International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 2, No. 4 (2013): 147-151
Research Article
Open Access

ISSN: 2320-0634

Identification and in silico observation of potential biomarkers in breast cancer region using bioinformatics approaches

Jyotsana Pandey[1]*, Alka Dwivedi[1], Vijay Laxmi Saxena[2] and Bipin Kumar Tripathi[3]

[1] *Department of Computer Science, Singhania University, Rajasthan-335515*
[2] *National Bioinformatics Infrastructure Facility Centre of (DBT) Ministry of Science and Technology (Govt. of India), D.G (PG) College civil lines Kanpur*
[3] *Department Of Computer Science, HBTI, Kanpur-208002*

*Corresponding author: Jyotsana Pandey; e-mail: imjyotsanapandey@gmail.com, alkabioinfo964@gmail.com

Received: 30 April 2013 Accepted: 23 May 2013 Online: 01 July 2013

ABSTRACT

Identification of *in silico* observation and development of high through put technologies are easier to predict cancer in early stages. So there is a need to establish biomarkers for prediction and detection of breast cancer at first stage. A biomarker is an indicator of a biological state in this project. In bioinformatics it can be used to study cancers to identify novel gene/protein which can be used as biomarker. Bioinformatics approaches can be used in effective identification of biomarkers. In this study a combined advance approach using gene expression analysis and bioinformatics were implemented to identify a novel biomarker for breast cancer. In this study we predicted gene ADIPOQ, associated with breast fibroepithelial neoplasia, CD36 and GHR associated with precancerous condition intraepithelial neoplasia. ADIPOQ and PTPN are associated with Myelodysplastic syndrome. All these marker genes can be used in prognosis of specific conditions of cancer.

Keywords: Breast Cancer, High throughput data analysis, Microarray Data Analysis, Biomarkers

INTRODUCTION

Cancer is a term used for diseases in which abnormal cells divide without control and are able to invade other tissues [1]. As we see from different surveys and news that Cancer caused 5,56,400 deaths in the country in 2010 and 71 per cent of those were middle age group i.e. aged between 30 and 69 years according to findings of the Million Death Study. WHO also says that around 13% of all deaths caused by cancer in 2008. The incidence of cancer is increasing in economically developing countries as a result of population aging and growth as well as, increasingly, an adoption of cancer-associated lifestyle choices. It can be interpreted by the fact that about 70% of all cancer deaths in 2008 occurred in low- and middle-income countries. Deaths from cancer worldwide are projected to continue to rise to over 13.1 million in 2030 [3]. This data not only provide us current status but also signifies the need to work upon it. Breast cancer is a malignant tumor that starts from cells of the breast. A malignant tumor is a group of cancer cells that may grow into (invade) surrounding tissues or spread (metastasize) to distant areas of the body. The disease occurs almost entirely in women, men can get it, too [4].

Breast cancer is the second leading cause of cancer death in women, exceeded only by lung cancer. Cancer is now appreciated as not only a highly heterogeneous pathology with respect to cell type and tissue origin but also as a disease involving deregulation of multiple pathways governing fundamental cell processes such as death, proliferation, differentiation and migration. So, it is important to study pathways involved in cancer, changes in them during cancer and their role in developing cancer. Here, the role of gene expression analysis comes into play to identify the differentially expressed genes in cancer. To study the pathways and its various components we use Systems biology. Systems biology studies biological systems by systematically perturbing them (biologically,

genetically, or chemically); monitoring the gene, protein, and informational pathway responses; integrating these data; and ultimately, formulating mathematical models that describe the structure of the system and its response to individual perturbations. Biomarkers represent major tools in determining the presence of cancer, its progression and the responses to treatments [5]. In cancer, a biomarker refers to a substance or process that is indicative of the presence of cancer in the body. The comparison of gene expression profiles of diseased vs. normal state of cancer can be used to find out differentially expressed genes in breast cancer which can be further implicated in network, to identify the gene which can be used as biomarker [10].

MATERIALS AND METHODS

Study material: NCBI (national centre of biotechnology information)

Databases: the Array Express archive is a database of functional genomics experiments including gene expression where you can query and download data collected to MIAME and MINSEQE standards. ID-E-GEOD-15852**Title**-Transcription profiling of human breast tumors and their paired normal tissues Assay-86, Organism-Homo sapiens, Date-2009-05-02

2.3Softwares

BRB-Array Tools (version 4.2.0)

It is integrated software package for the analysis of DNA microarray data. It was developed by the Biometric Research Branch of the Division of Cancer Treatment & Diagnosis of the National Cancer Institute [13]. It contains utilities for processing expression data from multiple experiments, visualization of data, multidimensional scaling, clustering of genes and samples, and classification and prediction of samples. BRB-Array Tools is used to analyze the microarray data taken from array express (E-GEOD-15852).

Cytoscape (version 2.7)

It is a project dedicated to building open-source network visualization and analysis software. Software "Core" provides basic functionality to layout and query the network and to visually integrate the network with state data. The Core is extensible through a plug-in architecture, allowing rapid development of additional computational analyses and features. Cytoscape is used to create relationship between the genes filtered through microarray data [7].

Reactome FI plugin

Reactome FI Cytoscape Plugin was designed to find network patterns related to cancer and other types of diseases. Reactome FI plugin is used to create network, finding modules and cancer gene annotation of the genes involved in network [8].The analysis starts with collection of data from Array Express database. E-GEOD-15852 archive of microarray data is taken as a sample. It includes patients of Malaysian, Indian and Chinese origin. Out of these three, only with malaises patients' origin are taken having infiltrating ductal carcinoma. This data is now analyzed using BRB-Array by GC-rma method using 2 fold difference and p-value >0.01.after that scatter plots are analyzed to identify Marker up regulated genes with the help of MS Access. These Marker genes are now analyzed using Cytoscape and Cytoscape plugin-Reactome FI.

RESULTS AND DISCUSSION

Microarray data analysis

Micro array data analysis of 25 patients (11 patients with class 2 carcinoma,12 patients with class 3 carcinoma and 2 patients with class 1 carcinoma) of Malaysian origin(taken from E-GEOD-15852 data) is done with the help of BRB-Array tool at the parameters of 2-fold expression and P-value =0.01. Scatter plots are used to retrieve up regulated genes from each tumor vs. normal array plot. These up regulated genes are analyzed using MS Access to identify Marker gene among different classes of infiltrating ductal carcinoma. This lead to identification of 26 Marker genes for Class I infiltrating ductal carcinoma, 4 Marker genes for Class II infiltrating ductal carcinoma and 2 Marker genes for Class III infiltrating ductal carcinoma (Table 1). For class 2 infiltrating ductal carcinoma age group of 45 to 60 year are considered with reference of occurrence of breast cancer in post-menopausal cases and age as a risk factor for breast cancer (Table 2).

Table 1. Marker genes in class 1, class 2 and class 3 breast cancer

Marker genes in class 1	Marker genes in class 2	Marker genes in class 3
1. ABCA8	1. ACACB	1. TNS1
2. ACSL1	2. ADIPOQ	2. LPL
3. ADH1B	3. ECM2	
4. ADIPOQ	4. SORBS1	
5. CAV2		
6. CFD		
7. CTNNA1		
8. DPT		
9. ECM2		
10. FBN1		
11. FOSB		
12. GBE1		
13. GYG2		
14. HBB		
15. LEP		
16. LGALS1		
17. MDH2		
18. PCOLCE2		
19. PRKAR2B		
20. RBP4		
21. SORBS1		
22. DUSP1		
23. SPARC		
24. MSN		
25. SCD		
26. SRPX		

Marker genes in class 2 breast cancer

Network of Marker genes is constructed using Cytoscape plug-in Reactome FI which excluded not linked genes and included genes which participate directly in network.

Different pathways involved in Marker genes

When the pathway enrichment of above 24 genes is done using Cytoscape plug-in Reactome FI, they are involved in 78 pathways. (Table 3) and the table IV explains Ranking of different modules and nodes.

Table 2. Marker genes for class 2 age group (45-60) (8 patients)

S. No.	Gene	Protein
1.	ABCA8	ATP-binding cassette sub-family A member 8
2.	ACACB	Acetyl-CoA carboxylase 2
3.	ACSL1	Long-chain-fatty-acid--CoA ligase 1
4.	ADIPOQ	Adiponectin precursor
5.	AKR1C2	
6.	ALDH2	Aldehyde dehydrogenase, mitochondrial precursor
7.	CAV2	Caveolin-2
8.	CD36	Platelet glycoprotein 4
9.	CFD	Complement factor D precursor
10.	DPT	Dermatopontin precursor
11.	ECM2	Extracellular matrix protein 2 precursor
12.	GBE1	1,4-alpha-glucan-branching enzyme
13.	GHR	Growth hormone receptor precursor
14.	GPX3	Glutathione peroxidase 3 precursor
15.	KANK1	KN motif and ankyrin repeat domain-containing protein 1
16.	LEP	Leptin precursor
17.	PCOLCE2	Procollagen C-endopeptidase enhancer 2 precursor
18.	PLA2G16	Group XVI phospholipase A2
19.	PRKAR2B	cAMP-dependent protein kinase type II-beta regulatory subunit
20.	QKI	Protein quaking
21.	RARRES2	Retinoic acid receptor responder protein 2 precursor
22.	RBP4	retinol binding protein 4, plasma
23.	RETSAT	All-trans-retinol 13,14-reductase precursor
24.	SERPINF1	Pigment epithelium-derived factor precursor
25.	SORBS1	sorbin and SH3 domain containing 1

The analysis is done for patients of age group 45-60 because the risk of breast cancer is high in post-menopausal woman. Identification of Marker genes for this particular group results in 24 genes. Out of these 24 genes 2 of them (SERPINF1 and GBE1) are requested for patent of breast cancer diagnostic array and another 2 genes (QKI and PLA2G16) are requested for patent in diagnostic array of thyroid cancer and small cell lung cancer. Reducing these four genes now we can target 20 genes which can be used in further study. The network of these genes was made with the help of Cytoscape to find out connectivity between these genes.

Table 3. Pathway enrichment of network of Marker genes

Gene set	Ratio of protein in Gene set	No. of protein in gene set	Protein in gene set	P-value	FDR	Nodes
Adipocytokine signaling pathway(K)	0.0089	67	4	0	1.00E-03	ACACB,ADIPOQ,ACSL1,CD36
PPAR signaling pathway(K)	0.0092	69	4	0	5.00E-04	ADIPOQ,ACSL1,CD36,SORBS1
Activated AMPK stimulates fatty-acid oxidation in muscle(R)	0.0021	16	2	0.0002	1.90E-02	ACACB,ACSL1
Insulin signaling pathway(K)	0.0182	137	3	0.0009	6.10E-02	ACACB,PRKAR2B,SORBS1
Muscarinic acetylcholine receptor 2 and 4 signaling pathway(P)	0.0008	6	1	0.0087	2.45E-01	PRKAR2B
Fatty acid biosynthesis(K)	0.0008	6	1	0.0087	2.45E-01	ACACB
Formation of Platelet plug(R)	0.0157	118	2	0.0123	2.98E-01	CD36,CFD
rho-selective guanine exchange factor akap13 mediates stress fiber formation(B)	0.0013	10	1	0.0145	2.90E-01	PRKAR2B
alternative complement pathway(B)	0.0013	10	1	0.0145	2.90E-01	CFD
phospholipase c-epsilon pathway(B)	0.0015	11	1	0.016	2.85E-01	PRKAR2B
protein kinase a at the centrosome(B)	0.0015	11	1	0.016	2.85E-01	PRKAR2B
akap95 role in mitosis and chromosome dynamics(B)	0.0016	12	1	0.0174	2.77E-01	PRKAR2B
attenuation of gpcr signaling(B)	0.0016	12	1	0.0174	2.77E-01	PRKAR2B
repression of pain sensation by the transcriptional regulator dream(B)	0.0017	13	1	0.0188	2.50E-01	PRKAR2B
transcription regulation by methyltransferase of carm1(B)	0.0017	13	1	0.0188	2.50E-01	PRKAR2B
Metabotropic glutamate receptor group II pathway(P)	0.0017	13	1	0.0188	2.50E-01	PRKAR2B
Hedgehog signaling pathway(P)	0.0019	14	1	0.0203	2.67E-01	PRKAR2B
akt signaling pathway(B)	0.0021	16	1	0.0232	3.04E-01	GHR
nitric oxide signaling pathway(B)	0.0023	17	1	0.0246	3.00E-01	PRKAR2B
gata3 participate in activating the th2 cytokine genes expression(B)	0.0023	17	1	0.0246	3.00E-01	PRKAR2B
cystic fibrosis transmembrane conductance regulator (cftr) and beta 2 adrenergic receptor (b2ar) pathway(B)	0.0025	19	1	0.0274	3.31E-01	PRKAR2B
regulation of bad phosphorylation(B)	0.0027	20	1	0.0289	3.37E-01	PRKAR2B
how progesterone initiates the oocyte maturation(B)	0.0029	22	1	0.0317	3.17E-01	PRKAR2B
growth hormone signaling pathway(B)	0.0029	22	1	0.0317	3.17E-01	GHR
stathmin and breast cancer resistance to antimicrotubuleagents(B)	0.0029	22	1	0.0317	3.17E-01	PRKAR2B

regulation of ck1/cdk5 by type 1 glutamate receptors(B)	0.0029	22	1	0.0317	3.17E-01	PRKAR2B
signaling pathway from g-protein families(B)	0.0029	22	1	0.0317	3.17E-01	PRKAR2B
regulation of eif-4e and p70s6 kinase(B)	0.0031	23	1	0.0331	3.09E-01	GHR
transcription factor creb and its extracellular signals(B)	0.0031	23	1	0.0331	3.09E-01	PRKAR2B
mtor signaling pathway(B)	0.0032	24	1	0.0345	3.16E-01	GHR
activation of camp-dependent protein kinase pka(B)	0.0035	26	1	0.0374	3.23E-01	PRKAR2B
mcalpain and friends in cell motility(B)	0.0035	26	1	0.0374	3.23E-01	PRKAR2B
Propanoate metabolism(K)	0.0044	33	1	0.0472	3.89E-01	ACACB
trefoil factors initiate mucosal healing(B)	0.0044	33	1	0.0472	3.89E-01	GHR
Glucagon signaling in metabolic regulation(R)	0.0045	34	1	0.0486	3.87E-01	PRKAR2B
chrebp regulation by carbohydrates and camp(B)	0.0047	35	1	0.05	3.89E-01	PRKAR2B
actions of nitric oxide in the heart(B)	0.0049	37	1	0.0528	3.95E-01	PRKAR2B
Proteogylcansyndecan-mediated signaling events(N)	0.0355	267	2	0.056	4.12E-01	CAV2,SORBS1
Pyruvate metabolism(K)	0.0053	40	1	0.057	4.04E-01	ACACB
activation of csk by camp-dependent protein kinase inhibits signaling through the t cell receptor(B)	0.0056	42	1	0.0597	4.01E-01	PRKAR2B
Fatty acid metabolism(K)	0.0056	42	1	0.0597	4.01E-01	ACSL1
Type II diabetes mellitus(K)	0.0062	47	1	0.0666	4.33E-01	ADIPOQ
Malaria(K)	0.0068	51	1	0.0721	4.58E-01	CD36
Starch and sucrose metabolism(K)	0.007	53	1	0.0748	4.61E-01	GBE1
Steroid hormone biosynthesis(K)	0.0074	56	1	0.0789	4.74E-01	AKR1C2
Endothelin signaling pathway(P)	0.0078	59	1	0.083	4.85E-01	PRKAR2B
Complement and coagulation cascades(K)	0.0092	69	1	0.0964	5.40E-01	CFD
Syndecan-2-mediated signaling events(N)	0.0093	70	1	0.0977	5.35E-01	CAV2
Metabolism of xenobiotics by cytochrome P450(K)	0.0094	71	1	0.0991	5.30E-01	AKR1C2
Bacterial invasion of epithelial cells(K)	0.0096	72	1	0.1004	5.24E-01	CAV2
Adherens junction(K)	0.0098	74	1	0.1031	5.24E-01	SORBS1
Peroxisome(K)	0.0104	78	1	0.1083	5.35E-01	ACSL1
Opioid Signalling(R)	0.0105	79	1	0.1097	5.27E-01	PRKAR2B
G2/M Transition(R)	0.011	83	1	0.1149	5.36E-01	PRKAR2B
ECM-receptor interaction(K)	0.0112	84	1	0.1162	5.31E-01	CD36
Apoptosis(K)	0.0117	88	1	0.1214	5.38E-01	PRKAR2B
Hematopoietic cell lineage(K)	0.0118	89	1	0.1227	5.35E-01	CD36
Innate Immunity Signaling(R)	0.0138	104	1	0.142	5.77E-01	CFD
Metabolism of lipids and lipoproteins(R)	0.0141	106	1	0.1446	5.73E-01	ACACB
Metabolism of carbohydrates(R)	0.0144	108	1	0.1471	5.69E-01	GBE1
Arf6 signaling events(N)	0.0185	139	1	0.1855	6.21E-01	SORBS1
Arf6 trafficking events(N)	0.0185	139	1	0.1855	6.21E-01	SORBS1
Insulin Pathway(N)	0.0185	139	1	0.1855	6.21E-01	SORBS1
EGFR1(C)	0.0198	149	1	0.1976	6.26E-01	CAV2
IGF1 pathway(N)	0.0199	150	1	0.1988	6.18E-01	SORBS1
Jak-STAT signaling pathway(K)	0.0206	155	1	0.2047	6.19E-01	GHR
Phagosome(K)	0.0209	157	1	0.2071	6.11E-01	CD36
Metabolic pathways(K)	0.1489	1120	3	0.2179	6.11E-01	ACACB,ACSL1,GBE1
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway(P)	0.0223	168	1	0.22	6.06E-01	PRKAR2B
Signaling events activated by Hepatocyte Growth Factor Receptor (c-Met)(N)	0.0225	169	1	0.2212	5.99E-01	SORBS1
Signalling by NGF(R)	0.023	173	1	0.2258	5.95E-01	PRKAR2B
Syndecan-1-mediated signaling events(N)	0.0235	177	1	0.2305	5.90E-01	SORBS1
Focal adhesion(K)	0.0266	200	1	0.2566	6.00E-01	CAV2
Endocytosis(K)	0.0271	204	1	0.261	5.94E-01	CAV2
Cytokine-cytokine receptor interaction(K)	0.0354	266	1	0.3271	6.17E-01	GHR
Neuroactive ligand-receptor interaction(K)	0.0362	272	1	0.3332	6.13E-01	GHR
Plasma membrane estrogen receptor signaling(N)	0.0412	310	1	0.3706	6.16E-01	SORBS1

Pathway enrichment analysis showed that these genes are involved in 78 different pathways. Out of these 78 pathway involved in cancer are PPAR signaling, focal adhesion pathway and EGFR1 signaling pathway(which were already established) includes ADIPOQ, ACSL1, CD36, SORBS1 and CAV2 genes .Out of these genes CD36 and CAV2 are important because CAV2 is already a candidate marker and CD36 is directly involved in PPAR signaling and connected with PPAR itself. Also CD36 is reported to involve in different cancers and its expression level is weak in normal tissue and moderate to strong in cancerous tissue of breast. When module analysis of network is done six modules are identified in it. Of these six modules, module 0 and module 1 are the most important modules because they had maximum number of nodes and module 1 include PPAR signaling genes in it which is linked with cancer specifically in mammary gland tumor development. When cancer gene index annotations are done for the network it is found that CD36 is only gene in the

network which is associated with benign breast cancer and also involved in breast cancer malignancies. So, it can be used as a novel diagnostic biomarker in breast cancer. CD36 is also associated with Breast Adenoma (tubular adenoma) and precancerous condition named as intraepithelial neoplasia.

Table 4. Ranking of different modules and nodes in them

Module	Nodes in module	Node percentage	Node list
0	9	0.3103	ACACB,AKR1C2,DPT,GBE 1,HIF1A,POU2F1,PRKAG1 ,SP1,TGFB1
1	9	0.3103	ACSL1,ADIPOQ,CD36,FAB P1,PPARA,PPARG,RBP4,R XRB,SORBS1
2	5	0.1724	BMP1,CAV2,CFD,COL1A1, ITGB1
3	4	0.1379	GHR,IRS1,PRKAR2B,PTP N7
4	2	0.069	RARG,SERPINF1

Some other important genes which linked with specific conditions are: **ADIPOQ-** It is associated with breast fibroepithelialneoplasia and fibroadenoma **CD36 and GHR-** Both of them are associated with precancerous condition Intraepithelialneoplasia. **CD36, ADIPOQ and PTPN** – They are associated with Myelodysplastic syndrome.So, these can be used in prognosis of specific conditions.

CONCLUSION

From this analysis we can conclude that CD36 gene can be used in diagnosis of breast cancer and ADIPOQ can be used in prognosis of specific breast cancer conditions. Besides this PTPN and GHR can be used in prognosis of Myelodysplastic syndrome and with precancerous condition intra epithelial neoplasia respectively.

ACKNOWLEDGEMENTS

I want to pay my sincere thanks to my principal investigator Dr. Bipin kumar Tripathi, Assosiate Professor, Dept of computer science, HBTI Kanpur who reviewed & encouraged my work.

REFERENCES

1. AebersoldRuedi, Auffray Charles et al. (2008)."How Systems Biology Can Advance Cancer Research". Report on EU–USA Workshop:567-89.
2. Marion T Weigel and Mitch Dowsett. (2010).Current and emerging biomarkers in breast cancer: prognosis and prediction, Endocrine-Related Cancer, 17 R245–R262
3. Mishra Alok. &VermaMukesh .(2010) . Cancer Biomarkers: Are We Ready for the Prime Time? , Cancers 2010, 2, 190-208;
4. Polanski Malu and Anderson N. Leigh.(2006). A List of Candidate Cancer Biomarkers for Targeted Proteomics ,Biomarker Insights,1: 1–48.
5. PtitsynAndrey A, Weil Michael M and Thamm Douglas H (2008)Systems biology approach to identification of biomarkers for metastatic progression in cancer BMC Bioinformatics , 9(Suppl 9):S8.
6. Saez Enrique, Rosenfeld John, Livolsi Antonia, et al. (2004). PPARg signaling exacerbates mammary gland tumor development , Genes Dev., 18: 528-540
7. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. (2003)Cytoscape User Manual Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* Nov; 13(11):2498-504.
8. Simon Richard, Lam Amy, Li Ming-Chung, Ngan Michael, MenenzesSupriya and Zhao Yingdong (2007) Analysis of Gene Expression Data Using BRB-Array Tools, Cancer Informatics , 3 :11–17.
9. SrinivasPothur R., VermaMukesh, Zhao Yinming and SrivastavaSudhir. (2002). Proteomics for Cancer Biomarker Discovery, Clinical Chemistry 48:8 1160–1169.
10. Weissleder AR, Ntziachristos V. (2003). Shedding light onto live molecular targets. Nat Med; 9 : 123-8.
11. Sawyers CL. (2008).The cancer biomarker problem. Nature; 452 : 548-52.

*****