British Journal of Pharmaceutical Research
15(3): 1-9, 2017; Article no.BJPR.32138
ISSN: 2231-2919, NLM ID: 101631759

SCIENCEDOMAIN *international*
www.sciencedomain.org

Human Drug Targets Identification in Breast Cancer by Computationally Based DNA Microarray Analysis

Syed Aun Muhammad[1*], Nighat Fatima[2], Hasan Afzaal[3], Ziaullah Shah[4] and Amjad Ali[5]

[1]Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University (BZU), Multan, Pakistan.
[2]Department of Pharmacy, COMSATS Institute of Information Technology, Abbottabad, 22060, Pakistan.
[3]Riphah Institute of Pharmaceutical Sciences, Riphah International University, Islamabad, Pakistan.
[4]Department of Biotechnology and Microbiology, Sarhad University of Science and Information Technology, Peshawar, Pakistan.
[5]Atta-ur-Rehman School of Applied Biosciences (ASAB), National University of Science and Technology, Islamabad, 44000, Pakistan.

Authors' contributions

This work was carried out in collaboration between all authors. Author SAM designed the study, collect the data and wrote the first draft of the manuscript. Authors NF and HA managed the analyses of the study. Authors ZS and AA managed the literature searches and proof read the article. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJPR/2017/32138
Editor(s):
(1) Rahul S. Khupse, Pharmaceutical Sciences, University of Findlay, USA.
(2) Rafik Karaman, Bioorganic Chemistry, College of Pharmacy, Al-Quds University, USA.
Reviewers:
(1) Asita Elengoe, Lincoln University College, Malaysia.
(2) César Luiz Da Silva Guimarães, Federal Universy of Rondônia State – Unir And Brazilian Institute of Environment and Natural Renewable Resources, Brazil.
(3) Nuket Ozkavruk Eliyatkın, Adnan Menderes University, Turkey.
(4) Yu Koyama, Niigata University Graduate School of Health Sciences, Japan.
(5) Carla Maria Ferreira Guerreiro da Silva Mendes, School of Health Portuguese Red Cross, Portuguese.
Complete Peer review History: http://www.sciencedomain.org/review-history/18457

Original Research Article

Received 9th February 2017
Accepted 5th March 2017
Published 1st April 2017

ABSTRACT

Aims: The availability of large cDNA datasets make it feasible to find new genetic variants. In this study, we focused to perform micro array differential analysis of breast cancer dataset to reveal genetic mutants of this disease.
Methodology: Human drug targets of breast cancer (BC) was found by comparing normal breast tissue samples and breast invasive cancer samples using GSE31138 DNA microarray dataset.

**Corresponding author: E-mail: aunmuhammad78@yahoo.com;*

The dataset was accessed from Gene Expression Omnibus (GEO) NCBI. The differential analysis was performed using "R" software and Bioconductor packages.
Results: In differential expressed genes (DEGs), LIFR and PSMD10 were significant BC-linked genes. Matrix analysis of these DEGs showed interdependencies between the probe levels of two groups. Gene ontology and interactomic analysis explored the functional enriched map. These critical protein targets are involved in ubiquitin-dependent protein degradation, cell morphogenesis, anti-apoptosis, and positive regulation of cell proliferation and their dysregulation are responsible for tumorigenesis.
Conclusion: These protein targets not only reveal the understandings about BC but can also progress into biological markers for diagnosis or treatment.

Keywords: Breast cancer; microarray; DEGs; functional analysis; PPI network.

1. INTRODUCTION

Genomics methods have transformed the way we do research in biology and medicine. We now can measure the majority of mRNAs, proteins, protein-protein interactions, genomic mutations, epigenetic alterations, and micro RNAs in a single experiment. Target-based drug discovery starts by recognizing important genes for which trepidation of activity can produce a required phenotype and find to discover the molecules that interact with these genes selectively to modify the disease state or symptoms in a positive way. Breast cancer (BC) is one of the most multifaceted diseases that accounts for 22.9% of all cancers in women worldwide. Prognosis rates of the disease vary and it depends on cancer type, stage, treatment, cultural and geographical sites [1]. Genetics and molecular based approaches showed that the pathological mechanism of BC is a developing procedure involving several stages and factors. However, the mechanism and full therapeutic role of these genes in pathophysiology, tumor growth rate, metastatic spread and molecular physiology of BC has not been entirely understood [2] and therefore substantial slits in clinical-results still raises questions, leading scientists to investigate the mechanisms for expansion and advancement [3,4].

Microarray technique is a valuable tool to find out variations in frequency and progress of cancer comprehensively [5]. Currently, diagnostic and treatment predictive biomarkers for BC are being established employing gene expression techniques. The gradual increase of data from gene expression studies on BC in public repositories provides a chance to construct pooled gene expression datasets containing a larger number of patients. Microarray based differential expression analysis of BC hold promise for future advances in diagnosis, treatment and prognosis. Therefore, in this study we accessed the breast cancer related microarray dataset, performed the differential and functional analysis which supports the hypothesis that differentially expressed genes could be valuable biomarker or therapeutic targets.

2. MATERIALS AND METHODS

2.1 Microarray Data

The CEL format files of DNA microarray dataset GSE31138 were accessed from NCBI-Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) [GEO: GSE31138], involving 03 breast invasive cancer samples and 03 normal samples. The genetic chip comprises GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix, Inc., Santa Clara, CA, 95051, USA, Technology: in situ oligonucleotide) and the functional annotation facts (hgu133plus2) of genetic probes were used to identify the genetic expression levels.

2.2 Analysis of Differential Expression

Raw and pheno-data files were changed into identifiable format, and redundancies are errors were removed [6]. The bioconductor packages of R software were used to execute the normalization and quality control [7,8,9]. RNA degradation analysis was done to assess the quality of RNA by using AffyRNAdeg, summary AffyRNAdeg, and plot AffyRNAdeg bioconductor packages on R software [10,11]. After normalization, statistical analysis on R software was executed to find the genetic variants (DEGs) by comparing normal with case samples and multiple testing corrections were completed by Benjamini-Hochberg method [12]. A false discovery rate (FDR) less than 0.05, p-value ≤0.05, Average Expression Level (AEL) ≥40%

and an absolute logFC>1 were considered as the substantial cutoffs [13].

2.3 Cluster Analysis

Cluster analysis [14] was performed based on expression data in individual sample to validate the variations in gene expression levels between BC tissue samples and normal samples.

2.4 Analysis of Functional Enrichment

Functional enrichment analysis uncovers the biological and cellular role of each gene in different physiological pathways [15]. The functional annotation of these DEGs was determined by using the online DAVID (Database for Annotation Visualization and Integrated Discovery) [16] and EMBL-EBI databases with p-value < 0.05 as the significant cutoff.

2.5 Construction of Protein-protein Interaction (PPI) Network

Proteins usually relate and interact each other to perform biological functions [17]. Therefore, interacting partners of the most important DEGs were anticipated, using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) [18] and HAPPI (Human Annotated and Predicted Protein Interaction) databases [19] and then the interaction networks of these aberrant genes with their interactors were constructed by using Cytoscape software [20].

3. RESULTS

3.1 Normalization and RNA Degradation Plots

The probe-level data that represent expression levels of genes was normalized (Fig. 1) and each gene in a probe set were arranged by location relative to the 5′-end of the sample RNA molecule. A side-by-side plot was produced by the function plot AffyRNAdeg (Fig. 2) and the function summary AffyRNAdeg produced a distinct statistic summary for individual array in the batch (Table 1), presenting an assessment of the severity of degradation and significance level.

Table 1. A summary statistic for each array in the batch, assessing the severity of RNA degradation and significance level

Arrays/Samples	Slope	P-value
GSM770819.CEL	5.74E+00	4.81E-07
GSM770820.CEL	5.16E+00	7.74E-07
GSM770821.CEL	7.76E+00	1.31E-07
GSM770822.CEL	5.90E+00	4.61E-10
GSM770823.CEL	5.69E+00	1.94E-09
GSM770824.CEL	6.81E+00	2.21E-07

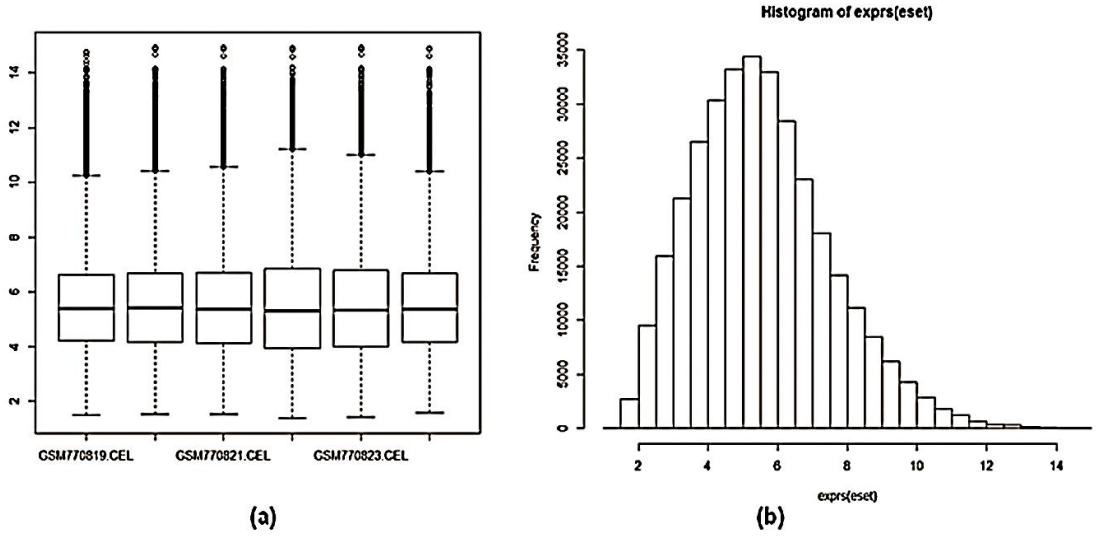

Fig. 1. Normalized probe-level gene expression data (a) Boxplot of gene expression data. The medians are almost at the same level, indicating high normalization performance (b) Histogram of log intensity

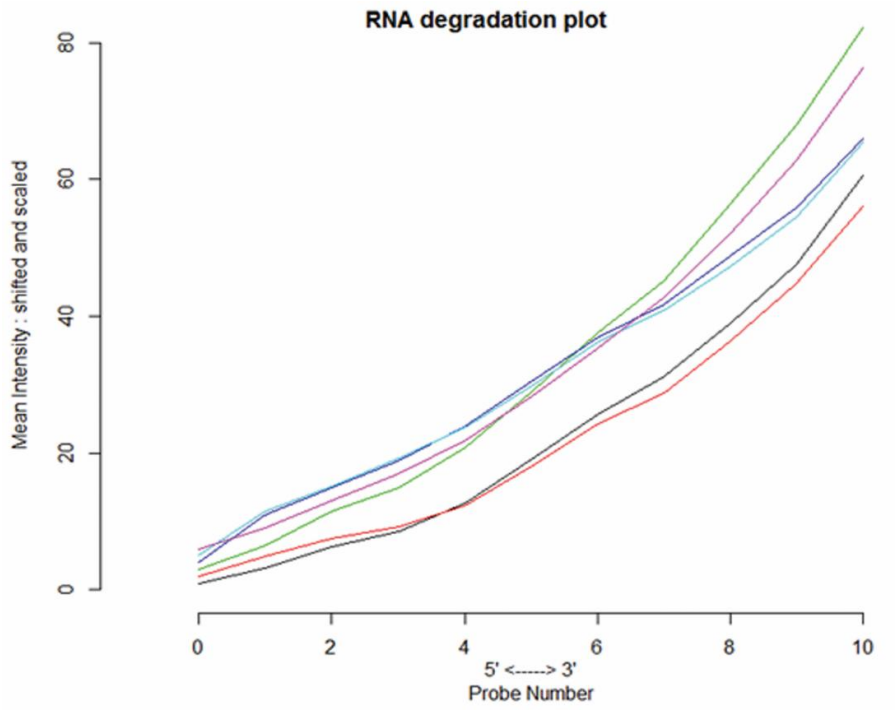

Fig. 2. Side-by-side plot produced by plotAffyRNAdeg representing 5' to 3' trend

3.2 Differentially Expressed Genes (DEGs)

A list of 50-DEGs were selected in BC samples associated with normal tissues, comprising 44 up-regulated and 6 down-regulated DEGs.

3.3 Cluster-analysis

In Fig. 3, cluster profile of gene expression level has been shown. The genetic expression of BC samples can be differentiated from the normal samples which indicated that there are understandable changes occurs between the comparative groups.

3.4 Analyzing Functional Enrichment

We performed the enrichment analysis of DEGs which showed that 10 and 16 terms for down-regulated and up-regulated genes respectively were significantly enriched. The total number of enriched terms was counted with their significant false discovery rate (Table 2).

3.5 Gene Expression Levels and Interaction Network

The expression levels of breast cancer associated differentially expressed genes PSMD10 (Uniprot ID: PSD10_Human), and LIFR (Uniprot ID: LIFR_Human) in each sample (normal and cancer) were determined. LIFR is down-regulated while PSMD10 gene is up-regulated in breast cancer samples as compared to normal expression levels (Fig. 4). The interacting molecules of these two source genes were obtained from HAPPI and STRING databases and their interaction-network were generated (Fig. 5). In total, 46, 14 and 11 genes were involved in the molecular network of PSMD10 (Uniprot ID: PSD10_Human), and LIFR (Uniprot ID: LIFR_Human) respectively. The network contained IL6RA_HUMAN, FST_HUMAN, STAT3_HUMAN, PPTN6_HUMAN, FSHB_HUMAN, MPIP1_HUMAN, ONCM_HUMAN, CNTFR_HUMAN and other genes.

4. DISCUSSION

Microarray data of normal and breast cancer samples (6 samples) were compared to identify the differentially expressed genes (DEGs). A total of 50 DEGs were obtained in BC samples. The functional enrichment analysis showed 10-downregulated and 16-upregulated genes contributing cancer progression. Cell proliferation, cell morphogenesis, leukemia inhibitory signaling and positive control of anti-apoptosis were expressively observed in the

down-regulated genes, whereas cell cycle process, mitosis, and control of cell death were enriched in the up-regulated genes. These factors are contributing in the progression of cancer [21]. During DEGs analysis, significant genes including PSMD10, and LIFR were screened that were curated to be involved in breast cancer. The expression levels of these genes in each sample showed that LIFR is down-regulated while PSMD10 genes are up-regulated in breast cancer samples as compared to normal samples.

In the regulation of cell proliferation and apoptosis, [22] identified LIFR as a breast cancer metastasis suppressor that functions through the HIPPO-YAP pathway. The aberrant function of this gene is associated with genetic instability and occurrence of breast cancer. PSMD10 has been reported to encode a regulatory factor of 26S proteasome which is essential for ubiquitin-dependent protein degradation. Similarly, this non-ATPase protein subunit has functional interaction with other proteins and therefore irregularity in the expression of this protein may progress to tumorigenesis [23].

Additionally, it has been observed that inhibin concentration in serum is associated with the enlargement of granulosa-cell tumors and can therefore be used as a prognostic biomarker for primary and recurring cancer [24,25]. These up and down regulated genes in a number of breast cancers might serve as therapeutic targets or diagnostic tools and these findings may provide useful future directions.

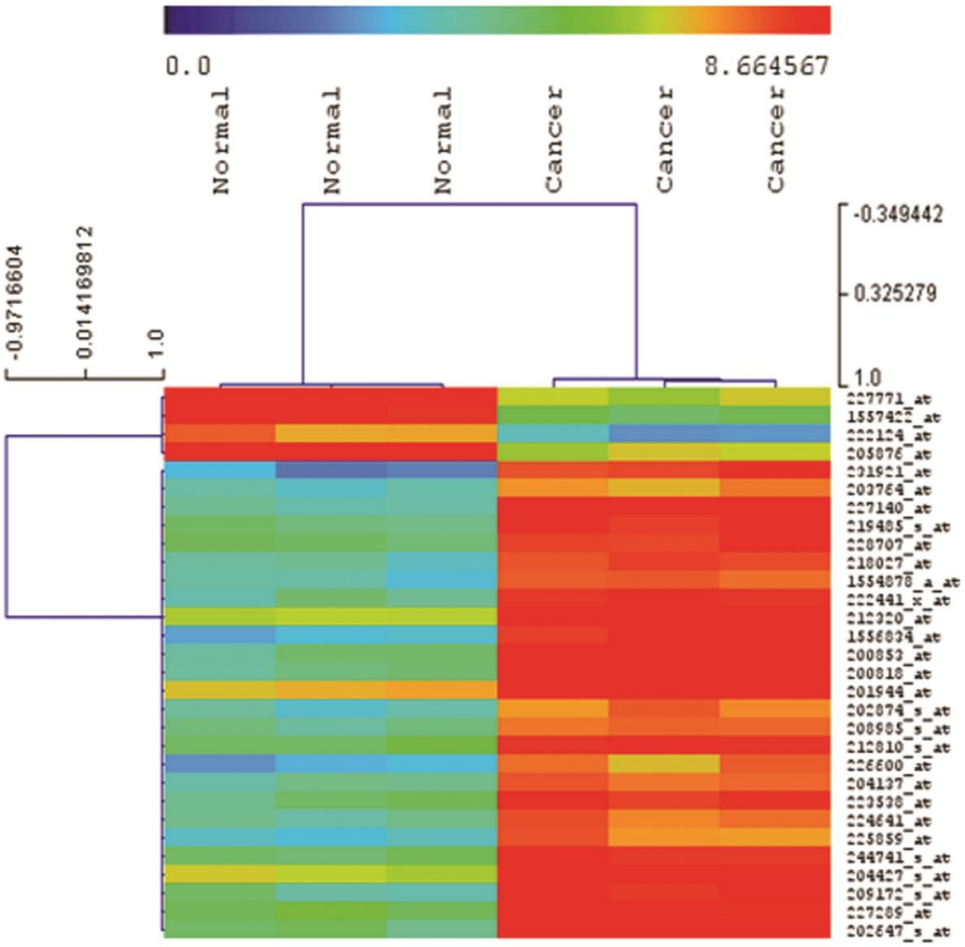

Fig. 3. Cluster analysis results for gene expression data. The expression values clustered in the red-shaded areas indicate over-expression, and the green-shaded areas indicate under-expression

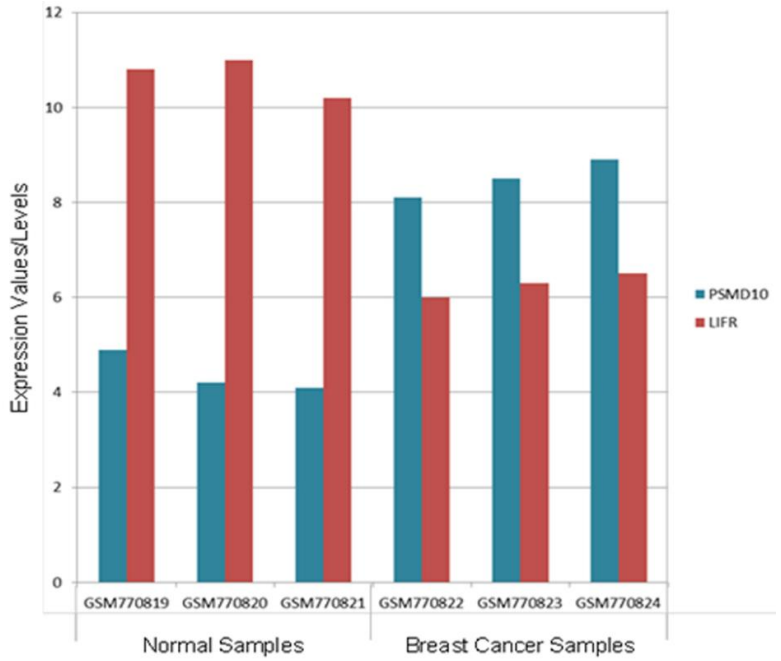

Fig. 4. Gene expression levels of PSMD10 and LIFR in each sample. LIFR is down-regulated while PSMD10 genes are up-regulated in breast cancer samples as compared to normal expression levels

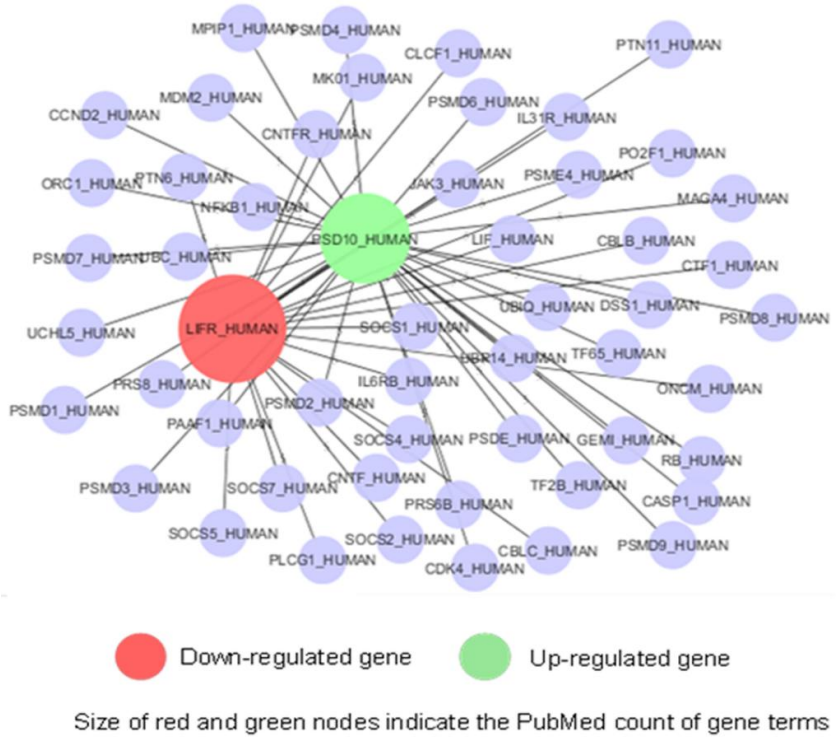

Fig. 5. Protein-Protein Interaction network of PSMD10 (Uniprot ID: PSD10_Human) and LIFR (Uniprot ID: LIFR_Human) based on HAPPY and STRING databases indicating high-confidence. Each node represents the protein while dark lines indicate the interaction

Table 2. Functional enrichment analysis of down-regulated and up-regulated differentially expressed genes (DEGs)

Serial #	GO* accession number	Count	FDR**
	Down-regulated DEGs		
1	[GO:0000902] Cell morphogenesis	18	2.63E-06
2	[GO:0007166] Cell surface receptor linked signal transduction	8	4.38E-06
3	[GO:0008284] Positive regulation of cell proliferation	20	5.71E-06
4	[GO:0010033] Response to organic substance	5	0.018694
5	[GO:0019221] Cytokine-mediated signaling pathway	8	7.80E-06
6	[GO:0030182] Neuron differentiation	9	8.34E-06
7	[GO:0032990] Cell part morphogenesis	12	8.72E-06
8	[GO:0042127] Regulation of cell proliferation	14	1.05E-05
9	[GO:0045768] Positive regulation of anti-apoptosis	10	1.12E-05
10	[GO:0048861] Leukemia inhibitory factor signaling pathway	20	1.18E-07
	Up-regulated DEGs		
1	[GO:0000278] Mitotic cell cycle	15	1.46E-05
2	[GO:0022402] Cell cycle process	5	0.004726
3	[GO:0043492] ATPase activity	8	2.63E-05
4	[GO:0007049] Cell cycle	10	7.80E-04
5	[GO:0022403] Cell cycle phase	10	6.46E-03
6	[GO:0050000] Chromosome localization	8	0.030496
7	[GO:0030071] Regulation of mitotic transition	12	5.71E-04
8	[GO:0007067] Mitosis	11	4.38E-04
9	[GO:0034622] Cellular macromolecular complex assembly	5	0.006950
10	[GO:0000279] M phase	6	0.014037
11	[GO:0051726] Regulation of cell cycle	8	0.014060
12	[GO:0042981] Regulation of apoptosis	10	1.84E-04
13	[GO:0043067] Regulation of programmed cell death	13	1.50E-03
14	[GO:0010941] Regulation of cell death	8	1.01E-02
15	[GO:0007059] Chromosome segregation	8	1.87E-03
16	[GO:0045786] Negative regulation of cell cycle	4	0.027306

*GO: Gene Ontology, **FDR: False Discovery Rate

The interaction network of the most significant genes (PSMD10 and LIFR) was constructed to analyze the interactors. It was shown that these genes are interacted with IL6_HUMAN [26], CDK4_HUMAN [27], STAT3_HUMAN [28], GRB2_HUMAN [29], and PLCG1_HUMAN [30] genes which have been reported to be involved in the progression of cancers and other diseases.

5. CONCLUSION

We found a number of DEGs through expression profiling of breast cancer samples with normal cases. The enrichment analysis particularly showed that PSMD10 and LIFR targets were closely associated with the pathogenesis. These genetic variants play significant roles in the development of BC. However, further genetic and experimental validation studies are required to specify their potential function in diagnostic and therapeutic applications.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. "World Cancer Report". International Agency for Research on Cancer; 2008. (Retrieved 2014-04-10)
2. Standis LJ, Sweet ES, Novack J. Breast cancer and the immune system. J. Soc. Integr. Oncol. 2008;6:158–168.

3. Sotiriou C. Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. J. Natl. Cancer Inst. 2006;98:262–272.
4. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. N. Engl. J. Med. 2009;360:790–800.
5. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, and Trent JM. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat. Genet. 1996;14:457–460.
6. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. Bioinformatics. 2001;17:520–525.
7. Bolstad BM, Irizarry RA., Astrand M and Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19:185-193.
8. Fujita A, Sato JR, de Oliveira Rodrigues L, Ferreira CE, Sogayar MC. Evaluating different methods of microarray data normalization. BMC Bioinformatics. 2006; 7:469.
9. Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M. Variant annotation: A bioconductor package for exploration and annotation of genetic variants. Bioinformatics. 2014;30: 2076-8.
10. Affymetrix. Affymetrix microarray suite user guide. Affymetrix, Santa Clara, CA. version 4th Edition; 1999.
11. Affymetrix. Affymetrix microarray suite user guide. Affymetrix, Santa Clara, CA. version 5th Edition; 2001.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 1995; 57:289–300.
13. Jin Y, Da W. Screening of key genes in gastric cancer with DNA microarray analysis. Europ. J. Med. Res. 2013;18: 37.
14. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA. 1998;95:14863–14868.
15. Nam D, Kim SY. Gene-set approach for expression pattern analysis. Brief Bioinform. 2008;9:189–197.
16. Da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 2009;4:44–57.
17. Li S, Armstrong CM, Bertin N, Ge H, Milstein S. A map of the interactome network of the metazoan *C. elegans*. Science. 2014;303:540–543.
18. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, and von Mering C. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011;39:D561–D568.
19. Chen JY, Mamidipalli SR, and Huan T. HAPPI: An online database of comprehensive human annotated and predicted protein interactions. BMC Genomics. 2009;10(Suppl 1):S16.
20. Cline MS, Smoot M, Cerami. Integration of biological networks and gene expression data using Cytoscape. Nat. Protoc. 2007;2: 2366-82.
21. Hunter T, Pines J. Cyclins and cancer II: Cyclin D and CDK inhibitors come of age. Cell. 1994;79:573–582.
22. Andreas T, Andrea K, Ingo K, Heinrich PC, Gerhard MN. A functional role of the membrane-proximal extracellular domains of the signal transducer gp130 in heterodimerization with the leukemia inhibitory factor receptor. Eur. J. Biochem. 2002;269:2716–262.
23. Lee H, Choi AJ, Kang GY, Park HS, Kim HC, Lim HJ et al. Increased 26S proteasome non-ATPase regulatory subunit 1 in the aqueous humor of patients with age-related macular degeneration. BMB. Rep. 2014;47:292-7.
24. Suzuki Y, Yoshitomo K, Maruyama K, Suyama A, Sugano S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene. 1997;200:149-156.
25. Available:www.ncbi.nlm.nih.gov/pubmed, www.ncbi.nlm.nih.gov/gene/3624 (Accessed on 2014-04-12)
26. Zilberstein A Ruggieri R, Korn JH, Revel M. Structure and expression of cDNA and genes for human interferon-beta-2, a distinct species inducible by growth-stimulatory cytokines. EMBO J. 1986;5: 2529-2537.
27. Elkahloun AG, Krizman DB, Wang Z, Hofmann TA, Roe BA, and Meltzer PS.

Transcript mapping in a 46-kb sequenced region at the core of 12q13.3 amplification in human cancers. Genomics. 1997;42: 295-301.

28. Yamamoto T, Sekine Y, Kashima K, Kubota A, Sato N, Aoki N, Matsuda T. The nuclear isoform of protein-tyrosine phosphatase TC-PTP regulates interleukin-6-mediated signaling pathway through STAT3 dephosphorylation. Biochem. Biophys. Res. Commun. 2002; 297:811-817.
29. Pandey P, Kharbanda S, Kufe D. Association of the DF3/MUC1 breast cancer antigen with Grb2 and the Sos/Ras exchange protein. Cancer Res. 1995;55: 4000-4003.
30. Tacconelli A, Farina AR, Cappabianca L, Desantis G, Tessitore A, Vetuschi A, Sferra R, Rucci N, Argenti B, Screpanti I, Gulino A, Mackay AR. TrkA alternative splicing: A regulated tumor-promoting switch in human neuroblastoma. Cancer Cell. 2004;6:347-360.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/18457