ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Exploring Molecular Pathways Significantly Associated with Breast Cancer Metastasis.

Vipin Tomar*[!], Kapil Chaudhary[*] Pramod K Yadava[#] and Anil Kumar![$#]

*Equal contribution
School of Life Sciences, Jawaharlal Nehru University, New Delhi, 110067
! Society for Science and Environmental Excellence, IGNOU Road, Delhi 110030
$ Department of Zoology, Miranda House, University of Delhi, Delhi, 110007

ABSTRACT

Worldwide breast carcinoma is recognized, one of the common malignancy afflicting female with a significant health problem. Breast cancer is believed to be a blended result of environmental and genetic factors. Here we performed the meta-analysis of microarray data to elucidate complex connections between genes and pathways in breast cancer metastasis. In the present study all datasets were normalized by robust multichip analysis (RMA) and statistical analysis were performed using R language. These datasets were retrieved from Gene Expression Omnibus (GEO) data repository (http://www.ncbi.nlm.nih.gov/geo/). In this meta-analysis study our results indicate the three core pathways, which are strongly affected during breast cancer metastasis: 'ECM-receptor interaction', 'Focal adhesion pathway' and 'pathways in cancer'. On further analysis of results, we found the fibronectin (Fn1) as a common gene present in all pathways, which indicate that the fibronectin is a major player in breast cancer metastasis. In conclusion this study not only illustrate the how standardize bioinformatics approaches can be used to study the microarray data in public repository but also offer analysis of vital pathways and genes, involved in breast cancer.

Keyword: Breast Cancer, Metastasis, Pathways, Bioinformatics, Fibronectin

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Worldwide breast carcinoma is recognized, one of the common malignancy afflicting female with a significant health problem. It account for more than 30 percent cancer cases in women therefore making it the most common form of cancer[1]. Breast cancer could have diverse spectrum of complex features which comprises a collection of heterogeneous disease with distinctive clinical, histo-pathological, and molecular features. Thus, it leads to encompass different entities with distinct subtypes, biological features and clinical outcomes[2–5]

The extensively used predictive clinico-pathological criteria are age, tumor size, lympho-vascular invasion, histologic grade, expression of steroid and growth factor receptors, estrogen-inducible genes like cathepsin D, protoonco genes like *ERBB2,* and mutations in the *TP53* gene are still the basis of treatment decision[2–7]. Unfortunately, subsequent clinical manifestations of patients cannot be predicted accurately by the breast cancer recurrence prognostic factors available in clinical practice[8].

During the progression of cancer many genes or their combination can be potentially involved. The differences in expression of individual gene and genes combinations can be seen among cancer progression stages. Individuals having some specific sequence alleles produced by multiple genetic alterations are comparatively more sensitive to cancer development. Presence and expression level of these alleles of these genes can be used as prognostic factors by calculating the chance of potential encounter by cancer for an individual. Cancer can be detected in very early stage, if the genetic expression features of these stages are known. Gene expression profiles can also be used to infer frequent progression pathways to estimate stage distances between tumors[1]. In advance cases the cancer cells start to spread distant sites such as bones, lungs, lymph nodes, liver and brain. The migration of cancer cells governs by a number of gene set and their expression pattern define the cell migratory property.

Thus, gene expression profiling of cancer can offer potential lead to development of new more sensitive prognostics, diagnostics marker. DNA microarray has been the method of choice for gene expression profiling and monitoring the complex expression patterns of uniquely identifiable markers and multi-genes 'signatures' which involved in breast carcinoma development. The correlation of expression patterns to specific features of phenotypic variation might provide the basis for an improved genetic and molecular taxonomy of cancer[9,10].

Meta-analysis can be used for summarizing and synthesizing studies to estimate the overall effect of global gene expression profile in breast cancer[11,12]. Cellular pathways and key genes involved in breast carcinoma development and progression can be determined by a broad class of models and bioinformatics tools during Meta-analysis including Quality control (QC) and analysis of general data up to the biological pathway level.

In the present study, we have analyzed the microarray data available in public repository with the help of 'Bio-Conductor' package of open source language R[13]. In addition, genetic profiling of carcinoma was summarized in a wide range of pathways arise from different datasets.

MATERIAL AND METHODS

Microarray dataset selection

Gene expression profiling data were obtained from National Center for Biotechnology Information Gene Expression Omnibus (GEO) data repository (http://www.ncbi.nlm.nih.gov/geo/). Standardized QC and preprocessing were done by open source R packages of Bioconductor 2.8, while subsequent pathway analysis were done by using online analysis tools GOEAST and WEB-based GEneSeTAnaLysis Toolkit (WebGestalt)(Planche et al. 2011). Datasets were included raw data CEL files as well as processed files for investigating expression profiling by array of human breast carcinoma tissue samples performed on the Affymetrix Gene Chip platform. Three different datasets shown in Table 1 were selected and downloaded.

ISSN: 0975-8585

Table 1. Selected datasets and their characteristics

Dataset	GO ID [ACCN]	Array type	Number of samples
Planche et al. (2011)	GDS4114	Affymetrix Human Genome U133 Plus 2.0 Array	24
Turashvili et al. (2011)	GSE5764	Affymetrix Human Genome U133 Plus 2.0 Array	30
Kretschmer et al., (2011).	GSE21422	Affymetrix Human Genome U133 Plus 2.0 Array	19

Gene expression profiles dataset by Planche *et. al*. (2011) [(14)] (GEO ID: GDS4114) consists of 24 tissue samples obtained from patients with clinically localized breast cancer and prostate cancer separately. In original study, out of 24 samples, 6 samples of stroma surrounding invasive breast primary tumors; 6 matched samples of normal stroma, 6 samples of stroma surrounding invasive prostate primary tumors; 6 matched samples of normal stroma are taken. As the focus of our current study was only breast cancer, 12 samples (6 from normal tissue and 6 from invasive breast primary tumors) were selected for further study.

The second dataset by Turashvili *et. al*., (2011) [(15)] (GEO ID: GSE5764) is composed of 66 Affymetrix Human Genome U133 2.0 arrays performed upon breast tissue. In this study 28 samples were collected at surgery from patients undergoing surgical resection for invasive breast cancer and 5 samples from reduction mammoplasty for normal tissue. All sample tissues were separated in normal and cancer tissues of stroma and epithelium which were subjected to microarray experimentation after total RNA extraction of total 66 samples separately. The aim of study was to compare cell type and disease state analysis of breast carcinoma. We have selected 5 normal stroma and 28 cancer stroma samples dataset for meta-analysis in current study.

The third dataset is a subset of 19 Affymetrix Human Genome U133A Plus 2.0 GeneChip originated from a gene expression experiment by Kretschmer *et. al*.. (2011) [(16)] GEO ID:GSE21422). The aim of original study was to identifythose marker genes, which show increased expression in ductal carcinoma in-situ and invasive ductal carcinomas, Out of 19 subset we have selected five sample of healthy breast and 5 samples of invasive ductal carcinoma for further analysis in current study.

DEGs analysis

The open source language R (version 2.13.0) and R packages of Bioconductor 2.8[(17)] was used for microarray data analysis. Downloaded raw CEL files of expression profiling data were read and normalized by robust multichip analysis (RMA). All genes corresponding to their probe sets were filtered by gene filter package of Bio-conductor with an intensity filter (the intensity of a gene should be above $\log_2(100)$ in at least 25 percent of the samples), and a variance filter (The inter-quartile range of $\log_2$ intensities should be at least 0.5).

In experiment design, all arrays were divided into two groups: normal and breast cancer carcinoma group based on condition (normal tissue versus advanced stage cancer tissue). Two different conditions were allotted to study namely control and carcinoma group. Significant deferentially expressed genes (DEGs) [(18,19)] between experimental groups were calculated by conducted the linear modeling with Limma package of Bio-cunductor.

P-Value was adjusted by with Benjamin and Hochberg (BH) method of false discovery rate (FDR) for multiple hypothesis testing. Genes with adjusted p-value ≤ 0.05 were selected as DEGs. The annotations of each probe set were obtained from extracted from corresponding hgu133plus2.db package.

Function annotation and gene set enrichment analysis

For the functional studies of microarray data, Gene Otology (GO) analysis is considered as a common and easy approach. GOEAST (Gene Ontology Enrichment Analysis Software Tollkit) was used to identify statistically overrepresented GO terms within given gene sets. Pathway Annotation and cluster analysis of

ISSN: 0975-8585

DEGs was carried on for all three studies by Geneset Analysis Toolkit V2 platform based on hyper geometric distribution with use of Kyoto Encyclopedia of Gens And Genomes (KEGG) pathway [20].

RESULTS

Data normalization

The quality of all three datasets was considerable for further analysis as modest differences between the arrays can be seen in all studies, when the boxplots of raw intensities and density histograms of log intensities before normalization have been plotted. The box plots and the density histogram of log intensities before normalization in Figure 1A and 1B are from Second Dataset by [15](Turashvili*et.al*.2011). All discrepancies between arrays were sufficiently removed by normalization. Box plots and the density histogram of log intensities after normalization are illustrated in Figure 1D and 1E.

The MA-plots before normalization of an example array selected from dataset of [15] (Turashvili et al., 2011) were shown in Figure 1C. Assuming that the majority of genes are unchanged, the MA-plots spread symmetrically around the x-axis (y = 0). Plots after normalization (Figure 1F) with little diversion from average log intensity of the arrays assign to X-axis showed the difference in logged intensity of one array to the reference median array. Furthermore, corrections for intensity-dependent biases by Robust Multi-array Average (RMA) normalization was done, which can be seen in MA-plots created for same example array as plot became centered and spread symmetrically around the x-axis with negligible deviations (y ≈ 0) (Figure 1F).

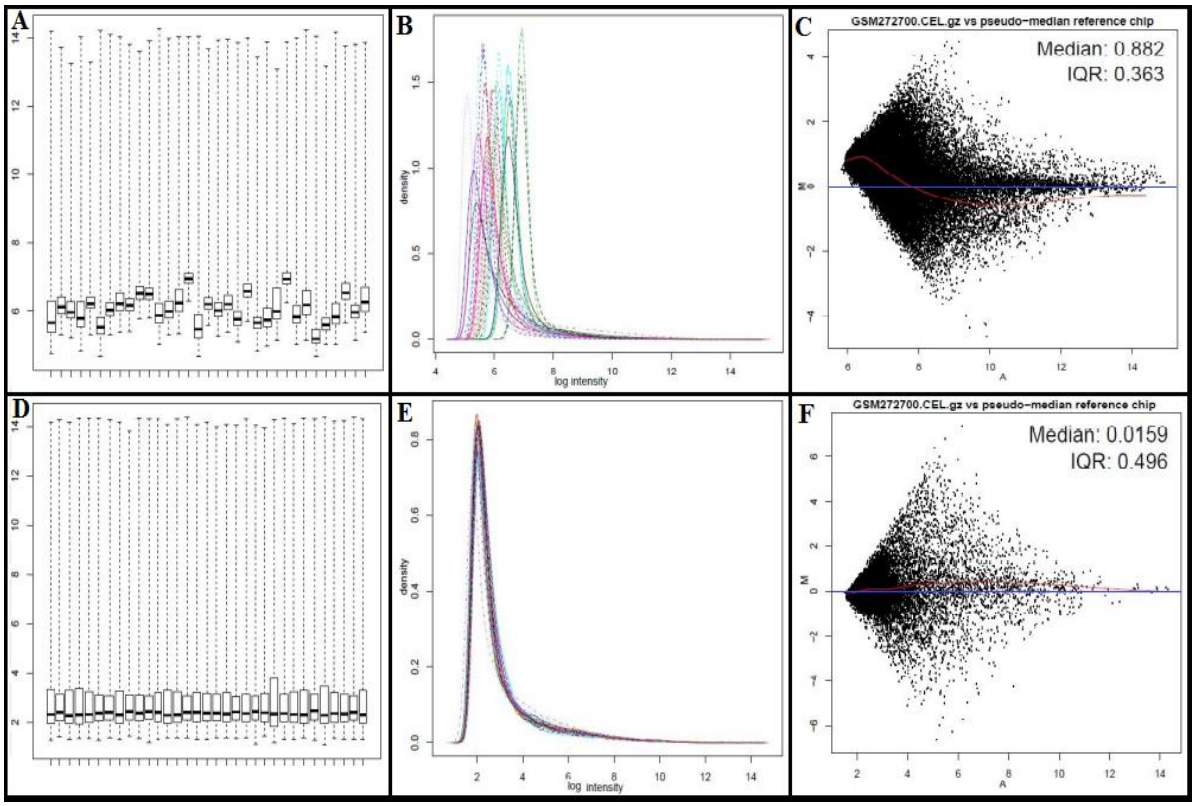


Figure 1: Quality control of dataset by Turashvili et al., 2011 ; (A) box plot before normalization (B) log intensity plot before normalization (C) MA plot of a sample array (GEO ID GSM272700)

ISSN: 0975-8585

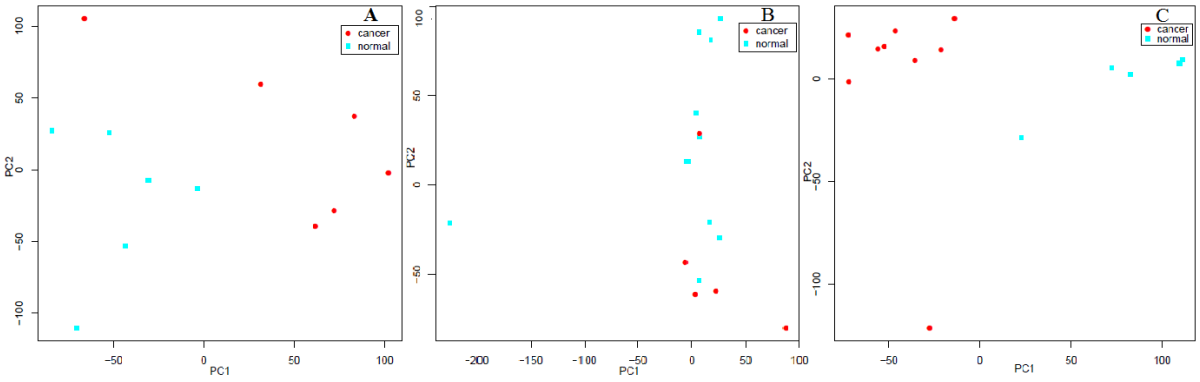


Figure 2: Principal Components Plots of A) dataset by (Planche et al., 2011); B) dataset by (Turashviliet al., 2011); C) dataset by (Kretschmer et al., 2011); clear evident groups can be seen in PCA among arrays in different conditions (normal versus cancer tissue).

Principal components analysis for all three studies is represented in Figure 2 which shows the clear evident grouping among the arrays according to the condition or state of disease. The two conditions (normal versus cancer tissue) were allocated to arrays.

Differentially expressed genes

Linear Models for Microarray (Limma) was used to identify genes differentially expressed between normal and cancer cells, with BH test corrections. At a FDR value of 0.05, a total of 2100 genes identified to be differentially expressed among 14000 probes from dataset of first study [14]. We have screened 878 up regulated genes and 1222 down regulated genes, according to the statistical analysis of microarray data. The top 20 significant up and down-regulated DEGs are shown respectively in Figure 3A and 3B.

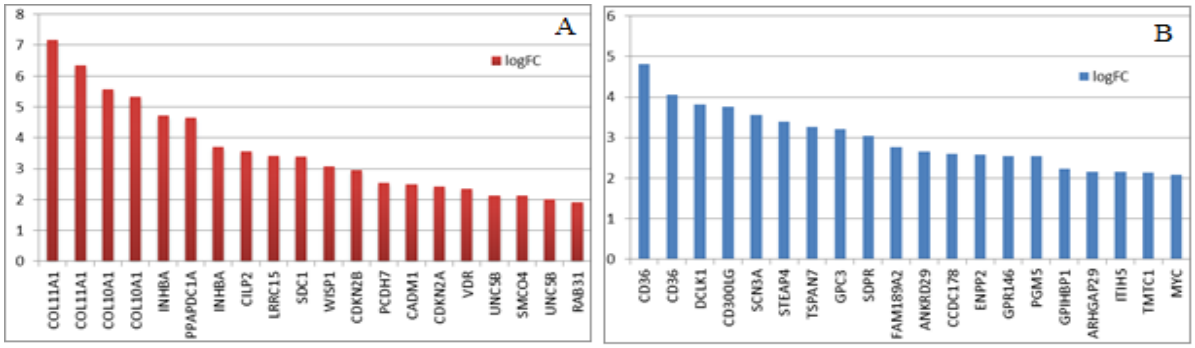


Figure 3: (A) top 20 up-regulated genes and; (B) top 20 down-regulated genes extracted from dataset by (Planche et al., 2011). Gene symbols are shown on X axis and log Fold Change is on y axis.

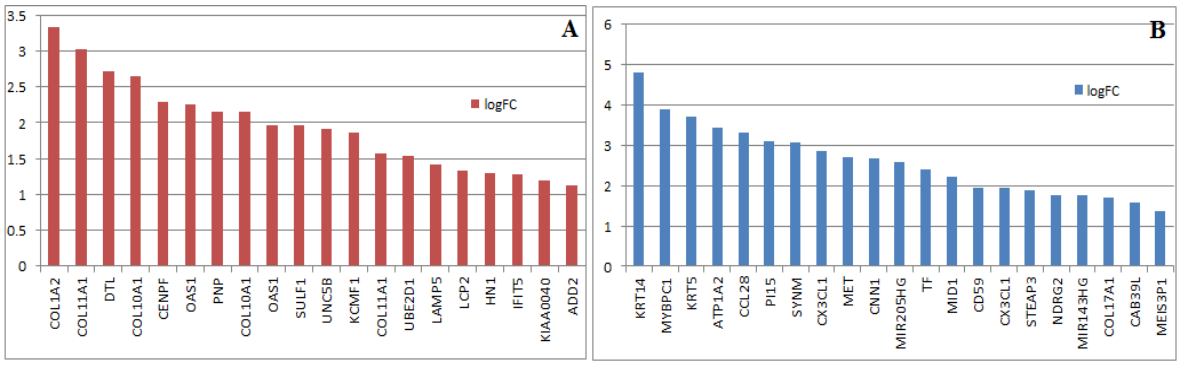


Figure 4: (A) top 20 up-regulated genes and (B) (A) top 20 down-regulated genes extracted from dataset by (Turashvili et al., 2011) gene symbol are shown on X axis and log Fold Change is on y axis.

ISSN: 0975-8585

969 differentially expressed genes including 337 up-regulated and 632 down-regulated genes were extracted with similar parameters for the statistical analysis of the dataset of second study, conducted by [15](Turashvili et al., 2011) included 54675 probes. Figure 4A shows top 20 significant up-regulated genes. Top 20 significant down-regulated genes are shown in Figure 4B.

The statistical analysis of microarray data between normal tissue sample and metastatic breast cancer sample in study by [16] (Kretschmeret al., 2011) revealed a total of 4610 genes differentially expressed among 54675 probes when threshold parameters were taken as similar to previous two studies. In which, 2031 genes were up-regulated while 2579 genes were down regulated. 20 top up and down-regulated genes are shown in Figure 5A and 5B respectively.

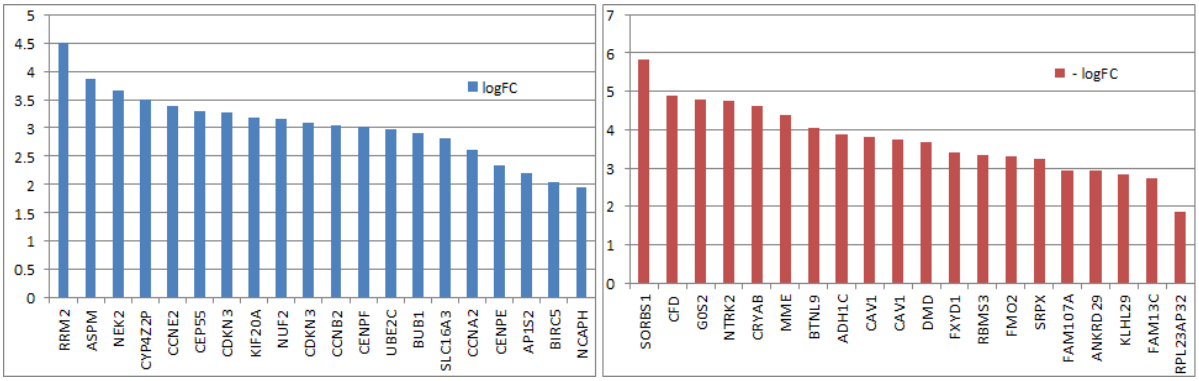


Figure 5: (A) top 20 up-regulated genes and(B) top 20 down-regulated genes extracted from dataset by (Kretschmeret al., 2011) gene symbol are shown on X axis and log Fold Change is on y axis.

Ontology and pathway analysis

Significant pathways corresponding to differentially expressed genes were identified by KEGG pathway enrichment analysis performed with Gene Set Analysis Toolkit V2 platform based on hyper geometric distribution.

The number of DEGs involved in different categories in ontology is represented in Figure 6. Total of 1350 enriched genes in the biological processes of first study, 786 enriched genes in biological regulation, 724 in metabolic process, 677 in response to stimulus and 587 in multi-cellular organism processes were mainly identified. The protein and ion bindings are most affected molecular functions in breast cancer corresponding 652 and 453 enriched genes respectively. Figure 7 shows the pathway diagram of enriched Gene Ontology (GO) in molecular function of the differentially expressed genes. In the terms of cellular component, 654 DEGs ware mapped as membrane proteins and 373 in nucleus. Statistical analysis of this study revealed that mainly 10 pathways were identified. Among these pathways, ECM-receptor interaction was identified most significant (FDR=4.32e-08) including 25 genes. Second significant pathway (FDR=2.07e-05) was Focal adhesion with 35 enriched genes. List of top 10 enriched pathways are shown in Table 2.

Table 2. Top 10 Significant pathways corresponding to dataset by Planche et al.,(2011)

Pathway Name	No. of Genes	Raw P Values	Adj P (FDR) Values
ECM-receptor interaction	25	2.47e-10	4.32e-08
Focal adhesion	35	2.37e-07	2.07e-05
Malaria	15	6.39e-07	3.73e-05
Pathways in cancer	47	1.31e-06	5.73e-05
Cytokine-cytokine receptor interaction	37	1.01e-05	0.0003
Complement and coagulation cascades	16	1.01e-05	0.0003
Cell adhesion molecules (CAMs)	24	7.96e-06	0.0003
Rheumatoid arthritis	17	6.32e-05	0.0014
Amoebiasis	19	0.0001	0.0016
Leukocyte transendothelial migration	20	8.08e-05	0.0016

ISSN: 0975-8585

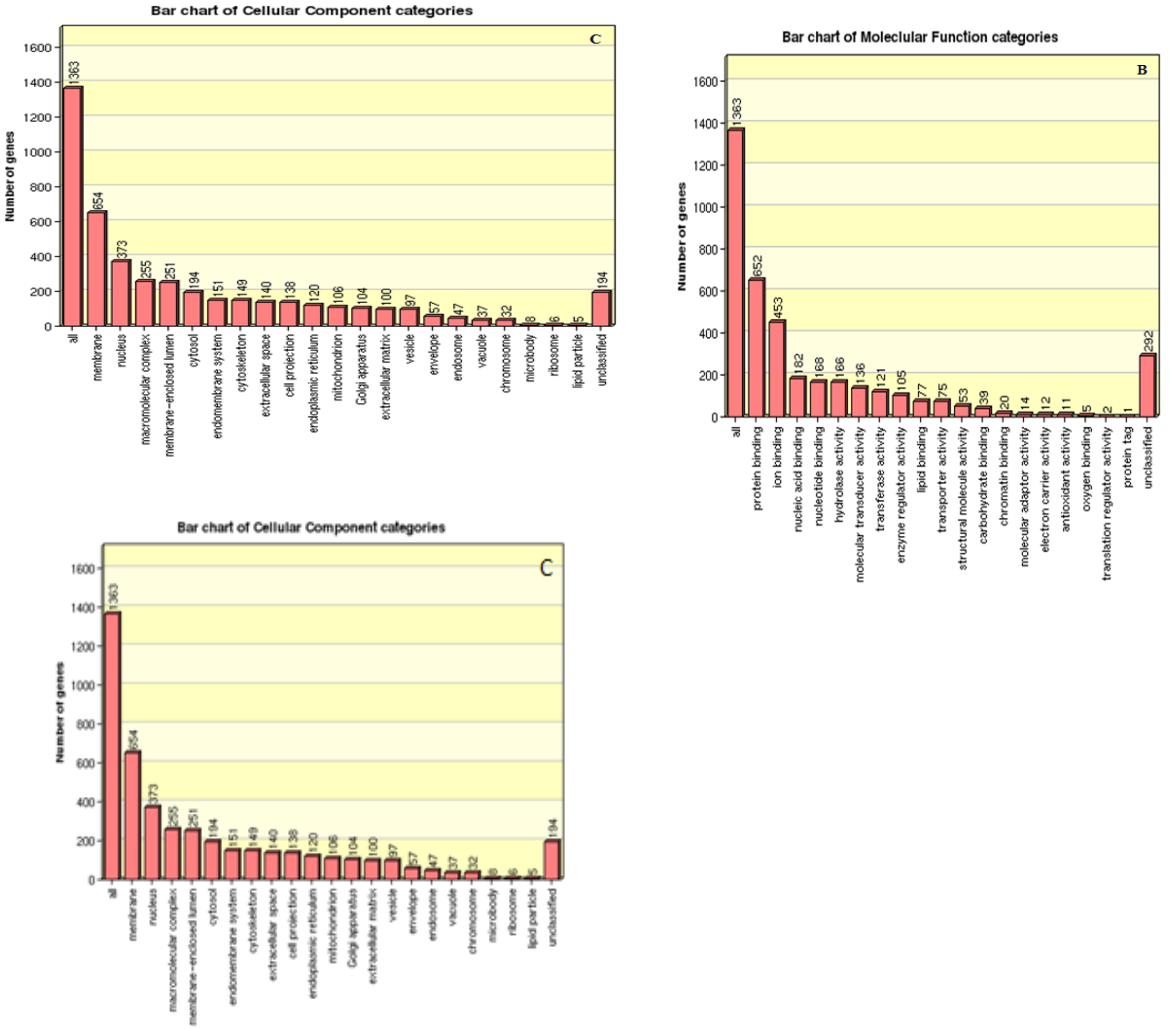
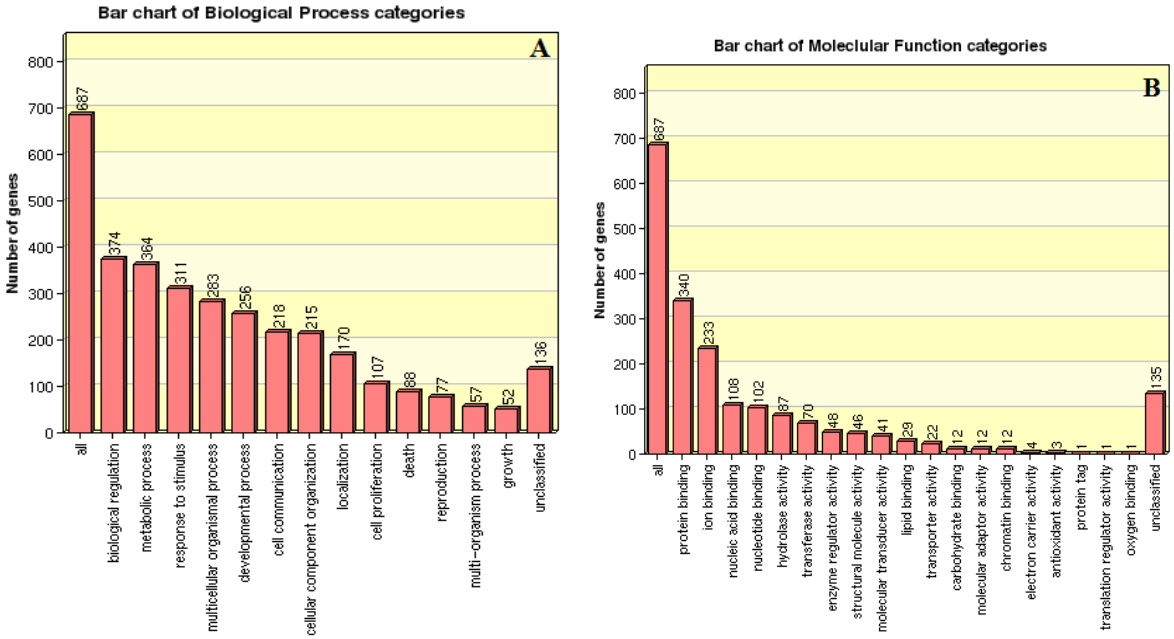


Figure 6: Gene Ontology Analysis of dataset by (Planche et al., 2011) (A) Biological Process; (B) Molecular Function; (C) Cellular Component



ISSN: 0975-8585

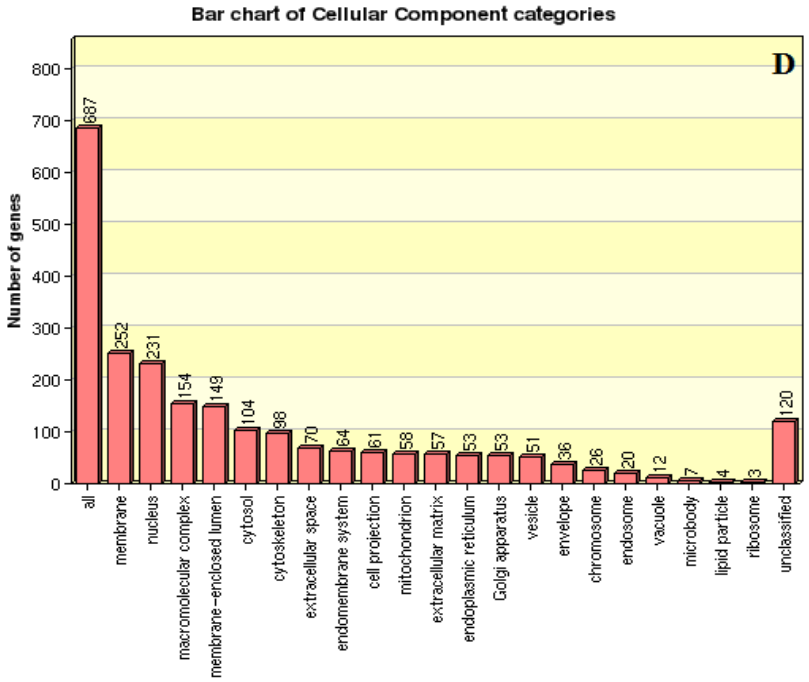


Figure 7: gene Ontology Analysis of dataset by Casey et al.; (A) Biological Process; (B) Molecular Function; (C) Cellular Component

In meta-analysis of dataset[15], total of 687 DEGs were enriched in ontology. Biological processes like biological regulation, metabolic process and response to stimulus contained the maximum number of enriched genes which were 374, 364 and 311 respectively (Figure 7). The enriched Gene Ontology (GO) of biological processes for the differentially expressed genes is shown in Figure 7A. The enriched Gene Ontology of the differentially expressed genes of molecular functions pathway is represented in figure 7B, which indicate 169 protein and ion binding enriched 103 genes. The enriched Gene Ontology of the differentially expressed genes in Cellular component is represented as pathway in figure 7C, in which the cellular component, 255 DEGs were mapped as membrane proteins and 231 in nucleus. The most significant identified pathway was ECM-receptor interaction with FDR 4.55e-07 containing 17 enriched genes. Focal adhesion was second significant pathway with approximately same FDR value then ECM-receptor interaction with 26 enriched genes. Protein digestion and absorption was third significantly enriched pathway (FDR = 0.0007) in meta-analysis of the dataset by Turashvili et al., (2011) [15] top 10 significantly enriched pathways identified were listed in Table 3.

Table 3. Top 10 Significant pathways corresponding to dataset by Turashvili et al. (2011)

Pathway Name	No. of Genes	Raw P Values	Adj. P (FDR) Values
ECM-receptor interaction	17	4.39e-09	4.55e-07
Focal adhesion	26	6.69e-09	4.55e-07
Protein digestion and absorption	12	1.45e-05	0.0007
p53 signaling pathway	11	2.29e-05	0.0008
Amoebiasis	13	8.73e-05	0.0024
Axon guidance	14	0.0002	0.0045
Pathways in cancer	22	0.0027	0.0525
Arrhythmogenic right ventricular cardiomyopathy	8	0.0040	0.0618
Glycine, serine and threonine metabolism	5	0.0048	0.0618
Drug metabolism - cytochrome P450	7	0.0045	0.0618

ISSN: 0975-8585

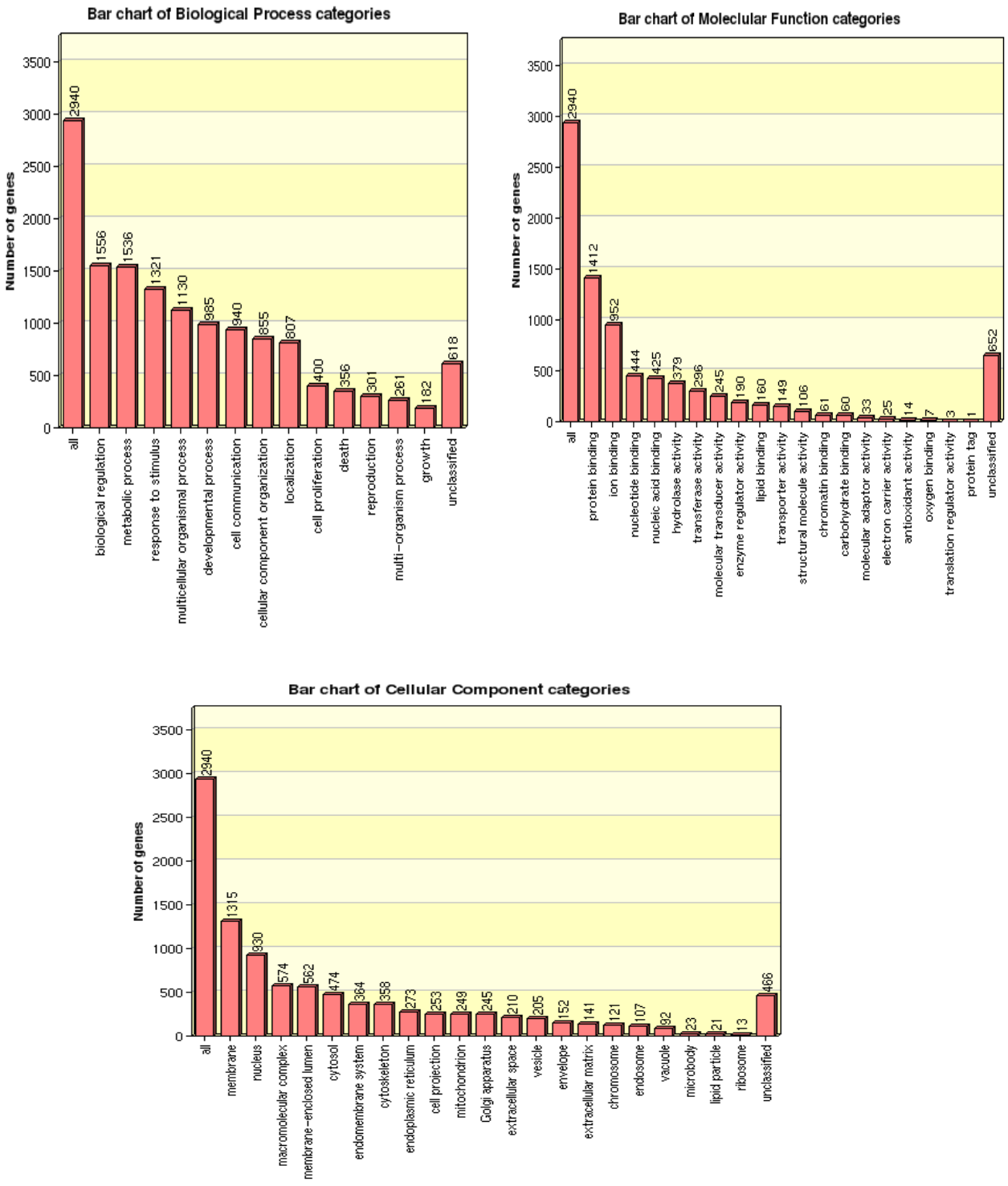


Figure 8: Gene Ontology Analysis of dataset by Kretschmer et al. 2011; (A) Biological Process; (B) Molecular Function; (C) Cellular Component

Figure 8 shows the Bar plots for number of DEGs involved in gene ontology for the dataset by [23]. In this study most of genes were enriched in biological regulation i.e. 1556 out of 2940 enriched genes, followed by 1536 in metabolic processes and 1320 in response to stimulus. The enriched Gene Ontology terms of biological progress of the differentially expressed genes (DEGs) are represented as pathway in figure 8A. In the molecular function pathway, 1412 protein and 952 ion binding genes were enriched. The enriched Gene Ontology molecular function pathway diagram is presented in figure 8B. 1315 genes were identified in

membrane while 930 genes are enriched for nucleus composition in the cellular component pathway (Figure 8C).

The comparison between normal tissue and breast cancer revealed 10 main biological pathways to be significantly changed. Focal adhesion was identified as most significant (FDR=1.07e-07) including 64 of genes. ECM-receptor interaction altered second most significantly with FDR = 3.37e-07 having 35 gene enriched while Systemic lupus erythematous was third significant pathway in meta-analysis of this study. Table 4 shows the pathways identified altered significantly.

Table 4. Top 10 Significant pathways corresponding to dataset by Kretschmer et al.,(2011).

Pathway Name	No. of Genes	Raw P Values	Adj. P (FDR) Values
Focal adhesion	64	5.20e-10	1.07e-07
ECM-receptor interaction	35	3.27e-09	3.37e-07
Systemic lupus erythematosus	31	6.63e-07	4.55e-05
Pathways in cancer	82	8.96e-07	4.61e-05
Cell adhesion molecules (CAMs)	40	3.23e-06	0.0001
PPAR signaling pathway	25	6.22e-06	0.0002
Tight junction	38	2.53e-05	0.0007
Axon guidance	37	4.20e-05	0.0011
Complement and coagulation cascades	23	6.72e-05	0.0015
Cell cycle	35	7.92e-05	0.0016

Common pathways among different studies

On the pathways enrichment of DEGs we find 10 main pathways for each study, involved in the progression of breast cancer. In all three studies which were undertaken in current meta-analysis, we find three common pathways namely the 'ECM-receptor interaction', 'Focal adhesion pathway' and 'pathways in cancer'. After that 'Amoebiasis' and 'Axon' guidance were common between first - second and second – third dataset. While, the two pathways were common between first and third dataset, namely 'cell adhesion molecules' and 'complement and coagulation cascades'. Venn diagram of common pathways among meta-analysis result and pathways of all three studies is shown in Figure 9.

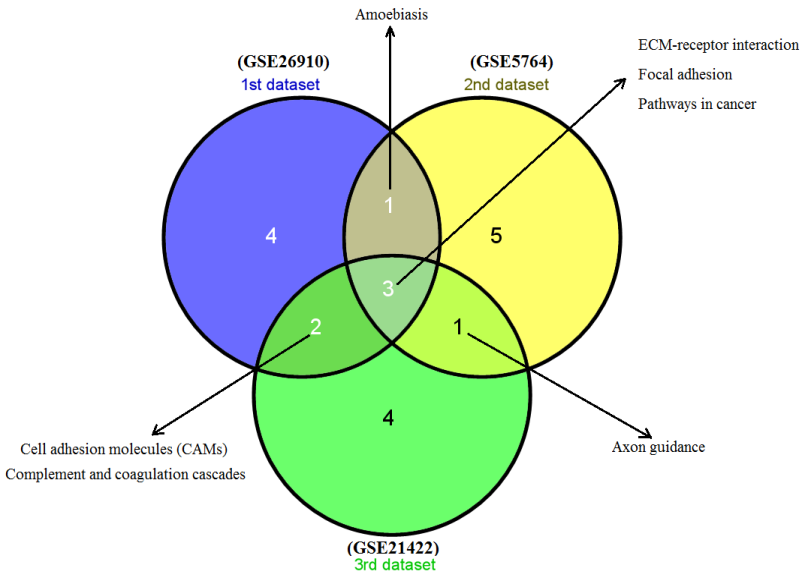


Figure 9: Venn representation of meta-analysis resultant common pathways of all three studies taken.

ISSN: 0975-8585

Most common genes among different pathways

On deep analysis of three dataset, we were able to find some most common genes associated with breast cancer metastasis. We find four genes common in ECM-Receptor Interaction pathway (Table 5), 7 genes in Focal Adhesion pathway (Table 6) and 8 genes were common in pathways in cancer (Table 7).

Table 5. Common genes present in ECM-Receptor Interaction pathway from all three studies

S.N.	Probe ID	Gene Symbol	Gene Name	Entrez Gene	Ensembl
1	202310_s_at	COL1A1	Collagen, type I, alpha 1	1277	ENSG00000108821
2	37892_at	COL11A1	Collagen, type XI, alpha 1	1301	ENSG00000060718
3	205713_s_at	COMP	Cartilage oligomeric matrix protein	1311	ENSG00000105664
4	211719_x_at	FN1	Fibronectin 1	2335	ENSG00000115414

Table 6. Common genes present in Focal Adhesion pathway from all three studies

S.N.	Probe ID	Gene Symbol	Gene Name	Entrez Gene	Ensembl
1	212097_at	CAV1	Caveolin 1, caveolae protein, 22kDa	857	ENSG00000105974
2	202310_s_at	COL1A1	Collagen,typeI,alpha 1	1277	ENSG00000108821
3	37892_at	COL11A1	Collagen,typeXI,alpha 1	1301	ENSG00000060718
4	205713_s_at	COMP	Cartilage oligomeric matrix protein	1311	ENSG00000105664
5	211719_x_at	FN1	Fibronectin 1	2335	ENSG00000115414
6	209542_x_at	IGF1	Insulin-like growth factor 1 (somatomedin C)	3479	ENSG00000017427
7	203510_at	MET	Met proto-oncogene (hepatocyte growth factor receptor)	4233	ENSG00000105976

Table 7. Common genes present in pathways in cancer from all three studies.

S.N.	Probe ID	Gene Symbol	Gene Name	Entrez Gene	Ensembl
1	205289_at	BMP2	Bone morphogenetic protein 2	650	ENSG00000125845
2	211719_x_at	FN1	Fibronectin 1	2335	ENSG00000115414
3	209189_at	FOS	FBJ murine osteosarcoma viral oncogene homolog	2353	ENSG00000170345
4	209542_x_at	IGF1	Insulin-like growth factor 1 (somatomedin C)	3479	ENSG00000017427
5	203510_at	MET	Met proto-oncogene (hepatocyte growth factor receptor)	4233	ENSG00000105976
6	203936_s_at	MMP9	Matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	4318	ENSG00000100985
7	236094_at	TCF7L2	Transcription factor 7-like 2 (T-cell specific, HMG-box)	6934	ENSG00000148737
8	205883_at	ZBTB16	Zinc finger and BTB domain containing 16	7704	ENSG00000109906

ISSN: 0975-8585

Common genes/gene among all studies

After the selection of genes present in common pathways from all of three studies, three lists of common genes were obtained. In further search, genes were search common from the lists of genes present in common pathways from all of three studies. In the result of this search, only one gene "fibronectin 1" was found common in different pathways from all of three studies.

DISCUSSION

In the present study, we have taken three studies simultaneously for the pathways enrichment analysis of micro-array data available in public repository, further we analyzed the 2030 DEGs of breast cancer. On the pathways enrichment of DEG we find 10 main pathways involves in the progression of breast cancer. In all the studies which taken in this study we find three common pathways namely the ECM-receptor interaction, Focal adhesion pathway and pathways in cancer. The ECM-receptor interaction pathway basically involved in complex mixture of structural and functional macromolecules and serves an important role in tissue and organ morphogenesis and in the maintenance of cell and tissue structure and function. The involvement of ECM-receptor interaction in breast cancer was also reported in number of studies[21,22] .

Gene expression data often lack statistical power due to several constraints, especially low sample number, as was the case in all three studies. This generally leads to underestimation of variances, which inflates the false-positive rate. The quality of the meta-analysis benefits from the number of single data sets analyzed.

Another important pathway predicted in this meta-analysis micro-array data is focal adhesion pathway, which is known to plays important role in cell motility, cell proliferation, cell differentiation, regulation of gene expression and cell survival. Basically, it mediated the adhesion links between integrin, proteoglycan and actin cytoskeleton. The involvement of focal adhesion molecules is well known in tumor metastasis and a number of workers reported its involvement in various cancer namely in pancreatic ductal adenocarcinoma [23], ovarian cancer cells[24], gastric cardia adenocarcinoma[25].

Pathways in cancer is also appears to be the third most important pathway in this analysis. There are several reports available in data base which indicates the alteration of these pathways leads to development of breast cancer[21].

On further analysis of our meta-analysis results we find 4 genes common in ECM receptor pathway, 7 genes in focal adhesion pathway and 8 gene common in pathways in cancer. Our data in indicates that these genes may play very crucial role in breast cancer metastasis. The Collagen, type I, alpha 1 and Collagen type XI, alpha1 are known to play a role in cancer invasiveness. Recently reported[26] COL11A1 as a potential biomarker of invasiveness in breast tumor lesions. There finding was based on two hundred and one breast Core Needle Biopsy samples, analyzed by immunohistochemistry against pro-COL11A1. The two other proteins in this pathways analysis are Cartilage oligomeric matrix protein and fibronectin were reported as major players in tumor invasiveness[27–29].The scaffolding protein of plasma membrane and structural protein of lysosomes and autosomes, caveolae, plays major roles in cellular processes like autophagyin lysosomes,endocytosis, mechanotransduction, signaling, lipid homeostasis and autolysosomes for degradation of intracellular proteins and organelles appears to be major player in breast cancer metastasis and the role of cav1 in breast cancer metastasis and invasions is highlight by many workers[30–32]. The abnormal activation of Insulin-like growth factor 1 (somatomedin C) and Met proto-oncogene (hepatocyte growth factor receptor) in breast cancer is associated with poor prognosis, and aberrantly active MET triggers tumor growth, formation of new blood vessels (angiogenesis). These new blood vessels in tumor are required for nutrient supply. A number of studies proposed the tumor aggressiveness is related to abnormal expression of these genes[33–36]. The Matrix metallopeptidase 9 (MMP-9), also known as 92 kDa type IV collagenase, 92 kDagelatinase or gelatinase B (GELB), is a family of the zinc-metalloproteinases family, encoded as signal peptide and is involved in normal physiological processes, such as embryonic development, reproduction, angiogenesis, bone development, wound healing, cell migration anddegradation of the extracellular matrix. The altered expression of this protein in development of cancer is well documented[37,38]. The two other genes, which are basically transcription factors namely TCF7L2 (Transcription factor 7-like 2 (T-cell specific, HMG-box and Zinc finger and

BTB domain-containing protein 16 (ZBTB16)regulates variety of cellular processes. The TCF7L2 gene is known to stimulate Wnt signaling pathway. There are number of studies, which relate the altered TCF7L2 expression with the development of cancer[39,40]. Transcription factor (ZBTB16) plays an important role in histone deacetylation. Specific instances of aberrant gene rearrangement at this locus have been associated with acute promyelocytic leukemia (APL). Some of very recent reports suggest that it might be an important factor in development of breast cancer[28,41].

Further analysis of data reveals that fibronectin (F1) is the genes which have a role in almost all pathways under consideration in this study. It suggests that fibronectin (Fn1) have crucial role in breast cancer metastasis[28,41].

The study helps in understanding the major pathways associated with breast cancer, reveals a common network of genes and helps in finding out central gene (Fn1) associated with breast cancer metastasis. The pathway network and associated gene combination may be used as early biomarker for breast cancer metastasis and furthermore it may help in management of breast cancer.

ACKNOWLEDGMENTS

AK is thankful to UGC, New Delhi for providing Dr. D.S. Kothari Postdoctoral Fellowship.

REFERENCES

[1] Park Y, Shackney S, Schwartz R. Network-Based Inference of Cancer Progression from Microarray Data. IEEE/ACM Trans. Comput. Biol. Bioinformatics, 2009; 6(2):200–212.

[2] Elston C w., Ellis I o. pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology, 1991; 19(5):403–410.

[3] Vollenweider-Zerargui L, Barrelet L, Wong Y, Lemarchand-Béraud T, Gómez F. The predictive value of estrogen and progesterone receptors' concentrations on the clinical behavior of breast cancer in women °clinical correlation on 547 patients. Cancer, 1986; 57(6):1171–1180.

[4] Foekens JA, Look MP, Vries JB, Gelder MEM, Putten WLJ van, Klijn JGM. Cathepsin-D in primary breast cancer: prognostic evaluation involving 2810 patients. British Journal of Cancer, 1999; 79(2):300–307.

[5] Slamon DJ. Studies of the HER-2/neu Proto-oncogene in Human Breast Cancer. Cancer Investigation, 1990; 8(2):253–254.

[6] Børresen A-L, Andersen TI, Eyfjörd JE, Cornelis RS, Thorlacius S, Borg Å, Johansson U, Theillet C, Scherneck S, Hartman S, Cornelisse CJ, Hovig E, Devilee P. TP53 mutations and breast cancer prognosis: Particularly poor survival rates for cases with mutations in the zinc-binding domains. Genes, Chromosomes and Cancer, 1995; 14(1):71–75.

[7] Bergh J, Norberg T, Sjögren S, Lindgren A, Holmberg L. Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy. Nature Medicine, 1995; 1(10):1029–1034.

[8] McGuire WL. Breast Cancer Prognostic Factors: Evaluation Guidelines. JNCI: Journal of the National Cancer Institute, 1991; 83(3):154–155.

[9] Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 2000; 403(6769):503–511.

[10] Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Raffeld M, Yakhini Z, Ben-Dor A, Dougherty E, Kononen J, Bubendorf L, et al. Gene-Expression Profiles in Hereditary Breast Cancer. New England Journal of Medicine, 2001; 344(8):539–548.

[11] Schneider J, Ruschhaupt M, Buneß A, Asslaber M, Regitnig P, Zatloukal K, Schippinger W, Ploner F, Poustka A, Sültmann H. Identification and meta-analysis of a small gene expression signature for the diagnosis of estrogen receptor status in invasive ductal breast cancer. International Journal of Cancer, 2006; 119(12):2974–2979.

[12] Mehra R, Varambally S, Ding L, Shen R, Sabel MS, Ghosh D, Chinnaiyan AM, Kleer CG. Identification of GATA3 as a Breast Cancer Prognostic Marker by Global Gene Expression Meta-analysis. Cancer Research, 2005; 65(24):11259–11264.

[13] Gristwood RE, Venables WN. Effects of Fenestra Size and Piston Diameter on the Outcome of Stapes Surgery for Clinical Otosclerosis. Annals of Otology, Rhinology & Laryngology, 2011; 120(6):363–371.
[14] Planche A, Bacac M, Provero P, Fusco C, Delorenzi M, Stehle J-C, Stamenkovic I. Identification of Prognostic Molecular Features in the Reactive Stroma of Human Breast and Prostate Cancer. PLOS ONE, 2011; 6(5):e18640.
[15] Turashvili G, McKinney SE, Goktepe O, Leung SC, Huntsman DG, Gelmon KA, Los G, Rejto PA, Aparicio SAJR. P-cadherin expression as a prognostic biomarker in a 3992 case tissue microarray series of breast cancer. Modern Pathology, 2011; 24(1):64–81.
[16] Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag PM, Kemmner W. Identification of early molecular markers for breast cancer. Molecular Cancer, 2011; 10:15.
[17] Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology, 2004; 5:R80.
[18] Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genomics, 2006; 7:252.
[19] Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol, 2004; 3(1).
[20] Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Research, 2013; 41(W1):W77–W83.
[21] Yang, Jia M, Li Z. Bioinformatics analysis of aggressive behavior of breast cancer via an integrated gene regulatory network. 2014.
[22] Huan JL, Gao X, Qi X. Screening for key genes associated with invasive ductal carcinoma of the breast via microarray data analysis. GMR | Genetics and Molecular Research, 2015.
[23] Muniyan S, Haridas D, Chugh S, Rachagani S, Lakshmanan I, Gupta S, Seshacharyulu P, Smith LM, Ponnusamy MP, Batra SK. MUC16 contributes to the metastasis of pancreatic ductal adenocarcinoma through focal adhesion mediated signaling mechanism. Genes & Cancer, 2016; 7(3–4):110–124.
[24] Chen Q, Xu R, Zeng C, Lu Q, Huang D, Shi C, Zhang W, Deng L, Yan R, Rao H, Gao G, Luo S. Down-Regulation of Gli Transcription Factor Leads to the Inhibition of Migration and Invasion of Ovarian Cancer Cells via Integrin β4-Mediated FAK Signaling. PLOS ONE, 2014; 9(2):e88386.
[25] Wang Y, Feng X, Jia R, Liu G, Zhang M, Fan D, Gao S. Microarray expression profile analysis of long non-coding RNAs of advanced stage human gastric cardia adenocarcinoma. Molecular Genetics and Genomics, 2014; 289(3):291–302.
[26] Freire J, García-Berbel L, García-Berbel P, Pereda S, Azueta A, García-Arranz P, De Juan A, Vega A, Hens Á, Enguita A, Muñoz-Cacho P, Gómez-Román J. Collagen Type XI Alpha 1 Expression in Intraductal Papillomas Predicts Malignant Recurrence. BioMed Research International, 2015; 2015.
[27] Zeng Z, Hincapie M, Pitteri SJ, Hanash S, Schalkwijk J, Hogan JM, Wang H, Hancock WS. A Proteomics Platform Combining Depletion, Multi-lectin Affinity Chromatography (M-LAC) and Isoelectric Focusing to Study the Breast Cancer Proteome. Analytical chemistry, 2011; 83(12):4845–4854.
[28] Agajanian M, Runa F, Kelber JA. Identification of a PEAK1/ZEB1 signaling axis during TGFβ/fibronectin-induced EMT in breast cancer. Biochemical and Biophysical Research Communications, 2015; 465(3):606–612.
[29] Fernandez-Garcia B, Eiró N, Marín L, González-Reyes S, González LO, Lamelas ML, Vizoso FJ. Expression and prognostic significance of fibronectin and matrix metalloproteases in breast cancer metastasis. Histopathology, 2014; 64(4):512–522.
[30] Shi Y, Tan S-H, Ng S, Zhou J, Yang N-D, Koo G-B, McMahon K-A, Parton RG, Hill MM, del Pozo MA, Kim Y-S, Shen H-M. Critical role of CAV1/caveolin-1 in cell stress responses in human breast cancer cells via modulation of lysosomal function and autophagy. Autophagy, 2015; 11(5):769–784.
[31] Pucci M, Bravatà V, Forte GI, Cammarata FP, Messa C, Gilardi MC, Minafra L. Caveolin-1, Breast Cancer and Ionizing Radiation. Cancer Genomics - Proteomics, 2015; 12(3):143–152.
[32] Han B, Tiwari A, Kenworthy AK. Tagging Strategies Strongly Affect the Fate of Overexpressed Caveolin-1. Traffic (Copenhagen, Denmark), 2015; 16(4):417–438.
[33] Kim YJ, Choi J-S, Seo J, Song J-Y, Eun Lee S, Kwon MJ, Kwon MJ, Kundu J, Jung K, Oh E, Shin YK, Choi Y-L. MET is a potential target for use in combination therapy with EGFR inhibition in triple-negative/basal-like breast cancer. International Journal of Cancer, 2014; 134(10):2424–2436.

[34] Akl MR, Ayoub NM, Mohyeldin MM, Busnena BA, Foudah AI, Liu Y-Y, Sayed KAE. Olive Phenolics as c-Met Inhibitors: (-)-Oleocanthal Attenuates Cell Proliferation, Invasiveness, and Tumor Growth in Breast Cancer Models. PLoS ONE, 2014; 9(5).

[35] Tomblin JK, Salisbury TB. Insulin like growth factor 2 regulation of Aryl Hydrocarbon Receptor in MCF-7 breast cancer cells. Biochemical and biophysical research communications, 2014; 443(3):1092–1096.

[36] Holland JD, Györffy B, Vogel R, Eckert K, Valenti G, Fang L, Lohneis P, Elezkurtaj S, Ziebold U, Birchmeier W. Combined Wnt/β-Catenin, Met, and CXCL12/CXCR4 Signals Characterize Basal Breast Cancer and Predict Disease Outcome. Cell Reports, 2013; 5(5):1214–1227.

[37] Eiró N, Fernandez-Garcia B, Vázquez J, del Casar JM, González LO, Vizoso FJ. A phenotype from tumor stroma based on the expression of metalloproteases and their inhibitors, associated with prognosis in breast cancer. Oncoimmunology, 2015; 4(7).

[38] Bottino J, Gelaleti GB, Maschio LB, Jardim-Perassi BV, Zuccari de C, Pires DA. Immunoexpression of ROCK-1 and MMP-9 as prognostic markers in breast cancer. Acta Histochem, 2014:1367–73.

[39] Chen C, Cao F, Bai L, Liu Y, Xie J, Wang W, Si Q, Yang J, Chang A, Liu D, Liu D, Chuang T-H, Xiang R, Luo Y. IKKβ Enforces a LIN28B/TCF7L2 Positive Feedback Loop That Promotes Cancer Cell Stemness and Metastasis. Cancer Research, 2015; 75(8):1725–1735.

[40] Mir R, Pradhan SJ, Patil P, Mulherkar R, Galande S. Wnt/β-catenin signaling regulated SATB1 promotes colorectal cancer tumorigenesis and progression. Oncogene, 2016; 35(13):1679–1691.

[41] an H-C, Jiang Q, Yu Y, Mei J-P, Cui Y-K, Zhao W-J. Quercetin promotes cell apoptosis and inhibits the expression of MMP-9 and fibronectin via the AKT and ERK signalling pathways in human glioma cells. Neurochemistry International, 2015; 80:60–71.