Study of Gene Expression and Methylation Profiles Associated with Early and Late Onset Indian Breast Cancer Patients

THESIS

Submitted in partial fulfilment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

SHRESHTHA MALVIA
ID. No. 2013PHXF111P

Under the Supervision of

Dr. Sunita Saxena

Co-Supervision **of Dr. S.A. Raju Bagadi and Dr. Uma S. Dubey**

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2018

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE,

PILANI, RAJASTHAN

CERTIFICATE

This is to certify that the thesis entitled **"Study of Gene Expression and Methylation Profiles Associated with Early and Late Onset Indian Breast Cancer Patients"** submitted by **Shreshtha Malvia,** ID No 2013PHXF111P for award of Ph.D. of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor
Name in capital letters — Dr. SUNITA SAXENA
Designation — Consultant, Ex Director,
National Institute of Pathology,
I.C.M.R, Safdarjung Hospital
Campus, New, Delhi

Date:

Signature of the Co Supervisor
Name in capital letters — Dr. S.A RAJU BAGADI
Designation — Scientist D
National Institute of Pathology,
I.C.M.R, Safdarjung Hospital
Campus, New, Delhi

Date:

Signature of the Co Supervisor
Name in capital letters — Dr. UMA. S. DUBEY
Designation — Associate Professor
Birla Institute of Technology&
Science Pilani, Rajasthan

Date:

Acknowledgement

With immense pleasure and profound gratitude I would like to take this opportunity to thanks all the people who have helped me through this course of my journey towards producing the thesis. First and foremost I sincerely want to thank my Supervisor ***Dr. Sunita Saxena****, Ex Director, National Institute of Pathology, ICMR. It has been a complete honour to be her Ph.D. student since her expertise; understanding and patience guided me continuously throughout my Ph.D. You have been a never ending source of inspiration, with your immense knowledge and skills in many areas which has oriented, supported and influenced me to the core of my heart. Your generous and constant support and encouragement has enlighten me throughout my journey for attaining this Ph.D. Your enthusiasm and joy for research was contagious and motivational for me, mainly during tough times in the Ph.D. pursuit. You have set an excellent example as a true mentor for me.*

This thesis would not have been possible without the help, support and patience of my co supervisor, Dr S.A. Raju Bagadi, Scientist C, National Institute of Pathology, ICMR. You have taught me, both consciously and unconsciously, how good experiments are conducted. Your passion to attain the in depth knowledge of data has driven me throughout the journey. I appreciate all your contributions of time, ideas and support to make my Ph.D. experience productive and stimulating.

This effort would have been incomplete without the support of Ex Director Dr. Sunita Saxena and acting Director Dr. Nasreen Z. Ehtesham, National Institute of Pathology for providing me the opportunity to pursue my PhD in the same institute.

I would like to extend my thanks and heartfelt gratitude to my co supervisor, Dr. Uma. S. Dubey, Associate professor, BITS Pilani campus for her overwhelming support, motivation and guidance towards attainment of Ph.D degree.

I am immensely thankful to Prof. Souvik Bhattacharyya, Vice chancellor of BITS Pilani for providing me the opportunity to pursue the PhD from the institute. I would also extend my gratitude to Prof S.K Verma , Dean, Academic Research Consultancy (ARC), BITS, Pilani for his constant official support and encouragement. I would like to thank my Doctoral Advisory Committee (DAC) members, Dr. Rajdeep Chowdhury and Dr Anupama Mittal who have

contributed immensely with good advice and comments for the improvement of my research work.

I am greatfull to our collaborators , Dr. Chintamani, Department of general surgery and Dr. Amar Bhatnagar, Department of cancer surgery, Safdarjung Hospital, New Delhi alongwith Dr Deepshikha , Dr Ramesh Sarin, Indraprastha Apollo Hospital, New Delhi for providing the tissue biopsies.

I am thankful to Dr. Dibyabhaba Pradhan, Scientist II , National institute of Pathology for providing me all the bioinformatics support and help for analyzing my data to the core. I would also like to thanks my senior Dr. Anurupa Chakraborty for the guidance which enabled me to pursue Ph.D through my tough times. My labmate Meena Lakhanpal, juniors Ananya Chowdhury, Prachi Gupta, Priyanka, Karuna have been encouraging and supportive. Thanks for laughs and keeping the environment pleasant and convenient to work.
I would like to thanks other members of my lab, Mrs Valsamma Mathew, Mr Jagat Singh for their assistance and help during my research work. I wish to thank all the students and staff that have passed through the institute for their inputs and support which enabled me to drive through the tenure.

Finally I would like to thanks the ones who are closest to my heart- my family for each and everything and most importantly believing in me when I doubted myself. My mother, father has been a strong pillar throughout my life for supporting me and making me what I am today. Thanks for instilling and inculcating these qualities in me that have motivated me all through the life. Special thanks to my mom and dad (in laws), sister in law for their patience and relentless support which enabled me to create a balance between personal and professional life. Without their support I would never have been able to complete my thesis.

Lastly my loving and caring husband who has been my guide, philosopher and friend for motivating me to attain and achieve my dreams. Thanks for being there for my failures, giving me strength and pushing to achieve success. Finally, the youngest members from my family my niece Ashi and my son Dhanwin have been like a boon to me. They have been like my pills to treat my mental frustration levels. Their laughs and hugs have given me strength to overcome all my difficulties and problems.

*Last but not the least I would like to thanks myself for pursuing my dreams and achieving it...**Well Done**...!!*

SHRESHTHA MALVIA

Abstract

Breast cancer is the most commonly reported cancers across world in females with leading mortality rates. There has been a progressive increase in the incidence of breast cancer over the last few decades especially in Asian countries including India. Indian subcontinent faces a high incidence of early onset breast carcinoma involving young women less than 40 years of age succumbing to the disease. Breast cancer is a heterogeneous disease with varying molecular and clinical characteristics. The underlying mechanisms leading to early onset breast cancer seems to be probably different from late onset cancer and have to be further elucidated.

In the present study whole genome expression profiling was performed in Indian patients in order to decipher their underlying molecular biology and mechanisms which are necessary to understand their role in carcinogenesis. To accomplish this 29 breast cancer patients and 9 controls were used for microarray using Illumina bead array. Differentially expressed genes in breast tumours were screened and further validated by Quantitative real-time. We found *COL10A1, COL11A1, MMP11, MMP1, MMP13, GJB2, CST1, KIAA1199, CEACAM6, BUB1* to be differentially upregulated and *PLIN, ADH1A, ADH1B, CIDEC, FABP4,* to be downregulated which were part of cell cycle, cell proliferation and metastasis, cell adhesion, ECM receptor interaction, focal adhesion pathways, PPAR. Further 9 genes leading to deregulation in extracellular matrix pathways, causing invasion and metastasis were uniquely deregulating early onset breast cancer. They consisted of *CDH15, VCAN, PVRL2, CDKN2A, SMAD3, IBSP, ITGA11, BRAF, and FZD2*. Genes such as *HRAS, SMAD3* were the most interactive hub top node identified in early onset breast cancer.

Matrixmetalloproteinases (MMPs) and Collagens were two gene families which showed highest fold expression in breast tumours and also their expression was differential between

early and late onset tumour groups. Validation of MMP genes by real time PCR was done, which in turn showed concordance with the microarray data. Molecular subtyping of the obtained profiles categorized our tumour samples and genes into the known molecular subtypes of breast cancer showing the presence of PAM50 breast classifiers in Indian breast cancer patients also. The Indian breast cancer patients showed higher percentage of Luminal subtypes which is in concordance with the western data. Early onset patients showed higher percentage of basal subtype, making them hormone negative and thus possibly explaining the reasons for aggressive behaviour of young breast cancer females in Indian subcontinent.

Further whole genome methylation profiling in 48 breast cancer tissue was also performed. Differentially methylated CpG sites in tumour were identified using beta normalization. 382 uniquely hypermethylated or hypomethylated genes in late and 385 in early were identified. The major top genes included *HHIP, RYR2, CLDN11, CDH5, TNFRSF10D* aberrantly methylated in tumours majorly playing role in apoptosis, cAMP signalling, cell adhesion, and cytokine cytokine recptor interactions. In early onset tumours *Bcl2, CFTR, TNNI3, LIPE, SELE* and *ICAM2* were aberrantly methylated which were further validated by MethHC software.

It is known that DNA methylation leads to reprogramming of gene expression and thus setting the hallmarks for the cancer initiation. Thus integrating our gene expression and methylation data identified common genes such as *SOCS2, TNFRSF10D, LHCGR* whose hypermethylation was leading to gene repression in both early and late onset breast tumours. Early onset breast cancer showed *STAT5A, ALDH1A2* as uniquely negatively correlating genes where their hypermethylation caused repression, deregulating processes such as proliferation and immune surveillance. Late onset tumours showed hypermethylation of *LEP* gene resulting in its gene repression and leading to deregulation of the fatty acid metabolism in breast cancer patients.

Thus critical challenge lies in identifying potential gene signatures which might serve as biomarkers for predicting clinical outcome and also have therapeutic utility. The genes identified in the present study have future prognostic values but have to be validated further in large cohort of Indian breast cancer patients to establish their possible role as biomarker.

Table of Contents

S.No.	Chapter	Page Number
1. Chapter 1	**Introduction**	**1-5**
2. Chapter 2	**Review of Literature**	**6-48**
2.1	**Cancer**	**6**
2.1.1	Worldwide scenario	**6**
2.1.2	Cancer scenario in India (2012-2014)	**8**
2.1.3	Global breast cancer scenario	**10**
2.1.4	Breast cancer scenario in India	**12**
2.2	**Normal anatomy of breast**	**14**
2.3	**Benign breast disease**	**15**
2.4	**Classification of breast carcinoma**	**16**
2.4.1	*In-situ* carcinoma	**16**
2.4.2	Invasive (infiltrating) breast carcinoma	**16**
2.4.3	Other types of breast cancer	**17**
2.5	**Staging of breast cancer**	**17**
2.6	**Grading of breast cancer**	**18**
2.7	**Sub-types in breast cancer**	**18**
2.7.1	Estrogen receptor (ER)	**19**
2.7.2	Progesterone Receptor (PR)	**19**
2.7.3	HER2/neu	**20**
2.8	**Molecular classification of breast cancer**	**20**
2.9	**Risk Factors for breast cancer**	**21**
2.9.1	Age	**21**
2.9.2	Race	**22**
2.9.3	Reproductive factor	**23**
2.9.3.1	Age at menarche and menopause	**23**
2.9.3.2	Nulliparity and age at first birth	**23**
2.9.3.3	Breast Feeding	**23**
2.9.4	Lifestyle breast cancer risk factors	**24**
2.9.4.1	Alcohol consumption	**24**

2.9.4.2	Diet	**24**
2.9.4.3	Obesity and physical activity	**24**
2.9.4.4	Tobacco and betel quid chewing	**25**
	Family history	**25**
2.10	**Molecular basis of breast cancer**	**26**
2.10.1	Oncogenes	**26**
2.10.2	Tumor suppressor genes	**27**
2.10.3	Growth factors	**27**
2.11	**Genome wide studies on breast cancer**	**29**
2.12	**Molecular classification of breast cancer via gene expression profiling**	**30**
2.13	**Gene expression in diagnostics**	**34**
2.13.1	Mammaprint	**35**
2.13.2	(Oncotype DX Breast Cancer Assay)	**35**
2.13.3	Theros-Breast Cancer Gene Expression Ratio Assay	**36**
2.13.4	Genomic Grade Index (MapQuant DX)	**36**
2.13.5	PAM50/Breast BioClassifier	**36**
2.14	**DNA Methylation**	**37**
2.14.1	Types of DNA methylation	**38**
2.14.1.1	CpGs	**38**
2.14.1.2	CpG islands and promoters	**38**
2.14.1.3	Repetitive elements	**39**
2.14.2	Mechanism of DNA methylation	**40**
2.14.2.1	DNMT1	**40**
2.14.2.2	DNMT3a and 3b	**40**
2.14.3	DNA methylation and gene expression	**40**
2.14.3.1	Hypermethylation and gene silencing	**41**
2.14.4	Aberrant DNA methylation changes in cancer	**41**
2.14.4.1	DNA hypomethylation	**41**
2.14.4.2	DNA hypermethylation	**41**
2.14.4.3	Imprinting disorders	**42**
2.14.5	Aberrant DNA methylation in breast cancer	**42**
2.14.6	Analysis of DNA methylation	**44**
2.14.7	Clinical relevance of DNA methylation	**44**

2.14.7.1 Diagnosis of early cancers **45**
2.14.8 Treatment and management of breast cancer **46**
2.14.8.1 Surgery **46**
2.14.8.2 Radiation therapy **46**
2.14.8.3 Chemotherapy **47**
2.14.8.4 Targeted therapy **47**
2.14.8.5 Endocrine therapy **48**
2.14.8.6 Gaps in existing research **48**
3. Chapter 3 Aim and Objectives 49-50
4. Chapter 4 To study the gene expression profiles in breast cancer in Indian women and compare the differentially expressed genes in early and late onset tumors 51
4.1 Introduction **51**
4.2 Materials and methods **52**
4.3 Results **60**
4.4 Discussion **108**
5. Chapter 5 To study methylation profiles of breast cancer in Indian women and compare the differential profiles in early and late onset tumours 122
5.1. Introduction **122**
5.2. Materials and methods **123**
5.3. Results **129**
5.4. Discussion **155**
6. Chapter 6 Conclusion and future Scope of study 161
7. Chapter 7 References 167
List of Publication and Presentation
Brief biography of candidate
Brief biography of supervisor
Brief biography of co-supervisor

List of Tables

Table Number	**Page Number**
Table 2.1 Summary statistics for world population depicting the number of cancer cases and deaths along with 5 year prevalence	**7**
Table 2.2 CR and AAR amongst different registries for male and female	**9-10**
Table 2.3 Showing estimated number of breast cancer incident and death cases along with 5 year prevalence	**12**
Table 2.4(a). Female breast Relative Proportion (%) and rank (r) (2012-2014 in all PBCR's) **Table 2.4** (b). Female breast Crude Rate (CR) and Age Adjusted Rate (AAR) per 100,000	**13**
Table 2.5 Summary of initial microarray studies related to human breast cancer	**32**
Table 4.1 Showing association of various clinicopathological characteristics of patients with age of onset of breast cancer	**61-62**
Table 4.2 Top 20 DEGs between breast cancer and normal tissues	**70**
Table 4.3 Pathways found differentially expressed in all tumors	**71**
Table 4.4 Pathways upregulated in all tumors	**71**
Table 4.5 Pathways downregulated in all tumors	**72**
Table 4.6 Top DEGs identified amongst ET vs CC a.) Upregulated genes b.) Downregulated genes	**78**
Table 4.7 Top DEGs identified amongst LT vs CC a.) Upregulated genes b.) Downregulated genes	**78-79**
Table 4.8 List of pathways deregulated in early onset tumours	**80**
Table 4.9 List of pathways deregulated in late onset tumors	**80**
Table 4.10 MMP pathway genes showing differential expression in ET and LT tumours along with their adjusted p value	**81**
Table 4.11 Cell adhesion molecule pathway genes differentially expressed in (a) Early onset tumours (b) Late onset tumours	**82**
Table 4.12 Cell cycle pathway genes differentially expressed in (a) Early onset tumours (b) Late onset tumours	**83**
Table 4.13 ECM receptor pathway genes differentially expressed uniquely in (a)	**84**

Early onset tumours (b) Late onset tumours	
Table 4.14 Pathways in cancer/ Metastasis genes differentially expressed uniquely in (a) Early onset tumours (b) Late onset tumours	**84**
Table 4.15 Leukocyte transendothelial migration genes differentially expressed uniquely in Late onset tumours	**85**
Table 4.16 Top Gene deregulated in various molecular subtypes of breast cancer	**88-89**
Table 4.17 Genes deregulated uniquely and playing functional role in various molecular subtypes of breast cancer	**89-90**
Table 4.18 List of pathways and genes deregulated uniquely in advanced stage of tumours	**95**
Table 4.19 List of pathways and genes deregulated uniquely in lower stage of tumours	**96**
Table.4.20 Unique gene sets specific to Indian breast cancer population	**99**
Table 4.21 Top commonly deregulated gens in both Indian and western population in breast cancer	**99-100**
Table 4.22 Genes getting commonly deregulated but with opposing directions in both populations	**100**
Table 4.23 Clinicopathological details of patients used in real time validation	**101**
Table 4.24 Showing expression MMP activity genes in all tumours a) micro array data and b) qPCR	**103**
Table 4.25 Fold change in breast cancer tumour tissue in age specific groups viz early onset (ET) and late onset (LT) breast cancer tumours as obtained from microarray	**103**
Table 4.26 Real time PCR validation in breast cancer tumour tissue in age specific groups *viz.* early onset (ET) and late onset (LT) breast cancer tumours	**103**
Table 5.1 Clinicopathological details of samples used for methylation microarray	**124**
Table 5.2 Beta values for the differentially methylated genes in breast tumours.	**135**
Table 5.3 Functional enriched genes getting deregulated in early tumors	**138**
Table 5.4 cAMP pathway genes showing their beta values in tumour and normal along with their diff score	**140**
Table 5.5 Cell adhesion pathway genes showing their beta values in tumour and normal along with their diff score	**140**

Table 5.6 Unique and common genes showing functional significance in ET and LT pathways	**141**
Table 5.7 List of genes showing their beta value and fold change for early integrome	**145**
Table 5.8 List of genes showing their beta value and fold change for late integrome	**146**
Table 5.9 Beta values of tumour and normal breast cancer tissues obtained after validation	**149**
Table 5.10 Median beta and p values for tumours and normals as observed in tumours and normals	**151**
Table 5.11 Interactome genes showing their correlation coefficient and corresponding p values	**152-153**

List of Figures

Figure Number	Page Number
Fig. 2.1 Estimated incidence, mortality and 5-year prevalence in both sexes across world	**8**
Fig. 2.2 Worldwide breast cancer ASR incidence and mortality rates	**11**
Fig. 2.3 Comparison of Age adjusted incidence rates for all PBCR's for breast cancer	**14**
Fig. 2.4 Anatomy of the human breast	**15**
Fig. 2.5 Trends for 5-year age distribution among different Indian registries	**22**
Fig. 2.6 Hallmarks of cancer	**28**
Fig. 2.7 Bead image showing its address bar and a probe bar which in turn attaches to labeled cRNA	**30**
Fig. 2.8 Mechanism of DNA Methylation	**38**
Fig. 4.1 Patient sample distribution for (a) microarray study (b) Real time validation	**56**
Fig.4.2 Images depicting changes in samples before and after normalization.	**57**
Fig. 4.3 Plot showing distribution of intensity values before and after normalization	**57**
Fig. 4.4 Gel electrophoresis image showing the quality of extracted RNA samples	**63**
Fig. 4.5 Bioanalyzer image showing high quality of RNA samples (RIN)	**63**
Fig. 4.6 Image of chip showing uniform hybridization of the probes	**64**
Fig. 4.7 Volcano plots showing genes differentially expressed in all breast tumour tissue vs. controls.	**68**
Fig. 4.8. (a) Heat map showing the clustering of all tumour tissues with respect to their normals. (b) Dendrogram for tumours and normals	**69**
Fig. 4.9 Heat Map of the top 50 genes that are differentially expressed in all tumours (Tumour Vs Control)	**73**
Fig. 4.10 Volcano plots showing genes differentially expressed in (a) ET vs CC and (b) LT vs. CC.	**74**

Fig. 4.11. (a) Heat map generated by clustering of ET and LT samples. (b) Sample dendrogram showing distinct ET and LT clusters	**76**
Fig. 4.12 Differential clusters identified amongst ET and LT (a) Cluster1 (b) Cluster 2 showing genes upregulated genes in ET and LT respectively.	**77**
Fig. 4.13 Venn diagram showing the unique and common genes amongst ET and LT	**79**
Fig. 4.14 Metallopeptidase genes pathway	**81**
Fig. 4.15 (a) Early unique network (b) Late unique network	**86**
Fig. 4.16 Early unique network of genes playing significant functional role in breast carcinogenesis	**87**
Fig. 4.17 Breast cancer predicted molecular subtype in Indian breast cancer patients. The subtypes LumimalA, LuminaB, Her2Nu, Basal and normal like were identified. a. Heat map. b. Sample Dendorgram	**91**
Fig. 4.18 Venn diagram representing the common and unique genes belonging to all the five predicted molecular subtypes of Indian breast cancer patients	**92**
Fig. 4.19 Network image for unique basal subtype genes	**92**
Fig. 4.20. Network image for unique HER2/neu subtype genes	**93**
Fig. 4.21 Network showing unique Luminal A interactive genes	**94**
Fig. 4.22 Network showing unique Luminal A interactive genes	**94**
Fig. 4.23 Venn diagram showing unique and common genes in advanced and lower stages	**95**
Fig. 4.24 Network showing unique higher stage interactive genes	**97**
Fig. 4.25 Network showing unique lower stage interactive genes	**98**
Fig. 4.26 Scatter plots showing the differential fold change amongst all tumour *vs.* all normal in breast cancer tissues.	**104-105**
Fig. 4.27 Scatter plots showing the differential fold change amongst early onset tumour *vs.* late onset tumours in breast cancer tissues.	**106-107**
Fig. 4.28 Breast carcinogenesis mechanism leading to increasing incidence of early onset breast cancer in Indian females.	**120**
Fig. 5.1 Agarose gel showing the good quality of DNA which was extracted from the tumour tissues	**129**
Fig. 5.2 Scan of the hybridized chip showing overall view of the probes present on chip	**130**

Fig. 5.3 Quality control assessment of the methylation microarray experiment	**131-132**
Fig. 5.4 Histogram depicting hypermethylated (red) and hypomethylated (green) CpG sites in tumours *vs*. controls	**132**
Fig. 5.5 Heat map of differentially methylated genes to identify the differential profiles between tumour vs controls	**133**
Fig. 5.6 Scatter plots of early and late onset tumours showing (a) hypermethylated and (b) hypomethylated CpG sites	**136**
Fig. 5.7 Venn diagram representing the differentially methylated genes in early and late onset breast cancer	**136**
Fig. 5.8 Showing Top nodes uniquely methylated in early onset tumours (a) Hypermethylated and (b) hypomethylated	**139**
Fig. 5.9 Top interactive nodes in late onset tumours (a) Hyper and (b) hypomethylated unique in LT	**142**
Fig. 5.10 Integration of gene expression and methylation data for breast cancer patients	**143**
Fig. 5.11 Early onset tumour integrome	**144**
Fig. 5.12 Late onset tumour integrome	**146**
Fig. 5.13 Box plots for average Beta values of tumours and normal controls for differentially methylated genes amongst various pathways in all tumours	**148**
Fig. 5.14 Box plots of tumours and normals for differentially methylated genes amongst various pathways in ET/LT	**150-151**
Fig. 5.15 Validation of significantly interacting and correlating genes as found in microarray experiments using MethHC software. (a) STATA5A (b) LEP (c) ALDH1A2 (d) LHCGR (e) SH2D3C (f) SOCS2 (g) TNFRSF10D.	**153-154**
Fig. 6.1 Interplay between gene expression and methylation results in early onset breast carcinogenesis deregulating several genes	**165**

List of Abbreviations

AAR : Age adjusted incidence rate

AJCC : American Joint Committee on Cancer

ANOVA : Analysis of variance

APC : Annual percentage change

AS : Advanced stage

ASR : Age standardized rate

BCS : Breast conserving surgery

CC : Control

CR : Crude rate

DCIS : Ductal carcinoma *in situ*

DEG : Differentially expressed genes

DNA : Deoxynucelic Acid

ER : Estrogen Receptor

ET : Early onset tumour

ETD : Extralobular terminal duct

FDR : False discovery rate

GO : Gene ontology

GSEA : Gene set enrichment analysis

HER2/Nu : Receptor tyrosine-protein kinase erbB-2

IDC : Invasive ductal carcinoma

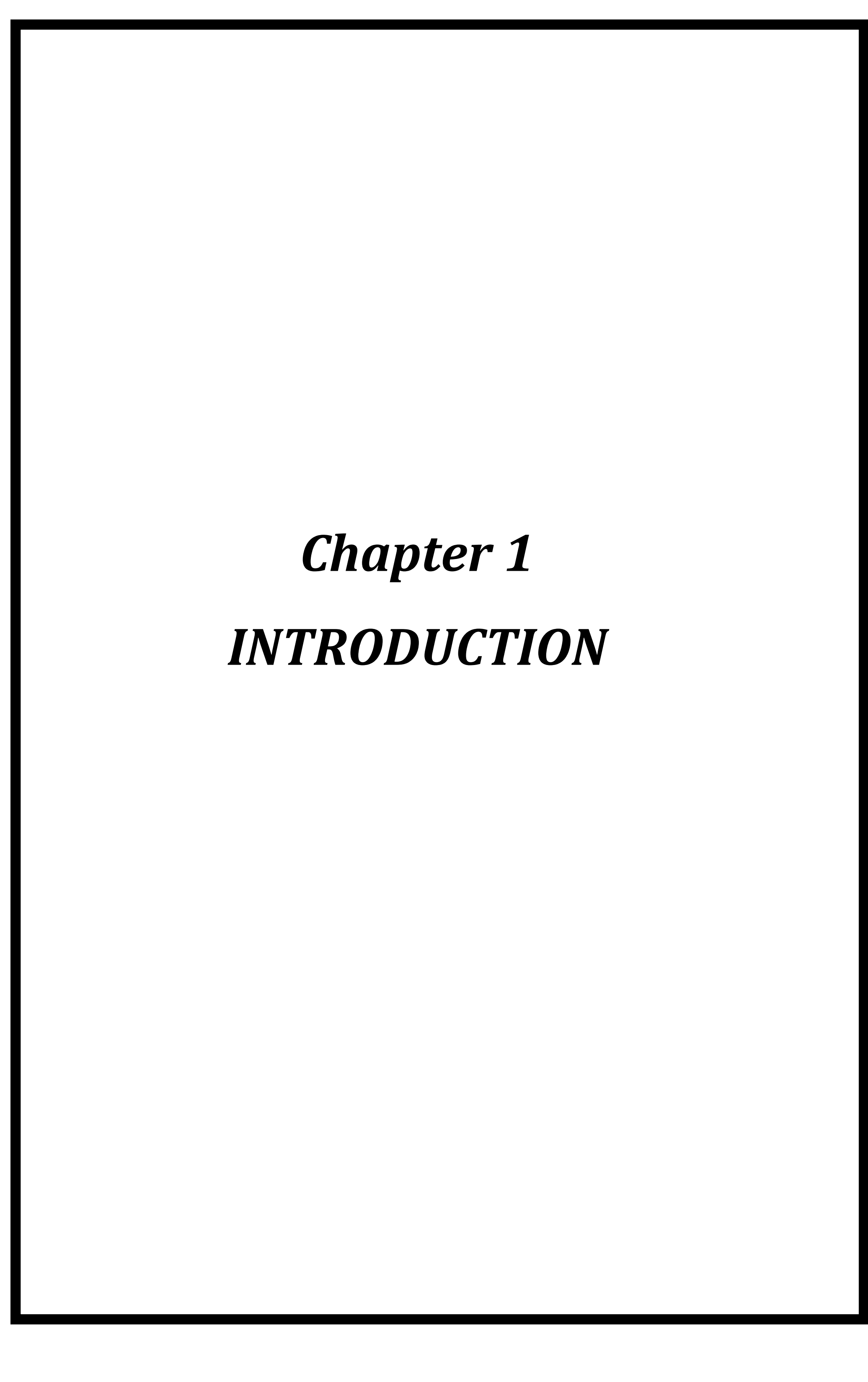
ITD : Intralobular terminal duct

IHC : Immunohistochemistry

KEGG : Encyclopedia of Genes and Genomes

LCIS : Lobular carcinoma *in situ*

LG	: Late onset tumour
LIMMA	: Linear models for microarray analysis
LS	: Lower Stage
mRNA	: Messenger RNA
methHC	: A database of DNA Methylation and gene expression in Human Cancer
NCRP	: National Cancer Registry Project
PBCR	: Population based cancer registry
PR	: Progesterone Receptor
Prev.	: Prevalence
RNA	: Ribonucleic Acid
RT	: Room Temperature
SERMS	: Selective estrogen receptor modulators
TDLU	: Terminal ductal lobular unit
TT	: Total tumour



Chapter 1

INTRODUCTION

Chapter 1

Introduction

Breast cancer is the most common malignancy amongst women worldwide with an age standardized incidence rate of 43.3 per 100,000 (1). In India it represents a major health concern with an age standardized incidence rate of 25.8 per 100,000 (1). Breast cancer has rapidly overtaken cervical cancer in frequency [Age adjusted rates(AAR) 21-37/100000 as compared to 12-18/100000 women] amongst all urban cancer registries *viz.* Delhi, Mumbai, Bangalore, Ahmedabad, Thiruvanthapuram, whereas in rural registry of Barshi cervical cancer is the still the most common cancer among women (2). It is estimated that by 2020, breast cancer will emerge as the leading cancer in Indian women amongst all registries. The breast cancer incidence in India is significantly lower as compared to western countries (AAR 21-37/100,000 *vs.*75-90/100,000), but mortality is at par (12.7 *vs* 17.1 per 100 000) with United Kingdom (1). Studies reported that majority of breast carcinoma cases in Asian countries were diagnosed at higher Stage III (11.3% vs. 7.0%, $p < 0.001$), Stage II (47.5% vs. 37.5%, $p < 0.001$) with a smaller percentage of women having stage I (36.3% vs. 51.0%, $p < 0.001$) as compared to Caucasian women (3-4). Increasing breast cancer incidence in Indian young women (<40 years) showed a higher proportion of early onset of breast cancer as compared to women of older ages >55 years (late onset) (5-6)**.** Trends for 5-year age distribution among different registries showed a peak relative proportion between 45 and 49 years in all registries except in north eastern registries where the peak is seen in even 10-year younger age group-35–39 (7).These patients tend to have highly aggressive biological behaviour and are also found to have clinical association with unfavourable prognosis compared with the disease arising in older women.

This variation in incidence of breast cancer within different regions of the world may be possibly attributed to lifestyle differences (diet and environmental exposures), genetic and epigenetic differences among populations. Several epidemiological studies conducted in different populations suggests potential risk factors for breast cancer as age, geographic location, socioeconomic status, reproductive events, oral contraceptives, lifestyle risk factors, history of benign breast disease, mammographic density, ionizing radiation, bone density, chemo preventive agents, height as well as genetic factors. Throughout life, women are exposed to hormones and it is this lifetime exposure to ovarian hormones that is believed to put a woman at risk for developing breast cancer.

Breast cancer, like other malignancies progresses by accumulation of a series of genetic or epigenetic changes in the genes regulating cellular proliferation, differentiation, cell death (apoptosis or necrosis), DNA repair, tissue compartmentalization, etc. Identification of early markers preceding clinical manifestation for disease is an important objective for disease prevention.

Emerging technologies, such as gene expression profiling, copy number variation, methylation profiling were increasingly valued as powerful tools for new biomarker identification since they enable simultaneous analysis of thousands of genes for identification of the underlying mechanism involved in tumourigenesis and progression of cancer. Initial microarray analysis of clinical breast cancer specimens has lead to classification of breast cancer at molecular level which tends to be associated with different clinical outcomes (8-12). These various subgroups are Luminal A, Luminal B, Basal and ERBB2 over-expressing and Normal-like subtypes. The Luminal A sub-type over expresses ESR1 and ER genes enabling better prognosis, response to endocrine therapy as compared to Luminal sub-type B. The remaining Basal and ERBB2 subgroups of patients display a more

aggressive form of cancer, ERBB2 over expressing patients responding well to Herceptin (12-13). Basal groups are devoid of hormone receptors and thus are non responders to therapies. Further these expression studies were translated to diagnostic assays to be used routinely on breast cancer patients for diagnosis and prognosis. They were later on approved by the US Food and Drug Administration (FDA; http://www.fda.gov/) also. Besides prediction of clinical outcome these studies can also show some therapeutic options for breast cancer patients.

Anders *et al*. (14), studied gene expression profiles of young women suffering from breast cancer and found pathways such as phosphatidylinositide 3-kinase and Myc getting deregulated in younger patients (15). Azim *et al.* did metanalysis in 2012 (16) and reported higher expression of the *BRCA1* mutation signature along with luminal progenitors and c-kit in younger patients. *RANKL* (Receptor activator of nuclear factor kappa-B ligand) gene expression was also found to be higher in young women which stimulates osteoclastogenesis (17). Although India faces higher incidence of younger onset breast cancer but no such Indian studies discussing the differential profiles in breast cancer in young woman have been reported or elucidated so far. Hence the rising proportion of early onset tumours in Indian and other Asian countries, suggests a need for analysis of molecular mechanism/etiology thereby involved causing early onset carcinogenesis. It would in turn enable identification of prognostic and predictive molecular signatures, which is of prime importance for diseases prevention (18).

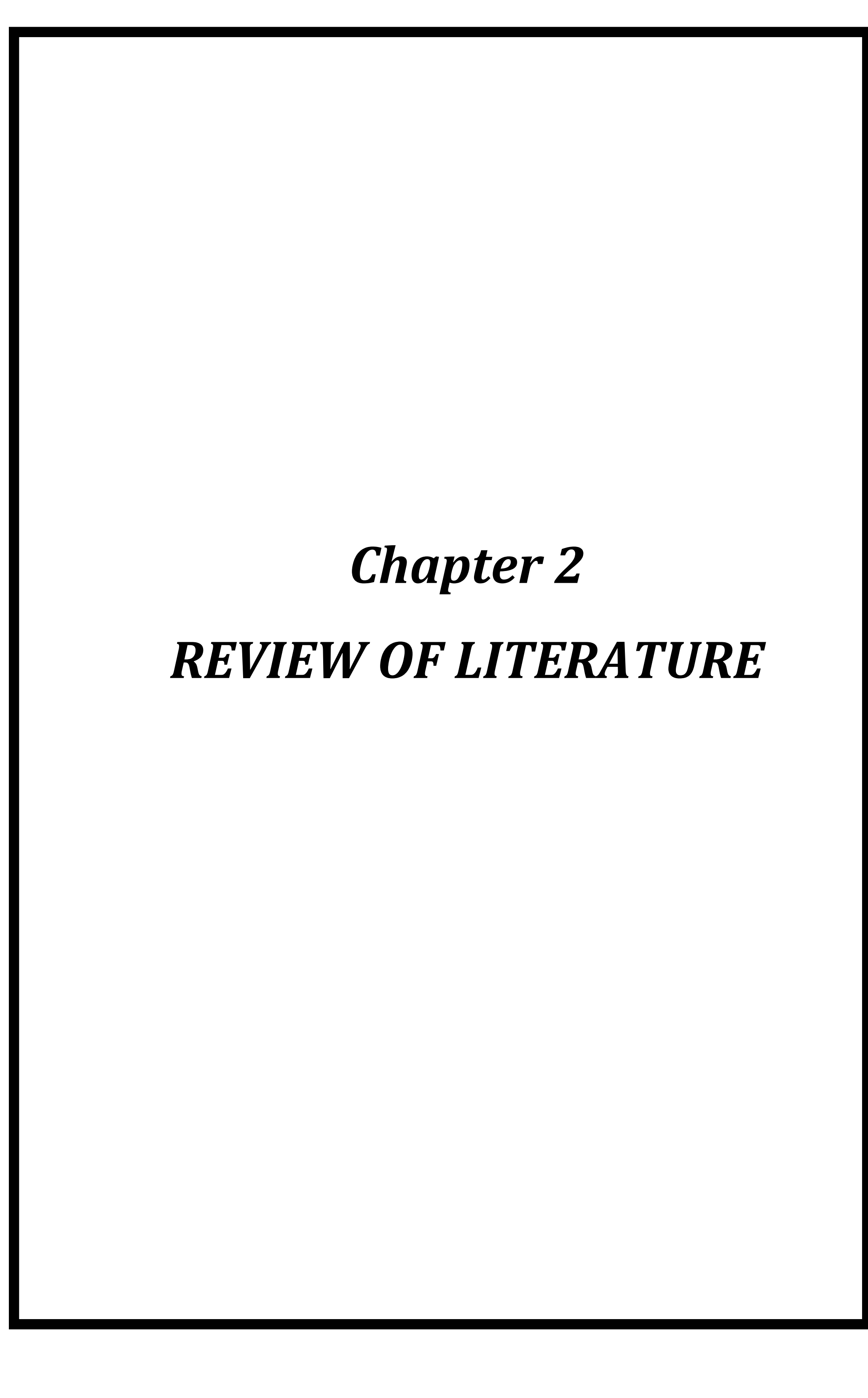
Another phenomenon regulating breast cancer is DNA methylation which is heritable epigenetic signature regulating gene expression. It occurs by addition of methyl group to DNA thereby modifying function of genes. In humans around 98% DNA methylation occurs in CpG dinucleotide in somatic cells and around a quarter methylation in non-CpG

dinucleotides in embryonic stem cells (19). The DNA methylation plays an important role in normal development along with several other key processes such as X-chromosome inactivation, genomic imprinting, and transcription and transposition, suppression of repetitive element etc. Epigenetic hallmarks of cancer are marked by global DNA hypomethylation and locus specific hypermethylation of CpG islands. This methylation is regulated by DNMTs (DNA Methyltransferase) and its dysregulation leads to diseases like cancer (20-21). The gene transcription requires free access to gene promoters to allow binding of transcription factors and methylation can result in prevention of transcription factor binding leading to changes in chromatin structure. The role of DNA methylation for normal and diseases development are widely studied in humans (20).

DNA methylation proves to be a robust biomarker, having more stability than RNA and proteins. Recently whole-genome approaches have allowed detection of DNA methylation signatures specific to breast-cancer, associated with specific clinical outcomes. It is a promising target for diagnosis, prognosis and development of new approaches for of breast cancer and other diseases (22). Targeted DNA methylation of candidate genes could be studied using techniques as methylation-sensitive restriction enzymes or by sequencing bisulfite-converted DNA. It depicted the differential nature of methylation patterns in tumour and normal controls. Genome-wide loss of methylation has been studied in many cancers including breast cancer (23). These specific methylation patterns can be used to identify various networks and pathways playing important role in breast cancers serving as diagnostic and prognostic tools in breast cancer.

Studying changes in gene expression and DNA methylation patterns at genome level would enable identification of complex molecular signatures which are crucial for cancer initiation, progression and metastasis. These altered patterns of DNA methylation have the ability to

modify the regulation of gene expression and are frequently observed in the early stages of breast tumourigenesis (24-25). However, till date the molecular changes leading to early and late onset breast cancer are not yet studied in Indian population. None of studies from India have reported the underlying pathways leading to epigenetic silencing or over expressing of genes in breast cancer. Hence this study was undertaken to elucidate the mechanisms leading to breast cancer in Indian females.



Chapter 2

REVIEW OF LITERATURE

Chapter 2

Review of Literature

2.1 Cancer

Cancer is a group of diseases which involves abnormal cell growth with the potential to invade to other parts of the body. Tumourigenesis is a multistep process involving a large number of molecular, biochemical and cellular steps which reflect genetic alterations that are key drivers of the progressive transformation of normal human cells to highly malignant derivatives (26-27). The main focus of cancer research is to identify genes (oncogenes, tumour suppressors etc) within cancer cells to gain knowledge on cancer biology and pathogenesis.

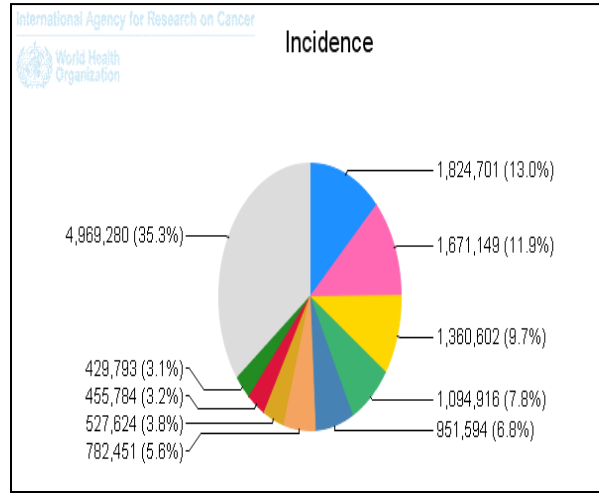
2.1.1 Worldwide Scenario

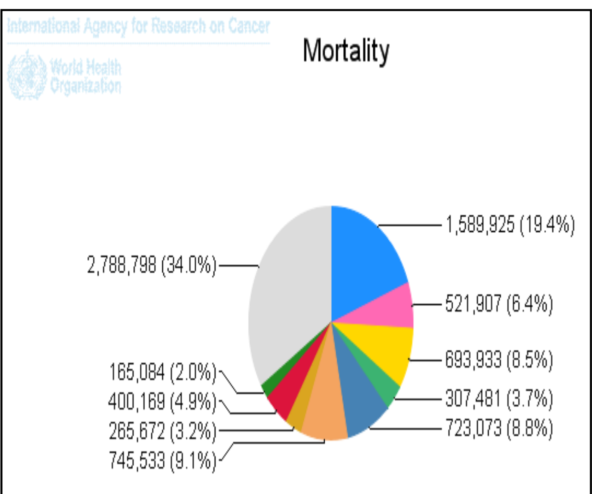
As per Globocan 2012, global cancer estimates for both sexes were 14.1 million new cases, 8.2 million cancer deaths, and five-year prevalence of 32.6 million cancers in individuals above the age of 15 years (Table.2.1) (1). Around 8 million new cancer cases, 5.3 million cancer deaths with 15.6 million five-year prevalence have been reported in the less developed regions of the world (Table.2.1) (1). It is predicted that by year 2025 increase in cancer cases would be 19.3 million (1). Globally, the five most common cancers reported in both sexes are cancers of the lung (1,824,701; 13%), breast (1,676,633; 11.9%), colorectum (1,360,602; 9.7%), prostate (1,111,689; 7.9%), and cervix uteri (527,624; 3.7%), (Fig. 2.1) comprising a total of 46.2% of the 28 cancer types reported (1). Further, deaths due to these five cancers were 3,378,622. In men cancers of lung (16.7%), prostate (15%), colorectum (10%), stomach (8.5%) and liver (7.5%) were most common, accounting for a total of 4,285,250 cancers cases, and 2,769,670 (1) cancer deaths. In women, the five most common cancers reported

were cancers of breast (25.2%), colorectum (9.2%), lung (8.8%), cervix uteri (7.9%), and corpus uteri (4.8%) accounting for a total of 3,721,266 cases and 1,675,069 deaths. Although overall incidence of cancer is lower in less developed countries than more developed countries but more than 70% of all cancer deaths occur in low and middle income countries, where the resources available for prevention, diagnosis and treatment of cancer are limited. It is predicted that by the end of year 2020, over 10 million people would die globally because of cancer with 70% deaths occurring from the developing countries only. However on global level World Health Organization has reported lower cancer incidence in India compared to those in developed countries including USA (1).

Table.2.1 Summary statistics of world population depicting the number of cancer incident cases and cancer deaths along with 5 year prevalence in both sexes

Estimated numbers (thousands)	**Cases**	**Deaths**	**5-year prev.**
World	14068	8202	32455
More developed regions	6054	2878	16823
Less developed regions	8014	5323	15632
WHO Africa region (AFRO)	645	456	1363
WHO Americas region (PAHO)	2882	1295	7958
WHO Europe region (EURO)	3715	1933	9701
WHO South-East Asia region (SEARO)	1724	1171	3278
WHO Western Pacific region (WPRO)	4543	2978	8956
IARC membership (24 countries)	7038	3470	18595
United States of America	1604	617	4775
China	3065	2206	5045
India	1015	683	1790
European Union (EU-28)	2635	1276	7157





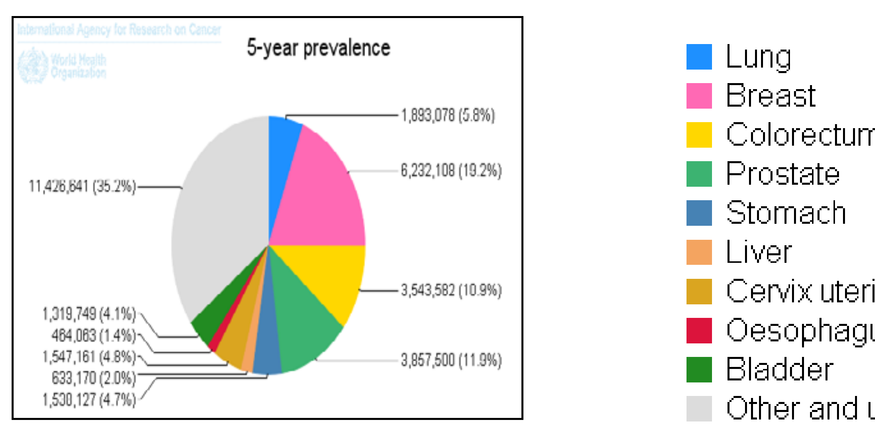


Fig.2.1. Estimated incidence, mortality and 5-year prevalence in both sexes across world *(Adapted from GLOBOCAN 2012(1))*

2.1.2 Cancer scenario in India (2012-2014)

India is a developing country having cancer incidence lower than the western countries. Maximum number of cases registered were from Delhi (19746) followed by Thiruvananthapuram District (15640), Mumbai (13357), Chennai (11659) and Kollam (11012) population based cancer registries (PBCRs) for the year 2012-2014 (2). Cancer cases in males were more in Delhi, Cachar District, Dibrugarh District, Kamrup Urban District, Mizoram State, Sikkim State, Ahmedabad Urban District, Kolkata, Kollam District,

Meghalaya State, Tripura State and Nagaland PBCRs. Female cancers were more in Bangalore, Chennai, Mumbai, Manipur State, Nagpur, Pune, Thiruvananthapuram District, Wardha and Patiala PBCRs. The crude rate (CR) and age adjusted rate (AAR) for different registries are mentioned in Table 2.2. Among males in Delhi PBCR the leading sites of cancer were: lung (10.5%), mouth (6.9%), prostate (6.7%), tongue (6.5%) and larynx (5.7%). The respective CR and AAR per 100,000 for the sites were: lung (11.8 and 17.2), mouth (7.8 and 9.5), prostate (7.6 and 12.4), tongue (7.3 and 9.3) and larynx (6.4 and 8.9) while in females for the same registry leading cancer sites was breast (28.6%) followed by cervix uteri (10.8%), gall bladder (7.9%), ovary (7.2%) and corpus uteri (3.5%). The respective CR and AAR per 100,000 population for the sites were: breast (34.8 and 41.0), cervix uteri (13.2 and 15.5), gall bladder (9.6 and 11.8), ovary (8.7and 10.0) and corpus uteri (4.3 and 5.5). Estimated number of new cancers diagnosed in India every year is 7 00,000 -900,000 (2).

Table.2.2 CR and AAR amongst different registries for male and female (Adapted from NCRP2012-2014(2))

Registry	Male		Female	
	CR	AAR	CR	AAR
Bangalore (2012)	82.8	105	107	126
Barshi Rural (2012-2014)	56.8	53.9	66	60.4
Barshi Expanded (2012)	39.9	40.9	54.6	52
Bhopal (2012-2013)	82.3	102	90.4	108
Chennai (2012-2013)	116	116	132	126
Delhi (2012)	113	149	122	145
Mumbai (2012)	98.1	113	117	119
Cachar District (2012-2014)	96.2	125	78.7	95.2
Dibrugarh District (2012-2014)	72.2	92.8	66.9	78.6
Kamrup Urban District (2012-2014)	143	206	123	174
Manipur State (MR) (2012-2014)	45.2	60.5	56	68.6
Mizoram State (MZ) (2012-2014)	147	212	122	166

Aizawl District (2012-2014)	205	271	167	208
Sikkim State (2012-2014)	71	90.7	76.2	100
Ahmedabad Urban (2012-2013)	87.4	98.5	73	76.5
Aurangabad (2012-2014)	56.7	72	60.4	73
Kolkata (2012)	119	101	121	103
Thi'puram District (2012-2014)	161	132	154	120

2.1.3 Global breast cancer scenario

Breast cancer is the most common cancer in the world amongst women with around 1.67 million new cancer cases diagnosed in 2012 which is 25% of all cancers (1). In more developed countries, the numbers of breast cancer cases (788,000) were slightly less than what was reported in less developed countries (883,000) (Table.2.3) (1). There is around four fold variation in breast cancer incidence across different regions of the world, with 27 per 100,000 in Middle Africa and Eastern Asia to 92 in Northern America (Fig.2.2) (1). Though breast cancer is the fifth cause of death amongst all cancer (522,000 deaths), but it is leading cause of death in less developed regions (324,000 deaths) than death for women in more developed regions (198,000 deaths). Developed regions have higher incidence of breast cancer but in turn they have better survival rates, which results in decreased mortality. The mortality rates range from 6 per 100,000 in Eastern Asia to 20 per 100,000 in Western Africa (1).

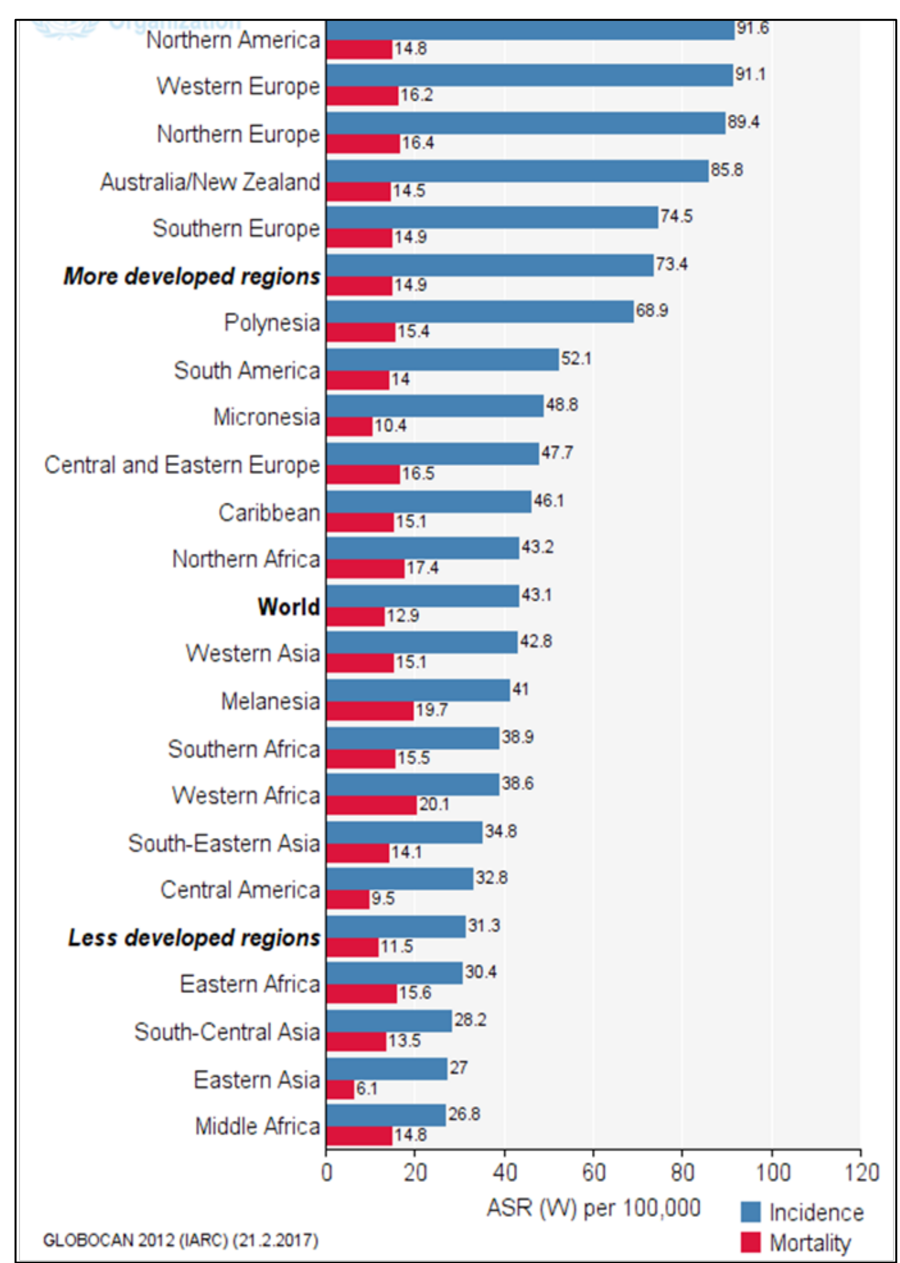


Fig 2.2 Worldwide breast cancer ASR incidence and mortality rates (Adapted form GLOBOCAN 2012(1))

Table 2.3 Showing estimated number of breast cancer incident and death cases along with 5 year prevalence (Adapted from GLOBOCAN 2012(1)

Estimated numbers	Incident cases (100,000	Deaths (100,000)	5-year prev.
World	1671	522	6232
United States of America	233	44	971
More developed regions	788	198	3201
Less developed regions	883	324	3032
WHO Africa region (AFRO)	100	49	318
WHO Americas region (PAHO)	408	92	1618
China	187	48	697
India	145	70	397
European Union (EU-28)	362	92	1444

2.1.4 Breast cancer scenario in India

Breast cancer is the most common cancer amongst Indian females with age standardized rate as high as 25.8 per 100,000 women and mortality 12.7 per 100,000 women. India faces a challenging situation due to 11.54% increases in incidence and 13.82% increase in mortality due to breast cancer during 2008 to 2012 (1, 28). Breast cancer attains top rank even in individual registries (Chennai, Mumbai, Bangalore, New Delhi and Dibrugarh) in females during the period of 2012-2014 (Table2.4a). ranging from 30.7% in Chennai to 19% in Dibrugarh (Table 2.4.b) (2). Increasing urbanization and westernization associated with changing lifestyle and food habits, has lead breast cancer to become number one in all major urban registries, while in Barshi rural registry still cervical cancer is at top position in females and cancer of breast holds a second position. Breast cancer Crude rate (CR) amongst different registries showed highest rate in Thiruvananthapuram 43.9 (per 100,000) followed by Chennai (40.6), New Delhi (34.8) and Mumbai (33.6). Among all the PBCR's top four places were occupied by Delhi with AAR 41.0 (per 100,000), Chennai 37.9, Bangalore 34.4

and Thiruvananthapuram District 33.7 (Fig.2.3)(6). Indian women with breast cancer are found a decade younger in comparison to western women suggesting that breast cancer occurs at a younger premenopausal age in India (3, 29-31). UK cancer registry showed an increasing trend for breast cancer from age 30-35 achieving highest peak during age 60-65 years. This suggested that an average woman in India under the age of 40 has a considerably higher chance of developing young age disease unlike UK (7). Reports suggest that western breast cancers are mostly represented in Stages I and II of disease, whereas Indian women tends to report in higher advanced stages (45.7%) (3-4). Disease presentation in such condition results in increased mortality in India.

Table 2.4(a). Female breast Relative Proportion (%) and rank (r) (2012-2014 in all PBCR's)

Breast	**%**	**R**
MUMBAI	28.8	1
BANGALORE	27.5	1
CHENNAI	30.7	1
THIRUVANANTHAPURAM	28.5	1
DIBRUGARH	19	1
NEW DELHI	28.6	1
BARSHI RURAL	20	2

Table 2.4(b). Female breast Crude Rate (CR) and Age Adjusted Rate (AAR) (2012-2014 in all PBCR's)

Breast	**CR per 100000**	**AAR per 100000**
MUMBAI	33.6	33.6
BANGALORE	29.3	34.4
CHENNAI	40.6	37.9
THIRUVANANTHAPURAM	43.9	33.7
DIBRUGARH	12.7	13.9
NEW DELHI	34.8	41
BARSHI RURAL	13.2	12.4

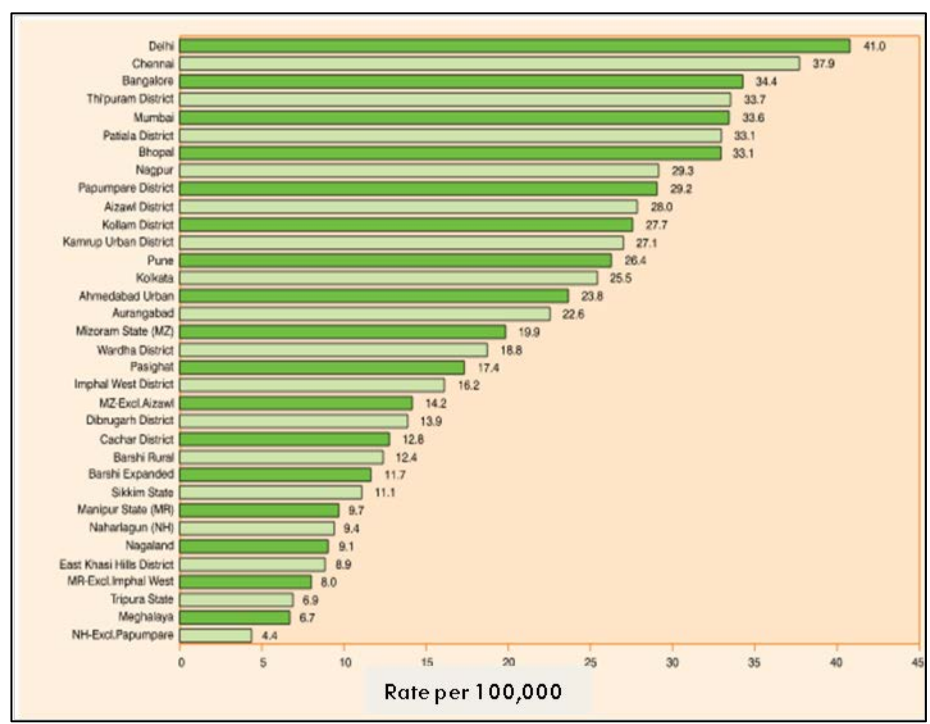


Fig.2.3 Comparison of Age adjusted incidence rates for all PBCR's for breast cancer (Adapted from PBCR 2012-2014 (2))

2.2 Normal anatomy of breast

The breast is composed of glandular ducts, lobules, connective tissue, and fat. The main morpho-functional unit of mammary gland is a single gland with a complex branching structure called the terminal ductal lobular unit (TDLU) which consists of extralobular terminal duct (ETD) which attaches the lobule to the ductal system; A intralobular terminal duct (ITD) which continues the duct system into the lobule and clusters of 10-100 sac-like acini that open into the ITD (Fig.2.4).

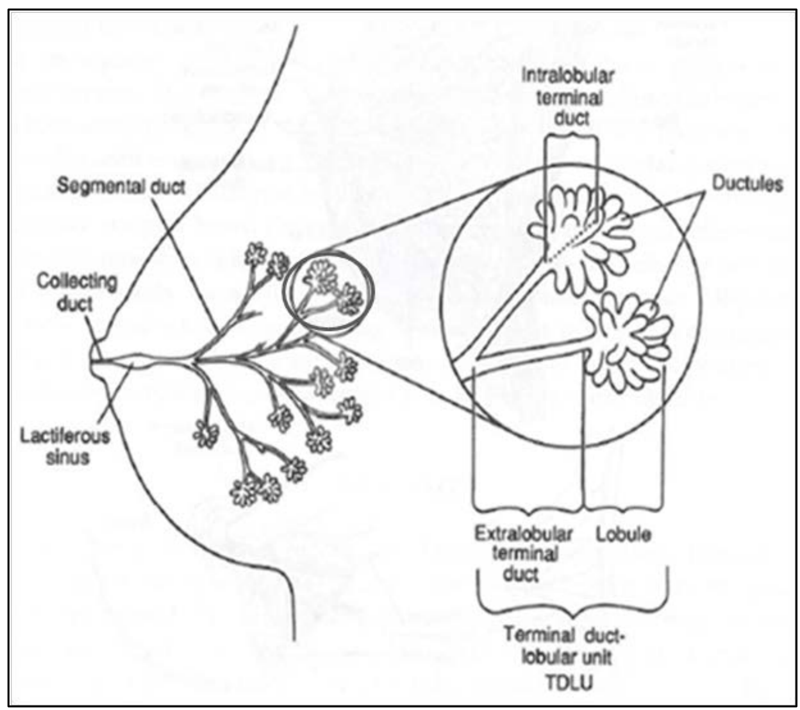


Fig.2.4 Anatomy of the human breast (http://wommen.org.uk/breast-anatomy/)

Acini and the terminal duct are the source of milk production. The majority of pathologic changes in the breast are believed to arise from the TDLU.

2.3 Benign Breast Disease

Benign breast diseases encompass a heterogeneous group of lesions that differ in their histopathologic feature and clinical behavior. It can be either non proliferative lesions, or proliferative lesions with or without atypia or atypical hyperplasias (32). Benign tumours are associated with trauma, pain and tenderness (mastalgia) and infection (mastitis) resulting in a palpable mass, swelling, skin dimpling, erythema, thickening, pain, nipple discharge, inversion/retraction (33-34). It could be in the form of either fibrosis/cysts or benign breast tumours. Fibrosis is firmness in the connective tissues, and cysts are fluid-filled sacs which can cause areas of lumpiness, thickening, or tenderness; nipple discharge; or pain in the breast. Non-cancerous benign breast tumours are hard and solid areas where abnormal division of breast cells results in a lump formation.

2.4 Classification of breast carcinoma

Carcinomas are broadly divided into *in situ* and invasive carcinoma. Carcinomas *in situ* are the tumours that are limited to ducts or lobules and confined to the basement/epithelial component of the organ (*in-situ*). In some cases cells extend beyond the basement membrane and invade the stroma and called as Invasive carcinoma.

2.4.1 *In-situ* carcinoma was originally classified as Ductal carcinoma *in-situ* (DCIS) or lobular carcinoma *in situ* (LCIS). Tumours either arising from or involving a duct is termed as DCIS while tumours either arising from or involving a lobule is termed as LCIS. Morphologically DCIS has been divide into 5 architectural subtypes namely, comedocarcinoma, solid, cribriform, papillary and micropapillary. The majority of cases of DCIS cannot be detected by either palpation or visual inspection of involved tissue. On other hand LCIS are mostly incidental findings in a biopsy performed for another reason as LCIS is not associated with calcifications or a stromal reaction which tends to form density in tumours.

2.4.2 Invasive (infiltrating) breast carcinoma is generally presented as a palpable mass and has stromal and basement membrane invasion. They are of two types: Invasive ductal carcinoma (IDC) and Invasive lobular carcinoma (ILC). IDC accounts for the majority of invasive breast cancers. It begins in the duct forming cells of breast with the gross appearance of an irregular stellate outline. Invasive lobular carcinoma (ILC) is characterized by the presence of small and relatively uniform tumour cells growing singly and in concentric (pagetoid) fashion around lobules involved by in situ lobular neoplasia. The incidence of lobular carcinomas has been reported to be increasing among postmenopausal women possibly due to use of hormone replacement therapy.

2.4.3 Other types of breast cancer includes inflammatory breast carcinoma, Paget's disease of the nipple, tubular, medullary, cribiform, metaplastic and apocrine carcinoma. Development of breast cancer manifests itself as a sequence of pathologically defined stages by initiating as a premalignant stage of atypical ductal hyperplasia (ADH, progresses into pre-invasive stage of ductal carcinoma in situ (DCIS) and culminates in the potentially lethal stage of invasive ductal carcinoma (IDC) (35).

2.5 Staging of breast cancer

Staging defines the extent to which disease is spread which in turn determines the treatment and estimates patients' prognosis. Staging can be done both clinically and pathologically. Presently staging is determined by the parameters laid by the American Joint Committee on Cancer (AJCC) 1977. The AJCC system is a clinical and pathologic staging system based on TNM nomenclature in which T refers to the tumour, N to lymph node and M refers to metastasis.

TNM Classification	
T= Primary Tumour	
Tx	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma in situ
T1	Greatest tumour dimensions, 2cm or less
T2	Greatest tumour dimension, more than 2 cm but less than 5 cm
T3	Greatest tumour dimension, more than 5 cm
T4	Any size tumour with extension to wall or skin
N	Regional Lymph Node
Nx	Regional Lymph Nodes cannot be assessed
N0	No regional nodes metastasis
N1	Metastases to movable ipsilateral axillary lymph node(s)
N2	Metastases to ipsilateral axillary lymph node(s) fixed to one another or other structure
N3	Metastases to ipsilateral internal mammary lymph node(s)
M	Distant metastasis
Mx	Presence of distant metastasis cannot be assessed
M0	No distant metastasis

M1	Distant metastasis
Stage Grouping	
Stage 0	TisN0M0
Stage 1	T1N0M0
Stage IIA+ IIB	T0N1M0,T1N1M0,T2N0M,T2N1M0,T3N0M0
Stage IIIA + IIIB	T0N2M0, T1N2M0, T2N2M0, T3N1M0, T3N2M0 +T4 Any N ,M0, any TN3M0
Stage IV	Any T Any N, M1

2.6 Grading of breast cancer

Histological grade is a widely accepted prognostic parameter for invasive breast carcinoma. The Nottingham method is the most accepted method of determining the histological grade. The grade is obtained by adding up the scores for tubule formation, nuclear pleomorphism, and mitotic count resulting in a score of 3-9 points which is translated into the final grade. For a score of 3-5 points the cancer has Grade I cancer; for 6-7 points cancer falls in Grade II while for 8-9 points cancer comes under Grade III. Grade I cancer are categorized as well differentiated cancer which are slow in growth. Grade II cancer are moderately differentiated with comparatively faster growth than grade I. Grade III are poorly differentiated cancer and are very fast growing and aggressive.

2.7 Sub-types in breast cancer

Breast cancer is highly heterogeneous group of cancers, arising from different cell types, showing different clinical implications. Hence breast cancer protein profiling for Estrogen Receptor, Progesterone Receptor and HER2/Nu proteins might helps in determination of the most appropriate treatment and predict the prognosis for individual patient.

2.7.1 Estrogen receptor (ER)

Estrogen receptor, which is a member of nuclear hormone family of intracellular receptors and gets activated by hormone 17β-estradiol (36). Estrogen Receptor mainly regulates gene expression(37). ER has two isoforms α encoded by ESR1 and β encoded by ESR2 genes(38). 70% of breast cancer cases tends to over express ER and are termed as "ER positive" tumours. of breast cancer cases, and are referred to as When ER binds to estrogen receptor, it stimulates proliferation of mammary cells resulting in an increased cell division, DNA replication & mutation rate which in turn causes disruption of the cell cycle, DNA repair processes and apoptosis, eventually leads to tumour formation. Expression of ERα expression is associated with more differentiated tumours in comparison to ERβ. However, research suggests that there is association of ERβ with proliferation and poor prognosis(39). Patients expressing higher levels of ER in tumour tissue were treated with endocrine therapy which consisted of ER antagonists called as Selective ER Modulators (SERMS) or aromatase inhibitors which inhibits the action of enzyme aromatase which functions by converting androgens into estrogens.

2.7.2 Progesterone Receptor (PR)

Progesterone receptor (PR) is an intracellular steroid receptor which binds to progesterone coded by PGR gene (40). There are two main forms A and B which differ in their molecular weight. PR is mainly expressed in reproductive tissue playing important role in folliculogenesis, implantation, ovulation and pregnancy (41). Estrogen is necessary for inducing progesterone receptors (PRs) activity (42). PRs binding to receptor leads to their hyperphosphorylation which in turn requires multiple serine kinases (43). This binding is followed by dimerization leading its entry to the nucleus where it binds to DNA and produces proteins. Around 65% ER-positive breast cancers are also PR-positive and while about only

5% of ER-negative breast cancers are PR-positive, but it is considered as hormone positive in both cases. Any mutation or change in expression of its co-regulators can affects the normal functions by disrupting the normal development and thereby leading to breast carcinogenesis (44).

2.7.3 HER2/neu

Human Epidermal growth factor Receptor 2 (HER2/neu also known as ERBB2) a member of transmembrane receptor tyrosine kinase being a protein associated with higher aggressiveness in breast cancers (45). It is a proto-oncogene mainly involved in signal transduction pathways that regulate cell growth, differentiation and proliferation (46). HER2/neu is important for its role in the breast cancer pathogenesis and as a target for treatment. Ligand binding results in dimerization of ErbB receptors with majorly HER2/neu being dimerisation partner of other members of the ErbB family(47). Its over-expression in breast cancer is associated with increased disease recurrence and worse prognosis. HER2/neu is known to co-localize with another protooncogene GRB7, and thus most of the time is co-amplified, with it (48). Clinically, HER2/Neu over expression is targeted by monoclonal antibody trastuzumab (marketed as Herceptin) which in turn increases p27, a protein that halts cell proliferation (49).

2.8 Molecular classification of breast cancer

There is a need to classify patients into Hormone positive or negative or HER-2-overexpressing categories, since targeted therapies for breast cancer are only effective in these subsets of patients. Hence these molecular signatures for a particular subset would also lead to the discovery of new therapeutic targets and treatments. Several high throughput technologies as gene-expression profiling have evolved which enabled the classification of breast cancer amongst various subtypes. Classification of breast cancer into 5 major molecular subtypes

was the result of first comprehensive study on gene expression profiling on breast cancer patients viz. luminal-like with ER/PR positive and HER2/neu negative, basal-like with ER/PR/ HER2/neu negative, normal-like, and HER2/neu over expressing with ER/PR negative and HER2/neu positive (9, 50). These subgroups mainly differed in prognosis and sensitivity to chemotherapy. Most favourable long term survival was represented by Luminal-like cancers since these were ER positive with low grade tumours responding well to endocrine therapy whereas basal-like and HER-2-positive tumours were more sensitive to chemotherapy (51-52).

2.9 Risk Factors for Breast Cancer

Risk factors mainly includes age, location, reproductive status events as (menarche, menopause, pregnancy, breastfeeding), hormones exposure as (hormone replacement therapy and oral contraceptives), lifestyle factors (diet, obesity, alcohol and physical activity), history of benign breast disease, mammographic density, density, ionizing radiation, prolactin levels, chemo preventive agents, and genetic factors (high and low penetrance breast cancer susceptibility genes).

2.9.1 Age

Age is a major risk factor for breast cancer. In US and Sweden it continues to rise up to 75 years while in Colombia, the age-specific increase is considerably smaller after age of 45 (53-54). In India women with breast cancer are found a decade younger in comparison to western women suggesting that breast cancer occurs at a younger premenopausal age (3, 29-31, 55). According to the National Cancer Registry Programme [NCRP] based on cancer registries at six hospitals, the average age of patients was found ranging between 44.2 years (Dibrugarh) to 49.6 years (Bangalore and Chennai) (NCRP 2001) whereas amongst US

caucasian females it was 61.0 years (56). Studies suggest that the disease peaks at 40-50 years in Indian women (57). These cancers are either HER2/Neu positive or ER/PR/ HER2/neu all three negative, having poor prognosis. Trends for five year age distribution amongst different registries showed a peak between 45-49 years amongst all registries except in north eastern registries where the peak was seen in even ten year younger age group–35-39 (Fig. 2.5) (7).

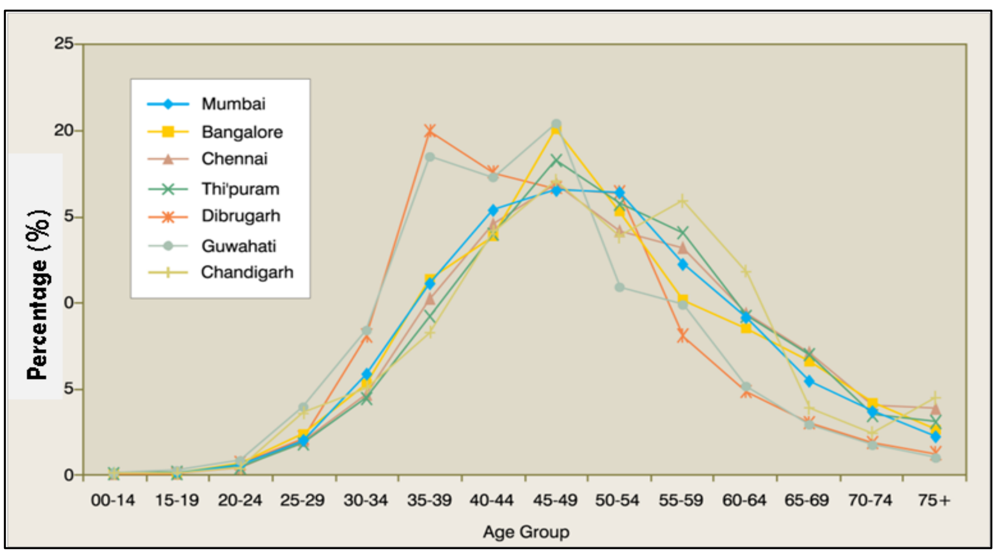


Fig. 2.5 Trends for 5-year age distribution among different Indian registries (Adapted from NCRP 2010–2012 (7))

2.9.2 Race

Ethnicity plays a role in breast cancer incidence and mortality. Black women were more likely to die of breast cancer followed by white, Hispanic, Asian/Pacific Islander, and American Indian/Alaska Native women. In India incidence of breast cancer is found to be more in urban than rural women. The average incidence rates ranges from 22-28 per 100,000 in urban settings to 6 per 100,0000 women in rural areas. Rural women are mainly adherent to rural lifestyle making them less vulnerable to disease although they are more exposed to pesticides (58). Higher incidence of breast cancer has been reported in parsi women than other religious communities. The rates and patterns of

cancer are strikingly different in the northeastern part of India which is ethnically isolated. Population in Northeast is more prone to cancers compared to remaining part of the country. Certain pockets of North eastern region have highest breast cancer incidence rate like Aizwal district in Mizoram and Kamrup district in Assam (North Eastern Regional Cancer Registry data 2006).

2.9.3 Reproductive factor

2.9.3.1 Age at menarche and menopause

Early menarche (less than 12 yrs vs. 14 yrs) has been found to be associated with increased breast cancer risk which is probably because of prolonged exposure of estrogens and progesterone to breast epithelium. Contrastingly, delayed menopause maximizes number of ovulatory cycles leading to an increased breast cancer risk (59-60).

2.9.3.2 Nulliparity and age at first birth

Nulliparity and old age at first birth contributes toward increased risk of developing breast cancer (61-62). Nulliparous women had a 2.2-fold higher risk than multiparous women and late age at first pregnancy (35 years and above) showed 5.4 fold higher risk compared with women having first pregnancy at age of 14 years or less (63). Moreover older age at first childbirth emerged as stronger risk factor for contralateral breast cancer (64). In western and southern India, females faced higher risk with increasing age while high parity was associated with 40-50% reduction in risk ($P < 0.01$) (63, 65-67).

2.9.3.3 Breast feeding

Prolonged lactation has a protective role in breast cancer development (68). There is a 4.3% decrease in the relative risk of breast cancer for every 12 months of breastfeeding, in addition to a decrease of 7.0% with each birth (69). Women from North India has revealed

strong association of risk factors like breast-feeding, location (urban/rural) and increased BMI to be associated with breast cancer ($p < 0.05$) (70).

2.9.4 Lifestyle breast cancer risk factors

2.9.4.1 Alcohol consumption

There was a positive association between risk of developing breast cancer with alcohol intake in both pre and postmenopausal women (71-72). Alcohol tends to increase sex hormone levels which in turn increase breast epithelial cell responsiveness which further leads to tumour initiation and promotion through leading to production of genotoxic metabolite acetaldehyde and oxidative stress. *In vitro* studies have reported alcohol stimulates invasion and migration of breast cancer cells by interfering with the epithelium–stroma interaction, enhancing EMT. Aberrant DNA methylation patterns could also lead to alcohol-induced cancer development (32).

2.9.4.2 Diet

Intake of fruits and vegetables along with other rich sources of natural antioxidants decreases breast cancer risk (73-75). The various cultural and religious practices in India introduces diversity in dietary patterns since past thousands of years (76-77). Every 10 years, nutrition recommendations are updated by ICMR according to surveys conducted by the National Institute of Nutrition (NIN), Hyderabad (78). NIN recommends diet including high intake of fresh vegetables and fruits and spices such as turmeric, in adequate amounts to prevent cancer (79).

2.9.4.3 Obesity and physical activity

Obesity and breast cancer have a complex relationship modulated by menopausal status.

Obesity induces high levels of endogenous estrogen, produced mainly by adipose tissues which are known to be a major source of estrogen (80). Physical activity delays the onset of menarche and thus reduces the risk of developing cancer by modulating the availability of hormone levels (81).

2.9.4.4 Tobacco and betel quid chewing

Breast cancer ranks top in north east India and has significant association with the factors as betel quid and tobacco chewing habits (p=0.003) (82-85). Environmental carcinogens tends to induce genetic alterations which differ from alterations observed in the non exposed ones (83). Betel Quid chewers faces more risk of developing breast cancer than the non-chewer with significant p value 0.0003 (95% CI 1.334-4.150) (86).

2.9.4.5 Family history

Family history of breast cancer increases lifetime risk two to three times for developing breast cancer (87) than a non familial case. Genetic factors play a major role in promoting familial breast cancer like mutations in BRCA1, BRCA2 gene inherited from parents (88). The frequency of BRCA1/2 genetic mutations ranged from 2.9% to 24.0% among Indian familial breast cancer patients (89). Also 2.8% of Indian early-onset breast cancer patients in were found to have BRCA1/2 mutations. Higher frequency of BRCA2 mutations than BRCA1 were reported in India(90). Although most of studies have shown distinct sequence variants both in BRCA1 and BRCA2,some of them being unique to Indian population however presence of 185delAG founder mutation reported in Ashkenazi Jews has also been reported by many Indian breast cancer patients (88, 91-94). Other gene mutations like ATM, TP53, CHEK2, PTEN, CDH11 and STK11 are also associated with breast cancer (95-97). Family history of cancer can enable identification of high risk groups History of cancers

among family members helps in identifying high risk groups, who can be further counseled for early diagnostic modalities and preventive therapies.

2.10 Molecular basis of breast cancer

Breast cancer, like other malignancies progresses by accumulation of a series of genetic and epigenetic changes in the genes regulating cellular proliferation, differentiation, cell death (apoptosis or necrosis), DNA repair and tissue compartmentalization, etc. The complex processes setting the hallmarks for cancer (Fig. 2.6) includes unlimited replicative potential, sustained angiogenesis, self-sufficiency in growth signals, insensitivity to growth inhibitory signals, evasion of apoptosis, tissue invasion metastasis as well the ability of the cancer to escape the immune response through several complex processes and events (26, 98-99). Further breast cancer stem cell identification led to elucidation about evolution and progression of breast cancer (100). Molecular basis of cancer evaluates biomarkers which enables identification of disease state and pathways to stake out disease progression. With the advent of gene expression profiling studies, our understanding of the molecular mechanisms involved in tumourigenesis and progression of breast cancer has improved. The basis of this heterogeneity lies in genetic variability (101-104). Early marker identification preceding clinical manifestation for disease is an important objective of molecular epidemiology (105). These genes may be classified into two main categories: (i) Oncogenes and ii) Tumour suppressor genes.

2.10.1 Oncogenes

Mutations, leading to gain of function, results in activation of proto oncogenes. Activated oncogenes start behaving in a dominant fashion leading to neoplastic development (106-107). Breast cancer genome analyses indicate that some genes are frequently mutated and

some get infrequently mutated, causing heterogeneity. Oncogenes are dominant in nature, since a single "hit" or alteration activates them. They get amplified and their protein products are over expressed leading to increase in product presence (107).They cause sporadic cancers, accounting for the majority breast cancers. To name a few important oncogenes that are reported to be involved in breast cancer are Erb-b2, Cyclin D1, Bcl2, MYC and H-RAS.

2.10.2 Tumour Suppressor Genes

Tumour suppressor genes are those that sustain loss-of function mutations in the development of cancer (108). They regulate vast number of cellular activities, like cell cycle check-point, repair of DNA damage, protein ubiquitination and degradation, mitotic signaling, and tumour angiogenesis. These genes stop cell from dividing during cell division, especially if DNA is damaged during its replication (109). Tumour-suppressor genes have three basic properties (a) they are recessive and undergo biallelic inactivation in tumours, (b) inheritance of a single mutant allele accelerates tumour susceptibility, because one additional mutation completes the loss of function, hence a germline mutation can be the underlying cause of a familial cancer syndrome that exhibits an autosomally dominant pattern of inheritance and (c) the same gene is frequently inactivated in sporadic cancers (110). They get deleted either physically or by loss via recombination or via hypermethytlation. TSGs involved in the breast cancer development includes, TP53, RB and PTEN, along with the susceptibility genes BRCA1 and BRCA2.

2.10.3 Growth Factors

GFs act as compact polypeptides, binding to transmembrane receptors leading to kinase activity, which stimulates specific combinations of intracellular signalling pathways, mitogen-

activated protein kinase (MAPK), the phosphatidylinositol 3-kinase (PI3K), phospholipase C-γ, and transcription factors like the signal transducers and activators of transcription (STATs) or SMAD proteins. A somatic mutation may initiate survival and growth advantage to a cell which may clonally expand in presence of growth factors like IGF1, EGF etc. Further this expanded cell goes through the process of invasion which involves loss in epithelial polarity, and acquisition of mesenchymal like motile phenotype. This critical phase of development is controlled by interplay between oncogenes and tumour suppressors, and GFs. Intravasation and extravasation of cancer cells through lymphatic and blood vessels leads to dissemination and metastasis of cancer to distant organs. The characteristic of cancer cells to acquire new mutations and produce GFs (autocrine loops) along with persistent angiogenesis, all together leads to establishment of large size tumours. The final phase involves large movement of tumour mass to target organs called metastasis. Some growth factors as EGF, epidermal growth factor, CSF-1, colony stimulating factor 1; HB-EGF, heparin-binding EGF, FGF, fibroblasts growth factor; NRG, neurogulin; TGF, transforming growth factor; VEGF, vascular endothelial growth factor plays a crucial role.

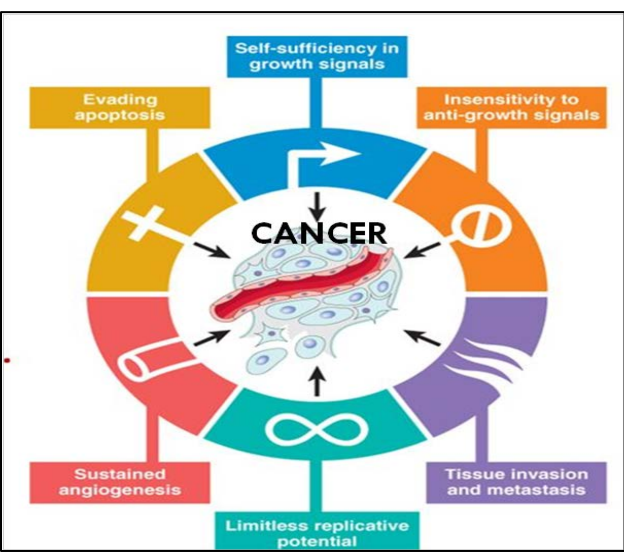


Fig. 2.6 Hallmarks of cancer (*Hanahan and Wienberg, 2011)(27))*

2.11 Genome wide studies on breast cancer

The genome wide association studies involve rapid screening of markers across the complete sets of DNA or genome to find out the genetic variation associated with diseases which can be further utilized to develop better strategies to detect, treat and prevent disease and also lead to the era of personalized medicine. The high throughput tools used for such studies are microarrays. Microarray is defined as a collection of microscopic features (most commonly DNA) which can be probed with target molecules (proteins, DNA, RNA, antibodies, carbohydrates or other chemical compounds) to produce either quantitative (gene expression) or qualitative (diagnostic) data.(111) There are various types of microarrays which are distinguished based upon the solid surface support used by them, nature of the probe, and the specific method used for probe addressing or target detection There are mainly two types of microarrays , Solid phase array where tiny spots of DNA termed as probes are spotted on a glass, plastic or silicon plate (112) which are used to bind to specific DNA sequence in test sample and Polystyrene bead array where DNA probes are placed on tiny plastic beads which can be mixed with test sample of DNA. Hundreds to thousands of DNA probes can be placed on single array with the help of robotic machine. A microarray chip mainly consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, each having a small amount probes which hybridize to cDNA or cRNA. To identify the relative abundance of nucleic acid sequences in the sample hybridization is detected and quantified by fluorescence-based detection. A microarray can be DNA microarray, Protein microarray, or Tissue microarray, with single or double channels. Microarray’s are used for detection and measuring gene expression at the mRNA or protein level, finding mutations, identifying copy number variations, miRNA profiling, methylation and histone profiling and locate chromosomal changes (CGH =comparative genomic hybridization) along with several other applications like gene discovery, disease diagnosis, drug discovery etc.

The, gene expression profiling by microarray enables simultaneously access to the expression profile of thousands of genes which enable understanding the multiple molecular events and mechanisms by creating a global picture of cellular functions for cancer development. Advanced technologies make use of expression Bead Chips to provide genome wide transcriptional coverage of well characterized genes. The oligos covalently attached to beads in Bead Chips having a 29-base address attached to 50-base gene-specific probe. This address helps in decoding the array, while the probe quantifies expression levels of transcripts. This probe is hybridized to labelled nucleic acid derived from total RNA (Fig.2.7.).

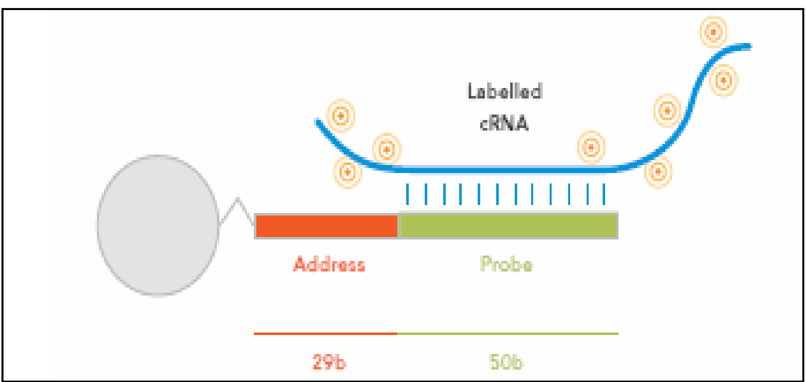


Fig.2.7. Bead image showing its address bar and a probe bar which in turn attaches to labelled cRNA (Adapted from Illumina)

Use of gene expression profiling in identification of breast cancer molecular subtypes can enable development of prognostic, predictive molecular signatures leading to better understanding of the biologic heterogeneity of breast cancer (18).

2.12 Molecular classification of breast cancer via gene expression profiling

Gene expression profiling using microarrays provided an opportunity to perform more individualized and detailed breast tumour characterization by identifying signature gene lists with potential for prediction of clinical outcome. It has divided breast cancer in four main molecular classes (8-11, 50). The "intrinsic" classification by Perou *et al*. proposed these four

classes as **basal-like** breast cancers, corresponding to ER/PR/ HER2/neu-negative tumours (also called as triple-negative" tumours); followed by **luminal-A** cancers, which consists of ER-positive and low-grade histologically; **luminal-B** cancers, which are ER-positive having low levels of hormone receptors and are high-grade; and **HER2/neu -overexpressing** cancers, showing amplification and high expression of the ERBB2 gene and several other genes of the ERBB2 amplicon. This subgroup correlates to clinical characterization on the basis of ER and HER2/neu status, as well as proliferation markers or histologic grade (12-13). The ERBB2 group of patients responds well to Herceptin. In 2001 data was reanalyzed by Sorlie *et al.* (52) on 85 tissue samples. Clinical outcomes (metastasis, death and survival) were found to be significantly different between the subtypes. Gene sets corresponding to each subtype were identified from the dataset which were further used for classification. Sorlie *et al*. added 38 new breast cancer tumour tissue arrays and rexamined 84 of the 85 arrays (9). The results were same since they found once again the same subtypes with significant differences in overall survival between the groups which consequently led to the conclusion that all breast cancers are not the same from molecular point of view. Further studies reported that these large-scale gene expression differences existing between ER-positive (mostly luminal-like) and ER-negative (mostly basal-like) cancers suggests that further molecular subsets also exists (9-10) varying in terms of prognosis and chemotherapy sensitivity. Luminal type cancers are the most favourable long-term survival group, whereas basal-like and ERBB2-positive tumours are more sensitive to chemotherapy (52, 113). Van't Veer *et al*. compared expression profiles of patients developing distant metastasis (n=34) within 5 years with patients who remained disease free (n=44) for at least 5 years (114-115). A 70 gene expression signature was developed which classified tumours into the good and poor prognosis groups. The results were later confirmed in a larger set of tumours (116-117). The summary of all microarray studies is represented below (Table.2.5).

Table.2.5 Summary of initial microarray studies related to human breast cancer

Study	**Findings/Conclusion**
Perou *et.al*. (2000) (50)	Classified tumours into 4 major subtypes alongwith a novel subgroup namely basal-like breast tumours was identified
Sorlie.*et.al*.(2001) (8)	Identification of novel luminal-type subclasses lumina A, luminal B and Basal-like, ERBB2, luminal subtype B possesd worse clinical outcome while luminal A subtype had a good clinical outcome
Sorlie *et.al*.(2003) (118)	Classification of breast cancer based on gene expression profiling captures the molecular complexity of tumours
Ma *et al*. (2003) (119)	Identification , in contrast to tumour stage, different tumour grades which are associated with distinct gene expression signatures
Zhao *et al.(*2004) (120)	Gene expression profiling has revealed distinct patterns of gene expression among ILCs and IDCs
West *et al.* (2001) (121)	Gene expression profile could predict metastatic potential even in the absence of reportedly positive nodes
Van‘t Veer *et al.(*2002) (114)	Signature that defines ER status can be used to decide on therapy and signature that reveals BRCA1 status may further improve diagnosis of hereditary breast cancer.
Van de Vijver *et al.* (2002)(116)	Classification of patients into high-risk and low-risk subgroups on basis of prognosis profile - useful means of guiding adjuvant therapy in patients with lymph-node–positive breast cancer
Huang *et al.(*2003)(122)	Group analysis of lymph-node risk defines metagene patterns that can accurately predict high- risk versus low-risk cases, in both internal and external validation studies
Ma *et al*. (2004)(123)	HOXB13:IL17BR expression ratio may be useful for identifying patients appropriate for alternative therapeutic regimens in early-stage breast cancer
Chang *et al*.(2005)(124)	Both overall survival and distant metastasis-free survival diminished in patients whose tumours expressed the wound response signature. Wound response signature improves risk stratification independently of known clinico- pathogenic risk factors.
Dai *et al*. (2005)(125)	By combining ER expression level and age, identified group of patients with relatively poor outcome Suggest that cell proliferation is the driving mechanism associated with poor outcome
Farmer *et al*.(2005)(126)	Divided mammary tumour cells into 3 groups based on steroid receptor activity: luminal (ER+ AR+), basal (ER- AR-) and molecular apocrine (ER- AR+)
Miller *et al*. (2005)(127)	Showed the primary importance of p53 functional status in predicting clinical breast cancer behavior
Nadari *et al*. (2007)(128)	Identified a prognostic classifier and gene set, which predicts the overall survival in three independent studies

Although carcinoma breast is predominantly a disease of aging, with only 5 to 7% of patients diagnosed below the age of 40 years in the developed world (129) but the developing countries show higher percentage of younger patients (20%) below 40 years of age (130-131). Younger patients tend to expresses key biomarkers differentially as absence of endocrine receptors, HER2/neu and high proliferation index. Younger women mostly presents with more aggressive subtypes. Young patients with luminal-B tumours had worst outcome compared to older (16, 132). These collective findings suggest that tumours arising in younger patients may have a different molecular pathology.

In a study by Azim and colleagues (16) 3,522 patients tumours were evaluated, of which 451 were ≤40 years age breast cancer diagnosis. Higher proportion of basal like breast tumours were shown by young people (34.3%) compared to age 41 to 52 (27.7%), 53 to 64 years (20.8%) and ≥65 years (17.9%) ($P < 0.0001$). HER2/neu-enriched tumours were also present in young patients. Luminal-A tumours were lesser in early onset tumours (17.2%) compared in other age groups 30.7%, 35.1% and 35.4% ($P < 0.0001$). Many studies based on the expression of ER and HER2/neu have discussed the differences in breast cancer subtypes according to age (<40, 40 to 49 and ≥50 years) (133).

Anders in 2008 (14), firstly attempted to describe the breast cancer biology in young women through gene expression profiling. Study included 200 patients having age ≤45 years compared to 211 patients having ≥65 years age. It was found that PI3K and Myc pathways were top deregulated in early onset patients. Later on after adjusting these datasets for potential differences a metanlysis was done (15), where no distinct signatures could be identified amongst 2 groups. In 2012 Azim and colleagues (16) metanalysis on gene expression identified 50 genes based on literature search related to early-onset breast cancer. RANKL was found to express significantly along with c-kit and BRCA1 mutation signatures

typically present in young breast cancer females. They also reported MAPK and PI3K and apoptotic pathways deregulation (134-135).

Studies suggest that young patients tend to have BRCA1 mutations and are mostly basal like, while previous literature suggested them to be of luminal origin regulated by c-kit (136-137). RANKL stimulates osteoclastogenesis and thus targeting it leads to reduction in risk of osteoporosis and related skeletal events secondary to bone metastases (17, 138-139). The potential advantage of improving tumour classification by expression profiling has been central to several large-scale breast cancer studies over the past few years.

It has been now proved that metastatic potential of tumour is the property of whole tumour (119, 140). These findings resulted in need of studies which aim at identifying genes which may be involved in metastasis, relapse and shorter survival prognostic models have been developed to predict long term relapse (114, 122, 141). As a result of continuous efforts and attempts prediction models have been developed which have been converted to diagnostic assays, approved by US Food and Drug Administration FDA for breast cancer patients. Apart from predicting clinical outcomes, these kits can also indicate therapeutic options for patients.

2.13 Gene expression in diagnostics

Gene-expression profiling technology aims providing better prediction of clinical outcome than the traditional clinical and pathological parameters. Gene predictor tests aid clinician to estimate outcome with local treatment along with the absolute benefits expected from systemic adjuvant endocrine therapy and chemotherapy. However, the challenge is to identify the low-risk patients who don't need adjuvant chemotherapy. To date, only two assays have received FDA clearance, while seven tests are marketed in the European Union and the US.

2.13.1 Mammaprint

A 70 gene classifier was developed by Van't Veer and colleagues called the MammaPrint dx signature. It accurately distinguishes patients who would likely remain free of distant metastases (good profile) from breast cancer patients at high risk of developing distant metastases (poor profile) within the 5 years after diagnosis.(116) This test is approved by US FDA as a prognostic assay for breast cancer patients who are 61 years or younger having stage I or II with positive or negative lymph nodes (142). ER positive cancers show a good prognostic risk discrimination but ER negative cancers have high risk score The prognostic risk discrimination is good among ER-positive cancers who benefit by chemotherapy (132, 143) but almost all ER-negative cancers are classified as high risk group.

2.13.2 Oncotype DX (Oncotype DX Breast Cancer Assay)

Currently, the most widely used prognostic assay is Oncotype DX (Genomic Health, Redwood City, CA, USA) for ER-positive cancers suggesting how likely the woman is going to benefit from chemotherapy in addition to tamoxifen therapy. This assay comprises of 16 cancer related genes and five housekeeping genes which utilizes expression measurements to compute a recurrence score from 0 to 100, which can be categorized into low-risk (score <18), intermediate-risk (score 18 to 30) or high-risk (score ≥31) groups corresponding to 10-year distant recurrence rates after 5 years of tamoxifen therapy to be <12%, from 12% to 21%, and from 21% to 33%, respectively (144). The commercially available Oncotype DX Breast Cancer Assay which uses reverse transcriptase-polymerase chain reaction (RT-PCR) evaluates the mRNA expression levels of 21 genes (16 cancer-related genes and 5 reference genes 8). These genes are main components of the ER pathway (ER, progesterone receptor, BCL2 and SCUBE2), proliferation (Ki67, STK15, Survivin, CCNB1 and MYBL2), HER2/neu amplicon (HER2/neu and GRB7), invasion (*MMP11* and CTSL2) and GSTM1,

CD68 and BAG1. The expression of these 21 genes is reported as a single Recurrence Score (RS).

2.13.3 Theros-Breast Cancer Gene Expression Ratio Assay

Theros is based on 3 predictive genes: the homeobox gene HOXB13, interleukin 17B receptor (IL17BR) and EST AI240933, which was identified in a microarray study conducted by Ma and co-workers. It was specifically developed for ER-positive breast cancer patients treated with tamoxifen.

2.13.4 Genomic Grade Index (MapQuant DX)

It is the first clinically validated molecular diagnostic test based on microarray which measures tumour grade as a indicator of tumour proliferation, metastatic risk and chemotherapy response. It allows distinction of grade2 tumours into either grade1 or grade 3. Higher GGI suggests increased sensitivity to neoadjuvant therapy with paclitaxel, fluorouracil, adriamycin, and cyclophosphamide in ER positive and negative patients; however it predicts worse survival in ER positive patients only. Hence this test implies the importance of tumour differentiation and proliferation genes in ER positive group of patients.

2.13.5 PAM50/Breast Bio Classifier

Patients outcome are predicted by a 50-gene qRT-PCR assay which classifies ER-positive and ER-negative breast cancers into various subtypes which predict patients outcome (high, medium and low risk groups). It generates a risk score which enables clinicians to take treatment decisions based on estimates of death risk. Consequently, chemotherapy can be avoided in case of a very good prognosis.

Hence the expression profiling has led to better understanding of heterogeneous nature of breast cancer even at genomic level and also prediction of clinical outcomes. Some of the validated and commercially available profiles include OncotypeDx® and Mammaprint® assay.

2.14 DNA Methylation

Epigenetic mechanisms have a potential role in cancer, mainly histone modifications & DNA methylation. DNA methylation is a heritable epigenetic mechanism regulating gene expression, which involves transfer of methyl group to the C-5 position of the cytosine ring of DNA by DNA methyltransferases (DNMTs) hence modifying the function of the genes (20). In human DNA methylation occurs at cytosine residues which form part of CpG dinucleotides (19), with some methylation occurring at CpNpG sequences, where N can be A, T or C. 98% of DNA methylation is known to occur in a CpG dinucleotide and around a quarter of all methylation occurs in a non-CpG sites in embryonic stem cells (ESCs) (19).Methylation patterns are removed during time of zygote formation and gets restablished at time of embryo implantation (145). DNA methylation plays a vital role in genomic imprinting, X-chromosome inactivation, and suppression of repetitive element transcription and transposition. It may also contribute to cancer when dysregulated (20-21). Approximately 41% of the average human genome content is G/C nucleotides which is different from the statically expected frequency (20%). This is mainly accounted to the fact that cytosines in CpGs becomes methylated, resulting in 5mCpG, which is highly susceptible to spontaneous deamination, resulting in thymine. Further DNA replication, repairs the mismatched T/G to T/A, hence producing a permanent alteration/mutation within the DNA sequence. As a result of this, the mutation rate of cytosine to thymine is much higher than the average nucleotide, estimated to be 8.5 times higher, resulting in the loss of many CpG dinucleotides over time.

High levels of these alterations are seen at CpG islands which are long repetitive sequences and regions which are often unmethylated.

2.14.1 Types of DNA Methylation

2.14.1.1 CpGs

DNA methylation generally refers to the addition of a methyl group to the 5’ carbon ring of cytosine in CpG (cytosine-phosphodiester bond-guanine) dinucleotides (Fig.2.8). The enzymes responsible for DNA methylation are called DNA methyltransferases (DNMTs). *DNMT3A*, *DNMT3B*, *DNMT3L* perform *de novo* methylation while *DNMT1* maintains DNA methylation by adding methyl groups to newly replicated DNA strands.

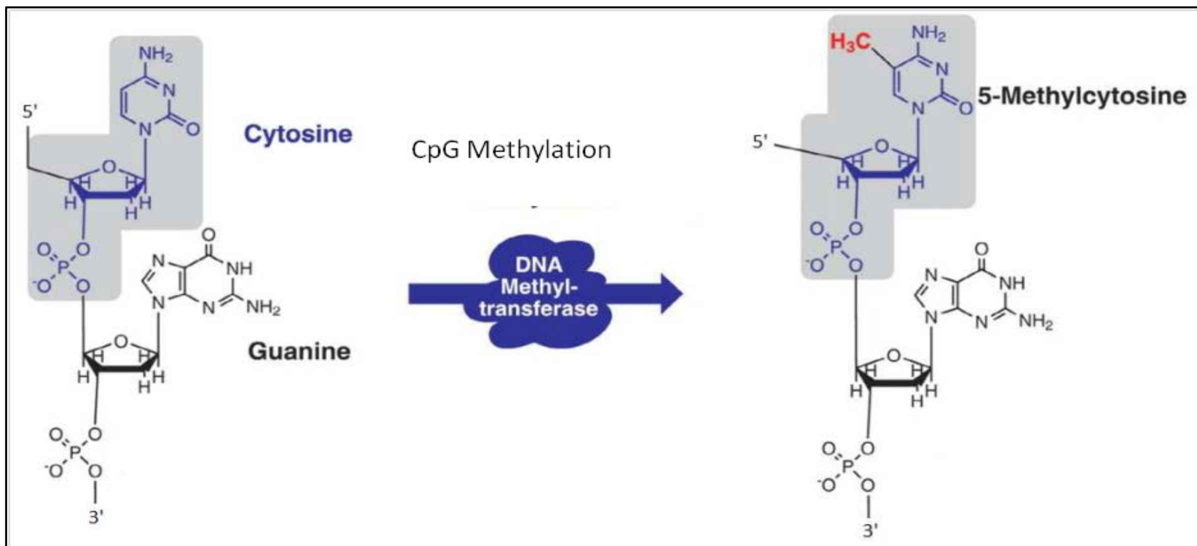


Fig. 2.8 Mechanism of DNA Methylation, a cytosine of the CpG dinucleotide is methylated via chemical modification facilitated by DNA methyltransferase adding a methyl group (CH3, depicted in red) to its 5’ carbon ring**.**

2.14.1.2 CpG Islands and Promoters

About 1% of human genome is comprised of highly dense regions of CpGs called CpG islands. with > 50% G+C concentration, > 200bp in length, and > 0.6 in the observed vs

expected ratio of CpG dinucleotides based G and C density (146). CpG islands are around 1kb in length and are usually devoid of DNA methylation which in turn protects them from the highly mutable events of 5mC deamination. Hence DNA sequences underlying CpG islands are evolutionarily conserved. Approximately 60 to 70% of gene promoters contain CpG islands, and these promoter CpG islands are known to regulate gene expression (147). DNA methylation at gene promoters usually is accounted to transcriptional silencing. Housekeeping genes often contain CpG islands at their promoters which lack methylation and hence these promoter CpG islands express ubiquitously. On other hand promoters of genes devoid of CpG islands are usually expressed in differentiated cells in a cell-type specific manner (148). CpG island shores are regions which are up to 2kb away from island, and shelves which are up to 4 kb away from the island, are sometimes suggested to be more functionally relevant than CpG islands themselves. Moreover, multiple non-malignant tissue types as well as cancerous cells were distinguishable from one another using only the methylation profiles of CpG island shores.

2.14.1.3 Repetitive Elements

Human genome consist of approximately 45% of repetitive elements, encompassing 52% of all CpGs. Somatic cells mainly consist of methylated repetitive elements which suppress their activities (e.g. retrotransposons) and maintain genome stability and integrity (149). Furthermore 25% of repeat elements in the human genome were suggested to play functional roles in cellular pluripotency and differentiation. Specifically, repeat element methylation was associated with genes that were active in human ESCs but are silent in differentiated cells.

2.14.2 Mechanism of DNA methylation

DNA methylation is accomplished with two classes of DNA methyltransferase enzyme namely DNMT1 and DNMT3a/DNMT3b. Though both of them uses, adenosyl methionine to methylate, but these two classes target different states of DNA.

2.14.2.1 DNMT1

DNMT1 methylates hemimethylated DNA on the nascent DNA following replication and hence it's also called as maintenance methyltransferase (150), It thus provides the mechanism by which methylation status of DNA is maintained in daughter cells. DNMT1 also maintains DNA methylation pattern in proliferating cells.

2.14.2.2 DNMT3a and 3b

DNMT3a and 3b are called as de novo methyltransferases which are required to establish new DNA methylation patterns. They interact with CENPC protein resulting in increased misalignment and segregation defects due to loss of protein during mitosis (151).

2.14.3 DNA Methylation and gene expression

DNA methylation can regulate gene expression in several ways, either directly or indirectly. Promoter CpG island methylation leads to downregulation of gene expression in cancers (152). Methylation causes X chromosome inactivation early in embryogenesis or genomic imprinting of either paternal or maternal allele, resulting in single allelic expression (153). Repression of expression occurs by either blocking of transcription factors or due to actions of methyl DNA binding protein (154). Directly, DNA methylation can physically block the binding of transcription factors. For example, the binding of N-Myc to the promoter of

epidermal growth factor receptor (*EGFR*) is hindered by DNA methylation of N-Myc binding site E-box (CACGTG).

2.14.3.1 Hypermethylation and Gene Silencing

Human genome expects the overall prevalence of CpG dinucleotide pair to be around 1% which is actually lower than the expected frequency (6%). High density concentrations are however present in promoter regions called as CpG islands extending between several hundreds to few thousands mainly associated with tumour suppression. These are usually unmethylated, but if they get methylated, they fail to transcribe downstream genes resulting in silencing of that gene.

2.14.4 Aberrant DNA methylation changes in cancer

2.14.4.1 DNA hypomethylation

Cancer consists of reduced or loss of genome wide methylation level/hypomethylation leading to global hypomethylation. The hallmark associated with hypomethylation occurs at repeat sequences of Alu and LINE1 elements juxtacentromeric satellite2 and tandem centromeric satellite alpha regions, leading to tumourigenesis due to genomic rearrangements and activating transposable elements. Loss of methylation can also lead to activation of proto-oncogenes such as *IGF2* in Wilms' tumour and *HOX11* in leukemia (155-156).

2.14.4.2 DNA hypermethylation

In 1986 DNA hypermethylation was known to be a form of gene inactivation. Down regulation of calcitonin gene in small cell lung cancer was accounted to its hypermethylation (157). Further other tumour suppressor genes such as *RB1* (retinoblastoma 1) in retinoblastoma, *VHL* (von Hippel-Lindau) in renal cell carcinoma and *CDKN2A* (*cyclin-dependent kinase inhibitor 2A*) were also methylated in multiple tumour types (158) resulting

in inactivation of certain tumour suppressor genes. Since then, many genes have been shown to be inactivated in this manner. Other genes as *CDH1* (E-cadherin), *RASSF1A* (*Ras association (RalGDS/AF-6) domain family member 1*) and *HIC1* (*hypermethylated in cancer 1*), were inactivated due to methylation besides deletion. RASSF1 is commonly methylated in lung, breast, glioma, colorectal and RCC. *HIC1* gene methylation was first observed in leukaemias as well as breast cancers (159) along with in multiple malignancies. For these reasons, DNA methylation has been used as a biomarker for the detection of tumours, assessment of tumour prognosis, and prediction of treatment responses (160).

2.14.4.3 Imprinting Disorders

Loss or gain of methylation can lead to imprinting disorders which is an unusual event that is associated with dysregulation of growth (e.g. Beckwith-Wiedemann syndrome or BWS), neurodevelopment (e.g. Angelman syndrome) or metabolism (e.g. transient neonatal diabetes mellitus or TNDM). Abnormal imprinting via methylation is sometimes due to an underlying genomic change. For example, patients can have both copies of one chromosome partially or entirely inherited from only one of the parent, a phenomenon known as uniparental disomy .

2.14.5 Aberrant DNA methylation in breast cancer

Epigenetic modifications shape the chromatin topology which determines the cellular phenotype and controls its gene expression patterns (161). Aberrant DNA methylation occurring in CpG islands in the proximity of promoters is responsible for the loss of gene function (162). A large number of genes have been shown to be either hypermethylated or hypomethylated in breast carcinomas (163). It has been found that cancer genomes are generally hypomethylated and this demethylation event is mostly associated with repeat regions rather than specific genes (23). *IL-10*(*Interleukin 10*), *MDR1* (*multidrug reistance 1*)

FEN1 (flap structure-specific endonuclease 1), *CDH3* (cadherin 3), *NAT1* (N-acetyltransferase1) are all reported to be hypomethylated in breast cancer (164-165). Hypomethylation of *TLR9* (Tol-like receptor 9) gene results in it's over expression leading to enhanced migratory ability of cells (166). Other classical tumour suppressor genes commonly methylated are *CDKN2A* (cyclin dependent kinase inhibitor2A), *BRCA1* (breast cancer1), *CCND2* (cyclin D2), *RASSF1A* (Ras association (RalGDS/AF-6 domain family member 1 isoform A)*, APC* (adenomatous polyposis coli) *and RARβ* (retinoic acid receptor β). *RASSF1A* is known to be methylated in 62% of breast cancers *CDKN2A and CCND2* are cell cycle tumour suppressor frequently methylated in breast cancer correlating with low expression at the mRNA and protein levels. *APC* another tumour suppressor gene involved in Wnt signalling is tumour-specifically methylated in 36% of breast cancers (167). *RARβ* is a commonly methylated gene in many cancers, including invasive and in situ breast carcinomas (161). Gene expression analysis of breast tumours has identified different breast cancer subgroups belonging to estrogen receptor (ER) negative basal-like and the ER positive luminal subgroups (50) with differences in outcome (8). In DCIS and early invasive breast cancer, FOXC1, GSTP1, ABCB1 and RASSF1A methylation were found to be associated with TP53 status and FOXC1, ABCB1, PPP2R2B and PTEN methylation to be associated to ER status (168). In locally advanced breast cancer patients treated with doxorubicin, methylation of GSTP1 and ABCB1 were found to be independent prognostic factors and we found DNA methylation patterns in single genes to be associated to TP53 and ER status as well as tumour expression subtypes (169). A concomitant methylation status was observed for several genes although residing on different chromosomes.

2.14.6 Analysis of DNA methylation

Analysis of DNA methylation can be done at individual gene level, or whole chromosome level or entire genome level. Most commonly used method to study DNA modification by methylation is by sodium bisulphite treatment. Denatured DNA when treated with sodium bisulphite leads to deamination of cytosine to uracil while 5-methylcytosine (5mC) remains unaffected. This is followed by PCR which incorporates T nucleotide wherever deamination has occurred and the methylated cytosine remains unaffected and incorporates C only. Bisulfite converts equally both 5-hydroxymethycytosine as well as 5-methylcytosine without any discrimination (170). Many techniques since then have been developed to assess the methylation levels in particular DNA sequences. Presently Illumina; GoldenGate, Infinium HumanMethylation27 and Infinium HumanMethylation450 are available arrays for methylation. They use bead chip technology and allow directly hybridizing the bisulphite modified DNA to the array. Correct binding of methylated or unmethylated DNA to the bead chip results in single base extension with the addition of a fluorescently labelled nucleotide. For each probed CpG locus, numerous methylated and unmethylated probes are scattered throughout the chip and measurement of the fluorescence of all the probes for each locus bead type gives a quantitative representation of the amount of methylated and unmethylated alleles in the analysed sample.

2.14.7 Clinical relevance of DNA methylation

DNA methylation mainly enables understanding of the molecular events leading to the evolution of cancer for the benefit to the clinician majorly to improve the accuracy of the diagnosis of cancer, prognostic information about the cancer, and offer a potential means for cancer therapy.

2.14.7.1 Diagnosis of Early Cancers

There is a urgent need to diagnose patients with cancer at a initial or preclinical stage. Diagnosing a disease early i.e before the invasion, lymph node metastasis, leads to improved surgical resection and has lesser side effects which is beneficial for the patients. The large population screening and surveillance programs are trying to identify these high risk patients. Presently the methods used for investigating cancers include serum tumour markers, radiologic procedures or endoscopic tests which are less sensitive or specific and expensive. Diagnosis of equivocal lesions in asymptomatic patient's leads to further series of investigations and surgery which might or might not be necessary or beneficial to patients. For example, screening mammography programs tend to indentify ductal carcinoma *in situ* incidentally which have a high risk of developing into invasive cancer, but with presently it's impossible to predict with any certainty which patients will have malignant disease and require surgery, versus which will remain quiescent, making potentially hazardous intervention less justifiable.

Methylation could prove as a potential tool for diagnosing premalignant lesions, since promoter regions methylation proves to be a consistent early event in the incipient cancer cells. It would allow accurate screening and surveillance by identifying higher-risk patients on a molecular basis. It would also provide a justification to all patients who need more definitive treatment associated with malignancy at molecular level rather at typical pathologic or microscopic level. Focal resection or abalation in such patients would prevent subsequent cancer in such patients. Surgically managing these cancers at a stage before invasion and metastasis involving minimum invasive procedures would be associated with lesser complications and benefit the patient.

2.14.8 Treatment and management of breast cancer

There are multiple modalities available for the treatment of breast cancer. It consists of both local as well as systemic treatment. Local treatment includes surgery or radiation while systemic treatment includes chemotherapy, radiotherapy, endocrine or targeted therapy.

2.14.8.1 Surgery

The first line of attack against breast cancer is surgery. Surgery can be either breast conserving surgery (BCS)/lumpectomy or mastectomy. When lumpectomy is performed only tumour tissue and its surrounding normal tissue is removed depending on size and tumour position. The tumour margins comprising of normal tissue should be pathologically free of cancer. Women undergoing BCS are later on subjected to radiation therapy too. If the whole breast is removed it is termed as Mastectomy. It could be subcutaneous mastectomy in which nipple and areola are left as such and only breast tissue is removed. On other hand if the entire breast along with level I and II axillary lymph nodes are also removed then it is Modified radical mastectomy. If the entire breast with level I, II, III axillary lymph nodes and the pectoralis muscle all are removed it is termed as Radical mastectomy.

2.14.8.2 Radiation Therapy

Radiation therapy is performed after surgery to ensure local control of clinically undetectable cancer cells. Radiation decreases the chance of breast cancer recurrence around 70%. Sites for therapy may consist of breast/chest wall or the axilla. Radiation therapy is done after breast conserving surgery or mastectomy in the patients with either a large tumour (greater than 5 cm) /with cancer in lymph nodes or cancer with chest wall involvement. The stimulation should be CT guided to insure even and maximum distribution of radiation to the tumour and minimal toxicity to the surrounding normal tissue. Therapy can be of traditional external radiations type targeting whole breast. This has to be done for 5-6 weeks with few minutes

every day. Intensity Modulated Radiation Therapy (IMRT) can be done to deliver radiations to whole breast with varying beam length which is assures more accuracy. It takes around four weeks for complete therapy. Hypo fractionated radiotherapy on the other hand has increased dose of radiation for patients in lesser time.

2.14.8.3 Chemotherapy

Chemotherapy, often termed as "chemo," is a systemic therapy affecting whole body through the bloodstream. It can be given before or after surgery and is called neoadjuvant or adjuvant therapy respectively. This therapy uses anticancer (cytotoxic) drugs to destroy and weaken cancer cells in the body present either at original cancer site or any other part of the body. Various drugs used in 1st line or 2nd line of treatment includes Cyclophosphamide, Epirubicin, Fluorouracil (5FU), Methotrexate, Mitomycin, Mitozantrone, Doxorubicin, Docetaxel, Taxotere, Gemcitabine(Gemzar).

2.14.8.4 Targeted Therapy

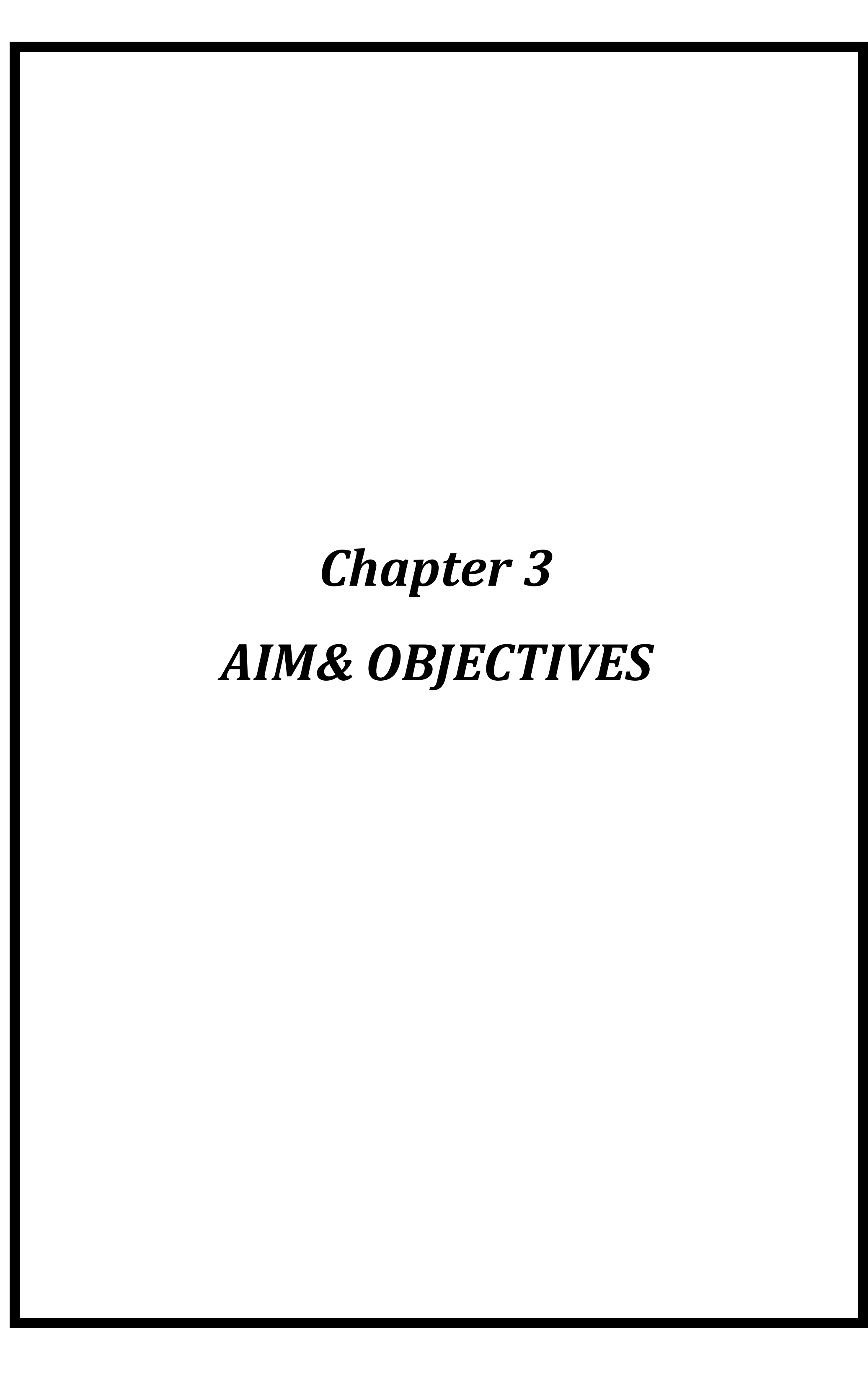
These therapies are meant to target specific characteristics of cancer cells such as protein which makes the cancer cells to grow rapidly and abnormally. They cause little or no damage to normal cells resulting in lesser cytotoxicity. In breast cancer targeted agents are commonly used to block the over expression of HER2/neu. A monoclonal antibody Trastuzumab (Herceptin®) interacts with the HER2/neu receptors on the cell surface and hence preventing them from activating the signaling pathways that causes cell proliferation. Other therapy for HER2/neu positive patients is treatment with monoclonal antibodies as Pertuzamab, Lapatinib.

2.14.8.5 Endocrine Therapy

Hormone therapy slows down the growth of hormone sensitive tumours by either blocking the body's ability to produce hormones by interfering with hormone action mainly used to slow or stop the growth of hormone-sensitive tumours by either blocking the body's ability to produce hormones or by interfering with hormone action. On the other hand hormone insensitive tumours fail to respond to hormone therapy. This either blocks ovarian function by blocking estrogen production or by blocking estrogen's effects. Ovarian suppression drugs as Zoladex/Lupron have been approved by FDA. The activity of aromatase is blocked by aromatase inhibitor like anastrozole or letrozole Selective estrogen receptor modulators (SERMs) specifically bind to estrogen receptors and hence prevent estrogen from binding to them. Some SERMs which are approved by the FDA are tamoxifen (Nolvadex®), raloxifene (Evista®), and toremifene (Fareston®). Since more than 30 years tamoxifen has been used to treat hormone receptor positive breast cancer or metastatic breast cancer.

2.14.8.6 Gaps in existing research

This lack of data from Indian subcontinent where incidence of early onset of breast cancer is on rise made us hypothesized that the underlying mechanism for early onset breast cancer is different from that of late onset and hence we studied the gene expression and methylation profiles of breast cancer patients. Once the profiles in gene expression and methylation are recognized the underlying genetic mechanism in early onset breast cancer in India will be elucidated, and would also assist in development of therapies for the early diagnosis screening and treatment. Therefore, the study focuses mainly to identify differential gene expression and methylation breast cancer profiles in Indian breast cancer patients, to understand the molecular pathogenesis.



Chapter 3

AIM& OBJECTIVES

Aim and Objectives

Aim

To understand molecular biology of breast cancer in Indian women with special emphasis on breast cancer in young Indian women.

The key questions include

1. Does gene expression profile in breast cancer in young women (≤ 40 years) differs from breast cancer in older women (≥ 55years) suggesting different genetic factors associated with both.
2. Does epigenetic factors vary in breast cancer in young women (early onset) compared to breast cancer in older women (late onset)

Objective

1. To study the gene expression profiles in breast cancer in Indian women and compare the differentially expressed genes in early and late onset tumours.

Whole genome gene expression profiling was performed in 29 breast cancer patients of which 12 tumours were early onset and 17 were from late onset patients. A comprehensive analysis was done to identify differentially expressed genes amongst the tumours vs normal (control) and further between early and late onset breast cancer patients, different molecular subtypes and stages. Clustering of differentially expressed genes and patients samples was performed. Gene ontology and pathway analysis was done to perform the functional classification of differentially expressed genes, followed by network buildup to

identify gene networks involved. All statistical analysis for expression data was done using bioconductor R with various packages available such as LIMMA, GENEFU. The gene expression profile was validated using qRTPCR for matrixmetallopeptidase pathway which showed differential expression between tumour and normal (control) tissue as well as early and late onset tumours.

2. To study differential methylation profile in Indian breast cancer women and compare the profiles in early and late onset tumours.

Whole genome methylation profiling was performed amongst 36 breast cancer tumours of which 19 tumours were of early onset and 17 tumours belonged to late onset along with 12 normal (control) tissues. The differential methylation profiles were identified and further they were also compared amongst early and late onset breast cancer patients. The extent of methylation was determined using beta values. The results were further validated using TCGA database. Later on gene expression and methylation was merged to understand the role of epigenetics in cancer. Genes and pathways correlating methylation and gene expression were also identified.

Chapter 4

To study the gene expression profiles in breast cancer in Indian women and compare the differentially expressed genes in early and late onset tumours

To study the gene expression profiles in breast cancer in Indian women and compare the differentially expressed genes in early and late onset tumours

4.1. Introduction

Breast cancer is the most common cancer among women worldwide representing nearly a quarter (25%) of all cancers with an estimated 1.67 million new cancer cases diagnosed in 2012. Women from less developed regions have slightly more number of cases (883 000 cases) compared to more developed (794 000) regions. In India, although age adjusted incidence rate of breast cancer is lower (25.8 per 100 000) than United States of America (93 per 100 000) but mortality is at par (12.7 *vs* 14.9 per 100 000) with United States of America (1). There is a significant increase in the incidence and cancer-associated morbidity and mortality in Indian subcontinent as described in global and Indian studies. In India more than 50% of women develop breast cancer at age below 40 yrs (4) suggesting greater burden of breast cancer due to early onset disease. An analysis of the time trends of breast cancer incidence in India by estimating annual percentage change (APC) for age of women by estimating change in age specific incidence rates (ASR) revealed 4.24% increase in the younger age group women (15-34 years) compared to 1.6 % and 0.8% in age groups 35-44 years and 45-54 years respectively suggesting significant increase in breast cancer incidence mainly in younger age group of patients during the last decade (6, 60). Early age onset breast cancer arising in younger women is characterized by a higher incidence of negative prognostic factors, higher recurrence rates and poorer overall survival despite aggressive therapies (171-175). The etiology of breast cancer is still poorly understood with known breast cancer risk factors which explain only a small proportion of cases. The heterogeneous nature of the disease coupled with the lack of robust markers for prediction, prognosis, and response to treatment has so far eluded our understanding of its complex nature.

Understanding of etiopathogenesis and molecular biology of the early onset breast is very important for disease control. Gene expression profiling is an advanced technology used initially for molecular classification of breast cancer for optimization of disease behaviour and improving therapeutic options. This study is planned to elucidate the gene signatures and molecular pathways involved in breast carcinogenesis in young women. This would help in confirming our hypothesis that breast cancer arising in young women is a unique disease entity driven by complex biologic processes extending beyond hormone receptors and hereditary cancer syndromes.

4.2 **Materials and Method**

4.2.1 **Clinical Specimen**

A total of 97 tumour tissue samples were collected from Safdarjung Hospital and Indraprastha Apollo Hospital, New Delhi, along with 38 distant normal tissues from histopathologically confirmed cases (Infiltrating Duct carcinoma) prior to any chemotherapy or radiotherapy treatment. The samples were staged according to American joint committee on cancer-AJCC. Expression of ER/PR/ERBB2 receptors had been studied in all the cases by immunohistochemistry The study group included 41 cases (42.3%) below 40 years of age (early onset) with a mean age of 35.9 and 56 cases (57.7%) above 55 with a mean age of 63.2 years (late onset). The patients aged between 41-54 yrs, who received any treatment viz chemotherapy or radiotherapy or patients who had reported with any other malignancies besides breast cancer were excluded from present study. Tissues were collected from tumour areas along with the distant normals either after incisional biopsies or (MRM) modified radical mastectomy and were snap frozen (RNA Later, Ambion) and stored at -80 immediately for histopathology, DNA and RNA extraction. Informed consent was obtained from all the patients prior to collection of samples which was approved by the human research ethical committee of Safdarjung hospital and Indraprastha Apollo hospital, New Delhi.

4.2.2 Isolation and purification of RNA from tumour and normal tissue

Isolation

Isolation of total RNA from tissues was carried out using standard protocols by trizol method. 1 mL TRIzol Reagent was added per 50–100 mg of tissue sample and homogenized using pestle and mortar in liquid nitrogen. Following homogenization 0.2 mL of chloroform per 1 mL of TRIzol Reagent was added and centrifuged at 12,000 × g for 15 minutes at 4°C. Aqueous phase was removed by angling the tube at 45° and pipetting the solution out in a new tube. 0.5 mL of 100% isopropanol was added to the aqueous phase, per 1 mL of TRIzol and sample was incubated at room temperature for 10 minutes followed by centrifugation at 12,000 × g for 10 minutes at 4°C.Supernatant was removed from the tube, leaving only the RNA pellet. Pellet was washed, with 1 mL of 75% ethanol per 1 mL of TRIzol Reagent used in the initial homogenization. Sample was vortexed, then centrifuged at 7500 × g for 5 minutes at 4°C. RNA pellet was air dried for 5–10 minutes. It was resuspended in 30µl DEPC treated water.

Purification

For purification Qiagen Rneasy Mini Kit was used. RNA sample was adjusted till total volume of 100µl with RNase free water and 350µl RLT buffer was added and mixed thoroughly. It was followed by addition of 250µl of absolute ethanol. Sample was passed through RNeasy mini spin column and centrifuged at >10,000 rpm. It was further washed by 500µl of buffer RPE and again centrifuged at >10,000 rpm for 1 minute. Wash step was repeated for 2 minutes followed by empty spin. Finally the RNA was eluted in a new tube in 30 µl of RNase free water after centrifuging it at speed >10,000 rpm for 2 minutes.

Quantitative/Qualitative Analysis

For each sample quantity was estimated by measuring RNA concentration and 260/280 nm ratios using a Nanodrop (Invitrogen, Carlsbad, California, USA).All the samples having RNA quantity above 300ng were used for the gene expression study. An approximate concentration of RNA was determined by electrophoresis of 1or 2 µl of RNA solution on ethidium bromide (0.5µg/ml) stained 1.8% agarose gel in 0.5X TBE buffer prepared in DEPC water with 100 bp DNA ladder. RIN was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California). Briefly, all the reagents were incubated at RT for 30 minutes, 550 µl gel was filtered by spinning at 1500g for 10 minutes. Then 65 µl filtered gel was taken and 1 µl of 6000 RNA nano dye concentrated was added and spinned at 13000g, 10 minutes at RT. Further 9 µl of this mix was pippetted in well marked as G in chip which was already primed in priming station and plunger was slowly pressed down until held by clip. After 30 seconds clip was released and gel dye mix was further added in all the remaining wells.5 µl of RNA Nano marker was added in each 12 sample well and in the well for ladder.

4.3 Gene expression microarray

Whole Genome Expression profiling was performed using Sentrix Human 6 and Human HT-12 v3 direct hybridization assay (Illumina, San Diego, CA) in 2 batches. The Human-6 Expression Bead Chip contains six arrays on a single BeadChip, each with >46,000 probes derived from human genes in the National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq) and UniGene databases while HT-12 chip contains 12 arrays on a single chip with more than 47,000 probes. Each array contains full-length 50-mer probes representing more than 48,000 well-annotated RefSeq transcripts, including 25,400 unique, curated, and up-to-date genes derived from the National Center for Biotechnology Information Reference Sequence (NCBI RefSeq) database.

4.3.1 Quantitation of RNA by Quant-iT RiboGreen RNA quantitation kit

A dilution of 1:200 was prepared of ribogreen into 1X TE using kit supplies. 500ng of total RNA was taken initially and converted by *in vitro* transcription step into cRNA, which were biotin labelled using Ambion Illumina Total prep RNA Amplification Kit (Ambion, Austin, TX). Further labelled probes were mixed with hybridization reagents and hybridized to the chips overnight. The labelled probes were then mixed with hybridization reagents and hybridized overnight to the Bead Chips. All protocols were performed following the manufacturer's recommendations. They were further followed by washing, staining and imaging using the Illumina Bead Array Reader which measures fluorescence intensity at each probe. The signal intensity corresponds to the mRNA quantity in the original sample.

Sample Distribution

From the total set of 97 tumours and 38 normal (non tumour controls), 29 tumour samples (12 early onset and 17 late onset cases) were used for gene expression microarray analysis along with 9 distant normal controls (Microarray Set). Further validation of results had been done by real time PCR in 67 samples (33 early onset and 34 late onset cases) (Validation set) and all normal controls i.e 38. Also 19 samples were kept in common between microarray set and validation set (Fig. 4.1a, b) to see the reproducibility of data. Remaining of the samples, along with validation and microarray set all (n=97) together were related analyzed for their association with parameters such as stage, lymph node status, metastasis, breast feeding, IHC staining for determination of ER status, PR status and HER2/neu status, phenotypes. The expression profiles have been deposited in NCBI's Gene Expression Omnibus (GEO) with GSE accession number **GSE 89116**

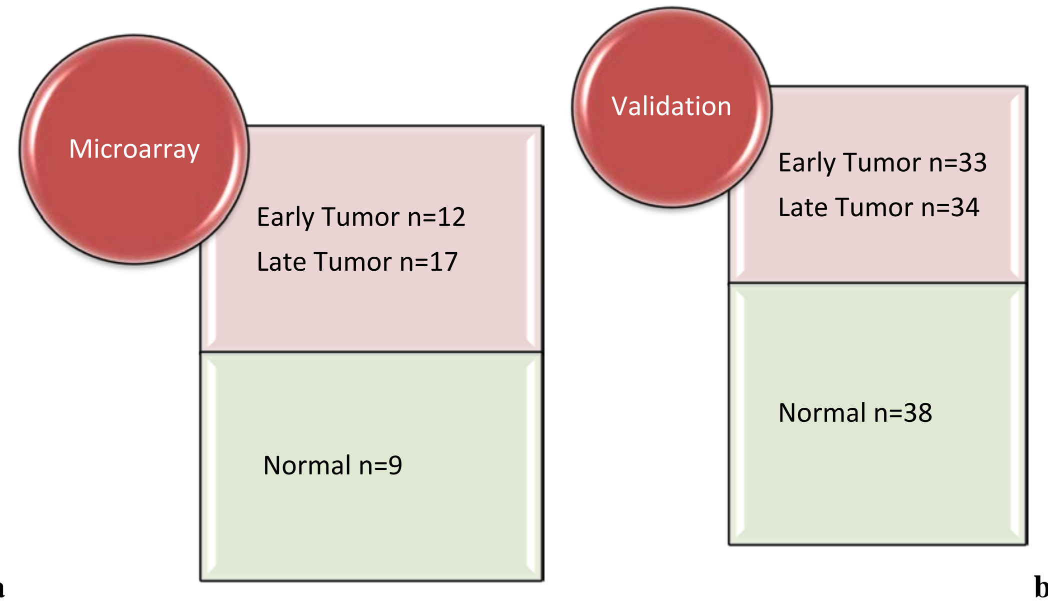


Fig.4.1 Patients sample distribution, for microarray and real time validation (a) Distribution of tissues used for microarray study. Tumour tissues from 29 cases were used for microarray experiment along with 9 distant normals. (b) Real time validation was performed on 67 tumour tissues along with 38 distant normals tissues

4.3.2 Raw data background subtraction and Normalization

Raw data was processed by Genome studio software. We obtained data for around 48,000 probes. In order to identify the probes which were common between the WG-6 and HT-12 chip we merged the data from both the platforms at probe level. Non specific hybridization generated background signals in the regions nearby the main spots. We further used Bioconductor R, LIMMA package for background subtraction and intensity correction by neqc command. This function matches distributions of gene counts across lanes.

True Intensity = Foreground Intensity – Background Intensity

Normalization of the background subtracted data was next step. Normalization is the process of balancing the intensities of the channels to account for variations in labelling and hybridization efficiencies. To achieve this, various adjustment strategies are used to force the distribution of all ratios to have a median (mean) of 1 or the log-ratios to have a median (mean) of 0. It removed the unwanted non biological variation that might existed between chips and microarray experiments. Data was quantile normalized to obtain the gene expression intensity values for all samples individually for 28,504 genes. The data was found to be distributed uniformly after quantile normalization as can be seen in Fig. 4.2 for samples.

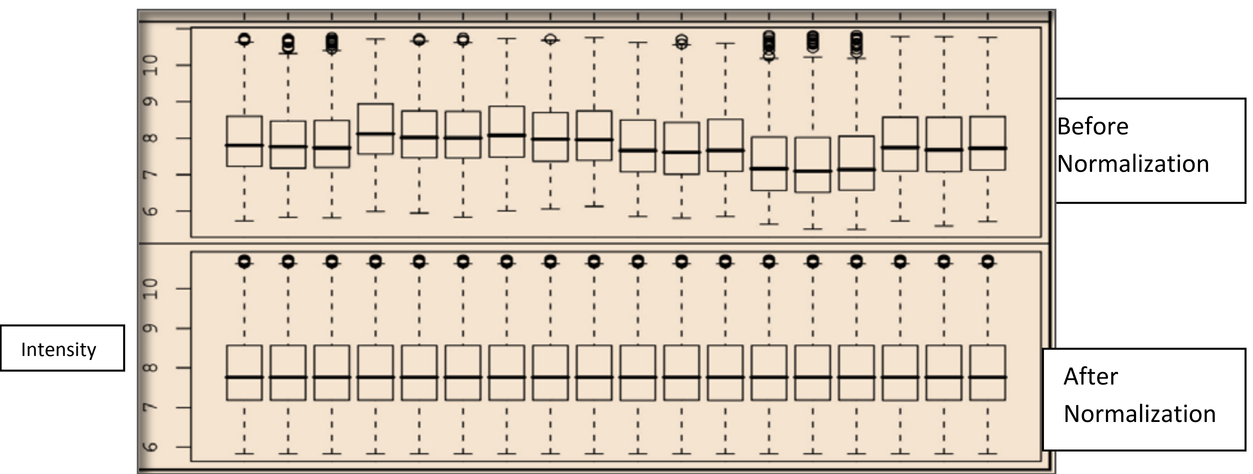


Fig.4.2 Images depicting changes in samples before and after normalization.

Further the normalized intensity values were log transformed for advanced analysis (Fig. 4.3).

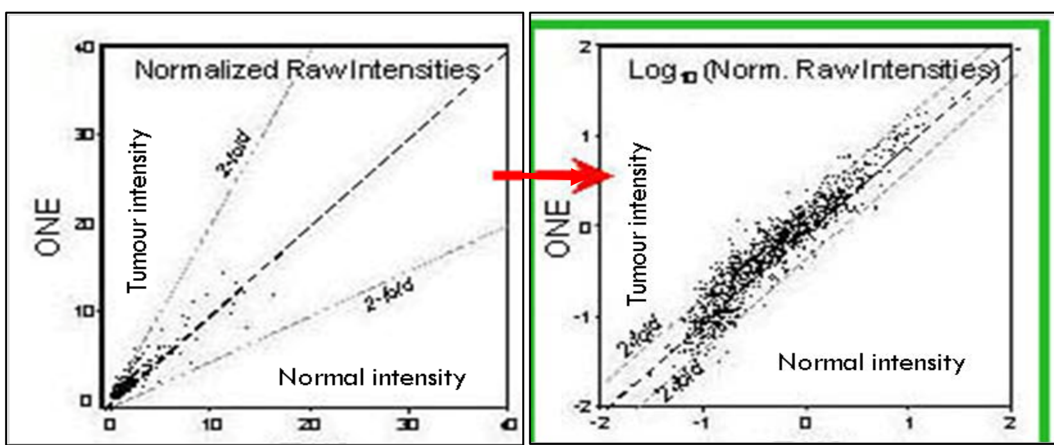


Fig. 4.3 Plot showing distribution of intensity values before and after normalization

Each sample had a unique average signal value along with a detection p value (a term used by Illumina). A small detection p value indicates that the measured intensity is very likely to be a true (significant) signal and not the background noise. The intensity associated with a detection p value ($\leq \pm 13$) corresponds to the respective upper quantile (percentile) of the (probability) distribution of the intensities of the negative controls.

4.4 Hierarchical Clustering

Hierarchical clustering is mathematical techniques whereby the analysed samples/genes are connected iteratively based on their similarity and are grouped together in form of dendrogram (or clustering tree). It enabled all samples with similar expression patterns to cluster together and which would further explain reasons for heterogeneity associated with various types of cancers. In order to determine the potential gene signatures from identified differentially expressed genes (DEGs) we performed unsupervised hierarchical clustering analysis in main groups viz. Total tumours (TT); Early tumours (ET) and Late tumours (LT). Clustering was also performed in molecular subtypes and according to stage of cancer viz. Lower Stage I,II vs Advanced Stages III, IV group to identify specific differential gene profiles. Cluster 3.0 software was used to accomplish clustering amongst these groups. We performed median centered clustering for centroid linkage using Pearson uncentered correlation. Clustering images were viewed using JavaTree View Software.

4.5 Gene ontology (GO) and pathway analysis

The transition from normal tissue to cancerous tissue is an important aspect in understanding the biology of cancer. GO analysis had been done to associate differentially expressed genes (DEGs) with GO function categories in order to understand the relevance and pathways associated with a particular gene using Pathwayexpress and Advanced pathway painter software. It generated various pathways involved significantly in cancer through Kyoto

Encyclopedia of Genes and Genomes (KEGG) database or Genemap databases. We also did enrichment of pathways using Gene set enrichment analysis (GSEA) to gain insights into the top significant pathways and their corresponding genes. It also generated a heat map for top 50 genes in each group.

4.6 Statistical analysis

Identification of DEGs using microarray was further performed by Bioconductor R package LIMMA. Raw data was processed by background subtraction and quantile normalization using neqc commands. Further genes which were differentially expressed were identified from normalized data by applying student's t test amongst all tumours vs total normal group; Analysis of variance (ANOVA) was done amongst three groups (Early tumour, Late tumour and Total normal control) which were corrected further by Benjamini Hocheberg based FDR correction. The significance of their association with various clinicopathological parameters was also identified by Chi square test. For identification of DEG's criteria of fold change ≥ 2 with a nominal p-value cutoff of 0.05 was used. Mann Whitney U test was used to determine the significance of differentially expressed genes in real time validation for groups TT/ET/LT.

4.7 Quantitative real time PCR

The reverse transcription reaction was performed using 1 μg of total RNA, random primers, and SuperScript® III RT (Invitrogen) in a total volume of 10 μL according to the manufacturer's protocol. Primers were purchased and designed using IDT software specific for real time PCR. The real-time RTqPCR reactions were prepared using power SYBR green (Applied Biosystems, Foster City, CA), with housekeeping gene 18sRNA, PSMC4 and beta actin as an endogenous control. 67 tumour tissues along with 38 distant normal tissues were used for validation of results (validation set). A normal pool was formed by combining RNA from all the normals upto 500ng. The cDNA was diluted to 5-fold for the real-time RT-PCR.

Each PCR reaction consisted of 5.0 µL of SYBR green mix, 4.5.0 µL diluted cDNA and 0.25 µL of individual primers making a total reaction to 10 µL. The following conditions were used: 95 °C for 2 min, 40 cycles of 95 °C for 10 s and 60 °C for 1 min.

Samples were run on the StepOne Real time PCR according to default parameters, with three replicate assays for each gene in each sample. Additionally we also procured commercial 2 vials of Normal breast RNA from Life technologies. Using the RQ Manager Stepone Software 2.3 (AB, Foster City, CA) the data was analyzed and the baseline and the threshold were verified for each gene of interest. Fold change of gene expression was calculated with the 2-ΔΔC T method, using β-actin and 18s and PSMC4 as the house keeping gene. For each assay, the average endogenous Ct (Cycle threshold) value in the SYBR qPCR assay was subtracted from the Ct of gene of interest to obtain a ΔCt value (gene of interest – endogenous), followed by ΔΔCt values and consecutively fold change (RQ mean) for the early and late group patients. Data analysis was performed using Datassist software and significant differential genes were identified with $p \leq 0.05$.

4.8 Results

4.8.1 Clinicopathological Details

In the present study 97 histopathologically confirmed cases of breast cancer were enrolled of which, 41 cases (42.3%) were below 40 years (Early onset-ET) having mean age 35.9 years while 56 cases (57.7%) had age ≥ 55 years (Late onset- LT) having mean age 63.2. Early onset group included patients with early menarche i.e.below 13 years (p=0.029) as compared to late onset group. Females not breast feeding their children were at higher risk of developing early onset breast cancer (p=0.023). Lymph node metastasis was also found significantly high in ET group of patients (p=0.05) compared to LT group. The cases for which the details were not available were excluded. On phenotypic characterization using immunohistochemistry it

was found that 19 cases (32%) were Luminal (ER+ PR+ HER2/neu -), 22 cases (24%) were HER2/neu over expressing (ER- PR- HER2/neu +) and 19 cases (21%) were Basal (ER- PR- HER2/neu -). Loss of estrogen receptor was found in 22 cases (55%) of early onset group and in 24 cases (48%) in late onset patients, however values did not reached statistical significance. HER2/neu positivity was found in 29 cases (58%) of late age onset group as compared to 18 patients (45%) in early onset group. In early age onset patients 25 cases (62.5 %) reported at advanced stages (p=0.07) as compared to 24 cases (45.3%) in late age onset group. Also 25% cases having early onset breast cancer belonged to Basal subtype as compared to 18% cases in late onset breast cancer group though the difference between the two groups was not found statistically significant (p=0.735) (Table 4.1).

Table 4.1 Showing association of various clinicopathological characteristics of patients with age of onset of breast cancer

Characteristics	**Early Group (≤ 40yrs)**		**Late Group (≥ 55yrs)**	**p value**
Tumour (n=97)	41 (42.3%)		56(57.7%)	
Age at menarche(yrs)				
≤13	23(56.1%)		19(34.5%)	0.029
≥14	18(43.9%)		36(65.5%)	
Data unavailable			1	
Breast Feed				
Yes	35(85.3%)		54(98.2%)	0.023
No	6 (14.6%)		1(1.8%)	
Data unavailable	1		3	
Tumour size				
=1,2cm	15(37.5%)		27(50.9%)	0.14
=3,4cm	25(62.5%)		26(49.1%)	
Data unavailable	1		3	

Lymph Node				
N=0	5(12.5%)		15(28.4%)	0.05
N=1	35(87.5%)		38(71.6%)	
Data unavailable	1		3	
Metastasis				
M=0	36(90.0%)		50(94.3%)	0.345
M=1	4(10%)		3(5.7%)	
Data unavailable	1		3	
Oestrogen receptor				
Negative	22(55.0%)		24(48.0 %)	0.327
Positive	18(45.0%)		26(52.0%)	
Data unavailable	1		6	
Progesterone receptor				
Negative	25(62.5%)		28(56.0%)	0.343
Positive	15(37.5%)		22(44.0%)	
Data unavailable	1		6	
HER2/neu expression				
Negative	22(55.0%)		21(42.0%)	0.155
Positive	18(45%)		29(58.0%)	
Data unavailable	1		6	
Phenotype				
Luminal (ER+ PR+ HER2/neu -)	9(23%)		10(20%)	0.735
Her2Nu (ER- PR- HER2/neu +)	9(23%)		13(26%)	
Basal (ER- PR- HER2/neu -)	10(25%)		9(18%)	

4.8.2 Qualitative and quantitative analysis of RNA

Gel electrophoresis for RNA

Quality of RNA extracted from all the cases was checked by running over 1.5% agarose gel in 0.5X TBE prepared in DEPC treated water. All the samples showed 2 bands for RNA of 28s and 18s for most of the samples suggesting high quality of RNA (Fig.4.4).

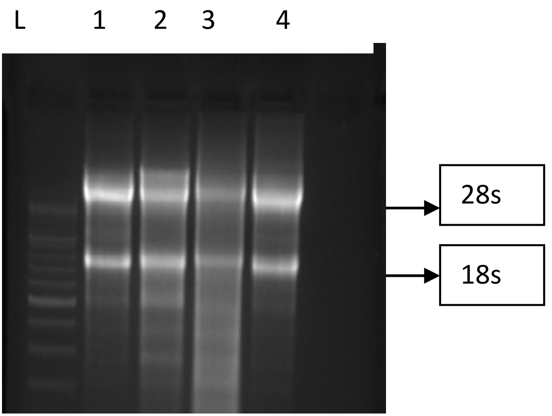


Fig.4.4 Gel electrophoresis image showing the quality of extracted RNA samples

Quality check by Bioanalyzer

All the samples were checked for their integrity by determining their RIN (RNA Integrity Number) on bioanalyzer. RIN measures the quality of RNA in the range of 1-10, RIN 1 suggesting the totally degraded RNA and 10 showing very high quality RNA. Samples having RIN above 7 were only included in the study. Maximun RIN for our tissue samples was 9.7(Fig.4.5).

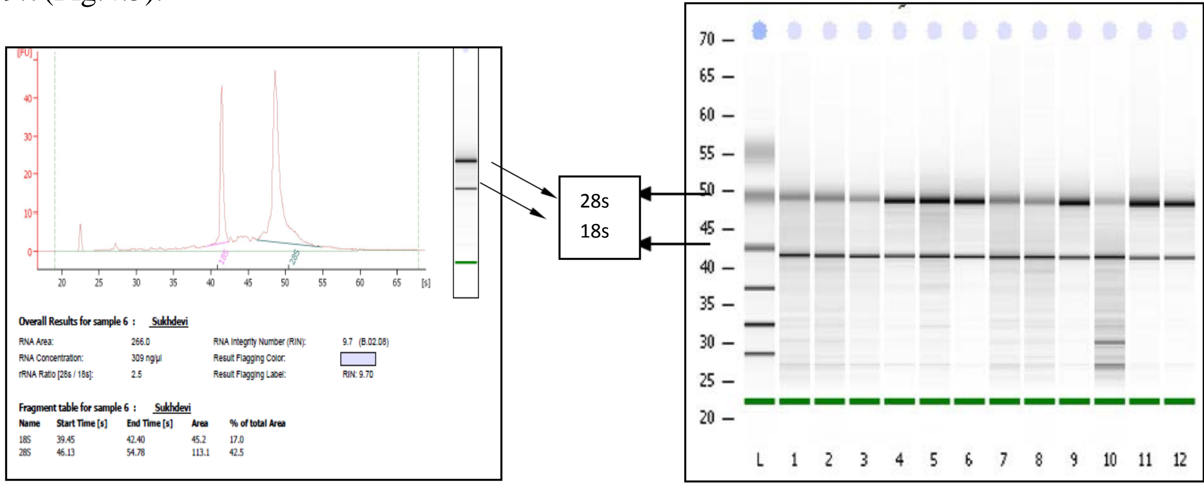


Fig.4.5 Bioanalyzer image showing high quality of RNA samples and 28s and 18s bands for other samples in gel using bionalyzer

4.8.3 Analysis of gene expression profiles of breast cancer

The total 29 samples were proceeded for microarray. The RNA was converted to cDNA followed by conversion to cRNA which was biotin labelled. It was hybridized to chip, followed by washing and scanning, resulting in green colour intensity signals.

A. Chip Hybridization Image

After the hybridization, raw data was collected and processed by Illumina's genome studio software. Microarray chip image data (Fig.4.6) was extracted and visualized to see the overall hybridization efficiency. All the probes on chip showed equal and uniform hybridization

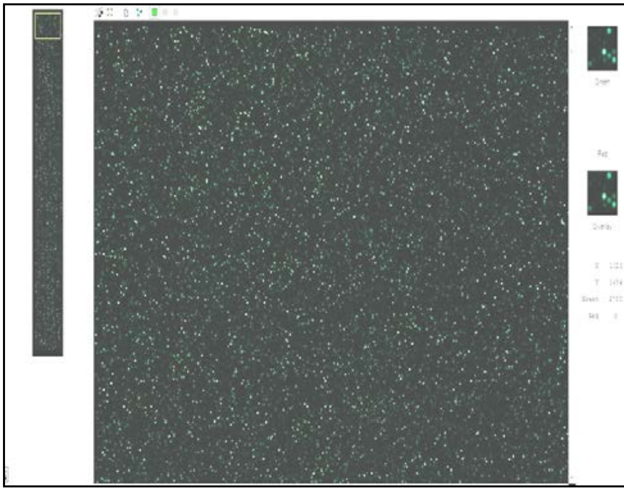


Fig.4.6 Image of chip showing uniform hybridization of the probes

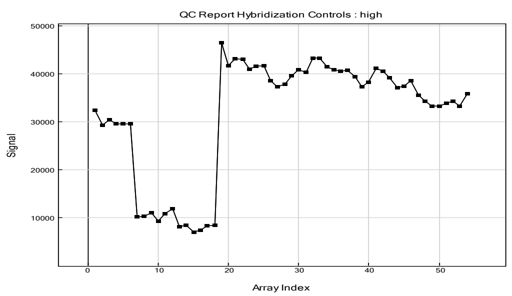
B. Quality Control parameters

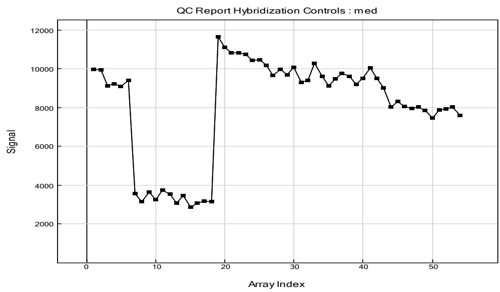
Quality control (QC) of data is an important step as it helps in identifying presence of any outliers or experimental failures. It is performed using internal control of chip by genome studio software itself. QC tools included plotting features as the control summary plot, box plots, scatter plots, cluster dendrograms. The sample independent metrics make use of oligonucleotide spiked into hybridization solution. Poor performance measured by these controls could indicate a general problem with hybridization, washing or staining. The sample

dependent metric is based on measurements from actual sample of interest and their poor performance may indicate problems related to sample or labelling.

Sample independent control- Hybridization control

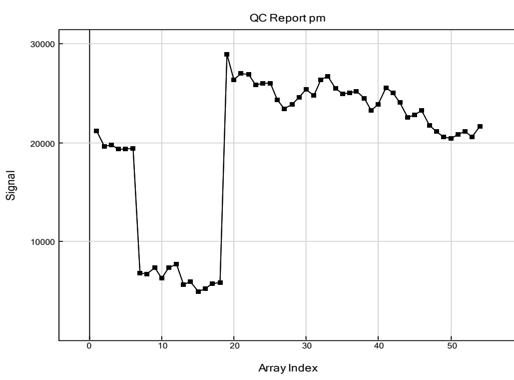
It is used to access efficiency of hybridization. This control worked fine at both high and medium scale for our samples

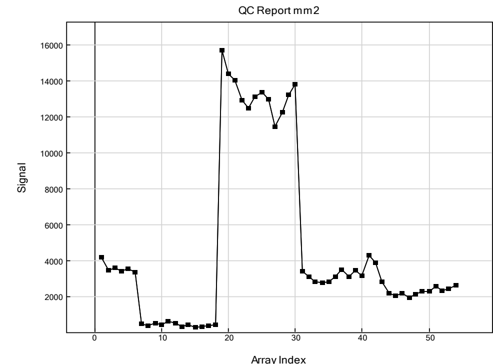




Stringency Control

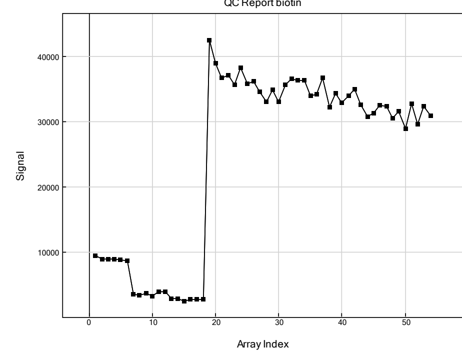
It determines the quality of the experiment with Perfect match signal > Mismatch signal. In all samples Perfect match signal > Mismatch signal inferred as lesser noise in experiment.

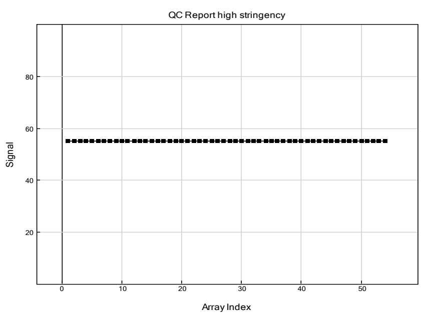




Biotin & high stringency control

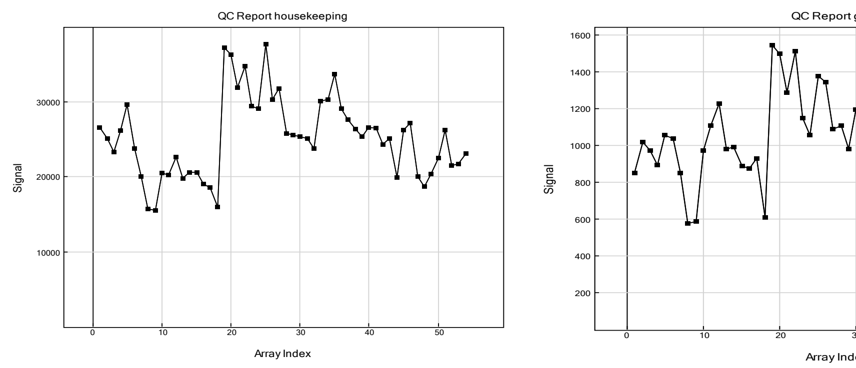
It is designed to determine efficiency in biotin labeling and stringency check. Here biotin labeling in all samples was good and efficient and all points were found to be in linear scale.





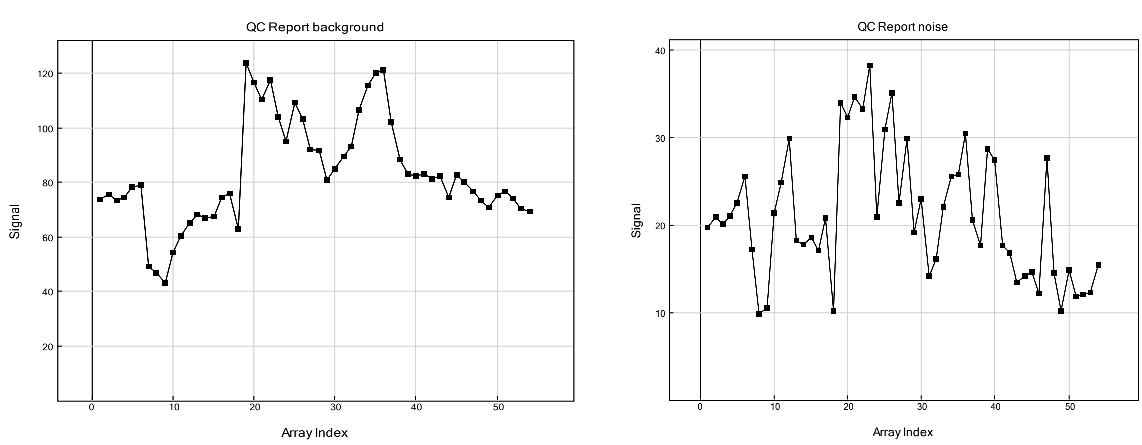
Sample dependent control-Gene intensity control

It detects sample quality and status with housekeeping control probes designed to show optimum or constant expression. Here housekeeping and control genes worked fine for all samples.



Negative Controls

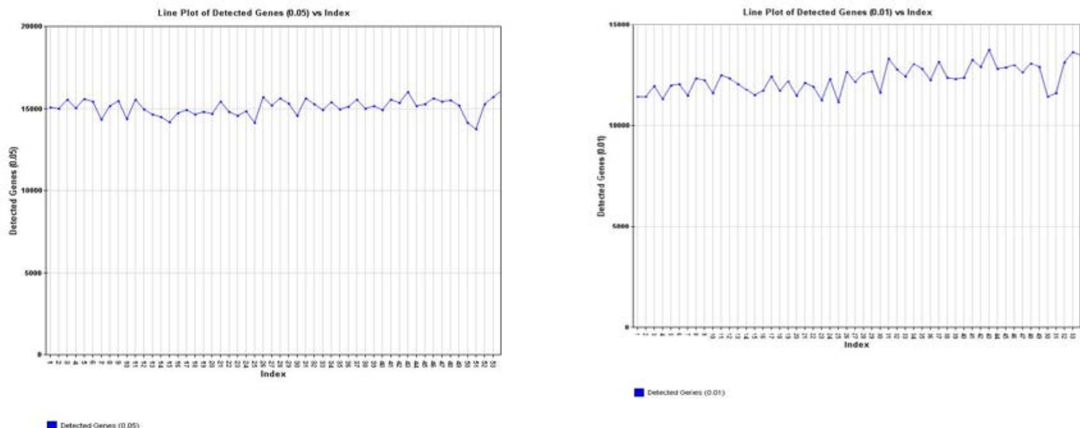
It detects contamination in the samples. Probes were designed which had no resemblance with human genome. Background and noise signal represented as nonspecific/negative signal expected to be least. In given experiment this negative control has worked fine



Detected genes p value 0.01 & 0.05

Genome studio calculated and reports a detection p-value, which represents the confidence that a given transcript is expressed above the background defined by negative control probes. This detection score determines whether a transcript on the array was called significantly

detected or not. Detected genes expected to be > 8000 in number. In all above samples detected transcript in expected range.



4.8.3 Differential Gene expression Analysis

The raw data after normalization showed 28,504 genes deregulated in all the samples. This normalized dataset was further used to identify the differentially expressed genes (DEG's) amongst the following 3 main groups-

1. Total Tumour (TT) vs. Normal Control (CC)
2. Early onset Tumour (ET) vs. Normal Control (CC)
3. Late onset Tumour (LT) vs. Normal Control (CC)

Also DEGs were determined for other 2 subgroups

1. Molecular subtypes (Luminal, HER2/neu over expressing, Basal)
2. Advanced Stage (III,IV) (AS) vs. Lower Stage (I,II) (LS)

4.8.4 Differential gene expression analysis

4.8.4.1 Total tumour (TT) vs. Normal Control (CC)

Volcano plots plotted for the TT vs. CC to get an overview of the deregulated genes showed (Fig. 4.7) higher number of the gene significantly downregulated ($p \leq 0.05$, FC cutoff 1.5) as compared to upregulated genes.

Total Tumour vs. Normal

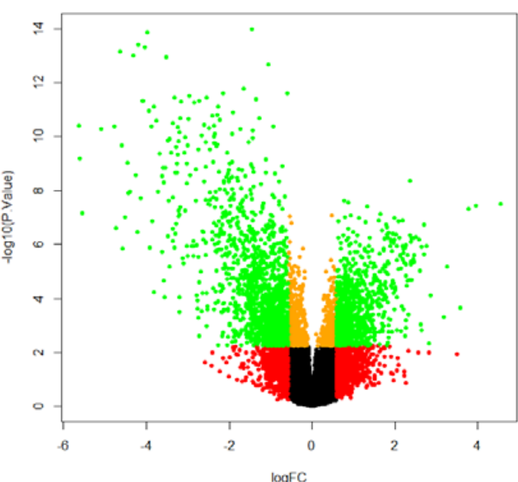


Fig. 4.7 Volcano plots showing genes differentially expressed in all breast tumour tissue vs. normals. All genes in green colour reached significance level with p≤0.05 and all red genes are those which did not reached significance level. The genes marked in black and orange colour are those genes that did not had a fold change greater than 1.5.All green genes were significant deregulated having a fold change cutoff of 1.5.

a. Hierarchical Clustering

Hierarchical clustering was performed for all the genes getting significantly deregulated in tumour tissues with respect to normal control tissues. Heat map showed separate and distinct clusters of tumour and normal tissue suggesting robustness of our data. It formed two separate gene clusters viz. Cluster 1 getting upregulated in control and Cluster 2 upregulated in tumours as compared to normals (Fig.4.8).

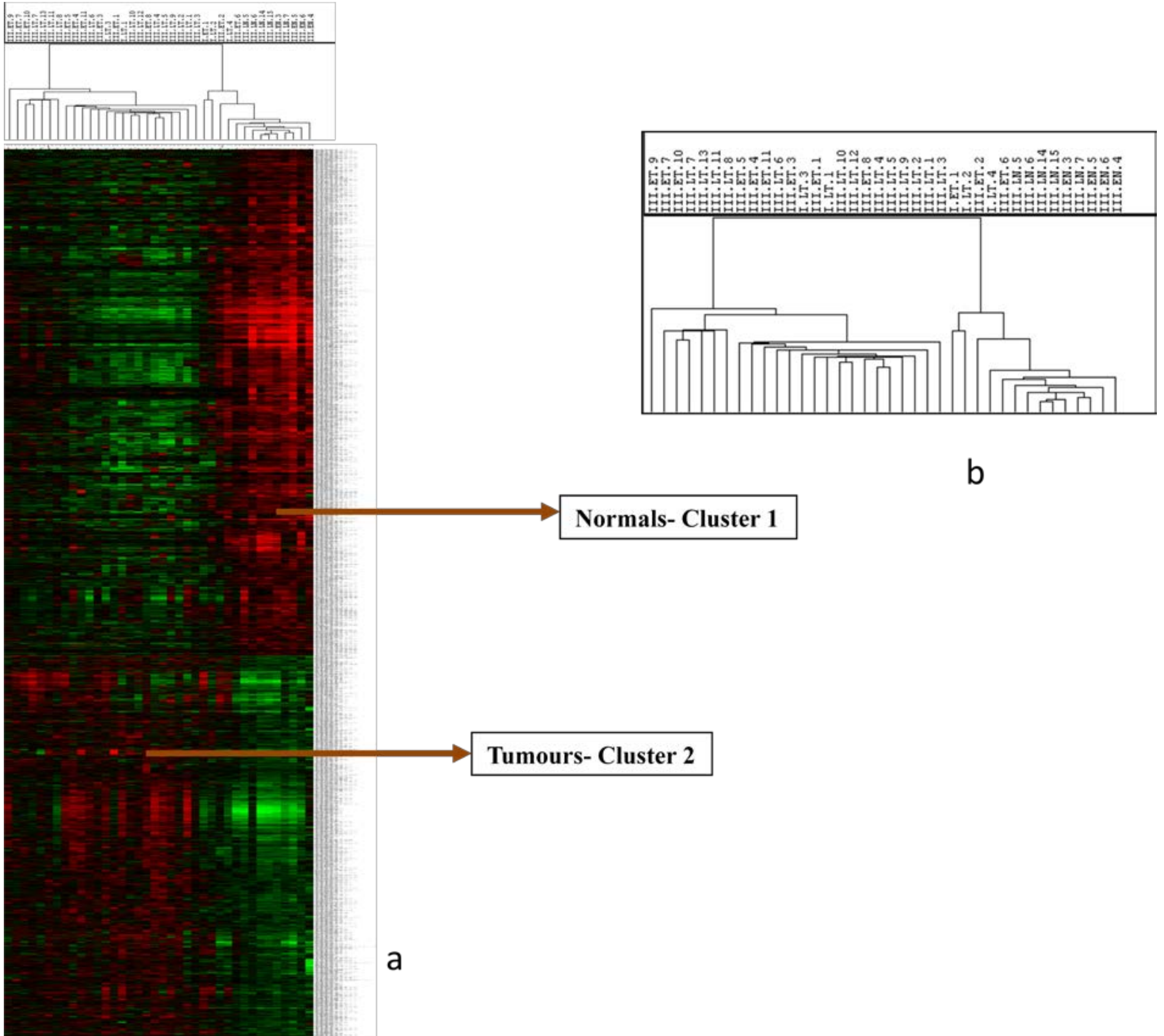


Fig.4.8. (a) Heat map showing the clustering of all tumour tissues with respect to their normal Differential patterns were observed amongst the tissues with red colour genes showing upregulation and green colour genes depicting downregulated. Cluster 1 includes genes downregulated in tumours and cluster 2 includes genes upregulated in tumours (b) dendrogram showing separate clusters of tumours and normals

b. Differentially Expressed Genes

Differential genes expression analysis using student t test, yielded 2413 deregulated genes at significance level $p \leq 0.005$ with a fold change cutoff of 1.5. Among these DEGs, 991 genes were found upregulated and 1422 were found downregulated. The top 20 DEGs are shown in Table. 4.2

Table 4.2 Top 20 DEGs between breast cancer and normal controls tissues

GENE	FC-UP	adj.P.Val
COL10A1	23.2815	5.96E-06
MMP11	15.45492	6.55E-06
GJB2	13.68954	8.27E-06
CST1	11.94889	0.004912
KIAA1199	9.509539	0.000351
MMP1	9.096327	0.008953
MMP13	7.297567	0.002203
CEACAM6	7.014203	0.042617
BUB1	6.795974	9.11E-05
ASPM	6.500159	2.39E-05
FAM83D	6.454691	0.000158
EPYC	6.265743	0.019523
CKAP2L	6.264448	9.48E-05
PITX1	6.117431	0.014046
TUBB3	6.044887	0.00046
CEP55	5.938681	6.02E-05
HSD17B6	5.64996	5.07E-05
KIF20A	5.529247	0.000161
CCNB2	5.48809	5.52E-05
MELK	5.47034	0.00025

GENE	FC-DOWN	adj.P.Val
PLIN	-49.7345	3.12E-08
KIAA1881	-48.6453	2.71E-07
ADH1A	-46.9269	1.12E-05
ADH1B	-33.9347	3.6E-08
CIDEC	-27.4842	3.18E-08
THRSP	-26.7033	3.01E-05
GPD1	-24.9122	3.83E-10
TIMP4	-24.1717	1.03E-07
FABP4	-23.8769	0.000104
C7	-22.871	1.39E-05
G0S2	-22.0594	3.68E-07
SCARA5	-21.6551	2.8E-06
C2ORF40	-21.1401	2.51E-06
ALDH1L1	-20.0129	4.43E-10
SLC19A3	-19.0548	7.94E-07
SAA1	-18.744	3.74E-05
AQP7P2	-18.3621	3.36E-10
PCOLCE2	-17.5336	4.09E-06
LIPE	-17.1976	7.36E-09
CA4	-17.0258	7.36E-09

c. Gene ontology analysis

Further, to identify the biological mechanism associated with breast cancer, pathway analysis was performed using pathway express software amongst tumours and normal controls. Major pathways found deregulated included- **Cell adhesion molecules (CAMs**) involving genes as *CADM3, CDH5, CLDN11, CLDN5*, **Cell cycle** with genes *BUB1, CCNA2, CDC2, CCNB2*, **Adherens junction** with genes as *ACVR1C, LEF1, PVRL2, PVRL4*, **pathways in cancer** involving genes as *MMP1, CCNE1, WNT2, BIRC5;* **PPAR signaling pathway** including genes as *MMP1, OLR1, PLIN, FABP4, AQP7* **; Focal adhesion pathway** with *COL11A1, COL1A1, COL1A2, COL5A1, COL5A2, CAV1, CAV2, FIGF* ; **ECM-receptor interaction i**ncluding genes as *COL11A1*, *SPP1*, **DNA replication** including genes *MCM2, MCM4, POLE2* etc. (Table 4.3.).

Table 4.3 Pathways found differentially expressed in all tumours

Rank	Pathway Name	Unique Pathway-id	Impact Factor	Input Genes in Pathway	Corrected p-value	Corrected gamma p-value
1	Cell adhesion molecules (CAMs)	1:04514	235.367	21	0.011143	1.43E-100
2	Cell cycle	1:04110	22.121	34	8.12E-10	5.71E-09
3	Adherens junction	1:04520	20.559	15	0.00458	2.54E-08
4	PPAR signaling pathway	1:03320	15.598	20	3.54E-06	2.79E-06
5	Pathways in cancer	1:05200	9.026	50	0.000696817	0.001205551
6	Focal adhesion	1:04510	14.939	40	5.09E-06	5.18E-06
7	ECM-receptor interaction	1:04512	14.846	23	1.69E-06	5.65E-06
8	DNA replication	1:03030	10.812	12	4.31E-05	2.38E-04

The pathways which were found upregulated in breast cancer with genes having Fold change above 2 included Cell cycle, Adherens junction, DNA replication, and ECM-receptor interaction (Table 4.4) while pathways found downregulated at less than -2 FC included leuckocyte transendothelial migration, Cell adhesion molecules, PPAR signaling, Adherens junctions, Focal adhesion etc.as shown in Table 4.5.

Table 4.4 Pathways upregulated in all tumours

Rank	**Unique Pathway-id**	**Pathway Name**	**Impact Factor**	**#Input Genes in Pathway**	**corrected gamma p-value**
1	1:04520	Adherens junction	33.507	2	9.68E-14
2	1:04512	ECM-receptor interaction	13.797	9	1.51E-05
3	1:03030	DNA replication	11.914	6	8.65E-05
4	1:04510	Focal adhesion	6.469	9	0.011583

Table 4.5 Pathways downregulated in all tumours

Rank	Unique Pathway-id	Pathway Name	Impact Factor	#Input Genes in Pathway	corrected gamma p-value
1	1:04670	Leukocyte transendothelial migration	322.933	12	1.83E-138
2	1:04514	Cell adhesion molecules (CAMs)	226.966	12	6.13E-97
3	1:04520	Adherens junction	20.562	8	2.53E-08
4	1:03320	PPAR signaling pathway	19.079	14	1.04E-07
5	1:04510	Focal adhesion	10.923	17	2.15E-04

d. Gene Set Enrichment Analysis (GSEA)

To further narrow down the genes hierarchical clustering of top 50 genes found differentially expressed between TT and CC, was done using GSEA, which showed two distinct clusters of upregulated and downregulated genes in all tumours (Fig.4.9). All genes marked red were upregulated while those marked blue were downregulated. The topmost over expressed genes in all breast tumours included *COL10A1, GJB2, MMP11, MMP13, MMP1, COL11A1* while the top under expressed genes included *GPD1, ALDH1L1, AQP7, PLIN, ADH1B, CIDEC*, (Fig. 4.9)

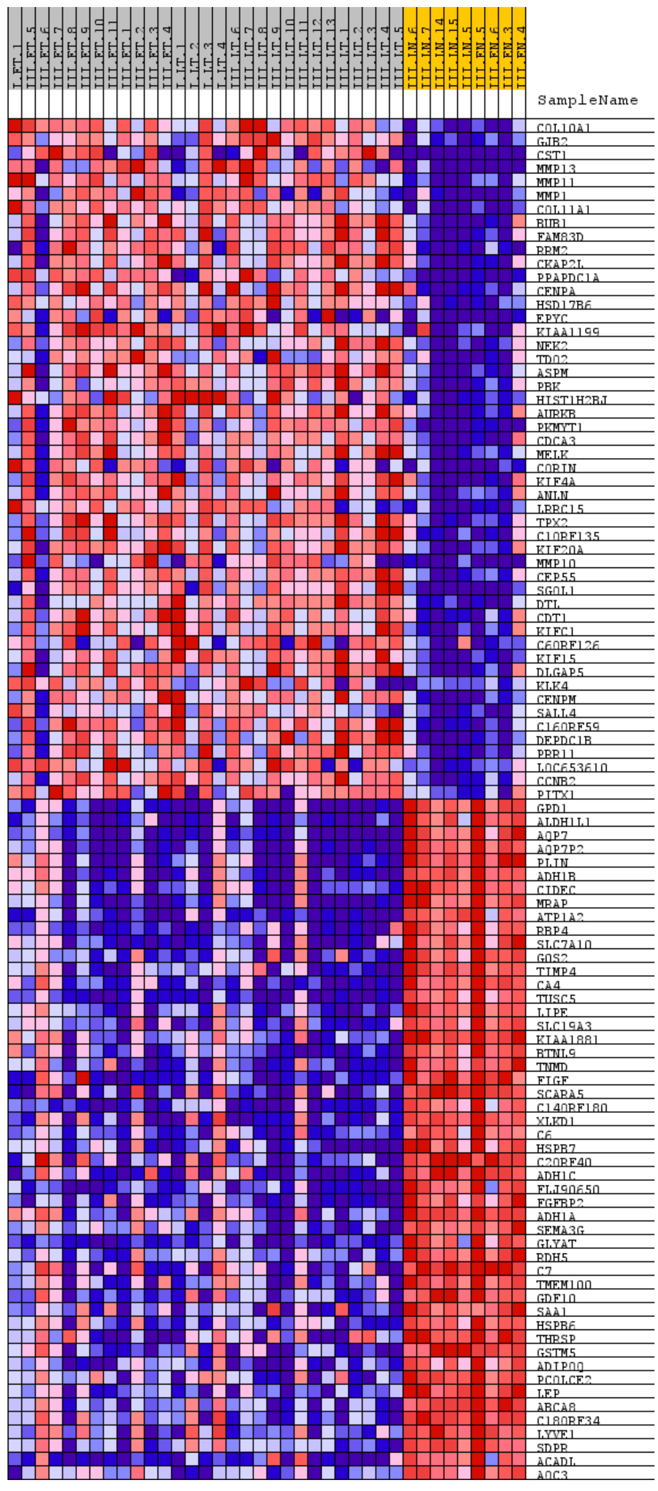


Fig 4.9 Heat Map of the top 50 genes that are differentially expressed in all tumours (Tumour Vs Normal). Red colour show upregulated genes and blue shows downregulated genes.

Despite identification of pathways and genes leading to breast carcinogenesis amongst Indian patients, it was important to determine the genes associated with early onset breast carcinoma. Hence we further analyzed the DEGs present in early and late onset groups of breast cancer patients.

4.8.4.2. Early onset Tumour (ET)/ Late onset Tumour (LT) vs. Normal Control (CC)

Since majority of breast cancer patients in India are below 40 years of age, belonging to early onset tumour group, study of DEGs in this group has been done to understand molecular biology.

a. Volcano Plots

Volcano plots were plotted for the ET vs. CC and LT vs CC to get an overview of the deregulated genes in both early and late onset tumours (Fig. 4.10). Higher numbers of the gene were found significantly ($p \leq 0.05$) downregulated ($FC \leq -1.5$) as compared to upregulated genes ($FC \geq 1.5$) in both ET and LT.

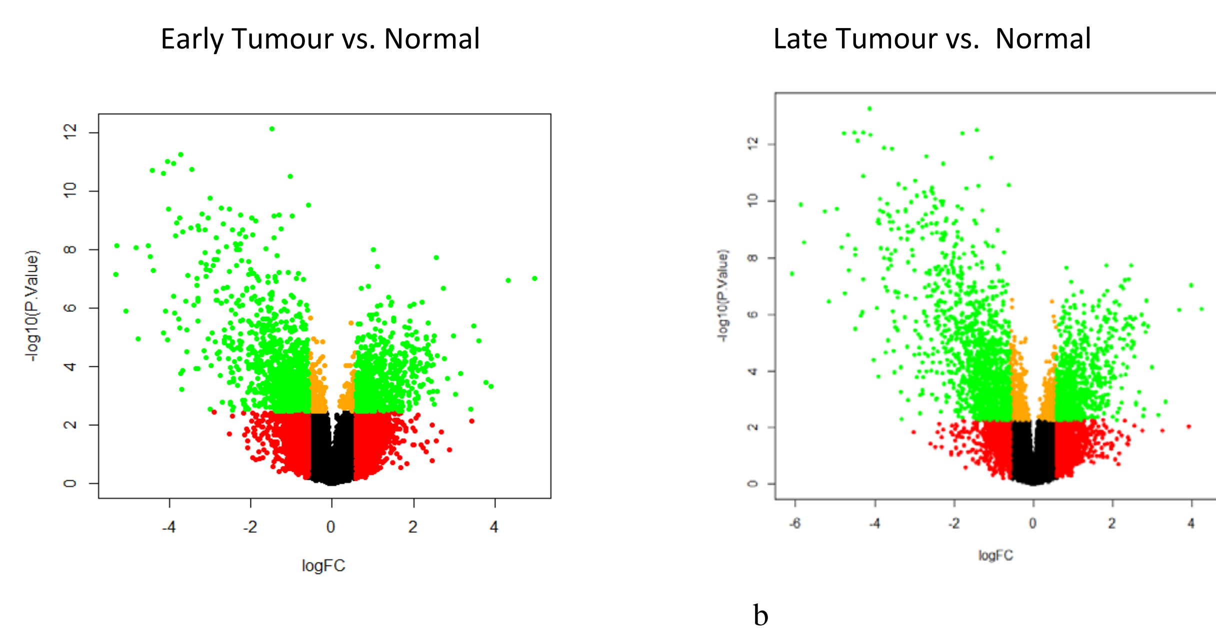


Fig.4.10 Volcano plots showing genes differentially expressed in (a) ET vs CC and (b). LT vs CC. All genes in green colour reached significance level with p≤0.05 and all red genes are

those which did not reached significance level. The genes marked in black and orange colour are below FC cutoff of 1.5.All green genes were significant deregulated ($p<0.005$, FC cut off 1.5).

b. Hierarchical Clustering

Hierarchical clustering was performed by Pearson uncentered algorithm with centroid linkage rule and heat map was generated between ET and LT using log ratios. The early tumours clustered distinctly from late tumours. However some samples from ET and LT clustered together showing grey zone and heterogeneous nature of breast cancer (Fig. 4.11). Cluster 1 showed upregulated genes while cluster 2 showed downregulated genes in ET as compared to LT where those genes showed reverse expression, separately shown in (Fig. 4.12).

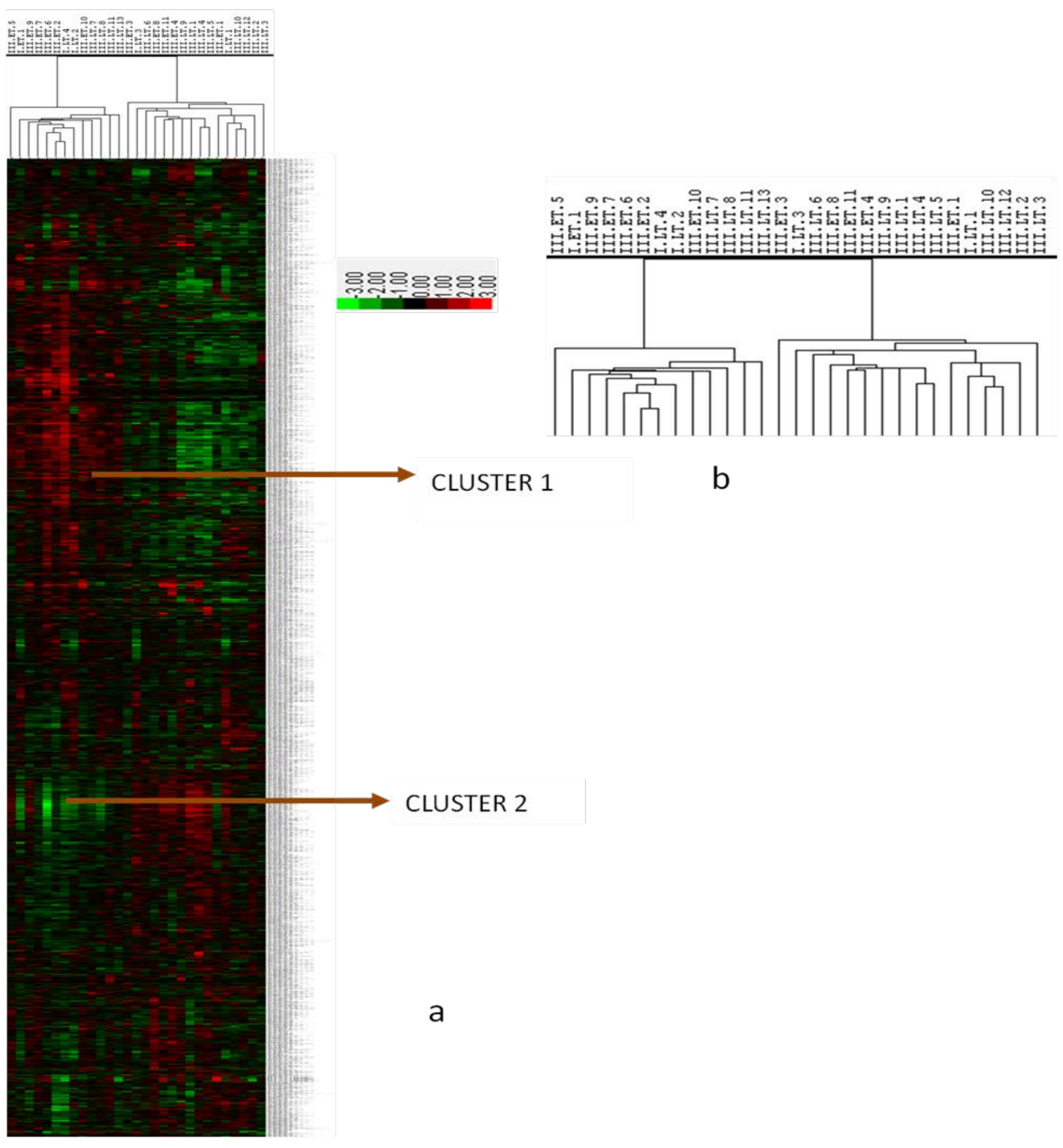
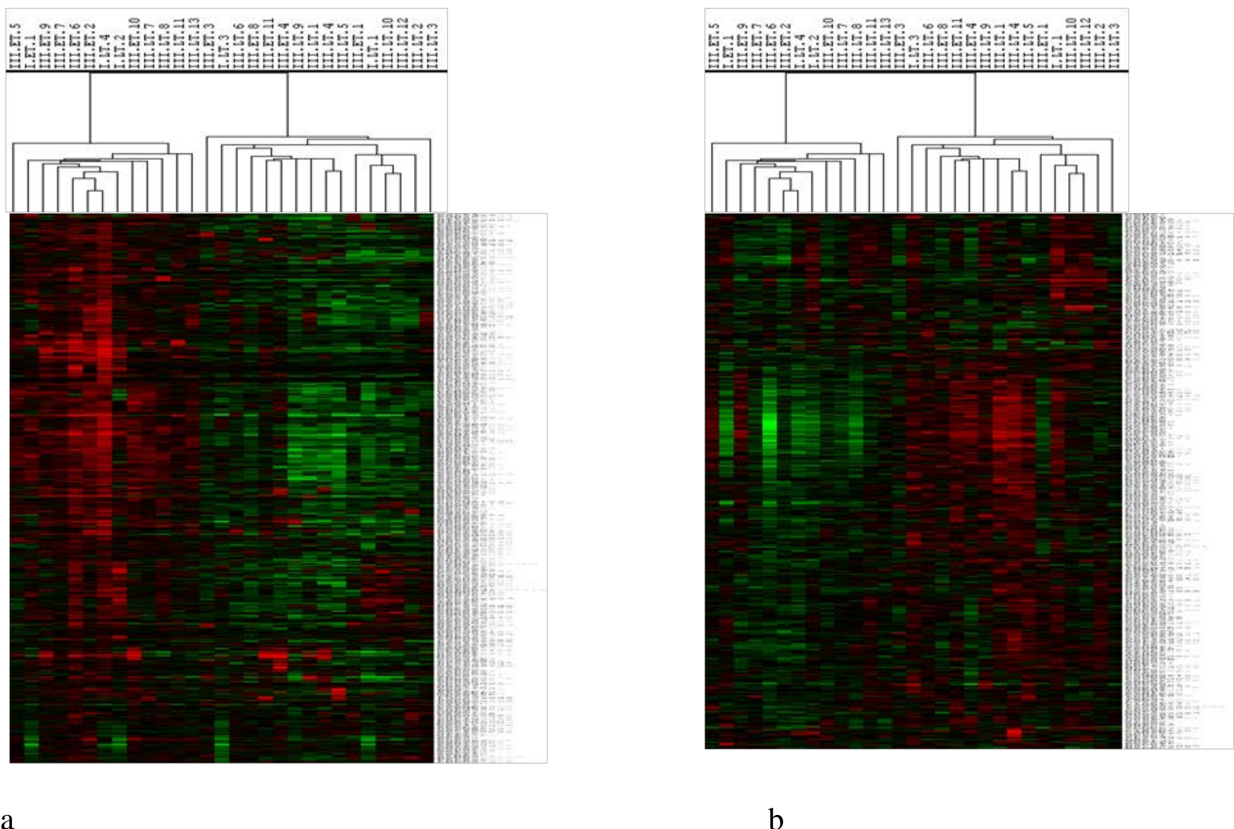


Fig.4.11. (a) Heat map generated by clustering of ET and LT samples. (b)Sample dendrogram showing distinct ET and LT clusters



a b

Fig.4.12 Differential clusters identified amongst ET and LT (a) Cluster1 (b) Cluster 2 showing genes upregulated genes in ET and LT respectively.

c. Differentially Expressed Genes

Differential expression was determined by using student t test, and 1685 significantly deregulated genes in ET ($p \leq 0.005$, FC cutoff 1.5) were identified of which 724 genes were upregulated and 961 genes were downregulated. The top 20 DEGs (upregulated /downregulated) are shown in Table 4.6

Table 4.6 Top DEGs identified amongst ET vs CC (a) Upregulated genes (b) Downregulated genes

GENES	FC_EARLY -Up	F.p.value
COL10A1	31.47292	1.46E-07
MMP11	20.16884	1.62E-07
CST1	15.09418	0.000998
MMP1	13.78096	0.001196
KIAA1199	12.24469	2.69E-05
GJB2	11.13843	2.76E-07
S100P	10.67661	0.011113
MMP13	8.923938	0.000343
PITX1	8.241605	0.00272
TUBB3	7.974669	2.51E-05
KLK4	7.194806	0.001101
FNDC1	6.822992	0.000239
PPAPDC1A	6.605047	9.53E-07
BUB1	6.013008	6.92E-06
SPP1	5.957017	0.001561
COL11A1	5.888016	2.28E-08
MMP3	5.746106	0.004459
HSD17B6	5.655976	4.00E-06
FAM83D	5.589527	1.29E-05
BAPX1	5.58275	0.002853

a

GENES	FC_EARLY -Down	F.p.value
KIAA1881	-40.5015	5.61E-09
PLIN	-39.6452	2.95E-10
THRSP	-33.9885	1.47E-06
ADH1B	-28.3934	4.42E-10
ADH1A	-27.5073	1.64E-07
CIDEC	-23.2443	3.64E-10
TIMP4	-22.5053	2.24E-09
GPD1	-21.6043	6.70E-13
G0S2	-21.3539	9.76E-09
ALDH1L1	-17.9509	1.08E-12
C7	-17.8379	5.32E-07
SCARA5	-17.304	7.51E-08
AQP7P2	-16.7673	4.77E-13
SAA1	-16.5605	2.44E-06
RBP4	-16.4476	1.22E-10
AQP7	-15.0834	5.85E-13
SLC19A3	-14.9731	1.45E-08
PCOLCE2	-14.8175	1.37E-07
SLC7A10	-14.3907	4.92E-10
LIPE	-14.2705	3.06E-11

b

Similarly we found 2379 genes differentially expressed in LT comprising of 988 upregulated and 1391 downregulated genes (Table 4.7). Table shows top 20 DEGs in LT.

Table 4.7 .Top DEGs identified amongst LT vs CC (a) Upregulated genes (b) Downregulated genes

GENES	FC_LATE1	F.p.value
COL10A1	18.81883	1.46E-07
GJB2	15.83475	2.76E-07
MMP11	12.80727	1.62E-07
CST1	10.13195	0.000998
CEACAM6	8.845062	0.012935
KIAA1199	7.955398	2.69E-05
BUB1	7.409286	6.92E-06
ASPM	7.250018	1.01E-06

GENES	FC_LATE1	F.p.value
ADH1A	-68.4175	1.64E-07
PLIN	-58.3665	2.95E-10
KIAA1881	-55.3614	5.61E-09
ADH1B	-38.4857	4.42E-10
FABP4	-36.043	1.88E-06
CIDEC	-30.9348	3.64E-10
C2ORF40	-28.6211	2.00E-08
GPD1	-27.5477	6.70E-13

FAM83D	7.144855	1.29E-05
CKAP2L	6.929208	6.64E-06
EPYC	6.883756	0.006325
MMP1	6.78446	0.001196
PBK	6.777075	1.21E-05
CEP55	6.63997	3.36E-06
MMP13	6.331392	0.000343
DIO1	6.011271	0.00376
C6ORF126	5.976349	0.005392
CCNB2	5.914008	3.61E-06
TOP2A	5.823935	9.92E-05
MELK	5.760493	2.72E-05

a

C7	-27.2571	5.32E-07
TIMP4	-25.4218	2.24E-09
SCARA5	-25.3702	7.51E-08
CA4	-22.8854	1.91E-12
SLC19A3	-22.5894	1.45E-08
G0S2	-22.5714	9.76E-09
THRSP	-22.5223	1.47E-06
ALDH1L1	-21.6094	1.08E-12
SAA1	-20.4564	2.44E-06
DARC	-19.7618	3.84E-06
PCOLCE2	-19.7456	1.37E-07
LIPE	-19.6185	3.06E-11

b

It was observed that there were 420 unique genes (270 upregulated; 150 downregulated) exclusively expressed in ET and 1114 unique genes (534 upregulated, 580 downregulated) exclusively expressed in LT and there were 1265 genes which were found deregulated in both ET and LT (Fig. 13.)

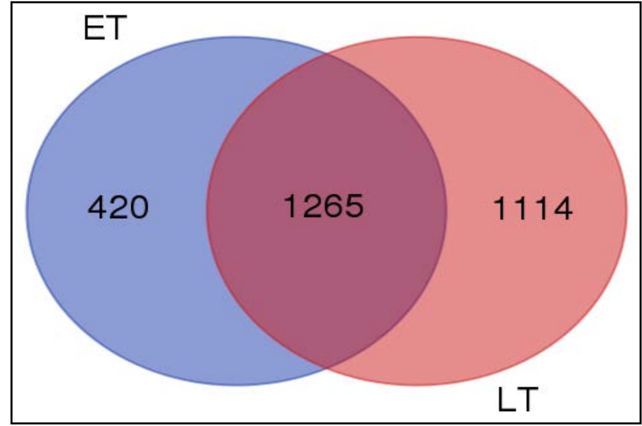


Fig.4.13 Venn diagram showing the unique and common genes amongst ET and LT

d. Gene Ontology

Further pathways were analyzed amongst ET and LT to understand the molecular pathogenesis associated with breast cancer in age specific cohort. GeneMapp identified **Metallopeptidases activity** (*MMP1*, *MMP13*, *MMP14*, *MMP11*, MMP3, *ADAMTS1*, *ADAMTS5*) and **Anion deregulation activity** (*COL10A1, COL12A1, COL5A1, COL1A1*) to

be top deregulated pathways. These pathways were of prime importance since they contained several genes which were differentially expressed in breast tumours and amongst ET / LT. Other deregulated pathways identified in KEGG are shown in Table 4.8 and Table 4.9.

Table 4.8 List of pathways deregulated in early onset tumours

Rank	**Pathway Name**	**Pathway-id**	**Impact Factor**	**#Input Genes in Pathway**	**Corrected p-value**
1	Cell adhesion molecules (CAMs)	1:04514	169.631	13	3.65E-72
2	Leukocyte transendothelial migration	1:04670	263.359	10	7.65E-172
3	ECM-receptor interaction	1:04512	16.951	18	0.00000078
4	Cell cycle	1:04110	15.375	21	0.00000344
5	PPAR signaling pathway	1:03320	15.374	15	0.00000345
6	Pathways in cancer	1:05200	10.153	34	0.000435

Table 4.9 List of pathways deregulated in late onset tumours

Rank	**Pathway Name**	**Unique Pathway-id**	**Impact Factor**	**#Input Genes in Pathway**	**corrected p-value**
1	Cell adhesion molecules (CAMs)	1:04514	260.699	13	1.58E-111
2	Leukocyte transendothelial migration	1:04670	345.414	14	3.37E-148
3	Adherens junction	1:04520	24.625	10	5.18E-10
4	Cell cycle	1:04110	21.738	25	8.24E-09
5	PPAR signaling pathway	1:03320	19.302	17	8.41E-08
6	ECM-receptor interaction	1:04512	16.445	18	1.26E-06
7	Pathways in cancer	1:05200	16.248	41	1.51E-06

The analysis showed that the key pathways found deregulated remained mostly same in both ET/LT groups to a larger extent however some of the genes showed deregulation in both the groups but with varied fold change of deregulation, few genes were found uniquely deregulated in each group. Some of the important pathways are discussed below ahead.

Metallopeptidase Activity Pathway

Metallopeptidase activity genes showed deregulation of top DEGs in all TT/ET/LT groups. These were found to be over expressed in breast cancer as well as they were differentially expressed in early and late onset tumours (Table 4.10) (Fig 4.14). MMPs are known to be over expressed in breast cancers including several other cancers, but their differential role in ET and LT carcinogenesis has not been elucidated.

Table 4.10 MMP pathway genes showing differential expression in ET and LT tumours along with their adjusted p value

GENES	FC_EARLY	FC_LATE	p.value
MMP1	13.78096	6.78446	0.001196
MMP11	20.16884	12.80727	1.62E-07
MMP13	8.923938	6.331392	0.000343
MMP14	3.75373	1.900651	1.26E-05
MMP3	5.746106	4.261598	0.004459
ADAMTS1	-3.99311	-6.96077	1.19E-08
ADAMTS5	-2.68595	-4.16648	3.20E-05

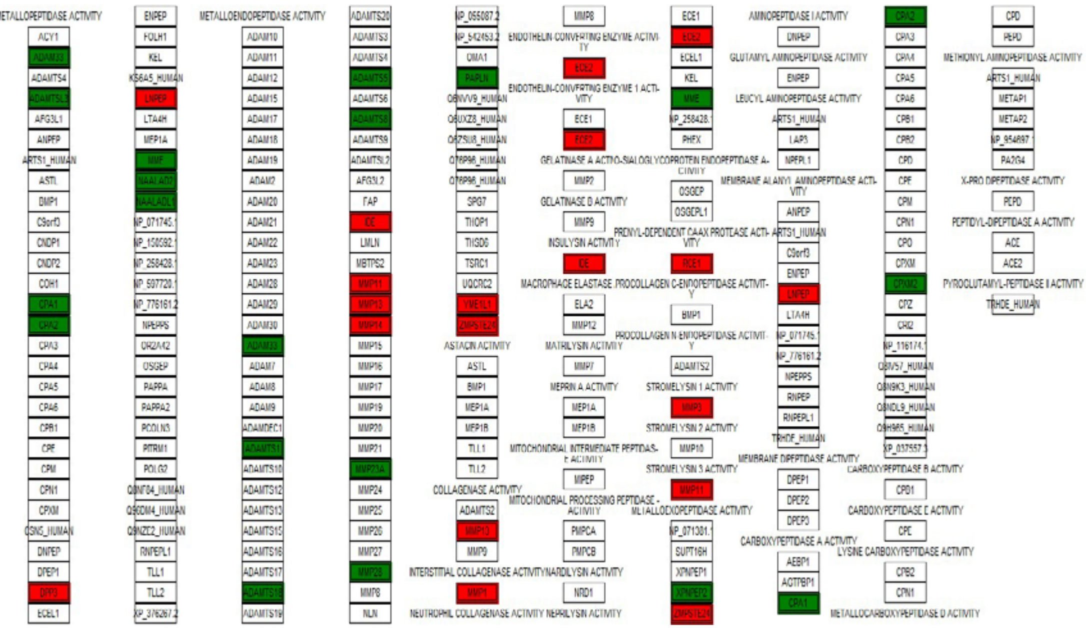


Fig.4.14 Metallopeptidase genes pathway showing differential expression of genes in early cancer.The red boxes showing upregulated genes and green boxes showing downregulated genes.

Cell adhesion molecule pathway

This pathway is mainly associated with binding of cell surface proteins with extracellular matrix and thus playing a critical role in wide array of biological processes that include homeostasis, the immune response, and inflammation. This is major deregulated pathway in both early and late onset breast cancer. Table (4.11) shows the genes uniquely deregulated in ET and LT group of patients

Table 4.11 Cell adhesion molecule pathway genes differentially expressed uniquely in (a) Early onset tumours (b) Late onset tumours

Early Unique	FC
CDH15(Cadherin 15)	2.452977
PVRL2(Poliovirus receptor-related 2)	2.192558
VCAN (Versican)	2.194877

Late unique	FC
ESAM(Endothelial cell adhesion molecule)	-2.99498
PTPRM(Protein tyrosine phosphatase, receptor type, M)	-2.22114
PVRL3(Poliovirus receptor-related 3)	-2.59583

Besides this some common genes such *as CLDN5 (*claudin 5*), CLDN11 (*claudin 11*), CDH5 (*cadherin 5*), SELE (*Selectin*), JAM2 (*junctional adhesion molecule 2), *ICAM2 (*intercellular adhesion molecule 2*)* were found under expressed in both ET and LT groups in this pathway.

Cell Cycle

A reproducible sequence of events is necessary to accomplish Mitotic cell cycle progression such as DNA replication (S phase) and mitosis (M phase). The genes found uniquely deregulated in ET included *CDKN2A* and *SMAD3* while DEGs in LT included *BUB1, CDKN2C, GADD45B, MAD2L1, MCM2, ORCIL* (Table 4.12)

Table. 4.12 Cell cycle pathway genes differentially expressed uniquely in (a) Early onset tumours (b) Late onset tumours

Early Unique	**FC**
CDKN2A(cyclin-dependent kinase inhibitor 2A)	2.034508
SMAD3(SMAD family member 3)	-2.06139
Late Unique	**FC**
BUB1B(budding uninhibited by benzimidazoles 1 homolog beta)	2.26076
CDKN2C(cyclin-dependent kinase inhibitor 2C)	-2.18721
GADD45B	-2.03287
MAD2L1(MAD2 mitotic arrest deficient-like 1)	3.074305
MCM2(minichromosome maintenance complex component 2)	2.873948
ORC1L(origin recognition complex, subunit 1-like)	2.343208

However genes such as *BUB1 (Budding uninhibited by benzimidazoles 1), CCNB2(cyclin B2), CCNA2(cyclin A2), CDC2 (cell division cycle 2, G1 to S and G2 to M), CCNB1(cyclin B1), MCM4(minichromosome maintenance complex component 4)* were all top upregulated genes in early and late onset tumours.

ECM receptor interaction pathway

The extracellular matrix (ECM) plays an important role in tissue and organ morphogenesis and thus maintains cell /tissue structure and function. The specific interactions amongst cells and ECM are mediated by transmembrane molecules, mainly integrins and also proteoglycans or other cell-surface-associated components which in turn control cellular activities such as adhesion, migration, differentiation, proliferation, and apoptosis. The unique DEGs in ET included *IBSP, ITGA11* while unique DEGs in LT included *LAMA4, LAMAB2, THBS4,* and *TNN* (Table 4.13)

Table 4.13 ECM receptor pathway genes differentially expressed uniquely in (a) Early onset tumours (b) Late onset tumours

Early Unique	**FC**
IBSP(integrin-binding sialoprotein)	2.003757
ITGA11(integrin, alpha 11)	3.87499
Late Unique	**FC**
LAMA4(laminin, alpha 4)	-2.39522
LAMB2(laminin, beta 2 (laminin S))	-2.05976
THBS4(thrombospondin 4)	-5.38855
TNN(tenascin N)	-2.24961

Other genes found upregulated in both the groups included *CD36 a*nd Collagens which included *COL11A1 (collagen, type XI, alpha 1), COL1A1 (collagen, type I, alpha 1), COL5A1(collagen, type V, alpha 1),COL5A2(collagen, type V, alpha 2*).

Pathways in cancer/Metastasis

This Kegg pathway termed Pathway in cancer consisted of a number of genes which were expressing exclusive to either ET/LT group (Table 4.14).

Table 4.14 Pathways in cancer genes differentially expressed uniquely in (a) Early onset tumours (b) Late onset tumours

Early Unique	**FC**
BRAF(v-raf murine sarcoma viral oncogene homolog B1)	2.016751
CDKN2A(cyclin-dependent kinase inhibitor 2A)	2.034508
FZD2(frizzled homolog 2)	2.199027
SMAD3(SMAD family member 3)	-2.06139
Late unique	**FC**
EPAS1(endothelial PAS domain protein 1)	-3.01042
TCF7L1(transcription factor 7-like 1)	-2.65815
WNT3(wingless-type MMTV integration site family, member 3)	2.859969
EGLN3(egl nine homolog 3)	2.94615

Other genes found upregulated in both ET and LT groups included such as *MMP1 (*matrix metallopeptidase 1 (interstitial collagenase)), *WNT2* (wingless-type MMTV integration site family member 2)*, E2F2* (E2F transcription factor 2).

Leukocyte transendothelial migration

Migration of Leukocyte from the blood into tissues is vital for immune surveillance and inflammation. During this diapedesis, leukocytes bind to endothelial cell adhesion molecules (CAM) and then migrate across the vascular endothelium. Hence it shares common genes with cell adhesion molecule pathway. No specific gene was found exclusively expressing in ET at FC of >2 while *CXCL2, ESAM, MYL9, and PXN* were exclusively downregulated in LT.

Table 4.15 Leukocyte transendothelial migration genes differentially expressed uniquely in Late onset tumours

Late unique	FC
CXCL12(chemokine (C-X-C motif) ligand 12)	-3.88799
ESAM(endothelial cell adhesion molecule)	-2.99498
MYL9 (myosin, light chain 9)	-3.3877
PXN(paxillin)	-2.21029

Caludins *(CLDN11, CLDN5 , JAM2(*junctional adhesion molecule 3*), MYL7*(myosin, light chain 7*), RAPGEF3 (*Rap guanine nucleotide exchange factor (GEF) 3*)* were the top genes which were found downregulated in both ET/LT groups.

PPAR signalling pathway

PPARs i.e Peroxisome proliferators-activated receptors are nuclear hormone receptors which are activated by fatty acids & their derivatives. PPAR plays significant role in clearance of cellular/ circulating lipids via regulating gene expression which is involved in lipid metabolism in the liver and skeletal muscle. It is also involved in lipid oxidation and cell

proliferation. This pathway was found downregulated in both ET and LT groups while GK (glycerol kinase) and PLTP (phospholipid transfer protein) expressing genes were found downregulated in LT, no gene was found exclusively expressed in ET group. Genes as *PLIN* (perilipin*), AQP7 (*aquaporin 7*), FABP4 (*fatty acid binding protein 4, adipocyte*), ACADL (*acyl-Coenzyme A dehydrogenase, long chain*), ADIPOQ (*adiponectin, C1Q and collagen domain) were found downregulated in both ET/LT. Besides this Matrix metallopeptidase 1, (interstitial collagenase) and *OLR1* (oxidized low density lipoprotein (lectin-like) receptor 1) were upregulated in both ET/LT groups.

e. Gene Networking

Networking was performed using network analyst software to identify top interacting differential nodes amongst various groups.

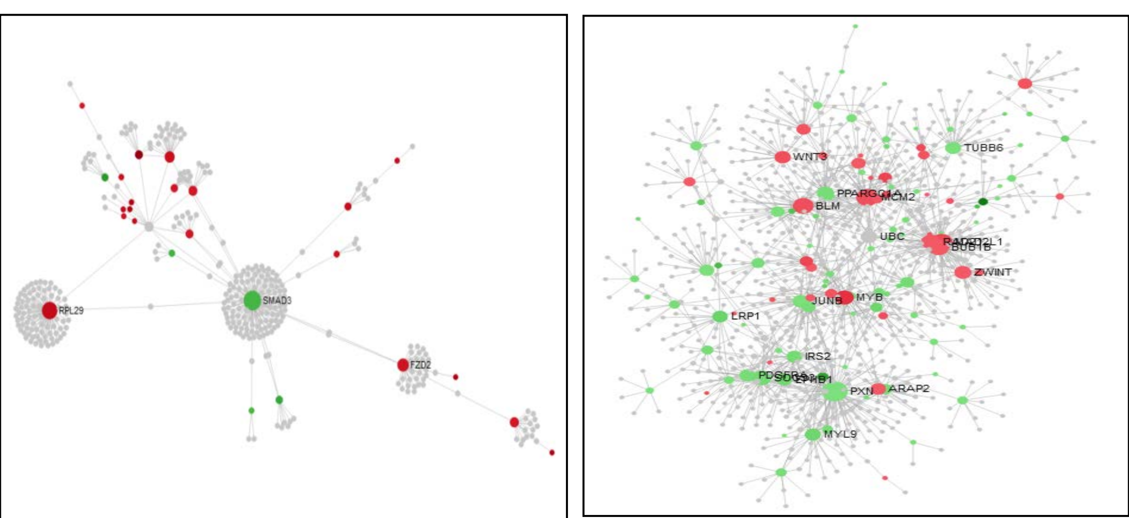


Fig.4.15 a Early unique network **b. Late unique network**

The unique nodes getting upregulated exclusively in ET consisted of *RPL29 and FZD2* while *SMAD3* was the top downregulated node (Fig.19 a). The top nodes getting upregulated exclusively in LT were *MCM2* , *MAD2L1* and top downregulated nodes were *PXN, SOCS3* (Fig. 19.b).

To elucidate the mechanism driving early onset breast carcinogenesis at molecular level, genes which were found to play significant and important functional role in various pathways only in early onset breast carcinogenesis were further evaluated for their interactive networks. These 9 genes were forming highly interactive 198 nodes. *SMAD3* was topmost interactive node and was downregulated while *BRAF* was topmost upregulated interactive node. (Fig.4.16)

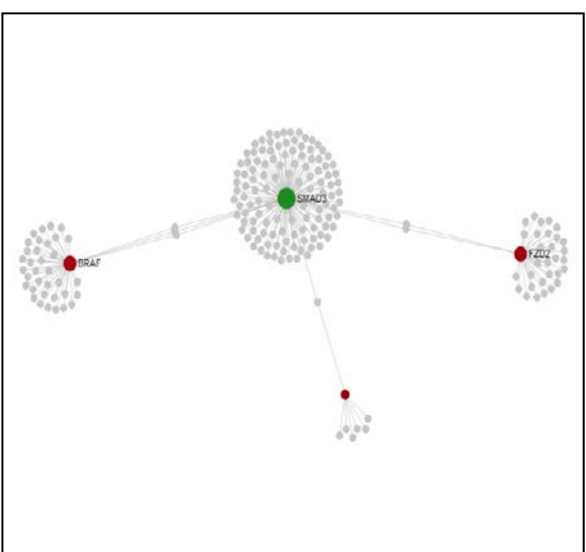


Fig.4.16 Early unique network of genes playing significant functional role in breast carcinogenesis

Further differentially expressed genes amongst various sub groups viz. Molecular subtypes and stagewise were also analyzed

4.8.4.3 Molecular Subtypes

To study the DEG profile in various molecular subgroups of breast cancer, Genefu package in R was used. It categorized 7 cases (24%) as Luminal A, 6 cases (20%) as Luminal B, 6 cases (20%) as HER2/neu over expressing, 9 cases (31%) as Basal and 1 case (3%) as Normal like class. Heat map shows the gene clusters formed by the unsupervised hierarchical clustering of tumour tissues into molecular subtypes (Fig. 4.17). 842, 705, 198, 415, 39 genes were found

deregulated exclusively in basal, HER2/neu, Luminal A, Luminal B and Normal like group of tumours respectively. A set of 534 genes was found commonly deregulated amongst all the 4 breast cancer subtypes viz. Basal, Luminal A, Luminal B and HER-2 excluding Normal like subtype (Fig 4.18) The top differential 20 genes exclusively deregulated to each group are mentioned in (table 4.16).

Table 4.16 Top deregulated genes in various subtypes of breast cancer

Basal	FC	HER2/neu	FC	Lum A	FC	Lum B	FC
MMP1	32.16716	*S100A7*	51.6068	*COL10A1*	36.40927	*TFF1*	45.97365
MMP12	19.28295	*EPYC*	34.41583	*CST1*	34.79067	*COL10A1*	35.18033
GJB2	14.85975	*COL10A1*	33.32071	*EEF1A2*	22.99457	*EEF1A2*	22.05647
KIAA1199	14.63984	*CEACAM6*	24.19824	*MMP11*	21.42561	*S100P*	19.21812
BUB1	14.13919	*S100A7A*	23.90803	*HS.573062*	16.37021	*HS.573062*	17.28372
CST1	12.8087	*MMP11*	20.71085	*TFF1*	13.29112	*MMP11*	17.14796
COL10A1	12.02824	*GJB2*	18.8979	*GJB2*	13.26494	*CST1*	14.48951
CKAP2L	11.79736	*CNTNAP2*	18.47607	*CEACAM6*	12.61417	*C1ORF64*	13.16269
FAM83D	11.39371	*CALML5*	15.72376	*ANKRD30A*	12.24614	*FOXA1*	11.16501
MELK	11.2226	*PNMT*	13.2562	*FOXA1*	12.0476	*AKR7A3*	10.93866
FABP4	-26.0292	*FABP4*	-23.5189	*ALDH1L1*	-19.0648	*FABP4*	-47.0658
GPD1	-29.2149	*SCARA5*	-23.5481	*SAA2*	-19.5392	*LYVE1*	-47.5
C7	-32.81	*SAA1*	-26.9342	*THRSP*	-20.0569	*SAA1*	-48.6369
TIMP4	-35.321	*C7*	-28.5069	*CIDEC*	-20.828	*ADH1B*	-50.1531
CIDEC	-36.1959	*TIMP4*	-29.4988	*ADH1A*	-21.333	*THRSP*	-50.2262
THRSP	-38.9895	*ADH1B*	-31.8829	*GPD1*	-21.6239	*SCARA5*	-55.2669
KIAA1881	-49.6841	*ADH1A*	-36.8982	*SAA1*	-25.292	*PCOLCE2*	-60.4794
ADH1B	-55.2248	*PLIN*	-40.2103	*KIAA1881*	-27.4985	*ADH1A*	-69.4295

PLIN	-76.9024	C2ORF40	-41.4724	*PLIN*	-29.0779	*PLIN*	-83.2191
ADH1A	-99.4174	*KIAA1881*	-62.857	*G0S2*	-30.6163	*KIAA1881*	-107.566

COL10A1, *GJB2, CST1* genes were robustly upregulated in all the 4 molecular subtypes (Basal, Luminal A, Luminal B, HER2/neu), suggesting up regulation of these genes in various cancer related pathways during breast carcinogenesis. *ADH1A, KIAA1881, PLIN* were downregulated amongst all breast cancer subtypes.

Pathway analysis was done for all the molecular subtypes. The unique genes pertaining to specific pathways are mentioned in 74.16 along with their FC.

Table 4.17 Genes deregulated uniquely and playing functional role in various molecular subtypes of breast cancer

Pathway	**Basal-Unique**	**FC**	**HER2/neu -Unique**	**FC**	**Lumina A-unique**	**FC**	**Luminal B-unique**	**FC**
Cell Adhesion Pathway	*ITGB8*	2.8	*CDH15*	3.1	*HLA-E*	-2.8	*CDH1*	3.6
	JAM3	-2.2	*CNTNAP2*	18.5			*CLDN23*	-2.7
							ITGA9	-2.2
Complement and coagulation pathway	*C4BPB*	-2.7	*FGB*	11.5	*A2M*	-2.5	*SERPING1*	-2.1
			FGG	6.5				
			PLAU	3.2				
ECM receptor	*THBS4*	-7.6	*ITGA10*	-2.1	*COL3A1*	2.2	*LAMA4*	-2.4
							LAMC1	-2.1
Pathways in cancer	*BRAF*	2.3	*BCL2*	-2	*WNT3*	3.4		
	CKS1B	2.1	*CSF3R*	3.1			*GSTP1*	-2.2
	E2F3	2.4	*EGLN3*	4.8			*IGF1R*	4.3
	JUN	-2.4	*ERBB2*	2.6			*RET*	3.1
	STAT1	2.1	*FGF2*	-6.9			*RXRA*	-2.1
			MYC	-3.9			*TCF7L1*	-3.8

			SMAD3	-2.2				
			TRAF4	2.1				
			WNT7B	2				
PPAR	*EHHADH*	-2.6	*FADS2*	6.2	*ME1*	-3.7	*RXRA*	-2.1
	GK	3.1						
	SLC27A1	-2.3						

Besides these differentially expressed unique genes, some common genes were also found specific to pathways such as Cell Adhesion (*CADM3, CDH5, CLDN5, ICAM2, and JAM2*), Complement and coagulation cascade (*CD59, CFD, CFH, F10, PROS1, VWF*), ECM receptor interaction (*COL11A1, COL1A1, COL5A2, ITGA7*), Pathways in Cancer (*ACVR1C,BMP2, EPAS1, EVI1, FIGF, FN1, IGF1)* PPAR Pathway (*ACADL, ACSL1, ADIPOQ, AQP7, CD36, FABP4, LPL, OLR1, PCK1, PLIN, PPARG, SORBS1* in all subgroups.

Interestingly, highest numbers of deregulated genes were present in Basal subtype followed by HER-2 and Luminal B and Luminal A. On other hand Normal-like subtype include lowest number of deregulated genes. All the normal samples clustered separately from luminal, basal, HER2/neu over expressing groups.

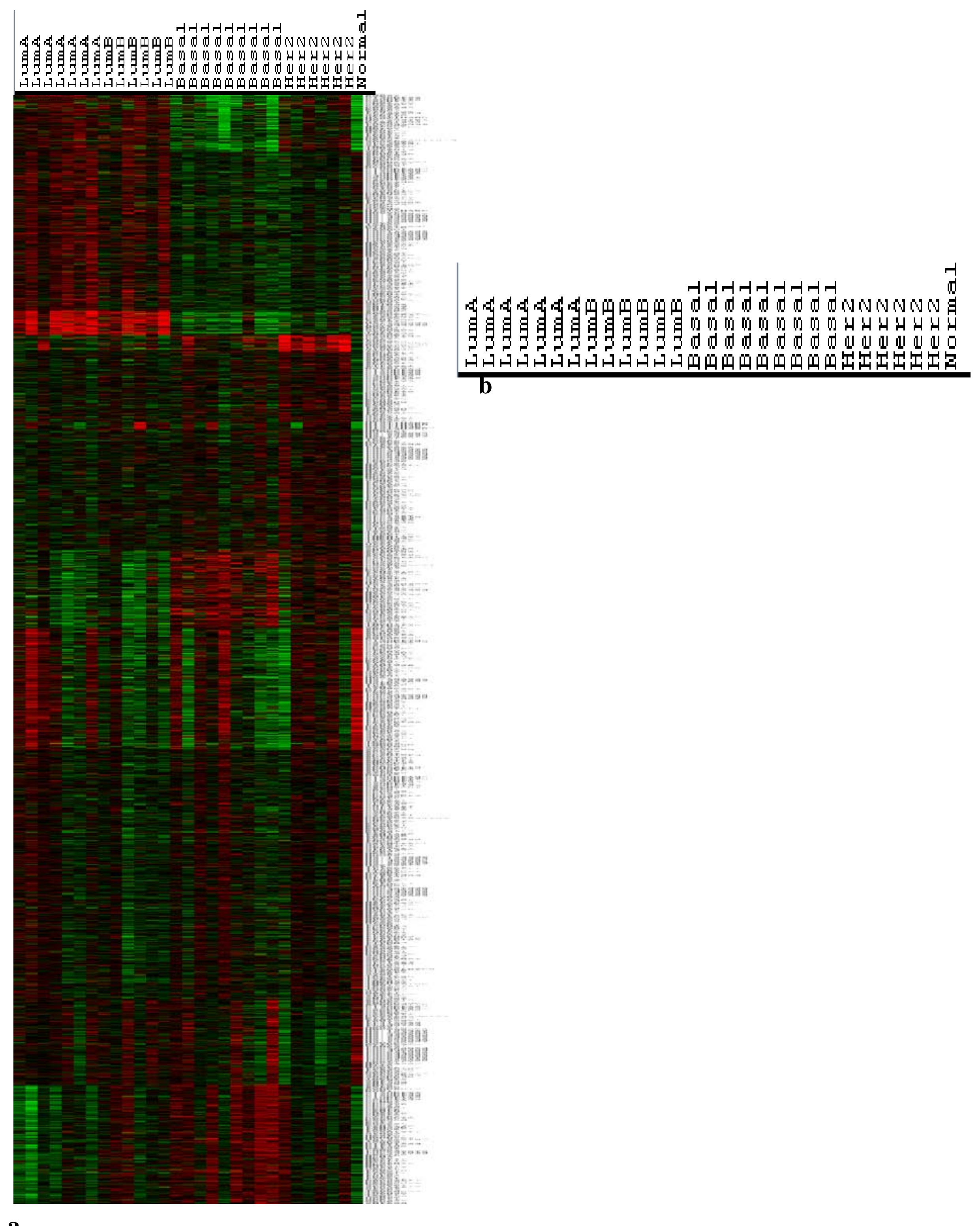


Fig.4.17 Breast cancer predicted molecular subtype in Indian breast cancer patients. The subtypes LumimalA, LuminaB, HER2/neu, Basal and normal like were identified. (a) Heat map (b) Sample Dendorgram

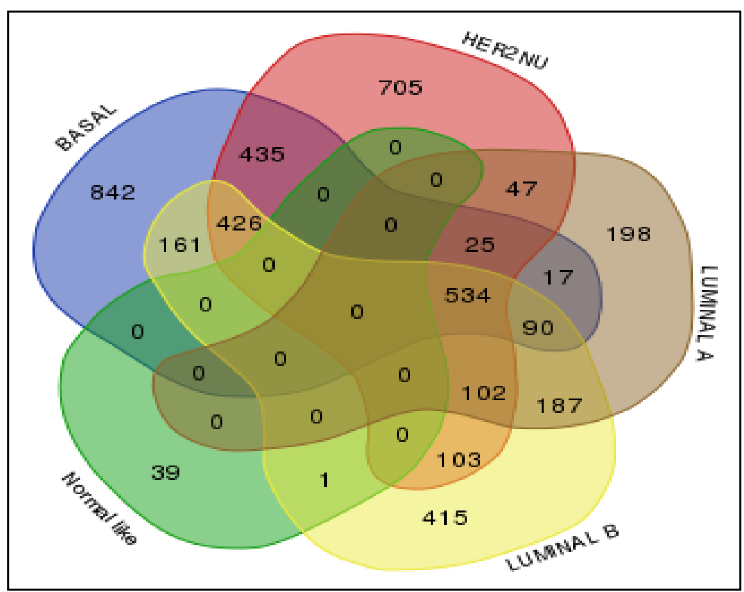


Fig.4.18 Venn diagram representing the common and unique genes belonging to all the five predicted molecular subtypes of Indian breast cancer patients

a. Basal Group

Since early onset breast tumours possess higher percentage of basal molecular subtypes of breast cancer hence we build up a network for basal group of tumours to identify the top interactive nodes.

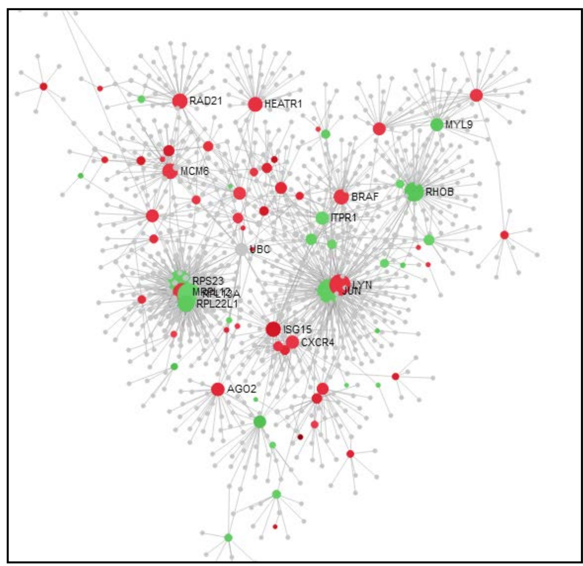


Fig. 4.19 Network image for unique basal subtype genes

JUN and *RPS23* were the top interactive downregulated nodes found in basal group of tumours. *LYN* was found to be the top upregulated node.

b. HER2/neu overexpressing group

Networking analysis of HER2/neu over expressing group of tumours identified top node as MYC which was found to be downregulated while next highly interactive node was ERBB2 which was upregulated node. Other upregulated nodes were CDC6, CSNK2A corresponding uniquely to Her2Nu over expressing group of tumours (Fig. 4.20).

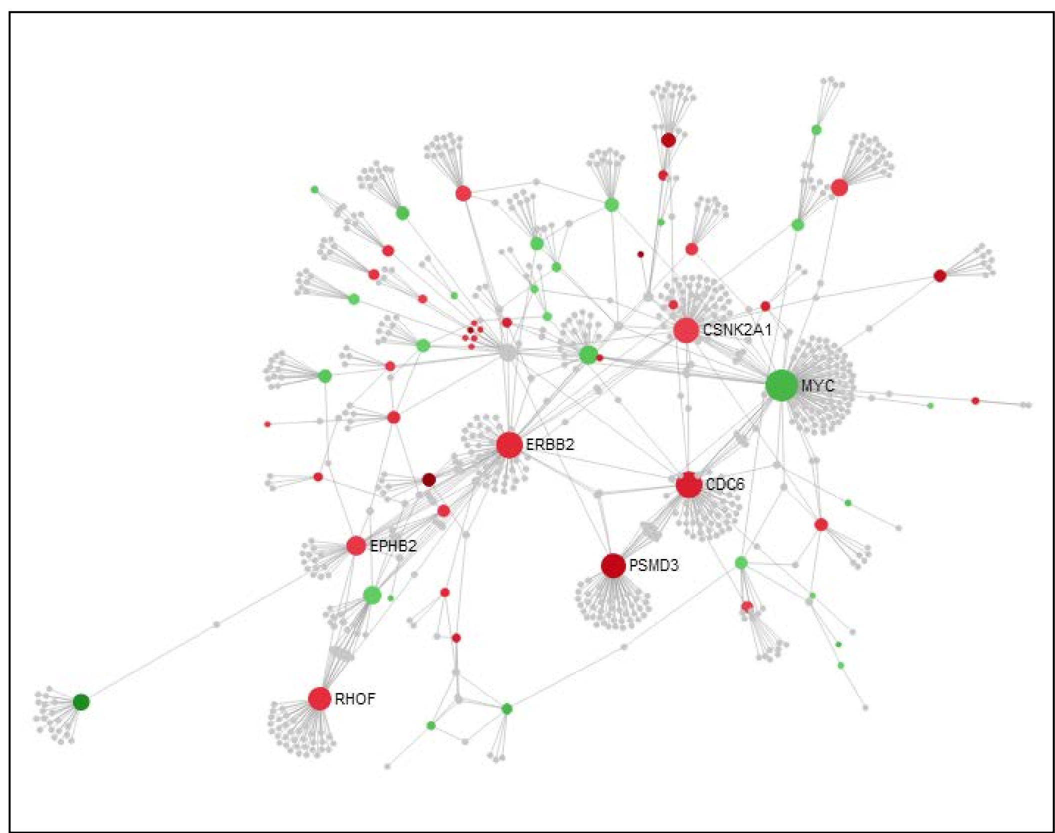


Fig.4.20**.** Network image for unique HER2/neu subtype genes

c. Luminal A and Luminal B subtypes

Lum A group of tumours possessed TLE1 as the highly interactive and top node which was found to be downregulated while WNT3 was top upregulated node (Fig. 4.19).

LumB group of tumours showed higher number of networks including top genes as IGF1R, CDH1, RET which were upregulated while SOCS3 was downregulated (Fig. 4.21, 4.22).

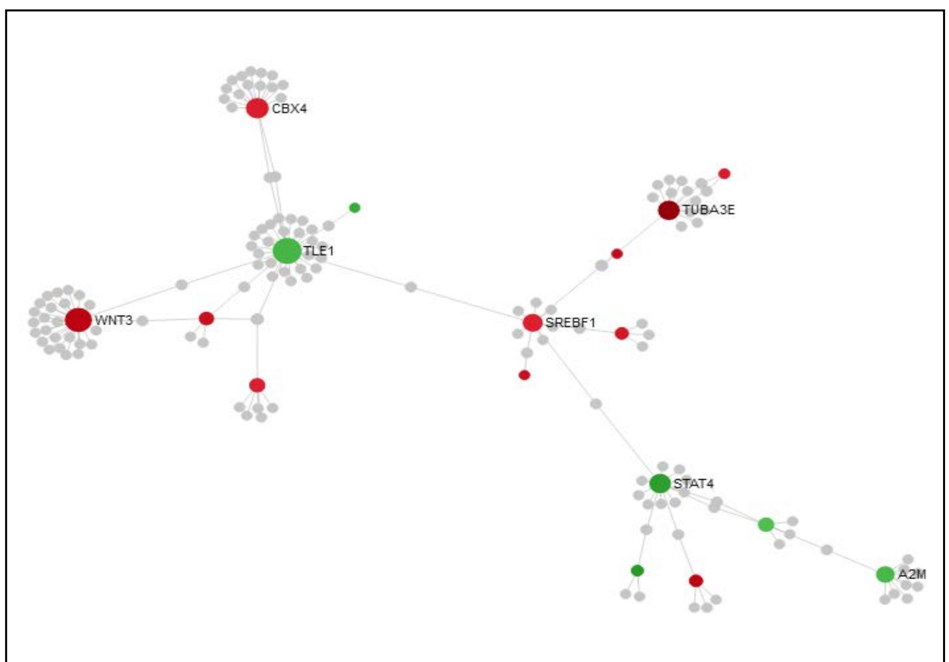


Fig. 4.21 Network showing unique Luminal A interactive genes

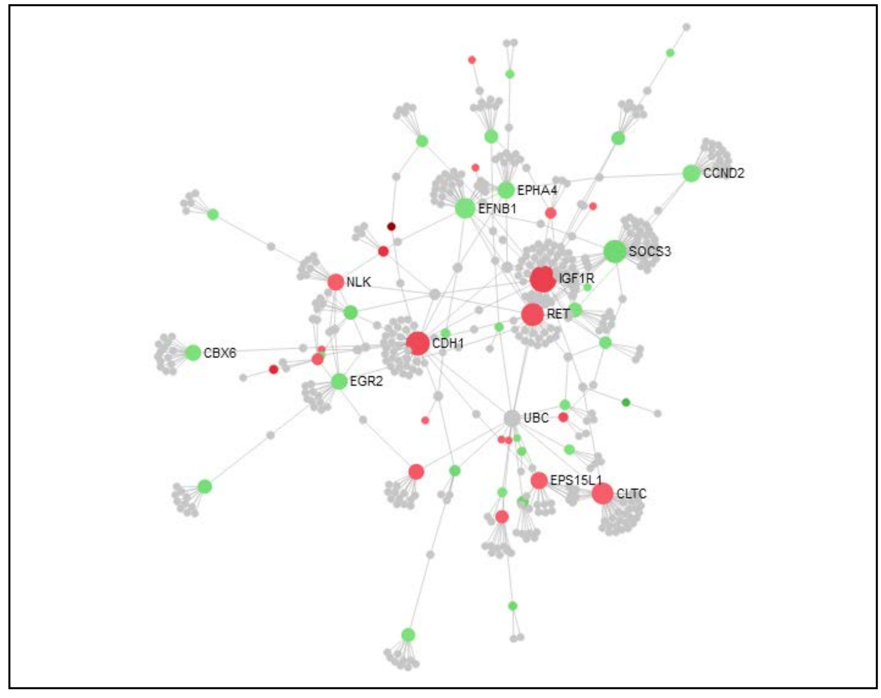


Fig.4.22. Network showing unique Luminal B interactive genes

4.8.4.4 Advanced stage (III, IV) vs Lower stage (I, II)

Categorization of DEG profile according to different tumour stage, identified 1386 deregulated genes unique to LS and 200 genes unique to AS while 1336 genes were commonly deregulated in both LS and AS (Fig. 4.23).

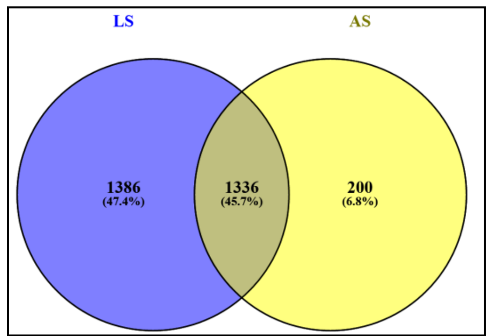


Fig. 4.23 Venn diagram showing unique and common genes in advanced and lower stages

Pathways getting deregulated in AS/LS were also identified It was observed that AS group of tumours possessed unique genes belonging to mainly 3 pathways (table 4.18) while LS also showed deregulation similar pathways but with different genes (table 4.18). Rest of the genes were present in both LS and AS tumours.

Table 4.18 List of pathways and genes deregulated uniquely in advanced stage of tumours

Pathway	**AS- Unique Genes**	**FC**
Cell Adhesion	*VCAN(versican)*	2.0194
ECM receptor interaction	*COL1A2(collagen, type I, alpha 2)*	2.406966
	COL3A1(collagen, type III, alpha 1)	2.217713
	ITGA11(integrin, alpha 11)	3.55656
	TNN(tenascin N)	-2.00255
Pathways in cancer	*JUN(jun oncogene)*	-2.38138
	MYC(v-myc myelocytomatosis viral oncogene homolog)	-2.51316

Table 4.19 List of pathways and genes deregulated uniquely in lower stage of tumours

Pathway	**LS Unique Genes**	**FC**
Cell Adhesion	*JAM3(junctional adhesion molecule 3)*	-2.2021
	NLGN1(neuroligin 1)	-3.88356
	PTPRM(protein tyrosine phosphatase, receptor type)	-2.35086
	PVRL3(poliovirus receptor-related 3)	-3.04221
Cell cycle	*BUB1B(budding uninhibited by benzimidazoles 1 homolog beta)*	2.733421
	CCND2(cyclin D2)	-2.219
	CCNE2(cyclin E2)	2.213775
	CDKN2C(cyclin-dependent kinase inhibitor 2C)	-2.16278
	MCM2(minichromosome maintenance complex component 2)	2.83024
ECM receptor	*LAMA3(laminin, alpha 3)*	-2.47639
	LAMA4(laminin, alpha 4)	-2.35986
	LAMB2(laminin, beta 2 (laminin S))	-2.25598
	THBS4(thrombospondin 4)	-5.15772

Genes such as *Cadherins (CDH5), claudins (CLDN5, CLDN11), Cyclin (CCNB1, CCNB2, CCNE1), Collagen (COL11A1, COL1A1, COL5A1, COL5A2)* were some of the commonly deregulated genes in both AS and LS.

As we wanted to study the biology of young onset breast cancer which in turn possesses advanced stage of tumours, hence we did networking for the AS group of tumours to identify the top deregulated nodes. Nodes Myc and Jun were the top highly interactive downregulated nodes present in AS of tumours.

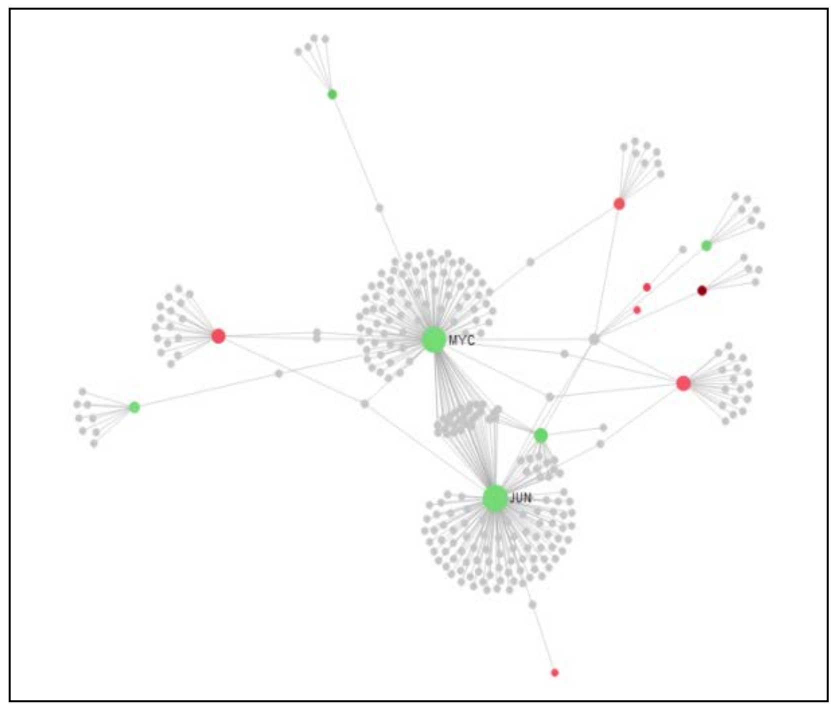


Fig. 4.24 Network showing unique higher stage interactive genes

a. Lower Stage tumours

Networking analysis was performed in lower group of tumours to identify the top nodes. Genes *RPL29, PRKDC* were the top upregulated nodes and *PRKCA* was top downregulated node (Fig. 4.23).

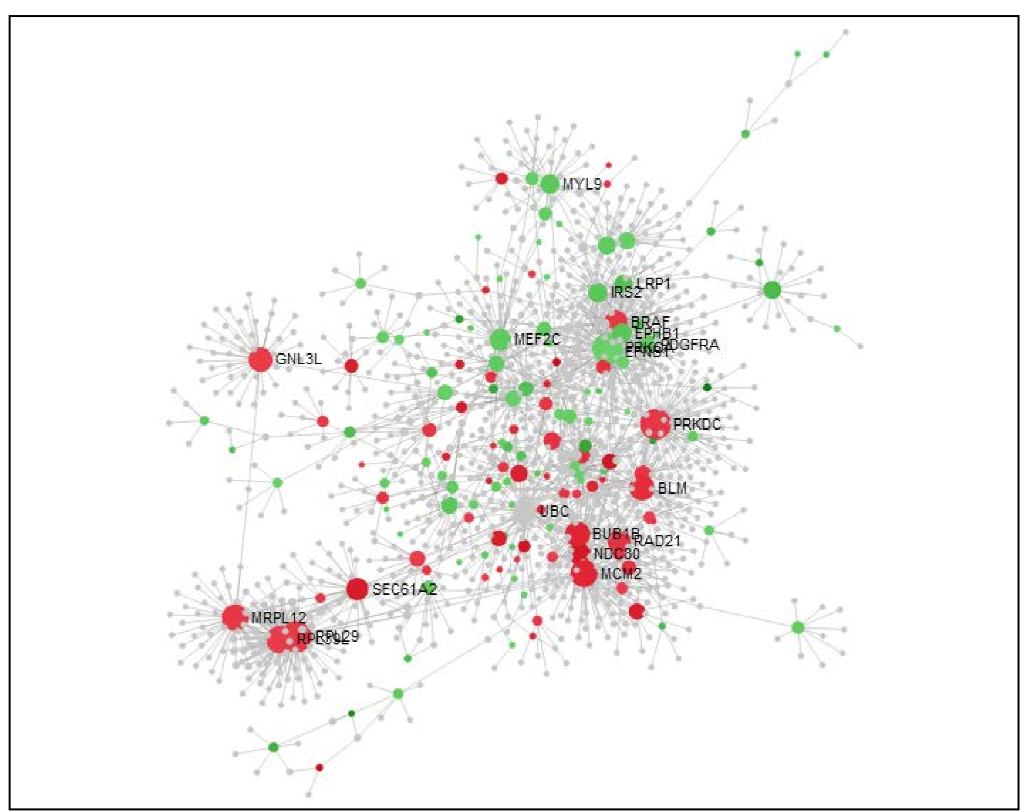


Fig. 4.25 Network showing unique lower stage interactive genes

4.8.4.5 Comparison of differentially expressed genes in Indian and western population

Molecular signatures found in breast cancer in Indian women were further compared with molecular signatures reported in western population. Using cutoff for FC> 2 and adjusted p value <0.05 we identified 5062 genes reported as significantly differentially expressed (3789 upregulated and 1273 downregulated) in tumours vs normals in western population(176-177). When compared to DEGs identified in our data we found 558 genes exclusive and unique to Indian population having 156 over expressed and 402 under expressed genes. These unique genes included *KIAA1199, CKAP2L, TUBB3, CENPA* significantly over expressed while *PLIN, KIAA1881, ADH1A, ALDH1L1, SAA1, AQP7P2, AQP7* genes were under expressed in Indian breast cancer patients (Table 4.20). Besides these top genes all the remaining genes were analyzed for various pathways which showed Leukocyte transendothelial migration

[*ESAM, MYL7]* Cytokine cytokine receptor interaction pathway [*IL17B, CNTFR, FIGF*, *MPL*, and *CCL21*], Adherens junction [*PVRL3, PVRL4*, and *TCF7l1*] majorly deregulated.

Table.4.20 Unique gene sets specific to Indian breast cancer population

KIAA1199	9.509539	0.000351
CKAP2L	6.264448	9.48E-05
TUBB3	6.044887	0.00046
CENPA	5.311936	5.86E-05
CDC45L	4.832954	0.001179
C6ORF126	4.722761	0.028214
PPAPDC1A	4.614044	0.000137
KLK4	4.404234	0.020902
CDC2	4.340372	0.000164
PTTG3P	4.210745	0.000407
FIGF	-10.6485	1.73E-07
SAA2	-11.9734	0.000866
XLKD1	-13.1265	2.6E-06
DARC	-15.1034	0.000104
AQP7	-16.3904	3.36E-10
AQP7P2	-18.3621	3.36E-10
SAA1	-18.744	3.74E-05
ADH1A	-46.9269	1.12E-05
KIAA1881	-48.6453	2.71E-07
PLIN	-49.7345	3.12E-08

A set of 715 genes were found to be commonly deregulated in Indian and western dataset, of which 247 genes were upregulated and 464 genes were downregulated (Table 4.21). However amongst them 4 genes *XIST, KLF6, CCNL1* and *CD59* were found upregulated in western data while they were downregulated in Indian data (Table 4.22).

Table 4.21 Top commonly deregulated gens in both Indian and western population in breast cancer

COL10A1	23.2815	5.96E-06
MMP11	15.45492	6.55E-06
GJB2	13.68954	8.27E-06
CST1	11.94889	0.004912
MMP1	9.096327	0.008953

MMP13	7.297567	0.002203
CEACAM6	7.014203	0.042617
BUB1	6.795974	9.11E-05
ASPM	6.500159	2.39E-05
FAM83D	6.454691	0.000158
C2ORF40	-21.1401	2.51E-06
SCARA5	-21.6551	2.8E-06
G0S2	-22.0594	3.68E-07
C7	-22.871	1.39E-05
FABP4	-23.8769	0.000104
TIMP4	-24.1717	1.03E-07
GPD1	-24.9122	3.83E-10
THRSP	-26.7033	3.01E-05
CIDEC	-27.4842	3.18E-08
ADH1B	-33.9347	3.6E-08

Table 4.22 Genes getting commonly deregulated but with opposing directions in both populations

Gene	FC-Western	FC-Indian
XIST	4.929642686	-2.142038315
KLF6	3.367593079	-2.14940167
CCNL1	3.043093301	-2.012116527
CD59	2.483964818	-2.579648816

We also compared our early onset patient's data with the previously identified young age onset gene signatures (16). 2 genes *BAX (1*.6 FC, p value 4.52E-07) and *ALDH2* (-3.36 FC, p value 0.000113) expressing differentially.

4.8.4.6 Validation of gene expression profiles

Genes which were found differentially expressed in microarray were validated in 67 breast cancer tissues (Table 4.23) and 38 distant control normals (13 early and 25 late) along with 2 commercial available breast total RNA by real Time PCR.

Table 4.23 Clinicopathological details of patients used in real time validation

S.NO	RT-ID	AGE	DIAGNOS	TNM	TSIZE	LYMPH N(	METASTA:	STAGE	ER	PR	HER2NU	MENARCH	BREAST FEE	FAMILY H	MARRIAG	CHEMOTH
1	ET-16	40	IDC	T3N0M0	T3	N0	M0	IIB	P	P	N	15	YES	NO	NO	NO
2	ET-29	40	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	N	15	YES	NO	NO	NO
3	ET-30	34	IDC	TXNXMX	TX	NX	MX	X	P	N	N	14	NO	NO	NO	NO
4	ET-32	40	IDC	T4N1M0	T4	N1	M0	IIIB	P	P	N	13	YES	NO	NO	NO
5	ET-33	35	IDC	T4bNxM1	T4	N1	M1	IV	P	P	N	14	YES	NO	NO	NO
6	ET-34	41	IDC	T4bN1M0	T4	N1	M0	IIIB	P	P	N	13	NO	NO	NO	NO
7	ET-36	25	IDC	T4N2M1	T4	N2	M1	IV	N	P	N	11	NO	NO	NO	NO
8	ET-37	40	IDC	T2N1M0	T2	N1	M0	IIB	N	P	N	12	YES	NO	NO	NO
9	ET-38	35	IDC	T4N2M0	T4	N2	M0	IIIB	P	N	N	13	YES	NO	NO	NO
10	ET-41	35	IDC	T2N1M0	T2	N1	M0	IIB	P	P	N	14	YES	NO	NO	NO
11	ET-47	40	IDC	T2N1M0	T2	N1	M0	IIB	N	P	N	13	YES	NO	NO	NO
12	ET-50	30	IDC	T2N0M0	T2	N0	M0	IIA	N	N	N	13	YES	NO	NO	NO
13	ET-51	32	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	N	13	YES	NO	NO	NO
14	ET-52	35	IDC	T4bN1M0	T4	N1	M0	IIIB	N	N	N	15	YES	NO	NO	NO
15	ET-55	40	IDC	T2N1M0	T2	N1	M0	IIB	P	P	N	14	YES	NO	NO	NO
16	ET-58	36	IDC	T4BN2M0	T4	N2	M0	IIIB	N	N	N	13	YES	NO	NO	NO
17	ET-59	36	IDC	T4BN1M0	T4	N1	M0	IIIB	N	N	N	13	YES	NO	NO	NO
18	ET-60	40	IDC	T2N1M0	T2	N1	M0	IIB	N	N	N	13	YES	NO	NO	NO
19	LT-18	57	IDC	T2N1M0	T2	N1	M0	IIB	N	N	N	13	yes	NO	NO	NO
20	LT-22	65	IDC	T2N0M0	T2	N0	M0	IIA	P	P	N	15	yes	NO	NO	NO
21	LT-24	65	IDC	T2N0M0	T2	N0	M0	IIA	P	P	N	16	yes	NO	NO	NO
22	LT-25	65	IDC	T4BN1M0	T4	N1	M0	IIIB	N	N	N	14	yes	NO	NO	NO
23	LT-29	70	IDC	T2N0M0	T2	N0	M0	IIA	P	P	N	15	yes	NO	NO	NO
24	LT-40	70	IDC	T1N0M0	T1	N0	M0	I	P	N	N	15	YES	NO	NO	NO
25	LT-53	65	IDC	T2N1M0	T2	N1	M0	IIB	N	N	N	15	YES	NO	NO	NO
26	LT-54	80	IDC	T2N1M0	T2	N1	M0	IIB	N	N	N	14	YES	NO	NO	NO
27	LT-55	57	IDC	T3N1M0	T3	N1	M0	IIIA	P	P	N	15	YES	NO	NO	NO
28	LT-62	70	IDC	T2N0M0	T2	N0	M0	IIA	N	P	N	15	YES	NO	NO	NO
29	LT-66	65	IDC	T4BN1M0	T4	N1	M0	IIIB	P	N	N	14	YES	NO	NO	NO
30	ET-18	35	IDC	T2N1M0	T2	N1	M0	IIB	P	N	P	14	yes	NO	NO	NO
31	ET-19	32	IDC	T2N1M0	T2	N1	M0	IIB	P	P	P	13	yes	NO	NO	NO
32	ET-20	33	IDC	T1N1M0	T1	N1	M0	IIA	N	N	P	13	yes	NO	NO	NO
33	ET-22	39	IDC	T3N1M0	T3	N1	M0	IIIA	P	P	P	13	yes	NO	NO	NO
34	ET-23	28	IDC	T4bN1M0	T4	N1	M0	IIIB	N	N	P	15	no	NO	NO	NO
35	ET-24	40	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	15	yes	NO	NO	NO
36	ET-25	30	IDC	T2N1M0	T2	N1	M0	IIB	N	N	P	14	yes	NO	NO	NO
37	ET-26	41	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	13	yes	NO	NO	NO
38	ET-27	37	IDC	T2N0M0	T2	N0	M0	IIA	N	N	P	13	yes	NO	NO	NO
39	ET-28	40	IDC	T2N0M0	T2	N0	M0	IIA	P	N	P	13	yes	NO	NO	NO
40	ET-46	35	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	14	YES	NO	NO	NO
41	ET-53	35	IDC	T3N1M0	T3	N1	M0	IIIA	P	P	P	13	YES	NO	NO	NO
42	ET-54	37	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	15	YES	NO	NO	NO
43	ET-57	E	IDC	T3N1M0	T3	N1	M0	IIIA	P	N	P	13	YES	NO	NO	NO
44	ET-61	E	IDC	T4BN1M0	T4	N1	M0	IIIB	N	N	P	15	YES	NO	NO	NO
45	LT-19	70	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	14	yes	NO	NO	NO
46	LT-23	60	IDC	T4bN1M0	T4	N1	M0	IIIB	N	N	P	14	yes	NO	NO	NO
47	LT-26	56	IDC	T2N1M0	T2	N1	M0	IIB	P	P	P	13	yes	NO	NO	NO
48	LT-31	71	IDC	T2N0M0	T2	N0	M0	IIA	P	P	P	14	yes	NO	NO	NO
49	LT-32	60	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	13	yes	NO	NO	NO
50	LT-33	59	IDC	T2N1M0	T2	N1	M0	IIB	P	N	P	15	yes	NO	NO	NO
51	LT-34	60	IDC	T3N0M0	T3	N0	M0	IIB	P	P	P	14	NO	NO	NO	NO
52	LT-35	55	IDC	T4BN1M0	T4	N1	M0	IIIB	N	N	P	13	YES	NO	NO	NO
53	LT-36	62	IDC	T4N1M0	T4	N1	M0	IIIB	N	N	P	14	YES	NO	NO	NO
54	LT-38	69	IDC	T4bN1M0	T4	N1	M0	IIIB	N	N	P	14	YES	NO	NO	NO
55	LT-39	65	IDC	T2N1M0	T2	N1	M0	IIB	N	N	P	13	YES	NO	NO	NO
56	LT-41	70	IDC	T2N1M0	T2	N1	M0	IIB	N	N	P	16	YES	NO	NO	NO
57	LT-47	63	IDC	T1N1M0	T1	N1	M0	IB	P	P	P	13	YES	NO	NO	NO
58	LT-49	58	IDC	T2N1M0	T2	N1	M0	IIB	P	P	P	14	yes	NO	NO	NO
59	LT-52	56	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	14	YES	NO	NO	NO
60	LT-56	75	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	13	YES	NO	NO	NO
61	LT-59	61	IDC	T2N1M0	T2	N1	M0	IIB	P	N	P	15	YES	NO	NO	NO
62	LT-61	57	IDC	T2N1M0	T2	N1	M0	IIB	N	N	P	13	YES	NO	NO	NO
63	LT-63	58	IDC	T4BN2M0	T4	N2	M0	IIIB	P	P	P	13	YES	NO	NO	NO
64	LT-64	62	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	12	YES	NO	NO	NO
65	LT-45	80	IDC	T4BN2M1	T4	N2	M1	IV	NA	NA	NA	15	YES	NO	NO	NO
66	LT-46	55	IDC	T2N1M0	T2	N1	M0	IIB	NA	NA	NA	13	YES	NO	NO	NO
67	LT-57	58	IDC	T3N1M0	T3	N1	M0	IIIA	NA	NA	NA	12	YES	NO	NO	NO

Genes from Metallopeptidase (*MMP1*, *MMP13*, *MMP14*, MMP3, *MMP11*, *ADAMTS1* and *ADAMTS5*) activity were selected for validation since several of these genes were found in top 50 genes in total tumour data set (TT) (Table 4.24 a) and also they were found differentially expressed in early onset tumours (Table 4.25). When compared to normal controls genes as *MMP1* (FC= 15.4 fold, p=.05), *MMP13* (FC=12.3 fold, p=0.018), *MMP11* (FC=6.8 fold, p=0.03) were all significantly over expressed in TT (Table 4.24 b) while gene *MMP14* (FC=0.48 fold, p=0.7) and MMP3 (FC= 6.8 , p= 0.214) expression did not reached the level of significance. Genes as *ADAMTS1* (FC= -9.4 fold, p=0.009), and *ADAMTS5* (FC=-5.7 fold, p=0.018) showed significant down regulation in tumours as compared to normal controls (Table.4.24 b). Scatter plots are plotted for all these genes (Fig.4.24). Further the expression of these genes was also found differential amongst ET and LT as seen in microarray data (Table.4.25), hence their expression was further validated amongst ET/LT. Tissue qPCR showed differential expression of all these genes between ET and LT but only *ADAMTS5* expression was found statistically significant (Table. 4.26). Down regulation of *ADAMTS5* was found to be -6.5 fold in late tumours as compare to -4.5 fold in early tumours (table.4.26) (p=0.013), suggesting its significant association with late onset tumours (Fig. 4.25). Expression of various MMP genes was although found upregulated in early onset tumours but it was not found statically significant. Further validation in large number of patient's samples is warranted to establish their role in early onset tumours.

Table 4.24 Showing expression MMP activity genes in all tumours (a) micro array data and (b) qPCR

GENE	FC-GX	p. Val
ADAMTS1	-5.53083	4.48E-06
ADAMTS5	-3.47433	0.000893
MMP1	9.096327	0.008953
MMP11	15.45492	6.55E-06
MMP13	7.297567	0.002203
MMP14	2.518856	0.006155
MMP3	4.822623	0.016434

a

GENE	FC-qPCR	p value
ADAMST1	-9.40506	0.009
ADAMST5	-5.66265	0.018
MMP1	15.42609	0.05
MMP11	6.820807	0.03
MMP13	12.34492	0.018
MMP14	0.484605	0.722
MMP3	6.683773	0.214

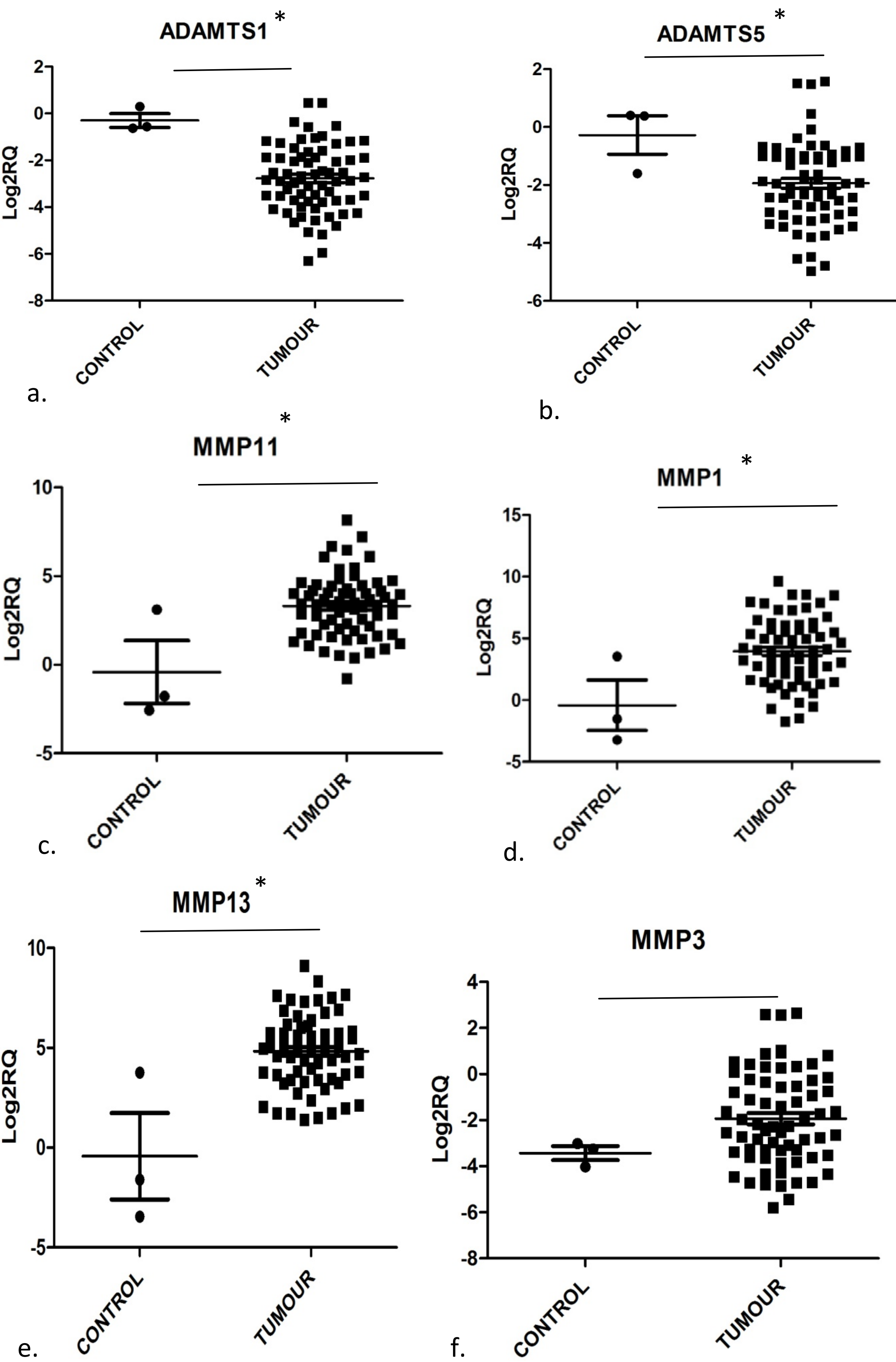
b

Table 4.25 Fold change in breast cancer tumour tissue in age specific group's viz early onset (ET) and late onset (LT) breast cancer tumours as obtained from microarray

GENES	FC-GX-EARLY	FC-GX-LATE	p.value
ADAMTS1	-3.99311	-6.96077	1.19E-08
ADAMTS5	-2.68595	-4.16648	3.20E-05
MMP1	13.78096	6.78446	0.001196
MMP11	20.16884	12.80727	1.62E-07
MMP13	8.923938	6.331392	0.000343
MMP14	3.75373	1.900651	1.26E-05
MMP3	5.746106	4.261598	0.004459

Table 4.26 Real time PCR validation in breast cancer tumour tissue in age specific groups viz early onset (ET) and late onset (LT) breast cancer tumours

GENE	qPCR-EARLY	qPCR-LATE	p value
ADAMST1	-7.76766	-10.9943	0.25
ADAMST5	-4.54458	-6.74784	0.013
MMP1	17.52682	13.44894	0.603
MMP11	7.807872	5.862773	0.554
MMP13	14.96795	9.799033	0.985
MMP14	0.542113	0.428789	0.985
MMP3	7.284474	6.10074	0.669


ADAMTS1 *
Log2RQ
2
0
-2
-4
-6
-8
CONTROL
TUMOUR
a.
ADAMTS5 *
Log2RQ
2
0
-2
-4
-6
CONTROL
TUMOUR
b.
MMP11 *
Log2RQ
10
5
0
-5
CONTROL
TUMOUR
c.
MMP1 *
Log2RQ
15
10
5
0
-5
CONTROL
TUMOUR
d.
MMP13 *
Log2RQ
10
5
0
-5
CONTROL
TUMOUR
e.
MMP3
Log2RQ
4
2
0
-2
-4
-6
-8
CONTROL
TUMOUR
f.

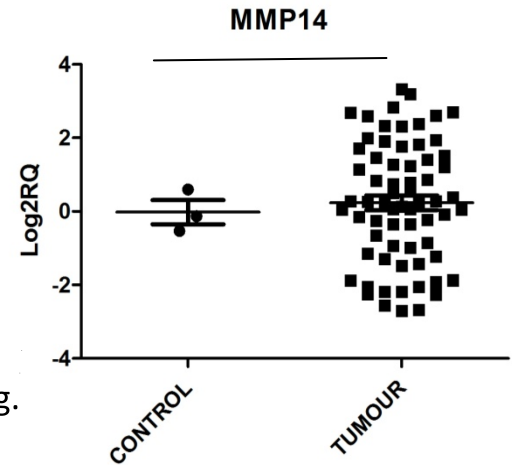
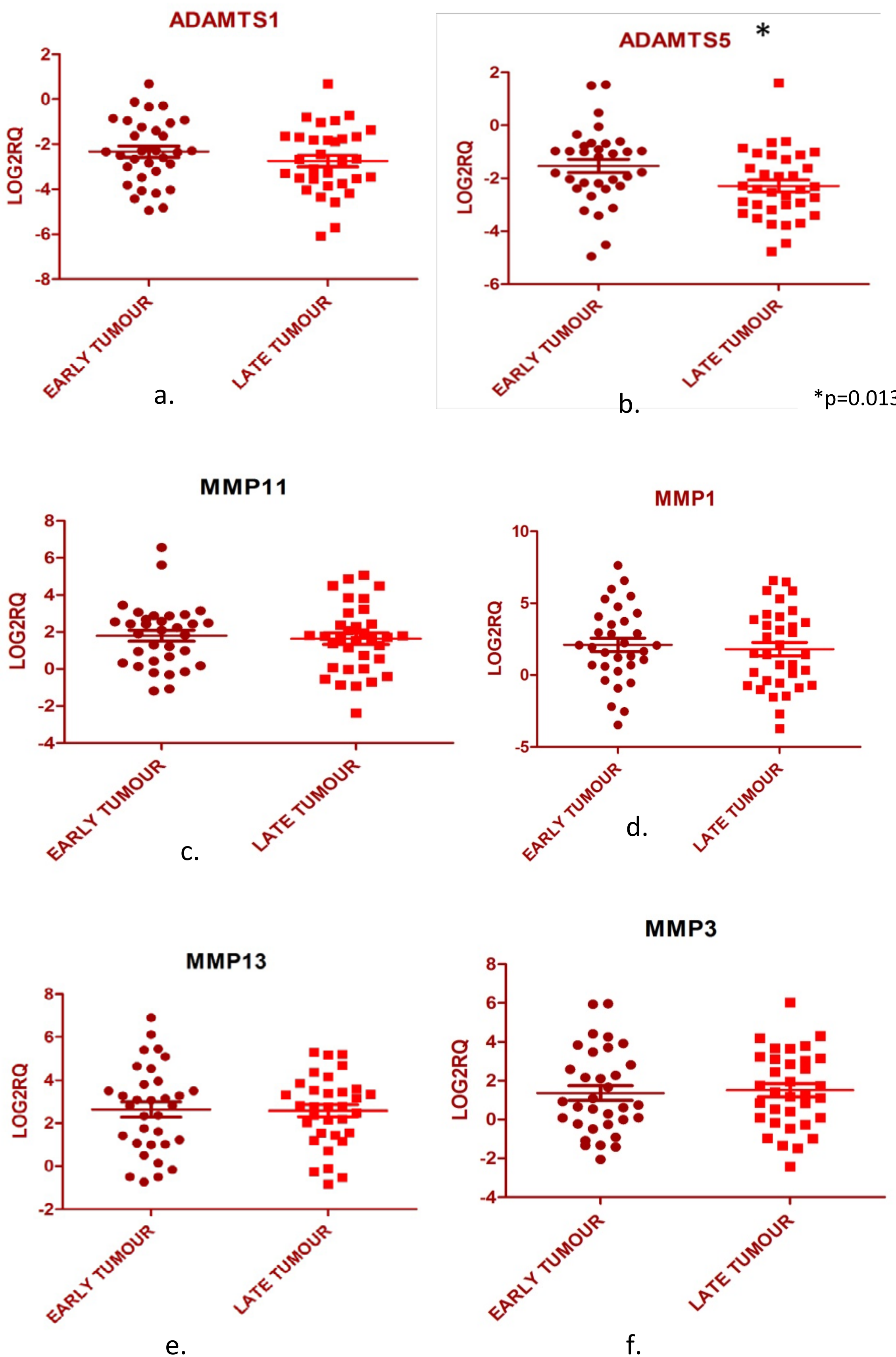


Fig. 4.26 Validation of microarray results by qRTPCR in Indian population. a,b) Scatter plots showing the significant downregulation of ADAMTS1 (p=0.009) and ADAMTS5 (p=0.018) and c,d,e,f,g) showing upregulation of MMP11 (p= 0.03), MMP1 (p=0.05), MMP13 (p= 0.018), MMP3 (p= 0.214), MMP14 (p= 0.722) respectively in all tumour compared to controls. The values are the mean of log fold change normalized to beta actin, 18sRNA and PSMC4 along with the standard error shown by vertical bars as obtained by Mann Whitney U test with the p value (* p≤ 0.05) as obtained from real time qPCR


ADAMTS1
LOG2RQ
EARLY TUMOUR
LATE TUMOUR
a.
ADAMTS5
*
b.
*p=0.013
MMP11
c.
MMP1
d.
MMP13
e.
MMP3
f.

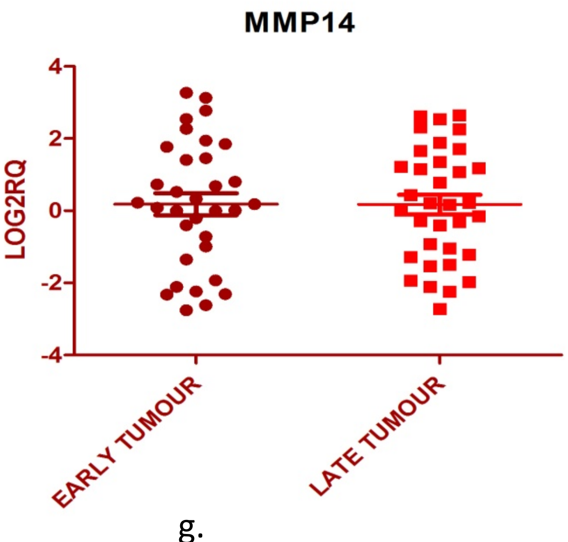


Fig. 4.27 Scatter plot showing the differential expression of MMP's in early and late onset tumours. Here figure a, b, c, d, e, f, g shows the differential fold changes for genes ADAMTS1, ADAMTS5, MMP11, MMP1, MMP13, MMP3, MMP14. Expression of ADAMTS5 was found to be significantly downregulated in early and late onset tumours (p=0.013) with higher fold downregulation in late onset tumours.

Further clinicopathological association with these MMPs was also identified. Expression of *MMP1* gene was found significantly associated with ER negative (p=0.01), PR negative (p=0.006) while expression of *MMP13* and *MMP14* was found correlated with HER2/neu positivity (p=0.023 and p=0.084 respectively). Metastasis was found to have significant positive association with over expression of *MMP11* gene (p=0.04).

In the present study we have identified pathways related to Cell cycle, Cell adhesion in all breast tumours, however, when we have analysed early and late onset tumours we found pathways cell adhesion, cell cycle, ECM receptor interaction, Pathways in cancer, Focal adhesion signalling involving 9 genes majorly altered in early onset tumours. The top altered genes exclusive to Indian females was identified. Thus we have shown in the present study the differential expression profile of early onset breast tumours, this is to our knowledge is the

first study analysing gene expression profiles of early and late onset tumours particularly form Indian continent.

4.8.4.7 Discussion

Over the past two decades there has been a significant increase in breast cancer incidence all over Asia, mainly South Eastern Asia (178-180). However the incidence of breast cancer in Asian women is much lower as compared to western countries, also the age at diagnosis peaks between 45-50 years for Asian women, while it ranges between 55-60 yrs for western women (4, 181). In developing countries populations are much younger on average, a higher proportion of patients are diagnosed below the age of 40, reaching as high as 20% (130-131). The underlying causes may be attributed to either epigenetic or environmental factors which make these women more prone to develop disease at a younger age (142) is a subject of ongoing research. In Indian women, breast cancer is now the most common cancer surpassing the cervical cancer, the mean age at diagnosis ranging between 45-50 years, a decade younger compared to western women.

Among various molecular subtypes, Triple negative (TNBC) breast cancer is considered the most aggressive subtype with worst prognosis. The present study showed 18% cases of TNBC which is higher in comparison to other Asian countries such as Korea (14.7%), China (12.6%) and Japan (8.4%) (182-184); whereas similar incidence had been reported by other Indian studies (11.8% to 31.9%) (185-191). Incidence of Basal, (ER- PR- HER2/neu -) (24%) has been found higher in younger patients (≤40 years) (Azim *et al.* 2012, (133) with higher proportion of advanced stage tumours (62.5%) and lymph node positivity (p=0.05) compared to late onset tumours as reported by others. Anders *et al.* (14) also demonstrated higher rates of lymphocytic and lymphovascular invasion (133, 192-194) in breast cancer. Breastfeeding had been reported to be protective against triple-negative breast cancer

(195-197) and a significant proportion of patients who have not breast feed their children (p=0.029) or who have attained their menarche before 13 years of age (p=0.023) had been found in early onset group of breast cancer. Breastfeeding had been reported protective for both early- and late-onset cancers in the high risk population and even BRCA1 carriers, in whom breastfeeding for 1 or 2 years was shown to be associated with a 32% and 49% reduction in breast cancer risk, respectively (198). These collective findings suggest that tumours arising in younger patients are probably more aggressive due to the biological differences.

Further the differences in gene expression profile of breast tumour enabled classification of the tissues into various molecular subtypes (12) viz Luminal A (24%), luminal B (20%), HER2/neu positive (20%), basal (31%) and normal-like (3%) which was similar to reported by Azim *et al.* (199). According to western literature various molecular subtypes associated with breast cancer includes Luminal A (50-60%), luminal B (15-20%), HER2/neu positive (15-20%), basal (8-37%) and normal-like (3%) (199). Early onset breast cancer consisted of 32% cases Luminal subtype, 24% HER2/neu over expressing and 21% cases basal subtype. Also IHC analysis for the phenotype viz. ER/PR/Her2/neu identified the subtypes which were showing the similar gene sets specific to each subtype as obtained by PAM50 classifier, hence showing the existence of molecular subtypes for the first time from Indian subcontinent.

Genome wide gene expression profiling in 29 breast cancer cases showed 2413 genes differentially expressed (adjusted p value ≤0.05) including *COL10A1, COL11A1, MMP1, MMP13, MMP11, GJB2, CST1, KIAA1199, CEACAM6, BUB1* top up regulated genes and *PLIN1, FABP4, LIPE, AQP7, LEP, ADH1A, ADH1B, CIDEC, THRSP, GPD1, TIMP4, KIAA1881* top down regulated genes. The up regulated genes *(COL10A1, COL11A1, MMP1, MMP13, MMP11, GJB2, CST1, KIAA1199, CEACAM6, BUB1)* are involved in various

cellular processes viz ECM, cell cycle/M phase, Focal Adhesion and Metastasis. *COL10A1* encodes a short chain collagen expressed by hypertrophic chondrocytes during endochondral ossification. The majority of human skeleton develops through the endochondral pathway, in which cartilage-forming chondrocytes proliferate and enlarge into hypertrophic chondrocytes that eventually undergo apoptosis and are replaced by bone. It acts through activating several molecules of Wnt signaling pathway and is direct transcriptional target of RUNX2 gene, a transcription factor that is expressed in cancer cells, and has been related to multiple cancers (200). GJB2 gene provides instructions for making a protein called gap junction beta 2/connexin 26, which permits the transport of nutrients, charged atoms (ions), and signaling molecules between adjoining cells. Connexins may interact with other junctional complexes to regulate overall cell adhesion and intercellular junction formation. It plays a role in the growth, maturation, and stability of the skin's outermost layer, the epidermis and is implicated in the mechanisms of Invasion of Ductal Breast Carcinomas. Cystatin SN (CST1), a cysteine protease inhibitor is upregulated and associated with tumor invasion and metastasis. It is a known tumour marker for colorectal cancer (CRC) (201) alongwith other cancer as gastric and esophageal squamous cell carcinoma (201-203). Cell cycle gene, as BUB1 functions in part by phosphorylating members of mitotic checkpoint complex and activating the spindle checkpoint complex leading to cell division. It is also reported to be associated with breast cancer prognosis in western population (204) and therapeutic target for colorectal (205), hepatocellular (206), gastric (207), non small cell lung (208) cancer in Chinese population. Over expression of BUB1 gene has been found associated with improved overall survival in breast cancer patients (209). Other cell cycle genes such as *CCNA2, CCNB2, CDC2* have also been reported over expressed in hepatocellular carcinoma (210), laryngeal squamous cell carcinoma (210) renal cell carcinoma and breast carcinoma with association with short-term disease-specific survival (211-212).

The focal adhesion genes (*COL11A1, COL10A1, and COL1A1*) have also been found upregulated in breast cancers which are important in maintaining integrity of tissues like skin, muscle, cartilage, tendons, and intervertebral disks. In invasive carcinomas, extracellular collagens are key to tumour behaviour and are subjected to continuous remodelling, inhibiting and/or promoting tumour progression. *COL11A1* is known to encode a minor collagen and its over expression had been reported in various microarray studies in breast cancer (213-214), solid tumours (215) and colon cancer (216-217). Over expression of *COL10A1* gene in breast tumours in Taiwanese population (213) had been reported to be associated with of tumour progression leading to invasion and metastasis (218). Increased expression of *COL10A1* gene had also been reported in diverse solid tumours (219-225), and had been associated with tumour vasculature (215). Li J. *et al.* had reported association of over expression of *COL10A1* gene with lower overall survival in gastric cancer (226) in Chinese population.

Genes playing important roles in extracellular matrix pathway/ metastasis such as *MMP1*, *MMP11*, and MMP13 were found to be over expressed in breast cancer (217, 227-228). *MMP1* has been reported as novel biomarkers for tumour progression in early stage cancer leading to even brain metastasis (229). MMP-11 promotes cancer development by inhibiting apoptosis as well as enhancing migration and invasion of cancer cells (230). Many Asian studies (Korea, China, Japan) have also reported their intricate role in breast and nasopharyngeal cancers (230-232). Over expression of *MMP13* has been reported in breast (232-233), oral and colorectal carcinoma (234-235) in many western and Asian studies. It also serves as a potential biomarker for early detection and prognostic assessment in esophageous squamous cell carcinoma (236).

Minichromosome maintenance complexes genes viz. *MCM2, MCM4* were also found upregulated in breast cancer, which are part of DNA replication process (237). Over

expression of *MCM2* gene in breast cancer has been found associated with poor outcome and poor recurrence free survival (238-241). Over expression of *MCM4* gene in breast cancer had been reported to be associated with shorter survival and used as prognostic biomarker and prediction of response to treatment in Chinese population (237, 242-243). *POLE2* gene, DNA polymerase epsilon, is involved in DNA repair and replication, and the protein encoded by this gene represents the small subunit (B). Our study for the first time has reported significant up regulation of *POLE2* in breast cancer, the differential expression of *POLE2* gene had been reported in bladder cancer (244) and cervical cancer earlier in Chinese population (245). Mutations in *POLE2* genes had been reported in colorectal cancer (246).

The top down regulated genes include *PLIN1, FABP4, LIPE, AQP7, LEP, ADH1A, ADH1B, CIDEC, THRSP, GPD1* and *TIMP4* genes which are involved mainly in Lipid Metabolism, Lipolysis, Oxidoreductase activity and PPAR pathway. PPAR pathway involved genes as *PLIN, FABP4, and AQP7*. Perilipin gene *(PLIN)* gene mainly coats lipid storage droplets in adipocytes. Not much is known about this genes and its functional role in cancer. Zhou C. *et al.* have shown down regulation of *PLIN* genes in breast cancer and proposed it to be potentially new target for gene therapy (247). *AQP7* is reported to be downregulated in hepatocellular carcinoma (248) and in breast cancer (249). These lipid metabolizing enzymes are shown to be differentially expressed in different subgroups of breast cancer and are associated with poor survival. (250).

Besides this several other important genes playing functional role in various pathways were also found under expressed such as *CDH5* which functions as a classical cadherin, imparting the ability to cells to adhere in a homophilic manner with major role in endothelial adherence junction assembly and maintenance. Expression of *CDH5* gene can distinguish metastatic breast cancer from non metastasizing type (251). Besides this *CDH5* gene mutation are

associated with TNBC which causes alterations in the RAS/RAF/MEK/ERK and PI3K/AKT/mTOR pathways leading to resistance to treatments (252). High *CDH5* expression has been reported in gastric cancer (253) also gene *CADM3* gene acts as a tumour suppressor gene and it encodes a calcium independent cell adhesion protein enabling adhesion activity. This is found downregulated in breast cancer patients and upregulated in oesophageal cancer in northeast Indian population (254). *CLDN11* is an intergral membrane protein and component of tight junctions. Loss of *CLDN11* gene has been reported in other cancers also such as cutaneous squamous cell carcinoma (255) oral leukoplakia (256) and prostate cancer (257) resulting in enhancement of cellular motility and invasiveness, however up regulation of *CLDN11* had been reported in gastric cancer (258). There is a single report which suggests its over expression in breast cancer making it a target for prognostic and therapeutic intervention in breast cancer (259). Deregulation of all these genes suggest a strong interplay with each other leading to complex process of breast cancer progression in Indian women.

Comparison of significant dysregulated genes in present study with that of western population identified few unique upregulated genes *KIAA1199, CKAP2L, TUBB3, CENPA* and unique downregulated genes *PLIN, KIAA1881, ADH1A, SAA1,* in Indian population. These genes are part of Cell cycle, Cytokine- cytokine receptor interaction, Lipolysis mainly involved in metastasis leading to breast cancer. Few genes which were found to be upregulated in western dataset as XIST, KLF6, CCNL1, CD59 were downregulated in our data. XIST is known to play a tumour supressive or oncogenic role. In breast cancer decreased XIST levels have been reported which act by increasing AKT phosphorylation and thus leading to tumourogenesis in Indian population as compared to western where it was found to be upregulated (260). KLF6 is a tumour suppressor gene and its downregulation is reported to play a role in development of CRC (261) which shows concordance with our data. Cyclin family gene CCNL1 is upregulated in various cancers and was associated with poor prognosis (262-263). CD59,

belonging to membrane complement regulatory proteins, inhibits the cytolytic activity of complement and is overexpressed in many types of solid cancers (264). The differences in the gene set, could be either attributed to either differences in ethinicity, platform variability or due difference in sample size between both studies which might have lead to bias in identifying the differential genes.

Few common genes amongst the two populations suggestive of some common mechanism promoting breast carcinogenesis included upregulated genes as *COL10A1*, *MMP11, CST1, GJB2, MMP1, MMP13, CEACAM6* and downregulated genes as *ADH1B1, CIDEC, THRSP, GPD1, TIMP4, FABP4, SCARA5.*

Validation of genes *MMP1*, *MMP13*, *MMP11*which were found up regulated and *ADAMTS1*, *ADAMTS5* genes which were found down regulated in 67 breast cancer cases by real time PCR showed concordance with our gene expression profiles. Significant up regulation of genes *MMP1*, *MMP13*, *MMP11* and down regulation of *ADAMTS1*, *ADAMTS5* was seen in breast tumours as compared with control. MMP's up regulation has already been reported in lung, prostate, breast cancer leading to increase in invasion and metastasis. There was also a significant higher fold up regulation of MMPs in ET; however it did not reach to level of significance in validation probably due to smaller number of samples. This is suggestive of aggressive nature of early onset tumours. *ADAMTS5* was found significantly differentially expressed between early and late groups, ($p=0.012$) suggesting its association with late onset tumours. Porter *et al.* reported *ADAMTS5* to be down-regulated in breast cancer. The significant association of these *MMP1*, *MMP13*, *MMP14* and *MMP11* was also seen with ER, PR negativity and HER2/neu positivity along with metastasis suggesting their significant role metastasis progression by epithelial mesenchymal transitions in breast cancer.

To understand the molecular biology of early onset breast cancer in Indian women, the significant genes and pathways involved in early onset tumours were also analyzed and

compared with those in late onset tumours. The pathways which were found involved include basic cellular pathways such as cell adhesion (*VCAN, CDH15, PVRL2*), cell cycle (*CDKN2A and SMAD3)*, ECM receptor (*IBSP, ITGA11*), pathways in cancer (*BRAF, FZD2).* All these genes were found over expressed except *SMAD3*gene which was found down regulated in early onset tumours. Versican (VCAN) is a member of the aggrecan/versican proteoglycan family. The protein encoded is a large chondroitin sulfate proteoglycan and is a major component of the extracellular matrix and is involved in cell adhesion, proliferation, proliferation, migration and angiogenesis playing a central role in tissue morphogenesis and maintenance. It accumulates in tumor stroma and plays a key role in both malignant transformation and tumor progression. Increased versican expression has been observed in a wide range of malignant tumours, and has been associated with both cancer relapse and poor patient outcomes in breast, prostate, and many other cancer types. Through negatively-charged chondroitin and dermatan sulfate side chains or interactions of the G1 and G3 domains, versican is able to regulate many cellular processes including cell adhesion, proliferation, apoptosis, migration, angiogenesis, invasion and metastasis. *VCAN i*s a known potential prognostic biomarker for colon cancer (265) and is associated with cancer relapse and poor patient outcomes in breast, prostate, and many other cancers.(266). VCAN was found to be overexpressed in early onset as compared to late onset showing association with aggressive behaviour, acting probably through EGFR/AKT/GSK-3β (267). *CDH15* gene is a member of the cadherin superfamily of genes, encoding calcium-dependent intercellular adhesion glycoproteins. Transcripts from this particular cadherin are expressed in myoblasts and upregulated in myotubule-forming cells. The protein is thought to be essential for the control of morphogenetic processes, specifically myogenesis, and may provide a trigger for terminal muscle cell differentiation. *CDH15* gene has been reported over expressed in hepatocelluar carcinoma (268), however its role in breast cancer has not yet elucidated. We

report for the first time over expression of *CDH15* gene in early onset breast cancer in Indian women which is known to play role in morphogenetic processes especially myogenesis (269). In aggressive cancers CHD15 are upregulated, as found in this study. *PVRL2* also called as Nectin2 are $Ca2^{+}$-independent immunoglobulin (Ig) superfamily proteins that participate in the organization of epithelial and endothelial junctions and regulate several cellular activities including the entry of some viruses. It has been found to be over expressed in breast tumours where it plays a major role in metastasis (270) and other cancers such as ovarian cancer and hepatocellular carcinoma (271) where its role as prognostic biomarker has been established through microarray and immunohistochemistry (272). Over expression of *CDKN2A* gene is mainly responsible for making tumour suppressor proteins p16 and p14ARF which in turn regulate cell cycle. It has been reported in ovarian cancer (273). *IBSP* gene is major structural protein of bone matrix and is found over expressed in osteosarcoma (274) in western population while Integrin *ITGA11* enables attachment and is mainly required to initiate collective invasion and metastasis during tumour progression. It promotes tumourigenesis and metastasis in non small cell lung cancer (275). It's over expression is associated with poor outcome in breast cancer (276). Significant up regulation of *ITGA11* in early onset tumours suggests its specific role in invasion, metastasis and aggressive behaviour in young Indian breast cancer patients. *BRAF* is a known proto-oncogene which is involved in sending signals for directing cell growth. Mutations in BRAF gene have been reported in many cancers (277). Epithelial mesenchymal transitions are promoted by *Frizzled2 (FZD2*) protein leading to metastasis of endometrial cancer (278), prostate cancer (276, 279) and breast cancer (280-281). Association of over expression of *BRAF* and *FZD2* genes have been found with metastasis. The SMAD3 gene normally provides instructions for making protein which is involved in transmitting chemical signals from the cell surface to the nucleus through transforming growth factor-beta (TGF-β) pathway. The signaling process begins when a TGF-

β protein attaches (binds) to a receptor on the cell surface, which activates a group of related SMAD proteins (including the SMAD3 protein). These SMAD proteins combine to form a protein complex, which then moves to the cell nucleus. In the nucleus, the SMAD protein complex binds to specific areas of DNA to control the activity of particular genes. Through the TGF-β signaling pathway, the SMAD3 protein also influences many aspects of cellular processes, including cell growth and division (proliferation), cell movement (migration), and controlled cell death (apoptosis). It functions as both positive and negative regulator of carcinogenesis depending on cell type and clinical stage of the tumour (282). It functions as tumour suppressor and is found downregulated in gastric cancer (283) and breast cancer in western population. In early onset breast cancer the downregulation of SMAD3 stops the antiproliferative activity of TGFbeta pathways and thus leading to carcinogenesis (284). It functions as tumour suppressor and is found downregulated in gastric cancer (283) and breast cancer in western population. Late onset tumours were found to deregulate almost similar pathways as that of early onset tumours, like cell adhesion, cell cycle, ECM receptor interactions, however the genes in the pathways were different in both the groups. In late onset cancer *MCM2, MAD2L1, BUB1B, WNT3* genes were found up regulated while *THBS4, EPAS1, ESAM, PVRL3, CXCL12, MYL9* genes were found down regulated suggesting distinct molecular events in both early and late onset breast cancer. They played a major role in angiogenesis, cell survival, invasion.

Study by Anders *et al.* (14) showed 367 biologically relevant gene significantly distinguished breast cancer arising in young women on gene set enrichment analysis (GSEA) from tumours arising in older women (>65yrs) at false discovery rate of < 0.25. The gene set unique to breast cancer in young women included those related to immune functions, mTOR/rapamycin pathway, hypoxia, BRCA1, stem cell, apoptosis, histone deacetylase and multiple oncogenic signalling pathways, holding therapeutic implications in breast cancer (14). In this analysis, a

higher probability of phosphatidylinositide 3-kinase (*PI3K*; P = 0.006) and *Myc* (P = 0.03) pathway deregulation was observed in tumours arising in younger patients. Since this analysis was not adjusted for potential differences in breast cancer molecular sub- types as well as other known prognostic factors, the same authors re-analyzed their data using two publicly available datasets with appropriate adjustment for molecular subtypes among other features and reported that younger patients have more basal-like tumours and after adjustment for subtype differences, no distinct molecular aberrations were found in early onset breast cancers. Azim *et al.* 2012 has reported 41 genes associated with young breast cancer signatures, after adjustments only 12 genes involved in biological processes like immature mammary cell population including *RANKL, c-kit, BRCA1*-mutated phenotype, mammary stem cells, and luminal progenitors cells were found significantly deregulated. Two genes- *BAX* (apoptosis related) and *ALDH2* (stem cell related) out of 41 genes which did not reached statistical significance in Azim *et al.* study were found significantly differentially expressed in present study.

Our study identified *RPL29, FZD2* as the most interactive topmost upregulated node in ET while *SMAD3* was topmost interactive downregulated node (Fig. 19.b). Early onset tumours mostly have aggressive basal subtype with higher and advanced stage of cancer which showed *LYN* as upregulated and *MYC, JUN, RPS23* as downregulated nodes. *RPL29* gene interacts with heparin and proteoglycans to modulate blood coagulation, cell adhesion and cancer cell invasion during metastasis (285). Liu J.J. *et al*. has reported over expression of *RPL29* in colon cancer (286). All these genes play important role in cell cycle progression and metastasis in turn leading to breast cancer progression.

To conclude, the molecular mechanism we hypothesize for early onset breast cancer in Indian women as shown in Fig. 4.28. Genes *IBSP, VCAN, PVRL2, FZD2, CDKN2A, B-RAF,*

ITGA11, CDH15, SMAD3 either activates several cancer pathways or enables epithelial mesenchymal transitions enabling adhesion, division and proliferation, angiogenesis, migration leading to breast cancer in young Indian females. Their over expression makes tumours highly aggressive in nature leading to bone metastasis which also correlates with poor outcome in breast cancer patients

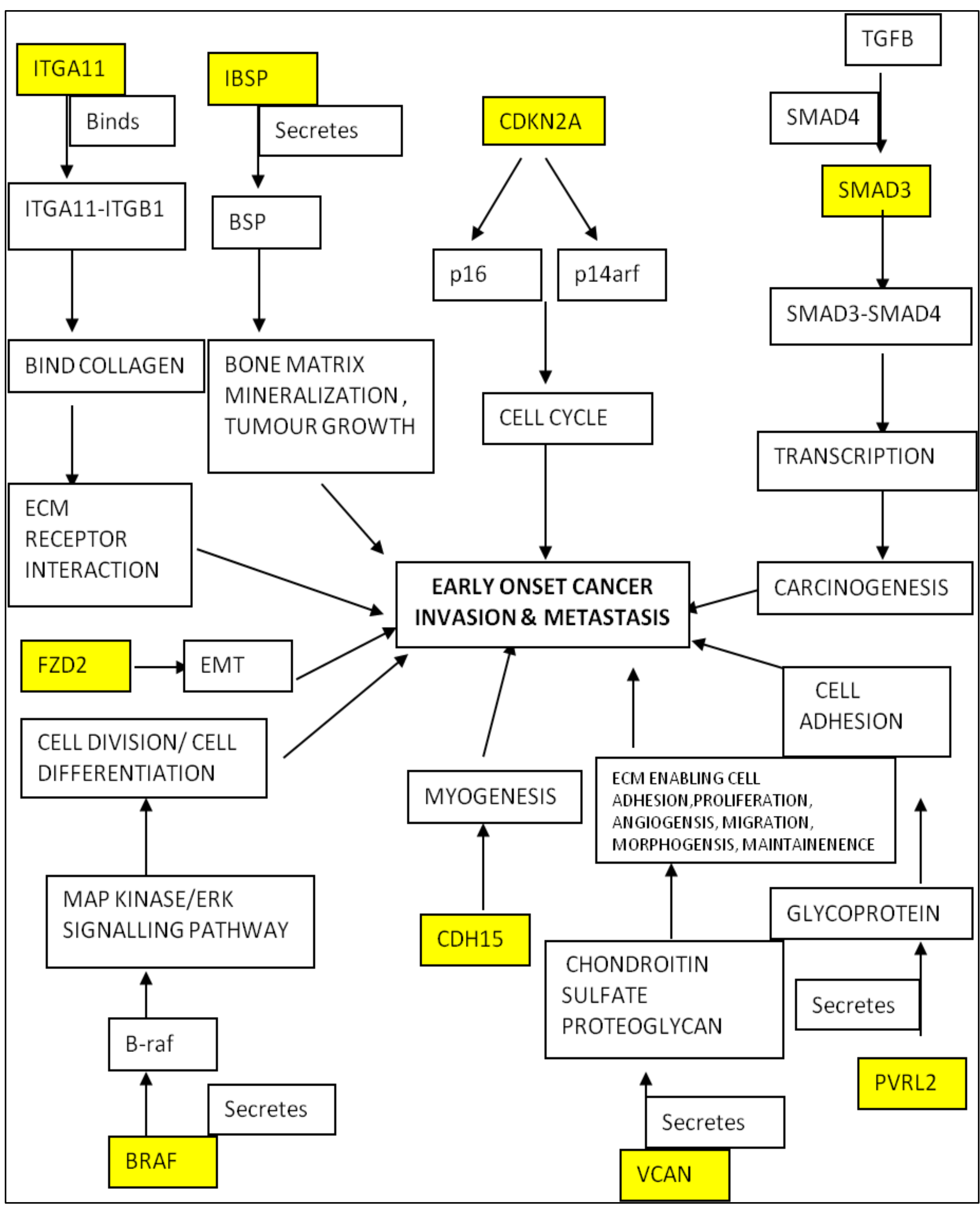


Fig.4.28 Breast carcinogenesis mechanism leading to increasing incidence of early onset breast cancer in Indian females.

Conclusion

Breast cancer is one of the most common cancers and the leading health crisis for women today. This continuous rise in breast cancer incidence has prompted the need to develop strategies for its prevention. Breast cancer appears to have a complex etiology, possibly with interplay of many causal factors including genetic, epigenetic, hormonal and environmental factors operating over a long period. Although several risk factors have been well defined, the interactions of the various etiological factors are yet to be completely understood. Moreover, analysis of parameters like early-onset could be helpful in screening high risk women for developing breast cancer, followed by planning of future preventive and treatment modalities. Younger patients more frequently exhibit aggressive features with young age itself being an independent predictor of adverse prognosis. This study had been initiated to understand molecular pathology and behaviour of early onset breast cancer and to identify potential biomarkers. MMPs could possibly serve as novel targets for early onset breast cancer but it has to be further validated in larger cohort of patients at protein level.

Chapter 5

To study differential methylation profile in Indian breast cancer women and compare the profiles in early and late onset tumours

To study methylation profiles of breast cancer in Indian women and compare the differential profiles in early and late onset tumours.

5.1 Introduction

Breast cancer is one of the most common cancer and second common cause of cancer-death among females in western countries (287). In India, it is most common cancer in women in metropolitan cities while it holds second position in rural population (6). Breast cancer is characterized by both genetic and epigenetic alterations (160, 288-291). Epigenetic hallmarks of cancer include global DNA hypomethylation and locus-specific hypermethylation of CpG islands which occurs at initiation or development of tumours (292-293). CpG methylation plays an important role in maintaining gene silencing, and this process is essential for a wide variety of cellular functions, such as tissue-specific gene expression, X chromosome inactivation and genomic imprinting (294-296). Although most CpG dinucleotides are methylated on cytosine residues, but CpGs present in promoter are protected from methylation. Many tumours exhibit excessive methylation of these CpG islands that are found near or inside the transcriptional promoters of mammalian genes contributing to breast cancer (297). Hypermethylation of CpG islands in specific gene promoters transcriptionally silence tumour suppressor gene expression and hence contributes to carcinogenesis by leading to the initiation and progression of cancer (298). A large number of genes are found to be aberrantly methylated in breast tumours as *RASSF1A*, *CDH1*, *RARβ*, *CYCLIND2*, *TWIST*, *BRCA1, HIN1* and *APC* (299-300). DNA methylation alterations are widely accepted as a potential source of early biomarkers for better diagnosis/prognosis (301-302). Aging increases the risk of cancer along with other diseases as cardiovascular diseases/neurodegenerative diseases gradually as body's ability to maintain a steady state is reduced. There are few studies on methylation status of breast cancer in Indian women (303-306) however none of them addressed

methylation specific changes in age specific cohort in breast cancer. In this objective we have studied DNA methylation in breast cancer in Indian women with special emphasis on early onset breast tumours.

5.2 Material and methods

5.2.1 Clinical Specimen

A total of 97 tumour tissue samples were collected from Safdarjung Hospital and Indraprastha Apollo Hospital, New Delhi, along with 38 distant normal control tissues from histopathologically confirmed Infiltrating Duct carcinoma cases prior to any chemotherapy or radiotherapy treatment (Table4.1). The samples were staged according to American joint committee on cancer-AJCC. Immunohistochemistry was performed to identify the receptor status for ER/PR/ERBB2 receptors. 41 cases were found to be below ≤40 yrs (early onset) and 56 cases were with age ≥55 years (late onset). A total of 48 samples were used for methylation microarray (Table 5.1) of which 19 tumours were of early onset and 17 tumours belonged to late age along with 12 normal control, 6 in each group. All tissue samples were snap frozen immediately after the modified radical mastectomy or after incision/true cut biopsy and immediately stored at -80 for RNA/ DNA isolations. Informed consent was obtained from all the patients prior to collection of samples and was approved by the ethical committee of Safdarjung hospital and Indraprastha Apollo hospital.

Table. 5.1 Clinicopathological details of samples used for methylation microarray

RT-ID	AGE	DIAGNOS	TNM	TSIZE	LYMPH NO	METASTASI	STAGE	ER	PR	HER2NU	MENARCH	BREAST F	FAMILY HI	MARRIAG	CHEMOT
ET-16	40	IDC	T3N0M0	T3	N0	M0	IIB	P	P	N	15	yes	NO	no	no
ET-18	35	IDC	T2N1M0	T2	N1	M0	IIB	P	N	P	14	yes	NO	no	no
ET-19	32	IDC	T2N1M0	T2	N1	M0	IIB	P	P	P	13	yes	NO	no	no
ET-20	33	IDC	T1N1M0	T1	N1	M0	IIA	N	N	P	13	yes	NO	no	no
ET-21	40	IDC	T2N1M0	T2	N1	M0	IIB	P	N	P	14	yes	NO	no	no
ET-22	39	IDC	T3N1M0	T3	N1	M0	IIIA	P	P	P	13	yes	NO	no	no
ET-23	28	IDC	T4bN1M0	T4	N1	M0	IIIB	N	N	P	15	no	NO	no	no
ET-24	40	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	15	yes	NO	no	no
ET-25	30	IDC	T2N1M0	T2	N1	M0	IIB	N	N	P	14	yes	NO	no	no
ET-29	40	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	N	15	yes	NO	no	no
ET-33	35	IDC	T4bNxM1	T4	N1	M1	IV	P	P	N	14	YES	NO	no	no
ET-36	25	IDC	T4N2M1	T4	N2	M1	IV	N	P	N	11	NO	NO	no	no
ET-37	40	IDC	T2N1M0	T2	N1	M0	IIB	N	P	N	12	YES	NO	no	no
ET-38	35	IDC	T4N2M0	T4	N2	M0	IIIB	P	N	N	13	YES	NO	no	no
ET-39	35	IDC	T3N1M0	T3	N1	M0	IIIA	P	P	P	12	NO	NO	no	no
ET-43	30	IDC	T3N1M0	T3	N1	M0	IIIA	N	P	P	13	YES	NO	no	no
ET-51	32	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	N	13	YES	NO	no	no
ET-52	35	IDC	T4bN1M0	T4	N1	M0	IIIB	N	N	N	15	YES	NO	no	no
ET-60	40	IDC	T2N1M0	T2	N1	M0	IIB	N	N	N	13	YES	NO	no	no
LT-61	70	IDC	X	X	X	X	X	P	P	N	13	YES	NO	no	no
LT-16	70	IDC	T2N1M0	T2	N1	M0	IIB	N	N	N	14	yes	NO	no	no
LT-17	80	IDC	T4bN1M0	T4	N1	M0	IIIB	N	N	P	14	yes	NO	no	no
LT-18	57	IDC	T2N1M0	T2	N1	M0	IIB	N	N	N	13	yes	NO	no	no
LT-19	70	IDC	T3N1M0	T3	N1	M0	IIIA	N	N	P	14	yes	NO	no	no
LT-21	62	IDC	T3N2M0	T3	N2	M0	IIIA	N	N	N	14	yes	NO	no	no
LT-23	60	IDC	T4bN1M0	T4	N1	M0	IIIB	N	N	P	14	yes	NO	no	no
LT-24	65	IDC	T2N0M0	T2	N0	M0	IIA	P	P	N	16	yes	NO	no	no
LT-29	70	IDC	T2N0M0	T2	N0	M0	IIA	P	P	N	15	yes	NO	no	no
LT-33	59	IDC	T2N1M0	T2	N1	M0	IIB	P	N	P	15	yes	NO	no	no
LT-35	55	IDC	T4BN1M0	T4	N1	M0	IIIB	N	N	P	13	YES	NO	no	no
LT-41	70	IDC	T2N1M0	T2	N1	M0	IIB	N	N	P	16	YES	NO	no	no
LT-44	60	IDC	T4BN1M0	T4	N1	M0	IIIB	NA	NA	NA	16	YES	NO	no	no
LT-45	80	IDC	T4BN2M1	T4	N2	M1	IV	NA	NA	NA	15	YES	NO	no	no
LT-46	55	IDC	T2N1M0	T2	N1	M0	IIB	NA	NA	NA	13	YES	NO	no	no
LT-49	56	IDC	T3N1M0	T3	N1	M0	IIIA	NA	NA	NA	12	YES	NO	no	no
LT-51	58	IDC	T2N0M0	T2	N0	M0	IIA	P	P	P	14	YES	NO	no	no

5.2.2 DNA isolation, purification from cancer tissues

DNA was isolated using Geneaid DNA tissue isolation kit following manufacturer's protocol. Upto 30mg of breast tissue was transferred to a 1.5 ml micro centrifuge tube. It was grounded to form a pulp with the help of micro pestle. After tissue homogenization 200 µl of GT Buffer was added to the tube and tissue was further homogenized by grinding. Next 20 µl of Proteinase K was added to the sample mixture and was incubated at 60ºC for 30 minutes. 200 µl of GBT Buffer was added and then shaken vigorously for 5 seconds. It was incubated at 60ºC for at least 20 minutes to ensure that the lysate becomes clear. Next 200 µl of absolute ethanol was added to the lysate which was immediately shaken vigorously for 10 seconds and the mixture (including any precipitate) was transferred to the GS Column and then centrifuged

at 14,000 x g for 2 minutes. For washing the column 400 µl of W1 Buffer was added to the GS Column and then centrifuged at 14,000 x g for 30 seconds. Another wash with 600 µl of Wash Buffer (make sure ethanol was added) was given to the GS Column. It was centrifuged at 14,000 x g for 30 seconds. Flow-through was discarded and then column was placed back in the 2 ml Collection Tube. It was centrifuged again for 3 minutes at 14,000 x g to dry the column matrix. DNA was eluted in 60 µl elution buffer by centrifuging it at 14,000 x g for 3 minutes.

5.2.2.1 Quantitation of DNA

Quantitation of double stranded DNA was performed by using Quant-iT™ PicoGreen® dsDNA Assay kit by Thermo fischer. Reagents from kit were thawed at RT in a light impermeable container. A standard DNA plate was made by diluting stock lambda DNA and plating its dilutions. Picogreen reagent was diluted 1:200 into 1X TE, vortexed and mixed. 195ul of this PicoGreen/1X TE dilution was dispensed into each well and dilutions of stock DNA were added. The sample fluorescence was measured to generate a standard curve and concentration of DNA samples was determined. All of the samples yielded high quantity of DNA per µl and were further proceeded for microarray.

5.2.3 Genomic Methylation study

Human whole genome methylation analysis was performed using Illumina's Human methylation 27 bead chips for 48 tissue samples in various batches. It interrogates over 27,000 highly informative CpG sites per sample at single-nucleotide resolution which includes approximately 13,000 well-annotated genes described in the NCBI Database1 (Genome Build 36) covering over 1,000 cancer related genes, 200 miRNA promoters, and 150 regions which are known to exhibit differential methylation status in various systems

5.2.3.1 Bisulphite Conversion of DNA

Around 500ng of DNA was taken for Bisulphite-conversion of the genomic DNA samples using the Zymo EZ DNA Methylation Kit according to manufacturer’s instructions. Briefly genomic DNA was denatured and CT conversion reagent was added followed by incubation at 95°C for 30 sec and 50°C for 1 hour for 16 cycles. The converted product was desulphonated by addition of desulphonation buffer. Bisulphite converted DNA was eluted using elution buffer.

5.2.3.2 Fragmentation/ Hybridization and Scanning of microarray

The bisulphite converted DNA samples were further denatured and neutralized by adding 4ul of 0.1N NaOH in the presence of buffer MA1 provided in the Illumina’s whole genome methylation kit. It was vortexed, centrifuged and incubated at RT for 10 minutes. Next 68ul of MA2 was dispensed along with 75ul of MSM solution into each well. Plate was sealed and incubated for 20-24 hrs at 37°C. Next the DNA was fragmented using end-point fragmentation. Finally DNA was precipitated using 2 propanol and PM1 solution and further resuspended into RA1 buffer. This DNA was now dispensed onto bead chips and they were kept into hybridization oven for homogenous hybridization overnight. Bead chips were further washed and labelled nucleotides were hybridized in order to extend the primers. Bead Chips were scanned using Illumina BeadArray Reader (Scanner).

5.2.4 Background correction

Hybridization of non specific labeled targets leads to fluorescence which results in generation of background. Background correction was performed by subtracting the average signal of the negative controls probes from the actual probe intensity signal. It also generated detection P-value which is a statistical calculation that provides the probability that the signal from a

given probe is greater than the average signal from the negative controls**.**

5.2.5 Normalization

Raw data normalization was initiated by selecting probes sets having uniform intensity among the tumour tissues samples. Illumina custom model which is analogues to t- test error model was used to find significant methylated CpG sites, based on the experimental conditions and the number of replicates available. Further Beta value was generated for each significantly identified probe, which in turn specifies the data into range between 0-1 and hence removes unnecessary high intensity variation which leads to noise. It is calculated by taking ratio of intensity of methylated bead and intensity of methylated and unmethylated bead for each probe used in experiment.

Beta Normalization

For the Infinium Methylation Assay, ß is calculated as:

$$\beta = \frac{Max(SignalB,0)}{Max(SignalA,0) + Max(SignalB,0) + 100}$$

Here signal A and signal B are produced by two different bead types viz. unmethylated and methylated beads.

Output also generated a **Diff score** which represent the magnitude of the difference between the Beta value of reference group and test group.

5.2.5 Differential Methylation

Post beta normalization, each gene was assigned a beta value (β). A gene was defined as having higher methylation level (Hypermethylated) if the average methylation level in tissue was at least 0.5 ($\beta \geq 0.5$). Similarly, a lower methylation level (hypomethylated) indicated that the average methylation level of a gene should be less than 0.5 ($\beta < 0.5$). The Diff score was used to identify the differentially methylated genes and gene sets amongst various groups

(TT/ET/LT) as compared to normal controls. Positive Diff score indicated hypermethylation and suggested that the methylation in tumour tissue was higher than methylation in control tissues while negative score suggested, hypomethylation of the gene having higher methylation in normal control tissue as compared to tumours. Higher Diff score suggests higher significance of the DMGs.

5.2.6 Hierarchical Clustering and Gene ontology analysis

Hierarchical clustering was done of differentially methylated genes by using Cluster 3 software by median centred clustering for centroid linkage using Pearson uncentered correlation. It segregated gene in clusters of hypermethylated or hypomethylated genes which were viewed by JavaTree View Software.

5.2.7 Validation

Validation of identified methylated profiles was performed in 740 human breast cancer patients along with 92 controls from TCGA database via MethHC (A database of DNA methylation and gene expression in human cancer) software which uses TCGA data portal. Box plots were generated for average tumour and normal beta values with corresponding p values for each genes. Also the gene expression and methylation integrated profiles were validated in 839 tumour samples and corresponding scatter plots were produced along with correlation coefficient and corresponding p value using MethHC software.

5.3 Results

5.3.1 DNA Methylation analysis

Gel Image

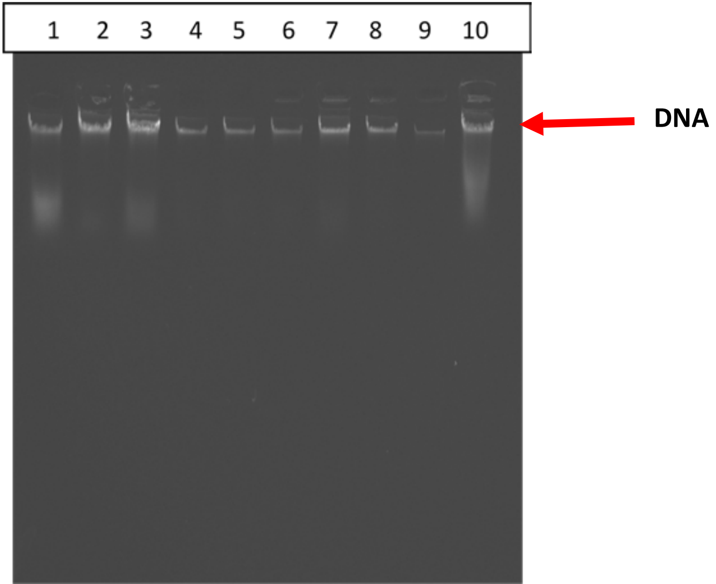


Fig. 5.1 Agarose gel showing the good quality of DNA which was extracted from the tumour tissues

Quality of DNA extracted from all cases were checked by running samples over 0.8% agarose gel in 0.5X TBE prepared in distilled water. All the samples showed a single discrete band for DNA suggesting its high quality (Fig.5.1).

5.3.2 Methylation Quality Control

5.3.2.1 Chip Hybridization

Infinium Human Methylation 27k bead chips were scanned to visualize the efficiency of hybridization. The hybridization of DNA was found to be uniform throughout the chip (Fig 5.2).

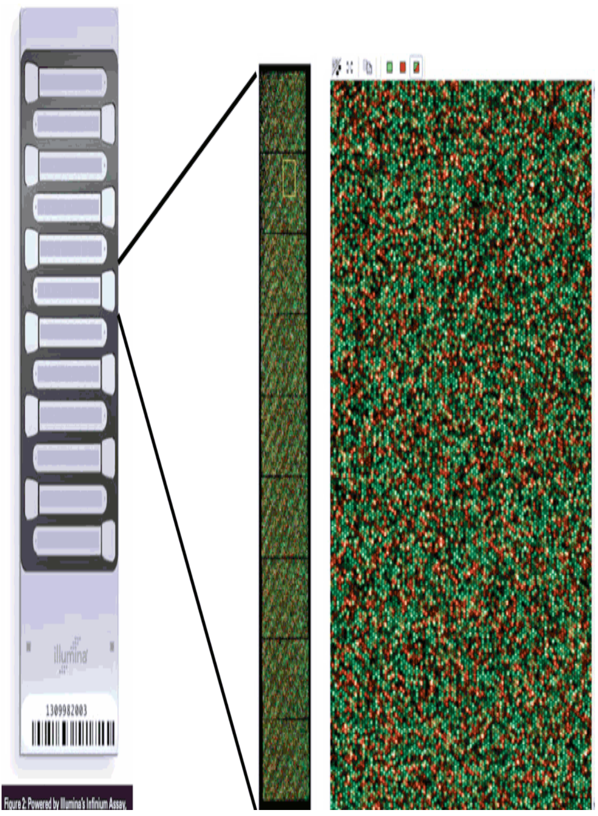


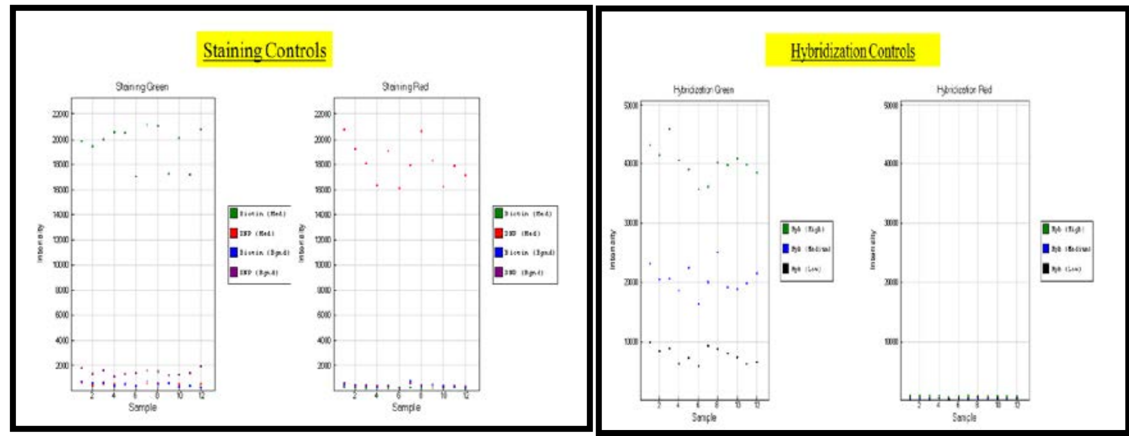
Fig.5.2 Scan of the hybridized chip showing overall view of the probes present on chip

5.3.2.2 Quality Control

To establish the quality of the microarray experiments, several quality control parameters were checked on each single array (Fig5.3).

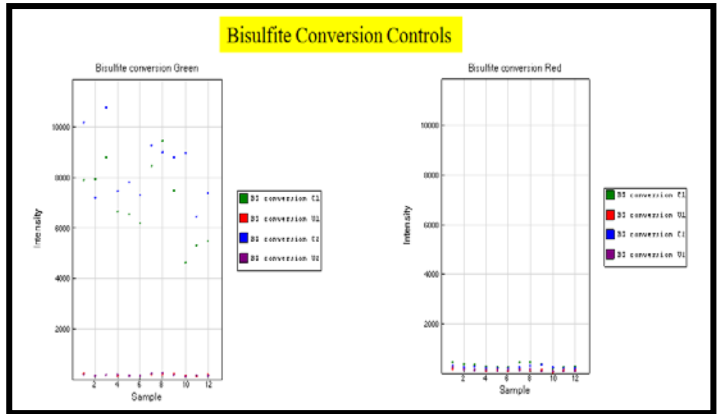
a. Staining and Hybridization Controls

Staining controls examine the efficiency of staining for all channels. The staining efficiency was found to be upto the mark for all samples used for profiling. Hybridization controls test the overall performance of the Infinium Assay using synthetic targets. These synthetic targets complement the sequence on the array perfectly allowing the probe to extend on the synthetic targets as a template. They are present in hybridization buffer (RA1) at three levels monitoring the response from high to low concentration.



b. Bisulphite conversion Controls

Bisulphite conversion controls monitor the efficiency of bisulphite conversion of genomic DNA. If the bisulfite conversion is successful the converted probes "C" matches the converted sequence and gets further extended. If the samples have unconverted DNA "U", unconverted probes gets extended. In our experiment all the converted controls worked well.



c. Negative Controls

Negative controls target bisulfite converted sequences that do not contain CpG dinucleotides. These assay probes are randomly permutated and does not hybridize to DNA template. The mean signal of these probes defines the system background. Mostly all of the probes showed low signal for background.

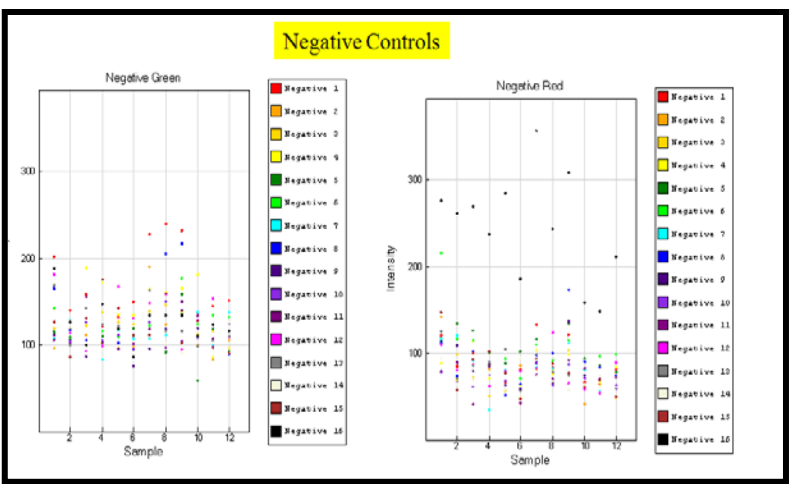


Fig.5.3 Quality control assessment of the methylation microarray experiment

5.3.3 Differential methylation Analysis

5.3.3.1 All tumours *vs*. normal controls

Methylation analysis in 36 tumours and 12 normals showed 11,016 CpG sites significantly differential methylated in all tumours. Of the total sites 5660 CpG sites were hypermethylated and 5356 sites were hypomethylated (Fig.5.4). Further filtration resulted in corresponding 910 genes which were aberrantly methylated.

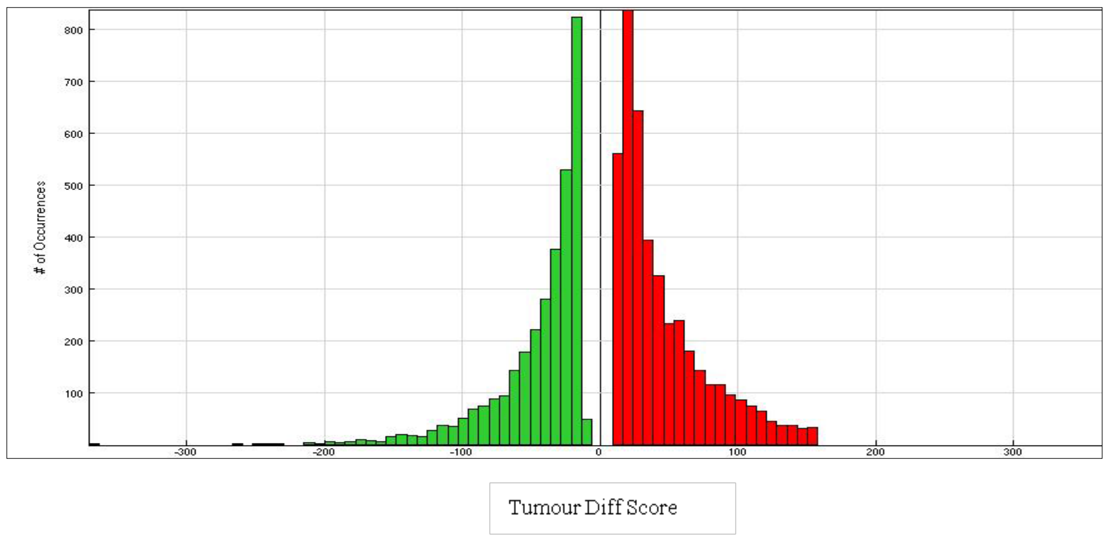


Fig. 5.4 Histogram depicting hypermethylated (red) and hypomethylated (green) CpG sites in tumours *vs.* normals

a. Hierarchical clustering

Unsupervised hierarchical clustering was performed using Beta values (range 0-1) for all individual genes and heat map was plotted. Two distinct differential gene clusters were visible separating hyper from hypo methylated genes amongst tumours and normals (Fig 5.5).

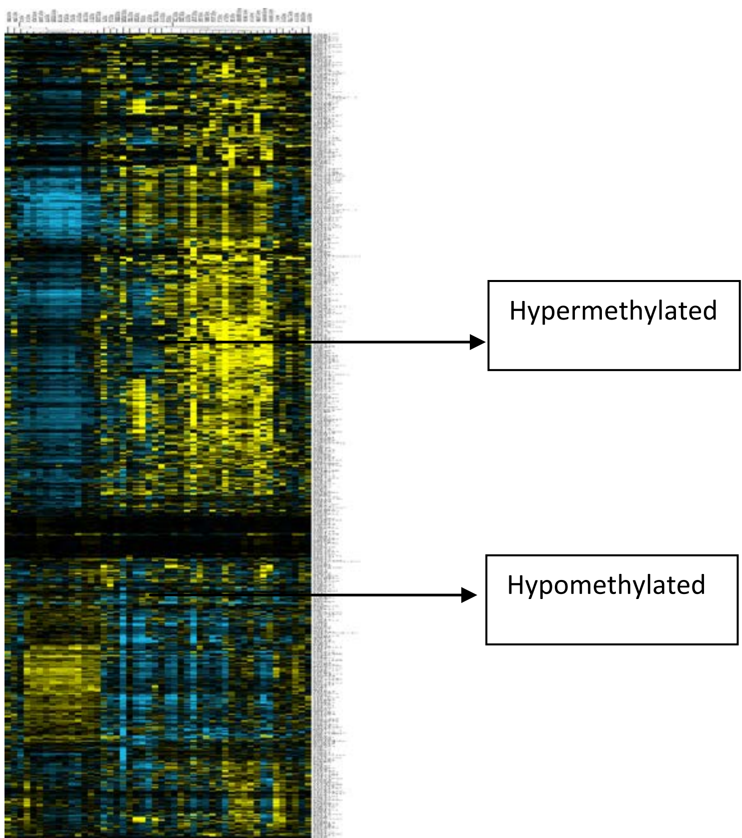


Fig.5.5 Heat map of differentially methylated genes between tumour *vs.* normals showing distinct clusters. Hypermethylated regions are represented as blue coloured genes while hypomethylated region are represented in yellow colour.

These differentially methylated genes were used further to explore their role in breast carcinogenesis.

b. GO Analysis

Pathway analysis was done to identify pathways associated with deregulated genes contributing to breast carcinogenesis. Major aberrant methylated genes were found involved in pathways such as cAMP signalling, cell adhesion, cytokine receptor interaction, neuroactive ligand receptor interaction etc.

cAMP signalling pathway showed hypermethylation of *HHIP* (hedgehog interacting protein), *RYR2* (ryanodine receptor 2) and *CFTR* (cystic fibrosis transmembrane conductance regulator) genes. The embryonic development is regulated by HH pathway, which is negatively regulated by *HHIP* (307). *RYR2* encodes a protein which is a major component of calcium channel which supplies calcium to cardiac muscle. They are mainly involved in signal transduction pathways in heart and their dysregulation leads to cancer and many other diseases (308). *CFTR* gene enables transport of various molecules across extracellular and intracellular membrane by functioning as chloride channel. Cell adhesion pathway showed hypermethylation of *CLDN11, CDH5* genes playing a mechanistic role in **cell adhesion** and hypomethylation of *CD86* and *CD28* genes. *CDH5* also known as vascular endothelial cadherin, plays an important role in homotypic cell-cell adhesion among epithelial cells (253). *CLDN11* are members of tight junction strands serving as a physical barrier which prevent solutes and water from passing freely through the paracellular space between epithelial or endothelial cell sheets, and maintains cell polarity and signal transductions. Hypomethylated genes *CD86* and *CD28* genes have role in immune regulation. The costimulatory signals reaching through *CD86* to its counter receptor *CD28* on T cells induce T-cell-mediated immunity against tumours (309) Further genes as *TNFRSF10D, CD40*, *TNFSF4* were found hypermethylated and act as mediators in regulation of **cytokine cytokine receptor interaction** in tumours. The tumour necrosis factor family genes play major role in adhesion of activated T cells to endothelial cells mediating immune and inflammatory response.

Neuroactive ligand receptor interaction pathway showed hypermethylation of *GALR1* (Galanin receptor1), *PTGDR (Prostaglandin D2 receptor), LEP genes (Leptin). GALR1* mediates many biological effects via interaction with G protein coupled receptor. Similarly *PTGDR* also mediates allergic inflammation via. G protein coupled receptor. *LEP* gene plays major role in regulation of body weight. Beta values and diff scores for all genes are mentioned in table 5.2.

Table. 5.2 Beta values for the differentially methylated genes in breast tumours.

GENE	**Normal. AVG_Beta**	**Tumour. AVG_Beta**	**Tumour.Diff Score**
CELL SIGNALLING			
HHIP	0.21737	0.43807	354.363
RYR2	0.14956	0.3697	354.363
CFTR	0.39799	0.60456	350.694
CELL ADHESION			
CLDN11	0.22769	0.43387	354.363
CDH5	0.42541	0.64487	354.363
CD86	0.73312	0.55526	-154.77
CD28	0.53185	0.33486	-140.76
CYTOKINE CYTOKINE RECEPTOR INTERACTION			
TNFRSF10D	0.11532	0.32766	354.363
CD40	0.18419	0.37311	350.694
TNFSF4	0.2	0.42	350.694
NEUROACTIVE LIGAND RECEPTOR			
GALR1	0.2	0.38	350.694
PTGDR	0.36	0.53	111
LEP	0.44	0.62	114

5.3.3.2 Early onset and Late onset tumours

Analysis of differentially methylated genes in ET (19) and LT tumours (17) showed a total of 6050 CpG sites differentially methylated in ET and 4966 sites differentially methylated in LT (Fig5.6 a,b)

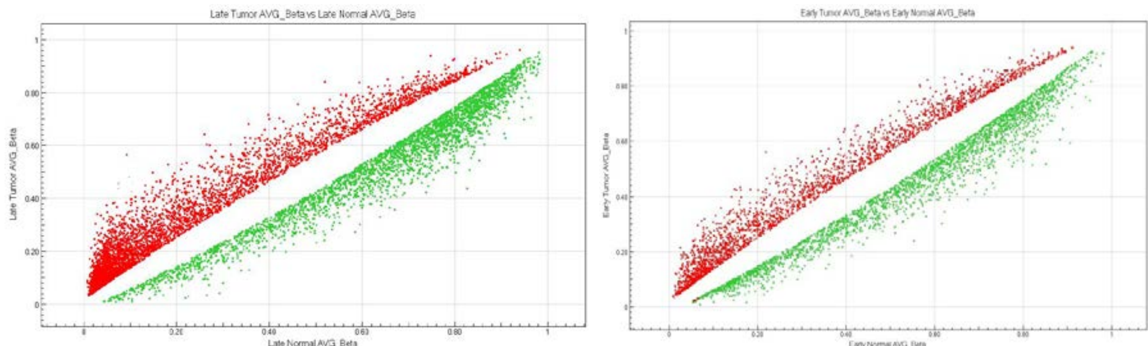


Fig. 5.6 Scatter plots of (a) early and (b) late onset tumours showing hypermethylated and hypomethylated CpG sites

These differential CpG sites corresponded to 528 genes (327 hypermethylated, 201 hypomethylated) in ET and 525 genes (466 hypermethylated , 59 hypomethylated) in LT. Comparison of these genes showed 385 (42%) unique genes aberrantly methylated in ET, 382 (42.3%) genes unique to LT and 143 (15.7%) genes found methylated in both early and late onset tumours (Fig. 5.7).

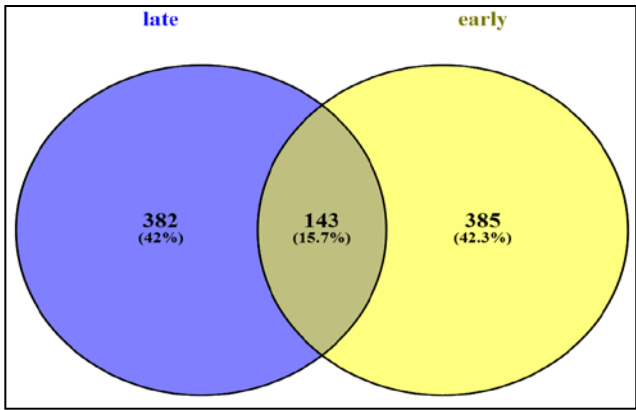


Fig.5.7 Venn diagram representing the differentially methylated genes in early and late onset breast cancer

a. Gene ontology analysis for ET

528 differentially methylated genes identified in ET were further explored for their pathways. Pathways such as Apoptosis, cAMP signalling pathway, cell adhesion molecule pathway, cell cycle, cytokine cytokine recptor interaction pathway were altered.

Apoptosis pathway

Epigenetic regulation of Apoptosis pathway resulted in hypermethylation of tumour supressors and hypomethylation of oncogenes leading to breast cancer progression. *DAB2IP* , *TNFRSF10D* were hypermethylated and *Bcl2, CAD* were hypomethylated. *DAB2IP,* **Disabled homolog 2-interacting protein** a tumour suppressor is a known to be hypermethylated in breast and prostate cancer. It is involved in several functions such as innate immune response, inflammation and cell growth inhibition, apoptosis, cell survival, cell migration and maturation. Hypermethylation of *TNFSRF10D* **Tumour necrosis factor receptor superfamily member 10D** gene has been reported in melanoma. *Bcl2,* **B-cell lymphoma 2** is a antiapoptotic gene and its over-expression along with activation of other proto-oncogenes such as myc causes cancer. *CAD*, **carbamoyl-phosphate synthetase 2** gene plays important role in de novo synthesis of pyrimidine nucleotides which is required for mammalian cells proliferation.

cAMP signaling pathway

Under normal physiological conditions, cyclic nucleotides regulate a myriad of biological processes such as cell growth and adhesion, energy homeostasis, neuronal signalling, and muscle relaxation. *CFTR,* ***cystic*** **fibrosis transmembrane conductance regulator** along with *LIPE*, **lipase** were found to be highly hypermethylated in tumours. Although *MYL9,* myosin light chain 9, *NPY,* neuropeptide Y genes were also hypermethylated but the difference between methylation levels among tumour and normals was marginal. *CFTR* has tumour

suppressive role and plays an essential role in anion regulation and tissue homeostasis of various epithelia which serve as source for many cancers.

Cell Adhesion molecules pathway

Hypermethylation of *CD40, SELE, ICAM2, CDH5* genes in cell adhesion pathway was found in early onset tumours. *CD40* plays a central role in both humoral and T-cell-mediated immunity. *SELE* mediates the adhesion of tumour cells to endothelial cells which is in turn associated with metastatic dissemination. *ICAM2* gene, Intercellular Adhesion Molecule 2 mediates adhesive interactions important for antigen-specific immune response. *CDH5,* Cadherin are calcium dependent cell adhesion proteins and play an important role in endothelial cell biology by controlling cohesion and organization of the intercellular junctions. *CD86* was found to be hypomethylated in early onset breast tumours which is known to play a role in T cell coinhibition (Table 5.3).

Table.5.3 Functional enriched genes getting deregulated in early tumours

SYMBOL	Early Normal AVG_Beta	Early Tumour.AVG_Beta	Early Tumour.DiffScore
BCL2	0.41488	0.257508	-94.9252
CAD	0.362022	0.287858	-21.611
TNFRSF10D	0.152177	0.256622	72.01141
DAB2IP	0.085654	0.313422	350.6937
TNNI3	0.19861	0.361443	350.6937
CFTR	0.397992	0.604564	350.6937
LIPE	0.460502	0.638042	115.5567
NPY	0.236208	0.347527	59.12702
MYL9	0.247866	0.378053	77.72686
CD40	0.184193	0.373106	350.6937
CNTN2	0.215315	0.309111	45.26828
SELE	0.571021	0.754095	135.7661
ICAM2	0.27316	0.448848	133.0642
CDH5	0.254247	0.383927	75.89573
CD86	0.73	0.55	-154

b. Network Analysis

Network analysis was performed for all the genes which were found to be uniquely hyper and hypomethylated in ET. The hypermethylated genes formed highly interactive 140 nodes. Genes as *PDGFRB and STAT5A, CCNA1* top hypermethylated nodes while *SYK, BCL2* were top hypomethylated nodes in early onset tumours (Fig. 5.8 a,b). *PDGFRB* gene encodes a cell surface tyrosine kinase receptor for members of the platelet-derived growth factor family. *CCNA1* gene is a part of highly conserved cyclin family gene which mainly targets activated CDKs during apoptosis. *STAT5*A mediates the response of many cell ligands and promotes carcinogenesis.

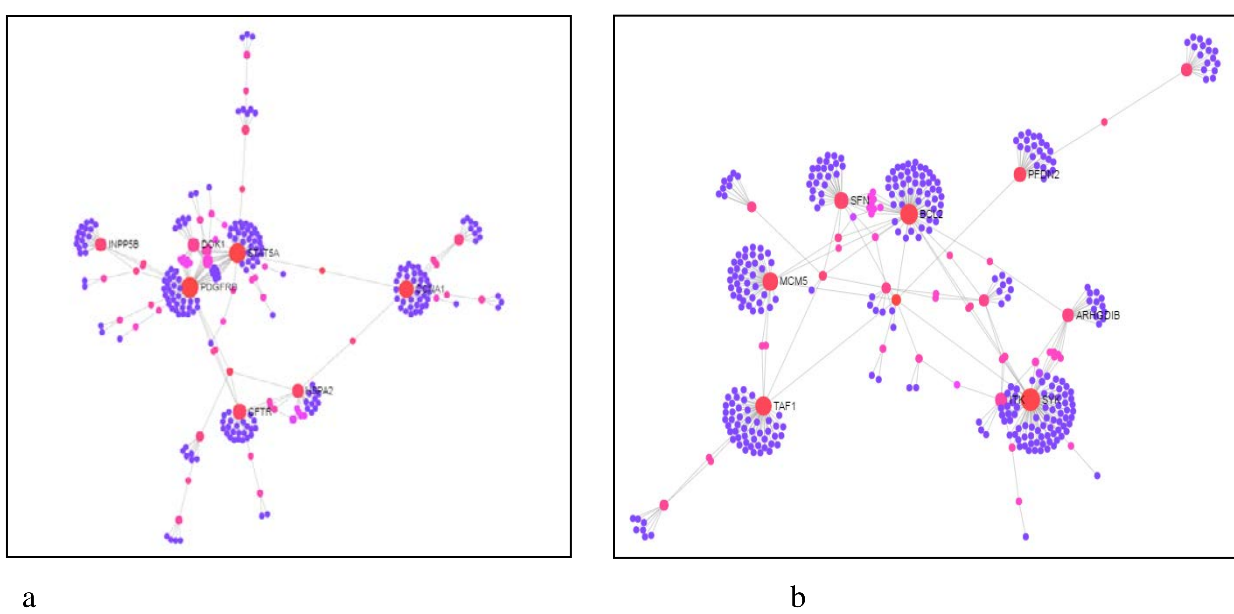


Fig. 5.8 Showing Top nodes uniquely methylated in early onset tumours a.) Hypermethylated and b) hypomethylated

c. Gene Ontology analysis in LT

Late onset tumours expressed 525 differentially methylated genes which were were similar to those found in early onset tumours, however the genes showing aberrant methylation in LT were different. The gene ontology analysis had been done to identify major functional pathways involved in carcinogenesis in late onset tumours. It involved cAMP signalling pathway and cell adhesion pathways (Table.5.4, 5.5)

cAMP signalling pathway showed hypermethylation of TNNI3, BRAF, NPY, RYR2 genes.

Table 5.4 cAMP pathway genes showing their beta values in tumour and normal along with their diff score

SYMBOL	Late Tumour.AVG_Beta	Late Normal_AVG_Beta	Late Tumour.DiffScore
HHIP	0.438071	0.217374	354.3631
NPY	0.384511	0.205567	354.3631
GRIA1	0.275704	0.102357	354.3631
RYR2	0.369698	0.149557	354.3631
TNNI3	0.350143	0.188131	137.0436
MYL9	0.381006	0.223029	118.2771
BRAF	0.373158	0.244154	74.56955

Cell adhesion pathway showed hypermethylation of *CLDN11, CDH5, ICAM2* genes in early onset tumours with hypomethylation of *CD28* gene late onset tumours.

Table 5.5 Cell adhesion pathway genes showing their beta values in tumour and normal along with their diff score

SYMBOL	Late Tumour.AVG_Beta	Late Normal.AVG_Beta	Late Tumour.DiffScore
CDH5	0.64487	0.425413	354.3631
CLDN11	0.433867	0.227685	354.3631
ICAM2	0.319995	0.211619	58.23082
CNTNAP2	0.336275	0.233516	48.81013
ITGA8	0.328822	0.252683	25.35596
CD28	0.334864	0.53185	-140.757

Comparison of aberrantly methylated genes among these pathways in early and late onset tumours lead to identification of significant unique and common genes in both groups (Table5.6).

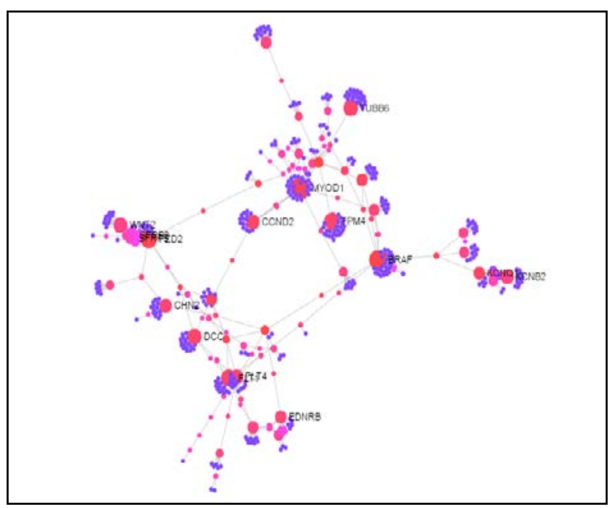
Table 5.6 Unique and common genes showing functional significance in ET and LT pathways

METHYLATION STATUS	**EARLY**		**LATE**
	UNIQUE	**COMMON**	**UNIQUE**
HYPERMETHYLATED	CFTR	TNNI3	RYR2
	LIPE	NPY	GRIA1
	CD40	MYL9	HHIP
			CLDN11
		CDH5	BRAF
HYPOMETHYLATED	CD86		CD28

Genes *CFTR, LIPE and CD40* epigenetically regulate early onset breast carcinogenesis. These play a major role in transport through ECM, growth and metastasis. Genes *RYR2, HHIP, CLDN11, BRAF* regulated late onset carcinogenesis by promoting growth and proliferation.

5.3.3.1.6 Network Analysis

Network analysis was performed for all the genes which were found to be uniquely hyper and hypomethylated in LT which consisted of *MYOD1, BRAF, FZD2* as top hypermethylated genes while *CCND1* was top hypomethylated gene.



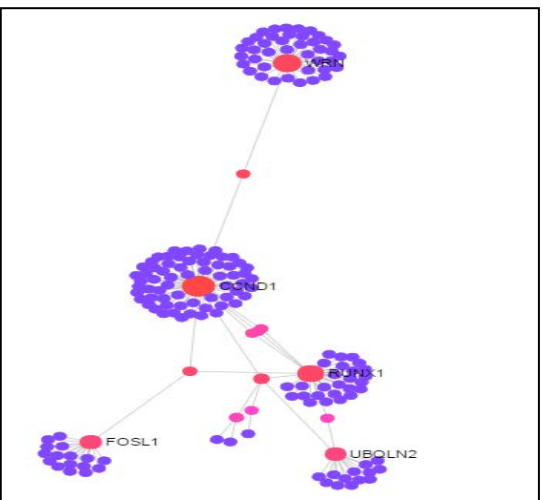


Fig. 5.9 Top interactive nodes in late onset tumours a.) Hyper and b) hypomethylated unique in LT

5.4 Integrome Analysis

Integrome analysis was done to develop a network map of the gene expression and methylation interactome in order to study biological relevance of their interactions and their potential functional role as biomarkers and further deduce biological meaningful conclusion.

Genes which were found differentially expressed and methylated were merged to develop interactome using bridge island software, which displayed interactive enriched genes and pathways. Two sets of genes which were hypermethylated and downregulated and those which were hypomethylated and upregulated were identified (293, 310). A total of 59 genes showed significant interaction between gene expression and methylation in early onset tumours, while 91 genes showed interaction in late onset tumours. Amongst them 42 genes were unique to ET and 74 to LT while 17 genes were found both in early and late onset tumours. Integration of filtered genes from expression and methylation set has been shown in Fig. 5.10

Carcinoma Patients/ Normal Controls

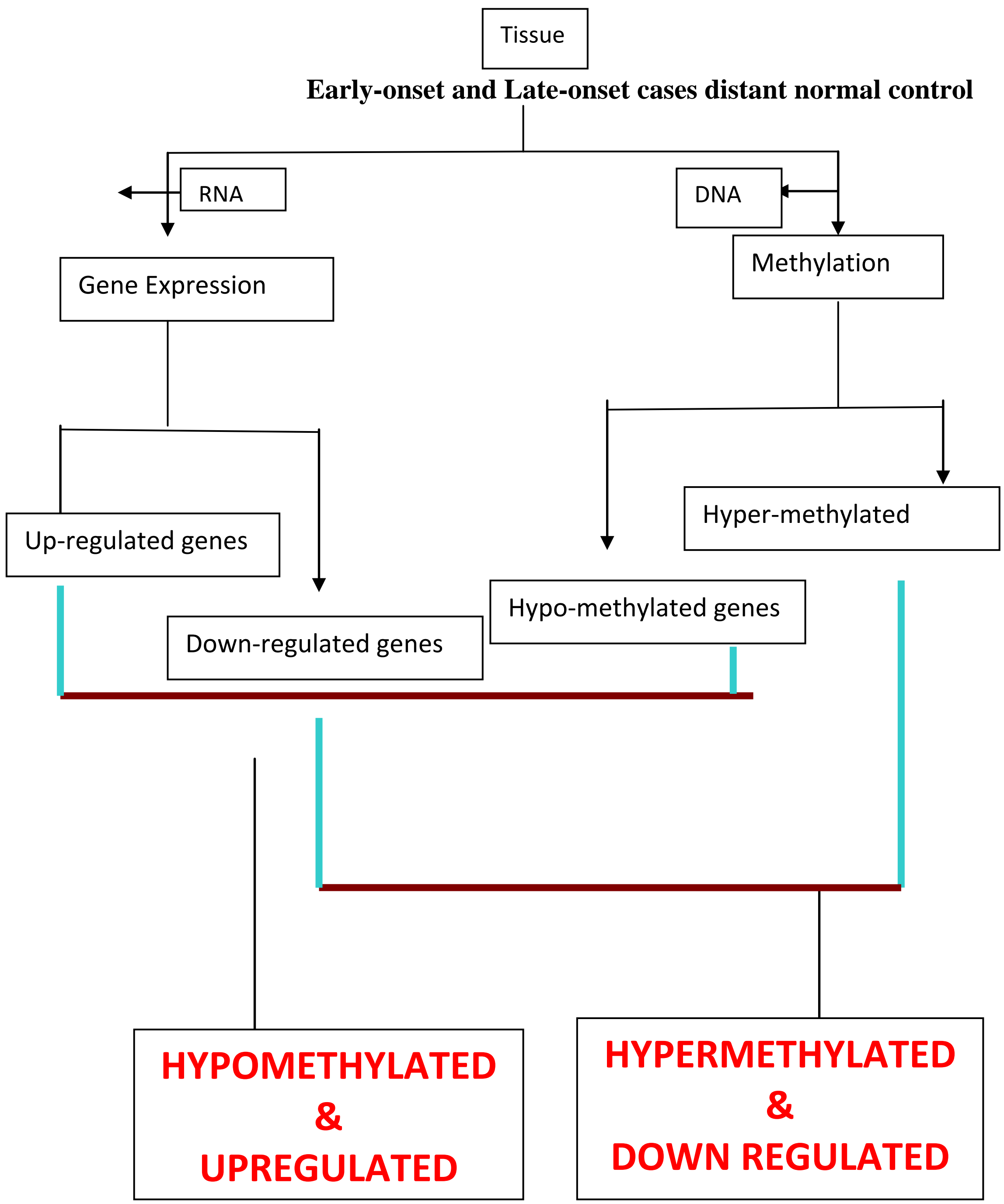


Fig. 5.10 Integration of gene expression and methylation data for breast cancer patients

5.4.1 Early onset tumour integrome

Early onset integrome uniquely consisted of *STAT5A, LIPE, ALDH1A2* genes which were hypermethylated leading to downregulation and hence loss of expression. These genes are mainly a part of pathways as Immune response, fatty acid metabolism and mesenchymal cell development. Genes as *MMP13* (Matrix metalloproteinase) was hypomethylated leading to it's over expression exclusively in early onset tumours

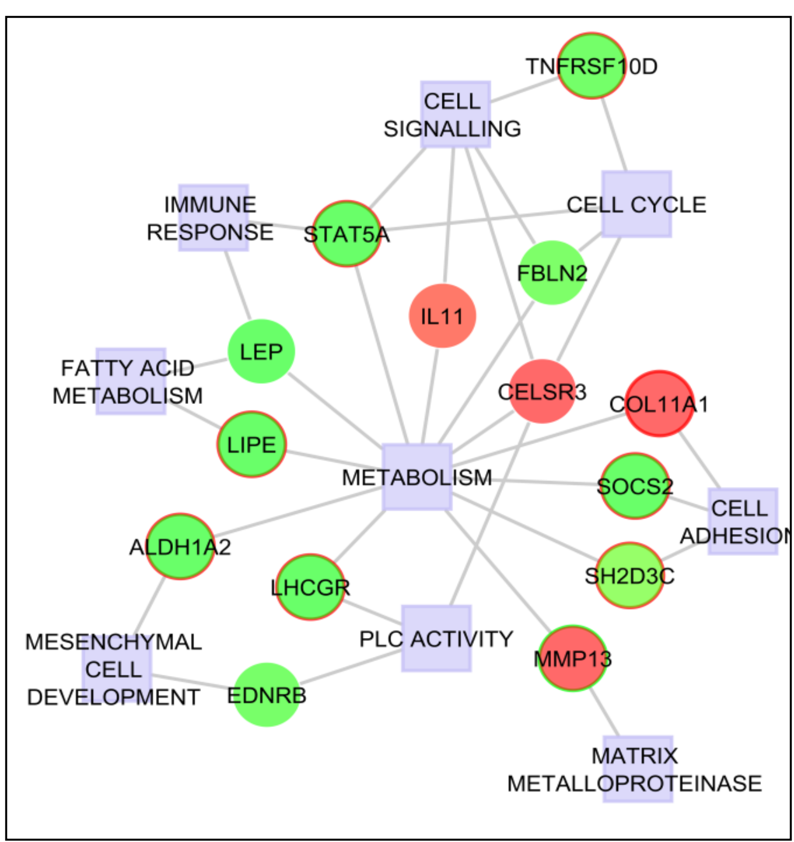


Fig.5.11 Early tumour integrome, here red indicates over expression while green shows under expression of the genes. Red circle indicated hypermethylation while green circle represents hypomethylation.

Table 5.7 List of genes showing their beta value and fold change for early integrome

Gene Symbol	Gx Fold Change	Tumour Beta Value	Normal Beta Value
ALDH1A2	-3.53612	0.326087	0.220613
LHCGR	-2.66263	0.368043	0.239029
LIPE	-14.2705	0.638042	0.460502
MMP13	8.923938	0.657276	0.831124
SH2D3C	-1.37949	0.364704	0.285132
SOCS2	-3.30331	0.202144	0.075554
STAT5A	-2.66207	0.33225	0.161668
TNFRSF10D	-1.85759	0.256622	0.152177

5.4.2 Late onset tumour integrome

Methylation of genes *LEP, EDNRB, FBLN2* was correlated with gene expression in late onset tumours. Theses play a major role in fatty acid metabolism, mesenchymal cell development and cell cycle. Two other genes *CELSR3, IL11* which belonged to cell signalling pathway and immune response pathway were over expressed due to hypomethylation uniquely in late onset tumours

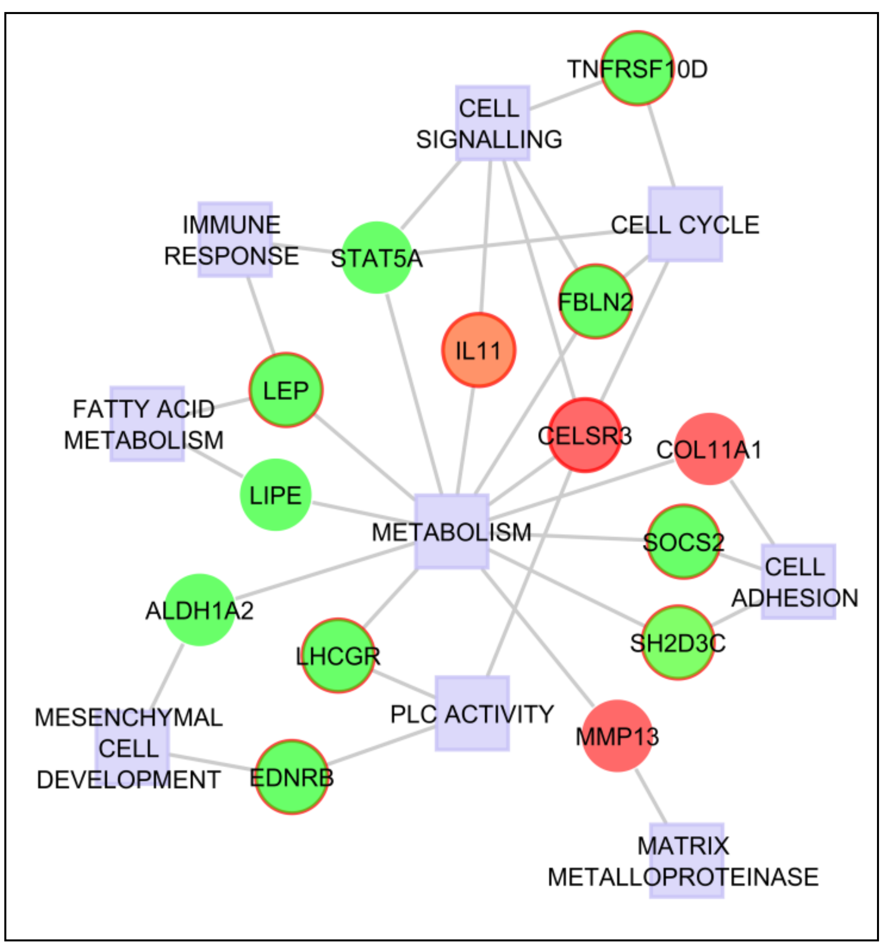


Fig.5.12 Late onset tumour integrome, here red indicates over expression while green shows under expression of the genes. Red circle indicated hypermethylation while green circle represents hypomethylation.

Table 5.8 List of genes showing their beta value and fold change for late integrome

Gene Symbol	**Gx Fold Change**	**Tumour Beta Value**	**Normal Beta Value**
EDNRB	-2.0594	0.395349	0.267877
FBLN2	-2.67356	0.315739	0.163669
LEP	-17.5896	0.624537	0.445898
LHCGR	-2.66263	0.379807	0.281122
SH2D3C	-1.6786	0.360766	0.255421
SOCS2	-3.30331	0.202144	0.075554
TNFRSF10D	-2.06288	0.327664	0.115319

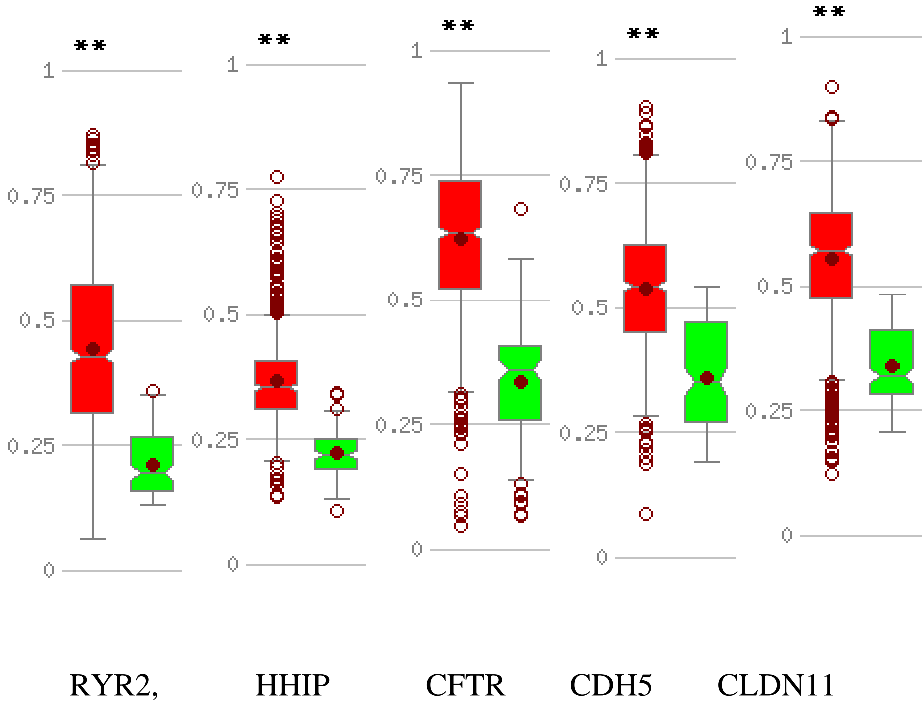
There was a common set of genes amongst early and late integrome consisting of *SOCS2, TNFRSF10D, SH2D3C* which were under expressed due to hypermethylation. They were a part of pathways as cell adhesion, cell signalling, PLC activity.

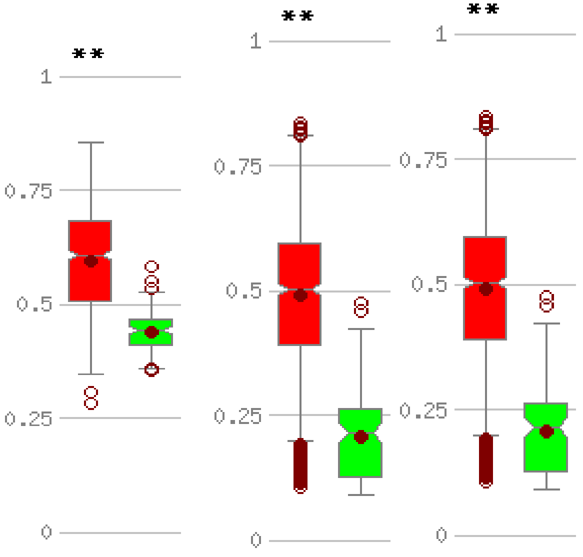
5.5 Validation

Validation of differentially methylated genes was done in 740 human breast cancer patients along with 92 Normal from TCGA database via MethHC (A database of DNA methylation and gene expression in human cancer) software. It was also observed that hyper methylation included mainly promoters of tumour supressors whose expression becomes repressed thereby facilitating cancer formation. Similarly hypomethylation at promoters of oncogenes results in their over expression and thus leads to cancer progression.

5.5.1 All tumours vs normal control

Genes such as *HHIP, RYR2, CFTR, CLDN11, CDH5, TNFRSF10D, CD40* which were found to be functionally relevant since they altered various pathways leading to breast carcinogenesis were selected for further validation. These genes were found to be hypermethylated in microarray and similar concordant results were produced by MethHC database which showed higher average tumour beta values than average normal control beta values, (Fig. 5.13, Table5. 9) as seen in box plots.





TNFRSF10D FLT4 CD40

Fig 5.13 Box plots for average Beta values of tumours and normal controls for differentially methylated genes amongst various pathways in all tumours. Red indicates average tumour beta values while green indicates average beta normal controls values. (** indicates p value <0.001)

The following table depicts the beta values obtained from MethHC database for tumour and normals along with the p values.

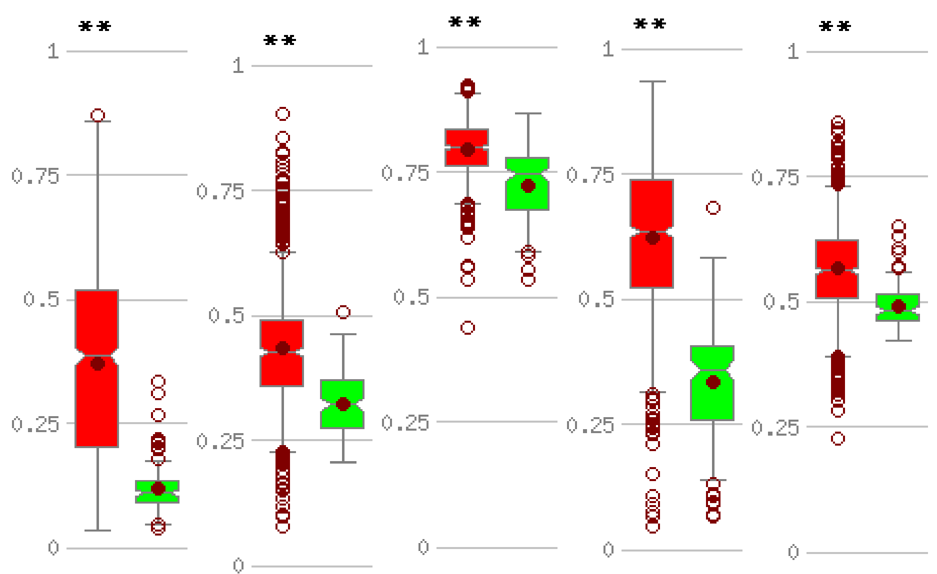
Table 5.9 Beta values of tumour and normals breast cancer tissues obtained after validation

GENE	ACCESSION	MEDIAN TUMOUR BETA	MEDIAN NORMAL BETA	P-VALUE
HHIP	NM_022475	0.3531	0.219	2.48E-14
RYR2	NM_001035	0.428	0.1937	5.96E-14
CFTR	NM_000492	0.63495	0.3591	7.66E-15
CLDN11	NM_001185056	0.5424	0.3511	3.02E-14
CDH5	NM_001795	0.57135	0.3176	5.40E-14
TNFRSF10D	NM_003840	0.60645	0.4426	7.51E-14
CD40	NM_001250	0.50245	0.2141	3.18E-14

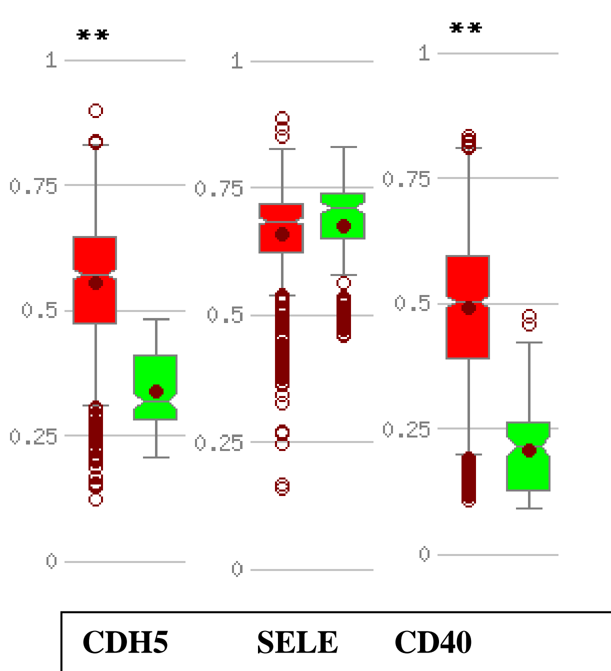
5.5.2 Pathways found differential amongst ET/LT groups

Further validation of genes differentially methylated in early and late tumours was done using MethHC database in all 740 tumour tissues compared to 92 normal controls (since the database did not showed distinction between early and late onset tumours).

Genes as *NPY, MYL9, LIPE, CFTR, TNNI3, CDH5, SELE, CD40, CAD, BCL2* playing functional role in early and late onset breast carcinogenesis were selected for validation. All genes were found to be hypermethylated in tumours samples except *Bcl2, CAD* which were found hypomethylated both in microarray and MethHC database. Another gene *SELE* which was hypermethylated in our data was found to be hypomethylated in MethHC database, however it did not reached to level of significance (Fig. 5.14 table 5.10). The observed median beta value for tumours and normals are mentioned in table along with the p value .



NPY MYL9 LIPE CFTR TNNI3



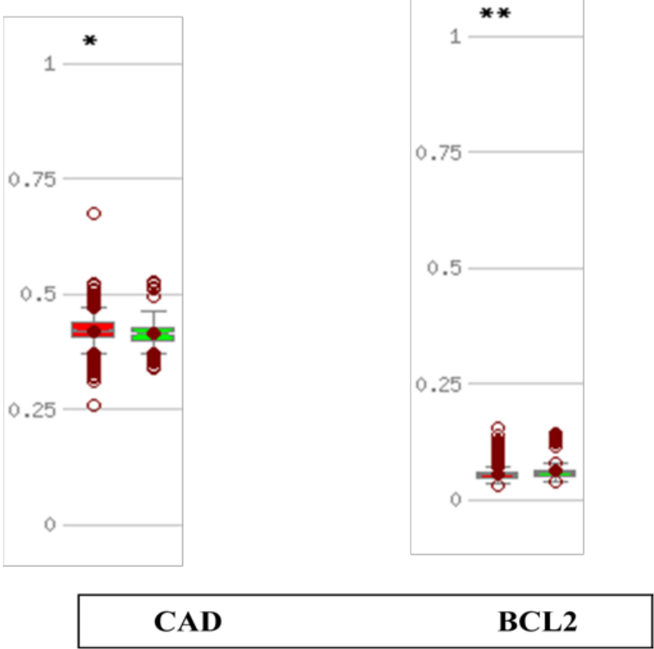


Fig 5.14 Box plots of tumours and normals for differentially methylated genes amongst various pathways in ET/LT. Red indicates average tumour beta values while green indicates average beta normal values. (** indicates p value <0.001, * indicates p value <0.05)

Table 5.10 Median beta value and p values for tumours and normal controls

GENE	ACCESSION	MEDIAN TUMOUR BETA	MEDIAN CONTOL BETA	P-VALUE
CFTR	NM_000492	0.63495	0.3591	7.66E-15
LIPE	NM_005357	0.7969	0.7441	8.39E-14
MYL9	NM_006097	0.42605	0.226325	1.18E-13
NPY	NM_000905	0.38405	0.1098	7.33E-14
TNNI3	NM_000363	0.56095	0.4827	1.36E-13
CD40	NM_001250	0.50245	0.2141	3.18E-14
CDH5	NM_001795	0.57135	0.3176	5.4E-14
ICAM2	NM_000873	0.6353	0.5221	6.36E-13
CNTN2	NM_005076	0.39775	0.3248	-9.5E-15
SELE	NM_000450	0.6827	0.709	0.147871
CAD	NM_004341	0.4196	0.4152	0.034588
BCL2	NM_000633	0.05245	0.0576	0.002475

The validation using MethHC software confirms our differential methylation profiles obtained using microarray.

Owing to the functional involvement of these genes in various pathways their hypermethylation or hypomethylation plays an important role in tumourigenesis and breast cancer progression.

5.6.2 Integrative gene expression and methylation profiles

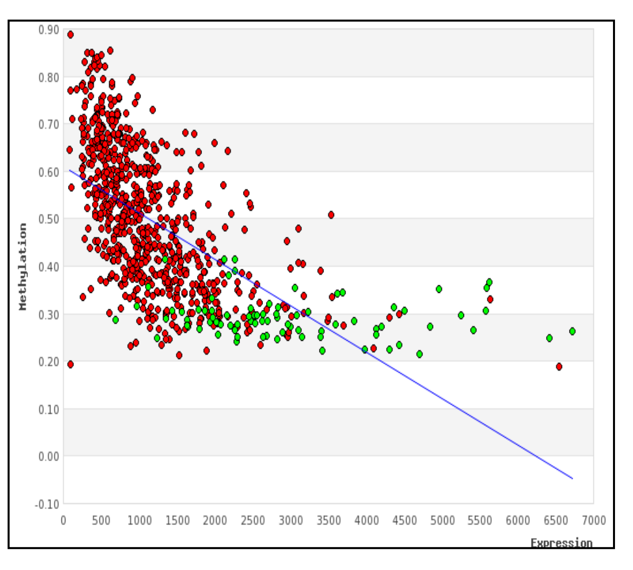
Validation of genes showing significant integration on integrome analysis was also done by MethHC database in 839 tumour tissues. Scatter plots were generated for tumours and controls gene expression and methylation intensity and beta values. Significant negative correlation was found for genes *TNFRSF10D, LEP, ALDH1A2, LHCGR, SOCS2, STAT5A, SH2D3C* which were hypermethylated resulting in their loss of expression. However correlation amongst remaining 4 genes *LIPE, FBLN2, MMP13, EDNRB* did not reached to level of significance (Table5.11).

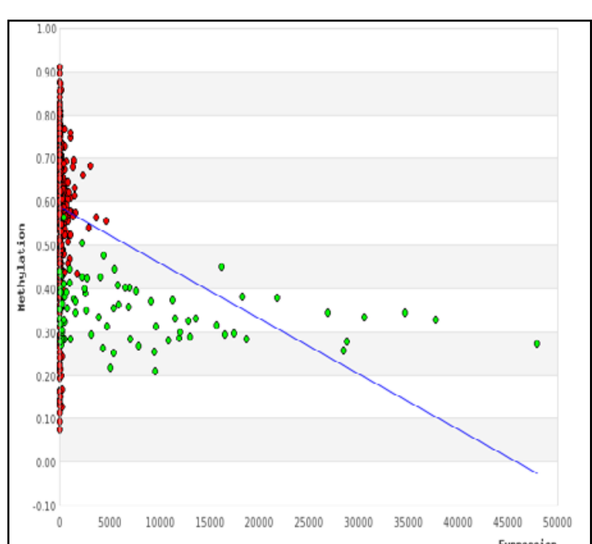
Of the genes showing significant correlation we found the expression of *STAT5A, ALDH1A2* genes to be significantly regulated through hypermethylation in early onset tumours while LEP gene was significantly repressed due to hypermethylation in late onset tumours (Table5.11) Besides this 4 genes *LHCGR, SOCS2, TNFRSF10D, SH2D3C* were also found hypermethylated leading to loss of its expression in both early and late onset tumours (Fig 5.15).

Table 5.11. Interactome genes showing their correlation coefficient and correspondimg p values

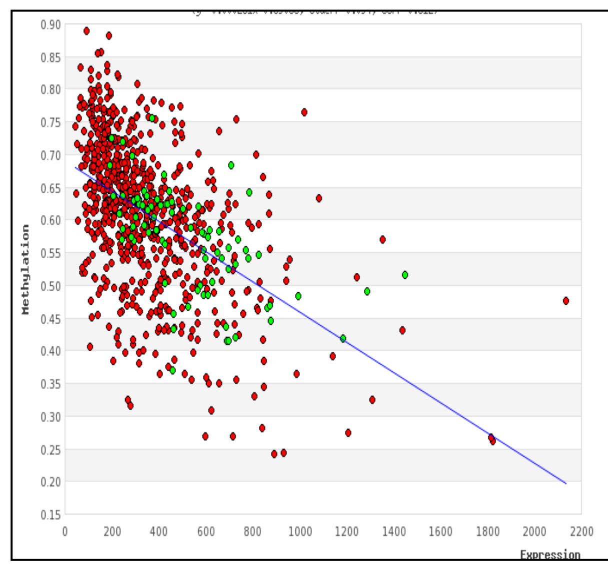
GENE	CORRELATION	P VALUE
STAT5A	**-0.507934609**	**1.11E-16**
LEP	**-0.10233753**	**2.22E-16**
ALDH1A2	**-0.146384096**	**3.33E-16**
LHCGR	**-0.066247526**	**7.11E-05**
SH2D3C	**-.5276209114**	**5.55E-16**

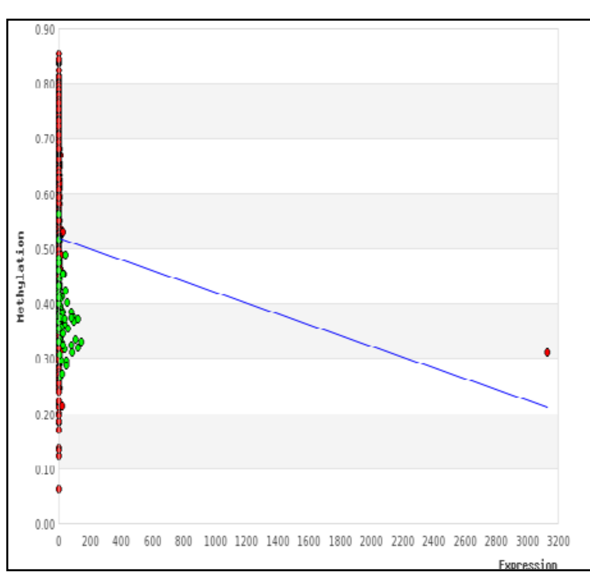
SOCS2	**-0.030401153**	**5.55E-16**
TNFRSF10D	**-0.360619155**	**6.66E-16**
MMP13	-0.205262142	0
FBLN2	-0.239788084	0
LIPE	-0.058124235	0
EDNRB	-0.077784999	0



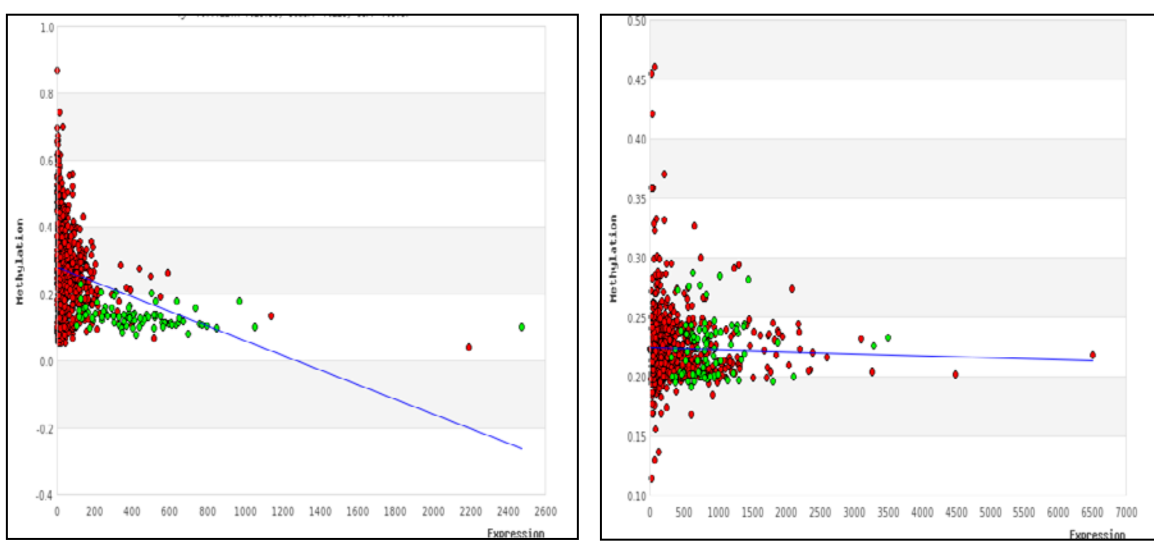


a b

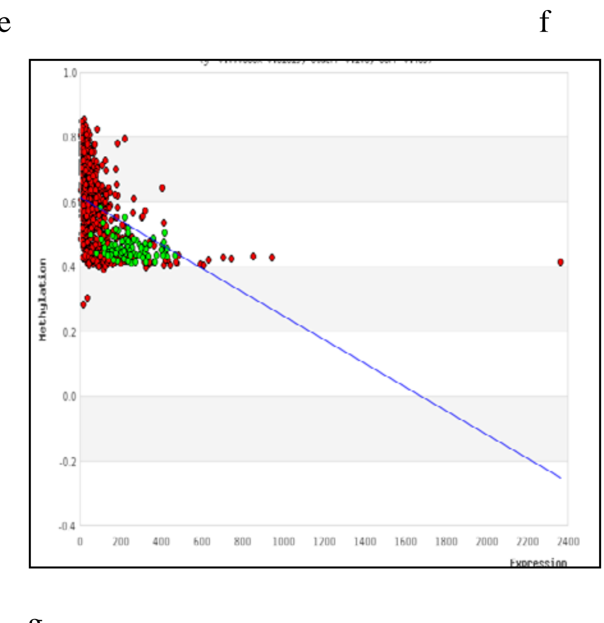




c d



e f



g

Fig. 5.15 Validation of significantly interacting and correlating genes as found in microarray experiments using MethHC software. a) STATA5A b) LEP c) ALDH1A2 d) LHCGR e) SH2D3C f) SOCS2 g) TNFRSF10D. Red dots shows the correlation for tumours and green dots shows that for normals

This study showed the differential methylation profiles obtained in breast cancer patients and the distinguishing profiles in early and late onset breast cancers. The results obtained by the methylation microarray were further validated by TCGA database. We identified *STAT5A,*

ALDH1A2 genes to be significantly downregulated through hypermethylation in early onset tumours while *LEP* was repressed due to hypermethylation in late onset tumours. Hence it is important to mention that these genes are epigenetically regulated and they might play a important role in tumour progression by activating specific pathways pertaining to breast cancer. These could also probably serve as potential drug targets in early and late onset breast carcinogenesis. Their functional role has to be further explored.

Discussion

The development and progression of breast cancer is a multi-step process modulated by genetic, epigenetic and environmental factors. DNA methylation modifies DNA resulting in dysfunction and deregulation of genes. As DNA methylation is a reversible process (311) it can result in gene re-expression which in turn can lead to normal gene regulation (312). Therefore it can serve as a potential therapeutic target. Hence it would be interesting to determine the changes in DNA methylation which plays an essential role in breast cancer development, emphasizing the crucial role of these epigenetic changes for future diagnosis, prognosis and prediction of response to therapies.(313-314)

In the present study whole genome methylation profiling of 48 breast cancer tissues had been studied to obtain distinct methylation profile of breast cancer in Indian women. A total of 910 genes were found aberrantly methylated, of which 667 genes were hypermethylated and 243 were hypomethylated. Pathways which were epigenetically regulated include Apoptosis, cAMP signalling, Cell adhesion, Cytokine cytokine receptor interaction, neuroactive receptor ligand interaction pathway. Various genes found hypermethylated among these pathways included *HHIP, RYR2, CFTR, CLDN11, CDH5, TNFRSF10D, CD40, GALR1* while *CAD, BCL2* genes were found hypomethylated. Downregulation of HHIP gene may activate hedgehog signalling pathway leading to EMT and stem cell activation. Song Y. *et al*. has

described HHIP to be hypermethylated in gastric cancer (307, 315) but there are no reports suggesting its role in breast cancer. Expression of *HHIP* is also considered as significant prognostic marker in glioblastoma patients (316). *RYR2* works as a key calcium channel by operating calcium influx during EMT and hence controls the transitions during carcinogenesis (317). It is reported to be aberrantly methylated in breast cancer (318). *CFTR* gene plays important role in carcinogenesis since it encodes members of ABC transporters family which enables transport of molecules across membrane. It is reported to be hypermethylated in prostate cancer (319) and is found to be associated with poorer survival in young patients in non small cell lung cancer (320). Cell adhesion genes as *CLDN11, CDH5* play critical role in formation of junctional proteins which are important for epithelial mesenchymal transitions. Agarwal R. *et al*. and Abe M. *et al*. have described hypermethylation of *CLDN11* in gastric and oral squamous cell carcinoma (256, 321). However its role in breast cancer is still unknown. *CDH5* is a known tumour suppressor gene and is known to be hypermethylated and silenced in many cancers (322-325). It has major role in imparting adherens and maintaining junctional assembly. *TNFRSF10D p*lays a inhibitory role in inducing TRAIL induced apoptosis. Li *et al*. has reported hypermethylation of *TNFRSF10D* in breast cancer. CD40 is hypermethylated in breast cancer and is known to generate different growth signals in tumours and controls leading to tumour cell proliferation (326-327). Chunlong Zhang *et al*. has reported negative correlation between *CD40* methylation and gene expression (327). *GALR1 a*ctivates variety of secondary intercellular pathways. Hypermethylation of GALR1 gene had been reported as predictive biomarkers for clinical outcome in head and neck cancer patients (328). Its hypermethylation also plays important role in regulation of neoplastic transformation from non-malignant intestinal metaplasia to cancer in gastric cancer (329).

Analysis of methylation profile in early onset breast cancer showed Apoptosis, cAMP signalling and cell adhesion pathways. Genes such as *DAB2IP, CFTR, TNNI3, LIPE, CD40,*

SELE, ICAM2 were found hypermethylated in ET. *Dab2* is a known tumour suppressor gene methylated in many cancers (330-331) and its hypermethylation in breast cancer in Indian women had been reported earlier also (332-333). Morbie *et al*., Son J.W. *et al*., Ashour N. have described hypermethylation of CFTR gene in many cancers (319-320, 334). *CFTR g*ene is a known tumour suppressor and modulates transport of chloride ion across epithelial cell membrane with its dysregulation leading to thickened mucus in lungs. *TNNI3* forms component of striated muscles and its role is not known. *LIPE* gene induces adiposity and is known to be hypermethylated in adipose tissue (335). *Selectin E* is hypermethylated and hence is lost during breast cancer progression. It is also reported to be hypermethylated in GBM (336-337). Another gene *BCl2* was hypomethylated in early onset breast cancer. Bcl2 is antiapoptotic and hence its hypomethylation results in its over expression leading to cancer progression. *CAD* is a strong tumour suppressor and was reported to be hypomethylated in ovarian cancer, breast cancer leading to its over expression (25, 338). In late onset tumours pathways as cAMP signalling and cell adhesion were major aberrantly methylated pathways which consisted of *RYR2, HHIP, CDH5, CLDN11*. These aberrant methylation profiles were further confirmed by MethHC database.

Network analysis in early onset tumours showed *CCNA1, PDGFRB, STAT5A* as top hypermethylated genes, while *SYK, BCL2* were top hypomethylated genes. CCNA1 gene which has the ability to control cell differentiation or suppress cell growth (339), is reported to be frequently hypermethylated in head and neck squamous cancer cells, nasopharyngeal cancer, cervical cancer and colon cancer (340-341) resulting in its silencing. Hypermethylation of *PDGFRB* in breast cancer is reported by Zibo Li *et al*. (342) also loss of SYK protein due to hypermethylation had been found associated with distant metastasis breast cancer patients (343). Most interactive node in Late onset tumours consisted of *MYOD1, BRAF* as top hypermethylated and *CCND1* as top hypomethylated node.

Altered DNA methylation patterns set the hallmarks for cancer by either repressing or overexpressing the genes which have mechanistic and functional role in cancer. It has also been established now that different gene expression patterns contribute to breast cancer heterogeneity (344). Merging of gene expression with methylation profile had been done to obtain the integrative profiles in breast cancer. Integration resulted in 59 and 91 correlated genes in ET/LT respectively. The early integrome comprised of significant hypermethylation of genes *STAT5A, ALDH1A2* genes resulting their downregulation. *STAT5A* deficiency in breast cancer results in impaired prolactin dependent differentiation of mammary cells. There is a correlation between low nuclear level of *STAT5A* and its association with unfavourable clinical outcomes and cancer progression. High *STAT5A* was suggested to be an inhibitor of invasion and metastasis and therefore an indicator of favourable clinical outcomes(345). *STAT5A* is a major part of Immune response pathway where it contributes to tumour development compromising immune surveillance (346-349).Studies have indicated key role of *STAT5*A in leukemia, breast, colon, head and neck, prostate cancers. Unphosphorylated or inactive *STAT5A* suppress tumour growth in colorectal cancer while active *STAT5A* expression has been shown as potential prognostic marker in oral squamous cell carcinoma (346, 350). *ALDH1A2* (aldehyde dehydrogenase 1 family, member A2) oxidizes retinaldehyde into retinoic acid. Kim *et al*. reported hypermethylation of *ALDH1A2* gene a candidate tumour suppressor and hypermethylated in primary prostate tumours as compared with normal prostate (351). Methylation of this gene has been associated with changes in alcohol metabolism (352). Its role in breast cancer has yet to be elucidated.

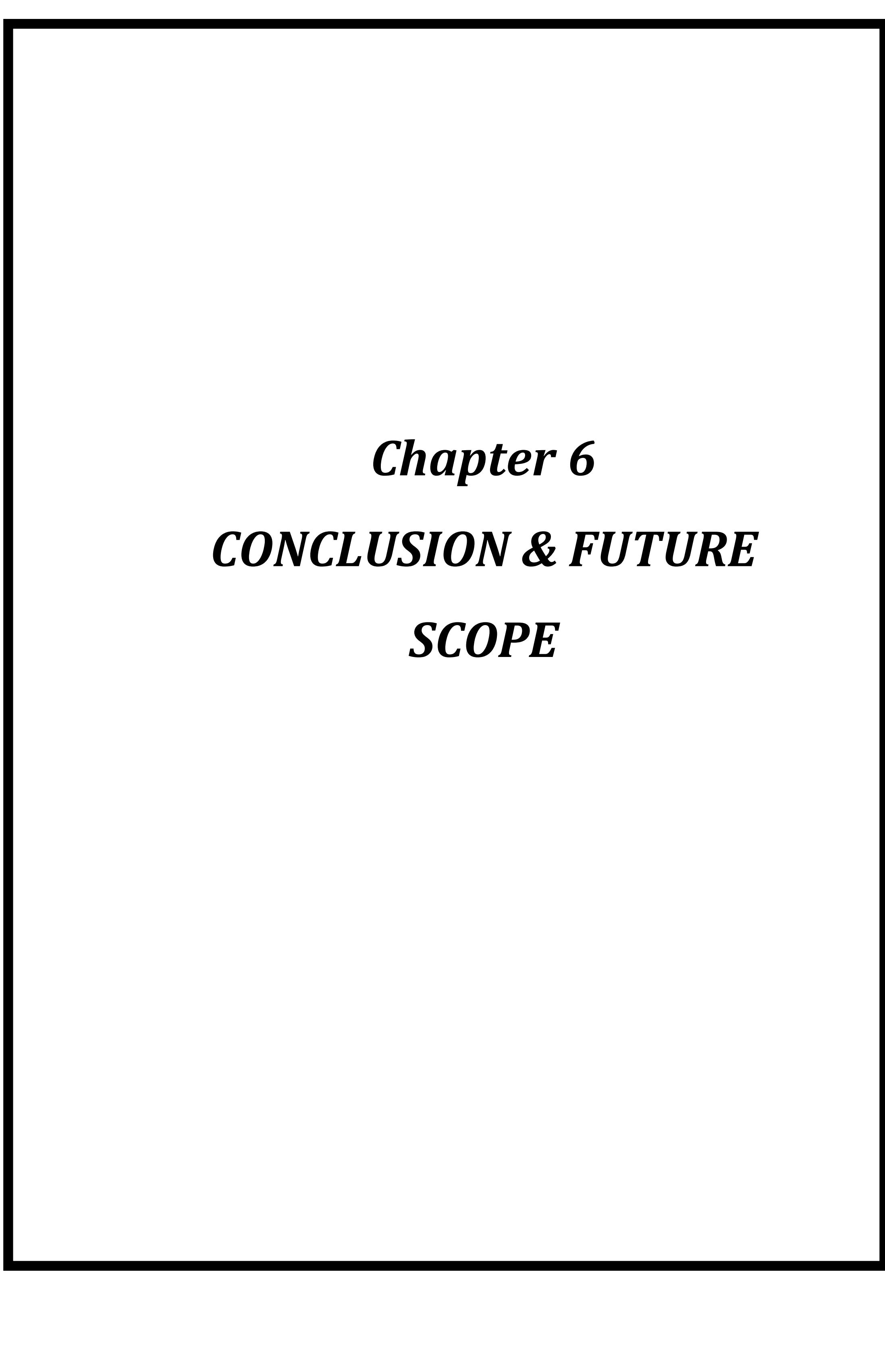
Late integrome consisted of hypermethylation of *LEP* gene. LEP gene has many endocrine functions and regulates immune and inflammatory responses, angiogenesis and wound healing. It is known to encode protein playing major role in body weight regulation. This protein acting through the leptin receptor, functions as part of a signalling pathway to

maintain consistency of the adipose mass. One of the mechanisms for prostate cancer involves methylation of circulating hormones associated with obesity and insulin resistance, such as leptin (353-354). Its hypermethylation resulted in significant loss of its expression in late onset tumours.

Some genes were commonly methylated genes which further lead to their repression such as The *SOCS2, TNFRSF10D, LHCGR. SH2D3C* genes were found hypermethylated in ET and LT. This family of genes activates JAk/STAT pathway. They interact with major molecules of signalling complexes to block further signal transduction by proteasomal depletion of receptors or signal-transducing proteins via ubiquitination. *SOCS2* gene is a tumour suppressor (355-356) and its low expression has been associated with hepatocellular, breast, pulmonary and ovarian cancers (355, 357-359). ***LHCGR*** (luteinizing hormone/choriogonadotropin receptor) gene activation is tightly regulated by the methylation status of its core promoter sequence. It is a central player for human gonadal maturation and functions. Its hypomethylation has been observed in tumours contributing to their increased aggressiveness(360). *TNFRSF10D belongs* to tumour necrosis receptor family protein and is involved in cell signalling and cell cycle. It is epigenetically silenced in human melanoma (222, 361), as well as in cancers of lung, breast, bladder, mesothelioma, prostate, cervix, brain ovary and in hematopoietic malignancies (362). SH2 domain containing 3C are DNA markers associated with fetal growth (363) and are also involved in cell migration. Its role is not known in breast cancer.

In present study, integrated analysis demonstrated that methylation status of different genomic regions plays a key role in establishing transcriptional patterns in human breast cancer. Hence understanding the functional and mechanistic impact of these distinct regions of DNA methylation on gene expression patterns may further provide additional insight into breast

cancer progression along with response to therapy, which is critical for improving management of the disease. Thus the critical challenge is to derive high-quality DNA methylation signatures that are confirmed in prospective studies as specific and sensitive predictors of clinical outcome and therapeutic responses. These specific genes leading to repression or over expression due to their hyper or hypo methylation were hence identified in breast cancer patients in this study. An additional question is to determine whether DNA methylation signatures would provide advantages over current histopathological and immunocytochemical methods. Hence these DNA methylation markers with independent prognostic value can be further evaluated for application in tailoring treatment to patient groups that today are receiving uniform treatment regimens.



Chapter 6

CONCLUSION & FUTURE SCOPE

Conclusion and Future Scope of Work

The rising incidence of breast cancer in India has enabled it to attain number one position with a higher proportion of young early onset breast cancer patients. Hence whole genome gene expression and methylation profiling was performed in Indian breast cancer patients in order to understand their underlying molecular pathogenesis with special emphasis to early onset breast carcinoma

1. Amongst total 97 histopathologically diagnosed cases of breast cancer 42.3% cases were early onset (≤40 years) and 35.9 were late onset cases (≥55 years). Early onset breast cancer had significant number of patients that did not breast fed (p=0.029) and have attained menarche before 13 years of age (p=0.023). Young patients with breast cancer presented significant association with Lymph node (p=0.05) and advanced stage (p=0.07). Higher number of these patients was estrogen receptor negative (55%) while less number of them was HER2/neu receptor positivity (45%). Basal subtype found as main representative molecular subtype in early onset patients than older patients (24 % vs. 18%). However the difference was not found significant probably because of less sample size corresponding to each sub group.
2. Gene expression profiling of 29 TT and 9 CC identified 2413 DEGs where the topmost upregulated genes such as *COL10A1, MMP11, GJB2, CST1, MMP1, MMP13, KIAA1199* and topmost downregulated genes such as *PLIN, KIAA1881, ADH1A,ADH1B, CIDEC, THRSP* . Pathway analysis for all significant deregulated genes in breast tumours identified Cell cycle, Cell adhesion molecules, PPAR, ECM pathway, DNA Replication to be majorly deregulated.

3. Gene set Enrichment analysis also identified genes as *COL10A1*, *GJB2, MMP11*, *MMP13*, *MMP1*, *COL11A1* as top upregulated and *AQP7, PLIN, ADH1B, CIDEC* topmost downregulated genes in breast cancer.
4. The top deregulated genes found in advanced stage of tumours included as *VCAN, COL1A2, ITGA11, MYC, JUN* while the top interactive nodes present in basal tumours included as *JUN, RPS23, LYN* genes.
5. Further expression analysis for ET and LT identified 1685 DEGs in ET of which 724 were upregulated and 961 downregulated while LT consisted of 2379 DEGs of which 988 were upregulated and 1391 downregulated. Further ET had 420 genes and LT had 1114 genes unique to these groups in addition to 1265 common gene set.
6. Several pathways were deregulated in ET and LT which had some differential genes unique, to each group along with a set of genes which were common to both ET and LT. Pathways deregulated in ET along with the set of uniquely expressing gens are cell adhesion (*CDH15, VCAN, PVRL2), Cell Cycle (CDKN2A, SMAD3), ECM receptor interaction (IBSP, ITGA11),* pathways in cancer metastasis (*BRAF, FZD2).*
7. Networks were created for the genes uniquely expressing in ET, it showed *RPL29* and *FZD2 as* top upregulated nodes. BRAF gene was the top upregulated node playing a important role in various pathways and mechanisms of early onset breast cancer. SMAD3 was top downregulated node it was the most interactive node playing significant functional role in early onset breast cancer. The top node found in basal group of breast cancer included *JUN* and *RSP23* as downregulated and *LYN* was upregulated. Advanced stages of cancer also showed *Myc* and *Jun* topmost interactive node. Networking of lower stage tumours showed *PRKDC , PRKCA* as the top deregulated node

8. For the first time we predict the gene sets from Indian patient using PAM50 classifier and also they showed the concordance with the profiles predicted by IHC profiles. Various molecular subtypes identified viz Luminal A (24%), luminal B (20%), HER2/neu positive (20%), basal (31%) and normal-like (3%). Basal subtype was most aggressive subtype and genes *ITGA8, BRAF, CKS1B, E2F3, STAT1* were uniquely upregulated in basal subtype. *JAM3, C4BPB, THBS4* were important genes which were downregulated leading to metastasis. HER2/neu over expressing group showed over expression of *CNTNAP2, FGB, ERBB2* and down regulation of *FGF2* gene. Luminal A showed over expression of *WNT3 a*nd *COL3A1 g*ene while *ME1 g*ene was found to be under expressed. Luminal B showed *over expression of CDH1 and IGF1R genes and under expression of TCF7L1*, LAMA4, LAMC1 genes.
9. Comparison with western population identified a set of unique genes found in Indian population comprising of *ESAM, MYL7, IL17B, CNTFR, PVRL3, and PVRL4*. However the topmost deregulated genes were common with those deregulated in western population suggesting some common mechanisms between Indian and western population. However when we compared our early onset signature genes with already known early signatures we found 2 common genes viz. BAX and ALDH2 in our data. Also comparison of data with that of Lebanese population (Asian), it was found that the a higher number of genes were common with Lebanese population than western population and thus in turn a larger similarity between the 2 populations, thus in turn
10. Validation of various top deregulated genes identified MMPs such as *MMP1*, *MMP11*, *MMP13, ADAMTS1 and ADAMTS5* significant differentially regulated in breast tumours as compared to normal controls however only *ADAMTS5* gene reached statically significance with late tumours when these were compared amongst ET and LT groups.

11. Methylation profiling in 48 samples (36 tumours and 12 normal controls) identified 910 differentially methylated genes playing important role in cAMP signalling, cell adhesion pathways, cytokine cytokine receptor interaction, neuroactiove ligand recptor interaction pathways involving genes as *RYR2, CFTR, HHIP, CLDN11, CDH5, GALR1, and LEP*.
12. Differential methylation analysis for early onset breast carcinoma showed 385 unique genes altering various pathways such as cAMP signalling, Cell adhesion, apoptosis, cytokine cytokine receptor interaction consisting of unique hypermethylated genes as *CFTR, LIPE, CD40* and hypomethylated genes as *CD86.* Network analysis showed *PDGFRB, STAT5A* and CCNA1 genes as top hypermethylated and *SYK, BCL2* genes as top hypomethylated interactive node.
13. In late onset tumours 382 unique genes involved in cAMP signalling pathway, cell adhesion pathway as differentially methylated consisting of unique hypermethylated genes as *RYR2,GRIA1, CLDN11* and *HHIP* and hypomethylated gene *CD28*. Network analysis showed *MYOD1, BRAF* and *FZD2* genes as top hypermethylated and *CCND1* as top hypomethylated gene.
14. Integrome analysis by merging gene expression and methylation profile showed unique/ interactive significant gene as *STAT5A, ALDH1A2* in ET while late integrome had *LEP a*s the significant correlated gene showing hypermethylation leading to loss of gene expression resulting in breast carcinogenesis in Indian patients as validated by MethHC database.
15. In conclusion; a complex interplay between gene expression and methylation leads to activation of gene sets leading to Invasion/Metastasis/ECM/stem cell proliferation in early onset tumours making them highly aggressive and metastatic in nature. In late

onset tumours mainly lipid metabolism genes play a crucial role in cancer progression (Fig. 6.1)

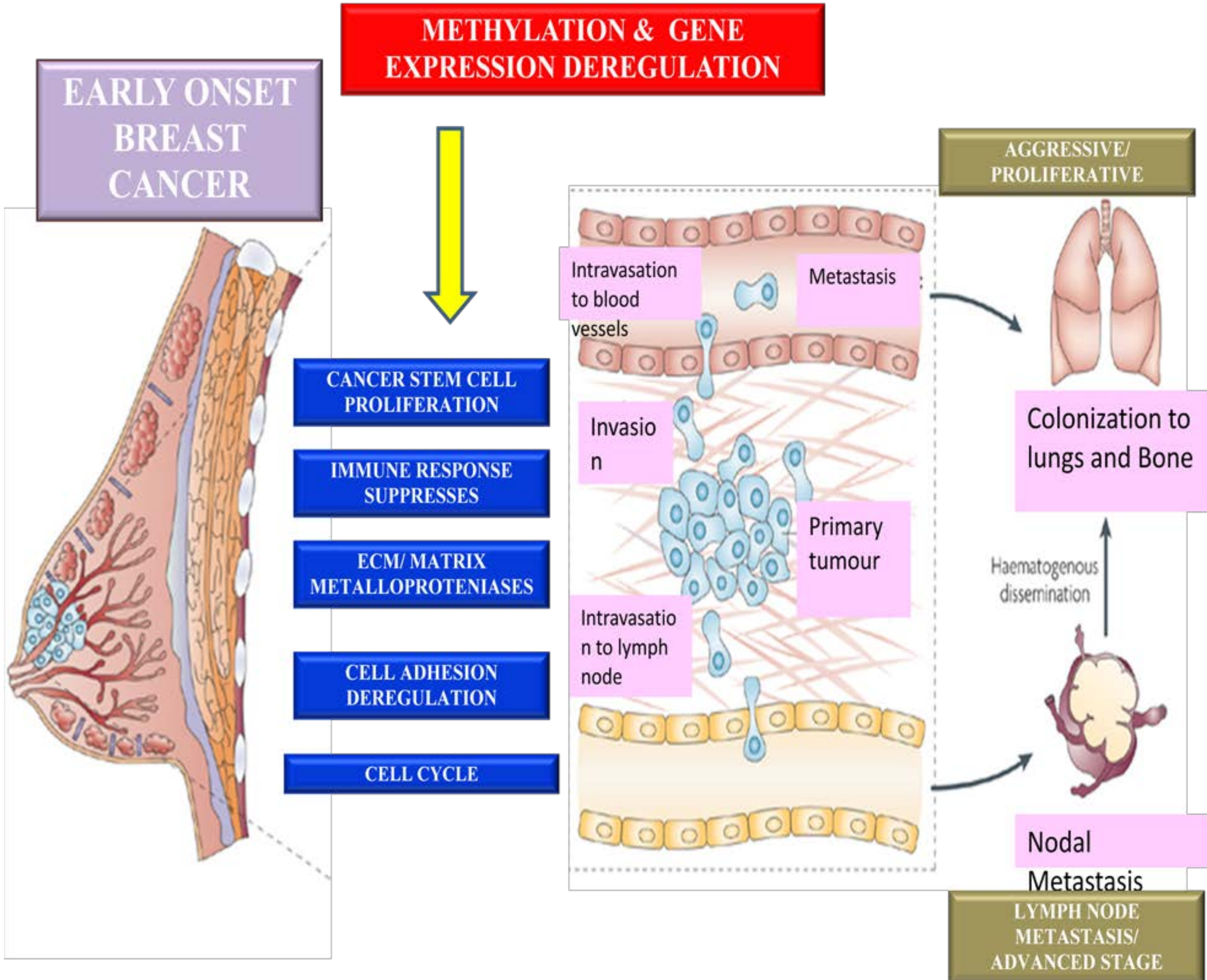
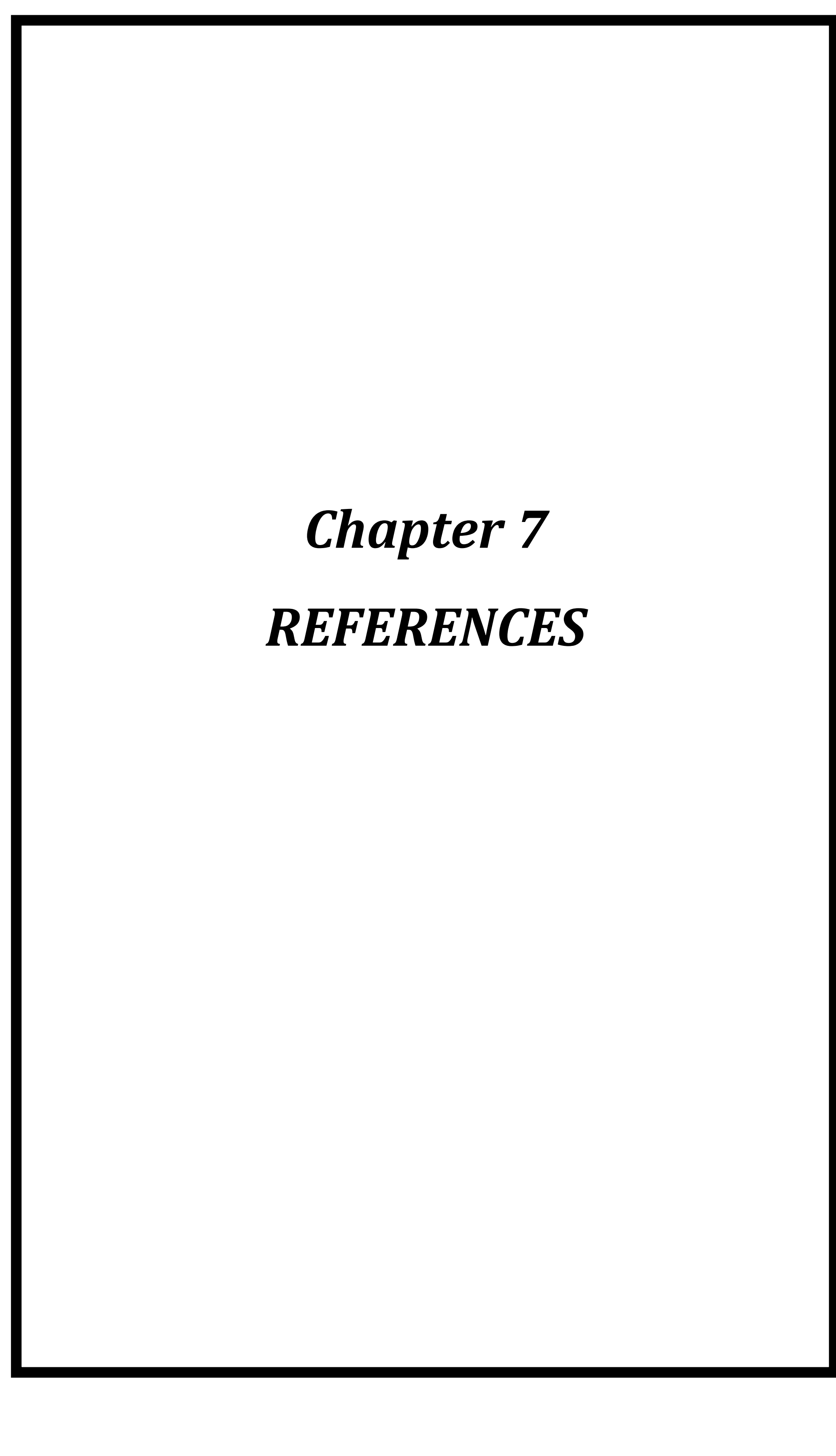


Fig. 6.1 Interplay between gene expression and methylation results in early onset breast carcinogenesis deregulating several genes

Future Scope

Gene expression and methylation profiling in Indian breast cancer patients has led to the discovery of several differentially expressed and differentially methylated genes. The top deregulated genes are common amongst both early and late onset breast cancer. However few

genes were found to express only in early onset patients mainly which were playing important role in ECM pathway leading to invasion and metastasis. The presence of molecular subtypes was first time identified in Indian breast cancer patients showing concordance in the gene sets as predicted by IHC for the various phenotypes of breast cancer *via.* expression profiles. The profiles were also compared with the existing western and Asian profiles and Indian profiles showed higher concordance with Asian population suggesting the similarity in the mechanisms of breast carcinogenesis in Asian countries. Methylation profiling identified the set of 3 genes (ALDH1A2, STAT5A, LEP) for which the gene expression and methylation profiles correlated significantly in early and late onset breast cancer Indian patients.



Chapter 7

REFERENCES

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015 Mar 1;136(5):E359-86.

2. Anonymous. Three Year Report of Population Based Cancer Registries 2012–2014. . Bangalore, India: Indian Council of Medical Research(ICMR); 2016.

3. Kakarala M, Rozek L, Cote M, Liyanage S, Brenner DE. Breast cancer histology and receptor status characterization in Asian Indian and Pakistani women in the U.S.--a SEER analysis. BMC Cancer. 2010;10:191.

4. Leong SP, Shen ZZ, Liu TJ, Agarwal G, Tajima T, Paik NS, et al. Is breast cancer the same disease in Asian and Western countries? World J Surg. 2010 Oct;34(10):2308-24.

5. Winchester DP, Osteen RT, Menck HR. The National Cancer Data Base report on breast carcinoma characteristics and outcome in relation to age. Cancer. 1996 Oct 15;78(8):1838-43.

6. Malvia S, Bagadi SA, Dubey US, Saxena S. Epidemiology of breast cancer in Indian women. Asia Pac J Clin Oncol. 2017 Feb 09.

7. Anonymous. National Cancer Registry Programme 2007-2011 Bangalore: Indian Council of Medical Research (ICMR); 2013.

8. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA. 2001;98.

9. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences. 2003 July 8, 2003;100(14):8418-23.

10. Sotiriou C, Neo S-Y, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proceedings of the National Academy of Sciences. 2003 September 2, 2003;100(18):10393-8.

11. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006;7:96.

12. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000 Aug 17;406(6797):747-52.

13. Sotiriou C, Pusztai L. Gene-Expression Signatures in Breast Cancer. New England Journal of Medicine. 2009;360(8):790-800.

14. Anders CK, Acharya CR, Hsu DS, Broadwater G, Garman K, Foekens JA, et al. Age-specific differences in oncogenic pathway deregulation seen in human breast tumors. PLoS One. 2008;3(1):e1373.

15. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-DeMore N, et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? J Clin Oncol. 2011 Jan 1;29(1):e18-20.

16. Azim HA, Jr., Michiels S, Bedard PL, Singhal SK, Criscitiello C, Ignatiadis M, et al. Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. Clin Cancer Res. 2012 Mar 1;18(5):1341-51.

17. Azim H, Azim HA, Jr. Targeting RANKL in breast cancer: bone metastasis and beyond. Expert Rev Anticancer Ther. 2013 Feb;13(2):195-201.

18. Stadler ZK, Come SE. Review of gene-expression profiling and its clinical use in breast cancer. Crit Rev Oncol Hematol. 2009 Jan;69(1):1-11.

19. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009 Nov 19;462(7271):315-22.

20. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005 Aug;6(8):597-610.

21. Gopalakrishnan S, Van Emburgh BO, Robertson KD. DNA methylation in development and human disease. Mutat Res. 2008 Dec 01;647(1-2):30-8.

22. Szyf M. DNA methylation signatures for breast cancer classification and prognosis. Genome Medicine. 2012;4(3):26-.

23. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009 Dec;1(2):239-59.

24. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003 Mar;33 Suppl:245-54.

25. Widschwendter M, Jones PA. DNA methylation and breast carcinogenesis. Oncogene. 2002 Aug 12;21(35):5462-82.

26. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 Jan 7;100(1):57-70.

27. Hanahan D, Weinberg Robert A. Hallmarks of Cancer: The Next Generation. Cell. [doi: 10.1016/j.cell.2011.02.013].144(5):646-74.

28. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010 Dec 15;127(12):2893-917.

29. Chopra B, Kaur V, Singh K, Verma M, Singh S, Singh A. Age shift: Breast cancer is occurring in younger age groups - Is it true? Clinical Cancer Investigation Journal. [Original Article]. 2014 November 1, 2014;3(6):526-9.

30. Thangjam S, Laishram RS, Debnath K. Breast carcinoma in young females below the age of 40 years: A histopathological perspective. South Asian J Cancer. 2014 Apr;3(2):97-100.

31. Sandhu D, Sandhu S, Karwasra R, Marwah S. Profile of breast cancer patients at a tertiary care hospital in north India. Indian Journal of Cancer. [Original Article]. 2010 January 1, 2010;47(1):16-22.

32. Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: a look at the evidence. Women's health (London, England). 2015;11(1):65-77.

33. Miltenburg DM, Speights VO, Jr. Benign breast disease. Obstet Gynecol Clin North Am. 2008 Jun;35(2):285-300, ix.

34. Guray M, Sahin AA. Benign breast diseases: classification, diagnosis, and management. Oncologist. 2006 May;11(5):435-49.

35. Bombonati A, Sgroi DC. The molecular pathology of breast cancer progression. J Pathol. 2011 Jan;223(2):307-17.

36. Dahlman-Wright K, Cavailles V, Fuqua SA, Jordan VC, Katzenellenbogen JA, Korach KS, et al. International Union of Pharmacology. LXIV. Estrogen receptors. Pharmacol Rev. 2006 Dec;58(4):773-81.

37. Levin ER. Integration of the extranuclear and nuclear actions of estrogen. Mol Endocrinol. 2005 Aug;19(8):1951-9.

38. Cowley SM, Parker MG. A comparison of transcriptional activation by ER alpha and ER beta. J Steroid Biochem Mol Biol. 1999 Apr-Jun;69(1-6):165-75.

39. Herynk MH, Fuqua SA. Estrogen receptor mutations in human disease. Endocr Rev. 2004 Dec;25(6):869-98.

40. Law ML, Kao FT, Wei Q, Hartz JA, Greene GL, Zarucki-Schulz T, et al. The progesterone receptor gene maps to human chromosome band 11q13, the site of the mammary oncogene int-2. Proc Natl Acad Sci U S A. 1987 May;84(9):2877-81.

41. Gadkar-Sable S, Shah C, Rosario G, Sachdeva G, Puri C. Progesterone receptors: various forms and functions in reproductive tissues. Front Biosci. 2005 Sep 01;10:2118-30.

42. Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli GA, et al. Progesterone receptor modulates ER[agr] action in breast cancer. Nature. [Article]. 2015;523(7560):313-7.

43. Durban J, Lazar R, Ofer G. The cracked container, the containing crack: chronic illness--its effect on the therapist and the therapeutic process. Int J Psychoanal. 1993 Aug;74 (Pt 4):705-13.

44. Gao X, Nawaz Z. Progesterone receptors - animal models and cell signaling in breast cancer: Role of steroid receptor coactivators and corepressors of progesterone receptors in breast cancer. Breast Cancer Res. 2002;4(5):182-6.

45. Quenel N, Wafflart J, Bonichon F, de Mascarel I, Trojani M, Durand M, et al. The prognostic value of c-erbB2 in primary breast carcinomas: a study on 942 cases. Breast Cancer Res Treat. 1995 Sep;35(3):283-91.

46. Zhou BP, Hung MC. Dysregulation of cellular signaling by HER2/neu in breast cancer. Semin Oncol. 2003 Oct;30(5 Suppl 16):38-48.

47. Olayioye MA. Update on HER-2 as a target for cancer therapy: intracellular signaling pathways of ErbB2/HER-2 and family members. Breast Cancer Res. 2001;3(6):385-9.

48. Vinatzer U, Dampier B, Streubel B, Pacher M, Seewald MJ, Stratowa C, et al. Expression of HER2 and the coamplified genes GRB7 and MLN64 in human breast cancer: quantitative real-time reverse transcription-PCR as a diagnostic alternative to immunohistochemistry and fluorescence in situ hybridization. Clin Cancer Res. 2005 Dec 01;11(23):8348-57.

49. Le XF, Pruefer F, Bast RC, Jr. HER2-targeting antibodies modulate the cyclin-dependent kinase inhibitor p27Kip1 via multiple signaling pathways. Cell Cycle. 2005 Jan;4(1):87-95.

50. Perou CM, Sorlie T, Eisen MB, Rijn Mvande, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406.

51. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A. 2003 Sep 02;100(18):10393-8.

52. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.

Proceedings of the National Academy of Sciences. 2001 September 11, 2001;98(19):10869-74.

53. Hulka BS, Moorman PG. Breast cancer: hormones and other risk factors. Maturitas. 2001 Feb 28;38(1):103-13; discussion 13-6.

54. Dumitrescu RG, Cotarla I. Understanding breast cancer risk -- where do we stand in 2005? J Cell Mol Med. 2005 Jan-Mar;9(1):208-21.

55. Imran Ali WAW, Kishwar Saleem. Cancer Scenario in India with Future Perspectives. Cancer Therapy 2011;Vol 8:56-70.

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

57. Chopra B, Kaur V, Singh K. Age shift: Breast cancer is occurring in younger age groups - Is it true? Clinical Cancer Investigation Journal. 2014;3(6):526-9.

58. Nagrani RT, Budukh A, Koyande S, Panse NS, Mhatre SS, Badwe R. Rural urban differences in breast cancer in India. Indian J Cancer. 2014 July-September;51(3):277-81.

59. Surakasula A, Nagarjunapu GC, Raghavaiah KV. A comparative study of pre- and post-menopausal breast cancer: Risk factors, presentation, characteristics and management. Journal of Research in Pharmacy Practice. 2014 Jan-Mar;3(1):12-8.

60. Murthy NS, Agarwal UK, Chaudhry K, Saxena S. A study on time trends in incidence of breast cancer - Indian scenario. Eur J Cancer Care (Engl). 2007 Mar;16(2):185-6.

61. McPherson K, Steel CM, Dixon JM. Breast cancer—epidemiology, risk factors, and genetics. BMJ : British Medical Journal. 2000;321(7261):624-8.

62. Bernstein L. Epidemiology of endocrine-related risk factors for breast cancer. J Mammary Gland Biol Neoplasia. 2002 Jan;7(1):3-15.

63. Rao DN, Ganesh B, Desai PB. Role of reproductive factors in breast cancer in a low-risk area: a case-control study. Br J Cancer. 1994 Jul;70(1):129-32.

64. Gajalakshmi CK, Shanta V, Hakama M. Risk factors for contralateral breast cancer in Chennai (Madras), India. Int J Epidemiol. 1998 Oct;27(5):743-50.

65. Rao DN, Ganesh B, Desai PB. Role of reproductive factors in breast cancer in a low-risk area: a case-control study. British Journal of Cancer. 1994;70(1):129-32.

66. Paul Augustine RJ, Arun Peter, Anoop Amrith Lal, Jem Prabhakar, Jayadevan S, Jeesha C Haran. Risk Factors of Breast Cancer in Kerala, India - A Case Control Study. Academic Medical Journal of India. 2014;2(1):7–13.

67. Takalkar UV, Asegaonkar SB, Kulkarni U, Saraf M, Advani S. Clinicopathological Profile of Breast Cancer Patients at a Tertiary Care Hospital in Marathwada Region of Westen India. Asian Pac J Cancer Prev. 2016;17(4):2195-8.

68. Lipworth L, Bailey LR, Trichopoulos D. History of Breast-Feeding in Relation to Breast Cancer Risk: a Review of the Epidemiologic Literature. Journal of the National Cancer Institute. 2000 February 16, 2000;92(4):302-12.

69. Beral V, Banks E, Reeves G. Evidence from randomised trials on the long-term effects of hormone replacement therapy. The Lancet. [doi: 10.1016/S0140-6736(02)11032-4].360(9337):942-4.

70. Pakseresht S, Ingle G, Bahadur A, Ramteke V, Singh M, Garg S, et al. Risk factors with breast cancer among women in Delhi. Indian Journal of Cancer. [Original Article]. 2009 April 1, 2009;46(2):132-8.

71. Singletary KW, Frey RS, Yan W. Effect of ethanol on proliferation and estrogen receptor-α expression in human breast cancer cells. Cancer letters. 2001;165(2):131-7.

72. Longnecker MP. Alcoholic beverage consumption in relation to risk of breast cancer: meta-analysis and review. Cancer Causes & Control. 1994;5(1):73-82.

73. Lee H, Lee J, Gourley L, Duffy S, Day N, Estève J. Dietary effects on breast-cancer risk in Singapore. The Lancet. 1991;337(8751):1197-200.

74. Howe GR, Hirohata T, Hislop TG, Iscovich JM, Yuan J-M, Katsouyanni K, et al. Dietary factors and risk of breast cancer: combined analysis of 12 case—control studies. Journal of the National Cancer Institute. 1990;82(7):561-9.

75. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. The lancet oncology. 2001;2(3):133-40.

76. McPherson K, Steel C, Dixon J. Breast cancer—epidemiology, risk factors, and genetics. Bmj. 2000;321(7261):624-8.

77. Sinha R, Anderson D, McDonald S, Greenwald P. Cancer risk and diet in India. Journal of postgraduate medicine. 2003;49(3):222.

78. Vijayaraghavan K, Rao DH. Diet & nutrition situation in rural India. Indian journal of medical Research. 1998;108:243.

79. Krishnaswamy K, Polasa K. Diet, nutrition & cancer--the Indian scenario. The Indian journal of medical research. 1995;102:200-9.

80. McTiernan A, Rajan KB, Tworoger SS, Irwin M, Bernstein L, Baumgartner R, et al. Adiposity and sex hormones in postmenopausal breast cancer survivors. J Clin Oncol. 2003 May 15;21(10):1961-6.
81. Hankinson SE, Colditz GA, Willett WC. Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones. Breast Cancer Res. 2004;6(5):213-8.
82. Kaushal M, Mishra AK, Raju BS, Ihsan R, Chakraborty A, Sharma J, et al. Betel quid chewing as an environmental risk factor for breast cancer. Mutat Res. 2010 Dec 21;703(2):143-8.
83. Kaushal M, Mishra AK, Sharma J, Zomawia E, Kataki A, Kapur S, et al. Genomic alterations in breast cancer patients in betel quid and non betel quid chewers. PLoS One. 2012;7(8):e43789.
84. Rajbongshi N, Mahanta LB, Nath DC, Sarma JD. A matched case control study of risk indicators of breast cancer in assam, India. Mymensingh Med J. 2015 Apr;24(2):385-91.
85. Paul S, Solanki PP, Shahi UP, Srikrishna S. Epidemiological Study on Breast Cancer Associated Risk Factors and Screening Practices among Women in the Holy City of Varanasi, Uttar Pradesh, India. Asian Pac J Cancer Prev. 2015;16(18):8163-71.
86. Rajbongshi N, Mahanta LB, Nath DC. Evaluation of Female Breast Cancer Risk Among the Betel Quid Chewer: A Bio-Statistical Assessment in Assam, India. Nepal J Epidemiol. 2015 Jun;5(2):494-8.
87. Thalib L, Wedren S, Granath F, Adami HO, Rydh B, Magnusson C, et al. Breast cancer prognosis in relation to family history of breast and ovarian cancer. Br J Cancer. 2004 Apr 5;90(7):1378-81.
88. Saxena S, Chakraborty A, Kaushal M, Kotwal S, Bhatanager D, Mohil RS, et al. Contribution of germline BRCA1 and BRCA2 sequence alterations to breast cancer in Northern India. BMC Med Genet. 2006;7:75.
89. Vaidyanathan K, Lakhotia S, Ravishankar HM, Tabassum U, Mukherjee G, Somasundaram K. BRCA1 and BRCA2 germline mutation analysis among Indian women from south India: identification of four novel mutations and high-frequency occurrence of 185delAG mutation. J Biosci. 2009 Sep;34(3):415-22.
90. Kim H, Choi DH. Distribution of BRCA1 and BRCA2 Mutations in Asian Patients with Breast Cancer. J Breast Cancer. 2013 Dec;16(4):357-65.

91. Saxena S, Szabo CI, Chopin S, Barjhoux L, Sinilnikova O, Lenoir G, et al. BRCA1 and BRCA2 in Indian breast cancer patients. Hum Mutat. 2002 Dec;20(6):473-4.

92. Valarmathi MT, A A, Deo SS, Shukla NK, Das SN. BRCA1 germline mutations in Indian familial breast cancer. Hum Mutat. 2003 Jan;21(1):98-9.

93. Valarmathi MT, Sawhney M, Deo SS, Shukla NK, Das SN. Novel germline mutations in the BRCA1 and BRCA2 genes in Indian breast and breast-ovarian cancer families. Hum Mutat. 2004 Feb;23(2):205.

94. Hedau S, Jain N, Husain SA, Mandal AK, Ray G, Shahid M, et al. Novel germline mutations in breast cancer susceptibility genes BRCA1, BRCA2 and p53 gene in breast cancer patients from India. Breast Cancer Res Treat. 2004 Nov;88(2):177-86.

95. Claus EB, Stowe M, Carter D. Breast carcinoma in situ: risk factors and screening patterns. J Natl Cancer Inst. 2001 Dec 5;93(23):1811-7.

96. Kabat GC, Kim MY, Woods NF, Habel LA, Messina CR, Wactawski-Wende J, et al. Reproductive and menstrual factors and risk of ductal carcinoma in situ of the breast in a cohort of postmenopausal women. Cancer Causes Control. 2011 Oct;22(10):1415-24.

97. Kerlikowske K, Barclay J, Grady D, Sickles EA, Ernster V. Comparison of risk factors for ductal carcinoma in situ and invasive breast cancer. J Natl Cancer Inst. 1997 Jan 1;89(1):76-82.

98. Bhardwaj N. Harnessing the immune system to treat cancer. J Clin Invest. 2007 May;117(5):1130-6.

99. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol. 2002 Nov;3(11):991-8.

100. Baar J. Biological therapy of breast cancer: recent clinical applications. Curr Opin Investig Drugs. 2007 Dec;8(12):987-95.

101. Morabito A, Magnani E, Gion M, Sarmiento R, Capaccetti B, Longo R, et al. Prognostic and predictive indicators in operable breast cancer. Clin Breast Cancer. 2003 Feb;3(6):381-90.

102. Pavelic K, Gall-Troselj K. Recent advances in molecular genetics of breast cancer. J Mol Med (Berl). 2001 Oct;79(10):566-73.

103. Beckmann MW, Niederacher D, Schnurch HG, Gusterson BA, Bender HG. Multistep carcinogenesis of breast cancer and tumour heterogeneity. J Mol Med (Berl). 1997 Jun;75(6):429-39.

104. Cuny M, Kramar A, Courjal F, Johannsdottir V, Iacopetta B, Fontaine H, et al. Relating genotype and phenotype in breast cancer: an analysis of the prognostic significance of amplification at eight different genes or loci and of p53 mutations. Cancer Res. 2000 Feb 15;60(4):1077-83.

105. Zaridze DG. Molecular epidemiology of cancer. Biochemistry (Mosc). 2008 May;73(5):532-42.

106. Yarbro JW. Oncogenes and cancer suppressor genes. Semin Oncol Nurs. 1992 Feb;8(1):30-9.

107. Osborne C, Wilson P, Tripathy D. Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. Oncologist. 2004;9(4):361-77.

108. Haber D, Harlow E. Tumour-suppressor genes: evolving definitions in the genomic age. Nat Genet. 1997 Aug;16(4):320-2.

109. Motoyama N, Naka K. DNA damage tumor suppressor genes and genomic instability. Curr Opin Genet Dev. 2004 Feb;14(1):11-6.

110. Sherr CJ. Principles of tumor suppression. Cell. 2004 Jan 23;116(2):235-46.

111. Miller MB, Tang Y-W. Basic Concepts of Microarrays and Potential Applications in Clinical Microbiology. Clinical Microbiology Reviews. 2009 October 1, 2009;22(4):611-33.

112. DeRisi J. Overview of nucleic acid arrays. Curr Protoc Neurosci. 2001 Nov;Chapter 4:Unit 4 25.

113. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, et al. Breast Cancer Molecular Subtypes Respond Differently to Preoperative Chemotherapy. Clinical Cancer Research. 2005;11(16):5678-85.

114. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002 Jan 31;415(6871):530-6.

115. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Bernards R, et al. Expression profiling predicts outcome in breast cancer. Breast Cancer Research. 2003;5(1):57-8.

116. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002 Dec 19;347(25):1999-2009.

117. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst. 2006 Sep 6;98(17):1183-92.
118. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003 Jul 08;100(14):8418-23.
119. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, et al. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A. 2003 May 13;100(10):5974-9.
120. Zhao H, Langerod A, Ji Y, Nowels KW, Nesland JM, Tibshirani R, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell. 2004 Jun;15(6):2523-36.
121. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci U S A. 2001 Sep 25;98(20):11462-7.
122. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, et al. Gene expression predictors of breast cancer outcomes. Lancet. 2003 May 10;361(9369):1590-6.
123. Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell. 2004 Jun;5(6):607-16.
124. Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A. 2005 Mar 08;102(10):3738-43.
125. Dai H, van't Veer L, Lamb J, He YD, Mao M, Fine BM, et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. Cancer Res. 2005 May 15;65(10):4059-66.
126. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, Larsimont D, et al. Identification of molecular apocrine breast tumours by microarray analysis. Oncogene. 2005 Jul 07;24(29):4660-71.
127. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A. 2005 Sep 20;102(38):13550-5.

128. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene. 2007 Mar 01;26(10):1507-16.
129. Brinton LA, Sherman ME, Carreon JD, Anderson WF. Recent trends in breast cancer among younger women in the United States. J Natl Cancer Inst. 2008 Nov 19;100(22):1643-8.
130. Akarolo-Anthony SN, Ogundiran TO, Adebamowo CA. Emerging breast cancer epidemic: evidence from Africa. Breast Cancer Res. 2010 Dec 20;12 Suppl 4:S8.
131. El Saghir NS, Khalil MK, Eid T, El Kinge AR, Charafeddine M, Geara F, et al. Trends in epidemiology and management of breast cancer in developing Arab countries: a literature and registry analysis. Int J Surg. 2007 Aug;5(4):225-33.
132. Cancello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Prognosis and adjuvant treatment effects in selected breast cancer subtypes of very young women (<35 years) with operable breast cancer. Ann Oncol. 2010 Oct;21(10):1974-81.
133. Keegan TH, DeRouen MC, Press DJ, Kurian AW, Clarke CA. Occurrence of breast cancer subtypes in adolescent and young adult women. Breast Cancer Res. 2012 Mar 27;14(2):R55.
134. Huzarski T, Byrski T, Gronwald J, Gorski B, Domagala P, Cybulski C, et al. Ten-year survival in patients with BRCA1-negative and BRCA1-positive breast cancer. J Clin Oncol. 2013 Sep 10;31(26):3191-6.
135. Young SR, Pilarski RT, Donenberg T, Shapiro C, Hammond LS, Miller J, et al. The prevalence of BRCA1 mutations among young women with triple-negative breast cancer. BMC Cancer. 2009;9:86.
136. Criscitiello C, Azim HA, Jr., Schouten PC, Linn SC, Sotiriou C. Understanding the biology of triple-negative breast cancer. Ann Oncol. 2012 Aug;23 Suppl 6:vi13-8.
137. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. 2009 Aug;15(8):907-13.
138. Asselin-Labat ML, Vaillant F, Sheridan JM, Pal B, Wu D, Simpson ER, et al. Control of mammary stem cell function by steroid hormone signalling. Nature. 2010 Jun 10;465(7299):798-802.
139. Gonzalez-Suarez E, Jacob AP, Jones J, Miller R, Roudier-Meyer MP, Erwert R, et al. RANK ligand mediates progestin-induced mammary epithelial proliferation and carcinogenesis. Nature. 2010 Nov 4;468(7320):103-7.

140. Weigelt B, Glas AM, Wessels LF, Witteveen AT, Peterse JL, van't Veer LJ. Gene expression profiles of primary breast tumors maintained in distant metastases. Proc Natl Acad Sci U S A. 2003 Dec 23;100(26):15901-5.

141. Karlsson E, Delle U, Danielsson A, Olsson B, Abel F, Karlsson P, et al. Gene expression variation to predict 10-year survival in lymph-node-negative breast cancer. BMC Cancer. [journal article]. 2008;8(1):254.

142. Chouchane L, Boussen H, Sastry KS. Breast cancer in Arab populations: molecular characteristics and disease management implications. Lancet Oncol. 2013 Sep;14(10):e417-24.

143. Fredholm H, Eaker S, Frisell J, Holmberg L, Fredriksson I, Lindman H. Breast cancer in young women: poor survival despite intensive treatment. PLoS One. 2009;4(11):e7695.

144. Copson E, Eccles B, Maishman T, Gerty S, Stanton L, Cutress RI, et al. Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study. J Natl Cancer Inst. 2013 Jul 3;105(13):978-88.

145. Zhu JK. Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet. 2009;43:143-66.

146. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. J Mol Biol. 1987 Jul 20;196(2):261-82.

147. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A. 2006 Jan 31;103(5):1412-7.

148. Vavouri T, Lehner B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. Genome Biol. 2012 Nov 27;13(11):R110.

149. Donnelly SR, Hawkins TE, Moss SE. A conserved nuclear element with a role in mammalian gene regulation. Hum Mol Genet. 1999 Sep;8(9):1723-8.

150. Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell. 1992 Nov 27;71(5):865-73.

151. Gopalakrishnan S, Sullivan BA, Trazzi S, Della Valle G, Robertson KD. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum Mol Genet. 2009 Sep 01;18(17):3178-93.

152. Feinberg AP. The epigenetics of cancer etiology. Semin Cancer Biol. 2004 Dec;14(6):427-32.

153. Arney KL, Fisher AG. Epigenetic aspects of differentiation. J Cell Sci. 2004 Sep 01;117(Pt 19):4355-63.

154. Clouaire T, Stancheva I. Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? Cell Mol Life Sci. 2008 May;65(10):1509-22.

155. Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. Biochim Biophys Acta. 2007 Jan;1775(1):138-62.

156. Watt PM, Kumar R, Kees UR. Promoter demethylation accompanies reactivation of the HOX11 proto-oncogene in leukemia. Genes, Chromosomes and Cancer. 2000;29(4):371-7.

157. Baylin SB, Hoppener JW, de Bustros A, Steenbergh PH, Lips CJ, Nelkin BD. DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. Cancer Res. 1986 Jun;46(6):2917-22.

158. Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A. 1994 Oct 11;91(21):9700-4.

159. Fleuriel C, Touka M, Boulay G, Guerardel C, Rood BR, Leprince D. HIC1 (Hypermethylated in Cancer 1) epigenetic silencing in tumors. Int J Biochem Cell Biol. 2009 Jan;41(1):26-33.

160. Esteller M. Epigenetics in cancer. N Engl J Med. 2008 Mar 13;358(11):1148-59.

161. Jovanovic J, Rønneberg JA, Tost J, Kristensen V. The epigenetics of breast cancer. Molecular Oncology. 2010;4(3):242-54.

162. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007 Feb 23;128(4):683-92.

163. Hinshelwood RA, Clark SJ. Breast cancer epigenetics: normal human mammary epithelial cells as a model system. J Mol Med (Berl). 2008 Dec;86(12):1315-28.

164. Singh P, Yang M, Dai H, Yu D, Huang Q, Tan W, et al. Overexpression and hypomethylation of flap endonuclease 1 gene in breast and other cancers. Mol Cancer Res. 2008 Nov;6(11):1710-7.

165. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010 Jan;31(1):27-36.

166. Berger R, Fiegl H, Goebel G, Obexer P, Ausserlechner M, Doppler W, et al. Toll-like receptor 9 expression in breast and ovarian cancer is associated with poorly differentiated tumors. Cancer Sci. 2010 Apr;101(4):1059-66.

167. Jin Z, Tamura G, Tsuchiya T, Sakata K, Kashiwaba M, Osakabe M, et al. Adenomatous polyposis coli (APC) gene promoter hypermethylation in primary breast cancers. Br J Cancer. 2001 Jul 06;85(1):69-73.

168. Muggerud AA, Ronneberg JA, Warnberg F, Botling J, Busato F, Jovanovic J, et al. Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. Breast Cancer Res. 2010;12(1):R3.

169. Dejeux E, Ronneberg JA, Solvang H, Bukholm I, Geisler S, Aas T, et al. DNA methylation profiling in doxorubicin treated primary locally advanced breast tumours identifies novel genes associated with survival and treatment response. Mol Cancer. 2010 Mar 25;9:68.

170. Huang KT, Dobrovic A, Yan M, Karim RZ, Lee CS, Lakhani SR, et al. DNA methylation profiling of phyllodes and fibroadenoma tumours of the breast. Breast Cancer Res Treat. 2010 Nov;124(2):555-65.

171. El Saghir NS, Seoud M, Khalil MK, Charafeddine M, Salem ZK, Geara FB, et al. Effects of young age at presentation on survival in breast cancer. BMC Cancer. [journal article]. 2006;6(1):194.

172. Adami HO, Malker B, Holmberg L, Persson I, Stone B. The relation between survival and age at diagnosis in breast cancer. N Engl J Med. 1986 Aug 28;315(9):559-63.

173. Nixon AJ, Neuberg D, Hayes DF, Gelman R, Connolly JL, Schnitt S, et al. Relationship of patient age to pathologic features of the tumor and prognosis for patients with stage I or II breast cancer. J Clin Oncol. 1994 May;12(5):888-94.

174. Holli K, Isola J. Effect of age on the survival of breast cancer patients. Eur J Cancer. 1997 Mar;33(3):425-8.

175. Kollias J, Elston CW, Ellis IO, Robertson JF, Blamey RW. Early-onset breast cancer--histopathological and prognostic considerations. Br J Cancer. 1997;75(9):1318-23.

176. Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, O'Driscoll L, et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. Carcinogenesis. 2013 Oct;34(10):2300-8.

177. Maubant S, Tesson B, Maire V, Ye M, Rigaill G, Gentien D, et al. Transcriptome analysis of Wnt3a-treated triple-negative breast cancer cells. PLoS One. 2015;10(4):e0122333.

178. Shin HR, Joubert C, Boniol M, Hery C, Ahn SH, Won YJ, et al. Recent trends and patterns in breast cancer incidence among Eastern and Southeastern Asian women. Cancer Causes Control. 2010 Nov;21(11):1777-85.

179. Wang N, Zhu WX, Xing XM, Yang L, Li PP, You WC. Time trends of cancer incidence in urban beijing, 1998-2007. Chin J Cancer Res. 2011 Mar;23(1):15-20.

180. Jung KW, Park S, Kong HJ, Won YJ, Lee JY, Park EC, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2008. Cancer Res Treat. 2011 Mar;43(1):1-11.

181. Toi M, Ohashi Y, Seow A, Moriya T, Tse G, Sasano H, et al. The Breast Cancer Working Group Presentation was Divided into Three Sections: The Epidemiology, Pathology and Treatment of Breast Cancer. Japanese Journal of Clinical Oncology. [doi: 10.1093/jjco/hyq122]. 2010;40(suppl_1):i13-i8.

182. Wang Y, Yin Q, Yu Q, Zhang J, Liu Z, Wang S, et al. A retrospective study of breast cancer subtypes: the risk of relapse and the relations with treatments. Breast Cancer Res Treat. 2011 Nov;130(2):489-98.

183. Kim MJ, Ro JY, Ahn SH, Kim HH, Kim SB, Gong G. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. Hum Pathol. 2006 Sep;37(9):1217-26.

184. Kurebayashi J, Moriya T, Ishida T, Hirakawa H, Kurosumi M, Akiyama F, et al. The prevalence of intrinsic subtypes and prognosis in breast cancer patients of different races. Breast. 2007 Dec;16 Suppl 2:S72-7.

185. Ambroise M, Ghosh M, Mallikarjuna VS, Kurian A. Immunohistochemical profile of breast cancer patients at a tertiary care hospital in South India. Asian Pac J Cancer Prev. 2011;12(3):625-9.

186. Krishnamurthy S, Poornima R, Challa VR, Goud YG. Triple negative breast cancer - our experience and review. Indian J Surg Oncol. 2012 Mar;3(1):12-6.

187. Gogia A, Raina V, Deo SV, Shukla NK, Mohanti BK. Triple-negative breast cancer: An institutional analysis. Indian J Cancer. 2014 Apr-Jun;51(2):163-6.

188. Rao C, Shetty J, Prasad KH. Immunohistochemical profile and morphology in triple - negative breast cancers. J Clin Diagn Res. 2013 Jul;7(7):1361-5.

189. Sharma M, Sharma JD, Sarma A, Ahmed S, Kataki AC, Saxena R, et al. Triple negative breast cancer in people of North East India: critical insights gained at a regional cancer centre. Asian Pac J Cancer Prev. 2014;15(11):4507-11.

190. Verma S, Bal A, Joshi K, Arora S, Singh G. Immunohistochemical characterization of molecular subtypes of invasive breast cancer: a study from North India. APMIS. 2012 Dec;120(12):1008-19.

191. Sharma B, Kalwar A, Sharma N, Kapoor A, Kumar N. Five year retrospective survival analysis of triple negative breast cancer in North-West India. Indian J Cancer. 2013 Oct-Dec;50(4):330-2.

192. Colleoni M, Rotmensz N, Robertson C, Orlando L, Viale G, Renne G, et al. Very young women (<35 years) with operable breast cancer: features of disease at presentation. Ann Oncol. 2002 Feb;13(2):273-9.

193. Gnerlich JL, Deshpande AD, Jeffe DB, Sweet A, White N, Margenthaler JA. Elevated breast cancer mortality in women younger than age 40 years compared with older women is attributed to poorer survival in early-stage disease. J Am Coll Surg. 2009 Mar;208(3):341-7.

194. Collins LC, Marotti JD, Gelber S, Cole K, Ruddy K, Kereakoglow S, et al. Pathologic features and molecular phenotype by patient age in a large cohort of young women with breast cancer. Breast Cancer Res Treat. 2012 Feb;131(3):1061-6.

195. Shinde SS, Forman MR, Kuerer HM, Yan K, Peintinger F, Hunt KK, et al. Higher parity and shorter breastfeeding duration: association with triple-negative phenotype of breast cancer. Cancer. 2010 Nov 01;116(21):4933-43.

196. Redondo CM, Gago-Dominguez M, Ponte SM, Castelo ME, Jiang X, Garcia AA, et al. Breast feeding, parity and breast cancer subtypes in a Spanish cohort. PLoS One. 2012;7(7):e40543.

197. Li CI, Beaber EF, Tang MT, Porter PL, Daling JR, Malone KE. Reproductive factors and risk of estrogen receptor positive, triple-negative, and HER2-neu overexpressing breast cancer among women 20-44 years of age. Breast Cancer Res Treat. 2013 Jan;137(2):579-87.

198. Kotsopoulos J, Lubinski J, Salmena L, Lynch HT, Kim-Sing C, Foulkes WD, et al. Breastfeeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. Breast Cancer Res. 2012 Mar 09;14(2):R42.

199. Azim HA, Jr., Partridge AH. Biology of breast cancer in young women. Breast Cancer Res. 2014 Aug 27;16(4):427.

200. Zheng Q, Zhou G, Morello R, Chen Y, Garcia-Rojas X, Lee B. Type X collagen gene regulation by Runx2 contributes directly to its hypertrophic chondrocyte-specific expression in vivo. J Cell Biol. 2003 Sep 1;162(5):833-42.

201. Yoneda K, Iida H, Endo H, Hosono K, Akiyama T, Takahashi H, et al. Identification of Cystatin SN as a novel tumor marker for colorectal cancer. Int J Oncol. 2009 Jul;35(1):33-40.

202. Choi EH, Kim JT, Kim JH, Kim SY, Song EY, Kim JW, et al. Upregulation of the cysteine protease inhibitor, cystatin SN, contributes to cell proliferation and cathepsin inhibition in gastric cancer. Clin Chim Acta. 2009 Aug;406(1-2):45-51.

203. Kim JT, Lee SJ, Kang MA, Park JE, Kim BY, Yoon DY, et al. Cystatin SN neutralizes the inhibitory effect of cystatin C on cathepsin B activity. Cell Death Dis. 2013 Dec 19;4:e974.

204. Wang Z, Katsaros D, Shen Y, Fu Y, Canuto EM, Benedetto C, et al. Biological and Clinical Significance of MAD2L1 and BUB1, Genes Frequently Appearing in Expression Signatures for Breast Cancer Prognosis. PLoS One. 2015;10(8):e0136246.

205. Yan M, Song M, Bai R, Cheng S, Yan W. Identification of potential therapeutic targets for colorectal cancer by bioinformatics analysis. Oncol Lett. 2016 Dec;12(6):5092-8.

206. Yan H, Li Z, Shen Q, Wang Q, Tian J, Jiang Q, et al. Aberrant expression of cell cycle and material metabolism related genes contributes to hepatocellular carcinoma occurrence. Pathol Res Pract. 2017 Jan 25.

207. Tong H, Wang J, Chen H, Wang Z, Fan H, Ni Z. Transcriptomic analysis of gene expression profiles of stomach carcinoma reveal abnormal expression of mitotic components. Life Sci. 2017 Feb 01;170:41-9.

208. Zhang C, Min L, Zhang L, Ma Y, Yang Y, Shou C. Combined analysis identifies six genes correlated with augmented malignancy from non-small cell to small cell lung cancer. Tumour Biol. 2016 Feb;37(2):2193-207.

209. Ocana A, Perez-Pena J, Diez-Gonzalez L, Sanchez-Corrales V, Templeton A, Seruga B, et al. Transcriptomic analyses identify association between mitotic kinases, PDZ-binding kinase and BUB1, and clinical outcome in breast cancer. Breast Cancer Res Treat. 2016 Feb;156(1):1-8.

210. Bednarek K, Kiwerska K, Szaumkessel M, Bodnar M, Kostrzewska-Poczekaj M, Marszalek A, et al. Recurrent CDK1 overexpression in laryngeal squamous cell carcinoma. Tumour Biol. 2016 Aug;37(8):11115-26.

211. Shubbar E, Kovács A, Hajizadeh S, Parris TZ, Nemes S, Gunnarsdóttir K, et al. Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. BMC Cancer. [journal article]. 2013;13(1):1.

212. Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. CCNA2 Is a Prognostic Biomarker for ER+ Breast Cancer and Tamoxifen Resistance. PLoS One. [doi:10.1371/journal.pone.0091771]. 2014;9(3):e91771.

213. Chang HJ, Yang MJ, Yang YH, Hou MF, Hsueh EJ, Lin SR. MMP13 is potentially a new tumor marker for breast cancer diagnosis. Oncol Rep. 2009 Nov;22(5):1119-27.

214. Karaglani M, Toumpoulis I, Goutas N, Poumpouridou N, Vlachodimitropoulos D, Vasilaros S, et al. Development of novel real-time PCR methodology for quantification of COL11A1 mRNA variants and evaluation in breast cancer tissue specimens. BMC Cancer. 2015 Oct 14;15:694.

215. Chapman KB, Prendes MJ, Sternberg H, Kidd JL, Funk WD, Wagner J, et al. COL10A1 expression is elevated in diverse solid tumor types and is associated with tumor vasculature. Future Oncol. 2012 Aug;8(8):1031-40.

216. Halsted KC, Bowen KB, Bond L, Luman SE, Jorcyk CL, Fyffe WE, et al. Collagen alpha1(XI) in normal and malignant breast tissue. Mod Pathol. 2008 Oct;21(10):1246-54.

217. Makoukji J, Makhoul NJ, Khalil M, El-Sitt S, Aldin ES, Jabbour M, et al. Gene expression profiling of breast cancer in Lebanese women. Sci Rep. 2016 Nov 18;6:36639.

218. Galvan JA, Garcia-Martinez J, Vazquez-Villa F, Garcia-Ocana M, Garcia-Pravia C, Menendez-Rodriguez P, et al. Validation of COL11A1/procollagen 11A1 expression in TGF-beta1-activated immortalised human mesenchymal cells and in stromal cells of human colon adenocarcinoma. BMC Cancer. 2014 Nov 23;14:867.

219. Wu X, Zhang W, Hu Y, Yi X. Bioinformatics approach reveals systematic mechanism underlying lung adenocarcinoma. Tumori. 2015 May-Jun;101(3):281-6.

220. Ibanez de Caceres I, Dulaimi E, Hoffman AM, Al-Saleem T, Uzzo RG, Cairns P. Identification of novel target genes by an epigenetic reactivation screen of renal cancer. Cancer Res. 2006 May 15;66(10):5021-8.

221. Hayashi M, Nomoto S, Hishida M, Inokawa Y, Kanda M, Okamura Y, et al. Identification of the collagen type 1 alpha 1 gene (COL1A1) as a candidate survival-related factor associated with hepatocellular carcinoma. BMC Cancer. 2014 Feb 19;14:108.

222. Bonazzi VF, Nancarrow DJ, Stark MS, Moser RJ, Boyle GM, Aoude LG, et al. Cross-platform array screening identifies COL1A2, THBS1, TNFRSF10D and UCHL1 as genes frequently silenced by methylation in melanoma. PLoS One. 2011;6(10):e26121.

223. Chai F, Liang Y, Zhang F, Wang M, Zhong L, Jiang J. Systematically identify key genes in inflammatory and non-inflammatory breast cancer. Gene. 2016 Jan 10;575(2 Pt 3):600-14.
224. Xiong G, Deng L, Zhu J, Rychahou PG, Xu R. Prolyl-4-hydroxylase alpha subunit 2 promotes breast cancer progression and metastasis by regulating collagen deposition. BMC Cancer. 2014 Jan 02;14:1.
225. Zekri AN, El-Sisi ER, Abdallah ZF, El-Din Ismail Abd El-Motaleb A, Barakat Barakat A. Gene expression profiling of circulating CD133+ cells of hepatocellular carcinoma patients associated with HCV infection. J Egypt Natl Canc Inst. 2017 Feb 28.
226. Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. World J Surg Oncol. 2016 Nov 29;14(1):297.
227. Delassus GS, Cho H, Eliceiri GL. New signaling pathways from cancer progression modulators to mRNA expression of matrix metalloproteinases in breast cancer cells. J Cell Physiol. 2011 Dec;226(12):3378-84.
228. Decock J, Hendrickx W, Drijkoningen M, Wildiers H, Neven P, Smeets A, et al. Matrix metalloproteinase expression patterns in luminal A type breast carcinomas. Dis Markers. 2007;23(3):189-96.
229. Wu K, Fukuda K, Xing F, Zhang Y, Sharma S, Liu Y, et al. Roles of the cyclooxygenase 2 matrix metalloproteinase 1 pathway in brain metastasis of breast cancer. J Biol Chem. 2015 Apr 10;290(15):9842-54.
230. Zhang X, Huang S, Guo J, Zhou L, You L, Zhang T, et al. Insights into the distinct roles of MMP-11 in tumor biology and future therapeutics (Review). Int J Oncol. 2016 May;48(5):1783-93.
231. Lee JY, Park K, Lee E, Ahn T, Jung HH, Lim SH, et al. Gene Expression Profiling of Breast Cancer Brain Metastasis. Sci Rep. 2016 Jun 24;6:28623.
232. Osako Y, Seki N, Kita Y, Yonemori K, Koshizuka K, Kurozumi A, et al. Regulation of MMP13 by antitumor microRNA-375 markedly inhibits cancer cell migration and invasion in esophageal squamous cell carcinoma. Int J Oncol. 2016 Dec;49(6):2255-64.
233. Celardo I, Antonov A, Amelio I, Annicchiarico-Petruzzelli M, Melino G. p63 transcriptionally regulates the expression of matrix metallopeptidase 13. Oncotarget. 2014 Mar 15;5(5):1279-89.

234. Pal SK, Nguyen CT, Morita KI, Miki Y, Kayamori K, Yamaguchi A, et al. THBS1 is induced by TGFB1 in the cancer stroma and promotes invasion of oral squamous cell carcinoma. J Oral Pathol Med. 2016 Nov;45(10):730-9.

235. Ou B, Zhao J, Guan S, Feng H, Wangpu X, Zhu C, et al. CCR4 promotes metastasis via ERK/NF-kappaB/MMP13 pathway and acts downstream of TNF-alpha in colorectal cancer. Oncotarget. 2016 Jul 26;7(30):47637-49.

236. Sedighi M, Aledavood SA, Abbaszadegan M, Memar B, Montazer M, Rajabian M, et al. Matrix Metalloproteinase-13 - A Potential Biomarker for Detection and Prognostic Assessment of Patients with Esophageal Squamous Cell Carcinoma. Asian Pac J Cancer Prev. 2016;17(6):2781-5.

237. Kwok HF, Zhang SD, McCrudden CM, Yuen HF, Ting KP, Wen Q, et al. Prognostic significance of minichromosome maintenance proteins in breast cancer. Am J Cancer Res. 2015;5(1):52-71.

238. Nieto-Jimenez C, Alcaraz-Sanabria A, Paez R, Perez-Pena J, Corrales-Sanchez V, Pandiella A, et al. DNA-damage related genes and clinical outcome in hormone receptor positive breast cancer. Oncotarget. 2016 Jul 28.

239. Yousef EM, Furrer D, Laperriere DL, Tahir MR, Mader S, Diorio C, et al. MCM2: An alternative to Ki-67 for measuring breast cancer cell proliferation. Mod Pathol. 2017 Jan 13.

240. Abe S, Yamamoto K, Kurata M, Abe-Suzuki S, Horii R, Akiyama F, et al. Targeting MCM2 function as a novel strategy for the treatment of highly malignant breast tumors. Oncotarget. 2015 Oct 27;6(33):34892-909.

241. Wu C, Zhu J, Zhang X. Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes. BMC Bioinformatics. 2012 Jul 28;13:182.

242. Chen J, Yang R, Zhang W, Wang Y. Candidate pathways and genes for nasopharyngeal carcinoma based on bioinformatics study. Int J Clin Exp Pathol. 2015;8(2):2026-32.

243. Zhai DK, Liu B, Bai XF, Wen JA. Identification of biomarkers and pathway-related modules involved in ovarian cancer based on topological centralities. J BUON. 2016 Jan-Feb;21(1):208-20.

244. Zekri AR, Hassan ZK, Bahnassy AA, Khaled HM, El-Rouby MN, Haggag RM, et al. Differentially expressed genes in metastatic advanced Egyptian bladder cancer. Asian Pac J Cancer Prev. 2015;16(8):3543-9.

245. Liu D, Zhang XX, Xi BX, Wan DY, Li L, Zhou J, et al. Sine oculis homeobox homolog 1 promotes DNA replication and cell proliferation in cervical cancer. Int J Oncol. 2014 Sep;45(3):1232-40.
246. Chubb D, Broderick P, Dobbins SE, Frampton M, Kinnersley B, Penegar S, et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nat Commun. 2016 Jun 22;7:11883.
247. Zhou C, Wang M, Zhou L, Zhang Y, Liu W, Qin W, et al. Prognostic significance of PLIN1 expression in human breast cancer. Oncotarget. 2016 Aug 23;7(34):54488-502.
248. Saha P, Heagerty PJ, Zheng Y: Introduction to risksetROC: An R Package for Riskset-ROC (Incident/Dynamic) Calculation. [http://faculty.washington.edu/heagerty/Software/SurvROC/].
249. Merdad A, Karim S, Schulten HJ, Jayapal M, Dallol A, Buhmeida A, et al. Transcriptomics profiling study of breast cancer from Kingdom of Saudi Arabia revealed altered expression of Adiponectin and Fatty Acid Binding Protein4: Is lipid metabolism associated with breast cancer? BMC Genomics. 2015;16 Suppl 1:S11.
250. Kim S, Lee Y, Koo JS. Differential expression of lipid metabolism-related proteins in different breast cancer subtypes. PLoS One. 2015;10(3):e0119473.
251. Fry SA, Robertson CE, Swann R, Dwek MV. Cadherin-5: a biomarker for metastatic breast cancer with optimum efficacy in oestrogen receptor-positive breast cancers with vascular invasion. Br J Cancer. 2016 Apr 26;114(9):1019-26.
252. Craig DW, O'Shaughnessy JA, Kiefer JA, Aldrich J, Sinari S, Moses TM, et al. Genome and transcriptome sequencing in prospective metastatic triple-negative breast cancer uncovers therapeutic vulnerabilities. Mol Cancer Ther. 2013 Jan;12(1):104-16.
253. Higuchi K, Inokuchi M, Takagi Y, Ishikawa T, Otsuki S, Uetake H, et al. Cadherin 5 expression correlates with poor survival in human gastric cancer. J Clin Pathol. 2017 Mar;70(3):217-21.
254. Singh V, Singh LC, Vasudevan M, Chattopadhyay I, Borthakar BB, Rai AK, et al. Esophageal Cancer Epigenomics and Integrome Analysis of Genome-Wide Methylation and Expression in High Risk Northeast Indian Population. OMICS. 2015 Nov;19(11):688-99.
255. Nissinen L, Siljamaki E, Riihila P, Piipponen M, Farshchian M, Kivisaari A, et al. Expression of claudin-11 by tumor cells in cutaneous squamous cell carcinoma is dependent on the activity of p38delta. Exp Dermatol. 2016 Dec 19.

256. Abe M, Yamashita S, Mori Y, Abe T, Saijo H, Hoshi K, et al. High-risk oral leukoplakia is associated with aberrant promoter methylation of multiple genes. BMC Cancer. 2016 Jun 03;16:350.
257. Hensley PJ, Desiniotis A, Wang C, Stromberg A, Chen CS, Kyprianou N. Novel pharmacologic targeting of tight junctions and focal adhesions in prostate cancer cells. PLoS One. 2014;9(1):e86238.
258. Lin Z, Zhang X, Liu Z, Liu Q, Wang L, Lu Y, et al. The distinct expression patterns of claudin-2, -6, and -11 between human gastric neoplasms and adjacent non-neoplastic tissues. Diagn Pathol. 2013 Aug 06;8:133.
259. Meng L, Xu Y, Xu C, Zhang W. Biomarker discovery to improve prediction of breast cancer survival: using gene expression profiling, meta-analysis, and tissue validation. Onco Targets Ther. 2016;9:6177-85.
260. Huang YS, Chang CC, Lee SS, Jou YS, Shih HM. Xist reduction in breast cancer upregulates AKT phosphorylation via HDAC3-mediated repression of PHLPP1 expression. Oncotarget. 2016 Jul 12;7(28):43256-66.
261. Reeves HL, Narla G, Ogunbiyi O, Haq AI, Katz A, Benzeno S, et al. Kruppel-like factor 6 (KLF6) is a tumor-suppressor gene frequently inactivated in colorectal cancer. Gastroenterology. 2004 Apr;126(4):1090-103.
262. Redon R, Hussenet T, Bour G, Caulee K, Jost B, Muller D, et al. Amplicon mapping and transcriptional analysis pinpoint cyclin L as a candidate oncogene in head and neck cancer. Cancer Res. 2002 Nov 1;62(21):6211-7.
263. Sticht C, Hofele C, Flechtenmacher C, Bosch FX, Freier K, Lichter P, et al. Amplification of Cyclin L1 is associated with lymph node metastases in head and neck squamous cell carcinoma (HNSCC). Br J Cancer. 2005 Feb 28;92(4):770-4.
264. Li B, Lin H, Fan J, Lan J, Zhong Y, Yang Y, et al. CD59 is overexpressed in human lung cancer and regulates apoptosis of human lung cancer cells. Int J Oncol. 2013 Sep;43(3):850-8.
265. Chida S, Okayama H, Noda M, Saito K, Nakajima T, Aoto K, et al. Stromal VCAN expression as a potential prognostic biomarker for disease recurrence in stage II-III colon cancer. Carcinogenesis. 2016 Sep;37(9):878-87.
266. Du WW, Yang W, Yee AJ. Roles of versican in cancer biology--tumorigenesis, progression and metastasis. Histol Histopathol. 2013 Jun;28(6):701-13.

267. Du WW, Fang L, Yang X, Sheng W, Yang BL, Seth A, et al. The role of versican in modulating breast cancer cell self-renewal. Mol Cancer Res. 2013 May;11(5):443-55.
268. Yamada S, Nomoto S, Fujii T, Takeda S, Kanazumi N, Sugimoto H, et al. Frequent promoter methylation of M-cadherin in hepatocellular carcinoma is associated with poor prognosis. Anticancer Res. 2007 Jul-Aug;27(4B):2269-74.
269. Suzuki A, Pelikan RC, Iwata J. WNT/beta-Catenin Signaling Regulates Multiple Steps of Myogenesis by Regulating Step-Specific Targets. Mol Cell Biol. 2015 May;35(10):1763-76.
270. Martin TA, Lane J, Harrison GM, Jiang WG. The Expression of the Nectin Complex in Human Breast Cancer and the Role of Nectin-3 in the Control of Tight Junctions during Metastasis. PLoS One. [doi:10.1371/journal.pone.0082696]. 2013;8(12):e82696.
271. Kurokawa Y, Honma K, Takemasa I, Nakamori S, Kita-Matsuo H, Motoori M, et al. Central genetic alterations common to all HCV-positive, HBV-positive and non-B, non-C hepatocellular carcinoma: a new approach to identify novel tumor markers. Int J Oncol. 2006 Feb;28(2):383-91.
272. Oshima T, Sato S, Kato J, Ito Y, Watanabe T, Tsuji I, et al. Nectin-2 is a potential target for antibody therapy of breast and ovarian cancers. Mol Cancer. 2013 Jun 12;12:60.
273. Dong Y, Walsh MD, McGuckin MA, Gabrielli BG, Cummings MC, Wright RG, et al. Increased expression of cyclin-dependent kinase inhibitor 2 (CDKN2A) gene product P16INK4A in ovarian cancer is associated with progression and unfavourable prognosis. Int J Cancer. 1997 Feb 20;74(1):57-63.
274. Sadikovic B, Thorner P, Chilton-Macneill S, Martin JW, Cervigne NK, Squire J, et al. Expression analysis of genes associated with human osteosarcoma tumors shows correlation of RUNX2 overexpression with poor response to chemotherapy. BMC Cancer. 2010 May 13;10:202.
275. Navab R, Strumpf D, To C, Pasko E, Kim KS, Park CJ, et al. Integrin alpha11beta1 regulates cancer stromal stiffness and promotes tumorigenicity and metastasis in non-small cell lung cancer. Oncogene. 2016 Apr 14;35(15):1899-908.
276. Westcott JM, Prechtl AM, Maine EA, Dang TT, Esparza MA, Sun H, et al. An epigenetically distinct breast cancer cell subpopulation promotes collective invasion. J Clin Invest. 2015 May;125(5):1927-43.
277. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. Nature. [10.1038/nature00766]. 2002;417(6892):949-54.

278. Bian Y, Chang X, Liao Y, Wang J, Li Y, Wang K, et al. Promotion of epithelial-mesenchymal transition by Frizzled2 is involved in the metastasis of endometrial cancer. Oncol Rep. 2016 Aug;36(2):803-10.

279. Sandsmark E, Hansen AF, Selnaes KM, Bertilsson H, Bofin AM, Wright AJ, et al. A novel non-canonical Wnt signature for prostate cancer aggressiveness. Oncotarget. 2016 Dec 24.

280. Gujral TS, Chan M, Peshkin L, Sorger PK, Kirschner MW, MacBeath G. A noncanonical Frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis. Cell. 2014 Nov 06;159(4):844-56.

281. Milovanovic T, Planutis K, Nguyen A, Marsh JL, Lin F, Hope C, et al. Expression of Wnt genes and frizzled 1 and 2 receptors in normal breast epithelium and infiltrating breast carcinoma. Int J Oncol. 2004 Nov;25(5):1337-42.

282. Millet C, Zhang YE. Roles of Smad3 in TGF-beta signaling during carcinogenesis. Crit Rev Eukaryot Gene Expr. 2007;17(4):281-93.

283. Han SU, Kim HT, Seong DH, Kim YS, Park YS, Bang YJ, et al. Loss of the Smad3 expression increases susceptibility to tumorigenicity in human gastric cancer. Oncogene. 2004 Feb 19;23(7):1333-41.

284. Kretzschmar M. Transforming growth factor-beta and breast cancer: Transforming growth factor-beta/SMAD signaling defects and cancer. Breast Cancer Res. 2000;2(2):107-15.

285. Sawazaki T, Minezawa M. The pig aminoacylase 1 (ACY1) and ribosomal protein L29 (RPL29)/heparin/heparan sulfate interacting protein (HIP) genes are located together at 13q21-->q22, corresponding to human 3p21.1. Cytogenet Cell Genet. 1999;86(1):56-61.

286. Liu JJ, Huang BH, Zhang J, Carson DD, Hooi SC. Repression of HIP/RPL29 expression induces differentiation in colon cancer cells. J Cell Physiol. 2006 May;207(2):287-92.

287. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin. 2009 Jul-Aug;59(4):225-49.

288. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006 Oct 13;314(5797):268-74.

289. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007 Apr;8(4):286-98.

290. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science. 2007 Nov 16;318(5853):1108-13.
291. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009 Dec 24;462(7276):1005-10.
292. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Research. 2012;22(2):246-58.
293. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008 Jun;9(6):465-76.
294. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotech. [Research]. 2010;28(10):1057-68.
295. Bird A. DNA methylation patterns and epigenetic memory. Genes & Development. 2002 January 1, 2002;16(1):6-21.
296. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature. [10.1038/nature02625]. 2004;429(6990):457-63.
297. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer. 2003 Apr;3(4):253-66.
298. Bird A. Perceptions of epigenetics. Nature. 2007 May 24;447(7143):396-8.
299. Swift-Scanlan T, Vang R, Blackford A, Fackler MJ, Sukumar S. Methylated genes in breast cancer: Associations with clinical and histopathological features in a familial breast cancer cohort. Cancer Biology & Therapy. 2011;11(10):853-65.
300. Brooks J, Cairns P, Zeleniuch-Jacquotte A. Promoter Methylation and the Detection of Breast Cancer. Cancer causes & control : CCC. 2009;20(9):1539-50.
301. Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, Perkmann E, et al. DNA methylation in serum of breast cancer patients: an independent prognostic marker. Cancer Res. 2003 Nov 15;63(22):7641-5.
302. Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, et al. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat Med. 1995 Jul;1(7):686-92.
303. Manjegowda MC, Gupta PS, Limaye AM. Hyper-methylation of the upstream CpG island shore is a likely mechanism of GPER1 silencing in breast cancer cells. Gene. 2017 May 30;614:65-73.

304. Islam S, Dasgupta H, Roychowdhury A, Bhattacharya R, Mukherjee N, Roy A, et al. Study of association and molecular analysis of human papillomavirus in breast cancer of Indian patients: Clinical and prognostic implication. PLoS One. 2017;12(2):e0172760.

305. Shilpi A, Bi Y, Jung S, Patra SK, Davuluri RV. Identification of Genetic and Epigenetic Variants Associated with Breast Cancer Prognosis by Integrative Bioinformatics Analysis. Cancer Inform. 2017;16:1-13.

306. Siddiqui S, Akhter N, Deo SV, Shukla NK, Husain SA. A study on promoter methylation of PTEN in sporadic breast cancer patients from North India. Breast Cancer. 2016 Nov;23(6):922-31.

307. Song Y, Zuo Y. Occurrence of HHIP gene CpG island methylation in gastric cancer. Oncol Lett. 2014 Nov;8(5):2340-4.

308. Mackrill JJ. Ryanodine receptor calcium channels and their partners as drug targets. Biochem Pharmacol. 2010 Jun 01;79(11):1535-43.

309. Li J, Yang Y, Inoue H, Mori M, Akiyoshi T. The expression of costimulatory molecules CD80 and CD86 in human carcinoma cell lines: its regulation by interferon gamma and interleukin-10. Cancer Immunol Immunother. 1996 Dec;43(4):213-9.

310. Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. Science. 2001 Aug 10;293(5532):1068-70.

311. Mascolo M, Siano M, Ilardi G, Russo D, Merolla F, De Rosa G, et al. Epigenetic disregulation in oral cancer. Int J Mol Sci. 2012;13(2):2331-53.

312. Qu Z, Fu J, Yan P, Hu J, Cheng SY, Xiao G. Epigenetic repression of PDZ-LIM domain-containing protein 2: implications for the biology and treatment of breast cancer. J Biol Chem. 2010 Apr 16;285(16):11786-92.

313. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nat Rev Genet. 2012 Oct;13(10):679-92.

314. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene. 2002 Aug 12;21(35):5400-13.

315. Song Y, Tian Y, Zuo Y, Tu JC, Feng YF, Qu CJ. Altered expression of PTCH and HHIP in gastric cancer through their gene promoter methylation: novel targets for gastric cancer. Mol Med Rep. 2013 Apr;7(4):1159-68.

316. Chang L, Zhang P, Zhao D, Liu H, Wang Q, Li C, et al. The hedgehog antagonist HHIP as a favorable prognosticator in glioblastoma. Tumour Biol. 2016 Mar;37(3):3979-86.

317. Davis FM, Parsonage MT, Cabot PJ, Parat M-O, Thompson EW, Roberts-Thomson SJ, et al. Assessment of gene expression of intracellular calcium channels, pumps and exchangers with epidermal growth factor-induced epithelial-mesenchymal transition in a breast cancer cell line. Cancer Cell International. [journal article]. 2013;13(1):76.

318. Pang JM, Dobrovic A, Fox SB. DNA methylation in ductal carcinoma in situ of the breast. Breast Cancer Res. 2013 Jun 28;15(3):206.

319. Ashour N, Angulo JC, Andres G, Alelu R, Gonzalez-Corpas A, Toledo MV, et al. A DNA hypermethylation profile reveals new potential biomarkers for prostate cancer diagnosis and prognosis. Prostate. 2014 Sep;74(12):1171-82.

320. Son JW, Kim YJ, Cho HM, Lee SY, Lee SM, Kang JK, et al. Promoter hypermethylation of the CFTR gene and clinical/pathological features associated with non-small cell lung cancer. Respirology. 2011 Nov;16(8):1203-9.

321. Agarwal R, Mori Y, Cheng Y, Jin Z, Olaru AV, Hamilton JP, et al. Silencing of claudin-11 is associated with increased invasiveness of gastric cancer cells. PLoS One. 2009 Nov 24;4(11):e8002.

322. Lin JC, Jeong S, Liang G, Takai D, Fatemi M, Tsai YC, et al. Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. Cancer Cell. 2007 Nov;12(5):432-44.

323. Fukuhara S, Chang I, Mitsui Y, Chiyomaru T, Yamamura S, Majid S, et al. DNA mismatch repair gene MLH1 induces apoptosis in prostate cancer cells. Oncotarget. 2014 Nov 30;5(22):11297-307.

324. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012 Jul 18;487(7407):330-7.

325. Berx G, van Roy F. Involvement of members of the cadherin superfamily in cancer. Cold Spring Harb Perspect Biol. 2009 Dec;1(6):a003129.

326. Shuang C, Dalin L, Weiguang Y, Zhenkun F, Fengyan X, Da P, et al. Association of CD40 gene polymorphisms with sporadic breast cancer in Chinese Han women of Northeast China. PLoS One. 2011;6(8):e23762.

327. Zhang C, Zhao H, Li J, Liu H, Wang F, Wei Y, et al. The identification of specific methylation patterns across different cancers. PLoS One. 2015;10(3):e0120361.

328. Misawa K, Mochizuki D, Endo S, Mima M, Misawa Y, Imai A, et al. Site-specific methylation patterns of the GAL and GALR1/2 genes in head and neck cancer: Potential utility as biomarkers for prognosis. Mol Carcinog. 2017 Mar;56(3):1107-16.

329. Sepulveda JL, Gutierrez-Pajares JL, Luna A, Yao Y, Tobias JW, Thomas S, et al. High-definition CpG methylation of novel genes in gastric carcinogenesis identified by next-generation sequencing. Mod Pathol. 2016 Feb;29(2):182-93.

330. Lam K, Pan K, Linnekamp JF, Medema JP, Kandimalla R. DNA methylation based biomarkers in colorectal cancer: A systematic review. Biochim Biophys Acta. 2016 Aug;1866(1):106-20.

331. Jou YC, Tsai YS, Chen SY, Hsieh HY, Tsai HT, Tzai TS. Loss of DAB2IP expression in human urothelial carcinoma is associated with poorer recurrence-free survival. Virchows Arch. 2016 Jun;468(6):733-40.

332. Bagadi SA, Prasad CP, Srivastava A, Prashad R, Gupta SD, Ralhan R. Frequent loss of Dab2 protein and infrequent promoter hypermethylation in breast cancer. Breast Cancer Res Treat. 2007 Sep;104(3):277-86.

333. Nayak BK, Das BR. Mutation and methylation status of p53 gene promoter in human breast tumours. Tumour Biol. 1999 Nov-Dec;20(6):341-6.

334. Moribe T, Iizuka N, Miura T, Kimura N, Tamatsukuri S, Ishitsuka H, et al. Methylation of multiple genes as molecular markers for diagnosis of a small, well-differentiated hepatocellular carcinoma. Int J Cancer. 2009 Jul 15;125(2):388-97.

335. Agha G, Houseman EA, Kelsey KT, Eaton CB, Buka SL, Loucks EB. Adiposity is associated with DNA methylation profile in adipose tissue. Int J Epidemiol. 2015 Aug;44(4):1277-87.

336. Lai RK, Chen Y, Guan X, Nousome D, Sharma C, Canoll P, et al. Genome-wide methylation analyses in glioblastoma multiforme. PLoS One. 2014;9(2):e89376.

337. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast Cancer Res. 2009;11(1):R7.

338. Wu X, Zhuang YX, Hong CQ, Chen JY, You YJ, Zhang F, et al. Clinical importance and therapeutic implication of E-cadherin gene methylation in human ovarian cancer. Med Oncol. 2014 Aug;31(8):100.

339. Yang B, Miao S, Zhang LN, Sun HB, Xu ZN, Han CS. Correlation of CCNA1 promoter methylation with malignant tumors: a meta-analysis introduction. Biomed Res Int. 2015;2015:134027.

340. Sriuranpong V, Mutirangura A, Gillespie JW, Patel V, Amornphimoltham P, Molinolo AA, et al. Global gene expression profile of nasopharyngeal carcinoma by laser capture

microdissection and complementary DNA microarrays. Clin Cancer Res. 2004 Aug 01;10(15):4944-58.
341. Muller-Tidow C, Ji P, Diederichs S, Potratz J, Baumer N, Kohler G, et al. The cyclin A1-CDK2 complex regulates DNA double-strand break repair. Mol Cell Biol. 2004 Oct;24(20):8917-28.
342. Li Z, Guo X, Wu Y, Li S, Yan J, Peng L, et al. Methylation profiling of 48 candidate genes in tumor and matched normal tissues from breast cancer patients. Breast Cancer Res Treat. 2015 Feb;149(3):767-79.
343. Blancato J, Graves A, Rashidi B, Moroni M, Tchobe L, Ozdemirli M, et al. SYK allelic loss and the role of Syk-regulated genes in breast cancer survival. PLoS One. 2014;9(2):e87610.
344. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct 04;490(7418):61-70.
345. Mukhopadhyay UK, Cass J, Raptis L, Craig AW, Bourdeau V, Varma S, et al. Dataset of STAT5A status in breast cancer. Data Brief. 2016 Jun;7:490-2.
346. Kar P, Supakar PC. Expression of Stat5A in tobacco chewing-mediated oral squamous cell carcinoma. Cancer Lett. 2006 Aug 28;240(2):306-11.
347. Liao Z, Lutz J, Nevalainen MT. Transcription factor Stat5a/b as a therapeutic target protein for prostate cancer. Int J Biochem Cell Biol. 2010 Feb;42(2):186-92.
348. Peck AR, Witkiewicz AK, Liu C, Klimowicz AC, Stringer GA, Pequignot E, et al. Low levels of Stat5a protein in breast cancer are associated with tumor progression and unfavorable clinical outcomes. Breast Cancer Res. 2012 Oct 04;14(5):R130.
349. Hong X, Chen G, Wang M, Lou C, Mao Y, Li Z, et al. STAT5a-targeting miRNA enhances chemosensitivity to cisplatin and 5-fluorouracil in human colorectal cancer cells. Mol Med Rep. 2012 May;5(5):1215-9.
350. Hu X, Dutta P, Tsurumi A, Li J, Wang J, Land H, et al. Unphosphorylated STAT5A stabilizes heterochromatin and suppresses tumor growth. Proc Natl Acad Sci U S A. 2013 Jun 18;110(25):10213-8.
351. Kim H, Lapointe J, Kaygusuz G, Ong DE, Li C, van de Rijn M, et al. The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer. Cancer Res. 2005 Sep 15;65(18):8118-24.

352. Harlaar N, Bryan AD, Thayer RE, Karoly HC, Oien N, Hutchison KE. Methylation of a CpG site near the ALDH1A2 gene is associated with loss of control over drinking and related phenotypes. Alcohol Clin Exp Res. 2014 Mar;38(3):713-21.

353. Liu L, Zhong R, Wei S, Xiang H, Chen J, Xie D, et al. The Leptin Gene Family and Colorectal Cancer: Interaction with Smoking Behavior and Family History of Cancer. PLoS One. [doi:10.1371/journal.pone.0060777]. 2013;8(4):e60777.

354. Wu Y, Sarkissyan M, Vadgama JV. Epigenetics in Breast and Prostate Cancer. Methods in molecular biology (Clifton, NJ). 2015;1238:425-66.

355. Sutherland KD, Lindeman GJ, Choong DY, Wittlin S, Brentzell L, Phillips W, et al. Differential hypermethylation of SOCS genes in ovarian and breast carcinomas. Oncogene. 2004 Oct 07;23(46):7726-33.

356. Letellier E, Schmitz M, Baig K, Beaume N, Schwartz C, Frasquilho S, et al. Identification of SOCS2 and SOCS6 as biomarkers in human colorectal cancer. Br J Cancer. [Molecular Diagnostics]. 2014;111(4):726-35.

357. Farabegoli F, Ceccarelli C, Santini D, Taffurelli M. Suppressor of cytokine signalling 2 (SOCS-2) expression in breast carcinoma. J Clin Pathol. 2005 Oct;58(10):1046-50.

358. Haffner MC, Petridou B, Peyrat JP, Revillion F, Muller-Holzner E, Daxenbichler G, et al. Favorable prognostic value of SOCS2 and IGF-I in breast cancer. BMC Cancer. 2007 Jul 25;7:136.

359. Qiu X, Zheng J, Guo X, Gao X, Liu H, Tu Y, et al. Reduced expression of SOCS2 and SOCS6 in hepatocellular carcinoma correlates with aggressive tumor progression and poor prognosis. Mol Cell Biochem. 2013 Jun;378(1-2):99-106.

360. Hudelist G, Wuelfing P, Czerwenka K, Knofler M, Haider S, Fink-Retter A, et al. Beta-hCG/LH receptor (b-HCG/LH-R) expression is increased in invasive versus preinvasive breast cancer: implications for breast carcinogenesis? J Cancer Res Clin Oncol. 2009 Feb;135(2):191-5.

361. Liu S, Ren S, Howell P, Fodstad O, Riker AI. Identification of novel epigenetically modified genes in human melanoma via promoter methylation gene profiling. Pigment Cell Melanoma Res. 2008 Oct;21(5):545-58.

362. Shivapurkar N, Toyooka S, Toyooka KO, Reddy J, Miyajima K, Suzuki M, et al. Aberrant methylation of trail decoy receptor genes is frequent in multiple tumor types. Int J Cancer. 2004 May 01;109(5):786-92.

363. Wang H-D, Hou Q-F, Guo Q-N, Li T, Wu D, Zhang X-P, et al. DNA methylation study of fetus genome through a genome-wide analysis. BMC Medical Genomics. 2014;7:18-.

List of Publications

1. **Malvia S**, Bagadi SA, Pradhan D, Chintamani C, Bhatnagar A, Arora D, Sarin R, Saxena S. Study of Gene Expression Profiles of Breast cancers in Indian Women, Scientific Reports (Under revision). **Impact factor 4.2**
2. **Malvia S**, Bagadi SA, Dubey US, Saxena S. Epidemiology of breast cancer in Indian women. Asia Pac J Clin Oncol 2017 **Impact factor 1.9**
3. Pandrangi SL, Raju Bagadi SA, Sinha NK, Kumar M, Dada R, Lakhanpal M, Soni A, **Malvia S**, Simon S, Chintamani C, Mohil RS, Bhatnagar D, Saxena S. Establishment and characterization of two primary breast cancer cell lines from young Indian breast cancer patients: mutation analysis. Cancer Cell Int. 2014 Feb 5;14(1):14. **Impact factor 2.74**

Presentations

1. "Whole genome gene expression profiling identifies key biological pathways differential in early and late onset breast cancer" for poster presentation in **HGM 2015, Kuala Lumpur**
2. "Differential expression profiling identifies pathways deregulated in early onset breast cancer" for poster presentation in **IACR-ACOS 2016, New Delhi**

Asia-Pacific Journal of
Clinical Oncology

Asia-Pacific Journal of Clinical Oncology 2017 doi: 10.1111/ajco.12661

REVIEW ARTICLE

Epidemiology of breast cancer in Indian women

Shreshtha MALVIA,[1] Sarangadhara Appalaraju BAGADI,[1] Uma S. DUBEY[2] and Sunita SAXENA[1]

[1]*National Institute of Pathology (ICMR), Safdarjung Hospital Campus, New Delhi* [2]*Birla institute of Technology, Pilani, Rajasthan, India*

Abstract

Breast cancer has ranked number one cancer among Indian females with age adjusted rate as high as 25.8 per 100,000 women and mortality 12.7 per 100,000 women. Data reports from various latest national cancer registries were compared for incidence, mortality rates. The age adjusted incidence rate of carcinoma of the breast was found as high as 41 per 100,000 women for Delhi, followed by Chennai (37.9), Bangalore (34.4) and Thiruvananthapuram District (33.7). A statistically significant increase in age adjusted rate over time (1982–2014) in all the PBCRs namely Bangalore (annual percentage change: 2.84%), Barshi (1.87%), Bhopal (2.00%), Chennai (2.44%), Delhi (1.44%) and Mumbai (1.42%) was observed. Mortality-to-incidence ratio was found to be as high as 66 in rural registries whereas as low as 8 in urban registries. Besides this young age has been found as a major risk factor for breast cancer in Indian women. Breast cancer projection for India during time periods 2020 suggests the number to go as high as 1797900. Better health awareness and availability of breast cancer screening programmes and treatment facilities would cause a favorable and positive clinical picture in the country.

Key words: breast, cancer, early age, India, PBCR, trends

INTRODUCTION

Breast cancer is the most common female cancer worldwide representing nearly a quarter (25%) of all cancers with an estimated 1.67 million new cancer cases diagnosed in 2012. Women from less developed regions (883 000 cases) have slightly more number of cases compared to more developed (794 000) regions.[1] In India, although age adjusted incidence rate of breast cancer is lower (25.8 per 100 000) than United Kingdom (95 per 100 000) but mortality is at par (12.7 *vs* 17.1 per 100 000) with United Kingdom.[2] There is a significant increase in the incidence and cancer-associated morbidity and mortality in Indian subcontinent as described in global and Indian studies.[3–7] Earlier cervical cancer was most common cancer in Indian woman but now the incidence of breast cancer has surpassed cervical cancer and is leading cause of cancer death, although cervical cancer still remains most common in rural India.[8]

To plan and formulate sound cancer control strategies based on scientific and empirical bases, authorities and policy makers need correct and complete knowledge of epidemiology. Although many epidemiologic studies have been conducted previously in cancers as prostate, gastric, oral including breast cancer[9–12] but this systematic review was conducted from evidences available on epidemiologic correlates of breast cancer addressing incidence, prevalence, and associated factors like age, time trends and other risk factors to understand disease burden and pattern in India. The objective of this review article is to bring together the information scattered in different Indian registries and studies to see a broader picture of breast cancer epidemiology in Indian subcontinent.

METHODS

Multiple sources from literature were used for gathering information and analysis of breast cancer. Information on crude rate (CR) and age adjusted rate (AAR) per 100,000 population was collected from National

Correspondence: Dr Sunita Saxena, Director Scientist-G, National Institute of Pathology (Indian Council of Medical Research), Safdarjung Hospital Campus, New Delhi 110029, India. Email: sunita_saxena@yahoo.com

Accepted for publication 3 December 2016.

© 2017 John Wiley & Sons Australia, Ltd

Pandrangi et al. Cancer Cell International 2014, 14:14
http://www.cancerci.com/content/14/1/14

PRIMARY RESEARCH Open Access

Establishment and characterization of two primary breast cancer cell lines from young Indian breast cancer patients: mutation analysis

Santhi Latha Pandrangi[1†], Sarangadhara Appala Raju Bagadi[1], Navin Kumar Sinha[2], Manoj Kumar[3], Rima Dada[3], Meena Lakhanpal[1], Abha Soni[1], Shreshtha Malvia[1], Sheeba Simon[1], Chintamani Chintamani[4], Ravindar Singh Mohil[4], Dinesh Bhatnagar[4] and Sunita Saxena[1*]

Abstract

Two novel triple negative breast cancer cell lines, NIPBC-1 and NIPBC-2 were successfully established from primary tumors of two young breast cancer patients aged 39 and 38 years respectively, diagnosed as infiltrating duct carcinoma of breast. Characterization of these cell lines showed luminal origin with expression of epithelial specific antigen and cytokeratin 18 and presence of microfilaments and secretary vesicles, microvilli, tight junctions and desmosomes on ultra-structural analysis. Both the cell lines showed anchorage independent growth and invasion of matrigel coated membranes. Karyotype analysis showed aneuploidy, deletions and multiple rearrangements in chromosomes 7, 9, X and 11 and isochromosomes 17q in both the cell lines. P53 mutational analysis revealed no mutation in the coding region in both the cell lines; however NIPBC-2 cell line showed presence of heterozygous C/G polymorphism, g.417 C > G (NM_000546.5) resulting in Arg/Pro allele at codon 72 of exon 4. Screening for mutations in BRCA1&2 genes revealed presence of three heterozygous polymorphisms in exon 11 of BRCA1 and 2 polymorphisms in exons 11, and14 of BRCA2 gene in both the cell lines. Both the cell lines showed presence of CD 44+/24-breast cancer stem cells and capability of producing mammosphere on culture. The two triple negative breast cancer cell lines established from early onset breast tumors can serve as novel *invitro* models to study mechanisms underlying breast tumorigenesis in younger age group patients and also identification of new therapeutic modalities targeting cancer stem cells.

Keywords: Breast cancer, Breast cancer cell line, Establishment

Introduction

Breast cancer is the leading cause of cancer deaths among women, accounting for 23% of the total cancer incidence and 14% cancer deaths globally [1]. In India breast cancer has emerged as most common cancer in women, which was earlier reported as second most common cancer after cancer of cervix [2], the age adjusted annual incidence rate (AAR) ranging from 25-33 cases per 100,000 women in urban population and 7.2 in rural areas [2]. Around 100,000 women are diagnosed with carcinoma breast every year in India, of which around 50,000 women die with the disease every year with a predicted rise to 131 000 cases by 2020, and increased concentration in urban areas [3]. The incidence of breast cancer in Indian population (1/35) is not as high as in the western countries (1/8) however, the incidence of early onset of breast cancer cases (<40 years) does not show significant variation in women worldwide (12-33 per 100,000 women); suggesting that a greater proportion of all breast cancers is mainly due to early onset of disease in Indian population [1,4]. The average age of onset of breast cancer in Indian patients ranges between 40-50 years compared to 60-70 in western countries. Breast cancer diagnosed at young age is well recognized as clinically different than breast cancers diagnosed at older ages [5]. Younger patients more frequently exhibit

* Correspondence: sunita_saxena@yahoo.com
[†]Equal contributors
[1]National Institute of Pathology (ICMR), Safdarjung Hospital Campus, New Delhi 110029, India
Full list of author information is available at the end of the article

© 2014 Pandrangi et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Brief Biography of Candidate

Shreshtha Malvia did her Masters in Biotechnology from University of Rajasthan. During this tenure, she did her dissertation from Dr. Reddy Laboratories, Hyderabad. After that she did 6 months Biotech Industrial Training Programme (Sponsored by Department of Biotechnology, Government of India) at Ranbaxy laboratories, Gurgaon. She worked as a Ranbaxy trainee for another 6 months. Then she worked as project assistant in NIPER (National Institute of Pharmeceutical Education and Research) Mohali. She joined the research team at the National Institute of Pathology (NIP) in 2009. Her research interests are in the area of Oncology, Cell culture, Biomarker discovery and Epigenetics. She worked as Project-SRF on breast cancer to identify novel biomarkers which could distinguish early onset breast carcinoma from late onset under supervision of Dr. Sunita Saxena at NIP, New Delhi. While doing SRF, she registered for PhD at Birla Institute of Science and Technology, Pilani. After finishing the project she joined as a SRF in another project where the leads from the earlier project was taken to further decipher the mutations in early onset patients by whole exome sequencing under the supervision of Dr. Sunita Saxena. Further, she was awarded as institutional ICMR Senior Research Fellow (SRF) to finish her pending work including validation of the identified biomarkers. She has presented her research work in national and international conferences such as HGM, Malaysia and IACR-ACOS, New Delhi. She has published 2 research articles on breast cancers in peer reviewed international journals.

Brief Biography of the Supervisor

Dr. Sunita Saxena is a Consultant and Ex-Director of National Institute of Pathology, one of the premier institutes of Indian Council of Medical Research in India. A pathologist by training, she has evolved beyond diagnostic pathology to using trans disciplinary approaches in understanding important cancers in India. She is one of the few molecular pathologists in country with keen interest in Oncopathology and Molecular oncology. She started her career in research working on medical diseases of the kidney and she has now diverted her specific research interest to tumour biology, specifically Breast Tumours, Tobacco Associated cancers and Genito urinary cancers. She has been trained at International Agency for Research in Cancer, Lyon, France for various genetic and molecular biology based technologies. Her current research interest is identification of genetic risk factors, prognostic and predictive biomarkers and novel drug targets for Breast cancer and Tobacco associated cancers in North east region using genome-wide approaches. Dr Sunita Saxena is recipient of Yamigawa-Yoshida fellowship of UICC, WHO fellowship, Novartis Oration and Dr. P.N.Wahi awards of ICMR and K.C.Basu Mullick award for best research work from Indian Association of Pathologists and Microbiologists. She is also Fellow of National Academy of Medical Sciences, Indian Association of Cancer Research, Proteomic Society of India, Human Genomic Organization and Indian College of Pathologists. She is also a life time International Member of Union against Cancer. She is a member of Member of Scientific Advisory committee and Project Review Committee of various reputed ICMR institute and funding body. She has more than 96 research articles in reputed journal and 2 book chapters.

Brief Biography of the Co-Supervisor

Dr. S.A Raju Bagadi is Scientist D has served at several capacities for past 10 years at National institute of pathology one of the premier institutes of Indian Council of Medical Research in India He has done his Masters in Biotechnology from Andhra University, Visakhapatnam, followed by Ph.D from Department of Biochemistry from AIIMS under the supervision of Dr. Ranju Ralhan. Dr SA Raju Bagadi is recipient of ACSB-UICC fellowship for pursuing his post doctoral research from University of Virginia, School of medicine, USA. He has established 5 breast cancer cell lines of Indian origin representing various cancer types and also for the first time demonstrated the loss of Dab2 and its epigenetic silencing in breast cancer. His current research interest is biomarker discovery; decipher role of epigenetics and role of non coding RNAs in breast cancer. The main focus is to understand the molecular pathogenesis of breast cancer, specifically early onset breast cancer in India. The underlying mechanisms of cancer stem cells which are responsible for development of resistance to chemotherapeutic drugs rendering difficulty in the treatment are being explored. He has 9 research publications in reputed journal.

Brief Biography of the Co-Supervisor

Dr. Uma S. Dubey is presently working as an Associate Professor in the Department of Biological Sciences at BITS Pilani-Pilani Campus. She has served this department for last 14 years at various teaching and research related posts. Earlier, she has conducted teaching and program development in the departments of Biotechnology, Microbiology and Environmental Sciences at the Institute of Life Sciences, Kanpur University. She also has a research experience of an year at the department of Plant Sciences, University of Alberta, Canada. She has done her Ph.D. in Immunology (under the Supervision of Prof. S.S. Agarwal) from Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow.

Teaching: She has taught more than 15 different courses in the department of Biological Sciences. She has been involved in the course planning and development of many of these. Besides this, she has initiated 2 new courses in Immunology and Cancer Biology. She has coordinated the course restructuring of M.Sc. and ME programs of the Biological Science Department as DCA Convener from 2012 -2014 and is continuing as a member of the same. She has coordinated the Science, Imagination and Discovery (SID) workshop at BITS Pilani. She has been Judge of various events of APOGEE from 2006-onwards in Biological Sciences and Medical Sciences categories and is a Life member of Indian Immunological Society and Member of My India Team, BITS Pilani.

Research: Her research interest is in both, theoretical and practical aspects of Immunology and Cancer Biology. Specifically in (i) Cellular Immune Responses: She has been involved in comparative studying on Lymphocyte proliferation, Natural Killer cell function and Antibody dependent cellular cytotoxicity at normal and febrile temperatures. Also the cell cycle proliferation kinetics of lymphocytes and cell lines is of interest. (ii) Mathematical Modelling of Immune system: She has been involved in studying the interaction of various components of the immune system with each other, with cancer cells and with infectious agents in the presence and absence of environmental toxicants. The analysis requires theoretic, mathematically and computational input. (iii) Cancer Biology: She is interested in studying the anticancer properties attributed to Camel milk and its associated mechanisms of action. She is also interested in alpha lactalbumin sequence analysis in various species. It is the primary component of HAMLET a recently discovered anticancer molecule (derived

from human milk) devoid of any side effect. She has 12 original research articles, 3 book chapters and 1 lab manual to her credit.