ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Vol 9, Suppl. 3, 2016

Online - 2455-3891
Print - 0974-2441

Research Article

DUCTAL CARCINOMA *IN SITU* AND INVASIVE BREAST CANCER-BASED DIFFERENTIAL GENE EXPRESSION STUDY FOR THERAPEUTIC DEVELOPMENT

HARSHITHA GOPISETTY RAMACHANDRA[1], INAMUL HASAN MADAR[2], SEETHALAKSHMI SAKTHIVEL[1], IFTIKHAR ASLAM TAYUBI[3], SKM HABEEB[1]*

[1]Department of Bioinformatics, School of Bio-Engineering, SRM University, Chennai - 603 203, Tamil Nadu, India. [2]Division of Chemical Engineering & Bio Engineering, Department of Bio Engineering, Hanyang University, Seoul, South Korea. [3]Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh-21911, Saudi Arabia. Email: habeeb_skm@yahoo.co.in

Received: 26 July 2016, Revised and Accepted: 27 August 2016

ABSTRACT

Objective: Breast cancer is the second most common cancer in women globally. Multiple inherited mutations in genes are predominantly associated with breast cancer. The gene expression profiling of breast tumors generated by DNA microarray analysis provides molecular phenotyping that determines and characterizes the classifications of these tumors.

Methods: In this work, we used gene expression profiling of breast cancer samples from Gene Expression Omnibus (GEO) database. The dataset GSE41194, retrieved from GEO, was used to investigate differential gene expression in ductal carcinoma *in situ* (DCIS) and invasive breast cancer (IBC). The dataset contains 26 DCIS and 24 IBC samples. The data were analyzed in R and Bioconductor. To normalize the data Robust Multiarray Average (RMA) method was applied, limma software was used to identify the differentially expressed genes (DEGs) in DCIS and IBC; an adjusted p value ≤0.05 was used to filter differentially expressed probe sets, and a fold change (FC) ≥ 2 to identify upregulated and ≤−2 for downregulated genes. The DEGs retrieved were clustered and annotated using Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources with an EASE score ≤0.1 and count 2.

Results: The analysis obtained 72 DEGs with a p≤0.05. The FC≥2 identified 38 upregulated probesets and FC≤−2 identified 34 downregulated probe sets. The up and downregulated genes obtained in various comparisons were characterized based on gene ontology (GO) and pathway analyses in DAVID, which retrieved six genes that had principal pathways targeting breast cancer.

Conclusion: Identification of these genes and pathways enhances the knowledge and progression of DCIS to IBC; paving a novel way for developing new therapies for treating patients with breast cancer.

Keywords: Molecular phenotyping, Gene Expression, Ductal carcinoma *in situ*, Invasive breast cancer.

 DOI: http://dx.doi.org/10.22159/ajpcr.2016.v9s3.14317

INTRODUCTION

Breast cancer is the most commonly diagnosed malignancy among females. While decrease in both, breast cancer incidence and mortality, have been apparent in recent years, the societal and economic impact of this malignancy continues to be enormous [1]. The cases of incidence were 1.8 million in 2013 and 464 thousand deaths approximately [2]. Nearly, 30% of all cancers in women occur in breast both in the developed and the developing world [3]. The genetic abnormalities such as variations in high-penetrance genes play a major role in about 90% of breast cancer cases. Several risk factors for breast cancer have been identified. Some mutations particularly in *BRCA1, BRCA2, p53*, *PTEN, ATM, NBS1,* and *LKB1* result in a very high risk for breast cancer [4].

Breast cancers are of two different types, invasive and non-invasive. Invasive cancers spread to other tissues in the breast from the milk ducts, whereas the non-invasive cancers do not invade other tissues in the breast. The non-invasive breast tumors are referred to as *"in situ."* These are classified as ductal carcinoma *in situ* (DCIS) or intraductal carcinoma and lobular carcinoma *in situ* [4].

DCIS is characterized by malignant epithelial cells confined to the ductal system of the breast, without evidence of invasion through the basement membrane into the surrounding stroma [5]. Once thought to be a rare breast lesion, DCIS now constitutes 20% of all newly diagnosed breast cancer cases (http://seer.cancer.gov, Accessed October 2013; http://www.cancer.org, Accessed June 1, 2012). Invasive breast cancer (IBC) starts in a milk duct of the breast, breaks through the wall of the duct and grows into the fatty tissue of the breast. At this point, it may be able to spread (metastasize) to other parts of the body through the lymphatic system and bloodstream [5]. Invasive breast carcinoma constitutes 70-85% of the incidence; the remaining 15-30% are *in situ* carcinomas, 80% of which are DCIS (http://www.cancer.org, Accessed June 1, 2012).

The factors that stimulate the breast cancer risk include gender, age, family history and additionally alcohol intake, dietary fat, obesity in postmenopausal age, and hormonal stimulations. These factors are said to have increased the progression of breast cancer along with the individual factors almost half a century. The dramatic increase in breast cancer research and its prevention has shown positivist approach in the current years [6].

With the advent of microarray technology, the procedure to measure gene expression on a genome-wide scale has transformed cancer biology by providing the tools to measure differences in diseases [7]. This technology utilizes differential gene expression patterns in cancer cells and normal cells or those of other subtypes of cancer to identify the genes that are over-expressed and under-expressed [8]. However, the analysis produces a large amount of data, which is challenging to interpret. With the employment of modern computational and statistical analysis packages and bioinformatics tools, the data analysis has been greatly flexible in the recent years. The microarray technology has been applied to a range of applications, including discovering novel disease subtypes, developing new diagnostic tools, and identifying underlying mechanisms of disease or drug response [9].

In this work, we studied the gene expression profiling of breast cancer samples from Gene Expression Omnibus (GEO) database. The dataset GSE41194 was retrieved from GEO, to investigate differential gene expression in DCIS and IBC. Gene expression profiling of DCIS and IBC was performed to discover uniquely expressed genes that also regulate the progression. Our study also focused on identifying pathways associated with the genes, which enables to develop novel treatments for DCIS and IBC.

METHODS

Data quality check for the samples in the dataset

To check the quality and detect the outliers within the samples in the dataset, diagnostic plots such as boxplots and density plots were plotted. These plots give a quick view of the normalized log2 intensities.

Gene expression in DCIS and IBC

To investigate the differential expression in DCIS and IBC, the dataset GSE41194 deposited by Lee *et al*. [10], titled differentially expressed genes (DEGs) regulating the progression of DCIS to IBC (Group 1) was downloaded from GEO database [11]. The dataset contains 26 DCIS samples and 24 IBC samples. The platform used was GPL8300 [HG_U95Av2] Affymetrix Human Genome U95 Version 2 Array. The original files (raw data) and the platform probe annotation files were downloaded.

Identification of DEGs

The original data were classified as DCIS and IBC groups and were analyzed using R software (v.3.0.1) [12] and Bioconductor (v.2.14) [13]. The multichip normalization method robust multiarray average was used for background correction, normalization across the chips, and summarization of probe level data [14]. Finally, Limma–Linear Models for Microarray Data [15], linear regression model software, were used to compare the differential expression on different classes of chips. To identify the differentially expressed genes in DCIS and IBC, an adjusted *p value ≤0.05 was used as the cut-off criterion. Furthermore, to filter the differentially expressed probe sets, a *fold change (FC)≥2 was used to identify upregulated genes and ≤−2 for downregulated genes.

Gene ontology (GO) of DEGs

To investigate the DEGs at a functional level: Primarily, Database for Annotation, Visualization and Integrated Discovery (DAVID)-v.6.7 [16] was used to functionally interpret gene lists, to analyze the GO classification of terms [17], for identification of cellular components (CC), biological process (BP), and molecular function (MF) and for visualizing genes and mapping pathways Kyoto Encyclopedia of Genes and Genomes (KEGG) [18] was used. The DEGs for stronger gene enrichment analysis were chosen with an *EASE Score Threshold of ≤0.1 for the maximum probability and a default *Count threshold (minimum count) of 2 for including the minimum number of genes for the corresponding GO term.

RESULTS

Quality analysis of samples in the dataset

The quality analysis involves the assessment of the data and detection of the outliers. In this analysis, boxplots and histograms were plotted to see whether the samples in the dataset had any outliers. The boxplot of the raw data (Fig. 1) represents the distribution of log2 intensities across all the samples. The boxplot of normalized signal intensities across all samples provides a certainty that the normalization step was accomplished (Fig. 2).

The density plot shows the biased log2 intensity distribution for all the samples (Fig. 3). The histogram obtained after normalization makes the distributions essentially the same across all the samples (Fig. 4).

Identification of DEGs

The limma package was used to build model matrix with defined contrasts and an adjusted false discovery rate to analyze the gene

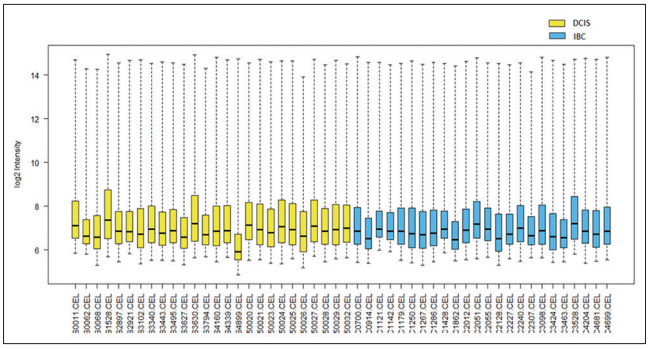

Fig. 1: The boxplot showing the summarized log2 intensities on the y-axis and the distribution of 26 ductal carcinoma *in situ* and 24 invasive breast cancer samples for the raw data

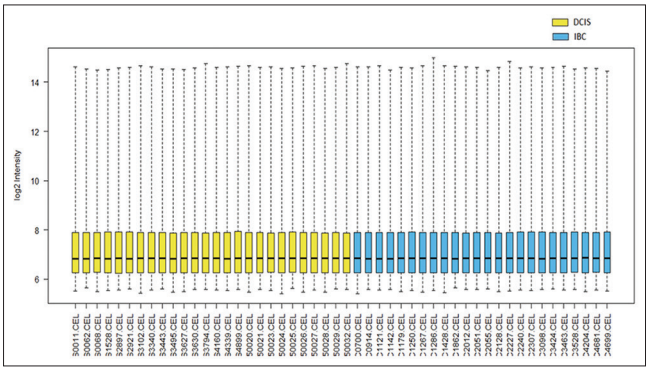

Fig. 2: The boxplot showing the normalized log2 intensities on the y-axis and the distribution of 26 ductal carcinoma *in situ* and 24 invasive breast cancer samples

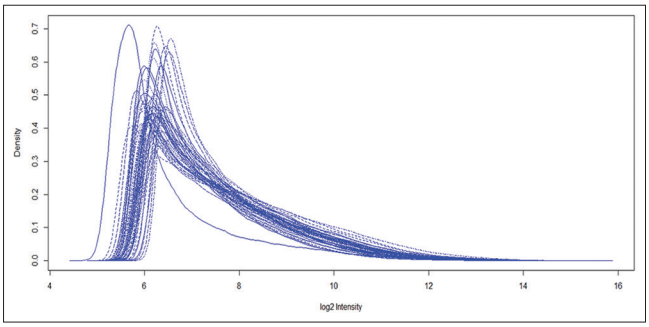

Fig. 3: The density plot showing histogram the log2 intensities plotted on the x-axis and distribution of density on the y-axis for the raw data

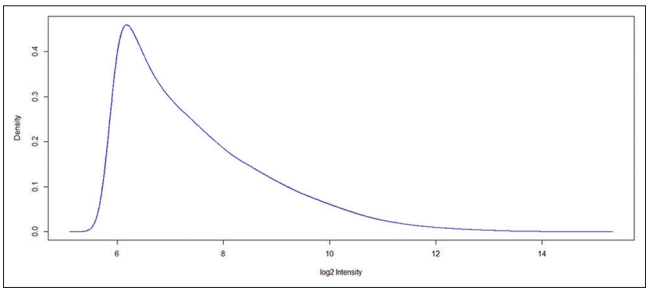

Fig. 4: The density plot showing histogram of the normalized log2 intensities plotted on the x-axis and distribution of density on the y-axis

expression profiles of DCIS and IBC. The analysis identified 72 genes that were found to be differentially expressed with the adjusted an *p≤0.05 and *FC≥2 and ≤−2. The *FC≥2 and ≤−2 revealed 38 probe sets that were upregulated and 34 probe sets that were downregulated.

GO clustering and pathway enrichment of DEGs

The functional classification of the obtained 72 DEGs was performed with the online biological classification tool-DAVID. The gene list was submitted with Affymetrix Human U133 chip as background and was provided for enrichment calculation. An *EASE Score Threshold (maximum probability), a modified *Fisher exact p value ≤0.1, was used for strong gene-enrichment. The*count threshold (minimum count) of 2 was used to retrieve minimum gene counts belonging to a GO term with its categories (classifications) - BP, CC, and MFs. The functional annotations of gene classifications, with their GO terms, p-value, count, and percentages that present study identified are detailed in Table 1. The DAVID analysis revealed six genes that were significantly associated with GO terms and pathways. The GO associated with the genes are shown in Table 2. The KEGG pathway associations for the obtained genes are reported in Table 3. Further investigation on these genes and pathways pave a novel way for developing new therapies for treating patients with breast cancer.

DISCUSSION

Breast cancer accounts the principle cause of death among women, with high incidence rates in Australia/New Zealand, North America, and several European countries [19,20]. It is estimated that 1.67 million breast cancer cases have been diagnosed in 2012, according to the Global Cancer Observatory series of the International Agency for Research on Cancer [21]. The steady increase in the morbidity of breast cancer in the recent years indicates a need for additional research on this disease.

Breast cancer has been described as an alarming health problem in India. It is the second most common cancer. A survey carried out by the Indian Council of Medical Research in the metropolitan cities, viz., Delhi, Mumbai, Bangalore, and Chennai; from 1982 to 2005, has shown that the incidences of breast cancer have doubled. Over the years, the incidences of breast cancer in India have steadily increased and as many as 100,000 new patients are being detected every year. A 12% increase has been registered by cancer registries from 1985 to 2001, which represented 57% rise of cancer burden in India [22].

The differential gene expression analysis of DCIS and IBC identified 72 genes with a significant *p value ≤0.05; in which 38 probe sets were upregulated, and 34 probe sets were downregulated. These DEGs retrieved are important for investigating the mechanism of disease development from DCIS to IBC. It is well known that breast cancer treatment has severe side effects which enable to find better chemotherapeutic agents [23]. The analysis of differentially expressed genes and their associated annotations provide useful information with genes function in all GO categories. These may be useful for determining and understanding the specific gene expression in the development of disease targets for the treatment of DCIS and IBC.

The results of GO functional annotation and pathway enrichment analysis of DEGs retrieved 6 genes and their associated pathways that were significantly enriched with GO terms, classifications and were associated with Mitogen-Activated Protein Kinase (MAPK) signaling

Table 1: GO categories with their corresponding GO terms

Category	Term	Count (%)	p value
UP_SEQ_FEATURE	Glycosylation site: N-linked (GlcNAc...)	4 (80)	0.033779
GOTERM_CC_FAT	GO: 0044420~extracellular matrix part	2 (40)	0.036119
SP_PIR_KEYWORDS	Glycoprotein	4 (80)	0.037616
SP_PIR_KEYWORDS	Extracellular matrix	2 (40)	0.049187
GOTERM_BP_FAT	GO: 0007423~sensory organ development	2 (40)	0.049932
SP_PIR_KEYWORDS	Actin-binding	2 (40)	0.050388
GOTERM_BP_FAT	GO: 0048666~neuron development	2 (40)	0.073315
GOTERM_BP_FAT	GO: 0030182~neuron differentiation	2 (40)	0.094028
GOTERM_CC_FAT	GO: 0005578~proteinaceous extracellular matrix	2 (40)	0.096454
GOTERM_MF_FAT	GO: 0003779~actin binding	2 (40)	0.09673

GO: Gene ontology

Table 2: GO terms associated with genes and their minimum count threshold and p value

GO term	Count (%)	p value
GO: 0007423~sensory organ development	2 (40)	0.049932
GO: 0048666~neuron development	2 (40)	0.073315
GO: 0030182~neuron differentiation	2 (40)	0.094028
GO: 0044420~extracellular matrix part	2 (40)	0.036119
GO: 0005578~proteinaceous extracellular matrix	2 (40)	0.096454
GO: 0003779~actin binding	2 (40)	0.09673

GO: Gene ontology

Table 3: KEGG pathway associations of the genes

Gene name	KEGG pathway
Collagen, Type XI, Alpha 1	Hsa04510: Focal Adhesion, Hsa04512: ECM-Receptor Interaction
Dystonin	
Neurotrophic Tyrosine Kinase, Receptor, Type 2	Hsa04010: MAPK signaling pathway, Hsa04722: Neurotrophin signaling pathway
Prolactin-induced protein	-
Tyrosine aminotransferase	Hsa00130: Ubiquinone and other terpenoid-quinone biosynthesis, Hsa00270: Cysteine and methionine metabolism, Hsa00350: Tyrosine metabolism, Hsa00360: Phenylalanine metabolism, Hsa00400: Phenylalanine, tyrosine and tryptophan biosynthesis

KEGG: Kyoto Encyclopedia of Genes and Genomes

pathway, neurotrophin signaling pathway, cysteine and methionine metabolism, tyrosine metabolism, phenylalanine metabolism, tyrosine and tryptophan biosynthesis cellular signaling pathways, and others. The MAPK pathway identified in our analysis is one of the principal targets for treating breast cancer [24]. This pathway is involved in various cellular functions, including cell proliferation, differentiation, and migration signaling [25]. Hence, it is necessary to carry out additional research for identifying potential targets as therapeutic agents that may directly or indirectly be involved in treating breast cancer.

CONCLUSION

Microarrays emerged as large-scale experimental studies to generate expression of thousands of genes parallelly. This technology makes biological observations more significant from a statistical point of view. Our study focused on analyzing and understanding breast tumor types - DCIS and IBC, as these data can provide us a wealth of information on the genetic susceptibility of disease through which decisive steps can be taken to translate these findings to clinical care. Gene expression profiling of breast tumors enables us to have a better understanding of tumor type and what markers certain tumors may have. Identifying gene profiles for DCIS and IBC tumors allows us to better group and classify these tumors. This enables development of better drug and treatment procedures. The gene expression affected by complex biochemical pathways and signaling events can be studied eventually.

ACKNOWLEDGMENTS

We would like to thank SRM University for all the support extended toward this research.

REFERENCES

1. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: An international comparison of screening, incidence, survival and mortality. Cancer Epidemiol 2012;36(3):237-48.
2. Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, MacIntyre MF, *et al.* The global burden of cancer 2013. JAMA Oncol 2015;1(4):505-27.
3. Smigal C, Jemal A, Ward E, Cokkinides V, Smith R, Howe HL, *et al.* Trends in breast cancer by race and ethnicity: Update 2006. CA Cancer J Clin 2006;56(3):168-83.
4. Kumar R, Sharma A, Tiwari RK. Application of microarray in breast cancer: An overview. J Pharm Bioallied Sci 2012;4(1):21-6.
5. Alteri R, Bandi P. Breast Cancer Facts & Figures 2011-2012. Atlanta: American Cancer Society, Inc.; 2011.
6. Cazzaniga M, Bonanni B. Breast cancer chemoprevention: Old and new approaches. J Biomed Biotechnol 2012;2012:985620.
7. Nevins JR, Potti A. Mining gene expression profiles: Expression signatures as cancer phenotypes. Nat Rev Genet 2007;8(8):601-9.
8. Kihara D, Yang YD, Hawkins T. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. Cancer Inform 2007;2:25-35.
9. Slonim DK, Yanai I. Getting started in gene expression microarray analysis. PLoS Comput Biol 2009;5(10):e1000543.
10. Lee S, Stewart S, Nagtegaal I, Luo J, Wu Y, Colditz G, *et al.* Differentially expressed genes regulating the progression of ductal carcinoma *in situ* to invasive breast cancer. Cancer Res 2012;72(17):4574-86.
11. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, *et al.* NCBI GEO: Archive for functional genomics data sets – update. Nucleic Acids Res 2013;41:D991-5.
12. R Development Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria; 2012. Available from: https://www.r-project.org. [Last accessed on 2012 Jun 01].
13. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, *et al.* Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol 2004;5(10):R80.
14. Irizarry RA, Hobbs B, Collin F, Beazer‐Barclay YD, Antonellis KJ. Uwe Scherf, and Terence P. Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4(2):249-64.
15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43(7):e47.
16. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4(1):44-57.
17. Gene Ontology Consortium. Gene ontology consortium: Going forward. Nucleic Acids Res 2015;43:D1049-56.
18. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: Back to metabolism in KEGG. Nucleic Acids Res 2014;42:D199-205.
19. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61(2):69-90.
20. Appiah-Opong R, Asante IK, Safo DO, Tuffour I, Ofori-attah E, Uto T, *et al*. Cytotoxic effects of *Albizia zygia* (Dc) J.F. Macbr, a ghanaian medicinal plant, against human T-lymphoblast-like leukemia, prostate and breast cancer cell lines. Int J Pharm Pharm Sci [S.L.] 2016;8(5):392-6.
21. 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;136(5):E359-86.
22. Ali I, Wani WA, Saleem K. Cancer scenario in India with future perspectives. Cancer Ther 2011;8(56):56-70.
23. Ismail MM, Soliman DH, Farrag AM, Sabour R. Synthesis, antitumor activity, pharmacophore modeling and Qsar studies of novel pyrazoles and pyrazolo [1, 5-A] pyrimidines against breast adenocarcinoma Mcf-7 cell line. Int J Pharm Pharm Sci [S.L.] 2016;8(7):434-42.
24. Dhillon AS, Hagan S, Rath O, Kolch W. MAP kinase signalling pathways in cancer. Oncogene 2007;26(22):3279-90.
25. Kumar P, Bolshette NB, Jamdade VS, Mundhe NA, Thakur KK, Saikia KK, *et al*. Breast cancer status in India: An overview. Biomed Prev Nutr 2013;3(2):177-83.