

Relevance of systems biological approach in the differential diagnosis of invasive lobular carcinoma & invasive ductal carcinoma

PK Rangunath^{1*}, B Vanaja Reddy², PA Abhinand¹ & Shiek SSJ Ahmed³

¹Department of Bioinformatics, Sri Ramachandra University, Porur, Chennai - 600 116, India; ²Department of Bioinformatics, Sathyabama University, Jeppiar Nagar Chennai - 600 119, India; ³Computational Neuroscience Laboratory, Indian Institute of Technology, Madras. Tamil Nadu, India; PK Rangunath - E-mail: sru.bioinformatics.research2@gmail.com; Phone: +919841351069, +91-44-45928603; *Corresponding author

Received April 02, 2012; Accepted April 27, 2012; Published April 30, 2012

Abstract:

Breast cancer is a malignant neoplasm originating from breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk. ILCs and IDCs vary from each other with respect to various histological, biological and clinical features. Remarkably, ductal tumors tending to form glandular structures, whereas lobular tumors are less cohesive and tends to invade in single file. The high degree of similarity in the prognoses of IDC and ILC makes it beneficial to develop a differential diagnostic protocol to classify the two conditions. The main goal of the study is to construct the genetic regulatory network from the microarray data using biological knowledge and constraint-based inferences, in order to explore the potential significant gene regulatory networks that can differentiate IDC and ILC and thereby understand the complex interactions that are influenced by the genetic networks. Out of the 54676 genes present on the GPL570 platform- 29 genes exhibited 4 fold up regulation in case of IDC and 22 in the case of ILC. The ductal and lobular tumors displayed a striking difference in the expression of genes associated with cell adhesion, protein folding, and protein phosphorylation and invasion. Construction of separate gene regulation networks for IDC and ILC on the basis of gene expression alteration can be utilized in understanding the distinction in the possible mechanism that underlies the pathological differences between the two, which can be exploited in identifying diagnostic or therapeutic targets.

Key Words: Invasive (or infiltrating) ductal carcinoma (IDC), Invasive lobular carcinoma (ILC), Differential diagnosis, Gene expression profiling, Pathoinformatics, Systems biology, Gene Networks.

Abbreviations: ILC: Invasive lobular c carcinoma, IDC: invasive ductal carcinoma, ER: Estrogen Receptor

Background:

Cancer is an abnormal growth of cells caused by multiple changes in gene expression leading to dysregulated balance of cell proliferation and cell death and ultimately evolving into a population of cells that can invade tissues and metastasize to

distant sites, causing significant morbidity and, if untreated, death of the host.

Breast cancer is a malignant neoplasm originating from breast tissue, most commonly from the inner lining of milk ducts or

the lobules that supply the ducts with milk [1]. Highly complex and heterogeneous nature of the disorder makes it exceedingly difficult to analyze and understand the disease in a comprehensive manner, in spite of strenuous efforts. Breast cancer is the second leading cause of cancer deaths in women today (after lung cancer) and is the most common cancer among women, excluding nonmelanoma skin cancers. According to the American Cancer Society, about 1.3 million women will be diagnosed with breast cancer annually worldwide and about 465,000 will die from the disease [2]. Breast MRI, biopsy, ultrasound, CT scan, Mammography, lymph node biopsy are the most common protocols employed in the diagnosis of breast cancer.

Out of ~ 20 pathological types that have been defined - invasive ductal (IDCs) and invasive lobular carcinomas (ILCs) are the most common malignancies of the breast. Invasive ductal and lobular breast carcinomas account for 80% and 15% of all invasive breast tumors, respectively [3]. Invasive (or infiltrating) ductal carcinoma (IDC) starts in a milk passage (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 metastasize to other parts of the body through the lymphatic system and bloodstream, whereas Invasive lobular carcinoma (ILC) starts in the milk-producing glands (lobules) and subsequently it can metastasize to other parts of the body.

ILCs and IDCs vary from each other with respect to various histological, biological and clinical features. Remarkably, ductal tumors tending to form glandular structures, whereas lobular tumors are less cohesive and tends to invade in single file. This feature has been associated with the frequent inactivation of the E-cadherin gene (CDH1) [4]. ILCs are predominantly estrogen receptor (ER), and progesterone receptor (PR) positive, and thus presumably more homogeneous than IDCs. Their pathological grade is generally lower than that of IDCs and they exhibit lower proliferation index [5]. ILCs are less sensitive to chemotherapy [6] and are more prone to form bone, gastrointestinal, peritoneal and ovarian metastases than IDCs [7]. They also have lower vascular endothelial growth factor expression [8]. Despite these differences, ILCs show similar prognoses as IDCs [9], and the diagnosis and treatment of ILCs and IDCs is similar.

A differential diagnosis is a systematic method used to identify unknowns. It is essentially a process of elimination used medical professionals to diagnose the specific disease in a patient. The high degree of similarity in the prognoses of IDC and ILC makes it beneficial to develop a differential diagnostic protocol to classify the two conditions. Therefore, it is crucial to gain deep insight into the molecular differences that distinguish the two pathological types in order to attempt differential diagnosis and to tailor specific treatment methods.

A few microarray studies have been performed to identify the differential gene expression between IDCs and ILCs, but they still utilize the traditional unsupervised clustering methods to realize the potential molecular variation between the two pathological types. Manual analysis of tumors based on expression array analyses can identify a gene set that distinguishes these two subtypes of breast tumors. Microarray data can reveal information pertaining to not only gene

expression but also regarding genetic networks of a particular biological process.

Networks are pervasive in biology [10]. These network data extend and compliment a great deal of other information available in the biomedical sciences. Although various datasets can appear quite different in quality and quantity, they all are reflections of the same underlying biological system and its responses [11]. Thus a network elucidating the molecular basis and interaction between the components can yield a possible insight for differentiating pathological types.

The main goal of the study is to construct the genetic regulatory network from the microarray data using biological knowledge and constraint-based inferences, in order to explore the potential significant gene regulatory networks that can differentiate IDC and ILC and thereby understand the complex interactions that are influenced by the genetic networks.

Methodology:

Dataset Collection

A comprehensive search of all eligible studies on differential gene expression of IDC and ILC (as on April 2010) was made by searching the electronic literature (PubMed database) for relevant published reports and by manual searching of reference lists of articles on this topic. Only human studies in the English language were included in the analysis. Based on the literature survey GEO record GDS2635 of the Platform GPL570. 5 diseased and 10 control samples for each of IDC & ILC were taken for analysis.

Gene expression analysis

A meta analysis of the chosen datasets was performed using the Gene Spring Gx 7.3 software. GeneSpring GX provides powerful, accessible statistical tools for fast visualization and analysis of expression and genomic structural variation data. Designed specifically for the needs of biologists, GeneSpring GX offers an interactive desktop computing environment that promotes investigation and enables understanding of microarray data within a biological context [12].

The datasets were normalized to standardize microarray data to enable differentiation between real (biological) variations in gene expression levels and variations due to the measurement process. The genes were filtered based on fold changes. The fold changes in gene expression levels between the disease samples control samples to check for the differential expression. Genes which were differentially up regulated by 4 fold were filtered out and their gene ontology was identified.

Gene Network Construction

Two gene networks to study the molecular differences between IDC and ILC were constructed using BisoGenet plugin [13] for Cytoscape was used for generation of biological networks. The Networks were generated taking as input an initial list of identifiers of genes filtered out on basis of fold change.

Network Analysis

The network obtained from the BisoGenet Server is analyzed using the plugin Network Analyzer. For every node in a network, NetworkAnalyzer computes its degree, its clustering coefficient, the number of self-loops, and a variety of other

parameters [14]. The Overrepresentation or underrepresentation of GO categories was assessed using The Biological Networks Gene Ontology tool (BiNGO)

Discussion:

Preliminary Analysis

The gene expression profiling of Invasive Ductal carcinoma and Invasive Lobular Carcinoma dataset containing 5 samples each was used to find the up regulated and down-regulated genes. First the up regulated and down-regulated genes were obtained by comparing Invasive Ductal carcinoma samples to control samples at a fold change of 4.

The gene expression data pertaining to both invasive ductal carcinoma and invasive lobular carcinoma was subjected to meta analysis. The genes which were upregulated by 4 fold were chosen. Out of the 54676 genes present on the gpl570 platform- 29 genes exhibited 4 fold up regulation in case of IDC and 22 in the case of ILC. **Table 1 & 2 (see supplementary material)** illustrate the list of the filtered genes in each of the two pathological states. These upregulated genes were analyzed systematically based on their ontology.

Accurate and precise diagnosis and subsequent treatment of IDC and ILC remain elusive and difficult owing to high degree of similarity in the prognosis of the two pathological types [9]. Gene expression profiling technique is widely in the measurement of the expression of thousands of genes at once and thereby creates a global picture of a biological phenomenon. The technology is employed in order to obtain a clear insight on how the expression of every individual gene is altered in a particular physiological state. Ideology was to look for subtle differences in gene expression may be responsible for the phenotypic differences between them.

The Gene expression profiling datasets available from previously concluded studies were used to find the list of gene which is significantly differentially expressed genes between the Invasive Ductal Carcinoma and Invasive Lobular Carcinoma. Detailed analysis of tumors was used to identify a gene set that distinguishes these two subtypes of breast tumors.

As indicated by the highly similar prognosis and physiological manifestation of the two pathological subtypes - It is observed that both the types of tumor exhibit a very similar expression level for numerous genes.

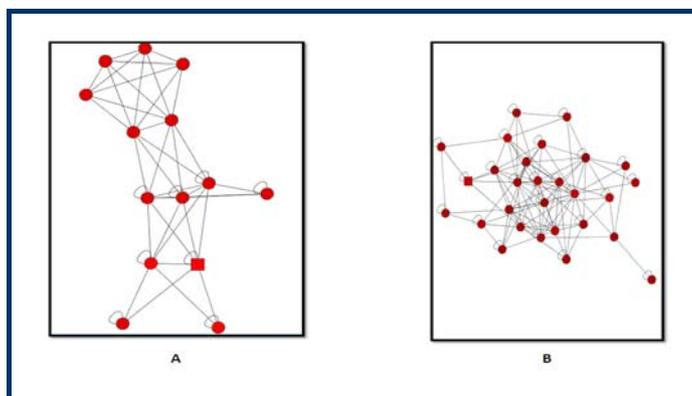


Figure 1: The top ranked gene dense cliques for A. IDC & B. ILC

Construction of Gene Regulatory Networks

Gene regulatory networks of the gene set compiled on the basis of expression level alterations can yield a heuristic insight into the molecular basis of the prognosis and thereby can be utilized to discover specific diagnostic and tailored disease targeting. The two complex gene networks that emerged as a resultant of the interaction between upregulated genes was broken down into smaller sub networks using mcode module of cytoscape. The top ranked gene dense cliques were found and subjected to further analysis by mapping each node to be corresponding ontology. The top ranked gene dense cliques are depicted in **(Figure 1)**. The **Tables 3 & 4 (see supplementary material)** show the list of genes which constitutes the top ranked dense clique along with their respective ontology.

Analyzing Gene Clusters

The gene clusters obtained contain only 3 genes in common - CDK1, HDAC and ESR. Cyclin dependent kinase 1 is a key player in cell cycle regulation. Cdk1 substrates frequently contain multiple phosphorylation sites that are clustered in regions of intrinsic disorder, and their exact position in the protein is often poorly conserved in evolution, indicating that precise positioning of phosphorylation is not required for regulation of the substrate [15]. Cdk1 interacts with nine different cyclins throughout the cell cycle. The interaction with cyclins is important for activation of its kinase activity [16]. CDK1 and ESR are key components of estrogen responsive protein efp controls cell cycle and breast tumors growth. Histone deacetylases (HDAC) are involved in removing the acetyl groups ($O=C-CH_3$) from an ϵ -N-acetyl lysine amino acid on a histone. The fact that acetylation is a key component in the regulation of gene expression has stimulated the study of HDACs in relation to the aberrant gene expression often observed in cancer. Although no direct alteration in the expression of HDACs has yet been demonstrated in human oncogenesis, it is now known that HDACs associate with a number of well characterized cellular oncogenes and tumour-suppressor genes [e.g. Mad and retinoblastoma protein (Rb)], leading to an aberrant recruitment of HDAC activity, which in turn results in changes in gene expression [17, 18].

A notable observation was that ductal and lobular tumors displayed a striking difference in the expression of genes associated with cell adhesion (PCDH9, IBSP, COL11A1, and CTNNA1), motility (S100P), apoptosis (HDAC1, IKBKB and BUB1), protein folding (CPB1, RAB26), extracellular matrix (COL11A1, PRMT1), and protein phosphorylation and invasion (ESRRA, SMAD3, HDAC2, STRN). This suggests that they may achieve invasive growth through separate mechanisms and evolve via distinct genetic pathways.

The ILC genes are differentially expressed compared to IDC genes and are involved in other biological processes such as nucleosome disassembly- HIST1H2AD, SMARCE1, intracellular signaling pathway - GSK3B, GRB2, CALM3, DNA repair and negative regulation of DNA binding - SUMO1, TOP2A, UBC, BRCA1 .

HIST1H2AD gene codes for Histone H2A type 1-D which is a core component of nucleosome involved in compacting DNA into chromatin and limiting DNA accessibility to the cellular machineries which require DNA as a template [19, 20].

SMARCE1 is involved in alteration of DNA-nucleosome topology, Involved in transcriptional activation and repression of select genes by chromatin remodeling [21]. The gene regulatory network constructed can be used as a hypothetical molecular framework to develop a diagnostic marker specific to each of the two pathological states of interest. The validity of the hypothesized biomarkers can be validated by employing available in vitro techniques.

Conclusion:

To sum up, we have utilized the Construction of separate gene regulation networks for IDC and ILC on the basis of gene expression alteration reveals clear distinction in the mechanism that underlies the pathological differences between the two. The molecular level understanding of the pathological manifestations can be exploited in future to find unique bio markers for diagnosis and to identify ideal therapeutic drug targets.

References:

- [1] Sariego J, *Am Surg.* 2010 **76**: 1397 [PMID: 21265355]
- [2] <http://www.cancer.org/Research/CancerFactsFigures/BreastCancerFactsFigures/breast-cancer-facts--figures-2009-2010>
- [3] James E *et al.* *Cancer Res.* 2003 **63**: 7167 [PMID: 14612510]
- [4] Berx G *et al.* *Embo J.* 1995 **14**: 6107 [PMID: 8557030]
- [5] Sastre-Garau X *et al.* *Cancer.* 1996 **77**: 113 [PMID: 8630916]
- [6] Katz A *et al.* *Lancet Oncol.* 2007 **8**: 55 [PMID: 17196511]
- [7] Lamovec J *et al.* *J Surg Oncol.* 1991 **48**: 28 [PMID: 1653879]
- [8] Lee AH *et al.* *J Pathol.* 1998 **185**: 394 [PMID: 9828838]
- [9] Toikkanen S *et al.* *Br J Cancer.* 1997 **76**: 1234 [PMID: 9365176]
- [10] Fowler JH *et al.* *Proc Natl Acad Sci.* 2009 **106**: 1720 [PMID: 19171900]
- [11] Smoot Me *et al.* *Bioinformatics.* 2011 **27**: 431 [PMID: 21149340]
- [12] Barr TL *et al.* *Neurology.* 2010 **75**: 1009 [PMID: 20837969]
- [13] Martin A *et al.* *BMC Bioinformatics.* 2010 **10**: 91 [PMID: 20163717]
- [14] Assenov Y *et al.* *Bioinformatics.* 2008 **24**: 282 [PMID: 18006545]
- [15] Moses AM *et al.* *Genome Biol.* 2007 **8**: R23 [PMID: 17316440]
- [16] Archambault V *et al.* *Cell Cycle.* 2005 **4**: 125 [PMID: 15611618]
- [17] Cress WD *et al.* *J Cell Physiol.* 2000 **184**: 1 [PMID: 10825229]
- [18] Timmermann S *et al.* *Cell Mol Life Sci.* 2001 **58**: 728 [PMID: 11437234]
- [19] Albig W *et al.* *Hum Genet.* 1998 **101**: 284 [PMID: 9439656]
- [20] El Kharroubi A *et al.* *Mol Cell Biol.* 1998 **18**: 2535 [PMID: 9566873]
- [21] Wang W *et al.* *Proc Natl Acad Sci U S A.* 1998 **95**: 492. [PMID: 9435219]

Edited by P Kanguane

Citation: Ragnath *et al.* *Bioinformation* 8(8): 359-364 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Gene list: upregulated genes in Invasive Ductal Carcinoma at Fold change 4

S. No.	Gene Symbol	Gene Ontology
1.	SPATA2	Cell differentiation
2.	S100P	Endothelial cell migration
3.	LOC145837	Hypothetical
4.	FAM40B	DNA binding
5.	ZNF670	Regulation of transcription, DNA-dependent
6.	MRPL35	Translation
7.	DTNA	Signal transduction and synaptic transmission
8.	C16orf54	Transmembrane Protein
9.	TNNT1	Muscle filament sliding and skeletal muscle contraction
10.	POLQc	DNA repair and DNA replication
11.	CPB1	Proteolysis
12.	C20orf114	Innate Immune Response
13.	GBP4	Guanylate Binding
14.	PCDH9	Cell adhesion
15.	INTS10	snRNA processing
16.	FBN1	Heart development and kidney development
17.	KIF20A	Cytokinesis
18.	TRIM59	DNA Binding
19.	HIST1H2AD	Nucleosome assembly
20.	BUB1	Apoptosis and cell division
21.	IBSP	Cell adhesion
22.	HMMR	cell motility
23.	SMEK1	May regulate the activity of PPP4C at centrosomal microtubule organizing centers.
24.	ASPM	Cell cycle and cell division
25.	COL11A1	Cell adhesion
26.	TOP2A	Chromosome segregation

Table 2: Gene list: upregulated genes in Invasive Lobular Carcinoma at Fold change 4

S. No.	Gene Symbol	Gene Ontology
1.	PRKRA	Immune response and induction of apoptosis
2.	SGCG	Cytoskeleton organization and muscle organ development
3.	CRTC3	Regulation of transcription
4.	NCOA3	Positive regulation of transcription, DNA-dependent
5.	SIX1	Cell division
6.	HSPA12A	Protein Folding
7.	NEK2	Cell division
8.	SIX1	Epithelial cell differentiation
9.	RAB26	Protein transport
10.	HOXB7	Multicellular organismal development
11.	COL10A1	Skeletal system development
12.	SPTBN1	Actin filament capping
13.	COCH	Sensory perception of sound
14.	NOX4	Cell aging, cell morphogenesis, negative regulation of cell proliferation
15.	TMEM55A	Degradation Of Phosphatidylinositol 4,5-Bisphosphate
16.	TOP2A	Chromosome segregation
17.	FBN1	Heart development and kidney development
18.	LRRC15	Unknown Function
19.	COL11A1	Extracellular matrix organization

Table 3: Gene list: Dense clique ranked 1 from Cytoscape plugin MCODE for Invasive Ductal Carcinoma

S. No.	Gene Symbol	Gene Ontology
1	FAM40A	DNA Binding
2	HDAC2	Negative regulation of cell cycle and positive regulation of cell proliferation
3	CTTNBP2NL	-
4	FAM40B	DNA binding
5	ESR2	Signal transduction, cell-cell signaling and gene expression
6	CDK1	Cell division, positive regulation of gene expression
7	MOBKL3	Transport
8	CALM2	Intracellular signaling pathway
9	CALM3	Intracellular signaling pathway
10	CSNK2A1P	Wnt receptor signaling pathway
11	STRN	Negative regulation of cell proliferation
12	CALM1	G-protein coupled receptor protein signaling pathway
13	CSNK2A1	Signal transduction and Wnt receptor signaling pathway
14	STRN3	Negative regulation of transcription

Table 4: Gene list: Dense clique ranked 1 from Cytoscape plugin MCODE for Invasive Lobular Carcinoma

S. No.	Gene Symbol	Gene Ontology
1	IKBKG	Induction of apoptosis
2	DHX9	Gene expression
3	NR1H2	Regulation of transcription from RNA polymerase II promoter and gene expression
4	MAPK1	Signal transduction and cell cycle
5	GRB2	Cell-cell signalling and cell differentiation
6	CDK1	Cell aging and cell division
7	IKBKB	Anti-apoptosis
8	ESRRA	Gene expression and regulation cell proliferation
9	SUMO1	DNA repair and negative regulation of DNA binding
10	PPARG	Negative regulation of transcription from RNA polymerase II promoter and signal transduction
11	GSK3B	Intracellular signaling pathway
12	TOP2A	DNA replication
13	EIF2AK2	Negative regulation of cell proliferation
14	CCND1	Regulation of cell cycle
15	SMARCE1	Nucleosome disassembly
16	SMAD3	Negative regulation of cell proliferation
17	PRMT1	Cell surface receptor linked signaling pathway
18	NR0B2	Regulation of transcription, DNA-dependent
19	STAT1	Signal transduction
20	EP300	Cell cycle
21	CREBBP	Signal transduction and cell proliferation
22	UBC	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest
23	JUN	Regulation of cell cycle and aging
24	BRCA1	DNA repair and cell cycle
25	ESR1	Signal transduction
26	NCOA1	-
27	HDAC1	Anti-apoptosis
28	CTNNB1	Negative regulation of cell proliferation, apoptosis and cell adhesion