, NME7, MARK1, ABCG1, CDC42BPA, KIF26B	

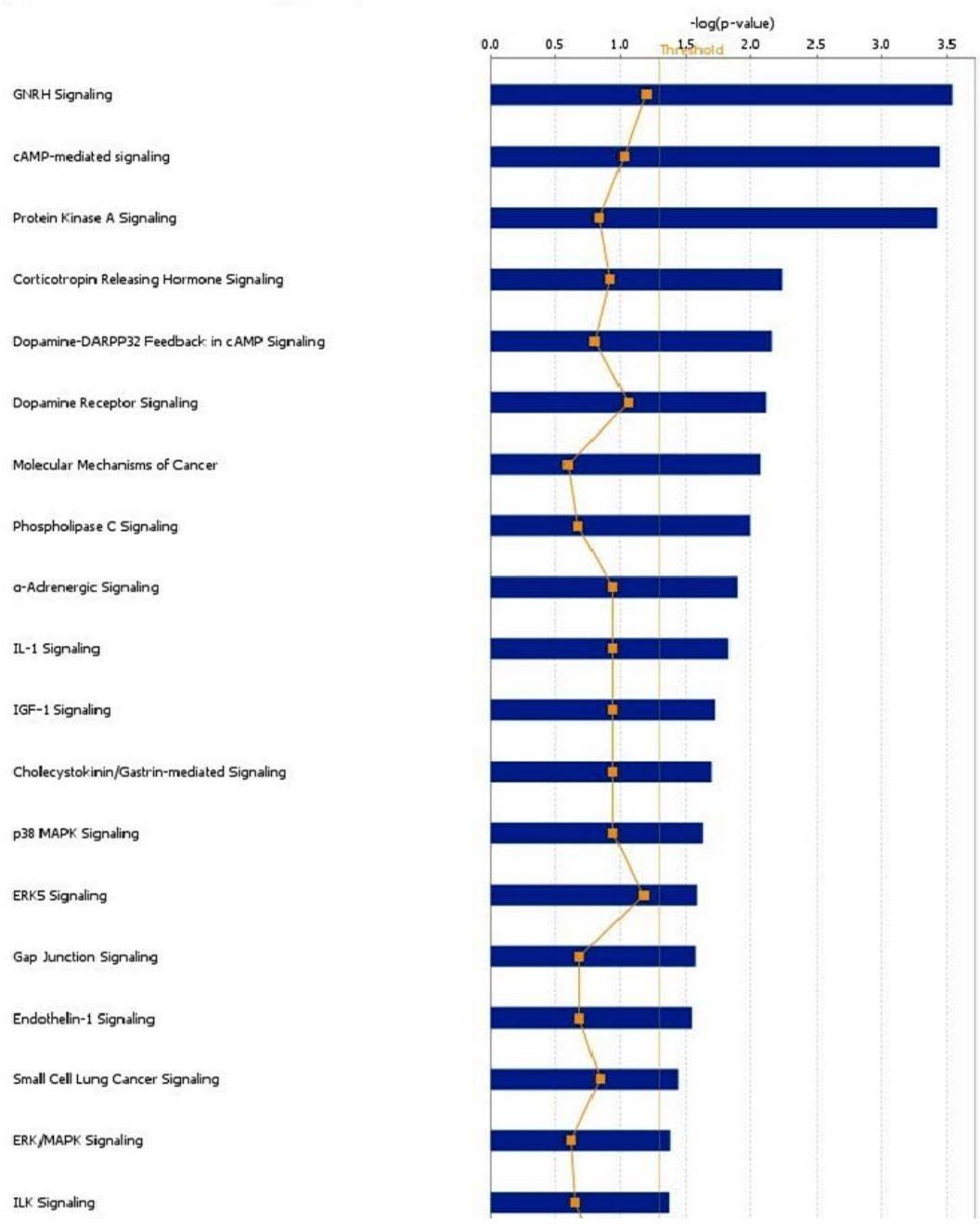


Figure 6.6: Cancer related Canonical pathways enriched from common genes.

6.5. Discussion

Two chromosomal regions, 7q33 and 21q22.3 presented more alterations in BQC tumors (gains) than NBQC tumors. Gain of 7q33 region has been previously reported in pancreatic and lung carcinoma (Nowak, N. J. et al. 2005 and Engelman, J. A. et al. (2007). Gain of 21q22.3 has previously been described in cholangiocarcinoma and as one of the predictive marker regions of systemic recurrence in breast cancer (Hwang, K. T. et al. 2008 and Muenphon, K. et al. 2006). GO terms, oxidoreductase and aldo-keto reductase activity were enriched with a single drug metabolism, molecular transport, nucleic acid metabolism network. AKR1B1and AKR1B10 genes were seen playing cardinal roles. AKR1B10 is overexpressed in colorectal, uterine, breast cancers. Considered a diagnostic marker in lung cancer, it may play a pathogenic role in hepatocellular carcinoma. Role of AKR1B10 in tobacco-related carcinogenesis is anticipated because of its overexpression observed in bronchial epithelium of smokers. Its expression which is stimulated by tobacco smoke condensate in normal human epidermal, oral and squamous cell carcinoma cells decreases with the cessation of smoking. Proposed AKR1B10-mediated tumorigenic mechanisms include retinoic acid depletion and cancer cell dedifferentiation as well as chemoresistance due to metabolism of carbonyl group–bearing anticancer drugs and activating pro-carcinogens and polycyclic aromatic hydrocarbon (PAH) transdihydrodiols to biologically reactive and redox-active *o*quinones (Liu, J. et al. 2009, Barski, O. A. et al. 2008 and Diez-Dacal, B. et al.

2011). Hence, the tobacco component in BQ may explain AKR1B10 gain rendering chemoresistance, dedifferentiation and DNA adduct formation in BQC leading to breast carcinogenesis. In addition, AKR1B1 contributes in regulating multiple inflammatory pathways and its inhibition has been shown to interrupt inflammation triggered by chemokines, growth factors and inflammatory cytokines such as TNF-α as depicted in our network (Diez-Dacal, B. et al. 2011). The ROS generated by the presence of slaked lime in BQ may amplify AKR1B1 gene rendering TNFα induced proliferation of breast cancer cells. Besides, the above network analysis also manifested role of ETS2 gene which maintains hTERT gene expression by interacting with the c-Myc transcription factor. It is a central driver of a transcriptional program in tumor associated macrophages that acts to promote lung metastasis of breast tumors (Xu, D. et al. (2008 and Zabuawala, T. et al. 2010).

Our data also presented with a high significance of gain in fragile sites in BQC tumors as compared with NBQC. Fragile sites form gaps, constrictions and breaks on chromosomes when exposed to partial replication stress and are rearranged in tumors. Frequency of fragile sites and sister chromatid exchanges have been found to be significantly higher in smokers in peripheral lymphocytes and bone marrow (Ban, S. et al. 1995 and Kao-Shan, C. S. et al. 1987). The above ascertains the potential of BQ carcinogens in causing chromosomal damage and instability leading to genetic alterations.

Since metabolic absorption of the ingredients of BQ directs the cancer-causing principles to other organs/tissues of the body, the evidence is growing to indicate that cancers other than oropharyngeal may also be caused by BQ chewing (Chatterjee, A. and S. Deb 1999). Tobacco related carcinogens can be stored in breast adipose tissue, metabolized and activated by human mammary epithelial cells (Terry, P. D. and T. E. Rohan 2002) . Moreover, the evidence that tobacco exposure (smoking) causes early gene expression changes in normal airway epithelial cells and many cancer types (Schembri, F. et al. 2009) the observed changes likely reflect the early carcinogenesis.

Among the NBQC tumors, two regions 19p13.3-19p12 and 20q11.22 presented more alterations. Gain observed in 19p13.3-19p12 has previously been reported in cutaneous and oral squamous cell carcinomas (Ambatipudi, S. et al. 2011and Purdie, K. J. et al. 2007). Gain at 20q11.2 has been observed in breast, colorectal and cervical cancers (Scotto, L. et al. 2008, Hodgson, J. G. et al. 2003 and Nakao, K. et al. 2004). GO terms, G-protein coupled receptor protein signaling pathway and cell surface receptor linked signal transduction were enriched with two networks, molecular transport, nucleic acid metabolism, small molecule biochemistry and cellular development, embryonic development, organismal development. RPN2, EMR3, VAV1, NNAT and MUC16 genes were recognized to have imperative functions. A recent study by Kimi Honma *et al* reported that RPN2 silencing and downregulation makes cancer cells hypersensitive in response to docetaxel a chemotherapeutic drug, proposing it as

a target for RNA interference–based therapeutics against drug resistance (Honma, K. et al. 2008). EMR-3 a G-protein coupled receptor, upregulated in glioblastoma is associated with poor survival and is a potential mediator of cellular invasion (Kane, A. J. et al. 2010). VAV1 contributes to tumorgenesis by regulating both cellular proliferation and cell survival pathways through the regulation of an EGF-Src-Vav1-Rac1-Pak1-NF-κ B-Cyclin D1 signaling axis. An increased and ectopic expression of VAV1 in lung and pancreatic tumors has been linked to large tumors and worse survival rate respectively (Fernandez-Zapico, M. E. et al. 2005 and Lazer, G. et al. 2009). mRNA expression of neuronatin (NNAT) has been reported in pituitary adenoma, prostatic cancer with neuroendocrine features, large cell neuroendocrine carcinoma lung and thyroid stimulating hormone-producing tumors in mice. High expression has been reported in a tamoxifen-resistant mammary carcinoma cell line (Uchihara, T. et al. 2007). Decreasing MUC16 levels are known to be of prognostic outcome in the post-operative and pre-operative neo-adjuvant chemotherapy especially in ovarian carcinoma (Vasudev, N. S. et al. 2011). Recent studies by Silke Reinartz et al and I Lakshmanan et al elucidated its central role in adhesion, migration and invasion in breast cancer. Overexpressed in breast cancer, it augments cell proliferation by interacting with JAK2 and inhibiting the apoptotic process through downregulation of TRAIL (Lakshmanan, I. et al. 2012 and Reinartz, S. et al. 2012). Since the NBQC tumors had no previous history of BQ and any other

environmental exposure, gain in genes regulating various facets of tumorigenesis can only be blamed as spontaneous instances arising in NBQC tumors.

Besides differences, both tumor groups shared twenty seven frequently altered regions. The IPA analyses resulted in three top networks and eight tumor associated canonical pathways. Extrapolating the data depicted ERK 1/2 and PTK2 (network 1), NFkB complex, SELP and NOV (network 2), MYC and YWHAZ (network 3) were key nodes in their respective networks.

A review by Ming Luo *et al* on PTK2 (FAK) described its principal role in breast carcinogenesis. As depicted in our network 1, PTK2 served as a mediator of cell cycle regulation by integrins through PTK2/Src complex formation in the focal contacts promoting ERK activation. Mechanistic studies indicate that PTK2 deletion in mammary tumor cells reduces the expression/phosphorylation of ERK1/2 contributing to the tumor dormancy in *vivo* and arrests growth in cultures suggesting PTK2 signaling through ERK-MAPK pathway is required to maintain tumor cell growth. In addition to Rac, PTK2 also mediates the activation of ERK to promote cell migration (Luo, M. and J. L. Guan 2010).

Network 2 witnessed SELP, NOV and NFκB complex as vital genes. NF-kB plays a key role in regulating the immune response and incorrect regulation of NF-kB has been linked to the development of cancer. Signaling pathways leading to tamoxifen resistance in breast cancer share a common mechanistic link with activation of nuclear factor-kB (NFkB) (Zhou, Y. et al. 2005). Elevated levels of

SELP have been observed in many cancers including melanoma, tongue, colon, gastric, lung and breast. SELP is an adhesion molecules that mediate cell-cell interactions among platelets) and endothelial cells. Its measurement may provide a sensitive tool for monitoring the clinical course of melanoma and lymphoma (Ferroni, P. et al. 2004). High expression levels of NOV are associated with endocrine therapy crossresistance in CL6.7 cells and endocrine therapy resistance in breast tumor samples proliferation (Ghayad, S. E. et al. 2009). NOV enhances migration of chondrosarcoma cells by increasing MMP-13 expression through αvβ3/αvβ5 integrin receptor, FAK, PI3K, Akt, p65, and NF-κB signal transduction pathway and regulates the differentiation of bone resident cells creating a resorptive environment that promotes the formation of osteolytic breast cancer metastases (Tzeng, H. E. et al. 2011 and Ouellet, V. et al. 2011).

YWHAZ (14-3-3ζ) seen in network 3, overexpressed in breast, lung and many other cancers is implicated in the initiation and progression of cancer (Neal, C. L. and D. Yu. 2010). Low level copy number gains in YWHAZ have been found in head and neck squamous cell carcinomas (Lin, M. et al. 2009). Previous studies documenting YWHAZ upregulation and a poor clinical outcome in tamoxifen treated breast cancer patients imply it to be a marker of poor prognosis in women with ER-positive breast cancers (Bergamaschi, A. and B. S. Katzenellenbogen 2012). The oncogenic Myc protein in network 3 plays an important role in breast cancer metastasis and several transcription factors are involved in the regulation of Myc expression. In breast cancer, amplification of c-

myc may correlate positively or negatively with alterations in other genes (Liao, D. J. and R. B. Dickson 2000). For e.g. as revealed by our network 3 heterodimerization with Max is necessary for c-Myc to mediate proliferation, transformation, and apoptosis (Sakamuro, D. and G. C. Prendergast 1999). Recent studies have indicated that Myc is an IKKs substrate and IKKs tightly regulate Myc expression in breast cancers as also seen in network 3 (Yeh, P. Y. et al.2011).

Alterations seen in the preceding genes can be seen as vital as they arise independently of the etiological factors, signifying the abovementioned genes importance in breast tumorigenesis. In addition, direct or indirect association of these key network genes to other cancer related genes (for example, MTDH, EXT1, ANGPT1, RAD21, EDARADD, TFF3, MARK1, DROSHA, etc seen in our networks) could create a permissive context activating or deactivating various facets of breast tumorigenesis. Super inducing these common alterations, AKR1B10, AKR1B1 and ETS2 alterations were BQ induced whereas alterations in RPN2, EMR3, VAV1, NNAT and MUC16 genes in NBQC tumors could only be termed as spontaneous. Diagrammatic illustration of the same is shown in *figure 6.7 and 6.8*

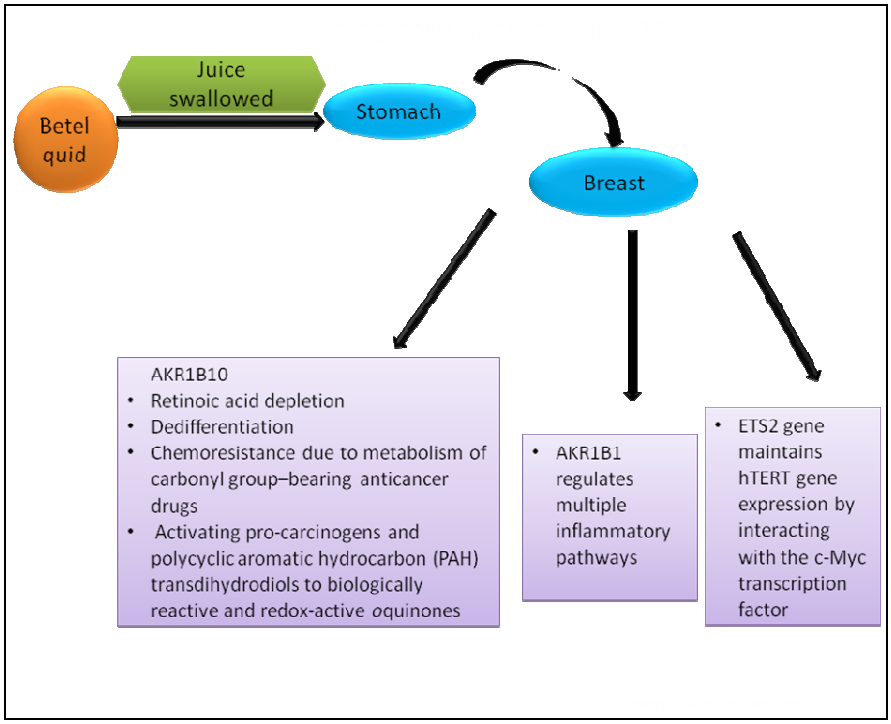


Figure: 6.7 Tumor genesis mechanism in BQC

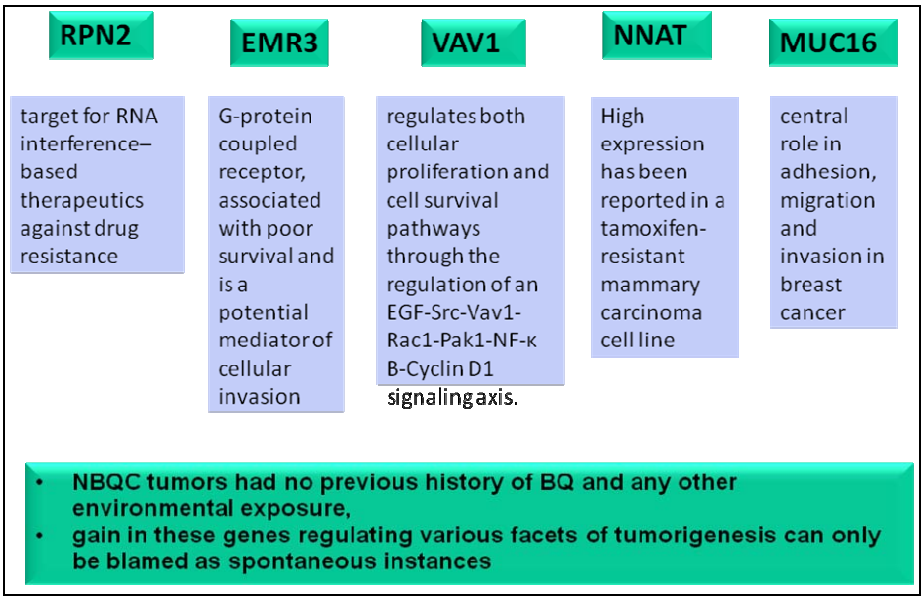


Figure: 6.8 Tumor genesis mechanisms in NBQC

It is important to acknowledge that apart from environmental factor such as betel quid being the prime focus of this study, genetic risk factors such as BRCA1 and BRCA2, lifestyle risk factors such as diet and reproductive risk factors also contribute to breast cancer. BRCA1 and BRCA2 mutant carriers impose a highly increased risk for hereditary or familial breast cancer. While our study is specifically based on sporadic tumors, BRCA2 mutation analysis performed on a larger set of samples in our unpublished study showed none of the tumors to be BRCA2 mutation positive. Therefore, likelihood of our samples containing BRCA1 mutations would still be minute if the probability of BRCA1 and BRCA2 mutations taken together is estimated to be 5%, equal to the proportion in total breast cancer incidence. Examination of impact of BRCA1 and BRCA2 mutations on copy number alteration illustrates a significant difference of genomic profiles between BRCA1 and sporadic tumors, followed by BRCA1 and BRCA2 tumors. BRCA2 and sporadic tumors (such as in our study) had very similar genomic profiles. Overall, BRCA1 tumors have a higher frequency of copy number alterations (Jonsson, G. et al. 2005) implying that high risk cases of BRCA1 mutant carriers if subjected to environmental toxicants like betel quid could exemplify the effects resulting in aggressive and early tumors. Furthermore, lifestyle factor like diet has been implicated as an important determinant of breast cancer. The diet pattern in Northeast population of India is mainly characterized by high intakes of dry fish and fermented soybean and vegetables (Das, A., J. and Deka, S., C. 2012). Such dietary pattern rich in

vegetables and fish, but poor in red meats and animal fats has been positively associated to a longer overall survival of breast cancer. However obese women have increased risk for breast cancer as they are exposed to high levels of estrogen additionally produced by adipose tissue (Dal Maso, L. et al. 2008). Reproductive factors, including age at menarche, age at first full-term pregnancy, number of live births and breast-feeding are related to a risk of breast cancer. Mechanism through which reproductive exposures influence breast cancer risk is their effect on lifetime number of menstrual cycles. Number of menstrual cycles influences the lifetime exposure to endogenous ovarian hormones like estrogen, which is strongly related to breast cancer risk (Beaber, E. F. et al. 2008). Estrogen when metabolized produces metabolites which further contribute to tumor initiation by activating estrogen receptor and generating DNA damaging molecular species (Sangrajrang, S. et al. 2009). In our previous study breast cancer risk was not associated with any of the reproductive factors and polymorphism in an estrogen synthesizing CYP17 gene in the Northeast population of India (Kaushal, M. et al. 2010). However, examination of the effect of lifestyle factors and reproductive factors on copy number alteration yet remains to be investigated. The foregoing further ascertains that the effects seen in the present study are due to betel quid chewing.

To our knowledge this is the first report of comparison of genomic alterations between BQ and NBQ chewer breast cancer patients. Overall our data agree well with previous genomic alteration analysis. The major strength of

this study is its homogeneous sample population, presence of only BQ as an environmental exposure variables and detailed demographic information. As a limitation, analysis of a larger sample set and cell systems is clearly needed to more precisely delineate the molecular basis for both BQC and NBQC breast tumors. Despite that the accuracy of our results is justified due to unbiased sample distribution in both groups and FDR adjustments. Since composition of BQ in this region consists of multiple components, assessing carcinogenic effect of individual constituent was not possible in this study. Application of high resolution arrays may elicit additional regions of differential alteration. Unfortunately, such studies are largely precluded by the relative rarity of appropriate specimens. However, biological information obtained from BQ exposed breast cancer subset is valuable. This subgroup is frequent in the North East Indian population as most of the women in this area are usually chewers. Given a unique set of underlying genomic changes, distinct approaches to treatment may be appropriate for this patient population and others where this habit is highly prevalent.

Chapter 7

Conclusions and Future Scope of Work

Chapter 7

Conclusions and Future Scope of Work

7.1. Conclusion

1. Betel quid chewing has identified as the main risk factor substantiated as the predominant risk factor by all three analytical approaches (Logistic regression, MDR and CART). Women with betel quid chewing history had five times the risk of developing breast cancer.
2. GSTT1 null and GSTM1 null genotypes conferred 41 per cent less and 55 per cent less reduced risk to breast cancer respectively.
3. Alcohol consumption was significantly associated with increased risk for breast cancer.
4. Homozygous AA genotype of CCND1 gene imparted significant risk in both NBQC and BQC subsets. CART analysis illustrated maximum risk amongst BQC subset with and CCND1 GG genotype followed by combination of CCND1 AA, AG genotypes and non smoker subset. In NBQC group risk was seen comprising of NBQC and non tobacco chewers
5. Homozygous AA genotype of TP53 gene imparted significant risk in NBQC subset whereas the C allele of RAD51 gene was significantly over-represented in the BQC subset along with mutations in the BRCA2 gene seen in two samples.

6. BQC tumors showed significantly higher total number of alterations, as compared with NBQC tumors, 48 ± 17 % versus 32 ± 25 respectively.
7. Incidence of gain in fragile sites in BQC tumors was significantly higher as compared with NBQC tumors, 34 versus 23% respectively.
8. 7q33 and 21q22.13 chromosomal regions were significantly associated with BQC tumors. One network “Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism” including genes AKR1B1, AKR1B10, ETS2 etc in BQC was enriched.
9. 19p13.3-19p12 and 20q11.22 chromosomal regions were significantly associated with NBQC tumors. Two networks “Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry” and “Cellular Development, Embryonic Development, Organismal Development” including genes RPN2, EMR3, VAV1, NNAT and MUC16 etc were seen in NBQC.
10. Common alterations (>30%) were seen in 27 regions. Three networks were significant in common regions with key roles of PTK2, RPN2, EMR3, VAV1, NNAT, MUC16, MYC and YWHAZ genes.

7.2. Future Scope of the Thesis

1. AKR1B10, AKR1B1 and ETS2 alterations which were betel quid induced could act as biomarkers for breast cancer susceptibility in areas with high tobacco usage .

2. Alterations in RPN2, EMR3, VAV1, NNAT and MUC16 genes seen in non betel quid chewer tumors illustrate additional breast cancer networks should be investigated further.

3. Direct or indirect association of the common key network genes to other cancer related genes (for example, MTDH, EXT1, ANGPT1, RAD21, EDARADD, TFF3, MARK1, DROSHA, etc seen in the networks) could act as candidates for drug targeting and therapy as they could create a permissive context activating or deactivating various facets of breast tumorigenesis.

4. Given a unique set of underlying genomic changes, distinct approaches to treatment may be appropriate for this patient population and others where this habit is highly prevalent.

References

~REFERENCES~

- Ahmad, N. and R. Kumar (2011) "Steroid hormone receptors in cancer development: a target for cancer therapeutics." Cancer Lett 300(1): 1-9.

- Alguacil, J. and D. T. Silverman (2004). "Smokeless and other noncigarette tobacco use and pancreatic cancer: a case-control study based on direct interviews." Cancer Epidemiol Biomarkers Prev 13(1): 55-8.

- Ambatipudi, S., M. Gerstung, et al.(2011) "Genomic profiling of advanced-stage oral cancers reveals chromosome 11q alterations as markers of poor clinical outcome." PLoS One 6(2): e17250.

- Anantharaman, D., P. M. Chaubal, et al. (2007). "Susceptibility to oral cancer by genetic polymorphisms at CYP1A1, GSTM1 and GSTT1 loci among Indians: tobacco exposure as a risk modulator." Carcinogenesis 28(7): 1455-62.

- Badve, S., D. J. Dabbs, et al.(2011) "Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists." Mod Pathol 24(2): 157-67.

- Ban, S., J. B. Cologne, et al. (1995). "Effect of radiation and cigarette smoking on expression of FUdR-inducible common fragile sites in human peripheral lymphocytes." Mutat Res 334(2): 197-203.

- Baras, A., Y. Yu, et al. (2009). "Combined genomic and gene expression microarray profiling identifies ECOP as an upregulated gene in squamous cell carcinomas independent of DNA amplification." Oncogene 28(32): 2919-24.

- Barski, O. A., S. M. Tipparaju, et al. (2008). "The aldo-keto reductase superfamily and its role in drug metabolism and detoxification." Drug Metab Rev 40(4): 553-624.

- Beaber, E. F., V. L. Holt, et al. (2008). "Reproductive factors, age at maximum height, and risk of three histologic types of breast cancer." Cancer Epidemiol Biomarkers Prev 17(12): 3427-34.

- Benjamini, Y., D. Drai, et al. (2001). "Controlling the false discovery rate in behavior genetics research." Behav Brain Res 125(1-2): 279-84.

- Bergamaschi, A. and B. S. Katzenellenbogen (2012)"Tamoxifen downregulation of miR-451 increases 14-3-3zeta and promotes breast cancer cell survival and endocrine resistance." Oncogene 31(1): 39-47.

- Bergamaschi, A., Y. H. Kim, et al. (2006). "Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer." Genes Chromosomes Cancer 45(11): 1033-40.

- Bhattacharjee, A., A. Chakraborty, et al. (2006). "Prevalence of head and neck cancers in the north east-An institutional study." Indian J Otolaryngol Head Neck Surg 58(1): 15-9.

- Bieche, I., M. Olivi, et al. (2002). "Prognostic value of CCND1 gene status in sporadic breast tumours, as determined by real-time quantitative PCR assays." Br J Cancer 86(4): 580-6.

- Boffetta, P. and M. Hashibe (2006). "Alcohol and cancer." Lancet Oncol 7(2): 149-56.

- Boffetta, P., B. Aagnes, et al. (2005). "Smokeless tobacco use and risk of cancer of the pancreas and other organs." Int J Cancer 114(6): 992-5.

- Boffetta, P., S. Hecht, et al. (2008). "Smokeless tobacco and cancer." Lancet Oncol 9(7): 667-75.

- Brennan, P., O. Bogillot, et al. (2001). "The contribution of cigarette smoking to bladder cancer in women (pooled European data)." Cancer Causes Control 12(5): 411-7.

- Briollais, L., Y. Wang, et al. (2007). "Methodological issues in detecting gene-gene interactions in breast cancer susceptibility: a population-based study in Ontario." BMC Med 5: 22.

- Brownson, R. C., L. W. Figgs, et al. (2002). "Epidemiology of environmental tobacco smoke exposure." Oncogene 21(48): 7341-8.

- Buchhop, S., M. K. Gibson, et al. (1997). "Interaction of p53 with the human Rad51 protein." Nucleic Acids Res 25(19): 3868-74.

- Cao, G., H. Lu, et al. (2008). "Lung cancer risk associated with Thr495Pro polymorphism of GHR in Chinese population." Jpn J Clin Oncol 38(4): 308-16.

- Catarino, R. J., E. Breda, et al. (2006). "Association of the A870G cyclin D1 gene polymorphism with genetic susceptibility to nasopharyngeal carcinoma." Head Neck 28(7): 603-8.

- Catarino, R., A. Matos, et al. (2005). "Increased risk of cervical cancer associated with cyclin D1 gene A870G polymorphism." Cancer Genet Cytogenet 160(1): 49-54.

- Cattaert, T., M. L. Calle, et al.(2011) "Model-based multifactor dimensionality reduction for detecting epistasis in case-control data in the presence of noise." Ann Hum Genet 75(1): 78-89.

- Ceschi, M., C. L. Sun, et al. (2005). "The effect of cyclin D1 (CCND1) G870A-polymorphism on breast cancer risk is modified by oxidative stress among Chinese women in Singapore." Carcinogenesis 26(8): 1457-64.

- Chacko, P., B. Rajan, et al. (2004). "CYP17 and SULT1A1 gene polymorphisms in Indian breast cancer." Breast Cancer 11(4): 380-8.

- Chakraborty, A., N. S. Murthy, et al. (2007). "CYP17 gene polymorphism and its association with high-risk north Indian breast cancer patients." J Hum Genet 52(2): 159-65.

- Chatterjee, A. and S. Deb (1999). "Genotoxic effect of arecoline given either by the peritoneal or oral route in murine bone marrow cells and the influence of N-acetylcysteine." Cancer Lett 139(1): 23-31.

- Chaturvedi, H. K., R. K. Phukan, et al. (2003). "The association of selected sociodemographic factors and differences in patterns of substance use: a pilot study in selected areas of Northeast India." Subst Use Misuse 38(9): 1305-22.

- Chaudru, V., M. T. Lo, et al. (2009). "Protective effect of copy number polymorphism of glutathione S-transferase T1 gene on melanoma risk in presence of CDKN2A mutations, MC1R variants and host-related phenotypes." Fam Cancer 8(4): 371-7.

- Chen, W. Y. (2008). "Exogenous and endogenous hormones and breast cancer." Best Pract Res Clin Endocrinol Metab 22(4): 573-85.

- Chiang, W. F., P. S. Hung, et al.(2011) "Increase of ZASC1 gene copy number in recurrent oral carcinoma." Oral Dis 17(1): 53-9.

- Chopra, R. (2001). "The Indian scene." J Clin Oncol 19(18 Suppl): 106S-111S.

- Comstock, C. E., M. A. Augello, et al. (2009). "Cyclin D1 splice variants: polymorphism, risk, and isoform-specific regulation in prostate cancer." Clin Cancer Res 15(17): 5338-49.

- Cordell, H. J. (2009). "Detecting gene-gene interactions that underlie human diseases." Nat Rev Genet 10(6): 392-404.

- Costa, S., D. Pinto, et al. (2008). "Importance of TP53 codon 72 and intron 3 duplication 16bp polymorphisms in prediction of susceptibility on breast cancer." BMC Cancer 8: 32.

- Coughlin, S. S. and M. Piper (1999). "Genetic polymorphisms and risk of breast cancer." Cancer Epidemiol Biomarkers Prev 8(11): 1023-32.

- Dal Maso, L., A. Zucchetto, et al. (2008). "Effect of obesity and other lifestyle factors on mortality in women with breast cancer." Int J Cancer 123(9): 2188-94.

- Das, A., J., Deka, S., C. (2012)" Fermented foods and beverages of the North-East India." International Food Research Journal 19(2): 377–392.

- David J. Hunter, Graham A. Colditz, Susan E. Hankinson, et al. (2010) "Oral Contraceptive Use and Breast Cancer: A Prospective Study of Young Women." Cancer Epidemol Biomarkers Prev 2010;19:2496-2502

- de Jong, M. M., I. M. Nolte, et al. (2002). "Genes other than BRCA1 and BRCA2 involved in breast cancer susceptibility." J Med Genet 39(4): 225-42.

- Diez-Dacal, B., J. Gayarre, et al.(2011) "Identification of aldo-keto reductase AKR1B10 as a selective target for modification and inhibition by prostaglandin A(1): implications for antitumoral activity." Cancer Res 71(12): 4161-71.

- Dunning, A. M., C. S. Healey, et al. (1999). "A systematic review of genetic polymorphisms and breast cancer risk." Cancer Epidemiol Biomarkers Prev 8(10): 843-54.

- Engelman, J. A., K. Zejnullahu, et al. (2007). "MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling." Science 316(5827): 1039-43.

- Evans, A. J., W. D. Henner, et al. (2004). "Polymorphisms of GSTT1 and related genes in head and neck cancer risk." Head Neck 26(1): 63-70.

- Fang, M., J. Toher, et al. (2011)"Genomic differences between estrogen receptor (ER)-positive and ER-negative human breast carcinoma identified by single nucleotide polymorphism array comparative genome hybridization analysis." Cancer.

- Fernandez-Zapico, M. E., N. C. Gonzalez-Paz, et al. (2005). "Ectopic expression of VAV1 reveals an unexpected role in pancreatic cancer tumorigenesis." Cancer Cell 7(1): 39-49.

- Ferroni, P., M. Roselli, et al. (2004). "Prognostic value of soluble P-selectin levels in colorectal cancer." Int J Cancer 111(3): 404-8.

- George, R. E., E. F. Attiyeh, et al. (2007). "Genome-wide analysis of neuroblastomas using high-density single nucleotide polymorphism arrays." PLoS One 2(2): e255.

- Ghayad, S. E., J. A. Vendrell, et al. (2009). "Identification of TACC1, NOV, and PTTG1 as new candidate genes associated with endocrine therapy resistance in breast cancer." J Mol Endocrinol 42(2): 87-103.

- Gochhait, S., S. I. Bukhari, et al. (2007). "Implication of BRCA2 -26G>A 5' untranslated region polymorphism in susceptibility to sporadic breast cancer and its modulation by p53 codon 72 Arg>Pro polymorphism." Breast Cancer Res 9(5): R71.

- Goode, E. L., C. M. Ulrich, et al. (2002). "Polymorphisms in DNA repair genes and associations with cancer risk." Cancer Epidemiol Biomarkers Prev 11(12): 1513-30.

- Gu, J., Y. Horikawa, et al. (2008). "Benzo(a)pyrene diol epoxide-induced chromosome 9p21 aberrations are associated with increased risk of bladder cancer." Cancer Epidemiol Biomarkers Prev 17(9): 2445-50.

- Gupta, P. C. and C. S. Ray (2004). "Epidemiology of betel quid usage." Ann Acad Med Singapore 33(4 Suppl): 31-6.

- He, X. F., J. Su, et al.(2011) "Association between the p53 polymorphisms and breast cancer risk: meta-analysis based on case-control study." Breast Cancer Res Treat 130(2): 517-29.

- Herzog, C. R., D. Desai, et al. (2006). "Array CGH analysis reveals chromosomal aberrations in mouse lung adenocarcinomas induced by the human lung carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone." Biochem Biophys Res Commun 341(3): 856-63.

- Hodgson, J. G., K. Chin, et al. (2003). "Genome amplification of chromosome 20 in breast cancer." Breast Cancer Res Treat 78(3): 337-45.

- Hong, Y., K. W. Eu, et al. (2005). "GG genotype of cyclin D1 G870A polymorphism is associated with increased risk and advanced colorectal cancer in patients in Singapore." Eur J Cancer 41(7): 1037-44.

- Honma, K., K. Iwao-Koizumi, et al. (2008). "RPN2 gene confers docetaxel resistance in breast cancer." Nat Med 14(9): 939-48.

- Hu, N., C. Wang, et al. (2006). "Genome-wide loss of heterozygosity and copy number alteration in esophageal squamous cell carcinoma using the Affymetrix GeneChip Mapping 10 K array." BMC Genomics 7: 299.

- Hu, Z., X. Li, et al.(2010) "Three common TP53 polymorphisms in susceptibility to breast cancer, evidence from meta-analysis." Breast Cancer Res Treat 120(3): 705-14.

- Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc 4(1): 44-57.

- Hung, R. J., P. Boffetta, et al. (2004). "GST, NAT, SULT1A1, CYP1B1 genetic polymorphisms, interactions with environmental exposures and bladder cancer risk in a high-risk population." Int J Cancer 110(4): 598-604.

- Hwang, K. T., W. Han, et al. (2008). "Genomic copy number alterations as predictive markers of systemic recurrence in breast cancer." Int J Cancer 123(8): 1807-15.

- IARC (2012) Available: http://monographs.iarc.fr/ENG/Monographs /vol85/index.php Accessed: 31 January 2012.

- Ihsan, R., P. S. Chauhan, et al. (2011) "Multiple analytical approaches reveal distinct gene-environment interactions in smokers and non smokers in lung cancer." PLoS One 6(12): e29431.

- Jakubowska, A., S. A. Narod, et al. (2003). "Breast cancer risk reduction associated with the RAD51 polymorphism among carriers of the BRCA1 5382insC mutation in Poland." Cancer Epidemiol Biomarkers Prev 12(5): 457-9.

- Jemal, A., F. Bray, et al. (2011) "Global cancer statistics." CA Cancer J Clin 61(2): 69-90.

- Jia, A., J. Gong, et al. (2008). "GG genotype of cyclin D1 G870A polymorphism is associated with non-cardiac gastric cancer in a high-risk region of China." Scand J Gastroenterol 43(11): 1353-9.

- Jonsson, G., T. L. Naylor, et al. (2005). "Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization." Cancer Res 65(17): 7612-21.

- Kane, A. J., M. E. Sughrue, et al.(2010) "EMR-3: a potential mediator of invasive phenotypic variation in glioblastoma and novel therapeutic target." Neuroreport 21(16): 1018-22.

- Kao-Shan, C. S., R. L. Fine, et al. (1987). "Increased fragile sites and sister chromatid exchanges in bone marrow and peripheral blood of young cigarette smokers." Cancer Res 47(23): 6278-82.

- Karami, S., P. Boffetta, et al. (2008). "Renal cell carcinoma, occupational pesticide exposure and modification by glutathione S-transferase polymorphisms." Carcinogenesis 29(8): 1567-71.

- Kaushal, M., A. K. Mishra, et al.(2010) "Betel quid chewing as an environmental risk factor for breast cancer." Mutat Res 703(2): 143-8.

- Kaushal, M., A. K. Mishra, et al.(2012) "Genomic alterations in breast cancer patients in betel quid and non betel quid chewers." PLoS One 7(8): e43789.

- Kaushal, M., I. Chattopadhyay, et al.(2009) "Contribution of germ line BRCA2 sequence alterations to risk of familial esophageal cancer in a high-risk area of India." Dis Esophagus 23(1): 71-5.

- Kelsey, K. T., S. E. Hankinson, et al. (1997). "Glutathione S-transferase class mu deletion polymorphism and breast cancer: results from prevalent versus incident cases." Cancer Epidemiol Biomarkers Prev 6(7): 511-5.

- Kenemans, P., R. A. Verstraeten, et al. (2004). "Oncogenic pathways in hereditary and sporadic breast cancer." Maturitas 49(1): 34-43.

- Key, T. J., N. E. Allen, et al. (2003). "Nutrition and breast cancer." Breast 12(6): 412-6.

- Key, T. J., P. K. Verkasalo, et al. (2001). "Epidemiology of breast cancer." Lancet Oncol 2(3): 133-40.

- Kim, W. J., H. Kim, et al. (2002). "GSTT1-null genotype is a protective factor against bladder cancer." Urology 60(5): 913-8.

- Korenaga, Y., H. Matsuyama, et al. (2005). "Smoking may cause genetic alterations at 5q22.2 approximately q23.1 in clear-cell renal cell carcinoma." Cancer Genet Cytogenet 163(1): 7-11.

- Krupa, R., A. Sobczuk, et al. (2011) "DNA damage and repair in endometrial cancer in correlation with the hOGG1 and RAD51 genes polymorphism." Mol Biol Rep 38(2): 1163-70.

- Kushman, M. E., S. L. Kabler, et al. (2007). "Protective efficacy of hGSTM1-1 against B[a]P and (+)- or (-)-B[a]P-7,8-dihydrodiol cytotoxicity, mutagenicity, and macromolecular adducts in V79 cells coexpressing hCYP1A1." Toxicol Sci 99(1): 51-7.

- LaFramboise, T., B. A. Weir, et al. (2005). "Allele-specific amplification in cancer revealed by SNP array analysis." PLoS Comput Biol 1(6): e65.

- Lakshmanan, I., M. P. Ponnusamy, et al.(2012) "MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells." Oncogene 31(7): 805-17.

- Landi, S. (2000). "Mammalian class theta GST and differential susceptibility to carcinogens: a review." Mutat Res 463(3): 247-83.

- Langerod, A., I. R. Bukholm, et al. (2002). "The TP53 codon 72 polymorphism may affect the function of TP53 mutations in breast carcinomas but not in colorectal carcinomas." Cancer Epidemiol Biomarkers Prev 11(12): 1684-8.

- Lash, T. L., B. D. Bradbury, et al. (2005). "A case-only analysis of the interaction between N-acetyltransferase 2 haplotypes and tobacco smoke in breast cancer etiology." Breast Cancer Res 7(3): R385-93.

- Lazer, G., Y. Idelchuk, et al. (2009). "The haematopoietic specific signal transducer Vav1 is aberrantly expressed in lung cancer and plays a role in tumourigenesis." J Pathol 219(1): 25-34.

- Lee, C. H., R. H. Lin, et al. (1996). "Mutual interactions among ingredients of betel quid in inducing genotoxicity on Chinese hamster ovary cells." Mutat Res 367(2): 99-104.

- Lee, S. Y., Y. Chung, et al. (2007). "Log-linear model-based multifactor dimensionality reduction method to detect gene gene interactions." Bioinformatics 23(19): 2589-95.

- Lemon, S. C., J. Roy, et al. (2003). "Classification and regression tree analysis in public health: methodological review and comparison with logistic regression." Ann Behav Med 26(3): 172-81.

- Lemos, M. C., E. Coutinho, et al. (2008). "Combined GSTM1 and GSTT1 null genotypes are associated with a lower risk of papillary thyroid cancer." J Endocrinol Invest 31(6): 542-5.

- Levy-Lahad, E., A. Lahad, et al. (2001). "A single nucleotide polymorphism in the RAD51 gene modifies cancer risk in BRCA2 but not BRCA1 carriers." Proc Natl Acad Sci U S A 98(6): 3232-6.

- Li, C. and W. Hung Wong (2001). "Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application." Genome Biol 2(8): RESEARCH0032.

- Liao, D. J. and R. B. Dickson (2000). "c-Myc in breast cancer." Endocr Relat Cancer 7(3): 143-64.

- Lin, M., C. D. Morrison, et al. (2009). "Copy number gain and oncogenic activity of YWHAZ/14-3-3zeta in head and neck squamous cell carcinoma." Int J Cancer 125(3): 603-11.

- Liu, G., D. P. Miller, et al. (2001). "Differential association of the codon 72 p53 and GSTM1 polymorphisms on histological subtype of non-small cell lung carcinoma." Cancer Res 61(24): 8718-22.

- Liu, J., G. Wen, et al. (2009). "Aldo-keto reductase family 1 member B1 inhibitors: old drugs with new perspectives." Recent Pat Anticancer Drug Discov 4(3): 246-53.

- Liu, S. Y., Y. C. Liu, et al. (2007). "Up-regulation of matrix metalloproteinase-8 by betel quid extract and arecoline and its role in 2D motility." Oral Oncol 43(10): 1026-33.

- Loo, L. W., D. I. Grove, et al. (2004). "Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes." Cancer Res 64(23): 8541-9.

- Lose, F., P. Lovelock, et al. (2006). "Variation in the RAD51 gene and familial breast cancer." Breast Cancer Res 8(3): R26.

- Lou, X. Y., G. B. Chen, et al. (2007). "A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence." Am J Hum Genet 80(6): 1125-37.

- Lu, H., X. Guo, et al. (2005). "The BRCA2-interacting protein BCCIP functions in RAD51 and BRCA2 focus formation and homologous recombinational repair." Mol Cell Biol 25(5): 1949-57.

- Luo, L. Z., K. M. Werner, et al. (2004). "Cigarette smoke induces anaphase bridges and genomic imbalances in normal cells." Mutat Res 554(1-2): 375-85.

- Luo, M. and J. L. Guan (2010) "Focal adhesion kinase: a prominent determinant in breast cancer initiation, progression and metastasis." Cancer Lett 289(2): 127-39.

- Lynch, B. M., H. K. Neilson, et al.(2011) "Physical activity and breast cancer prevention." Recent Results Cancer Res 186: 13-42.

- Magnusson, C., S. Wedren, et al. (2007). "Cigarette smoking and breast cancer risk: a population-based study in Sweden." Br J Cancer 97(9): 1287-90.

- Malhotra, G. K., X. Zhao, et al.(2010) "Histological, molecular and functional subtypes of breast cancers." Cancer Biol Ther 10(10): 955-60.

- Martin, A. M. and B. L. Weber (2000). "Genetic and hormonal risk factors in breast cancer." J Natl Cancer Inst 92(14): 1126-35.

- Martin, M. B., R. Reiter, et al. (2007). "Effects of tobacco smoke condensate on estrogen receptor-alpha gene expression and activity." Endocrinology 148(10): 4676-86.

- Marx, G. (1997). "Possible function found for breast cancer genes." Science 276(5312): 531-2.

- Matthias, C., K. Branigan, et al. (1998). "Polymorphism within the cyclin D1 gene is associated with prognosis in patients with squamous cell carcinoma of the head and neck." Clin Cancer Res 4(10): 2411-8.

- McPherson, K., C. M. Steel, et al. (2000). "ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics." Bmj 321(7261): 624-8.

- Moore, L. E., A. H. Smith, et al. (2002). "Arsenic-related chromosomal alterations in bladder cancer." J Natl Cancer Inst 94(22): 1688-96.

- Mudur, G. (2005). India has some of the highest cancer rates in the world. BMJ. 330: 215.

- Muenphon, K., T. Limpaiboon, et al. (2006). "Amplification of chromosome 21q22.3 harboring trefoil factor family genes in liver fluke related cholangiocarcinoma is associated with poor prognosis." World J Gastroenterol 12(26): 4143-8.

- Nair, J., H. Ohshima, et al. (1985). "Tobacco-specific and betel nut-specific N-nitroso compounds: occurrence in saliva and urine of betel quid chewers and formation in vitro by nitrosation of betel quid." Carcinogenesis 6(2): 295-303.

- Nair, U., H. Bartsch, et al. (2004). "Alert for an epidemic of oral cancer due to use of the betel quid substitutes gutkha and pan masala: a review of agents and causative mechanisms." Mutagenesis 19(4): 251-62.

- Nakao, K., K. R. Mehta, et al. (2004). "High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization." Carcinogenesis 25(8): 1345-57.

- Narayan, S., A. S. Jaiswal, et al. (2004). "Cigarette smoke condensate-induced transformation of normal human breast epithelial cells in vitro." Oncogene 23(35): 5880-9.

- NCRPINDIA (2006) website, Available: http://ncrpindia.org/Three-Year Reports of Population Based Cancer Registries 2006–2008. Accessed: 31 January 2012.

- Neal, C. L. and D. Yu. (2010) "14-3-3zeta as a prognostic marker and therapeutic target for cancer." Expert Opin Ther Targets 14(12): 1343-54.

- Nowak, N. J., D. Gaile, et al. (2005). "Genome-wide aberrations in pancreatic adenocarcinoma." Cancer Genet Cytogenet 161(1): 36-50.

- O'Donovan, P. J. and D. M. Livingston (2010) "BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair." Carcinogenesis 31(6): 961-7.

- Ouellet, V., K. Tiedemann, et al. (2011) "CCN3 impairs osteoblast and stimulates osteoclast differentiation to favor breast cancer metastasis to bone." Am J Pathol 178(5): 2377-88.

- Pang, H., M. Hauser, et al.(2011) "Pathway-based identification of SNPs predictive of survival." Eur J Hum Genet 19(6): 704-9.

- Parkin, D. M.(2011) "15. Cancers attributable to reproductive factors in the UK in 2010." Br J Cancer 105 Suppl 2: S73-6.

- Pharoah, P. D., A. M. Dunning, et al. (2004). "Association studies for finding cancer-susceptibility genetic variants." Nat Rev Cancer 4(11): 850-60.

- Pharoah, P. D., J. Tyrer, et al. (2007). "Association between common variation in 120 candidate genes and breast cancer risk." PLoS Genet 3(3): e42.

- Phukan, R. K., E. Zomawia, et al. (2005). "Tobacco use and stomach cancer in Mizoram, India." Cancer Epidemiol Biomarkers Prev 14(8): 1892-6.

- Phukan, R. K., M. S. Ali, et al. (2001). "Betel nut and tobacco chewing; potential risk factors of cancer of oesophagus in Assam, India." Br J Cancer 85(5): 661-7.

- Pollack, J. R., T. Sorlie, et al. (2002). "Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors." Proc Natl Acad Sci U S A 99(20): 12963-8.

- Purdie, K. J., S. R. Lambert, et al. (2007). "Allelic imbalances and microdeletions affecting the PTPRD gene in cutaneous squamous cell carcinomas detected using single nucleotide polymorphism microarray analysis." Genes Chromosomes Cancer 46(7): 661-9.

- Qiu, Y. L., W. Wang, et al. (2008). "Genetic polymorphisms, messenger RNA expression of p53, p21, and CCND1, and possible links with chromosomal aberrations in Chinese vinyl chloride-exposed workers." Cancer Epidemiol Biomarkers Prev 17(10): 2578-84.

- Rani, M., S. Bonu, et al. (2003). "Tobacco use in India: prevalence and predictors of smoking and chewing in a national cross sectional household survey." Tob Control 12(4): e4.

- Reinartz, S., S. Failer, et al.(2012) "CA125 (MUC16) gene silencing suppresses growth properties of ovarian and breast cancer cells." Eur J Cancer 48(10): 1558-69.

- Reynolds, P., S. Hurley, et al. (2004). "Active smoking, household passive smoking, and breast cancer: evidence from the California Teachers Study." J Natl Cancer Inst 96(1): 29-37.

- Ritchie, M. D., L. W. Hahn, et al. (2001). "Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer." Am J Hum Genet 69(1): 138-47.

- Rivenson, A., D. Hoffmann, et al. (1988). "Induction of lung and exocrine pancreas tumors in F344 rats by tobacco-specific and Areca-derived N-nitrosamines." Cancer Res 48(23): 6912-7.

- Saama, P. M., O. V. Patel, et al. (2006). "Novel algorithm for transcriptome analysis." Physiol Genomics 28(1): 62-6.

- Sakamuro, D. and G. C. Prendergast (1999). "New Myc-interacting proteins: a second Myc network emerges." Oncogene 18(19): 2942-54.

- Samson, M., R. Swaminathan, et al. (2007). "Role of GSTM1 (Null/Present), GSTP1 (Ile105Val) and P53 (Arg72Pro) genetic polymorphisms and the risk of breast cancer: a case control study from South India." Asian Pac J Cancer Prev 8(2): 253-7.

- Sangrajrang, S., Y. Sato, et al. (2009). "Genetic polymorphisms of estrogen metabolizing enzyme and breast cancer risk in Thai women." Int J Cancer 125(4): 837-43.

- Saxena, A., V. S. Dhillon, et al. (2009). "Detection and relevance of germline genetic polymorphisms in glutathione S-transferases (GSTs) in breast cancer patients from northern Indian population." Breast Cancer Res Treat 115(3): 537-43.

- Saxena, S., A. Chakraborty, et al. (2006). "Contribution of germline BRCA1 and BRCA2 sequence alterations to breast cancer in Northern India." BMC Med Genet 7: 75.

- Saxena, S., C. I. Szabo, et al. (2002). "BRCA1 and BRCA2 in Indian breast cancer patients." Hum Mutat 20(6): 473-4.

- Schembri, F., S. Sridhar, et al. (2009). "MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium." Proc Natl Acad Sci U S A 106(7): 2319-24.

- Scotto, L., G. Narayan, et al. (2008). "Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression." Genes Chromosomes Cancer 47(9): 755-65.

- Seitz, H. K., C. Pelucchi, et al. (2012) "Epidemiology and pathophysiology of alcohol and breast cancer: Update 2012." Alcohol Alcohol 47(3): 204-12.

- Sen, S.,Talukder, G., Sharma, A. (1987) Potentiation of betel-induced alterations of mouse glandular stomach mucosa by tobacco in studies simulating betel addiction Pharmaceutical Biology, 25 (4). pp. 209-215.

- Shadeo, A. and W. L. Lam (2006). "Comprehensive copy number profiles of breast cancer cell model genomes." Breast Cancer Res 8(1): R9.

- Shen, M., Q. Lan, et al. (2006). "Polymorphisms in genes involved in DNA double-strand break repair pathway and susceptibility to benzene-induced hematotoxicity." Carcinogenesis 27(10): 2083-9.

- Shetty, P. (2012) "India faces growing breast cancer epidemic." Lancet 379(9820): 992-3.

- Shimada, T., H. Yamazaki, et al. (1996). "Activation and inactivation of carcinogenic dihaloalkanes and other compounds by glutathione S-transferase 5-5 in Salmonella typhimurium tester strain NM5004." Chem Res Toxicol 9(1): 333-40.

- Siddique, M. and K. Sabapathy (2006). "Trp53-dependent DNA-repair is affected by the codon 72 polymorphism." Oncogene 25(25): 3489-500.

- Singh, B., V. B. Wreesmann, et al. (2002). "Chromosomal aberrations in patients with head and neck squamous cell carcinoma do not vary based on severity of tobacco/alcohol exposure." BMC Genet 3: 22.

- Smith, T. R., E. A. Levine, et al. (2003). "DNA-repair genetic polymorphisms and breast cancer risk." Cancer Epidemiol Biomarkers Prev 12(11 Pt 1): 1200-4.

- Smith, T. R., E. A. Levine, et al. (2008). "Polygenic model of DNA repair genetic polymorphisms in human breast cancer risk." Carcinogenesis 29(11): 2132-8.

- Spector, D., M. Mishel, et al. (2009). "Breast cancer risk perception and lifestyle behaviors among White and Black women with a family history of the disease." Cancer Nurs 32(4): 299-308.

- Srivastava, A., K. L. Sharma, et al.(2012) "Significant role of estrogen and progesterone receptor sequence variants in gallbladder cancer predisposition: a multi-analytical strategy." PLoS One 7(7): e40162.

- Taioli, E. (2008). "Gene-environment interaction in tobacco-related cancers." Carcinogenesis 29(8): 1467-74.

- Tan, S. M., A. J. Evans, et al. (2007). "How relevant is breast cancer screening in the Asia/Pacific region?" Breast 16(2): 113-9.

- Terry, P. D. and T. E. Rohan (2002). "Cigarette smoking and the risk of breast cancer in women: a review of the literature." Cancer Epidemiol Biomarkers Prev 11(10 Pt 1): 953-71.

- Tiemersma, E. W., R. E. Omer, et al. (2001). "Role of genetic polymorphism of glutathione-S-transferase T1 and microsomal epoxide hydrolase in aflatoxin-associated hepatocellular carcinoma." Cancer Epidemiol Biomarkers Prev 10(7): 785-91.

- Tsai, J. F., L. Y. Chuang, et al. (2001). "Betel quid chewing, cigarette smoking and alcohol consumption related to oral cancer in Taiwan." J Oral Pathol Med 24(10): 450-3.

- Turner, N., M. B. Lambros, et al.(2010) "Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets." Oncogene 29(14): 2013-23.

- Tzeng, H. E., J. C. Chen, et al.(2011) "CCN3 increases cell motility and MMP-13 expression in human chondrosarcoma through integrin-dependent pathway." J Cell Physiol 226(12): 3181-9.

- Uchihara, T., C. Okubo, et al. (2007). "Neuronatin expression and its clinicopathological significance in pulmonary non-small cell carcinoma." J Thorac Oncol 2(9): 796-801.

- Vaarala, M. H., H. Mattila, et al. (2008). "The interaction of CYP3A5 polymorphisms along the androgen metabolism pathway in prostate cancer." Int J Cancer 122(11): 2511-6.

- Valarmathi, M. T., M. Sawhney, et al. (2004). "Novel germline mutations in the BRCA1 and BRCA2 genes in Indian breast and breast-ovarian cancer families." Hum Mutat 23(2): 205.

- van Bladeren, P. J. (2000). "Glutathione conjugation as a bioactivation reaction." Chem Biol Interact 129(1-2): 61-76.

- Vannini, I., W. Zoli, et al. (2008). "Role of p53 codon 72 arginine allele in cell survival in vitro and in the clinical outcome of patients with advanced breast cancer." Tumour Biol 29(3): 145-51.

- Vasudev, N. S., I. Trigonis, et al.(2011) "The prognostic and predictive value of CA-125 regression during neoadjuvant chemotherapy for advanced ovarian or primary peritoneal carcinoma." Arch Gynecol Obstet 284(1): 221-7.

- Virnig, B. A., T. Shamliyan, et al. (2009). "Diagnosis and management of ductal carcinoma in situ (DCIS)." Evid Rep Technol Assess (Full Rep)(185): 1-549.

- Werbrouck, J., K. De Ruyck, et al. (2008). "Single-nucleotide polymorphisms in DNA double-strand break repair genes: association with head and neck cancer and interaction with tobacco use and alcohol consumption." Mutat Res 656(1-2): 74-81.

- Wikman, H., S. Ruosaari, et al. (2007). "Gene expression and copy number profiling suggests the importance of allelic imbalance in 19p in asbestos-associated lung cancer." Oncogene 26(32): 4730-7.

- Williams, J. A. and D. H. Phillips (2000). "Mammary expression of xenobiotic metabolizing enzymes and their potential role in breast cancer." Cancer Res 60(17): 4667-77.

- Wiseman, H. and B. Halliwell (1996). "Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer." Biochem J 313 (Pt 1): 17-29.

- Wong, K. K., Y. T. Tsang, et al. (2004). "Allelic imbalance analysis by high-density single-nucleotide polymorphic allele (SNP) array with whole genome amplified DNA." Nucleic Acids Res 32(9): e69.

- Wu, M. T., D. C. Wu, et al. (2004). "Constituents of areca chewing related to esophageal cancer risk in Taiwanese men." Dis Esophagus 17(3): 257-9.

- Xiong, P., M. L. Bondy, et al. (2001). "Sensitivity to benzo(a)pyrene diol-epoxide associated with risk of breast cancer in young women and modulation by glutathione S-transferase polymorphisms: a case-control study." Cancer Res 61(23): 8465-9.

- Xu, D., J. Dwyer, et al. (2008). "Ets2 maintains hTERT gene expression and breast cancer cell proliferation by interacting with c-Myc." J Biol Chem 283(35): 23567-80.

- Xue, F., W. C. Willett, et al. (2011) "Cigarette smoking and the incidence of breast cancer." Arch Intern Med 171(2): 125-33.

- Yao, J., S. Weremowicz, et al. (2006). "Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression." Cancer Res 66(8): 4065-78.

- Yarnold, J. (2009). "Early and locally advanced breast cancer: diagnosis and treatment National Institute for Health and Clinical Excellence guideline 2009." Clin Oncol (R Coll Radiol) 21(3): 159-60.

- Ye, Z. and J. M. Parry (2002). "The CYP17 MspA1 polymorphism and breast cancer risk: a meta-analysis." Mutagenesis 17(2): 119-26.

- Yeh, P. Y., Y. S. Lu, et al.(2011) "IkappaB kinases increase Myc protein stability and enhance progression of breast cancer cells." Mol Cancer 10: 53.

- Zabuawala, T., D. A. Taffany, et al.(2010) "An ets2-driven transcriptional program in tumor-associated macrophages promotes tumor metastasis." Cancer Res 70(4): 1323-33.

- Zhai, R., F. Chen, et al.(2010) "Interactions among genetic variants in apoptosis pathway genes, reflux symptoms, body mass index, and smoking indicate two distinct etiologic patterns of esophageal adenocarcinoma." J Clin Oncol 28(14): 2445-51.

- Zhang, Z., M. Wang, et al. (20101)"P53 codon 72 polymorphism contributes to breast cancer risk: a meta-analysis based on 39 case-control studies." Breast Cancer Res Treat 120(2): 509-17.

- Zhao, M., R. Lewis, et al. (2001). "No apparent association of GSTP1 A(313)G polymorphism with breast cancer risk among postmenopausal Iowa women." Cancer Epidemiol Biomarkers Prev 10(12): 1301-2.

- Zhao, X., C. Li, et al. (2004). "An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays." Cancer Res 64(9): 3060-71.

- Zhou, Y., S. Eppenberger-Castori, et al. (2005). "The NFkappaB pathway and endocrine-resistant breast cancer." Endocr Relat Cancer 12 Suppl 1: S37-46.

- Zhu, H., Lei, X. et al. (2012) "l contraceptive use and risk of breast cancer: A meta-analysis of prospective cohort studies." Eur J Contracept Reprod Health Care. (6):402-14

- Zienolddiny, S., D. Campa, et al. (2006). "Polymorphisms of DNA repair genes and risk of non-small cell lung cancer." Carcinogenesis 27(3): 560-7.

Appendix

Appendix

Preparation of reagents

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.

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

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.

List of Publications

~List of Publications~

1. ***Mishi Kaushal,*** Ashwani K Mishra, Jagannath Dev Sharma, Eric Zomawia, Amal Kataki, Sujala Kapur, Sunita Saxena Genomic alterations in breast cancer patients in betel quid and non betel quid chewers PloS ONE Accepted Aug 2012 Impact factor 4.1
2. ***Mishi Kaushal***, Ashwani K. Mishra, B.S. Raju, Rakhshan Ihsan, Anurupa Chakraborty, Jaganath Sharma, Eric Zomawia, Yogesh Verma, Amal Kataki, Sujala Kapur, Sunita Saxena Betel quid chewing as an environmental risk factor for breast cancer Mutation Research 703 (2010) 143–148 Impact factor 3.035 Citations 1
3. ***M. Kaushal***, I. Chattopadhyay, R. Phukan, J. Purkayastha, J. Mahanta, S. Kapur, S. Saxena Contribution of germ line *BRCA2* sequence alterations to risk of familial esophageal cancer in a high-risk area of India Diseases of the Esophagus (2010) **23**, 71–75 Impact factor 1.81 Citations 1
4. Rakhshan Ihsan, Pradeep Singh Chauhan, Ashwani Kumar Mishra, Dhirendra Singh Yadav, ***Mishi Kaushal***, Jagannath Dev Sharma, Eric Zomawia, Yogesh Verma, Sujala Kapur, Sunita Saxena Multiple Analytical Approaches Reveal Distinct Gene-Environment Interactions in Smokers and Non Smokers in Lung Cancer PLoS ONE 6(12): e29431 Impact factor 4.41 Citations 1
5. Regina Devi Thoudam, Dhirendra Singh Yadav, Ashwani Kumar Mishra, ***Mishi Kaushal***, Rakhshan Ihsan,Indranil Chattopadhyay, Pradeep Singh Chauhan, Jagannath Sarma, Eric Zomawia, Yogesh Verma,A. Nandkumar, Jagadish Mahanta, Rupkumar Phukan, Sujala Kapur, and Sunita Saxena Distribution of Glutathione S-Transferase T1 and M1 Genes Polymorphisms in North East Indians: A Potential Report Genetic testing and molecular biomarkers Volume 14, Number 2, 2010 Pp. 163–169 ^{a} Mary Ann Liebert, Inc. Impact factor 1.11 Citations 3
6. Dhirendra Singh Yadav, Thoudam Regina Devi, Rakhshan Ihsan, Ashwani Kumar Mishra, ***Mishi Kaushal***, Pradeep Singh Chauhan, Sarangadhara A.R. Bagadi, Jagannath Sharma, Eric Zamoawia, Yogesh Verma, Ambakumar Nandkumar, Sunita Saxena, and Sujala Kapur Polymorphisms of Glutathione-S-Transferase Genes and the Risk of

Aerodigestive Tract Cancers in the Northeast Indian Population Genetic testing and molecular biomarkers Volume 14, Number 5, 2010 Pp. 715–723 [a] Mary Ann Liebert, Inc. Impact factor 1.11 Citations 5

7. Pradeep Singh Chauhan, Rakhshan Ihsan, Dhirendra Singh Yadav, Ashwani Kumar Mishra, Bharat Bhushan, Abha Soni, ***Mishi Kaushal***, Thoudam Regina Devi, Sumita Saluja, Dipendra Kumar Gupta, Vishakha Mittal, Sunita Saxena, Sujala Kapur Association of Glutathione *S*-Transferase, *EPHX*, and *p53* codon 72 Gene Polymorphisms with Adult Acute Myeloid Leukemia DNA and Cell Biology. January 2011, 30(1): 39-46. Impact factor 1.11 Citation 1
8. Sunita Saxena[*], Anurupa Chakraborty, ***Mishi Kaushal***, Sanjeev Kotwal, Dinesh Bhatanager, Ravindar S Mohil, Chintamani Chintamani, Anil K Aggarwal, Veena K Sharma, Prakash C Sharma, Gilbert Lenoir, David E Goldgar and Csilla I Szabo Contribution of germline *BRCA1* and *BRCA2* sequence alterations to breast cancer in Northern India *BMC Medical Genetics* 2006, **7**:75 Impact factor 2.3 Citations 32

~*Conferences*~

Sno.	Title	Year and particulars of Publication/Presentation
1.	Betel Quid Chewing A Risk Factor For Breast Cancer: Study Of Genomic Alterations	Thirteenth Human Genome Meeting (HUGO) 20012,
2.	Chromosomal copy-number alterations in breast cancer significantly affect gene modules involved in cancer progression.	AACR's New Horizons in Cancer Research conference 2012, Gurgaon
3.	Genomic alterations in breast cancer patients from using 10K SNP arrays	'Beyond the Genome: The true gene count, human evolution and disease genomics' at Harvard Medical School, Boston, USA on 11-13 October 2010
4.	Genome-wide analysis of DNA copy number variations in Indian breast cancer patients using high-density SNP arrays	20th Asia Pacific Cancer
5	Genome-wide analysis of genetic alterations in	Organisation for Oncology and Translational Research (OOTR) 6th

	breast cancer patients from using 10K SNP arrays	Annual Conference,
6.	"Significance of TP53 codon 72 polymorphism in lung and breast cancer showing different xenobiotic potential".	Thirteenth Human Genome Meeting (HUGO) 2008,
7.	"Assessment of Breast cancer risk: Genotype polymorphism in estrogen synthesizing and metabolizing genes and their contribution in breast cancer susceptibility.	"27th annual convention of indian association for cancer research", from Feb. 6 – 9 , 2008
8.	"Study of Interactions between Glutathione-S-Transferase Metabolic Enzymes and Smoking in Lung Cancer"	IACRCON-2008 and 27th Annual Convention of Indian Association For Cancer Research.
9.	Role of high and low penetrance genes in susceptibility to breast cancer in patients from	International Symposium on Cancer Biology November 14-16, 2007,

Biography of Candidate

B-7/5138, Vasant Kunj New Delhi 110070
Mobile No. 9717129911
Email: mishi.wasson@gmail.com

Mishi Wasson

CAREER OBJECTIVE

With my academic background, strengthened by practical experience, exposure to research and research motivation, I would make an original contribution in the field of Biochemistry and Medicine. In accordance with this goal, I would like to pursue a career in research pertaining to Life Science.

ACADEMIC PROFILE

Course: Ph.D. (Doctor of Philosophy), 2006-2012

University : Birla Institute of Technology and Science, Pilani, INDIA.

Thesis Title : Contribution of alterations in putative susceptibility genes and genomic imbalances in the occurrence of Breast Cancer

Specialization : Microarray, Biostatistics, Data Mining, Polymorphic Studies, DHPLC, Sequencing, Molecular Techniques

Course: M.Sc. (Master of Science) (Biochemistry), 2003-2005

University : Jamia Hamdard, New Delhi, INDIA.

Specialization : Biochemistry

Division : I

Course: B.Sc. (Bachelor of Science) (Zoology), 2000-2003

University : Delhi University, New Delhi, INDIA.

Specialization : **Zoology,** Botany, Chemistry

Division : I I

Course: Senior School Certificate Examination (XII), 2000

University : Central Board of Secondary Education, INDIA.

Subjects : English, Mathematics, Physics, Chemistry. Biology

Division : I

B-7/5138, Vasant Kunj New Delhi 110070
Mobile No. 9717129911
Email: mishi.wasson@gmail.com

Mishi Wasson

RESEARCH PROFILE

A. Articles published in National/International Journals: 8
B. National/International Conferences Attended: 9
C. Abstracts Published in National/International Conferences: 9
D. Workshops attended: 2

(List Enclosed)

The research articles have been published in the journals of various international organization of very high repute viz., **PLOS One, Mutation Research, Diseases of the Esophagus, Genetic testing and molecular biomarkers, DNA and Cell Biology, BMC Medical Genetics**

SUMMARY OF TECHNIQUES KNOWN

A. Polymorphism Studies
PCR,RFLP, DHPLC, Sequencing

B. Microarray
Microarray analysis, Copy number variation analysis on Affymetrix Platform

C. Statistical Analysis
SPSS, High order analysis using MDR, CART software, Microarray analysis using Dchip, Data mining using DAVID and Ingenuity Pathway

PERSONAL PROFILE

DOB	:	10th Nov' 1982
Nationality	:	Indian
Marital status	:	Married
Husband's Name	:	Satvik Wasson

PASSPORT DETAILS

Passport No.	:	J5273357
Date of Issue	:	30/12/2010
Place of Issue	:	New Delhi, INDIA

LANGUAGE(s) PROFICIENCY

Proficient in both written & spoken English & Hindi

SKILLS KNOWN

Statistical Software like SPSS, MDR, CART, DAVID, Dchip known
Computer literate: Editing and Typing Text
Proficient in MS Word, PowerPoint, Excel
Internet (Send & Receive E-mail, Browsing)

B-7/5138, Vasant Kunj New Delhi 110070
Mobile No. 9717129911
Email: mishi.wasson@gmail.com

Mishi Wasson

INTERETS

Related to Genetics and susceptibility

STRENGTHS

- Self-motivated and hard worker with a high level of integrity to meet deadlines.
- Enthusiastic and committed to do all types of work initiatives.
- Ability to adapt to fast changing work environment.
- Strong interpersonal, highly motivated and interviewing skills.

REFERENCES

Dr Sunita Saxena
Director, National Institute of pathology, (ICMR)
Safdarjung Hospital Campus
New Delhi 110029
Email: sunita_saxena@yahoo.com

Dr Usha Agarwal
Scientist E, National Institute of pathology, (ICMR)
Safdarjung Hospital Campus
New Delhi 110029
Email: uburra@gmail.com

Dr. Ashwani K Mishra
Assistant Professor Biostatistics
National Drug Dependence Treatment Centre (NDDTC)
All India Institute of Medical Sciences (AIIMS)
Ansari Nagar, New Delhi-110029
Mobile: +91-9810222807
emailid: ashwanikm@yahoo.com
ashwaniiop@gmail.com

DECLARATION

I hereby solemnly affirm that all details provided above are true to the best of my knowledge and belief and that all the time, I shall carry myself in a manner that lends dignity to the organization and worthy enough of the person.

DATE:

PLACE: New Delhi

B-7/5138, Vasant Kunj New Delhi 110070
Mobile No. 9717129911
Email: mishi.wasson@gmail.com

Mishi Wasson

LIST OF PUBLICATIONS AND RESEARCH PROFILE

A. Articles in International Journals

1. ***Mishi Kaushal,*** Ashwani K Mishra, Jagannath Dev Sharma, Eric Zomawia, Amal Kataki, Sujala Kapur, Sunita Saxena Genomic alterations in breast cancer patients in betel quid and non betel quid chewers PloS ONE Accepted Aug 2012 Impact factor 4.1

2. ***Mishi Kaushal***, Ashwani K. Mishra, B.S. Raju, Rakhshan Ihsan, Anurupa Chakraborty, Jaganath Sharma, Eric Zomawia, Yogesh Verma, Amal Kataki, Sujala Kapur, Sunita Saxena Betel quid chewing as an environmental risk factor for breast cancer Mutation Research 703 (2010) 143–148 Impact factor 3.035 Citations 1

3. ***M. Kaushal***, I. Chattopadhyay, R. Phukan, J. Purkayastha, J. Mahanta, S. Kapur, S. Saxena Contribution of germ line *BRCA2* sequence alterations to risk of familial esophageal cancer in a high-risk area of India Diseases of the Esophagus (2010) **23**, 71–75 Impact factor 1.81 Citations 1

4. Rakhshan Ihsan, Pradeep Singh Chauhan, Ashwani Kumar Mishra, Dhirendra Singh Yadav, ***Mishi Kaushal***, Jagannath Dev Sharma, Eric Zomawia, Yogesh Verma, Sujala Kapur, Sunita Saxena Multiple Analytical Approaches Reveal Distinct Gene-Environment Interactions in Smokers and Non Smokers in Lung Cancer PLoS ONE 6(12): e29431 Impact factor 4.41 Citations 1

5. Regina Devi Thoudam, Dhirendra Singh Yadav, Ashwani Kumar Mishra, ***Mishi Kaushal***, Rakhshan Ihsan,Indranil Chattopadhyay, Pradeep Singh Chauhan, Jagannath Sarma, Eric Zomawia, Yogesh Verma,A. Nandkumar, Jagadish Mahanta, Rupkumar Phukan, Sujala Kapur, and Sunita Saxena Distribution of Glutathione S-Transferase T1 and M1 Genes Polymorphisms in North East Indians: A Potential Report Genetic testing and molecular biomarkers Volume 14, Number 2, 2010 Pp. 163–169 [a] Mary Ann Liebert, Inc. Impact factor 1.11 Citations 3

6. Dhirendra Singh Yadav, Thoudam Regina Devi, Rakhshan Ihsan, Ashwani Kumar Mishra, ***Mishi Kaushal***, Pradeep Singh Chauhan, Sarangadhara A.R. Bagadi, Jagannath Sharma, Eric Zamoawia, Yogesh Verma, Ambakumar Nandkumar, Sunita Saxena, and Sujala Kapur Polymorphisms of Glutathione-S-Transferase Genes and the Risk of Aerodigestive Tract Cancers in the Northeast Indian Population Genetic testing and molecular biomarkers Volume 14, Number 5, 2010 Pp. 715–723 [a] Mary Ann Liebert, Inc. Impact factor 1.11 Citations 5

B-7/5138, Vasant Kunj New Delhi 110070
Mobile No. 9717129911
Email: mishi.wasson@gmail.com

7. Pradeep Singh Chauhan, Rakhshan Ihsan, Dhirendra Singh Yadav, Ashwani Kumar Mishra, Bharat Bhushan, Abha Soni, ***Mishi Kaushal***, Thoudam Regina Devi, Sumita Saluja, Dipendra Kumar Gupta, Vishakha Mittal, Sunita Saxena, Sujala Kapur Association of Glutathione *S*-Transferase, *EPHX*, and *p53* codon 72 Gene Polymorphisms with Adult Acute Myeloid Leukemia DNA and Cell Biology. January 2011, 30(1): 39-46. Impact factor 1.11 Citation 1
8. Sunita Saxena*, Anurupa Chakraborty, ***Mishi Kaushal***, Sanjeev Kotwal, Dinesh Bhatanager, Ravindar S Mohil, Chintamani Chintamani, Anil K Aggarwal, Veena K Sharma, Prakash C Sharma, Gilbert Lenoir, David E Goldgar and Csilla I Szabo Contribution of germline *BRCA1* and *BRCA2* sequence alterations to breast cancer in Northern India *BMC Medical Genetics* 2006, **7**:75 Impact factor 2.3 Citations 32

B. Conferences Attended

Sno.	Title	Year and particulars of Publication/Presentation
1.	Betel Quid Chewing A Risk Factor For Breast Cancer: Study Of Genomic Alterations	Thirteenth Human Genome Meeting (HUGO) 20012,
2.	Chromosomal copy-number alterations in breast cancer significantly affect gene modules involved in cancer progression.	AACR's New Horizons in Cancer Research conference 2012, Gurgaon
3.	Genomic alterations in breast cancer patients from using 10K SNP arrays	'Beyond the Genome: The true gene count, human evolution and disease genomics' at Harvard Medical School, Boston, USA on 11-13 October 2010
4.	Genome-wide analysis of DNA copy number variations in Indian breast cancer patients using high-density SNP arrays	20th Asia Pacific Cancer
5	Genome-wide analysis of genetic alterations in breast cancer patients from using 10K SNP arrays	Organisation for Oncology and Translational Research (OOTR) 6th Annual Conference,
6.	"Significance of TP53 codon 72 polymorphism in lung and breast cancer showing different xenobiotic potential".	Thirteenth Human Genome Meeting (HUGO) 2008,

B-7/5138, Vasant Kunj New Delhi 110070
Mobile No. 9717129911
Email: mishi.wasson@gmail.com

Mishi Wasson

7.	"Assessment of Breast cancer risk: Genotype polymorphism in estrogen synthesizing and metabolizing genes and their contribution in breast cancer susceptibility.	"27th annual convention of indian association for cancer research", from Feb. 6 – 9 , 2008
8.	"Study of Interactions between Glutathione-S-Transferase Metabolic Enzymes and Smoking in Lung Cancer"	IACRCON-2008 and 27th Annual Convention of Indian Association For Cancer Research.
9.	Role of high and low penetrance genes in susceptibility to breast cancer in patients from	International Symposium on Cancer Biology November 14-16, 2007,

C. Workshops Attended

1. National worshop on Micrarray technology 16-18 April 2007 National Institute of pathology, New Delhi

2. Workshop on Genetic Epidemiological Methods for dissection of complex human traits 23-28 February 2009, TCG-ISI Center for population genomics , Kolkatta, India

Biography of the Supervisor

Curriculum Vitae

Name	:	**Dr. (Mrs.) Sunita Saxena**
Date of Birth	:	16th September, 1952
Designation	:	**Director**
Address	:	Institute of Pathology-ICMR Safdarjang Hospital Campus, Post Box No.4909, New Delhi – 110029
Academic Qualifications	:	**M.B.B.S., D.C.P., M.D.(Path)**

Degree	Institute	Year	Remarks
M.B.B.S.	M.L.N.Medical College, Allahabad	1974	
D.C.P. (Clinical Pathology	L.L.R.M.Medical College Meerut	1978	**Received Gold Medal & Merit Certificate**
M.D. (Pathology)	-do-	1981	

Details of Employment

Post	Duration	Institute
Research Officer	April 1981 to Dec.1985	Institute of Pathology, New Delhi.
Sr. Research Officer	Jan.1986 to Jan.1991	-do
Asst. Director	9th Jan.1991-2nd May, 1994	-do-
Deputy Director	2nd May, 1994 to 2nd May,'99	-do-
Deputy Director (Sr.Gr)	3rdMay, 99 to 8thApril, 2002	
Deputy Director (Sr.Gr) & Officer In charge	9thApril, 2002to 13th Dec., 2004	-do-
Director	14th Dec., 2004 till date	-do-

Areas of Specialization : **Molecular Oncology, Oncopathology**

Areas of Interest **Breast Tumors,**

Tobacco Associated cancers

Genito urinary cancers

Membership of National and International bodies

International: - Life Member International Union against Cancer (U.I.C.C.)

National: Life Member: Indian Association of Cancer Research (IACR).

Life Member: Indian Association of Pathologists and Microbiologists (IAPM).

Life Member: Association for the promotion of DNA fingerprinting and other DNA technologies (ADNAT)

Life Member: Proteomic Society of India

Life member: Human Genomic Organization (HUGO)

Trainings Received:-

1. Trained for **'Culture of fastidious cells and modern techniques of cell manipulation'** at *National Facility for animal Tissue and cell culture, Pune* in Dec., 1991.
2. Trained in **'Genetic mutation detection techniques for BRCA 1 and BRCA2 genes in genomic DNA of Breast cancer patients'** at *Unit Genetic Epidemiology* at *International Agency for Research on Cancer, Lyon, France* in 1998.
3. Attended **Hands-on Training Course on Proteomics and DNA Micro arrays** held from 25th February to 10th March 2003 at CCMB, Hyderabad.

Fellowships:-

1. Awarded **WHO fellowship** to work on **Renal Pathology** at **St. Vincent's Hospital, Melbourne, Australia**, 1987.
2. Awarded **Yamigawa Yoshida Fellowship of U.I.C.C.** to work on "*Mutation analysis of BRCA1 and BRCA2 genes in Breast Cancer Patients***"** at unit of Genetic Epidemiology **International Agency for Research on Cancer, Lyon, France** in 1998.

Awards and Honors Received:-

1. Awarded **Gold Medal** and **Merit Certificate** for securing highest marks in **Diploma in Clinical Pathology** (D.C.P.).

2. **K. C. Basu Mullick award** for best research work by **Indian Association of Pathologists and Microbiologists for year 2008.**

3. Received '***NOVARTIS ORATION AWARD 2006***" of Indian Council of Medical Research for her work on Breast cancer on 18th Sept. 2009.

4. Elected ***Fellow of National Academy of Medical Sciences in 2010.***

5. Elected ***Fellow of Indian College of Pathologist in 2010***

6. Paper entitled "Flow cytometric analyses of Th1 and Th2 cytokine production as a parameter of immunologic dysfunction in patients with superficial transitial cell Carcinoma" **received special appreciation award for the Best Poster presentation at 24th Annual Convention of Indian Association for Cancer Research & International Symposium of Human Papilloma virus and cervical cancer held at ICPO from 9th – 12th Feb., 2005.**

7. **Best Poster Award for poster entitled** "What Androgen Receptor CAG repeats polymorphism and p53 mutations/polymorphisms have to do with prostate cancer risk and progression? at "XXXIII Annual Conference of the ISHG and International Symposium on Genetics Revisited: the Genomics and Proteomics Advantage", Vishakhapatnam, India from 11-13 Feb 2008.

Research Grants received:-

1. Awarded research grant to work on **"Genetics of Breast Cancer in Indian Women"** by *Indo-French Centre for Promotion of Advanced Research*, **New Delhi (1999-2002).**
2. Awarded Research grant from *Deptt. of Science and Technology* for proposal entitled **" A new approach to the management of superficial bladder cancer. Role of in vitro Cytotoxicity assessment and immunologic enhancement"** (2000-2003).
3. **"Comprehensive study of carcinoma esophagus at North-East India-Multidiscipline approach"**. Collaborative, Multicentric ICMR Task Force project (2004-2007).
4. **"Microsatellite instability in androgen receptor gene and mismatch repair system in Prostate cancer in Indian males"** – Department of Science Technology (2005-2008).
5. **"Role of tobacco use in causation of cancer in north-east India"** – Indian Council of Medical Research Task Force project (2005-2008).

6. **"Effects of pesticide exposure in causation of cancer in north-east India"** – Indian Council of Medical Research Task Force project (2005-2008).
7. **"Establishment of Cell lines from Primary Breast Cancer" –** Indian Council of Medical Research. Task force project (2007-2010).
8. **"Study on Gene Expression and Hypermethylation Profiles in Early Onset Breast Cancer"** Department of Biotechnology (2008-2011)
9. "**Characterization of host immune factors associated with progression of superficial TCC of bladder by microarray analysis"** Indian Council of Medical Research (2009-2012)
10. "**Immunogenetic profile of Nasopharyngeal Cancer in a high prevalence region of Northeast India"** Department of Biotechnology (2010-2013)
11. **"Comparative study of Genetic, Clinical and Epidemiological Factors of Breast Cancer in Rural and Urban Area of India"** Indian Council of Medical Research Task force project (2009-2012).
12. **"Epigenetic studies in esophageal cancer in high risk region of Northeast India"** Department of Biotechnology, Twining Program for NER (2011-2013)
13. "**Genome wide Analysis of Genetic alterations in patients with Esophageal Cancer from Northeast India using Single Nucleotide Polymorphism arrays"** Indian Council of Medical Research (2011-2013)
14. **"Study on miRNA signatures associated with Breast cancer stem like cells (CSC) and their role in drug response"** Indian Council of Medical Research (2012-2015)
15. **"Targeted sequencing of Breast cancer specific genes in early onset breast carcinoma."** Indian Council of Medical Research (2012-2015)

Academic Experience:-

a. Diplomat of National Board Examinations (N.B.E.) – Institute is accredited for running this program since 1993.
    - ***Core member, Supervisor & examiner***.
    - Guided 20 DNB dissertations as supervisor and co supervisor.
    - Nominated as Inspector, Examiner for theory & Practical examinations and paper setter.

b. Ph.D. - ***Supervisor/Mentor***

- ***Supervisor*** of student registered under GGSIPS University, Delhi, BITS, Pilani.
- External Examiner for Ph.D. candidates of AIIMS, PGI, Chandigadh, Agra University, BITS, Pilani.
- 4 students completed Ph.D and 6 are registered.

c. Member of project Review Committee of Divisions of ***Non Communicable Diseases of Indian Council of Medical Research.***

d. Member of project Review Committee of Divisions of ***Basic Medical Sciences of Indian Council of Medical Research.***

e. Member of ***Scientific Advisory committee of Institute Of Cytology and Preventive Oncology, Noida***

f. Reviewer of papers of ***Indian Journal of Medical Research, Human Mutation, Cancer Immunology Immunotherapy, World Journal of Surgical Oncology***, **BMC Cancer,** ***Cancer Detection and Prevention, British Journal of Urology International, Indian Journal of Medical Research***

g. Member of ***Scientific Advisory committee of National Jalma Institute for Leprosy and other Mycobacterial diseases, Agra.***

h. Member of ***Scientific Advisory committee of Regional Medical Research Centre, Dibrugadh***

i. Member of ***Scientific Advisory Group of Publication & Information Division Of Indian Council Of Medical Research***

j. ***Chairperson of Institutional Ethical Committee of Safdarjang Hospital, New Delhi.***

k. Member of ***Technical committee of Indian Council of Medical Research***

l. Member of **DBT sponsored DSMB on Curcumin Trial in Cancer Cervix**

m. **Senate member of BITS, Pilani.**

n. Nominated as the expert member of **Task force on Leprosy at ICMR.**

o. Nominated as the member of Scientific Advisory Group for creating a new centre for Environmental Health and Bhopal Gas Tragedy at Bhopal.

p. Nominated as the member of the Data Safety Monitoring Board (DSMB) on "BASANT Clinical Trial" of DBT, New Delhi.

q. Appointed as **Appraiser and Inspector** by National Board of Examination for assessment of DNB students and institutes.

q. Dr. Sunita Saxena has been nominated as the expert member of " ICMR-ICAR Joint Task force on the Epidemiology of Human and Animal Brucellosis".

r. Dr. Sunita Saxena has been nominated as nodal officer for getting ICMR university status.

International Conferences attended.

Presented a paper on **"Pattern of lymphokines in minimal change Nephrotic syndrome"** in 5th *Asia Pacific Congress of Nephrology* held in New Delhi during 9-12th Dec., 1992.

1. Presented paper on **'Role of Proto-oncogene, Growth Factor Receptor and Steroid Hormones on Malignant Human Mammary Epithelial Cancer Cells in vitro and vivo"** in XVI, *International Cancer Congress (U.I.C.C.)* at New Delhi, 30th-5th Nov., 1994.

2. **'Stage A carcinoma of Prostate'** paper presented at *first conference of Nephrology, Urology and Transplantation Society of SAARC Countries* held at A.I.I.M.S. , New Delhi during 24th-26th March, 1995.

3. **'Mutation profile of BRCA 1 / 2 genes in Indian patients'** paper presented at *XV Asia Pacific Cancer Congress* held at Chennai during Dec. 12-15, 1999.

4. **'Mutation profile of BRCA 1 / 2 mutations in worldwide population. The MAGIC project'** paper presented in meeting of *American Association of Cancer Research AACR*, 2001, LA, USA.

5. Attended the 7th International Symposium on Molecular Basis of Predictive Oncology and Intervention Strategies' and presented a paper **BRCA1 and BRCA2 Genes in Indian Breast Cancer Patients** held at Nice, France from 7th 10th Feb. 2004.

6. Attended the **"UICC World Cancer Congress and Centre for Disease Control and Prevention (CDC)"** held during 8th to 13th July, 2006 at Washington DC, U.S.A and presented paper **"Study of candidate genes associated with Breast Cancer Susceptibility in the Indian Women".**

7. Attended the NCRI Cancer Conference held at International Convention Centre in Birmingham, UK from 30th September - 3rd October 2007 and presented paper (oral and poster) entitled **"Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis"**

8. Attended World Cancer Congress-2008 held in Shanghai, China during 12th-17th June 2008 and presented paper **"Differential gene expression in familial and tobacco associated esophageal cancers in north-east region of India".**

9. Attended HUGO's 13th Human Genome Meeting, September 27 - 30 2008, Hyderabad. and presented papers-
    - **Prognostic value of TP53 Codon 72 polymorphism in oral cancer and stomach cancer in high risk region of India**
    - **Differential gene expression profile of stomach and oral cancer in high risk region of India.**
    - **Differential expression of MAPK and GPCR pathway in esophageal cancer of North-east region of India**
    - **Significance of TP53 codon 72 polymorphism in breast and lung cancer showing different xenobiotic potential spectrum**
10. Attended **First Symposium on HPV Vaccination in the Asia Pacific and Middle East Region held at Seoul, Korea** during1st to 3rd June'09.
11. Visited **University of Minnesota, USA as member of expert team of Indian Scientists on Cancer** for collaborative research projects.
12. Presented papers in **Conference of the Organisation for Oncology and Translational Research (OOTR), 6th Annual Conference on 26 and 27 February, 2010 at Kyoto Japan**
    - Genetic alterations in patients with esophageal cancer from high-risk region in India by SNP array. *Sujala Kapur, Indranil Chattopadhyay, Rupkumar Phukan, Joydeep Purkayastha, Vikki Marshal, Amal Kataki, Jagdish Mohanta, David Bowtell, Sunita Saxena*
    - Genome-wide analysis of genetic alterations in breast cancer patients from Northeast India using 10K SNP arrays. *Sunita Saxena, Mishi Kaushal, Indranil Chatterjee, A. Bhatnagar, Chintamani, D. Bhatnagar, Sujala Kapur*

13. Presented papers in **20th Asia Specific Cancer Conference, November 12-14, 2009,Japan.**
    - GENOME-WIDE ANALYSIS OF GENETIC ALTERATIONS IN ESOPHAGEAL SQUAMOUS CELL CARCINOMA BY SNP ARRAY. Sujala Kapur, Indranil Chattopadhyay, Rupkumar Phukan, Joydeep Purkayastha, Vikki Marshal, Amal Kataki, Jagdish Mohanta, David Bowtell, Sunita Saxena.
    - GENOME-WIDE ANALYSIS OF DNA COPY NUMBER VARIATIONS IN INDIAN BREAST CANCER PATIENTS USING HIGH-DENSITY SNP ARRAYS. Sunita Saxena, Mishi Kaushal Wasson, Indranil Chatterjee, A Bhatnagar, Dr Chintamani, D Bhatnagar, Sujala Kapur.

14. Presented paper entitled "**Molecular Profile of Esophageal Cancer in High Risk Region of India"** in 10th World Congress of OESO held at Boston, USA during 28-31 Aug'2010. (Abstract J. Clin. Gastroenterology. Vol: 45 (2), 2011.
15. Presented paper entitled "**Genomic alterations in breast cancer patients from Northeast India using 10K SNP arrays"** in BMC group conference **Beyond the**

Genome: The true gene count, human evolution and disease genomics at Harvard Medical School, Boston, USA during11th-13th Oct'2010

16. Presented paper entitled "**Betel Quid Chewing A Risk Factor For Breast Cancer: Study Of Genomic Alterations"** at 16th Human Genome Meeting 2012 held at Sydney, Australia during 11th-14th March, 2012

Invited Speaker

1. **"Morphological patterns of Childhood Nephrotic Syndrome. Ultrastructure and Immunohistologic study"** at meeting of Delhi State Chapter of Indian Association of Pathologists & Microbiologists held in Feb. 1985.

2. **"Patterns of Lymphokines in Minimal changes nephritic Syndrome"** at meeting of Delhi State chapter of IAPM held in July, 1991.

3. **"Study of Differentiating markers in Breast Cancer" and 'Pathobiology of Prostate tumors'** in the workshop held on **"An overview of tumor Biology"** at Institute of Pathology in May 1993 in collaboration with Tata Memorial Hospital and Cancer Research Institute, Bombay.

4. **"Pathobiology of Prostate Malignancies"** at IX annual conference of Delhi regional Chapter of IAPM, 1994 held at Ram Manohar Lohia Hospital, New Delhi.

5. **"Tumor Markers in Paediatric Malignancies – Clinical Application"** in National Seminar on Paediatric malignancies in Feb. 96 at Safdarjang Hosptial, New Delhi.

6. **"Breast Cancer – Diagnosis"** popular lecture for 85th Centenary Celebration of ICMR in May'96 at Institute of Pathology, New Delhi.

7. **"Determinants of cell behavior in Breast Cancer"** in Sept.'97 at Delhi Breat Group meeting at Batra Hospital, New Delhi.

8. **"Genetic Predisposition of Breast Cancer in Indian Women"** talk given at Institute of Cytology and Preventive Oncology in Jan.'2000.

9. **"Genetic Predisposition of Breast Cancer in Indian Women – Clinical Significance"** invited talk at meeting of Delhi Breast Group in Feb.'2000.

10. **"Bioinformatics : Opportunities and Challenges for New Millennium"** at ICMR-WHO workshop on use of informatics in Biomedical Research, 13-15th Dec., 2000, New Delhi.

11. **"Breast Cancer : Genetics, Risks and Strategies"** in CME Programme on Oncopathology at A.H. Regional Cancer Center, Cuttack, 17-18th March, 2001.

12. **"Genetics and Prognostic Markers in Colorectal Cancer"** in XVIII Annual Conference of Association of Surgeons of India (Delhi Chapter), 24th March 2001, New Delhi.

13. **"Breast cancer Genetics : Risk assessment to prognostic implication"** in XI UP Chapter of Indian Association of Pathologists and Microbiologists held at LLRM Medical College, Meerut on 6-7th Oct. 2001.

14. **"Determinants of cell behavior in Breast cancer"** at meeting of Delhi State Chapter of IAPM held on Ist Dec. 2001 at Safdarjang Hospital, New Delhi.

15. Delivered a plenary talk on **"Early Onset Breast Cancer in Indian women and Genetic Susceptibility and Molecular characterstics"** on **the 24th Annual Convention of Indian Association for Cancer Research (IACR)** & International Symposium on Human Paillomavirus and Cervical Cancer at ICPO, NOIDA from 9th to 12th February, 2005.

16. Attended Symposium on "Biological and clinical relevance of placenta" and delivered talk on "**Hormones & Receptors in Placenta – Role in Fetal development Trophoblastic neoplasms"** at Department of Anatomy, Vardhman Mahavir Medical College & Safdarjang Hospital, New Delhi on 15th February, 2006.

17. Delivered talk on "**Study of candidate Genes Associated with Breast Cancer Susceptibility in Indian Women"** at International Symposium on Preventive and Predictive Molecular Diagnostics held on Januarly 21-22, 2006 at Dhirubhai Ambani Life Sciences, Mumbai.

18. Invited as a guest speaker in CME entitled "Ancillary Techniques in Anatomic Pathology from digital imaging to confocal imaging and laser microdissection" at 56th Annual Conference of the Indian Association of Pathologists and Microbiologists [APCON 2007] held at PGIMER, Chandigarh from 26th -29th November 2007 and delivered a talk on **" Molecular Biology of Cancer and Laser Capture Microscopy"**

19. Delivered a talk on "**Breast Cancer risk factors in North-East Indian women"** in Breast Con-2008 held at Guwahati on 7th March 2008.

20. Chaired session on Non-communicable diseases during conference on **"Show casing Science by Indian Women Scientist"** held on 8th – 10th March, 2008 and delivered a talk on **"Genetic Profile of Breast Cancer in Indian women".**

21. Delivered a talk on **"Genome-wide approach to identify prognostic markers for Esophageal Cancer"** in seminar on 'Prognostic and Predictive Factors in Cancer Management' at Foundation Day celebration & Workshop on Brachytherapy,

Department of Radiotherapy at Convention Center, CSM Medical University, Lucknow on 14th December 2008.

22. Delivered a talk on **"Breast cancer in Indian women : Risk and prevention"** in 32nd Session of Indian – Social Science Congress (ISSC) held at Department of Biotechnology, Jamia Millia Islamia University, New Delhi on 18th December 2008.

23. Invited to deliver a talk on **"Understanding molecular biology of cancer using Genomic approaches"** at 63rd IAPM Kerala Chapter Meeing & 6th National CME in Pathology during 14-15th Feb. 2009 organised by Department of Pathology, Amrita Institute of Medical Sciences, Kochi.

24. Invited as **Chief Guest** to attend Conference on **"Emerging Trends in Life Sciences Research"** organized by BITS, Pilani and delivered Key-note address on **"Genome-wide approach to identify biomarkers for Esophageal cancer in North East India"** on 6th March 2009.
25. Delivered a talk on "**Molecular biology of Cancer by Genome-wide approaches"** at CME Pathology held on 18th Nov' 2009 at Maulana Azad Medical College, New Delhi. Invited as **Guest faculty** to give talk on **" Esophageal Cancer in North East region- Contribution of genetic vis-à-vis environmental factors "** at 29th Annual Convention of Indian Association of Cancer Research held at Amrita Institute of Medical Sciences, Cochin during 21st-23rd Feb.10.
26. Invited to give a talk on **"Understanding molecular biology of cancer using Genomic approaches"** at **National Symposium on Current Trends in Genomics and Proteomics** organized by Deshbandhu Gupta College, University of Delhi during 4th-5th Feb'10.
27. Invited as Guest faculty to give talk on**" Trends of Cancer in Indian Women : Can I prevent cancer"** for the International Conference on**" Empowering Women in Developing Countries through better health care and Nutrition"** held at BITS, Pilani during 22nd -24th April' 10. Dr Sunita Saxena was invited to deliver talk on **" Role of Electronmicroscopy in Renal Pathology"** at the International Conference entitled "**Renal Pathology for the Nephrologists**" organized by Department of Pathology, G.B. Pant Hospital, New Delhi during 22nd & 23rd January, 2011.
28. Dr. Saxena was invited as faculty member to1st Indo-USA initiative on "**Translational Cancer Prevention and Biomarkers workshop 2011**" held at Mazumdar-Shaw cancer Center,Bangaluru during 13th to 16th February, 2011 and
    1. Chaired the session on "**Prevention of Breast Cancer"**
    2. Acted as panel discussant in session on "**Genetic and Population Epidemiology**"and presented the talk "**Identification of Geo-ethnic variation in North-East India and their association with cancer risk**."
29. Invited to give talk on "**Applications of Cancer Genomics**" in the workshop "New Frontiers in Medicine" organized by Deptt. Of Pathology, M.L.N.Medical College, Allhabad on 10th Feb'2011.

30. Delivered talk on "**Applications of Cancer Genomics**" at workshop on "**Hands on Training in basic Molecular Biology Techniques**" at the Institute of Pathology, New Delhi during 1st to 4th March, 2011, as part of ICMR Centenary Celebrations.
31. Invited as Chief Guest at **National conference on "Current Trends in Advanced Biomedical Technology (CTA B-II)** organized by Department of Biosciences, Nehru Arts and Sciences College, Coimbatore, during 20th to 21st September, 2011 and delivered Key note address on "**Understanding molecular biology of cancer using Genomic approaches"**
32. Invited to deliver talk on "**Genome wise approach to identify Cancer Biomarker : Role of Microarray Technology**" in **1st Biennial Conference of IASN** organized by Agra Medical College in association of National JALMA Institute for Leprosy and Other Microbacterial Disease (ICMR), Agra held in Agra during 30th November – 1st December, 2011.
33. Invited to deliver talk on "**Application of Tissue Microarray in Pathology**"in "**Symposium on Antibody – based proteomics**" at annual meeting of **International Academy of Pathologists** held at Govt. Medical College Patiala during 1st-4th Dec. 2011
34. Invited to deliver talk on "**Applications of Tissue Microarray in Pathology and Research**" in "**Workshop on Applications of Molecular Biology in Cancer Diagnostics"** organized by Dr. B. Borooah Cancer Institute Guwahati and National Institute of Pathology, New Delhi during 28th Feb. -1st March 2012 at BBCI Guwahati

Publications: -

Research Articles**:**

1. Chandra M., **Saxena, S.**, Dave P.K., Kaza R. M. and Saxena, H.M.K.: **Proliferative myositis (a pseudosarcomatous lesion of skeletal muscle) occurring in an infant**. *Ind. J. Path. & Microbiol :* 26, 213, 1983.

2. Nagar P., **Saxena S.** (nee Sinha). Pratap V.K. and Mehrotra M.L.: **Malignant haemangiopericytoma of the uterus**. **A case report with brief review of literture**. *J. of Obst. Gynaec. Of India : 34,753, 1984.*

3. **Saxena S.**, Andal, A. and Saxena H.M.K.: **Ultrastructure study of minimal change nephrotic syndrome – a clinico-morphologic correlation**. *Ind. J. Med. Res.* 82: 171, 1985.

4. **Saxena S.,** Andal A, and Saxena H.M.K. **Stereomicroscopic examination of kidney tissue for rapid identification of glomerulus**. *Nephron:* 45: 249, 1987.

5. **Saxena S.,** Mehrotra M.L.: **Host tissue response in soft tissue sarcomas.** *Ind. J. Path. & Microbiol*. 30:97, 1987.

6. **Saxena S.,** Andal A, and Saxena H.M.K. **Idiopathic nephrotic syndrome of childhood: Ultrastructural immunohistologic and Clinicocomorphologic correlation.** *Ind. J. Path. & Microbiol.* 31 (3) 195, 1988.

7. Andal A, **Saxena S**, Chellani H.K. and Sharma S. **Pure Mesangioproliferative Glomerulonephritis. A Clinicomorphologic analysis and its possible role in morphological transition of minimal change lesion to Focal glomerulosclerosis**. *Nephron:* 51(3): 314, 1989.

8. **Saxena S,** Davies D.J., Krisner R.L.G. **Thin basement membrane in minimally abnormal glomeruli.** *J. Clin. Pathol*. 43: 32, 1990.

9. **Saxena S.**, Andal A, Saxena R.K., Sharma S, Chandra M, Saxena H.M.K. **Immune status of children suffering from Minimal change nephrotic syndrome**. *Ind. J. Path. & Microbiol.* 35(3) 171, 1992.

10. **Saxena. S.**, Davies D.J., **Glomerular alterations in Idiopathic haematuria– Ultrastructural and Morphometric analysis**. *Ind. J. Path. & Microbiol.* 35(4), 326-332, 1992.

11. **Saxena. S.**, Andal. A., Sharma. S, Saxena H.M.K., Chandra M. **Immunomodulation by measles vaccine in children with Minimal change nephrotic syndrome**. *Indian J. of Nephrology* 2, 141-146, 1992

12. Verma. A.K., Tandon, R.., **Saxena. S.**, Pandy, J., Talib. V.H. **Aspiration Cytology of maxillary myxoma**. *Diagnostic Cytopathology* 9(2), 202-204, 1993.

13. **Saxena S**, Mital. A, Andal A,: **Pattern of interleukins in MCNS of childhood**. *Nephron* 65(1) 56-61, 1993

14. **Saxena S.**, Bhargawa R., Mohanty N.K., Talwar M: **Primary adenocarcinoma of the urinary bladder. A case report with review of literature** *Ind J Pathol and Microbiol.* 37(4), 453, 1994.

15. Saha T.K., Jolly B B., Mohanty N.K., **Saxena S.**, Dawson. L. **Multiple stones in Ectopic megaureter with Dysgenetic kidney – A case report**. *Ind. J. Nephrol*. 4(2). 61, 1994.

16. **Saxena S**. **Cytokine growth factors and childhood nephrotic syndrome**. *Jr. of Nephrol*. Vol. 8(6), 287, 1995.

17. Mohanty NK, Jolly BB, **Saxena S**, Dawson L. **Squamous cell carcinoma of peripheral urethrostomy**. *Urol. Int.* 1995, 55: 118-119.

18. Mohanty NK, Jolly BB, Talwar M, **Saxena S**, Dawson L. **Aspergillosis kidney. A case report**. *Indian Jr. of Nephrol*. 6(2), 56-58, 1996.

19. **Saxena S**, Jain A K, Pandey K K, Dewan A K. **Study on role of Steroid Hormone Receptors, Growth factor/receptors and Proto-oncogenes on behavior of Human Mammary Epithelial cancer cells in vitro.** *Pathobiology* 65(2), 75-82, 1997.

20. **Saxena S**, Mohanty N K, Talwar M, Jain A K. **Screening of Prostate Cancer in males with prostatism**. *Ind. J. of Path & Microbiol*. 40(4), 441-450, 1997.

21. Mohanty N K, Gulati P, **Saxena S**. **Role of interferon α-2b in the prevention of superficial carcinoma of bladder recurrence**. *Urol. Intern*.59: 194-196, 1997.

22. Mohanty N.K., Jha AK, **Saxena S**, Kumar S., Arora RP. **Ten years experience with Adujuvant Intravesical Immunotherapy in management of superficial transitional cell carcinoma of Urinary bladder – A review**. *Ind. J. of urology* 2001, 17, 127.

23. **Saxena S**, Jain A.K., Bhatnagar D. **Study of events leading to cellular Senescence to Human Mammary epithelial cancer cells in vitro**. *Indian. J. Cancer*. 38: 103-116, 2001

24. **Saxena S**., Beena KR, Bansal A, Bhatnagar A. **Emperipolesis: Significance of an unusual phenomenon in common breast malignancy**. *Acta Cytologica.* 46: 883-886, 2002

25. **Saxena S**, Szabo C, Barjhoux H, Chopin S, Siniliniova O, Lenoir G, Goldgar D, Bhatnagar D**. BRCA 1 and BRCA 2 in Indian Breast Cancer Patients.** *Human mutation.* 20 (6): 473-74, 2002.

26. Chintamani, Sharma R D, Bardan R, Singhal V, **Saxena S,** Bansal A **Sweat gland Adenocarcinoma – a rare clinical dilemma**. *World J Sur. Oncol*. 1: 13, 2003

27. Bharat R, Burra U, Vidyadharan G, **Saxena S**. **Morphogical spectrum of cysticercus cellulose on cytology in case of malnutrished child**. *J Cytol*. 21 (2): 95-06, 2004

28. **Saxena S,** Bansal A, Mohil R S, Bhatnagar D. **Metaplastic carcinoma of the breast-A rare breast tumor.** *Ind J Pathol and Microbiol.* 47(2): 217-220, 2004

29. **C**hintamani, Shankar M, Singhal V, Singh J P, **Saxena S**. **Squamous cell carcinoma developing in the scar of fournier's gangrene-case report.** *BMC Cancer*. 4:16, 2004.

30. Bharat R, **Saxena S,** Burra U. **Fine needle aspiration cytology of Dermato fibrosarcoma protuberans**. *J Cytol*. 21(3), 2004

31. **C**hintamani, Singhal V, Singh J P, Bansal A, **Saxena S**, Lyall A. **Is drug induced cytotoxicity a good predictor of response to new adjuvantchemotherapy in breast cancer? A prospective clinical study.** *BMC Cancer*. Aug 13; 4(1): 48, 2004

32. Mukherji A, Madholia V, Malhotra S, Singh P, Rekhi B, **Saxena S**, Aggarwal Y, Bhowmik K.T. **Multiple Myeloma Of The Breast –An Unusual Case Of Multiple Myeloma Of The Breast With Pathological Fractures Of Humerus and Femur**. *Jr. Of Clinical Radiotherapy and Oncology.4 (4):27-30, 2004.*

33. Mohanti. N.K, **Saxena . S**, Goyal .N.K, Singh. U.P, Arora .R.P **Delayed Cystectomy for T1G3 TCC of Urinary Bladder Managed initially by TURBT & Intravesical Immunotherapy (BCG+Interferron)—Rationale & our results.** *Indian Journal of Urology, 2004, vol. 20: 2*.

34. **C**hintamani, Singhal V, Singh J P, Bansal A, **Saxena S**. **Half versus full vacuum suction drainage after modified radical mastectomy for breast cancer-A prospective randomized clinical trial** (ISRCT N24484328). *BMC Cancer*. 5:11, 2005**.**

35. Rekhi B, Bansal A, Bhatnagar D, Bhatnagar A, **Saxena S. Cytomorphological study of soft tissue neoplasms: role of fluorescent immunocytochemistry in diagnosis** *Cytopathology, 16(5) :219-26, 2005.*

36. Rekhi B, **Saxena S,** Chintamani**. Gastric outlet obstruction and cutaneous metastasis in Adenocarcinoid Tumor of Stomach- Unusual presentations with cytologic and Ultrastructural findings.** *Indian J Cancer. ,42(2):99-101,2005*

37. Ghai R, Rekhi B, **Saxena S**, Kapoor S. **An unusual presentation of Primary Lympoma of the Thyroid in a young male patient—A case report.** *I.J.P.M.,48(3): 385-387,2005*

38. Murthy NS, Chaudhary K, Saxena **S**. **Trends in Incidence of Cervical Cancer – Indian Scenario.** *Euro. J Can Prev. 2005 Dec; 14(6):513-8.*

39. A. Agarwal, S. Verma, U. Burra, NS Murthy, NK Mohanty and S. Saxena Ca"**Flow Cytometric analysis of Th1 and Th2 cytokines in PBMCs as a parameter of immunological dysfunction in patients of Superficial Transitional cell carcinoma of bladder".** *Cancer Immunology and Immunotherapy. 2006, 55(6), 734-743.*

40. N.K.Mohanty, **Sunita Saxena**, Uday Pratap Singh, Neeraj K.Goyal, R.P.Arora "Lycopene **as a chemoprevetvive agent in the treatment of High Grade Prostate Intraepithelial Neoplasia.**" *Urol Oncol. 2005 Nov-Dec; 23(6):383-5.*

41. Chintamani, Jai Prakash Singh, Mahesh K Mittal, **Sunita Saxena**, Anju Bansal, Ashima Bhatia, Pranjal Kulshreshtha **The role of P--glycoprotein expression in predicting response to neoadjuvant chemotherapy in breast cancer-a prospective clinical study.** *World Journal of Surgical Oncology* 2005 **3**:61

42. **Sunita Saxena** ; Bharat Rekhi ; Anju Bansal ; Ashok Bagga ; Chintamani C and N.S.Murthy: **Clinico-morphological patterns of Breast Cancers Including family history in a Delhi hospital, India- A Cross-sectional study** *World Journal of Surgical Oncology 2005, **3:**67*

43. Burra UK, Singh A, **Saxena S**. **Eccrine porocarcinoma (malignant eccrine poroma): a case report.** *Dermatol Online J. 2005 Aug 1; 11(2):17.*

44. **Sunita Saxena**, Usha Agrawal, Abhilasha Agarwal*, Saurabh Verma, NS Murthy*, NK Mohanty **"Adjuvant Intravesical Therapy Based on In Vitro Cytotoxicity Assay In Management Of Superficial Transitional Cell Cancer of Urinary Bladder"** *BJU International 2006 Vol. 98 (5), 1012*

45. Bharat Rekhi, **Sunita Saxena "New Pot-pourri of Markers related to Invasive Breast Cancer"** *JIMSA 2006 19(1), Jan-Mar.*

46. Rekhi B, **Saxena S "Cytomorphology of Basal Cell Type Of Solid Ameloblastoma-A Case Report"** *J Of Cytology 2006;23:83-85*

47. **Sunita Saxena**, Anurupa Chakraborty, Mishi Kaushal Sanjeev Kotwal, Dinesh Bhatnagar, RS Mohil, Chintamani Chintamani, AK Aggarwal, Veena Sharma, PC Sharma, Gilbert Lenior and David Goldgar, Csilla Szabo **Contribution of germline BRCA1 and BRCA2 sequence alterations in to breast cancer in Northern India.** *BMC Medical Genetics 2006, 7:75.*

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

49. Singh A, **Saxena S**. **"Infiltrating Duct Carcinoma Of Breast, Metastatic to Axillary Lymph Nodes Harbouring Primary Tuberculous Lymphadinitis"** *Pathology Oncology Research,12(3),2006*

50. Singh Avninder, Amar Bhatnagar, Usha Agrawal and **Sunita Saxena**. **Isolated splenic metastasis from colorectal mucinous carcinoma: a case report** *International Journal of Gastrointestinal Cancer 2006;37(2-3):98-101*

51. N S Murthy, Usha K Burra, K Chaudhry, and **S Saxena**" **Trends in incidence of breast cancer-Indian Scenario".** *Europian Jr. Of Cancer Care. doi:10.1111/j.1365-2354.2006.*

52. Anurupa Chakraborty[1], N.S. Murthy[2], Chintamani[3], D Bhatnagar[3], R.S. Mohil[3], A. Bhatnagar[3], P.C. Sharma[4]**, Sunita Saxena[1] CYP 17 gene polymorphism and its association with high-risk North-Indian breast cancer patients"** *Journal of Human Genetics 52(2):159-165,2007*

53. IndranilChatterjee, SujalaKapur, JoydeepPurkayastha, Rupkumar Phukan, Amal Kataki, Jayanta Mahanta, **Sunita Saxena**. **Gene expression profile of esophageal cancer in North East India by cDNA microarray analysis.** *World Jr of Gastroentrology 2007; 13(9):1438-1444.*

54. Chintamani, Binita P Jha, Anju Bansal, **Sunita Saxena** and Dinesh Bhatnagar **The expression of mismatched repair genes and their correlation with clinicopathological parameters and response to neo-adjuvant chemotherapy in breast cancer** *International Seminars in Surgical Oncology.20074:5.*

55. Chintamani, Pranjal Kulshreshtha, Nidhi Sugandhi, Anju Bansal, Dinesh Bhatnagar and **Sunita Saxena. Is an aggressive approach justified in the management of an aggressive cancer-the squamous cell carcinoma of thyroid?** *International Seminars in Surgical Oncology* 2007, 4:8 doi:10.1186/1477-7800-4-8

56. Chintamani, Rohan Khandelwal, Aliza Mittal, Sai Saijanani, Amita Tuteja, Anju Bansal, Dinesh Bhatnagar and **Sunita Saxena. Qualitative and quantitative dermatoglyphic traits in patients with breast cancer: a prospective clinical study**
*BMC Cancer 2007, **7:**44*

57. Chintamani, Vinay Singhal, Anju Bansal, Dinesh Bhatnagar and **Sunita Saxena. Isolated colostomy site recurrence in rectal cancer -two cases with review of literature**
*World Journal of Surgical Oncology 2007 **5:**52*

58. Tyagi I, Agarwal U, Amitabh V, Jain A K, **Saxena S. Thickness of Glomerular and Tubular basement membranes in preclinical and clinical stages of Diabetic Nephropathy.** *Indian Jr of Nephrology* 2008; 18(2):60-65.

59. Singh, A., Kapur, S. and Saxena, S. **Cytokeratins and gastrointestinal cancer: A brief review.** *Gastroenterol Today 12 (2008) 115.*

60. Sharma, M., Chintamani, Saxena, S. and Agarwal, U. **Squamous cell carcinoma arising in unilateral Warthin's tumor of parotid gland.** *J Oral Maxillo Facial Pathol 12 (2008).*

61. Anurupa Chakraborty[1], A.K Mishra[1], Abha Soni[1], Thodum Regina[1], D Bhatnagar,[2] A Bhatnagar[2], Chintamani,[2] Sunita Saxena[1] **VDR gene polymorphism(s) and breast cancer risk in North Indian Population.** *Cancer Detection and Prevention* **32** (2009) pp. 386-394

62. Chattopadhyay I, Phukan R, Vasudevan M, Singh A, Purkayastha J, Hewitt S, Kataki A, Mahanta J, Kapur S, Saxena S; **"Molecular profiling to identify molecular mechanism in esophageal cancer with familial clustering"** *Oncology Reports 21:1135-1146,2009*

63. Chintamani, T. Aeron, M. Mittal, D. Bhatnagar, U. Agarwal, S. Saxena **Are the structures preserved in functional neck dissections truly preserved functionally? – A prospective study of patients with head and neck cancer at a tertiary cancer care center** *Oral Oncology Supplement, Volume 3, Issue 1, July 2009, Page 175*

64. Murthy, N.S., Chaudhry, K., Nadayil, D., Agarwal, U.K. and Saxena, S. **Changing trends in incidence of breast cancer: Indian Scenario.** *Indian J Cancer 46 (2009) 73.*

65. Avninder Singh, ; Sujala Kapur,; Indranil Chattopadhyay,; Joydeep Purkayastha,; Jagannath Sharma,; Ashwani Mishra,; Stephen M. Hewitt; Sunita Saxena,. **Cytokeratin immunoexpression in esophageal squamous cell carcinoma of high-risk populationin Northeast India.** *Applied Immunohistochemistry & Molecular Morphology* , 17(5):419-424,Oct.2009

66. Mishi Kaushal, Indranil Chattopadhyay, Rupkumar Phukan, Joydeep Purkayastha, Jagadish Mahanta, Sujala Kapur, Sunita Saxena. **Contribution of germline *BRCA2* sequence alterations to risk of familial esophageal cancer in high-risk area of India.** *Disease of the Esophagus*. DOI:10.1111/j.1442-2050.2009.00975.x (published online),2010:23(1) 71-5.

67. Agrawal A, Agrawal U, Verma S, Mohanty N.K and SaxenaS. **Serum Th1 and Th2 cytokine balance in patients of superficial transitional cell carcinoma of bladder pre and post intravesical combination immunotherapy.** *Immunopharmacology and Immunotoxicology.2010:32(2)348-56*

68. Regina Devi T, Yadav DS, Mishra AK, Kaushal M, Ihsan R, Chattopadhayay I, Chauhan P, Sarma J, Zomawia E, Verma Y, Nandkumar A, Mahanta J, Phukan R, Kapur S, Saxena S: **Distribution of Glutathione S- transferase T1 and M1 genes polymorphisms in North East Indians. A potential report.** *Genetic Testing and Molecular Biomarkers.* 14(2);163-169,2010

69. Chintamani Chintamani, Rohan Khandelwal, Megha Tandon, Yashwant K, Pranjal Kulshreshtha, Tushar Aeron, Dinesh Bhatnagar, Anju Bansal, Sunita Saxena **Carcinoma developing in a fibroadenoma in a woman with a family history of breast cancer: a case report and review of literature** *Cases Journal 2009, 2:9348*

70. Chattopadhyay I, Singh A, Phukan R, Purkayastha J, Kataki A, Mahanta J, Saxena S, Kapur S .**Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma exposed to tobacco and betel quid from high-risk area in India** Mutation Research - Genetic Toxicology and Environmental Mutagenesis 696 (2010), pp. 130-138 DOI information: 10.1016/j.mrgentox.2010.01.001
71. Rakshan I, Chattopadhyay I, Phukan R, Mishra A K, Purkayastha J, Sharma J, Zomawia E, Verma Y, Nandkumar A, Mahanta J, Saxena S, Kapur S. **Role of EPHX1 gene polymorphisms in esophageal cancer of high-risk area in India.** Jr. of Gastroenterology and Hepatology 2010 Aug; 25(8):1456-62.

72. Usha Agrawal, Ashwani K Mishra, Payal Salgia, Saurabh Verma, Nayan K Mohanty, **Sunita Saxena. Role of Tumor Suppressor and Angiogenesis Markers in Prediction of Recurrence of Non Muscle Invasive Bladder Cancer.** Pathology and Oncology Research 17(1); 91-101,2011, DOI: 10.1007/s12253-010-9287-1

73. Dhirendra Singh. Yadav, Thoudam Regina Devi, Rakhshan Ihsan, AK Mishra, Mishi Kaushal, Indranil Chattopadhyay, Pradeep Singh Chauhan, Jagannath Sharma, Eric Zomawia, Yogesh Verma, A. Nandkumar, Jagadish Mahanta, Rupkumar Phukan, Sunita Saxena., Sujala Kapur. **Polymorphisms of Glutathione-S-transferase (GST) genes and the risk of aerodigestive cancers in Northeast Indian population.** Genetic Testing and Molecular Biomarker 14(5);1-9,2010

74. Pradeep Singh Chauhan1, Rakhshan Ihsan1, Dhirendra Singh Yadav1, Ashwani Kumar Mishra1, Bharat Bhushan 1, Abha Soni1, Mishi Kaushal1, Thoudam Regina Devi1,Sumita Saluja2, Dipendra Kumar Gupta2, Vishakha Mittal2, Sunita Saxena1,Sujala Kapur. **Association of *GST, EPHX* and *p53 codon 72* gene polymorphism with adult acute myeloid leukaemia.** DNA and Cell Biology 2011 Jan; 30(1): 39-46.

75. Chintamani**,** Pranjal Kulshreshtha, Anurupa Chakraborty, L C Singh, Ashwani K Mishra, Dinesh Bhatnagar, Sunita Saxena **Androgen receptor status predicts response to chemotherapy, not risk of breast cancer in Indian women** *World Journal of Surgical Oncology* 2010, **8**:64

76. Kaushal M, Mishra AK,Raju BS, Ihsan R, Chakraborty A, Sharma J, Zomawia E, Verma Y, Kataki A, Kapur S. Saxena S: **Betel quid chewing as an environmental risk factor for breast cancer** Mutation Research - Genetic Toxicology and Environmental Mutagenesis 703(2010), 143-148
77. Rakhshan Ihsan, Thoudam Regina Devi, Dhirendra Singh Yadav, Ashwani Kumar Mishra, Jagannath Sharma, Eric Zomawia, Yogesh Verma, Rupkumar Phukan, Jagadish Mahanta, Amal Chandra Kataki, Sujala Kapur, Sunita Saxena. **Investigation on the role of p53 codon 72 polymorphism and interactions with tobacco, betel quid and alcohol in susceptibility to cancers in a high risk population from north east India** DNA and Cell Biology 2011 March; 30(3): 163-171
78. .Chintamani, Jp Singh, Megha Tandon, Rohan Khandelwal, Tushar Aeron, Sidharth Jain, Nikhil Narayan, Rahul Bamal, Yashwant Kumar, S Srinivas, Sunita Saxena **Vulval elephantiasis as a result of tubercular lymphadenitis: two case reports and a review of the literature** *Journal of Medical Case Reports* 2010 **4**:369

79. Abha Soni, Anju Bansal, L C Singh, Ashwani Kumar Mishra, Thoudam Regina, N K Mohanty, Sunita Saxena. **Gene expression profile and mutational analysis of DNA mismatch repair genes in carcinoma prostate in Indian population.** *OMICS: 2011 Feb 24* (Epub ahead of print)
80. Simmy Soni, Gayatri Rath, Chandra Prakash Prasad, Sudha Salhan, Arun Kumar Jain, Sunita Saxena. **Fas-Fas System in Molar Pregnancy.** *American Journal of Reproductive Immunology.* (In Press)

81. Simmy Soni, Gayatri Rath, Chandra Prakash Prasad, Sudha Salhan, Sunita Saxena, Arun Kumar Jain. **Apoptosis and Bcl-2 protein expression in Human Placenta over the course of Normal Pregnancy**. *Anatomia. Hisologia Embryologia.39(2010): 426-431*

82. Chintamani**,** Megha Tandon**,** Ashwini Mishra**,** Usha Agarwal and **Sunita Saxena. Sentinel lymph node biopsy using dye alone method is reliable and accurate even after neo-adjuvant chemotherapy in locally advanced breast cancer- a prospective study** World Journal of Surgical Oncology 2011, 9:19

83. Agarwal S, Agrawal U, Mohanty NK, Saxena S. **Multilocular Cystic Renal Cell Carcinoma: A case report of rare entity.** Arch Pathol Lab Med—Vol135, March 2011

84. L C Singh, Anurupa Chakraborty, Ashwani K Mishra, Thoudam Regina Devi, Nidhi Sugandhi[2], Chintamani, Dinesh Bhatnagar, Sujala Kapur, **Sunita Saxena[*]"Study on predictive role of AR and EGFR family genes with**

response to Neo-adjuvant Chemotherapy in Locally Advanced Breast Cancer in Indian women" Medical Oncology 2012, 29(2) : 539-546

85. Singh A, Mishra A K, Ylaya K, Hewitt S M, Sharma K C, Saxena S. **Wilms Tumor-1, Claudin-1 and Ezrin are useful Immunohistochemical markers that helps to distinguish Schwannoma from Fibroblastic Meningiomas.** Pathol. Oncol. Res. DOI 10.1007/s12253-011-9456-x

86. Anju Bansal, Abha Soni, Punita Rao, LC Singh, Ashwini Mishra, N K Mohanty, Sunita Saxena. **Implication of DNA repairs genes in prostate carcinogenesis in Indian men.** Indian J Med Res 136, October 2012, pp 622-632

87. Ihsan R, Chauhan PS, Mishra AK, Yadav DS, Kaushal M, Sharma JD, Zomawia E, Verma Y, Kapur S, Saxena S. **Multiple Analytic Approaches reveal distinct Gene- Environment interactions in Smokers in Lung Cancers. PLoS One**, 6(12): e29431, 2011.

88. Bansal A, Bhatnagar A, Saxena S. **Metastasizing granular cell ameloblastoma.** J Oral Maxillofac Pathol 2012;16:122-4.

89. Abha Sony, Anju Bansal, Aswini Kumar Mishra, Jyotsna Batra, L.C. Singh, Anurupa Chakraborty, Dhirendra Singh Yadav, N. K. Mohanty, Sunita Saxena **"Association of Androgen Receptor, Prostate Specific Antigen and CYP19 gene polymorphisms to Prostate Carcinoma and Benign Prostatic Hyperplasia in North Indian population"** Genetic Testing and Molecular Biomarker (In Press)

90. Mishra AK, Agrawal U, Negi S, Bansal A, Bhatnagar A, Bhatnagar D, Chintamani, Mohil R, Saxena S. **Study on expression of AR in Breast Cancer and its correlation with other steroid receptors and growth factors**. Indian J Med Res 135, June 2012, pp 843-852

91. Chauhan PS, Ihsan R, MishraAK,Yadav DS, Saluja S, Mittal V, Saxena S, and Kapur S: **High Order Interactions of Xenobiotic Metabolizing Genes and P53 Codon 72 Polymorphisms in Acute Leukemia.** Environmental and Molecular Mutagenesis (2012)

92. Mishi Kaushal, Ashwani. K. Mishra, Jaganath Sharma, Eric Zomawia , Amal Kataki, Sujala Kapur, Sunita Saxena **Genomic alterations in breast cancer patients in betel quid** 1 **and non betel quid chewers. PLoS One 2012**

Books-

1. Mishra AK, Chakraborty A, Saxena S: Significance of Vitamin D receptor polymorphisms in Breast Cancer—Multinomial Logistic Regression Analysis in Vitamin D,: Nutrition, Side effects and supplements. Editor: Stephanie R. Malone
2. Sunita Saxena: Molecular and Genetic Aspects of Lung Cancer. Chapter 13 pp 141-148 In Pathological and Occupational Lung Health. Editors: V.K.Vijyan, H.K.Tazelaar and Ritu Kulshrestha

OPEN ACCESS Freely available online

PLOS ONE

Genomic Alterations in Breast Cancer Patients in Betel Quid and Non Betel Quid Chewers

Mishi Kaushal[1], **Ashwani K. Mishra**[1], **Jagannath Sharma**[2], **Eric Zomawia**[3], **Amal Kataki**[2], **Sujala Kapur**[1], **Sunita Saxena**[1]*

1 National Institute of Pathology, ICMR, Safdarjung Hospital Campus, New Delhi, India, **2** Dr. B. Borrooah Cancer Institute, Guwahati, India, **3** Civil Hospital, Aizwal, India

Abstract

Betel Quid (BQ) chewing independently contributes to oral, hepatic and esophageal carcinomas. Strong association of breast cancer risk with BQ chewing in Northeast Indian population has been reported where this habit is prodigal. We investigated genomic alterations in breast cancer patients with and without BQ chewing exposure. Twenty six BQ chewers (BQC) and 17 non BQ chewer (NBQC) breast cancer patients from Northeast India were analyzed for genomic alterations and pathway networks using SNP array and IPA. BQC tumors showed significantly ($P<0.01$) higher total number of alterations, as compared with NBQC tumors, 48±17% versus 32±25 respectively. Incidence of gain in fragile sites in BQC tumors were significantly ($P<0.001$) higher as compared with NBQC tumors, 34 versus 23% respectively. Two chromosomal regions (7q33 and 21q22.13) were significantly ($p<0.05$) associated with BQC tumors while two regions (19p13.3–19p12 and 20q11.22) were significantly associated with NBQC tumors. GO terms oxidoreductase and aldo-keto reductase activity in BQC tumors in contrast to G-protein coupled receptor protein signaling pathway and cell surface receptor linked signal transduction in NBQC tumors were enriched in DAVID. One network "Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism" including genes AKR1B1, AKR1B10, ETS2 etc in BQC and two networks "Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry" and "Cellular Development, Embryonic Development, Organismal Development" including genes RPN2, EMR3, VAV1, NNAT and MUC16 etc were seen in NBQC. Common alterations (>30%) were seen in 27 regions. Three networks were significant in common regions with key roles of PTK2, RPN2, EMR3, VAV1, NNAT, MUC16, MYC and YWHAZ genes. These data show that breast cancer arising by environmental carcinogens exemplifies genetic alterations differing from those observed in the non exposed ones. A number of genetic changes are shared in both tumor groups considered as crucial in breast cancer progression.

Citation: Kaushal M, Mishra AK, Sharma J, Zomawia E, Kataki A, et al. (2012) Genomic Alterations in Breast Cancer Patients in Betel Quid and Non Betel Quid Chewers. PLoS ONE 7(8): e43789. doi:10.1371/journal.pone.0043789

Editor: Reiner Albert Veitia, Institut Jacques Monod, France

Received May 24, 2012; **Accepted** July 24, 2012; **Published** August 24, 2012

Funding: This study was supported by grants from Indian Council of Medical Research (ICMR), New Delhi, India(Ref no,123/4/RMRC/NE/2005-NCD-II/III. Mishi Kaushal is a recipient of a Senior Research Fellowship of Indian Council of Medical Research (ICMR), New Delhi, India (Ref. No.3/2/2/183/2009/NCD-III). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sunita_saxena@yahoo.com

Introduction

Breast cancer is the most common malignancy worldwide among women attributed to various genetic and environmental factors [1]. In India it constitutes 22.2% of all cancers with approximately 115,000 incident cases reported in 2008 [2]. The several fold difference in incidence rates between different geographical regions suggest that environmental factors influence breast cancer risk significantly [3]. Both high and low age-adjusted breast cancer incidence rates (AAR) have been observed in Northeast India (23.3 in Aizwal to 12.1 in Dibrugarh in 2008) which has steadily increased [4].

A previous case control study on assessment of various environmental and genetic factors in Northeast Indian population illustrated significant increase in breast cancer risk in women who consumed Betel Quid (BQ) [3]. In the Northeast region of India BQ is consumed as a mixture of areca nut (Areca catechu), catechu (Acacia catechu) and slaked lime (calcium oxide and calcium hydroxide) wrapped in betel leaf (Piper betel) and tobacco [5].

BQ independently contributes to the risk of oropharyngeal cancer, oral mucosal lesions, oral leukoplakia, oral submucous fibrosis, liver cirrhosis and hepatocellular carcinoma [6]. In vitro and in vivo experiments have shown that BQ consumption can also cause micronuclei and DNA adducts formation, chromosomal aberrations, allelic imbalances and sister chromatid exchange in oral mucosa cells [7]. Carcinogens in BQ lead to accumulation of genetic alterations at 3q26.3 locus particularly in recurrent oral tumors [8] besides accelerating tumor migration by stimulating MMP-8 expression through MEK pathway [9].

In addition, calcium hydroxide a major content of slaked lime in the presence of areca nut is responsible for the formation of ROS (reactive oxygen species) known to cause oxidative damage in the DNA of buccal mucosa cells of BQ chewers. Presence of iron and copper transition metals are also involved in the catalytic process of ROS generation [5]. This ROS generation leads to structural alterations in DNA, including rearrangements, deletions, insertions and sequence amplification, affect cytoplasmic and nuclear signal transduction pathways, modulate the activity of the proteins and genes that respond to stress and act to regulate genes related to cell proliferation, differentiation and apoptosis [10].

Tobacco chewing with BQ results in increased exposure (~1000 µg/day) to carcinogenic tobacco-specifc nitrosamines

(TSNAs). High levels of TSNAs have been found in saliva samples of BQ chewers collected from India. N'-nitrosonornicotine (NNN), 4-(N-methyl-N-nitrosamino)-1-(3-pyridyl)-1-butanone (NNK), N-nitrosoanabasine (NAB), N-nitrosodimethylamine and N-nitrosodiethylamine have been detected in saliva of BQ with tobacco chewers [5], breast tissue of women workers and are known to induce mammary tumors in rodents and anaphase bridges via DNA double stranded breaks causing genomic imbalances in human cells [11,12]. Regions like 7p11.2 (epidermal growth factor receptor) and 11q13.3 (cyclin D1) playing a role in pathogenesis of tobacco-related human squamous cell carcinoma has been identified by SNP array [13]. Examination of genomic alteration due to tobacco carcinogens depicts gain on chromosomes 6 and 8, and losses on chromosomes 11 and 14 in mouse lung adenocarcinomas [14] and gains of 1p and 3q in patients with tobacco exposure history in head and neck squamous cell carcinomas [15]. In addition, Benzo(*a*)pyrene [B(a)P] diolepoxide (BPDE), a carcinogen present in cigarette smoke, induces chromosomal 9p21 aberrations seen to be significantly higher in peripheral blood lymphocytes of bladder cancer cases than that of controls [16]. Allelic imbalance at 5q22.2~q22.3 (LOX gene) is significantly higher among smokers than nonsmokers in clear cell renal carcinomas indicating that tobacco may cause genetic alterations [17].

Earlier studies on genomic alterations in breast cancer have investigated copy number changes between different subtypes and BRCA predisposed breast tumors and cell lines [18 19 20]. Although, the literature suggests role of BQ carcinogens in mediating genomic alterations, there is dearth of evidence suggesting its role in breast carcinogenesis. The present study has been undertaken to elucidate the genetic alterations induced by BQ chewing leading to breast carcinogenesis utilizing whole genome SNP array and Ingenuity pathway analysis in breast cancer patients with and without BQ chewing history.

Results

10K SNP Array Profiles of Overall Breast Cancer Patients

Forty-three tissue samples of breast carcinoma and fourteen matched samples of germline DNA were analyzed for copy number alterations. The mean age of cases was 44.4±9.6 years and maximum cases were between 40–49 years. Twenty six patients were with BQ chewing history (BQC) and seventeen patients were without BQ chewing history (NBQC). Among the total cases, 23 cases were premenopausal and 20 cases were postmenopausal. Stage IV tumors were more, followed by stage III and II tumors whereas stage for six tumors was unknown. The association between above groups of patients with regard to patient age at diagnosis, tumor stage and menopausal status was statistically insignificant and no sample had a previous family history of cancer, alcohol drinking and tobacco smoking (Table S1). 110 recurrent altered regions were identified ranging from 0.15 Mb to 51 Mb in size (Table S2) with more gains than losses. More than 40% alterations were observed in 30 regions which were essentially gains (1q24.1, 1q25.2, 1q31.1, 1q32.1,1q41, 1q42.2–1q42.3, 1q43–44, 2p11.2, 5p13.3, 5p15.2, 7p12.3–7p12.1, 7p14.1–7p12.3, 7p14.3, 7p21.2, 7p21.3–7p21.2, 7q33, 8q12.1–8q12.2, 8q13.2, 8q22.1, 8q22.2, 8q22.3, 8q24.11, 8q24.21, 12q22.3, 16p13.3–16p13.2, 17q23.2, 17q23.3, 20q13.2, 20q13.33, 21q22.1. Most of the recurrent alterations observed were focal amplification (<10 Mb). Although, most chromosomes depicted multiple regions of alteration, 10, 22 and X chromosomes were altered in only single region with no alteration in chromosome 18. Frequent gains were observed in regions of long arm of chromosome 1 and 8 with genes implicated in cancer. 67 and 50 percent samples presented gain at 8q22.1 and 8q24 respectively. 1q43–44 and 1q41 regions presented with gain in 60 and 51 percent samples respectively. The remaining 80 regions were seen to be altered in 39 to 16 percent samples. The key regions comprising tumor associated genes were 6p25.3, 14q21.3, 1q21.1, 15q25.1, 1p13.2,20q13.11, 15q22.2–15q23, 19q13.11, 9p23,11q13.3, 11p14.3, 17q25.1, 3p24.2–3p23, 9q34.11–9q34.3, 5p15.33–5p12, 5q35.1–5q35.3, 22q12.1. Loss in single regions was more frequent than recurrent loss (losses in ≥15% of samples) (Table S2 and S3).

Genetic Alterations Different between BQC and NBQC

Total number of alterations varied considerably between BQC and NBQC tumors from 12 to 93 alterations among BQC tumors, and from 6 to 63 alterations among NBQC tumors. The frequency plot of alterations per chromosome 1–X is shown in Figure S1. The BQC tumors showed a significantly ($P<0.01$, T test) higher total number of alterations, as compared with NBQC tumors (48±17% versus 32±25, respectively) (Table S3). One of the important finding was significantly high incidence of gain in fragile sites in BQC tumors ($P<0.001$, T test) as compared in NBQC tumors, 34 versus 23%, respectively (Table S2). Significant ($P<0.05$) differential genetic alterations were found in twelve chromosomal regions among BQC and NBQC tumors (Table 1, Figure 1). Among the twelve regions seven chromosomal regions (3p26.3, 3q26.1–3q27.2, 4p16.1, 5q11.2–5q12.1, 6q25.3, 7q33 and 21q22.13) presented more gain in BQC tumors while five regions (16p13.12–16p11.2, 17q11.2, 19p13.3–19p12, 19q13.32–19q13.43, 20q11.22) showed more gain in NBQC tumors. The alterations observed were chiefly gains of sizes ranging between 0.65 Mb to 22 Mb. Multiple testing was controlled using the false discovery rate (FDR) q-value method. The FDR cutoff up to 0.2 has been commonly used in case-control GWAS studies [21,22]. FDR correction is likely to be conservative considering the relatively small number of cases, but four differentially altered regions at various chromosomes remained significant, as indicated by relatively low FDR values. The FDR value of 0.26 as for regions (3p26.3, 3q26.1–3q27.2, 4p16.1, 5q11.2–5q12.1, 6q25.3, 16p13.12–16p11.2, 17q11.2, 19q13.32–19q13.43) indicates that the relevance of these finding should be interpreted with caution, and we therefore focused particularly on the regions with P-values = 0.01 and low FDR values. More than 50% BQC tumors presented with gain at 7q33 and 21q22.13 in contrast to just 17% gain in NBQC tumors. Among the regions altered more in NBQC tumors, 52% NBQC tumors had gain at 19p13.3–19p12 in comparison to gain in 11% BQC tumors and 47% of NBQC tumors had gain at 20q11.22 in comparison to gain in 3% BQC tumors.

Gene Ontology (GO) and Network Analyses of Associated Regions

Genes associated with BQC regions, 7q33 and 21q22.1 were enriched for oxidoreductase ($p<0.001$) and aldo-keto reductase activity ($p = 0.015$) in contrast to G-protein coupled receptor protein signaling pathway ($p = 0.005$) and cell surface receptor linked signal transduction ($p = 0.012$) for 19p13.3–19p12 and 20q11.22 NBQC associated regions. IPA (Ingenuity Pathway Analysis) analysis for BQC associated regions revealed one top network (score = 20) "Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism" encompassing genes like AKR1B1, AKR1B10, AKR1B15, ERG, ETS2 (Figure 2). IPA analysis for NBQC genes revealed two top networks (score = 29) "Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochem-

Table 1. Chromosomal areas with gain those are significantly different between betel quid chewers (BQC) and non betel quid chewers (NBQC) breast cancer patients.

Cytoband	BQC (26)	NBQC (17)	P value	Q value (FDR)	Start Site	End Site	Size (Mb)
3p26.3	7	0	0.03	0.26	653347	2264798	1.61
3q26.1–3q27.2	13	3	0.05	0.26	165409849	167801377	2.39
4p16.1	12	2	0.02	0.26	10760950	11857265	1.09
5q11.2–5q12.1	7	0	0.03	0.26	57466589	58659721	1.19
6q25.3	12	2	0.02	0.26	155713132	157738990	2.02
7q33	**16**	**3**	**0.005**	**0.10**	**133281372**	**135010987**	**1.72**
21q22.13	**15**	**3**	**0.01**	**0.10**	**37974454**	**40484883**	**2.51**
16p13.12–16p11.2	2	6	0.04	0.26	10529386	33498455	22.96
17q11.2	2	6	0.04	0.26	22436842	23092917	0.65
19p13.3–19p12	**3**	**9**	**0.005**	**0.10**	**3542590**	**17471210**	**13.92**
19q13.32–19q13.43	4	8	0.03	0.26	51160543	63437743	12.27
20q11.22	**1**	**8**	**0.001**	**0.08**	**31982015**	**35933409**	**3.95**

Frequency of chromosomal regions with significantly different ($P<0.05$; see Materials and methods for the statistical test) alterations between TBC and NTBC tumors are depicted. Most significant regions, based on the criteria of $P<0.05$ and a relatively low FDR value, are indicated in bold. FDR = false discovery rate.
doi:10.1371/journal.pone.0043789.t001

istry" and "Cellular Development, Embryonic Development, Organismal Development" (Figure 3) encompassing genes like RPN2, EMR3, BLCAP and VAV1, NNAT and MUC16 respectively.

Genetic Alterations Similar between BQC and NBQC

Twenty seven common regions of gain were illustrated between BQC and NBQC tumors. Regions demonstrating gain in minimum 30% cases from each group were considered as similarly altered [23]. Both groups exhibited gain on chromosomes 1q, 5p, 7p, 8q, 12q, 16p, 17q and 20q (Table S4). Gain in more than 50% samples was seen in six regions (1q31.1, 1q42.2–1q42.3, 1q43–44, 8q22.1, 8q22.2, 8q24.11). Gain in more than 45% samples was seen at 1q24.1, 1q41, 7p12.3–7p12.1, 8q24.21 and 20q13.2 regions. Other regions encompassing probable tumor associated genes were 1q32.1, 1q21.1, 7p21.3–7p21.2, 12q22.3, 16p13.3–16p13.2, 17q23.3.

Gene Ontology (GO) and Network Analyses of Similar Regions

Enrichment and IPA was performed to investigate the function of genes associated with these regions. Regions were mainly enriched for activation of protein kinase activity ($p = 0.009$), cell junction ($p = 0.01$). IPA analysis revealed three top networks (Table 2, (Figure 4). Network 1 functions in Cellular Movement, Connective Tissue Development and Function, Cellular Assembly

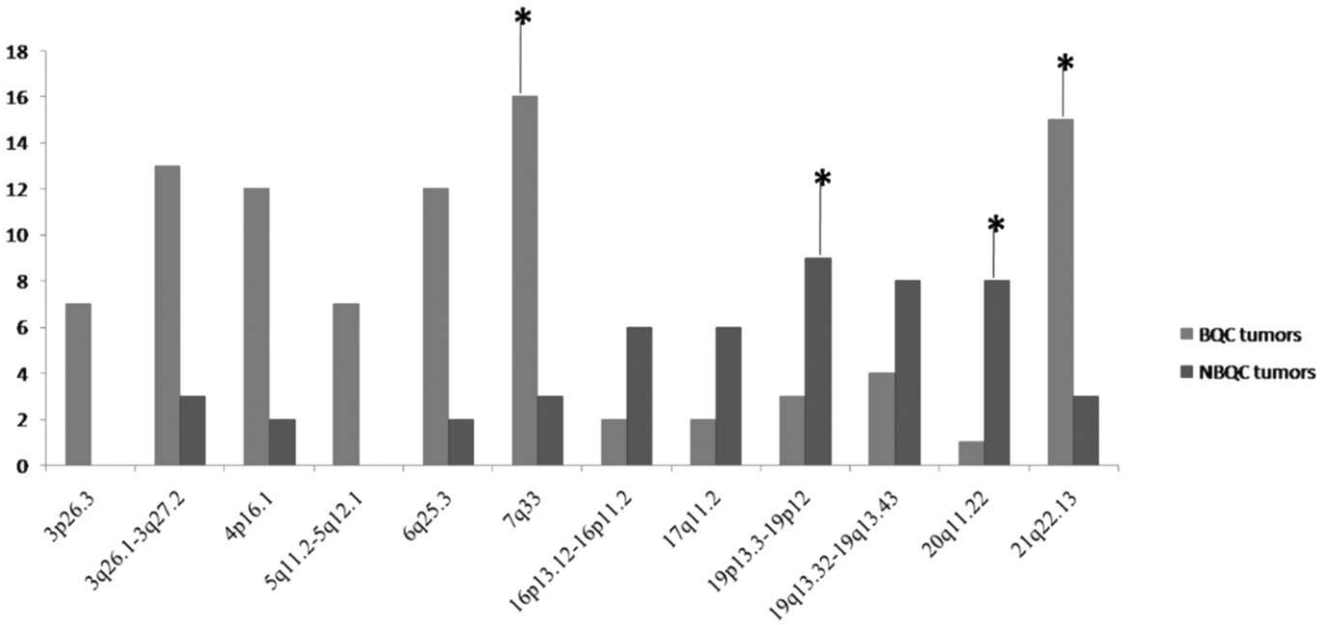


Figure 1. Chromosomal regions altered differently between BQC and NBQC breast tumors. *Regions significant after FDR correction.
doi:10.1371/journal.pone.0043789.g001

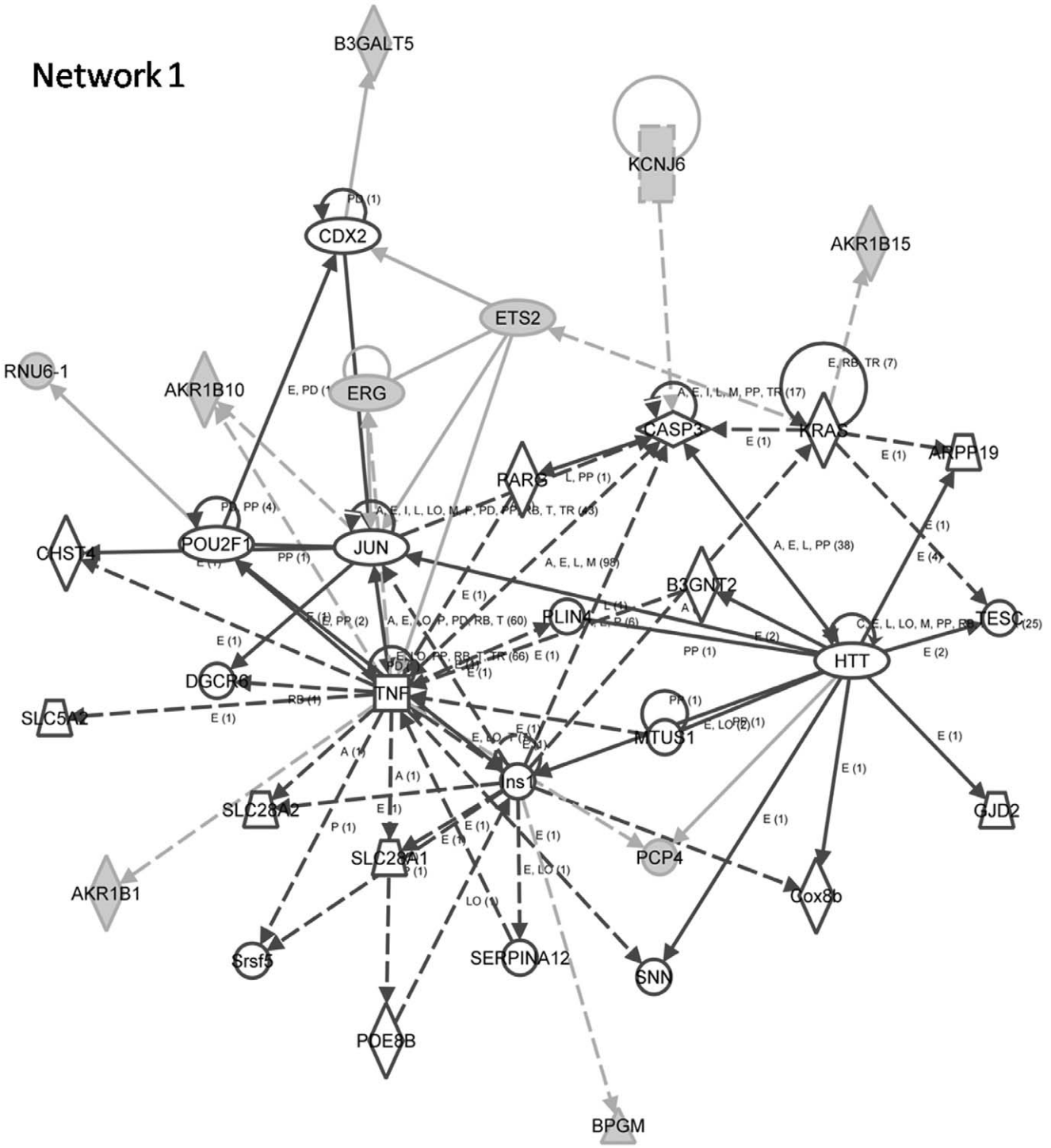


Figure 2. BQC Network 1 Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism.
doi:10.1371/journal.pone.0043789.g002

and Organization (score = 43) with key role played by PTK2. Network 2 functions in Cell-To-Cell Signaling and Interaction, Tissue Development, Organismal Injury and Abnormalities (score = 43) with RPN2, EMR3, VAV1, NNAT and MUC16 important genes. Network 3 functions in Cell Morphology, Cellular Assembly and Organization, Cellular Compromise (score = 32) with key roles played by MYC and YWHAZ. Among all the tumor associated canonical pathways enriched were GNRH signaling (p = 2.92E−04), cAMP-mediated signaling (p = 3.60E−04), Protein Kinase A signaling (p = 3.77E−04), CXCR4 signaling (p = 4.99E−03), molecular mechanisms of cancer (p = 8.58E−03), phospholipase C Signaling (p = 1.01E−02), RAR Activation (p = 3.16E−02), ILK Signaling (p = 4.21E−02)(Table S5, Figure 5).

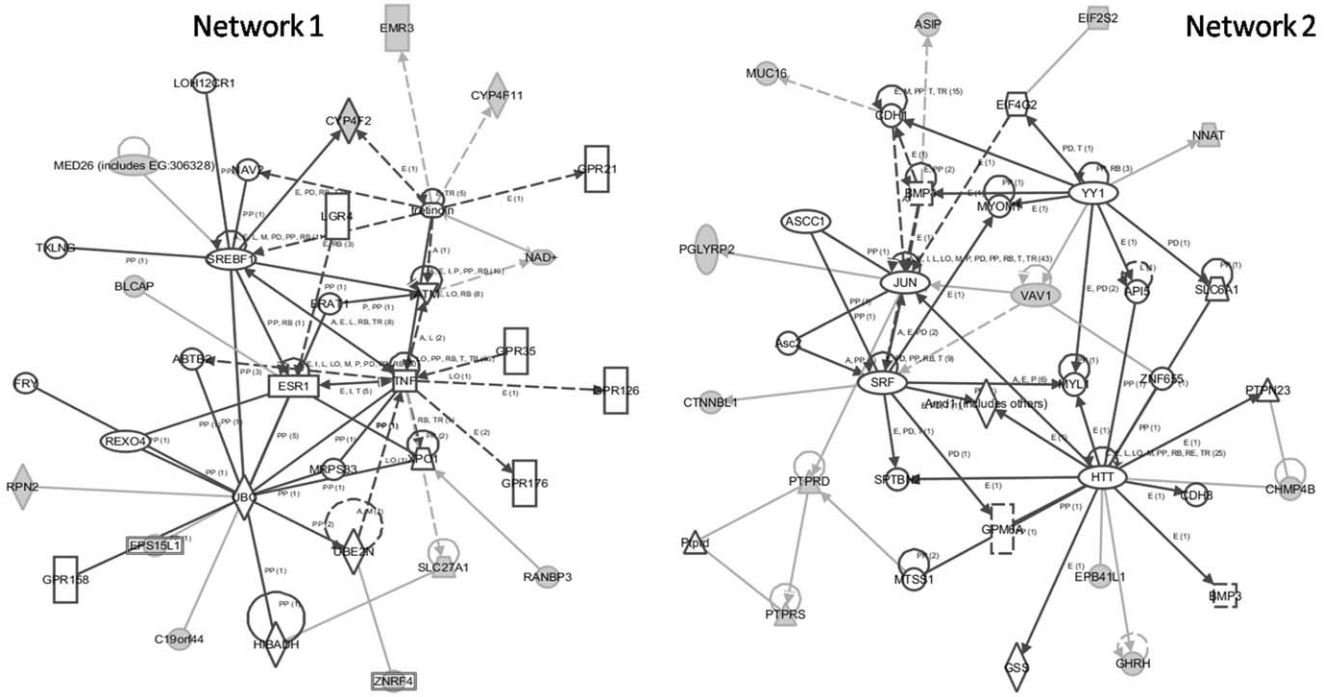


Figure 3. NBQC Networks: Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry (Network 1) and Cellular Development, Embryonic Development, Organismal Development (Network 2).
doi:10.1371/journal.pone.0043789.g003

Table 2. Significant signaling pathway networks observed in BQC, NBQC and Common altered genomic regions.

Networks	Nodes (genes) in Network	Score	Nodes	Identified Nodes	Top Functions
BQC Network 1	AKR1B1,AKR1B10,AKR1B15,ARPP19, B3GALT5,B3GNT2, BPGM,CASP3,CDX2,CHST4, Cox8b,DGCR6,ERG,ETS2,GJD2,HTT,Ins1, JUN, KCNJ6,KRAS,MTUS1,PARG,PCP4,PDE8B, PLIN4, POU2F1,RNU6-1,SERPINA12,SLC28A1,SLC28A2, SLC5A2,SNN,Srsf5,TESC,TNF	28	34	10	Drug Metabolism, Molecular Transport, Nucleic Acid Metabolism
NBQC Network 1	ABTB2,ATM,BLCAP,BRAT1,C19orf44, CYP4F2,CYP4F11,EMR3, EPS15L1, ESR1,FRY,GPR21,GPR35,GPR126, GPR158,GPR176,HIBADH,LGR4, LOH12CR1,MED26 (includes EG:306328), MRPS33,NAD+, NAV2,RANBP3,REXO4, RPN2,SLC27A1,SREBF1,TNF,tretinoin, TXLNG,UBC,UBE2N,XPO1,ZNRF4	29	34	12	Molecular Transport, Nucleic Acid Metabolism, Small Molecule Biochemistry
NBQC Network 2	Amd1 (includes others),API5,Asc2,ASCC1, ASIP,BMP3, BMP4,CDH1,CDH8,CHMP4B, CTNNBL1,EIF2S2,EIF4G2,EPB41L1,GHRH,GPM6A, GSS,HTT,JUN,MTSS1,MUC16,MYL1,MYOM1,NNAT, PGLYRP2,PTPN23,PTPRD,Ptprd,PTPRS,SLC6A1, SPTBN2,SRF,VAV1,YY1,ZNF655	29	34	12	Cellular Development, Embryonic Development, Organismal Development
COMMON Network 1	ADCY1,ADCY10,Alpha tubulin,ANGPT1,Arf, ASAP1,ATP2B4,CACNA1E,Calpain, CAPN9, CDC42BPA,CYP24A1,DISC1,EIF3H,ERK1/2, EXT1,GALNT2,Integrin,KIFAP3, NADPH oxidase, NPHS2,Pdgf (complex),PFDN4,Pld,PTK2 (includes EG:14083), Rac,RAD21,Rap1,Rxr, RXRG,RYR2,TNFRSF11B,TNS3,TRIO,TSH	43	35	25	Cellular Movement, Connective Tissue Development and Function, Cellular Assembly and Organization
COMMON Network 2	ABCG1,ADAMTS12,AEBP1,AKAP1, Alpha catenin,APOH,CDH6,CHN2, Collagen type I,Collagen type IV,DOK5, Ecm,EDARADD,Fibrinogen,GRB10, Growth hormone, GTPASE,HAS2,HDL,LDL, MTDH,Mucin,NFkB (complex),NID1,NOV, PKP1,Pro-inflammatory Cytokine,RAB3GAP2, RGS7,SELP,SNX13,SUMO2,SUMO3,TFF3,WIPI1	43	35	24	Cell-To-Cell Signaling and Interaction, Tissue Development, Organismal Injury and Abnormalities
COMMON Network 3	26s Proteasome,Actin,Ck2,DROSHA,ENPP2, ERMAP,FSH,HEATR1,HELZ,HISTONE, Histone h3,Histone h4,IKK (complex),IKZF1, IPO9,Jnk,MARK1,Max-Myc,MYBPH,MYC, NBPF11 (includes others),NCALD,P38 MAPK, PSEN2,RAI14,RNA polymerase II,SRRM2,STRADA, TARBP1,TBCE,Ube2-ubiquitin,UBE2D4,UBE2G2, Ubiquitin,YWHAZ	32	35	20	Cell Morphology, Cellular Assembly and Organization, Cellular Compromise

doi:10.1371/journal.pone.0043789.t002

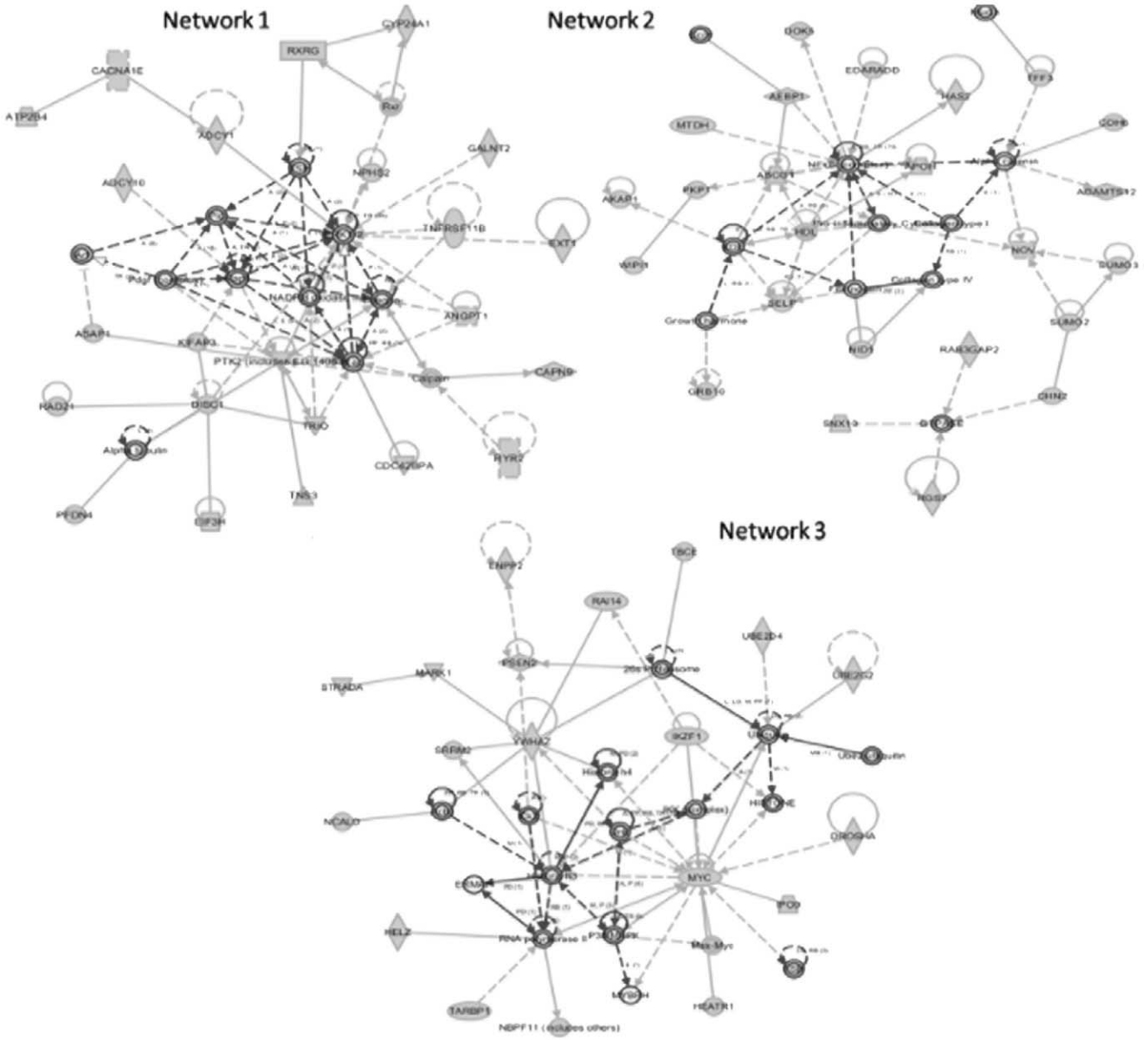


Figure 4. Common Networks: Cellular Movement, Connective Tissue Development and Function, Cellular Assembly and Organization (Network 1), Cell-To-Cell Signaling and Interaction, Tissue Development, Organismal Injury and Abnormalities (Network 2), Cell Morphology, Cellular Assembly and Organization, Cellular Compromise (Network 3).
doi:10.1371/journal.pone.0043789.g004

Discussion

Two chromosomal regions, 7q33 and 21q22.3 presented more alterations in BQC tumors (gains) than NBQC tumors. Gain of 7q33 region has been previously reported in pancreatic and lung carcinoma [23,24]. Gain of 21q22.3 has previously been described in cholangiocarcinoma and as one of the predictive marker regions of systemic recurrence in breast cancer [25,26]. GO terms, oxidoreductase and aldo-keto reductase activity were enriched with a single drug metabolism, molecular transport, nucleic acid metabolism network. AKR1B1and AKR1B10 genes were seen playing cardinal roles. AKR1B10 is overexpressed in colorectal, uterine, breast cancers. Considered a diagnostic marker in lung cancer, it may play a pathogenic role in hepatocellular carcinoma. Role of AKR1B10 in tobacco-related carcinogenesis is anticipated because of its overexpression observed in bronchial epithelium of smokers. Its expression which is stimulated by tobacco smoke condensate in normal human epidermal, oral and squamous cell carcinoma cells decreases with the cessation of smoking. Proposed AKR1B10-mediated tumorigenic mechanisms include retinoic acid depletion and cancer cell dedifferentiation as well as chemoresistance due to metabolism of carbonyl group–bearing anticancer drugs and activating pro-carcinogens and polycyclic aromatic hydrocarbon (PAH) transdihydrodiols to biologically reactive and redox-active *o*quinones [27,28,29]. Hence, the tobacco component in BQ may explain AKR1B10 gain rendering chemoresistance, dedifferentiation and DNA adduct formation in BQC leading to breast carcinogenesis. In addition, AKR1B1 contributes in regulating multiple inflammatory pathways and its inhibition has been shown to interrupt inflammation triggered by

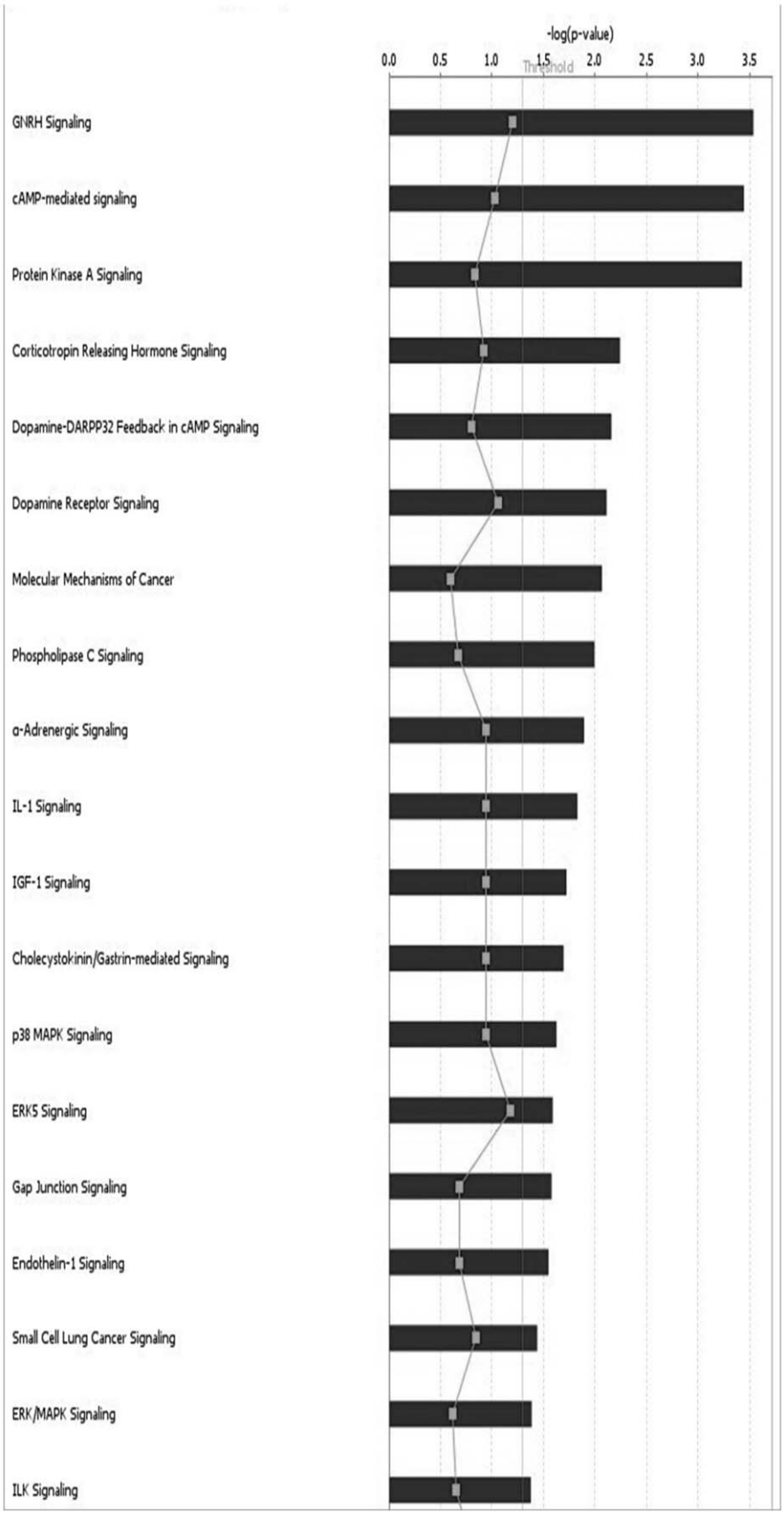


Figure 5. Cancer related Canonical pathways enriched from common genes.
doi:10.1371/journal.pone.0043789.g005

chemokines, growth factors and inflammatory cytokines such as TNF-α as depicted in our network [29]. The ROS generated by the presence of slaked lime in BQ may amplify AKR1B1 gene rendering TNFα induced proliferation of breast cancer cells. Besides, the above network analysis also manifested role of ETS2 gene which maintains hTERT gene expression by interacting with the c-Myc transcription factor. It is a central driver of a transcriptional program in tumor associated macrophages that acts to promote lung metastasis of breast tumors [30,31].

Our data also presented with a high significance of gain in fragile sites in BQC tumors as compared with NBQC. Fragile sites form gaps, constrictions and breaks on chromosomes when exposed to partial replication stress and are rearranged in tumors. Frequency of fragile sites and sister chromatid exchanges have been found to be significantly higher in smokers in peripheral lymphocytes and bone marrow [32,33]. The above ascertains the potential of BQ carcinogens in causing chromosomal damage and instability leading to genetic alterations.

Since metabolic absorption of the ingredients of BQ directs the cancer-causing principles to other organs/tissues of the body, the evidence is growing to indicate that cancers other than oropharyngeal may also be caused by BQ chewing [34]. Tobacco related carcinogens can be stored in breast adipose tissue, metabolized and activated by human mammary epithelial cells [35]. Moreover, the evidence that tobacco exposure (smoking) causes early gene expression changes in normal airway epithelial cells and many other cancer types [36], the aforesaid observed changes likely reflect early carcinogenesis.

Among the NBQC tumors, two regions 19p13.3–19p12 and 20q11.22 presented more alterations. Gain observed in 19p13.3–19p12 has previously been reported in cutaneous and oral squamous cell carcinomas [37,38]. Gain at 20q11.2 has been observed in breast, colorectal and cervical cancers [39,40,41]. GO terms, G-protein coupled receptor protein signaling pathway and cell surface receptor linked signal transduction were enriched with two networks, molecular transport, nucleic acid metabolism, small molecule biochemistry and cellular development, embryonic development, organismal development. RPN2, EMR3, VAV1, NNAT and MUC16 genes were recognized to have imperative functions. A recent study by Kimi Honma *et al* reported that RPN2 silencing and downregulation makes cancer cells hypersensitive in response to docetaxel a chemotherapeutic drug, proposing it as a target for RNA interference–based therapeutics against drug resistance [42]. EMR-3 a G-protein coupled receptor, upregulated in glioblastoma is associated with poor survival and is a potential mediator of cellular invasion [43]. VAV1 contributes to tumorgenesis by regulating both cellular proliferation and cell survival pathways through the regulation of an EGF-Src-Vav1-Rac1-Pak1-NF-κ B-Cyclin D1 signaling axis. An increased and ectopic expression of VAV1 in lung and pancreatic tumors has been linked to large tumors and worse survival rate respectively [44,45]. mRNA expression of neuronatin (NNAT) has been reported in pituitary adenoma, prostatic cancer with neuroendocrine features, large cell neuroendocrine carcinoma lung and thyroid stimulating hormone-producing tumors in mice. High expression has been reported in a tamoxifen-resistant mammary carcinoma cell line [46]. Decreasing MUC16 levels are known to be of prognostic outcome in the post-operative and pre-operative neo-adjuvant chemotherapy especially in ovarian carcinoma [47]. Recent studies by Silke Reinartz *et al* and I Lakshmanan *et al* elucidated its central role in adhesion, migration and invasion in breast cancer. Overexpressed in breast cancer, it augments cell proliferation by interacting with JAK2 and inhibiting the apoptotic process through downregulation of TRAIL [48,49]. Since the NBQC tumors had no previous history of BQ and any other environmental exposure, gain in genes regulating various facets of tumorigenesis can only be blamed as spontaneous instances arising in NBQC tumors.

Besides differences, both tumor groups shared twenty seven frequently altered regions. The IPA analyses resulted in three top networks and eight tumor associated canonical pathways. Extrapolating the data depicted ERK 1/2 and PTK2 (network 1), NFkB complex, SELP and NOV (network 2), MYC and YWHAZ (network 3) were key nodes in their respective networks.

A review by Ming Luo *et al* on PTK2 (FAK) described its principal role in breast carcinogenesis. As depicted in our network 1, PTK2 served as a mediator of cell cycle regulation by integrins through PTK2/Src complex formation in the focal contacts promoting ERK activation. Mechanistic studies indicate that PTK2 deletion in mammary tumor cells reduces the expression/phosphorylation of ERK1/2 contributing to the tumor dormancy in *vivo* and arrests growth in cultures suggesting PTK2 signaling through ERK-MAPK pathway is required to maintain tumor cell growth. In addition to Rac, PTK2 also mediates the activation of ERK to promote cell migration [50].

Network 2 witnessed SELP, NOV and NFκB complex as vital genes. NF-kB plays a key role in regulating the immune response and incorrect regulation of NF-kB has been linked to the development of cancer. Signaling pathways leading to tamoxifen resistance in breast cancer share a common mechanistic link with activation of nuclear factor-kB (NFkB) [51]. Elevated levels of SELP have been observed in many cancers including melanoma, tongue, colon, gastric, lung and breast. SELP is an adhesion molecules that mediate cell-cell interactions among platelets) and endothelial cells. Its measurement may provide a sensitive tool for monitoring the clinical course of melanoma and lymphoma [52]. High expression levels of NOV are associated with endocrine therapy crossresistance in CL6.7 cells and endocrine therapy resistance in breast tumor samples proliferation [53]. NOV enhances migration of chondrosarcoma cells by increasing MMP-13 expression through αvβ3/αvβ5 integrin receptor, FAK, PI3K, Akt, p65, and NF-κB signal transduction pathway and regulates the differentiation of bone resident cells creating a resorptive environment that promotes the formation of osteolytic breast cancer metastases [54,55].

YWHAZ (14-3–3ζ) seen in network 3, overexpressed in breast, lung and many other cancers is implicated in the initiation and progression of cancer [56]. Low level copy number gains in YWHAZ have been found in head and neck squamous cell carcinomas [57]. Previous studies documenting YWHAZ upregulation and a poor clinical outcome in tamoxifen treated breast cancer patients imply it to be a marker of poor prognosis in women with ER-positive breast cancers [58]. The oncogenic Myc protein in network 3 plays an important role in breast cancer metastasis and several transcription factors are involved in the regulation of Myc expression. In breast cancer, amplification of c-*myc* may correlate positively or negatively with alterations in other genes [59]. For e.g. as revealed by our network 3 heterodimerization with Max is necessary for c-Myc to mediate proliferation, transformation, and apoptosis [60]. Recent studies have indicated that Myc is an IKKs substrate and IKKs tightly regulate Myc expression in breast cancers as also seen in network 3 [61].

Alterations seen in the preceding genes can be seen as vital as they arise independently of the etiological factors signifying the abovementioned genes importance in breast tumorigenesis. In addition, direct or indirect association of these key network genes to other cancer related genes (for example, MTDH, EXT1, ANGPT1, RAD21, EDARADD, TFF3, MARK1, DROSHA, etc

seen in our networks) could create a permissive context activating or deactivating various facets of breast tumorigenesis. Super inducing these common alterations, AKR1B10, AKR1B1 and ETS2 alterations were BQ induced whereas alterations in RPN2, EMR3, VAV1, NNAT and MUC16 genes in NBQC tumors could only be termed as spontaneous.

It is important to acknowledge that apart from environmental factor such as betel quid being the prime focus of this study, genetic risk factors such as BRCA1 and BRCA2, lifestyle risk factors such as diet and reproductive risk factors also contribute to breast cancer. BRCA1 and BRCA2 mutant carriers impose a highly increased risk for hereditary or familial breast cancer. While our study is specifically based on sporadic tumors, BRCA2 mutation analysis performed on a larger set of samples in our unpublished study showed none of the tumors to be BRCA2 mutation positive. Therefore, likelihood of our samples containing BRCA1 mutations would still be minute if the probability of BRCA1 and BRCA2 mutations taken together is estimated to be 5%, equal to the proportion in total breast cancer incidence. Examination of impact of BRCA1 and BRCA2 mutations on copy number alteration illustrates a significant difference of genomic profiles between BRCA1 and sporadic tumors, followed by BRCA1 and BRCA2 tumors. BRCA2 and sporadic tumors (such as in our study) had very similar genomic profiles. Overall, BRCA1 tumors have a higher frequency of copy number alterations [62] implying that high risk cases of BRCA1 mutant carriers if subjected to environmental toxicants like betel quid could exemplify the effects resulting in aggressive and early tumors. Furthermore, lifestyle factor like diet has been implicated as an important determinant of breast cancer. The diet pattern in Northeast population of India is mainly characterized by high intakes of dry fish and fermented soybean and vegetables [63]. Such dietary pattern rich in vegetables and fish, but poor in red meats and animal fats has been positively associated to a longer overall survival of breast cancer. However obese women have increased risk for breast cancer as they are exposed to high levels of estrogen additionally produced by adipose tissue [64]. Reproductive factors, including age at menarche, age at first full-term pregnancy, number of live births and breast-feeding are related to a risk of breast cancer. Mechanism through which reproductive exposures influence breast cancer risk is their effect on lifetime number of menstrual cycles. Number of menstrual cycles influences the lifetime exposure to endogenous ovarian hormones like estrogen, which is strongly related to breast cancer risk [65]. Estrogen when metabolized produces metabolites which further contribute to tumor initiation by activating estrogen receptor and generating DNA damaging molecular species [66]. In our unpublished study breast cancer risk was not associated with any of the reproductive factors and polymorphism in an estrogen synthesizing CYP17 gene in the Northeast population of India (3). However, examination of the effect of lifestyle factors and reproductive factors on copy number alteration yet remains to be investigated. The foregoing further ascertains that the effects seen in the present study are due to betel quid chewing.

To our knowledge this is the first report of comparison of genomic alterations between BQ and NBQ chewer breast cancer patients. Overall our data agree well with previous genomic alteration analysis. The major strength of this study is its homogeneous sample population, presence of only BQ as an environmental exposure variables and detailed demographic information. As a limitation, analysis of a larger sample set and cell systems is clearly needed to more precisely delineate the molecular basis for both BQC and NBQC breast tumors. Despite that the accuracy of our results is justified due to unbiased sample distribution in both groups and FDR adjustments. Since composition of BQ in this region consists of multiple components, assessing carcinogenic effect of individual constituent was not possible in this study. Application of high resolution arrays may elicit additional regions of differential alteration. Unfortunately, such studies are largely precluded by the relative rarity of appropriate specimens. However, biological information obtained from BQ exposed breast cancer subset is valuable. This subgroup is frequent in the North East Indian population as most of the women in this area are usually chewers. Given a unique set of underlying genomic changes, distinct approaches to treatment may be appropriate for this patient population and others where this habit is highly prevalent.

Materials and Methods

Patient Recruitment and Sample Collection

Ninety two patients with breast tumors histopathologically confirmed as breast cancer at the Dr. B. Borrooah Cancer Institute, Guwahati and Civil Hospital, Aizwal India between November 2005 and December 2008 were registered for this study. Besides collecting tumor tissues in formalin for histopathology, tumor tissue in RNAlater and 5 ml blood in EDTA vials were collected for copy number analysis. Demographics, including age, sex, menopausal status, BQ history, tobacco history, alcohol drinking, family history and area of residence were obtained for each case. To quantify betel quid chewing we defined a habitual BQ chewer who chewed one betel quid or more daily for no less than ten years. Details of betel quid chewing history for 26 BQC samples are given in supplementary table S6. The ingredients of BQ included areca nut (Areca catechu), catechu (Acacia catechu) and slaked lime (calcium oxide and calcium hydroxide) wrapped in a betel leaf (Piper betle) and tobacco. Thirty two patients with locally advanced breast cancer were given neoadjuvant chemotherapy therefore were excluded. DNA was extracted from the fresh frozen tumor tissue and blood. Specimens with lower than 70% cancer cellularity, inadequate DNA concentration (<50 ng/mL), or a smearing pattern in gel electrophoresis were not included for genotyping. On this basis 43 cases of breast cancer cases were selected and analyzed for copy number assessment which included 26 BQC with only BQ chewing history and 17 NBQC with no history of tobacco chewing, tobacco smoking and alcohol consumption. All 43 cases were morphologically infilterating ductal carcinoma, not otherwise specific. Control germline DNA extracted from blood lymphocytes was used from age matched 14 breast cancer patients. All samples were collected with the patient's written informed consent and the study was approved by the institutional ethics committee of Regional Medical Research Centre, North East Region (Indian Council of Medical Research).

Single Nucleotide Polymorphism Array

Genechip Mapping 10 K early access array analysis The Single Primer Assay Protocol (labeling, hybridization, washing, staining and scanning) was performed according to the manufacturer's instructions (Affymetrix, Santa Clara, CA, USA).

Data Analysis

The primary experimental data was normalized to a baseline array with median signal intensity by applying invariant set normalization method. Copy number change was measured based on comparing the signal intensities at each probe locus between control and tumor samples by applying the hidden Markov Model using the dChip software, with a sliding window of 3 SNPs. Copy

number gain was defined as >2.8 copies and loss was defined as less than 1.2 copies in at least 3 consecutive SNPs [67]. Recurrent altered regions were identified as regions with gain or loss in ≥3SNPs in not less than 15% of samples [68]. Mapping information of SNP locations and cytogenetic band were based on curation of Affymetrix and University of California Santa Cruz hg 17 (http://genome.ucsc.edu). To identify exposure-related aberrations, the data from individual patients were analyzed at group level by comparing gene copy number ratios of the tumors of chewer and nonchewer patients. In each region, we considered a 3×2 contingency table, with the rows representing number of patients with copy number gain, copy number loss or normal copy number in that region and the column representing BQC and NBQC breast cancer patients. Significant regions ($p<0.05$) were identified by comparing the copy number changes in the 26 BQC versus 17 NBQC breast cancer patients using a Fisher's Exact Test based on the 3×2 table in each region. FDR was calculated using Benjamini and Hoeschbergs using the Q value software in R package [69].

Gene Ontology (GO), Pathway and Network Analyses

Functional annotation analysis was performed using the DAVID (Database for Annotation, Visualization and Integrated Discovery) Functional Annotation Tool and Database [70]. A modified, more conservative Fisher's exact p-value, or EASE score, is used to determine if there is a significant level of enrichment in the gene set. To determine pathways and networks those were significantly enriched in the two groups we performed pathway analysis using the Ingenuity Pathway Analysis (IPA) program (http://www.ingenuity.com).

Supporting Information

Figure S1 Prevalence (%) of patients with ≥3 copies (red) and ≤1 copies (blue) in BQC and NBQC tumors, respectively. The x-axis represents the positions in genome/chromosomes, and the y-axis represents the prevalence.
(TIF)

Table S1 Patient and tumor characteristics in relation to betel quid chewing.
(DOC)

Table S2 Total 110 regions seen to be altered in overall samples.
(XLS)

Table S3 Chromosomal gains and deletions in breast tumors from 26 BQC and 17 NBQC. High-level amplifications are in boldface.
(XLS)

Table S4 Regions with chromosomal alterations frequent in betel quid chewers and non chewers breast cancer patients.
(XLS)

Table S5 DAVID analysis of genes in BQC, NBQC and Common regions.
(XLS)

Table S6 Details of betel quid chewing history for 26 BQC samples.
(DOC)

Author Contributions

Conceived and designed the experiments: MK SK SS. Performed the experiments: MK. Analyzed the data: MK AKM. Contributed reagents/materials/analysis tools: SK SS. Wrote the paper: MK. Performed pathological review: JDS EZ AK.

References

1. Tchatchou S, Burwinkel B (2008) Chromosome copy number variation and breast cancer risk. Cytogenet Genome Res 123: 183–187.
2. IARC website, Available at http://globocan.iarc.fr accessed on 31 January 2012.
3. Kaushal M, Mishra AK, Raju BS, Ihsan R, Chakraborty A, et al. (2010) Betel quid chewing as an environmental risk factor for breast cancer. Mutat Res 703: 143–148.
4. NCRPINDIA (2006) website, Available: http://ncrpindia.org/Three-Year Reports of Population Based Cancer Registries 2006–2008. Accessed: 31 January 2012.
5. Nair U, Bartsch H, Nair J (2004) Alert for an epidemic of oral cancer due to use of the betel quid substitutes gutkha and pan masala: a review of agents and causative mechanisms. Mutagenesis 19: 251–262.
6. Lin CF, Wang JD, Chen PH, Chang SJ, Yang YH, et al. (2006) Predictors of betel quid chewing behavior and cessation patterns in Taiwan aborigines. BMC Public Health 6: 271.
7. IARC (2012) Available: http://monographs.iarc.fr/ENG/Monographs/vol85/index.php Accessed: 31 January 2012.
8. Chiang WF, Hung PS, Liu SY, Yuan TC, Chang KW, et al. (2011) Increase of ZASC1 gene copy number in recurrent oral carcinoma. Oral Dis 17: 53–59.
9. Liu SY, Liu YC, Huang WT, Huang GC, Chen TC, et al. (2007) Up-regulation of matrix metalloproteinase-8 by betel quid extract and arecoline and its role in 2D motility. Oral Oncol 43: 1026–1033.
10. Wiseman H, Halliwell B (1996) Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. Biochem J 313 (Pt 1): 17–29.
11. Narayan S, Jaiswal AS, Kang D, Srivastava P, Das GM, et al. (2004) Cigarette smoke condensate-induced transformation of normal human breast epithelial cells in vitro. Oncogene 23: 5880–5889.
12. Luo LZ, Werner KM, Gollin SM, Saunders WS (2004) Cigarette smoke induces anaphase bridges and genomic imbalances in normal cells. Mutat Res 554: 375–385.
13. Baras A, Yu Y, Filtz M, Kim B, Moskaluk CA (2009) Combined genomic and gene expression microarray profiling identifies ECOP as an upregulated gene in squamous cell carcinomas independent of DNA amplification. Oncogene 28: 2919–2924.
14. Herzog CR, Desai D, Amin S (2006) Array CGH analysis reveals chromosomal aberrations in mouse lung adenocarcinomas induced by the human lung carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. Biochem Biophys Res Commun 341: 856–863.
15. Singh B, Wreesmann VB, Pfister D, Poluri A, Shaha AR, et al. (2002) Chromosomal aberrations in patients with head and neck squamous cell carcinoma do not vary based on severity of tobacco/alcohol exposure. BMC Genet 3: 22.
16. Gu J, Horikawa Y, Chen M, Dinney CP, Wu X (2008) Benzo(a)pyrene diol epoxide-induced chromosome 9p21 aberrations are associated with increased risk of bladder cancer. Cancer Epidemiol Biomarkers Prev 17: 2445–2450.
17. Korenaga Y, Matsuyama H, Hirata H, Nagao K, Ohmi C, et al. (2005) Smoking may cause genetic alterations at 5q22.2 approximately q23.1 in clear-cell renal cell carcinoma. Cancer Genet Cytogenet 163: 7–11.
18. Loo LW, Grove DI, Williams EM, Neal CL, Cousens LA, et al. (2004) Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. Cancer Res 64: 8541–8549.
19. Fang M, Toher J, Morgan M, Davison J, Tannenbaum S, et al. (2011) Genomic differences between estrogen receptor (ER)-positive and ER-negative human breast carcinoma identified by single nucleotide polymorphism array comparative genome hybridization analysis. Cancer 117(10): 2024–34.
20. Jonsson G, Naylor TL, Vallon-Christersson J, Staaf J, Huang J, et al. (2005) Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res 65: 7612–7621.
21. Saama PM, Patel OV, Bettegowda A, Ireland JJ, Smith GW (2006) Novel algorithm for transcriptome analysis. Physiol Genomics 28: 62–66.
22. Pang H, Hauser M, Minvielle S (2011) Pathway-based identification of SNPs predictive of survival. Eur J Hum Genet 19: 704–709.
23. Nowak NJ, Gaile D, Conroy JM, McQuaid D, Cowell J, et al. (2005) Genome-wide aberrations in pancreatic adenocarcinoma. Cancer Genet Cytogenet 161: 36–50.

24. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, et al. (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. Science 316: 1039–1043.
25. Hwang KT, Han W, Cho J, Lee JW, Ko E, et al. (2008) Genomic copy number alterations as predictive markers of systemic recurrence in breast cancer. Int J Cancer 123: 1807–1815.
26. Muenphon K, Limpaiboon T, Jearanaikoon P, Pairojkul C, Sripa B, et al. (2006) Amplification of chromosome 21q22.3 harboring trefoil factor family genes in liver fluke related cholangiocarcinoma is associated with poor prognosis. World J Gastroenterol 12: 4143–4148.
27. Liu J, Wen G, Cao D (2009) Aldo-keto reductase family 1 member B1 inhibitors: old drugs with new perspectives. Recent Pat Anticancer Drug Discov 4: 246–253.
28. Barski OA, Tipparaju SM, Bhatnagar A (2008) The aldo-keto reductase superfamily and its role in drug metabolism and detoxification. Drug Metab Rev 40: 553–624.
29. Diez-Dacal B, Gayarre J, Gharbi S, Timms JF, Coderch C, et al. (2011) Identification of aldo-keto reductase AKR1B10 as a selective target for modification and inhibition by prostaglandin A(1): implications for antitumoral activity. Cancer Res 71: 4161–4171.
30. Xu D, Dwyer J, Li H, Duan W, Liu JP (2008) Ets2 maintains hTERT gene expression and breast cancer cell proliferation by interacting with c-Myc. J Biol Chem 283: 23567–23580.
31. Zabuawala T, Taffany DA, Sharma SM, Merchant A, Adair B, et al. (2010) An ets2-driven transcriptional program in tumor-associated macrophages promotes tumor metastasis. Cancer Res 70: 1323–1333.
32. Ban S, Cologne JB, Neriishi K (1995) Effect of radiation and cigarette smoking on expression of FUdR-inducible common fragile sites in human peripheral lymphocytes. Mutat Res 334: 197–203.
33. Kao-Shan CS, Fine RL, Whang-Peng J, Lee EC, Chabner BA (1987) Increased fragile sites and sister chromatid exchanges in bone marrow and peripheral blood of young cigarette smokers. Cancer Res 47: 6278–6282.
34. Chatterjee A, Deb S (1999) Genotoxic effect of arecoline given either by the peritoneal or oral route in murine bone marrow cells and the influence of N-acetylcysteine. Cancer Lett 139: 23–31.
35. Terry PD, Rohan TE (2002) Cigarette smoking and the risk of breast cancer in women: a review of the literature. Cancer Epidemiol Biomarkers Prev 11: 953–971.
36. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, et al. (2009) MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proc Natl Acad Sci U S A 106: 2319–2324.
37. Ambatipudi S, Gerstung M, Gowda R, Pai P, Borges AM, et al.(2011) Genomic profiling of advanced-stage oral cancers reveals chromosome 11q alterations as markers of poor clinical outcome. PLoS One 6: e17250.
38. Purdie KJ, Lambert SR, Teh MT, Chaplin T, Molloy G, et al. (2007) Allelic imbalances and microdeletions affecting the PTPRD gene in cutaneous squamous cell carcinomas detected using single nucleotide polymorphism microarray analysis. Genes Chromosomes Cancer 46: 661–669.
39. Scotto L, Narayan G, Nandula SV, Arias-Pulido H, Subramaniyam S, et al. (2008) Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. Genes Chromosomes Cancer 47: 755–765.
40. Hodgson JG, Chin K, Collins C, Gray JW (2003) Genome amplification of chromosome 20 in breast cancer. Breast Cancer Res Treat 78: 337–345.
41. Nakao K, Mehta KR, Fridlyand J, Moore DH, Jain AN, et al. (2004) High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. Carcinogenesis 25: 1345–1357.
42. Honma K, Iwao-Koizumi K, Takeshita F, Yamamoto Y, Yoshida T, et al. (2008) RPN2 gene confers docetaxel resistance in breast cancer. Nat Med 14: 939–948.
43. Kane AJ, Sughrue ME, Rutkowski MJ, Phillips JJ, Parsa AT. (2010) EMR-3: a potential mediator of invasive phenotypic variation in glioblastoma and novel therapeutic target. Neuroreport 21: 1018–1022.
44. Fernandez-Zapico ME, Gonzalez-Paz NC, Weiss E, Savoy DN, Molina JR, et al. (2005) Ectopic expression of VAV1 reveals an unexpected role in pancreatic cancer tumorigenesis. Cancer Cell 7: 39–49.
45. Lazer G, Idelchuk Y, Schapira V, Pikarsky E, Katzav S (2009) The haematopoietic specific signal transducer Vav1 is aberrantly expressed in lung cancer and plays a role in tumourigenesis. J Pathol 219: 25–34.
46. Uchihara T, Okubo C, Tanaka R, Minami Y, Inadome Y, et al. (2007) Neuronatin expression and its clinicopathological significance in pulmonary non-small cell carcinoma. J Thorac Oncol 2: 796–801.
47. Vasudev NS, Trigonis I, Cairns DA, Hall GD, Jackson DP, et al. (2011) The prognostic and predictive value of CA-125 regression during neoadjuvant chemotherapy for advanced ovarian or primary peritoneal carcinoma. Arch Gynecol Obstet 284: 221–227.
48. Lakshmanan I, Ponnusamy MP, Das S, Chakraborty S, Haridas D, et al. (2012) MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells. Oncogene 31(7): 805–17.
49. Reinartz S, Failer S, Schuell T, Wagner U (2012) CA125 (MUC16) gene silencing suppresses growth properties of ovarian and breast cancer cells. Eur J Cancer 48(10): 1558–69.
50. Luo M, Guan JL (2010) Focal adhesion kinase: a prominent determinant in breast cancer initiation, progression and metastasis. Cancer Lett 289: 127–139.
51. Zhou Y, Eppenberger-Castori S, Eppenberger U, Benz CC (2005) The NFkappaB pathway and endocrine-resistant breast cancer. Endocr Relat Cancer 12 Suppl 1: S37–46.
52. Ferroni P, Roselli M, Martini F, D'Alessandro R, Mariotti S, et al. (2004) Prognostic value of soluble P-selectin levels in colorectal cancer. Int J Cancer 111: 404–408.
53. Ghayad SE, Vendrell JA, Bieche I, Spyratos F, Dumontet C, et al. (2009) Identification of TACC1, NOV, and PTTG1 as new candidate genes associated with endocrine therapy resistance in breast cancer. J Mol Endocrinol 42: 87–103.
54. Tzeng HE, Chen JC, Tsai CH, Kuo CC, Hsu HC, et al. (2011) CCN3 increases cell motility and MMP-13 expression in human chondrosarcoma through integrin-dependent pathway. J Cell Physiol 226: 3181–3189.
55. Ouellet V, Tiedemann K, Mourskaia A, Fong JE, Tran-Thanh D, et al. (2011) CCN3 impairs osteoblast and stimulates osteoclast differentiation to favor breast cancer metastasis to bone. Am J Pathol 178: 2377–2388.
56. Neal CL, Yu D (2010)14-3-3zeta as a prognostic marker and therapeutic target for cancer. Expert Opin Ther Targets 14: 1343–1354.
57. Lin M, Morrison CD, Jones S, Mohamed N, Bacher J, et al. (2009) Copy number gain and oncogenic activity of YWHAZ/14–3-3zeta in head and neck squamous cell carcinoma. Int J Cancer 125: 603–611.
58. Bergamaschi A, Katzenellenbogen BS (2012) Tamoxifen downregulation of miR-451 increases 14–3-3zeta and promotes breast cancer cell survival and endocrine resistance. Oncogene 31: 39–47.
59. Liao DJ, Dickson RB (2000) c-Myc in breast cancer. Endocr Relat Cancer 7: 143–164.
60. Sakamuro D, Prendergast GC (1999) New Myc-interacting proteins: a second Myc network emerges. Oncogene 18: 2942–2954.
61. Yeh PY, Lu YS, Ou DL, Cheng AL (2011) IkappaB kinases increase Myc protein stability and enhance progression of breast cancer cells. Mol Cancer 10: 53.
62. Jonsson G, Naylor TL, Vallon-Christersson J, Staaf J, Huang J, et al. (2005) Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res 65: 7612–7621.
63. Das AJ and Deka SC. (2012) Fermented foods and beverages of the North-East India. International Food Research Journal 19(2): 377–392.
64. Dal Maso L, Zucchetto A, Talamini R, Serraino D, Stocco CF, et al. (2008) Effect of obesity and other lifestyle factors on mortality in women with breast cancer. Int J Cancer 123: 2188–2194.
65. Beaber EF, Holt VL, Malone KE, Porter PL, Daling JR, et al. (2008) Reproductive factors, age at maximum height, and risk of three histologic types of breast cancer. Cancer Epidemiol Biomarkers Prev 17: 3427–3434.
66. Sangrajrang S, Sato Y, Sakamoto H, Ohnami S, Laird NM, et al. (2009) Genetic polymorphisms of estrogen metabolizing enzyme and breast cancer risk in Thai women. Int J Cancer 125: 837–843.
67. George RE, Attiyeh EF, Li S, Moreau LA, Neuberg D, et al. (2007) Genome-wide analysis of neuroblastomas using high-density single nucleotide polymorphism arrays. PLoS One 2: e255.
68. Turner N, Lambros MB, Horlings HM, Pearson A, Sharpe R, et al.(2010) Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. Oncogene 29: 2013–2023.
69. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125: 279–284.
70. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57.

Mutation Research 703 (2010) 143–148

Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Betel quid chewing as an environmental risk factor for breast cancer

Mishi Kaushal[a], Ashwani K. Mishra[a], B.S. Raju[a], Rakhshan Ihsan[a], Anurupa Chakraborty[a], Jaganath Sharma[b], Eric Zomawia[c], Yogesh Verma[d], Amal Kataki[b], Sujala Kapur[a], Sunita Saxena[a,*]

[a] Institute of Pathology, ICMR, Safdarjung Hospital Campus, New Delhi 110029, India
[b] Dr. B. Borrooah Cancer Institute, Guwahati, India
[c] Civil Hospital, Aizwal, India
[d] Sir Thutab Namgyal Memorial Hospital, Gangtok, Sikkim, India

ARTICLE INFO

Article history:
Received 20 May 2010
Received in revised form 29 July 2010
Accepted 15 August 2010
Available online 20 August 2010

Keywords:
Breast cancer
Betel quid
GST
Polymorphisms
MDR

ABSTRACT

Northeast region of India shows high incidence of tobacco-related cancer with widespread consumption of betel quid and tobacco in different forms. There is an increasing incidence of breast cancer and eminent use of tobacco in females in this region. Thus, we analysed the role of tobacco exposure and polymorphisms in detoxification enzymes in breast cancer risk. Polymorphisms in five gene variants (GSTT1, GSTM1, GSTP1, TP53 and CYP17) and four environmental exposure variables (tobacco smoking, tobacco chewing, betel quid chewing, alcohol) were analysed in 117 breast cancer cases and 174 cancer free controls. Multifactor dimensionality reduction identified betel quid chewing as the single main risk factor and women with betel quid chewing history had five times the risk of developing breast cancer [4.78 (2.87–8.00) 0.001]. In logistic regression analysis, GSTT1 null and GSTM1 null genotypes conferred 41% less [0.59 (0.34–1.03) 0.06] and 55% less [0.58 (0.30–1.02) 0.05] reduced risk to breast cancer, respectively. However, the risk increased in women with GSTP1 variant G allele which conferred 1.43 times [(0.96–2.11) 0.07] more risk to breast cancer. In conclusion this study suggests betel quid chewing as a significant risk factor for developing breast cancer. Moreover, the lack of detoxification enzymes GSTT1 and GSTM1 are associated with reduced breast cancer risk.

1. Introduction

Breast cancer is the commonest cause of cancer death among women worldwide [1]. The several fold difference in incidence rates between different geographical regions suggest that environmental factors influence breast cancer risk significantly. Among the identified environmental risk factors in general for cancers, tobacco exposure has been reported as the leading preventable risk factor [2]. In India breast cancer is the second most common cancer among women [3]. The Northeast districts of India have the highest incidence of cancers associated with both smoking and smokeless tobacco [4]. The Northeast Indian population is estimated to be at high risks for oesophageal, gastric and oral cancers due to its high incidence of tobacco consumption [5–7]. The mean age for tobacco use initiation in this region is 18.5 years and the prevalence of tobacco use is estimated as 41% that includes a large number of female chewers too apart from male smokers [8]. Method and form of tobacco consumption in this region is reported to be different from rest of the India. Also, this region reports high risk for developing oesophageal, gastric cancer with betel quid chewing, another form for tobacco consumption [5]. Betel quid is a combination of betel leaf, areca nut slaked lime and tobacco. There is a great spectrum of ingredients and ways of preparing betel quid which differs with different geographical region. Studies from Pakistan and Mainland China also report betel quid chewing as the major aetiological factor for oral leukoplakia and oral submucous fibrosis. It is also commonly used in South and Southeast Asia and Asia Pacific [9].

Epidemiological perspective suggests an increased risk associated with exposure to genotoxic agents during breast development, as the undifferentiated ductal elements of the breast are more susceptible to the action of genotoxins early in life [10]. Environmental genotoxic stress like tobacco smoke and smokeless tobacco contain polycyclic aromatic hydrocarbons (PAHs), tobacco-specific nitrosamines, nitrosamino acids, aldehydes, metals, aromatic and heterocyclic amines and other genotoxic carcinogens [11,12]. The concomitant use of betel quid also leads to a 50-fold increase in reactive oxygen species generated [13].

In keeping with the polygene hypothesis of breast cancer [14], the genes responsible for metabolizing the tobacco carcinogens appears to be prime candidates for the investigative search of breast cancer susceptibility genes. As the northeast region has a

* Corresponding author. Tel.: +91 11 26198402/26198405; fax: +91 11 26198401.
E-mail address: sunita_saxena@yahoo.com (S. Saxena).

1383-5718/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrgentox.2010.08.011

Table 1
Selected epidemiological variables and their risk association in our study population.

Variable	Categories	*N*	Cases (117) (%)	Controls (174) (%)	χ^2 *p* value	Logistic regression analysis OR (95% CI)
Tobacco smoking	No	240	95 (81.2)	145 (83.3)	0.63	1
	Yes	51	22 (18.8)	29 (16.7)		1.15 (0.62–2.13)
Tobacco chewing	No	168	67 (57.3)	101 (58)	0.89	1
	Yes	123	50 (42.7)	73 (42)		1.03 (0.64–1.65)
Betel quid chewing	No	144	32 (27.4)	112 (64.4)	0.001	1
	Yes	147	85 (72.6)	62 (35.6)		4.78 (2.87–8.00)
Alcohol consumption	No	276	108 (92.3)	168 (96.6)	0.10	1
	Yes	15	9 (7.7)	6 (3.4)		2.33 (0.80–6.74)
Distribution of age	<30	38	13 (11.1)	25 (14.36)	0.59	
	30–39	58	26 (22.2)	32 (18.3)		
	40–49	80	39 (33.3)	41 (23.5)		
	50–59	63	18 (15.3)	45 (25.8)		
	>60	52	21 (17.9)	31 (17.8)		
Family history	Yes		83 (70.9)	69 (39.6)		
	No		14 (11.9)	32 (18.3)		
	Missing		20 (17.0)	73 (41.9)		

very high and typical usage of tobacco we selected genes related to catabolism and detoxification of xenobiotics (GSTM1, GSTT1, and GSTP1), tumour suppressor gene (TP53) and oestrogen biosynthesis (CYP17), to explore their contribution for breast cancer.

The existing studies in Indian setting on the above genes have adopted a case control study design and applied regression approach to estimate the risk for a particular genotype and environmental factors [15–17]. However the interaction between gene–gene and gene–environment factors in such study settings can be enormous and lead to a biased estimate of the regression coefficients. Under such a situation the regression approach are not designed to test the high-order interaction and therefore one need to employ advanced methodologies such as multifactor dimensionality reduction (MDR) method for estimating the risk of cancer accredited to such interactions.

We present single factor and multifactorial analyses of high-order gene–gene and gene–environment interactions, and discuss the findings.

2. Methods

2.1. Patient recruitment and sample collection

The study was conducted as multicentric study between Institute of Pathology, Indian Council of Medical Research (ICMR) and Population Based Cancer Registries (PBCRs) of Northeast India, by adopting a case control study design after taking approval from the Ethical Committee and the Institutional Review Board of the Hospital. Out of all the incident breast cancer cases only 117 cases agreed to participate in the present study and answered the questionnaire. All suspected cases of cancer of the breast were directed to the social investigator(s) of the project for interview before referral to the medical consultant. Only cases confirmed by microscopy and for whom the breast was the primary site of cancer were included in the study. At the same time information was collected from the attendants who accompanied cancer patients and who provided a readily available and cooperative source of controls from the same socio-economic background as the patients. A final group of matched controls were selected by random pairing of the cases with subjects from the pool of controls after matching for sex and age (within ±5 years). The study included 117 cases and 174 controls between November 2005 and December 2008.

All subjects including cases and controls were resident of the north-eastern part of India at the time of recruitment for the past 5 years and belonged to the same ethnicity. Subjects were excluded if, according to self-report, they had other cancers. All of the 117 cases were of infiltrating ductal type of breast carcinoma.

Details of age and sex and various demographic variables were collected in the course of the interviews as well as details of personal habits that included tobacco smoking and the consumption of alcohol as well as chewing practices. A pre-designed, pretested questionnaire was designed specifically for the study. The selection of controls from among the persons bringing the patients to hospital is likely to have minimised differences of socio-economic conditions and also of adequacy of nutrition between the patients and controls and these have not been investigated further.

2.2. Genotyping assays

5-ml venous blood sample was collected from each participant in EDTA coated vials. The blood was stored at −20 °C and was transported to Institute where the study was performed. Genomic DNA was isolated by standard phenol chloroform method. Genotyping for *GSTT1*, *GSTM1*, *GSTP1*, *TP53* and *CYP17* polymorphisms in 117 Northeast Indian breast cancer cases and 174 matched controls was performed. Genotyping of *GSTM1* and *GSTT1* deletion polymorphisms was carried out by multiplex polymerase chain reaction with three pairs of primers [18]. *GSTP1 Ile*^{105} *Val*, *CYP17 T > C* and *TP53* codon 72 polymorphism was analysed by PCR-restriction fragment length polymorphism (PCR-RFLP) [19–21]. In addition, about 10% random samples were rechecked by the same method (multiplex PCR or PCR-RFLP).

2.3. Statistical analysis

A χ^2 test was used to assess whether the genotypes were in Hardy–Weinberg equilibrium (HWE) among case and control subjects. Odds ratios and their corresponding 95% confidence intervals were calculated by both unadjusted and adjusted logistic regression analysis as a measure for association with the risk for the genes and environmental factors considered. A two side $p \leq 0.05$ was considered as statistically significant. Data for family history was missing for 17% of cases and 42% of controls hence family history was not incorporated in the analysis. The gene–gene and gene–environment interaction was examined using the multifactor dimensionality reduction (MDR) method.

With MDR, genotype and environmental factors were pooled into high- and low-risk groups, effectively reducing the multifactor predictors from multidimensions to single dimension. The new one-dimensional multifactor variable was evaluated for its ability to classify and predict disease status through cross-validation and permutation testing [22]. MDR ultimately selected one genetic model, either single or multilocus, that most successfully predicted phenotype or disease status. The data was then randomly divided into 10 equal parts. A training set of 9/10 of the data was used to search for the best model. The remaining 1/10 of the data is the testing set. Here, we also used 10-fold cross-validation, and the analysis was repeated 10 times with different random seeds to reduce the possibility of biased results due to the chance divisions of the data into training and testing sets [23]. Finally, all the variables in the best model were combined and dichotomized according to the MDR software and their ORs and 95% CIs in relation to breast cancer risk were calculated in logistic regression models.

3. Results

The mean age was 45.5 ± 12.86 years for the cases and 45.98 ± 14.44 years for the controls. There were no significant differences between case and control subjects in terms of distributions of tobacco smoking ($p = 0.63$), tobacco chewing ($p = 0.89$), alcohol consumption ($p = 0.10$) and they were not found to be associated with breast cancer risk. However, women with a betel quid chewing history had around five times the risk of developing breast cancer [4.78 (2.87–8.00) 0.001] (Table 1).

The genotypic distribution of the genetic markers under study was found to be in HWE both in cases and controls (Table 2 gives the *p* values). The *GSTP1*, *CYP17* and *TP53* genes were not associated with breast cancer risk when adjusted for age, tobacco smoking,

Table 2
Association between genotypes, alleles of GSTT1, GSTM1, GSTP1, CYP17 and TP53 polymorphisms and the risk of breast cancer.

	Case (117)		Control (174)		Logistic regression analysis				χ^2, df, p
	Count (%)	Allele probabilities (p value HWE)	Count (%)	Allele probabilities (p value HWE)	Unadjusted OR (95% CI)	p	Adjusted OR (95% CI)	p	
GSTM1									
Present	94 (80.3)		122 (70.1)		1.00		1.00		3.82, 1, 0.05
Null	23 (19.7)		52 (29.9)		0.57 (0.32–1.00)	0.05	0.55 (0.30–1.02)	0.05	
GSTT1									
Present	84 (71.8)		105 (60.3)		1.00		1.00		4.02, 1, 0.04
Null	33 (28.2)		69 (39.7)		0.59 (0.36–0.99)	0.04	0.59 (0.34–1.03)	0.06	
GSTP1									
AA	62 (53.0)	A: 0.73, G: 0.26 (0.56)	108 (62.1)	A: 0.79, G: 0.20 (0.15)	1.00		1.00		4.03, 2, 0.13
AG	48 (41.0)		62 (35.6)		1.34 (0.82–2.20)	0.23	1.26 (0.74–2.16)	0.38	
GG	7 (6.0)		4 (2.3)		3.04 (0.85–10.82)	0.08	2.01 (0.53–7.66)	0.30	
AA + AG					0.37 (0.10–1.29)	0.11	0.54 (0.14–2.03)	0.36	
AG + GG					2.70 (0.77–9.45)	0.11	1.83 (0.49–6.87)	0.36	
A	172		278		1.000				
G	62		70		1.43 (0.96–2.11)	0.07			
CYP17									3.12, 2, 0.21
A1A1	33 (28.2)	A1: 0.52, A2: 0.47 (0.95)	44 (25.3)	A1: 0.53, A2: 0.46 (0.08)	1.00		1.00		
A1A2	58 (49.6)		98 (56.3)		0.78 (0.45–1.37)	0.40	.072 (.39–1.32)	0.29	
A2A2	26 (22.2)		32 (18.3)		1.08 (0.54–2.15)	0.81	1.23 (0.58–2.63)	0.57	
A1	124		186		1.				
A2	110		162		1.01 (0.73–1.41)	0.91			
A1A1 + A1A2					0.78 (0.44–1.41)	0.42	0.652 (0.34–1.23)	0.19	
TP53									
AA	28 (23.9)	A: 0.47, P: 0.52 (0.42)	38 (21.8)	A: 0.47, P: 0.52 (0.84)	1.00		1.00		0.54, 2, 0.76
AP	54 (46.2)		88 (50.6)		0.83 (0.46–1.50)	0.54	0.72 (0.37–1.39)	0.33	
PP	35 (29.9)		48 (27.6)		0.99 (0.51–1.90)	0.97	0.84 (0.41–1.73)	0.65	
AA + AP					0.893 (0.53–1.49)	0.66	0.94 (0.54–1.65)	0.85	
AP + PP					1.12 (0.66–1.87)	0.66	1.05 (0.60–1.84)	0.85	
A	110		164		1.00				
P	124		184		1.00 (0.72–1.40)	0.97			

Fig. 1. Interaction dendrogram for the breast cancer dataset: graphical representation of interactions between nine attributes (GST1 (T1), GSTM1 (M1), GSTP1, CYP17, TP53 (p53), tobacco smoking (smk), tobacco chewing (chw), betel quid chewers (bet) and alcohol consumption (alc)) from the multifactor dimensionality reduction analysis using an 'interaction dendrogram'. Synergy, the interaction between two attributes provides more information than the sum of the individual attributes; redundancy, the interaction between attributes provides redundant information.

tobacco chewing, betel quid chewing, alcohol consumption and tobacco exposure. However, women with GSTT1 null polymorphism were 41% less susceptible [0.59 (0.34–1.03) 0.06] for having breast cancer. Women with GSTM1 null polymorphism were also 55% less susceptible [0.55 (0.30–1.02) 0.05] for having breast cancer.

A marginally significant risk was observed among women having G/G genotype of *GSTP1* gene [3.04 (0.85–10.82) 0.08], but when adjusted for the exposure variables significance was lost. Allele frequencies of different alleles of *GSTP1*, *CYP17* and *TP53* genes were also compared. The G allele of the *GSTP1* gene was found to be over represented in cases as compared to the controls indicating that G allele might be a risk factor for breast cancer [1.43 (0.96–2.11) 0.07]. The frequencies of alleles of other genes did not differ significantly.

MDR analysis revealed, betel quid chewing to be the single factor imparting the main effect [testing accuracy of 0.6851 and cross-validation consistency 10/10, $p = 0.05$]. The combination of betel quid chewing × alcohol consumption [testing accuracy 0.6808 and CVC of 10/10 ($p = 0.05$)] and GSTP1 × betel quid chewing [testing accuracy of 0.6851 and CVC of 10/10 ($p = 0.05$)] were seen as the best two factor interaction models. The MDR analysis gave a three factor interaction model which added tobacco smoking and tobacco chewing to betel quid chewing increasing the test accuracy to 0.7193 and CVC of 10/10. Four way (GSTP1 × tobacco smoking × tobacco chewing × betel quid chewing) interaction model was also identified which showed a lower cross-validation consistency 9/10 but a significant testing accuracy. The addition of TP53 to the four factor interaction model gave a five factor interaction model (GSTP1 × TP53 × smoking × chewing × betel quid chewing) which decreased its testing accuracy to 0.6625 but increased the CVC to 10/10. Among the remaining two five order interaction models found, GSTM1 × TP53 × smoking × chewing × betel quid chewing was found to be significant with a CVC of 10/10 and 0.658 training accuracy. CYP17 was not found to interact with the environmental or the genotypic factors in any of the significant interaction models found, suggesting its minor role towards breast cancer development in this population.

Fig. 1 depicts the interactions between nine attributes from the MDR analysis via a graphical representation of a 'dendrogram'. It shows betel quid chewing, GSTT1 and GSTM1 on a separate branch imparting there independent effects to breast cancer risk.

4. Discussion

There has been increasing interest in the association between tobacco exposure and increased breast cancer risk. The role of smoking in breast cancer aetiology has been extensively studied. Yet, the association remains equivocal and much debated. Smoking has been proposed to increase breast cancer risk, based on studies showing breast epithelial genotoxicity of tobacco-related compounds [24]. Several explanations for the lack of consistency in previous studies have been suggested. Included among these is the possibility that the observed associations are not causal, in which case chance or bias might have driven some of the previous findings in either direction from the null [2]. Tobacco smoke and smokeless tobacco are well known risk factors for pancreatic, bladder and hepatic cancer which occur at sites that are not in direct contact with them [25–28]. Smokeless tobacco has been extensively investigated in both oral and oesophageal cancers. Most of the carcinogenic contents in smokeless and tobacco smoke are similar and it should to be examined in other cancers as well to get a better understanding of the pathogenesis.

Our study provides evidence that betel quid chewing is a very important independent risk factor for breast cancer. Since betel quid chewing has not been shown to be a risk factor for breast cancer earlier, it is important to validate that our finding is not due to confounding bias. The bias may result from the control selection, information bias, or by un-controlling confounding factor. The estimated prevalence of current betel quid chewers reported in control females in the same population was found to be 38% which is almost similar to the prevalence found in our study. Moreover, there was no significant difference in the prevalence rates of habitual alcohol drinking between our controls and those (4% for alcohol) found in another case control study [9]. Based on the information mentioned above, our controls seem be representative for the same population, and make bias unlikely from control selection or under-reporting of life-style habits.

As shown in Table 3, although betel quid chewing is identified as the main risk factor, the interactions with other factors (smoking, tobacco chewing, alcohol consumption and GSTM1, GSTT1, GSTP1, CYP17, TP53) only modified the risk. Interactions conferred insignificant risk to breast cancer on removal of betel quid chewing from MDR analysis (data not shown). This confirms the major contribution of betel quid to breast cancer risk. Betel quid chewers swallow the betel quid juice (saliva extract of betel quid produced by chewing) [29] which gets absorbed from the intestine and pass through the blood stream to various organs like kidney, pancreas and breast. The carcinogens present in it can be stored in breast adipose tissue and then get metabolized and activated by human mammary epithelial cells [2]). Presence of tobacco-related DNA adducts has been demon-

Table 3
Multifactor dimensionality reduction (MDR) models of selected gene and environmental factors.

	Models	Testing balanced accuracy	CVC	Odds ratio (95% CI)[a]	p value of χ^2 test[a]
1 order	Bet	0.6851	10/10	4.79 (0.95, 24.17)	0.05
2 order	bet alc	0.6808	10/10	4.59 (0.92, 22.94)	0.05
2 order	GSTP1 bet	0.6851	10/10	4.79 (0.95, 24.17)	0.05
3 order	smk chw bet	0.7193	10/10	6.8 (1.24, 37.28)	0.02
4 order	GSTP1 smk chw bet	0.7107	9/10	6.10 (1.19, 31.24)	0.02
5 order	GSTP1 TP53 smk chw bet	0.6625	10/10	3.89 (0.81, 18.63	0.08
5 order	M1 p53 smk chw bet	0.6967	10/10	5.28 (1.04, 26.60)	0.03
5 order	T1 p53 smk chw bet	0.658	10/10	3.70 (0.77, 17.70)	0.09

Bet, betel quid chewers; smk, tobacco smoking; chw, tobacco chewing; T1, GSTT1; M1, GSTM1; p53, TP53; CVC, cross-validation consistency.

[a] Risk estimate and χ^2 test were based on the combination and dichotomization of the distribution of genetic factors according to the MDR software.

strated in both breast tumour and adjacent normal tissue [30]. Saliva and urine of betel quid chewers have also shown presence of cancer causing nitrosamines like N′-nitrosonornicotine (1.0–51.7 ng/ml), N′-nitrosoanatabine, 4-(methyl-nitrosamino)-1-(3-pyridyl)-l-butanone (0–2.3 ng/ml), nicotine and cotinine [31]. Animals fed with arecoline, a major constituent of betel quid is shown to develop genotoxcity in the ovary [32]. Animals exposed to nitrosamines, by chronic oral administration and by drinking water, produced during betel quid chewing are shown to develop lung, pancreatic tumours and intestinal metaplasia [33,34]. Moreover carcinogens derived from betel quid chewing are shown to induce p53 mutation and over-expression of c-myc protein with activated ras oncogene and subsequent over-expression of cell cycle regulatory protein, cyclin D1 in oral cancer [28]. However, similar phenomenon induced by betel quid consumption leading to breast carcinogenesis needs elucidation.

GSTs are predominantly involved in the detoxification of xenobiotics and genetic variations in them have been implicated in the etiology of numerous cancers. However, substrate bioactivation reactions by GSTs are known to occur through formation of conjugates which are activated through cysteine conjugate beta-lyase, redox cycling and/or release of the original reactive parent compound [35]. Generally, detoxification by GST leads to formation of less reactive products that are readily excreted. However, in specific tissues and with certain exposures, the products formed are found to be more reactive than the parent compound [36]. For example, detoxification of dichloromethane by GSTT1 enzyme results in biproducts which are found to be carcinogenic in mouse [37]. Toxic metabolites formed during CYP activation of reactive diol-epoxides are not detoxified by the GSTM1 enzyme [38]. Glutathione conjugation of halogenated compounds by GST is known to serve as a substrate for renal cysteine conjugate *b*-lyase which forms reactive chlorothioketenes found to directly damage the tissue. Therefore, an active GST enzyme conjugates the substrate and forms more reactive intermediate that directly damages the tissue. Conversely, the deleted variant GST genotype forms an inactive enzyme, metabolizing the compounds through oxidation, without formation of reactive intermediates [36].

An increase in risk of kidney and liver tumours in humans with the active GSTT1 genotype following exposures to halogenated compounds has been reported [39]. In another study increase risk to renal cell carcinoma was reported to be associated with pesticide exposure exclusive to individuals with active GSTM1/T1 genotypes [36]. Previous studies have also revealed a decrease risk to breast cancer among premenopausal women with the absence of GSTT1 enzyme (null genotype) [40]. Similar findings have been reported in head and neck [41], bladder [42], melanoma [42], and thyroid [43] cancer. Similar mechanisms have been proposed to operate through tobacco carcinogenesis among GSTT1 and GSTM1 positive individuals [44]. In this study a protective role has been observed by the absence of GSTT1 and GSTM1 enzymes for development of breast cancer (Table 2). The presence of the enzymes might have led to the activation of certain known as well as unknown procarcinogens present in the betel quid chewers [13] leading to breast carcinogenesis.

GSTP1 catalysis the conjugating reactions of PAHs and their electrophilic compounds to facilitate their excretion. A polymorphic adenine to guanine transition at nucleotide 313 (*A*313*G*) in exon 5 results in an isoleucine to valine substitution in codon 105 (I105V). This codon is located in the substrate-binding site of GSTP1, and the corresponding allozymes exhibit differential catalytic activities towards diverse substrates [19]. In the present study marginal risk conferred by the variant genotype and its overrepresentation in cases suggests that due to the lower activity of this enzyme the detoxification of the carcinogens was hindered leading to breast carcinogenesis.The most obvious limitation of the present study is its small sample size. However, the sample studied was well characterized and was from a homogenous Northeast Indian population, which reduces the risk of population stratification and false associations. As with all statistical analyses, replication and validity of findings are necessary to separate true relationships from chance findings. One advantage of the MDR method is that false-positive results due to multiple testing are minimized. This is primarily due to the cross-validation strategy used to select optimal models. MDR facilitates the simultaneous detection and characterization of multiple variables associated with a discrete clinical endpoint. This is accomplished by reducing the dimensionality of the multilocus data. Another advantage of MDR is that it can overcome sample size limitation since it is nonparametric and that it assumes no particular genetic model and no mode of inheritance need to be specified. This is important for diseases, such as breast cancer, in which the mode of inheritance is unknown and very complex [45]. However, logistic regression analysis revealed a protective role of GSTT1 and GSTM1 genes which were not observed in the MDR analysis as it only gave the high-order risk estimates.

The incidence of breast cancer in Asia has been steadily increasing over the years [46]. The habit of betel quid chewing is known and has been reported from many Asian countries such as Sri Lanka, Bangladesh, Thailand, Cambodia, Malaysia, Indonesia, and China [9]. These populations report betel quid associated increase in risk for cancers other than breast cancer also [5,47]. This suggests need for investigating the mechanism of betel quid-induced carcinogenesis in breast cancer.

5. Conclusions

In summary, our data provides evidence that betel quid consumption seems to impose strong environmental effects and appears to be an independent risk factor of breast cancer. This case control study also suggests that GSTT1 and GSTM1 enzymes null polymorphism are inversely associated with the risk of breast cancer. The results also demonstrate the need for more epidemiological and genetic studies demonstrating and confirming the role of betel quid in breast cancer.

Conflict of interest statement

The authors declare that they have no competing interests.

Acknowledgement

The authors are thankful to Indian Council of Medical Research for their financial support.

References

[1] T.J. Key, P.K. Verkasalo, E. Banks, Epidemiology of breast cancer, Lancet Oncol. 2 (2001) 133–140.
[2] P.D. Terry, T.E. Rohan, Cigarette smoking and the risk of breast cancer in women: a review of the literature, Cancer Epidemiol. Biomarkers Prev. 11 (2002) 953–971.
[3] R. Chopra, The Indian scene, J. Clin. Oncol. 19 (2001) 106S–111S.
[4] G. Mudur, India has some of the highest cancer rates in the world, BMJ 330 (2005) 215.
[5] R.K. Phukan, M.S. Ali, C.K. Chetia, J. Mahanta, Betel nut and tobacco chewing; potential risk factors of cancer of oesophagus in Assam, India, Br. J. Cancer 85 (2001) 661–667.
[6] A. Bhattacharjee, A. Chakraborty, P. Purkaystha, Prevalence of head and neck cancers in the north east—an institutional study, Indian J. Otolaryngol. Head Neck Surg. 58 (January–March (1)) (2006).
[7] R.K. Phukan, E. Zomawia, K. Narain, N.C. Hazarika, J. Mahanta, Tobacco use and stomach cancer in Mizoram, India, Cancer Epidemiol. Biomarkers Prev. 14 (2005) 1892–1896.
[8] H.K. Chaturvedi, R.K. Phukan, J. Mahanta, The association of selected sociodemographic factors and differences in patterns of substance use: a pilot study in selected areas of Northeast India, Subst. Use Misuse 38 (2003) 1305–1322.

[9] P.C. Gupta, C.S. Ray, Epidemiology of betel quid usage, Ann. Acad. Med. Singapore 33 (2004) 31–36.
[10] J.A. Williams, D.H. Phillips, Mammary expression of xenobiotic metabolizing enzymes and their potential role in breast cancer, Cancer Res. 60 (2000) 4667–4677.
[11] P. Boffetta, S. Hecht, N. Gray, P. Gupta, K. Straif, Smokeless tobacco and cancer, Lancet Oncol. 9 (2008) 667–675.
[12] T.L. Lash, B.D. Bradbury, J.B. Wilk, A. Aschengrau, A case-only analysis of the interaction between N-acetyltransferase 2 haplotypes and tobacco smoke in breast cancer etiology, Breast Cancer Res. 7 (2005) R385–393.
[13] D. Anantharaman, P.M. Chaubal, S. Kannan, R.A. Bhisey, M.B. Mahimkar, Susceptibility to oral cancer by genetic polymorphisms at CYP1A1, GSTM1 and GSTT1 loci among Indians: tobacco exposure as a risk modulator, Carcinogenesis 28 (2007) 1455–1462.
[14] P.D. Pharoah, A.M. Dunning, B.A. Ponder, D.F. Easton, Association studies for finding cancer-susceptibility genetic variants, Nat. Rev. Cancer 4 (2004) 850–860.
[15] M. Samson, R. Swaminathan, R. Rama, V. Sridevi, K.N. Nancy, T. Rajkumar, Role of GSTM1 (null/present), GSTP1 (Ile105Val) and P53 (Arg72Pro) genetic polymorphisms and the risk of breast cancer: a case control study from South India, Asian Pac. J. Cancer Prev. 8 (2007) 253–257.
[16] A. Saxena, V.S. Dhillon, M. Raish, M. Asim, S. Rehman, N.K. Shukla, S.V. Deo, A. Ara, S.A. Husain, Detection and relevance of germline genetic polymorphisms in glutathione S-transferases (GSTs) in breast cancer patients from northern Indian population, Breast Cancer Res. Treat. 115 (2009) 537–543.
[17] P. Chacko, B. Rajan, B.S. Mathew, T. Joseph, M.R. Pillai, CYP17 and SULT1A1 gene polymorphisms in Indian breast cancer, Breast Cancer 11 (2004) 380–388.
[18] E.W. Tiemersma, R.E. Omer, A. Bunschoten, P. van't Veer, F.J. Kok, M.O. Idris, A.M. Kadaru, S.S. Fedail, E. Kampman, Role of genetic polymorphism of glutathione-S-transferase T1 and microsomal epoxide hydrolase in aflatoxin-associated hepatocellular carcinoma, Cancer Epidemiol. Biomarkers Prev. 10 (2001) 785–791.
[19] M. Zhao, R. Lewis, D.R. Gustafson, W.Q. Wen, J.R. Cerhan, W. Zheng, No apparent association of GSTP1 A(313)G polymorphism with breast cancer risk among postmenopausal Iowa women, Cancer Epidemiol. Biomarkers Prev. 10 (2001) 1301–1302.
[20] A. Chakraborty, N.S. Murthy, C. Chintamani, D. Bhatnagar, R.S. Mohil, P.C. Sharma, S. Saxena, CYP17 gene polymorphism and its association with high-risk north Indian breast cancer patients, J. Hum. Genet. 52 (2007) 159–165.
[21] G. Liu, D.P. Miller, W. Zhou, S.W. Thurston, R. Fan, L.L. Xu, T.J. Lynch, J.C. Wain, L. Su, D.C. Christiani, Differential association of the codon 72 p53 and GSTM1 polymorphisms on histological subtype of non-small cell lung carcinoma, Cancer Res. 61 (2001) 8718–8722.
[22] G. Cao, H. Lu, J. Feng, J. Shu, D. Zheng, Y. Hou, Lung cancer risk associated with Thr495Pro polymorphism of GHR in Chinese population, Jpn. J. Clin. Oncol. 38 (2008) 308–316.
[23] M.H. Vaarala, H. Mattila, P. Ohtonen, T.L. Tammela, T.K. Paavonen, J. Schleutker, The interaction of CYP3A5 polymorphisms along the androgen metabolism pathway in prostate cancer, Int. J. Cancer 122 (2008) 2511–2516.
[24] C. Magnusson, S. Wedren, L.U. Rosenberg, Cigarette smoking and breast cancer risk: a population-based study in Sweden, Br. J. Cancer 97 (2007) 1287–1290.
[25] J. Alguacil, D.T. Silverman, Smokeless and other noncigarette tobacco use and pancreatic cancer: a case–control study based on direct interviews, Cancer Epidemiol. Biomarkers Prev. 13 (2004) 55–58.
[26] P. Boffetta, B. Aagnes, E. Weiderpass, A. Andersen, Smokeless tobacco use and risk of cancer of the pancreas and other organs, Int. J. Cancer 114 (2005) 992–995.
[27] P. Brennan, O. Bogillot, E. Greiser, J. Chang-Claude, J. Wahrendorf, S. Cordier, K.H. Jockel, G. Lopez-Abente, A. Tzonou, P. Vineis, F. Donato, M. Hours, C. Serra, U. Bolm-Audorff, W. Schill, M. Kogevinas, P. Boffetta, The contribution of cigarette smoking to bladder cancer in women (pooled European data), Cancer Causes Control 12 (2001) 411–417.
[28] J.F. Tsai, L.Y. Chuang, J.E. Jeng, M.S. Ho, M.Y. Hsieh, Z.Y. Lin, L.Y. Wang, Betel quid chewing as a risk factor for hepatocellular carcinoma: a case–control study, Br. J. Cancer 84 (2001) 709–713.
[29] Y.C. Ko, Y.L. Huang, C.H. Lee, M.J. Chen, L.M. Lin, C.C. Tsai, Betel quid chewing, cigarette smoking and alcohol consumption related to oral cancer in Taiwan, J. Oral Pathol. Med. 24 (1995) 450–453.
[30] P Xiong, M.L. Bondy, D. Li, H. Shen, L.E. Wang, S.E. Singletary, M.R. Spitz, Q. Wei, Sensitivity to benzo(a)pyrene diol-epoxide associated with risk of breast cancer in young women and modulation by glutathione S-transferase polymorphisms: a case-control study, Cancer Res. 61 (2001) 8465–8469.
[31] J. Nair, H. Ohshima, M. Friesen, A. Croisy, S.V. Bhide, H. Bartsch, Tobacco-specific and betel nut-specific N-nitroso compounds: occurrence in saliva and urine of betel quid chewers and formation in vitro by nitrosation of betel quid, Carcinogenesis 6 (1985) 295–303.
[32] C.H. Lee, R.H. Lin, S.H. Liu, S.Y. Lin-Shiau, Mutual interactions among ingredients of betel quid in inducing genotoxicity on Chinese hamster ovary cells, Mutat. Res. 367 (1996) 99–104.
[33] A. Rivenson, D. Hoffmann, B. Prokopczyk, S. Amin, S.S. Hecht, Induction of lung and exocrine pancreas tumors in F344 rats by tobacco-specific and Areca-derived N-nitrosamines, Cancer Res. 48 (1988) 6912–6917.
[34] S. Sen, G. Talukder, A. Sharma, Potentiation of betel-induced alterations of mouse glandular stomach mucosa by tobacco in studies simulating betel addiction, Pharm. Biol. 25 (1987) 209–215.
[35] P.J. van, Bladeren, Glutathione conjugation as a bioactivation reaction, Chem. Biol. Interact. 129 (2000) 61–76.
[36] S. Karami, P. Boffetta, N. Rothman, R.J. Hung, T. Stewart, D. Zaridze, M. Navritalova, D. Mates, V. Janout, H. Kollarova, V. Bencko, N. Szeszenia-Dabrowska, I. Holcatova, A. Mukeria, J. Gromiec, S.J. Chanock, P. Brennan, W.H. Chow, L.E. Moore, Renal cell carcinoma, occupational pesticide exposure and modification by glutathione S-transferase polymorphisms, Carcinogenesis 29 (2008) 1567–1571.
[37] W.J. Kim, H. Kim, C.H. Kim, M.S. Lee, B.R. Oh, H.M. Lee, T. Katoh, GSTT1-null genotype is a protective factor against bladder cancer, Urology 60 (2002) 913–918.
[38] M.E. Kushman, S.L. Kabler, S. Ahmad, J. Doehmer, C.S. Morrow, A.J. Townsend, Protective efficacy of hGSTM1-1 against B[a]P and (+)- or (−)-B[a]P-7,8-dihydrodiol cytotoxicity, mutagenicity, and macromolecular adducts in V79 cells coexpressing hCYP1A1, Toxicol. Sci. 99 (2007) 51–57.
[39] S. Landi, Mammalian class theta GST and differential susceptibility to carcinogens: a review, Mutat. Res. 463 (2000) 247–283.
[40] K.T. Kelsey, S.E. Hankinson, G.A. Colditz, K. Springer, M. Garcia-Closas, D. Spiegelman, J.E. Manson, M. Garland, M.J. Stampfer, W.C. Willett, F.E. Speizer, D.J. Hunter, Glutathione S-transferase class mu deletion polymorphism and breast cancer: results from prevalent versus incident cases, Cancer Epidemiol. Biomarkers Prev. 6 (1997) 511–515.
[41] A.J. Evans, W.D. Henner, K.M. Eilers, M.A. Montalto, E.M. Wersinger, P.E. Andersen, J.I. Cohen, E.C. Everts, J.E. McWilliams, T.M. Beer, Polymorphisms of GSTT1 and related genes in head and neck cancer risk, Head Neck 26 (2004) 63–70.
[42] V. Chaudru, M.T. Lo, F. Lesueur, C. Marian, H. Mohamdi, K. Laud, M. Barrois, A. Chompret, M.F. Avril, F. Demenais, B. Bressac-de, Paillerets, Protective effect of copy number polymorphism of glutathione S-transferase T1 gene on melanoma risk in presence of CDKN2A mutations, MC1R variants and host-related phenotypes, Fam. Cancer 8 (2009) 371–377.
[43] M.C. Lemos, E. Coutinho, L. Gomes, F. Carrilho, F. Rodrigues, F.J. Regateiro, M. Carvalheiro, Combined GSTM1 and GSTT1 null genotypes are associated with a lower risk of papillary thyroid cancer, J. Endocrinol. Invest. 31 (2008) 542–545.
[44] T. Shimada, H. Yamazaki, Y. Oda, A. Hiratsuka, T. Watabe, F.P. Guengerich, Activation and inactivation of carcinogenic dihaloalkanes and other compounds by glutathione S-transferase 5-5 in *Salmonella typhimurium* tester strain NM5004, Chem. Res. Toxicol. 9 (1996) 333–340.
[45] M.D. Ritchie, L.W. Hahn, N. Roodi, L.R. Bailey, W.D. Dupont, F.F. Parl, J.H. Moore, Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer, Am. J. Hum. Genet. 69 (2001) 138–147.
[46] S.M. Tan, A.J. Evans, T.P. Lam, K.L. Cheung, How relevant is breast cancer screening in the Asia/Pacific region? Breast 16 (2007) 113–119.
[47] M.T. Wu, D.C. Wu, H.K. Hsu, E.L. Kao, J.M. Lee, Constituents of areca chewing related to esophageal cancer risk in Taiwanese men, Dis. Esophagus 17 (2004) 257–259.

Diseases of the Esophagus (2010) **23**, 71–75
DOI: 10.1111/j.1442-2050.2009.00975.x

DISEASES OF THE ESOPHAGUS

Original article

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

M. Kaushal,[1] I. Chattopadhyay,[1] R. Phukan,[2] J. Purkayastha,[3] J. Mahanta,[2] S. Kapur,[1] S. Saxena[1]

[1]*Institute of Pathology, Indian Council of Medical Research, Safdarjung Hospital Campus, New Delhi,*
[2]*Regional Medical Research Center, Dibrugarh, and* [3]*Dr B. Borooah Cancer Institute, Guwahati, Assam, India*

SUMMARY. **The incidence of esophageal squamous cell carcinoma (ESCC) is very high in the northeast region of India. 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 of ESCC in high-risk region of India are not known. In the current study of 317 cases of esophageal cancer, 92 (29%) cases had a family history of esophageal and/or other cancers. Of these 92 patients, 45 (49%) patients had a family history of esophageal cancer. The risk of developing esophageal cancer was higher in cases where family history showed occurrence of cancers in first-degree relatives (odds ratio [OR]: 3.1; confidence interval [CI]: 1.9–5.3) than in second-degree relatives (OR: 1.3; CI: 0.25–3.2). Moreover, the risk of developing esophageal cancer was higher in subjects whose predegree suffered from esophageal cancer (OR: 2.4; CI: 1.1–4.1) than from any 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). Screening for mutations of the *BRCA2* gene in the germ line DNA was carried out for 20 familial and 80 nonfamilial ESCC patients. One hundred unrelated healthy controls from the same population were included in this study. Nonsynonymous variants in exon 18 (K2729N) and exon 27 (I3412V) of *BRCA2* gene were found in 3 of 20 patients with familial ESCC. No sequence alterations were found in 80 nonfamilial ESCC cases ($P = 0.01$) and 100 healthy controls ($P = 0.0037$), suggesting that germ line *BRCA2* gene mutation may play a role in familial aggregation of ESCC in high-risk region of India.**

KEY WORDS: ***BRCA2* mutation, familial esophageal cancer.**

INTRODUCTION

Esophageal cancer is among the 10 most common malignancies worldwide and ranks as the sixth leading cause of death from cancer.[1] 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.[2] The esophageal cancer belt is a geographic area of high incidence that stretches from north-central China westward through central Asia to northern Iran.[1] 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] 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.[5] A high prevalence of esophageal cancer has also been reported from Assam in the northeast region of India, with an age-adjusted rate of 33/100,000 males.[6,7]

Contribution of *BRCA2* mutations for the development of esophageal squamous cell carcinoma (ESCC) has been reported in high-risk Chinese population 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% (6 of 44) patients with ESCC with a family history of esophageal cancer in a high-risk

Address correspondence to: Dr Sunita Saxena, MD, Institute of Pathology, Indian Council of Medical Research, Safdarjung Hospital Campus, PO Box 4909, New Delhi 110029, India.
Email: sunita_saxena@yahoo.com

area of China.[8] In Turkmen population of Iran, K3326X nonsense variant has been reported in eight ESCC cases with a family history of esophageal cancer.[5]

BRCA2 was identified as a breast cancer susceptibility gene in 1995 by Wooster *et al.*[9] Germ line mutations of *BRCA2* gene contribute to the development of breast, ovarian, prostrate, and pancreatic cancers.[5,8] *BRCA2* is mainly involved in homologous recombination repair (HRR) through control of *RAD51* recombinase and interacts with many other proteins involved in various cellular functions, including cell cycle regulation, transcription regulation, cytokinesis, and control of cell proliferation.[10]

In the current study, screening for germ line mutations in *BRCA2* gene has been carried out in 20 ESCC cases having family history of esophageal cancer and 80 nonfamilial 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, Guwahati, Assam, during the year of 2005–2006, 92 (29%) cases had a family history of esophageal and/or other cancers besides habit of tobacco and betel quid chewing. Of 92 patients, 45 patients (49%) had a family history of esophageal cancer. Out of 45 patients with familial ESCC, 26 patients had esophageal cancer in first-degree relatives, 2 patients had esophageal cancer in second-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 first- or second-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 nonfamilial 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 ages 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.

Mutation detection

Genomic DNA was extracted by QIAamp DNA Blood Maxi Kit (Qiagen GmbH, Hilden, Germany). The complete coding regions (27 exons) and exon–intron boundaries for *BRCA2* gene were screened for DNA sequence variants by Heteroduplex (HDX) analysis of polymerase chain reaction (PCR) amplicons by using exon specific primers.[11,12] 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-mmol KCl), 1.5-mM $MgCl_2$, 5-mM dNTP mix, 10 μmol of both forward and reverse primer, 0.2 U platinum Taq (Invitrogen, Carlsbad, CA, USA), and 0.4-μCi [α-P33] dATP (BRIT, Department of Atomic Energy, Mumbai, India). An initial denaturation 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.

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 × mutation enhancement gel solution [MDE], 0.6 × *Tris*-base boric acid [TBE], 4% glycerol, 400 μL 10% APS, 40 μL N, N, N′, N′-tetramethylethylenediamine [TEMED]) and run at 8–10 mA for 16–20 h in 0.6 × TBE at room temperature. Gels were dried under vacuum at 80°C for 2 h and exposed to film (Kodak BioMax-MR Amersham, Rochester, NY, USA) for 10–12 h with an intensifying screen. For the possibility of PCR fidelity artifacts to be ruled out, both PCR amplification and gel-based HDX analysis were carried out twice for samples that showed altered mobility on HDX gels.

PCR products showing an aberrant banding pattern were re-amplified and directly sequenced on 3130XL Genetic Analyzer (Applied Biosystems Inc., Foster City, CA, USA) by using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA). Data analysis was carried out by Sequencing Analysis Software V5.2 Patch2 (Applied Biosystems) and ABI PRISM SeqScape Software Version 2.X (Applied Biosystems).

RESULTS

Epidemiological information of enrolled patients and estimation of risk factors

Of 92 patients, 45 (49%) patients had a family history of esophageal cancer. In 64% (59/92) of the cases, cancer occurred in the first-degree relatives, whereas, in 11% (10/92) of the 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%)

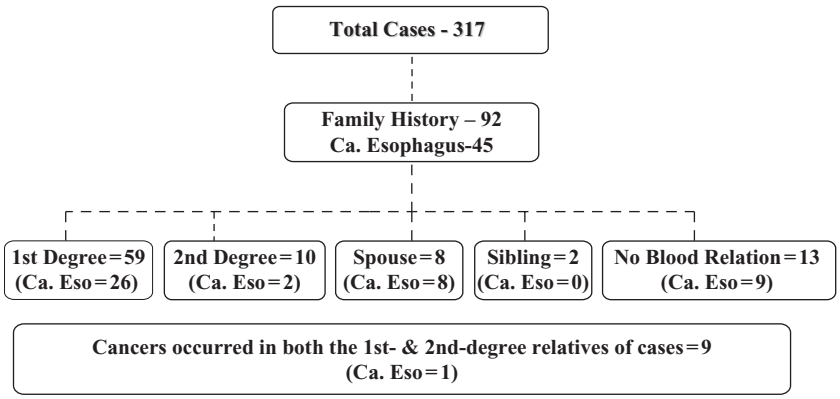


Fig. 1 Status of family history of cancer among esophageal squamous cell carcinoma patients from high-risk area of India registered at Dr Bhubaneshwar Borooah Cancer Institute during the year of 2005–2006.

cases had cancer in family with no blood relation (Fig. 1). 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 predegree had a positive family history of cancers. 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, that is, parents, brother, and sister (OR: 3.1; CI: 1.9–5.3), than among second-degree relatives, that is, paternal and maternal grandparents (OR: 1.3; CI: 0.25–3.2). Moreover, the risk of developing esophageal cancer was higher in subjects whose predegree 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). Demographic, lifestyle cancer risk factors such as smoking, chewing, and alcohol drinking and family history of cancer in 20 cases are shown in Table 1. No associated breast or ovarian cancers were reported in the family members of these patients.

BRCA2 gene sequence variation in germ line DNA

Two germ line variants in exons 27 and 18 of *BRCA2* gene were found in three patients with familial ESCC. 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 (esophageal cancer [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 (EC-85) with a family history of esophageal cancer in the maternal grandfather (Fig. 2). No

Table 1 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 grandfather
EC-187	58	M	Yes	Yes	Yes	Yes	Father

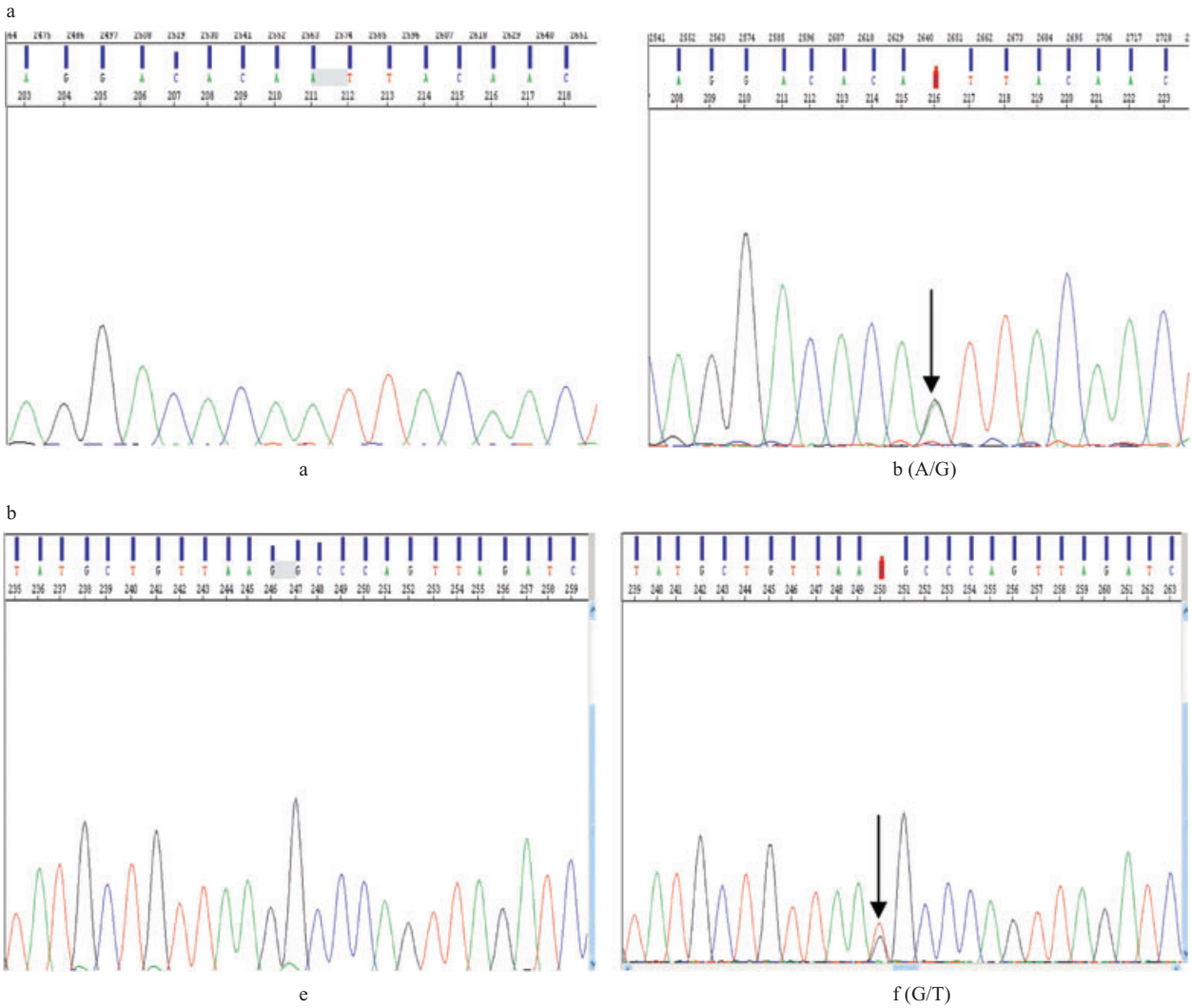


Fig. 2 Sequence chromatograms of missense mutation in exon 27 (a) and exon 18 (b) of *BRCA2* gene as determined by automated sequence analysis. Arrow indicates the position of changes in nucleotide sequences of *BRCA2* gene. Wild type (a: ATT) and mutant (b: GTT) forward sequence of exon 27 of *BRCA2* gene. Wild type (e: AAG) and mutant (f: AAT) forward sequence of exon 18 of *BRCA2* gene.

sequence alterations were found in control group ($P = 0.0037$) and 80 nonfamilial ESCC cases ($P = 0.01$).

DISCUSSION

Germ line 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 cancers, 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 biologic response to DNA damage, as evidenced by the finding that mutations in *BRCA2* lead to chromosomal instability because of defects in repairing double- and single-strand DNA breaks.[8] 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 germ line DNA of 20 ESCC patients with a family history of esophageal cancer. I3412V variant results 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.[8] 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.[5] Thus, this variant is common in ESCC cases from high-risk populations of China, Iran, and India.

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.[8] 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 (esophageal cancer reported in both father and mother) in Turkmen population of Iran.[5,8] 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.[12] 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 fold 1.[13] Oligosaccharide-binding fold 1 is also a site of interaction with FANCG protein, a proposed regulator for unfolding of RAD51 from BRCA2 to the damaged DNA.[14] 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.[15] 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.[5]

Esophageal epithelial cells are exposed to exogenous carcinogens, some of which produce interstrand DNA links that cannot be repaired in cells with defective 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 germ line *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.[5]

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 aerodigestive tract cancers, can significantly modify cancer risk in the presence of an inherited cancer-susceptibility gene.[17]

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.

Acknowledgment

The authors are thankful to Ms. Richa Dubey, Jiwaji University, Gwalior, MP, India, for the technical support.

References

1 Parkin D M, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005; 55 (2): 74–108.
2 Akbari M R, Malekzadeh R, Nasrollahzadeh D *et al.* Familial risks of esophageal cancer among the Turkmen population of the Caspian littoral of Iran. Int J Cancer 2006; 119: 1047–51.
3 Su H, Hu N, Shih J, Hu Y *et al.* 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: 3872–6.
4 Nyren O, Adami H-O. Esophageal cancer. In: Adami H-O, Hunter D, Trichopoulos D, (eds). Text Book of Cancer Epidemiology. Oxford: Oxford University Press, 2002; 137–55.
5 Akbari M R, Malekzadeh R, Nasrollahzadeh D *et al.* Germline BRCA2 mutations and the risk of esophageal squamous cell carcinoma. Oncogene 2008; 27: 1290–96.
6 Phukan R K, Mahanta J, Hazarika N C. Annual Report. Dibrugarh, Assam: Regional Medical Research Center, 2005–2006; 21–4.
7 Phukan R K, Ali M S, Chetia C K *et al.* Betel nut and tobacco chewing; potential risk factors of cancer of oesophagus in Assam, India. Br J Cancer 2001; 85: 661–7.
8 Hu N, Wang C, Han X Y *et al.* Evaluation of BRCA2 in the genetic susceptibility of familial esophageal cancer. Oncogene 2004; 23: 852–8.
9 Wooster R, Bignell G, Lancaster J *et al.* Identification of the breast cancer susceptibility gene BRCA2. Nature 1995; 378: 789–92.
10 Pellegrini L, Yu D S, Lo T *et al.* Insights into DNA recombination from the structure of RAD51-BRCA2 complex. Nature 2002; 420: 287–93.
11 Saxena S, Szabo C I, Chopin S *et al.* BRCA1 and BRCA2 in Indian breast cancer patients. Hum Mutat 2002; 20: 473–4.
12 Saxena S, Chakraborty A, Kaushal M *et al.* Contribution of germline BRCA1 and BRCA2 sequence alterations to breast cancer in Northern India. BMC Med Genet 2006; 7: 75.
13 Yang H, Jeffrey P D, Miller J *et al.* BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science 2002; 297: 1837–48.
14 Hussain S, Witt E, Huber P A, Medhurst A L, Ashworth A, Mathew C G. Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1. Hum Mol Genet 2003; 12: 2503–10.
15 Tian X X. BRCA2 suppresses cell proliferation via stabilizing MAGE-D1. Cancer Res 2005; 65: 4747–53.
16 Hemminki K, Li X. Familial risks of cancer as a guide to gene identification and mode of inheritance. Int J Cancer 2004; 110: 291–4.
17 Hemminki K, Rawal R, Chen B, Bermejo J L. Genetic epidemiology of cancer: from families to heritable genes. Int J Cancer 2004; 111: 944–50.