Interdisciplinary Sciences: Computational Life Sciences
https://doi.org/10.1007/s12539-020-00360-0

ORIGINAL RESEARCH ARTICLE

An Integrated Systems Biology and Network-Based Approaches to Identify Novel Biomarkers in Breast Cancer Cell Lines Using Gene Expression Data

Abbas Khan[1,2,3] · Zainab Rehman[4] · Huma Farooque Hashmi[5] · Abdul Aziz Khan[4] · Muhammad Junaid[1,2,3] · Abrar Muhammad Sayaf[6] · Syed Shujait Ali[6] · Fakhr Ul Hassan[7] · Wang Heng[1,2,3] · Dong-Qing Wei[1,2,3]

Received: 17 August 2019 / Revised: 31 December 2019 / Accepted: 18 January 2020

Abstract

Breast cancer is the most common cause of death in women worldwide. Approximately 5%–10% of instances are attributed to mutations acquired from the parents. Therefore, it is highly recommended to design more potential drugs and drug targets to eradicate such complex diseases. Network-based gene expression profiling is a suggested tool for discovering drug targets by incorporating various factors such as disease states, intensities based on gene expression as well as protein–protein interactions. To find prospective biomarkers in breast cancer, we first identified differentially expressed genes (DEGs) statistical methods *p*-value and false discovery rate were initially used. Of the total 82 DEGs, 67 were upregulated while the remaining 17 were downregulated. Sub-modules and hub genes include VEGFA with the highest degree, followed by 15 CCND1 and CXCL8 with 12-degree score was found. The survival analysis revealed that all the hub genes have important role in the development and progression of breast cancer. Enrichment analysis revealed that most of these genes are involved in signaling pathways and in the extracellular spaces. We also identified transcription factors and kinases, which regulate proteins in the DEGs PPI. Finally, drugs for each hub genes were identified. These results further expanded the knowledge regarding important biomarkers in breast cancer.

Keywords Breast cancer · Hub genes · PPI · Cytohubba · Differentially expressed genes (DEGs)

1 Introduction

Breast cancer (BC) is the most common cause of death in women worldwide. The symptoms of this heterogeneous disease ranges from a bump in the breast, a shift in the form of the breast, skin dimpling, liquid from the nipple, a newly inverted nipple to a yellow or scaly hair patch. The outcomes of breast cancer greatly depend on the form of cancer, the magnitude of the disorder and the age of the person [1]. The survival rates against breast cancer are higher in the developed

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12539-020-00360-0) contains supplementary material, which is available to authorized users.

✉ Wang Heng
Wangheng0802@sjtu.edu.cn

✉ Dong-Qing Wei
dqwei@sjtu.edu.cn

[1] Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

[2] Peng Cheng Laboratory, Vanke Cloud City Phase I Building 8, Xili Street, Nashan District, Shenzhen 518055, Guangdong, People's Republic of China

[3] Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education, Beijing, People's Republic of China

[4] Laboratory of Animal and Human Physiology, Department of Animal Sciences, Quaid-I-Azam University, Islamabad 45320, Pakistan

[5] Department of Biochemistry, Faculty of Biological Sciences, Quaid-I-Azam University, Islamabad 45320, Pakistan

[6] Centre for Biotechnology and Microbiology, University of Swat, Swat, Khyber Pakhtunkhwa, Pakistan

[7] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

nation (80–90%) whereas in developing the survival chances are lower. Risk factors for breast disease are depression, absence of physical activity, smoking, liquor, hormone substitution treatment during menopause, ionizing radiation, and premature death at first menstruation, having kids late or not at all, elderly era, and family history. Approximately 5%–10% of instances are attributed to mutations acquired from the parents. Therefore, it is highly recommended to design more potential drugs and drug targets to eradicate such complex diseases [2]. In this regard, in silico methods to predict structural implications of mutations will be extremely useful in understanding mechanisms of drug resistance for quantitative estimation of the phenotypic resistance outcomes.

On the other hand, network-based gene expression profiling is a suggested tool for discovering drug targets by incorporating various factors such as disease states, intensities based on gene expression as well as protein–protein interactions [3]. Systems biology is one of those methods that depend on a worldwide strategy by evaluating the entire network that interacts rather than studying a particular protein, gene or enzyme. Systems biology has revealed that cellular protein does not work alone, but these genes/proteins are linked together to build an interconnected molecular network to execute a particular role.

To find prospective biomarkers in breast cancer, proteomics and transcriptomic modelling of molecular networks from microarray information have not yet been established for some datasets. Therefore, we used gene expression information to define potential therapeutic gene/protein biomarkers for the treatment of breast cancer using systematic system biology strategy centered on microarray study. To identify differentially expressed genes (DEGs) statistical methods *p*-value and false discovery rate (FDR) were initially used. Also, the subnetwork modules were built, and the DEGs acquired were evaluated for biological processes, molecular components, KEGG pathways and interpretation of cellular components. Addition of hub genes identification and their drugs interactions predicted enormous information in the understanding of disease mechanism and potential therapeutic targets.

2 Materials and Methods

2.1 Data Retrieval for Meta-Analysis

Breast cancer dataset search was performed in the GEO database of NCBI [4]. Breast cancer dataset with accession number GSE14335 [5] was downloaded, which has a total of 24 samples, including wild and mutants. This data is based on GPL96 Affymetrix Human Genome U133A Array [HG-U133A] and is based on mutations in the ER-alpha gene, which is the most important parameter in breast cancer measurement. Differential expression analysis, principal component analysis (PCA) for clustering of the samples, heat map analysis of the DEGs, construction of the DEGsPPI, sub-networks identification, hub genes identification, protein-drug interaction network and finally molecular docking was performed in this study.

2.2 Preprocessing and Differential Expression Analysis of the Dataset

An online webserver NetworkAnalyst tool (https://www.networkanalyst.ca/NetworkAnalyst/faces/home.xhtml) has been used to screen the dataset for differential expression genes based on the statistically significant parameters [6]. The dataset was defined as rows and columns where the rows represent each gene entry, and the columns represent the samples. An even distribution of the samples, 12 wild type and 12 mutants, based on experimental data was carried out, and all gene probe IDs were transformed to Entrez IDs [7]. The meta-analysis of the microarray was carried out using Network Analyst, an integrative meta-analysis web tool [6]. Each dataset was normalized with log2, VSN and quantile normalization methods and normality were further assured through box plots and PCA-plots inspection. The differential expression testing for all individual datasets was performed using $p < 0.05$, FDR (false discovery rate) ≥ 2(by Benjamin–Hochberg method) and a Limma algorithm-based *t*-test (LAT) [8].

2.3 Enrichment Analysis of Gene Set

Both DEGs and modules function and pathway enrichment analysis was done with the DAVID server (https://david.ncifcrf.gov/). The modules DEGs and hub genes were uploaded, and various parameters; mode-function, Entrez gene ID, species-*Homo sapiens*, molecular function, cellular component, ontology/pathways-biological process, and evidence from the respective gene ontology and KEGG database were set to analyze the function and pathway enrichment. The two-sided hypergeometric test (Benjamin-Hochberg method) with kappa score 0.96 and cutoff -value > 0.005 was performed for enrichment calculation.

2.4 Protein–Protein Interaction (PPI) Network Generation

The STRING database was used for PPI networking of all DEGs [9, 10]. The database currently includes total 18,838 human proteins, with 25,914,693 interactions. The interactions in the Cytoscape v3. 4 have been filled and analyzed using various integral attributes. Protein interactions were initially uploaded into the Cytoscape v 3.4 and were assessed using integrated functions [11]. The highest score

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

of confidence interaction was 0.9 in this analysis [12]. A search was made to recognize the direct involvement of each DEG with the first-order interactors. The search resulted in a large subnetwork (continent) and several small (islands). In subnetwork, a minimum of three nodes, edges, and seed genes were found. A complete information list of both DEGs was uploaded, and the network design was limited to original DEGs, i.e. zero-order interactions selection was made to allow precise visualization of PPI and to exclude hairball effect. Different topological variables are present too for analysis and comparing the network. Although Cytoscape is freely accessible, that brings an integrated "Network Analyzer" function for analyzing the network of genes/proteins. In this stud, we used Network Analyzer, and main variables that are analyzed involves clustering coefficient, node distribution power law, node degree distribution, network centralization, and density to distinguish the three developed networks [13].

2.5 Molecular Complex Detection Analysis (MCODE)

MCODE an automated algorithm can be utilized in a Cytoscape-integrated plug-in to identify strongly linked complex subnetwork in a PPI/gene subnetworks [14] and was used to cluster the total DEGs subnetworks. For further evaluation, the interconnected nodes in the subgraphs were chosen based on the node number. We also used more than ten parameters for strongly interconnected sub-networks.

2.6 Identification of Hub Genes

Cytohubba is a prominent integrated Cytoscape that basically analyze the features and ranks the nodes accordingly [15]. It uses 11 different approaches for analyzing the network functions including along with the network hub genes/nodes identification. We, therefore, used Cytohubba to figure out the hub genes that might be the possible new drug targets for breast cancer treatment.

2.7 Transcription Factor and Regulatory Network Association

X2K (Expression2Kinases) web tool (https://amp.pharm.mssm.edu/X2K/) was utilized to find the association between the transcription factor(s) and linked target genes. The DEGs full list with specified gene symbols was uploaded into X2K web server [16]. Based on Fischer test *p*-value the ten most significant TFs and kinases enrichment scores have been determined using TF and kinase module, which build the exploited chip-x from the ChEA69 database. A regulatory network was established, and the "graphml" file visualized in the Cytoscape [17]. The regulatory network ensures that the acquired network must have sufficiently associated edge nodes during the development of the network. If TFs and kinases are not linked, it automatically increases the length of the path to allow the TFs to be linked with enough transitional proteins.

2.8 Survival Analysis

Median quartiles were calculated in Kaplan–Meier web plotter tool, with 54,675 survival nodes and 10,188 tumor samples to determine the importance of hub nodes [18]. The patients were separated into two groups according to their gene expression level. The patient sex, grade, histology, stage and smoking status were used in multivariate Cox analysis. The total cases of lung cancer were 2437, the data on gene expression and survival of these samples were collected from GEO, Cancer Biomedical Informatics Grid (caBIG), and The Cancer Genome Atlas (TCGA). The hazard ratios (HRs) (with confidence intervals of 95% and log-rank *p*-value) were quantified in *R* using the Bioconductor Survival Analysis Package [19]. We analyzed the breast cancer survival outcomes in diverse expression levels of hub genes and downloaded survival curves from the website.

2.9 Construction of PDI Network

To evaluate gene interactions with drugs, the ten best hub genes were used. The information on drug and their targets was extracted from the Drug Bank database (Version 5.0) that was embedded with Network Analyzer. The drug database has 11,091 drugs; approved small molecule drugs (2550), approved protein/peptide-based drugs (960), and approved nutraceuticals (112), and more than 5117 medicines currently pass through experimental stages. The data entry of each drug contains sufficient information (about 200 in each case) on drug chemical nature and their target [20].

3 Results

3.1 Quality Control and Principle Component Analysis

Genes and their products provide bases for a biological system, in which they interact randomly to develop a complicated network. Genes, as well as protein expression, is very informative to understand the mechanism of immunity, defense, signaling, transport and disease. Normalization of variation prone Microarray expression analysis was performed on quantile normalization option. Figure 1 shows the Boxplot, density plot and means of the data before and after normalization. It can be seen from the figure that the means of each sample is uniform and any noise in the data is removed during the process of normalization. PCA

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

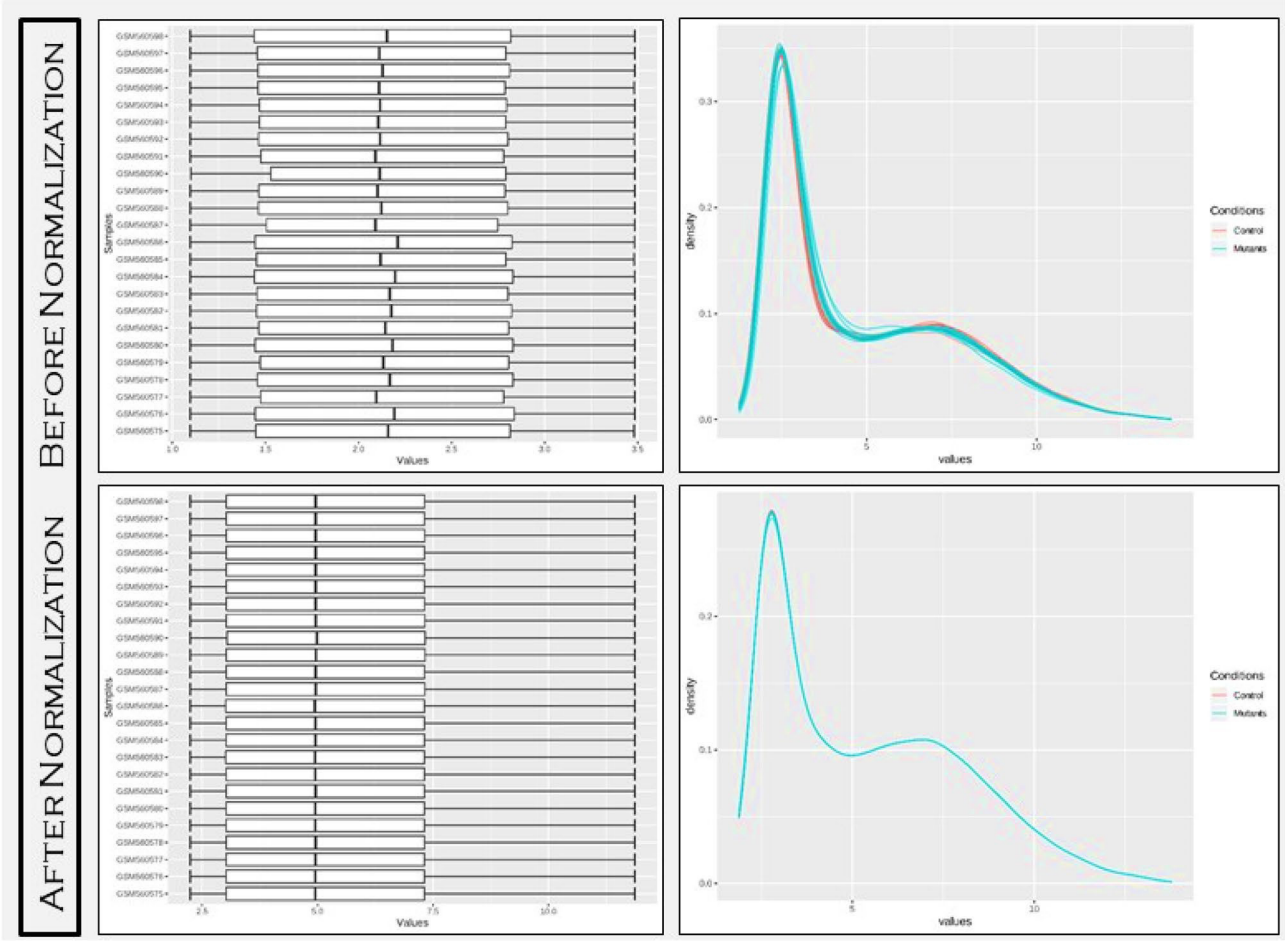


Fig. 1 Presentation of the samples using a box plot and a density plot. The figure is showing the boxplot and density plots before and after normalization. Quantile normalization was used to correct the means and remove the noise from the data

revealed the distinct clusters in the data. PCA clustered the controlled and mutant samples in the dataset according to the gene expression intensities. The PCA plot and mean expression plot is given in Fig. S1 in the supplementary materials.

3.2 Differential Expression Analysis and PPI Network Analysis

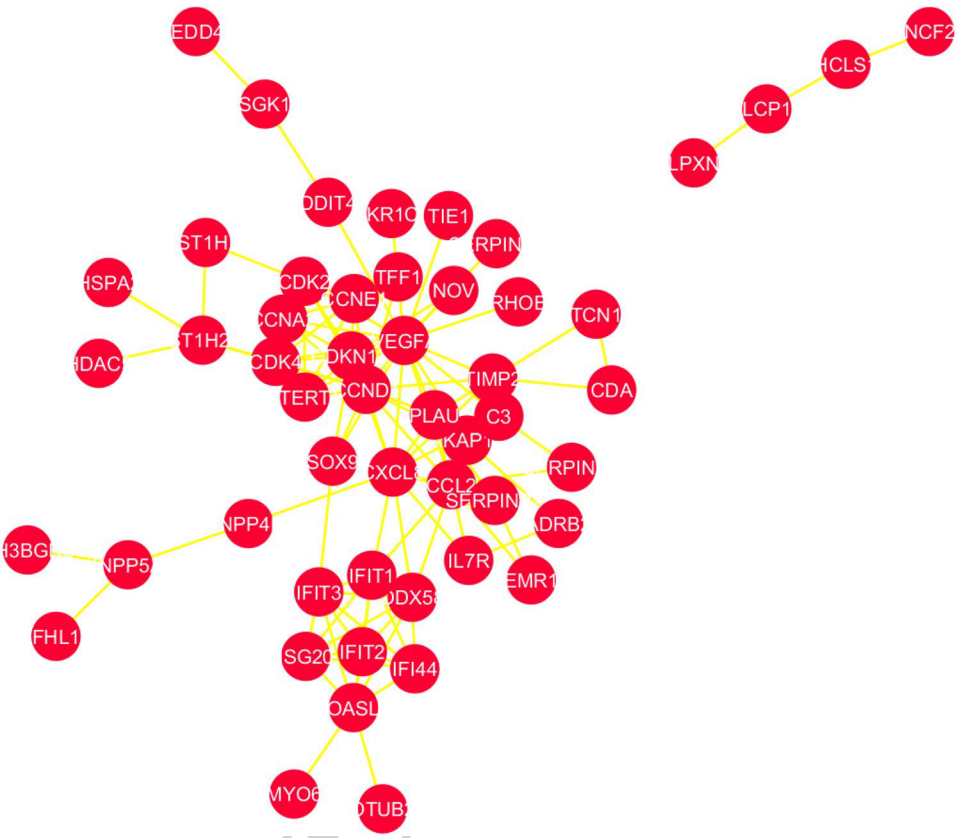
The control and diseased datasets were analyzed with the help of various statistical tests such as the Pearson correlation test, Benjamin-Hochberg method, and student's t-test resulted in the identification of 82 DEGs. In these DEGs, 65 genes were upregulated and 17 genes were downregulated, respectively. The final DEGs were ranked on the adjusted *p*-value. Protein–protein interaction network is a very important approach to highlight the cellular networking mechanism. The changes in protein cellular network in healthy and disease condition provide information about the severity of the situation. In this analysis, nodes represent protein and genes, while edges indicate interactions of these proteins and genes. The interaction network for the identified DEGs was constructed by mapping of these DEGs. Cytoscape was used to visualize the mapped DEGs. String database was accessed with medium confidence of 0.40 to download all the interaction of the DEGs. The network constructed of all the DEGs is given in Fig. 2.

A total of 81 nodes with 84 edges were constructed. The topological network parameters revealed that that average node degree was 2.07 while the average local clustering coefficient was found to be 0.425, which shows that our network has significant interactions. The heat map and volcano plot of these DEGs are given in Fig. 3.

3.3 GO Function and KEGG Pathway Enrichment Analysis

Functional enrichment analysis, including GO cellular components, molecular function, biological processes and

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020



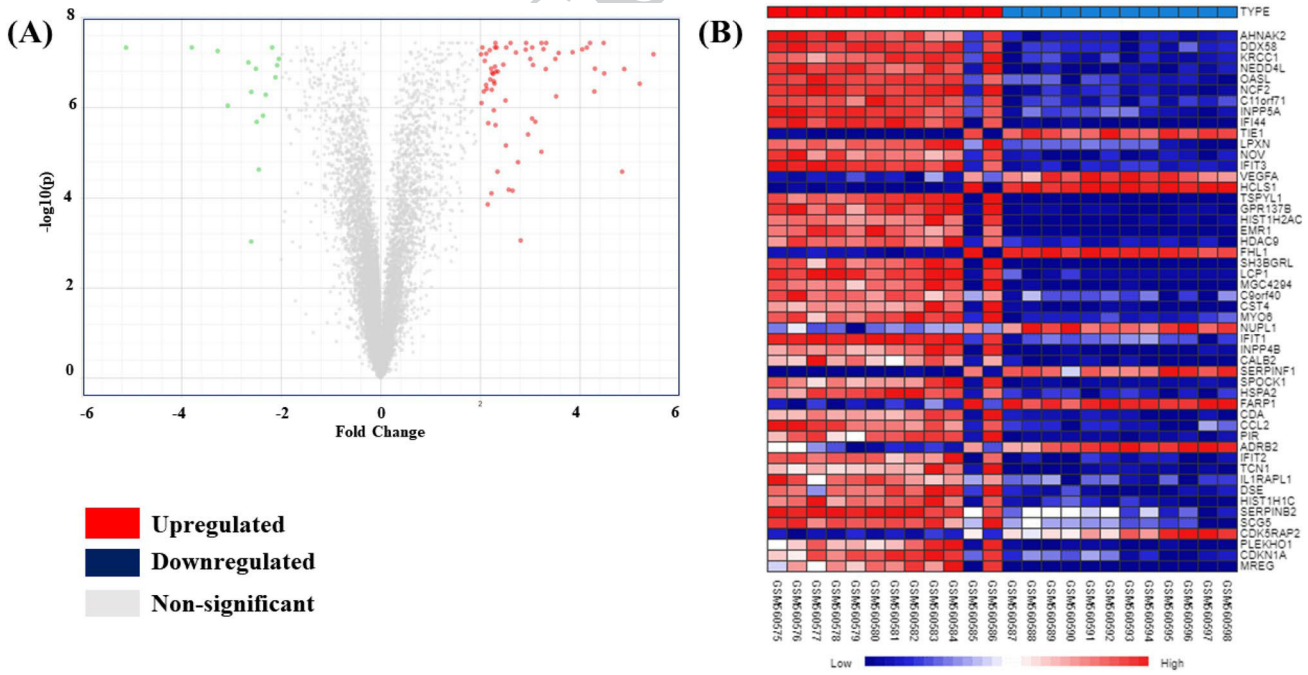


Fig. 3 **a** Showing the distribution of the intensities of all the genes on volcano plot. Based on FDR > 2 upregulated genes are colored as red, downregulated are colored as blue while the non-significant genes in the samples are colored as grey. **b** Panel B plots the genes in against each sample as a heat map based on the expression value. The legend represents the distribution of the expression from low (blue) to high (red)

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

KEGG pathways, was performed using FDR as highly significant criteria for the terms. GO cellular components of the DEGs revealed that our proteins were significantly distributed in extracellular space, extracellular region part, extracellular region and ruffle membrane. In the case of molecular function, 11 significant molecular functions in which our proteins are involved was found. Among the total 11, the top four highly significant molecular functions include endopeptidase inhibitor activity, enzyme inhibitor activity, molecular function regulator and serine-type endopeptidase inhibitor activity was found.

On the other hand, significant KEGG pathways primarily belong to cancer-related processes. Among the total reported pathways complement and coagulation cascades, salivary secretion, bladder cancer, cytokine signaling in the immune system, interferon-alpha/beta signaling, signaling by interleukins and interferon signaling pathways are highly enriched.

The enriched biological processes of DEGs were identified in negative regulation of biological process, response to chemical negative regulation of endopeptidase activity, negative regulation of cellular process, response to organic substance, response to cytokine, negative regulation of hydrolase activity, cellular response to organic substance, cellular response to cytokine stimulus, response to stimulus, cellular response to chemical stimulus, regulation of localization, regulation of cellular process and negative regulation of molecular function.

3.4 Sub-Network Analysis

The highly dense interconnected module for DEGs was identified using MCODE method. The K-score, node score cutoff and maximum depth for seed node were kept 2, 0.2, and 100, respectively for efficiency. Only one sub-network was identified having node degree value > 10 shown in Fig. 4.

In the subnetwork, 19 proteins from the DEGs were involved. Both upregulated and downregulated genes, which are also important as found in the hub genes, are frequently involved in the subnetwork (Tables 1, 2).

3.5 Hub Genes Identification

The connectivity between hub genes and nodes were determined and ranked. The red nodes are highly connected. Cytohubba calculated the degree of nodes and the nodes with degree value > 10 was considered hub nodes (Table 3). VEGFA with the highest degree, followed by 15 CCND1 and CXCL8 with 12-degree score was found. Among the other CCL2, CDKN1A, CDK4, CDK2, IFIT1, OASL and DDX58were reported with lower degree scores. The shortest path network of these hub genes interactions and color based on degree is given in Fig. S2 in the supplementary materials.

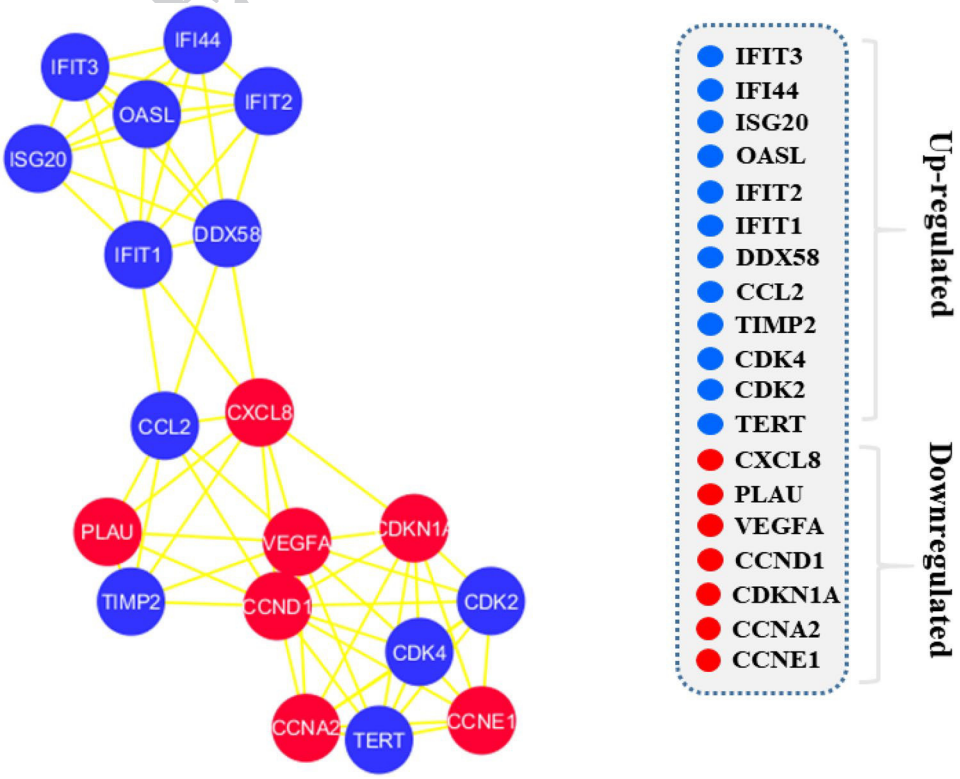


Fig. 4 the identified only sub-network from the total DEGs network using default parameters. The blue colored nodes are upregulated while red are downregulated genes

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

Table 1 Showing the distribution of DEGs in highly significant cellular components, molecular functions and KEGG pathways based on FDR significant value

GO ID	Term description	False discovery rate	Matching proteins in the network (labels)
Cellular components			
GO:0,005,615	Extracellular space	0.0033	C3,CCL2,CST1,CST2,CST4,CXCL8,INHBB,KAL1,KISS1,PLAU,SERPINB2,SERPIND1,SPOCK1,TFF1,TIMP2,VEGFA
GO:0,044,421	Extracellular region part	0.0033	C3,CCL2,CST1,CST2,CST4,CXCL8,INHBB,KAL1,KISS1,NOV,PLAU,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TFF1,TIMP2,VEGFA
GO:0,005,576	Extracellular region	0.0418	C3,CCL2,CDA,CST1,CST2,CST4,CXCL8,IL7R,INHBB,KAL1,KISS1,NOV,PLAU,SCG5,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TCN1,TFF1,TIMP2,VEGFA
GO:0,032,587	Ruffle membrane	0.0418	DDX58,LCP1,MYO6,PLEKHO1
Go molecular function			
GO:0,004,866	Endopeptidase inhibitor activity	1.30E−06	C3,CST1,CST2,CST4,KAL1,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TIMP2
GO:0,004,857	Enzyme inhibitor activity	6.80E−06	C3,CDKN1A,CST1,CST2,CST4,KAL1,SCG5,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TIMP2
GO:0,098,772	Molecular function regulator	4.57E−05	ADRB2,C3,CCL2,CDKN1A,CST1,CST2,CST4,CXCL8,FARP1,INHBB,KAL1,NCF2,NEDD4L,NOV,SCG5,SERPINB2,SERPIND1,SERPINF1,SGK1,SPOCK1,TFF1,TIMP2,VEGFA
GO:0,004,867	Serine-type endopeptidase inhibitor activity	0.0031	KAL1,SERPINB2,SERPIND1,SERPINF1,SPOCK1
GO:0,004,869	Cysteine-type endopeptidase inhibitor activity	0.0058	CST1,CST2,CST4,SPOCK1
GO:0,030,234	Enzyme regulator activity	0.0124	C3,CDKN1A,CST1,CST2,CST4,KAL1,NCF2,SCG5,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TIMP2
GO:0,005,488	Binding	0.0348	ADRB2,AKAP12,ARID5B,C3,CALB2,CCL2,CDA,CDK5RAP2,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,DDX58,EMR1,FARP1,FHL1,HCLS1,HDAC9,HIST1H1C,HIST1H2AC,HSPA2,IFIT1,IFIT3,IL1RAPL1,IL7R,INHBB,INPP5A,ISG20,KAL1,KISS1,KRCC1,LCP1,LPXN,MYO6,NCF2,NEDD4L,NMRK1,NOV,NUPL1,OASL,OTUB2,PIR,RAB20,RHOB,SCG5,SERPIND1,SGK1,SH3BGRL,SOX9,SPOCK1,TCN1,TERT,TFF1,TIE1,TIMP2,TMEM2,TOB1,TSPYL1,VEGFA,ZMIZ1,ZNF266
GO:0,005,515	Protein binding	0.0458	ADRB2,AKAP12,C3,CCL2,CDA,CDK5RAP2,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,DDX58,FARP1,FHL1,HCLS1,HDAC9,HIST1H2AC,HSPA2,IFIT3,IL1RAPL1,INHBB,INPP5A,KISS1,KRCC1,LCP1,MYO6,NCF2,NEDD4L,NOV,OASL,OTUB2,SCG5,SH3BGRL,SOX9,TERT,TFF1,TIMP2,TOB1,TSPYL1,VEGFA
GO:0,015,459	Potassium channel regulator activity	0.0458	ADRB2,NEDD4L,SGK1
GO:0,046,030	Inositol trisphosphate phosphatase activity	0.0458	INPP4B,INPP5A
GO:0,008,179	Adenylate cyclase binding	0.0482	ADRB2,AKAP12
KEGG Pathways			
hsa04610	Complement and coagulation cascades	0.0452	C3,PLAU,SERPINB2,SERPIND1
hsa04970	Salivary secretion	0.0452	ADRB2,CST1,CST2,CST4
hsa05219	Bladder cancer	0.0452	CDKN1A,CXCL8,VEGFA
HSA-1280215	Cytokine Signaling in Immune system	2.53E−05	CCL2,CDKN1A,CXCL8,DDX58,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,ISG20,LCP1,OASL,PTPN7,SERPINB2,VEGFA
HSA-168256	Immune system	0.0013	C3,CCL2,CDA,CDKN1A,CXCL8,DDX58,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,ISG20,LCP1,NCF2,NEDD4L,OASL,PLAU,PTPN7,SERPINB2,TCN1,TIMP2,VEGFA
HSA-909733	Interferon alpha/beta signaling	0.0013	IFIT1,IFIT2,IFIT3,ISG20,OASL

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

Table 1 (continued)

GO ID	Term description	False discovery rate	Matching proteins in the network (labels)
HSA-449147	Signaling by Interleukins	0.0072	CCL2,CDKN1A,CXCL8,IL1RAPL1,IL7R,LCP1,PTPN7, SERPINB2,VEGFA
HSA-913531	Interferon signaling	0.0094	DDX58,IFIT1,IFIT2,IFIT3,ISG20,OASL
HSA-8983711	OAS antiviral response	0.0458	DDX58,OASL
HSA-2262752	Cellular responses to stress	0.0484	CDKN1A,CXCL8,HIST1H1C,HIST1H2AC,HSPA2,NCF2,VEGFA
HSA-6785807	Interleukin-4 and Interleukin-13 signaling	0.0484	CCL2,CDKN1A,CXCL8,VEGFA

3.6 Transcription Factors Analysis

In this analysis, the most significant TFs and protein kinases associated with DEGs were listed depending on their relation to regulatory network progression. The regulatory network was established between TFs, kinases, and their transient proteins implicated in the development of the regulatory complex. In the first step, integrated target genes for transcription factors as determined by ChIP-seq experiments (ChEA) to predict the top transcription factors (TFs) and then mapped on PPI. The TFs predicted here, and the PPI network is shown in Fig. 5 panels (A) and (B). Here the top transcription factors include PPARG, NFE2L2, CEBPB, CEBPD AR and GATA2 based on the hypergeometric *p*-value. Kinases that are likely the regulators of the expanded protein–protein interaction network were also identified and mapped on the PPI network. Figure 5c, d shows the tope predicted kinases and their PPI network. Based on the hypergeometric *p*-value CSNK2A1, MAPK14, MAPK1, CDK1, MAPK3, GSK3B, ERK2, ERK1, HIPK2 and CDK2were found to be the topmost Kinases in these DEGs. All the TFs and kinases identified with their scores are given in Tables S2 and S3 in the supplementary materials.

3.7 Survival Analysis of the hub genes

Survival analysis of all the hub genes based on *p*-value revealed that all the hub genes predicted has important role in the disease prediction and diagnosis. These could be experimentally prioritized genes when accessing the disease prediction and treatment process. Figure 6 is showing the survival profile of each hub gene predicted by the KMPlot server.

Furthermore, we also checked the survival status of each of this hub gene in Luminal A, Luminal B, HEr2, and basal group breast cancer. The following table (Table 4) is showing the *p*-values of each of the hub gene obtained.

3.8 Protein-Drugs Interactions

The identified top Hub genes were subjected to drugs identification from drug bank database. In this study, we manually searched each target and identified their respective drugs. Drug from different categories such as FDA approved, investigational, and nutraceuticals were added to the search panel. We searched drugs for each protein identified as hub genes manually to find the interacting drugs. For all these 157 drugs were identified as a sum. For VEGFA 17 drugs, for CCND1 and IFIT1 4 drugs each, three drugs for CXCL8 and CCL2 each, six drugs for CDK4, 117 drugs for CDK2 while one drug for CDKN1A, OASL andDDX58 each was found. The identified drugs for all these targets are given in Tables S4 in the supplementary materials.

4 Discussion

Many computational methods, such as Single Nucleotide Polymorphism (SNPs) assessment, Genome-Wide Association Studies (GWAS), diseasomes and microarray analysis of gene expression, in particular, are accessible to evaluate distinct genomic data and to access vital information about the disease, from diagnosis to therapy [21]. The profiling of normal, tumor cell activity and Illumina systems contribute to the evaluation of mRNA and the profiling of gene expression, offering advantages adapted to any layout of the research [22, 23]. All of these methods can be used to obtain the various information accessible from distinct stages including genomics, proteomics, transcriptomic, metagenomics, epigenetics, and metabolomics to frequently assist both predictive and prognostic biomarkers in their forecast and growth. Analysis of the PPI network has been commonly used to assist the method of explaining the mechanism of various diseases, finding targets for drugs and metabolic processes [24]. In different biological processes, the structural relationships of distinct proteins varying from normal to disease phenotypes play a significant role [25, 26]. Analyzing the data set of microarray gene expression and identifying differentially

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

Table 2 Showing the distribution of DEGs in highly significant biological processes based on FDR significant value

GO ID	Term description	False discovery rate	Matching proteins in the network (labels)
GO:0048519	Negative regulation of biological process	1.27E−06	ADRB2,AKR1C3,ARID5B,C3,CCL2,CDA,CDK5RAP2,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,DDX58,FARP1,FHL1,HCLS1,HDAC9,HIST1H1C,HIST1H2AC,HSPA2,IFIT1,IFIT3,IL1RAPL1,IL7R,INHBB,ISG20,KAL1,KISS1,LPXN,NEDD4L,NOV,OASL,PLAU,PTPRE,RHOB,SERPINB2,SERPIND1,SERPINF1,SOX9,SPOCK1,TERT,TFF1,TIE1,TIMP2,TOB1,VEGFA
GO:0042221	Response to chemical	1.86E−06	ADRB2,AKR1C3,ARID5B,CCL2,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,DDX58,HCLS1,HDAC9,HSPA2,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,ISG20,KAL1,LCP1,MYO6,NEDD4L,NOV,OASL,PLAU,PTPN7,RAB20,RHOB,SERPINB2,SERPIND1,SERPINF1,SOX9,TCN1,TERT,TFF1,TIE1,TIMP2,TOB1,VEGFA
GO:0010951	Negative regulation of endopeptidase activity	6.16E−06	C3,CST1,CST2,CST4,KAL1,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TIMP2,VEGFA
GO:0048523	Negative regulation of cellular process	6.64E−06	ADRB2,AKR1C3,ARID5B,C3,CCL2,CDA,CDK5RAP2,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,FARP1,FHL1,HCLS1,HDAC9,HIST1H1C,HIST1H2AC,HSPA2,IFIT3,IL1RAPL1,IL7R,INHBB,KAL1,KISS1,LPXN,NEDD4L,NOV,PTPRE,RHOB,SERPINB2,SERPIND1,SERPINF1,SOX9,SPOCK1,TERT,TFF1,TIE1,TIMP2,TOB1,VEGFA
GO:0010033	Response to organic substance	1.28E−05	ADRB2,AKR1C3,ARID5B,CCL2,CDKN1A,CXCL8,DDIT4,DDX58,HCLS1,HDAC9,HSPA2,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,ISG20,KAL1,LCP1,OASL,PTPN7,RAB20,SERPINB2,SERPINF1,SOX9,TERT,TFF1,TIE1,TIMP2,TOB1,VEGFA
GO:0034097	Response to cytokine	1.36E−05	ARID5B,CCL2,CDKN1A,CXCL8,HCLS1,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,ISG20,LCP1,OASL,PTPN7,RAB20,SERPINB2,SOX9,TIMP2,VEGFA
GO:0051346	Negative regulation of hydrolase activity	1.36E−05	C3,CST1,CST2,CST4,FARP1,IFIT1,KAL1,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TIMP2,VEGFA
GO:0071310	Cellular response to organic substance	1.36E−05	ADRB2,AKR1C3,ARID5B,CCL2,CDKN1A,CXCL8,DDIT4,DDX58,HCLS1,HDAC9,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,ISG20,KAL1,LCP1,OASL,PTPN7,RAB20,SERPINB2,SERPINF1,SOX9,TERT,TOB1,VEGFA
GO:0071345	Cellular response to cytokine stimulus	1.64E−05	ARID5B,CCL2,CDKN1A,CXCL8,HCLS1,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,ISG20,LCP1,OASL,PTPN7,RAB20,SERPINB2,SOX9,VEGFA
GO:0050896	Response to stimulus	2.70E−05	ADRB2,AKAP12,AKR1C3,ARID5B,C3,CCL2,CDA,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,DDX58,EMR1,HCLS1,HDAC9,HSPA2,IFI44,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,INPP4B,INPP5A,ISG20,KAL1,KISS1,KRCC1,LCP1,LPXN,MYO6,NCF2,NEDD4L,NOV,OASL,PLAU,PTPN7,PTPRE,RAB20,RHOB,SCG5,SERPINB2,SERPIND1,SERPINF1,SGK1,SOX9,TCN1,TERT,TFF1,TIE1,TIMP2,TOB1,VEGFA
GO:0070887	Cellular response to chemical stimulus	2.86E−05	ADRB2,AKR1C3,ARID5B,CCL2,CDKN1A,CXCL8,DDIT4,DDX58,HCLS1,HDAC9,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,ISG20,KAL1,LCP1,NOV,OASL,PTPN7,RAB20,RHOB,SERPINB2,SERPINF1,SOX9,TERT,TOB1,VEGFA
GO:0032879	Regulation of localization	0.0001	ADRB2,C3,CCL2,CDKN1A,CXCL8,DDX58,FHL1,HCLS1,HDAC9,HSPA2,IL1RAPL1,INHBB,KISS1,LCP1,MYO6,NEDD4L,NOV,NUPL1,PLAU,RAB20,RHOB,SCG5,SERPINF1,SGK1,SOX9,TERT,TIE1,VEGFA

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

Table 2 (continued)

GO ID	Term description	False discovery rate	Matching proteins in the network (labels)
GO:0050794	Regulation of cellular process	0.0001	ADRB2,AKAP12,AKR1C3,ARID5B,C3,C9orf40,CCL2,CDA,CDK5RAP2,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,DDX58,EMR1,FARP1,FHL1,HCLS1,HDAC9,HIST1H1C,HIST1H2AC,HSPA2,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,INPP4B,INPP5A,ISG20,KAL1,KISS1,KRCC1,LCP1,LPXN,MYO6,NCF2,NEDD4L,NOV,OASL,PIR,PLAU,PLEKHO1,PTPRE,RAB20,RHOB,SCG5,SERPINB2,SERPIND1,SERPINF1,SGK1,SH3BGRL,SOX9,SPOCK1,TERT,TFF1,TIE1,TIMP2,TOB1,VEGFA,ZMIZ1,ZNF266
GO:0044092	Negative regulation of molecular function	0.00011	ADRB2,C3,CDKN1A,CST1,CST2,CST4,FARP1,IFIT1,IFIT2,KAL1,NEDD4L,SCG5,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TIMP2,VEGFA
GO:0050789	Regulation of biological process	0.00011	ADRB2,AKAP12,AKR1C3,ARID5B,C3,C9orf40,CCL2,CDA,CDK5RAP2,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,DDX58,EMR1,FARP1,FHL1,HCLS1,HDAC9,HIST1H1C,HIST1H2AC,HSPA2,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,INPP4B,INPP5A,ISG20,KAL1,KISS1,KRCC1,LCP1,LPXN,MYO6,NCF2,NEDD4L,NOV,NUPL1,OASL,PIR,PLAU,PLEKHO1,PTPRE,RAB20,RHOB,SCG5,SERPINB2,SERPIND1,SERPINF1,SGK1,SH3BGRL,SOX9,SPOCK1,TERT,TFF1,TIE1,TIMP2,TMEM2,TOB1,VEGFA,ZMIZ1,ZNF266
GO:0007165	Signal transduction	0.00012	ADRB2,AKAP12,AKR1C3,ARID5B,C3,CCL2,CDA,CDKN1A,CXCL8,DDIT4,DDX58,EMR1,HCLS1,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,INPP4B,INPP5A,ISG20,KAL1,KISS1,KRCC1,LCP1,LPXN,MYO6,NCF2,OASL,PLAU,PTPRE,RHOB,SCG5,SERPINB2,SGK1,SOX9,TIE1,TOB1,VEGFA
GO:0065007	Biological regulation	0.00012	ADRB2,AKAP12,AKR1C3,ARID5B,C3,C9orf40,CALB2,CCL2,CDA,CDK5RAP2,CDKN1A,CST1,CST2,CST4,CXCL8,DDIT4,DDX58,EMR1,FARP1,FHL1,HCLS1,HDAC9,HIST1H1C,HIST1H2AC,HSD11B1,HSPA2,IFIT1,IFIT2,IFIT3,IL1RAPL1,IL7R,INHBB,INPP4B,INPP5A,ISG20,KAL1,KISS1,KRCC1,LCP1,LPXN,MYO6,NCF2,NEDD4L,NOV,NUPL1,OASL,PIR,PLAU,PLEKHO1,PTPRE,RAB20,RHOB,SCG5,SERPINB2,SERPIND1,SERPINF1,SGK1,SH3BGRL,SOX9,SPOCK1,TERT,TFF1,TIE1,TIMP2,TMEM2,TOB1,VEGFA,ZMIZ1,ZNF266
GO:0043086	Negative regulation of catalytic activity	0.00015	C3,CDKN1A,CST1,CST2,CST4,FARP1,IFIT1,KAL1,SCG5,SERPINB2,SERPIND1,SERPINF1,SPOCK1,TIMP2,VEGFA
GO:0065009	Regulation of molecular function	0.00016	ADRB2,ARID5B,C3,CCL2,CDKN1A,CST1,CST2,CST4,CXCL8,DDX58,FARP1,FHL1,HCLS1,HSPA2,IFIT1,IFIT2,INHBB,KAL1,NCF2,NEDD4L,NOV,PLAU,SCG5,SERPINB2,SERPIND1,SERPINF1,SGK1,SPOCK1,TERT,TFF1,TIMP2,VEGFA

expressed genes in a diseased situation relative to normal offers a route to target distinct nodes for the development of new drugs. Biomarkers linked to disease are revealed more precise and robust through molecular network relationships [27]. Previously such studies are reported to be useful in forecasting the hub nodes and their important role in different diseases [28–31].

Herein, we also used microarray data from breast cancer and identified differentially expressed genes in it. Following the enrichment analysis, we identified ten hub genes and possible therapeutics targets. Vascular endothelial growth factor (VEGF) was found to be the top with the highest degree. The role of VEGF in breast cancer has been explored by different studies and reported that in a study of 1788 samples from breast cancer, 72.5% of cases were positive for VEGF. It has been studied that VEGF expression is a prognostic but not predictive marker of hormonal response in non-metastatic invasive breast cancer [32].CCND1 has

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

Table 3 Top 10 in the DEGs network from string protein–protein interaction database ranked by Degree method

Rank	Gene name	Score	logFC	Adjusted *p*-value
1	VEGFA	19	−2.177	4.72E−08
2	CCND1	15	−3.143	3.56E−08
3	CXCL8	12	−3.121	2.17E−07
4	CCL2	11	4.301	1.39E−07
5	CDKN1A	10	2.271	2.64E−07
6	CDK4	8	2.2963	1.25E−07
6	CDK2	8	−2.1122	2.17E−07
6	IFIT1	8	3.0049	8.41E−08
6	OASL	8	4.4767	3.75E−08
6	DDX58	8	2.9199	3.76E−08

The fold change and adjusted values of each of these genes are also given in the table

been reported to have important implication in breast cancer progression. Of the total 63%, samples by a study reported having an important role of CCDN1 in breast cancer [33, 34]. CXCL8 has been reported as a therapeutic target as CXCL8 has been reported to play multiple roles in cancer, such as increased proliferation, angiogenesis, invasion, and metastases and specifically as Cancer stem-like cells (CSC) regulator in breast cancer [35]. CCL2 can be profoundly expressed in breast carcinomas in both the tumor and the surrounding stromal cells [36]. It has been reported that CDKN1A/p21 and low TGFBR2 expression was closely correlated with adverse pathological parameters and poor prognosis in breast cancer [37]. The other targets we identified are also of extreme significance in the progression of breast cancer. The enrichment analysis and the construction of subnetworks and the availability of all the hub genes in the subnetworks clarified that important role of these hub genes. KEGG pathway analysis, molecular function, cellular components and biological processes explained the role of these genes in different related pathways. Identification of transcription factors and their role in the expanded network helped in the identification of regulatory function.

Previously it has also been reported that the interferon and cytokines pathways are important in breast cancer. Herein, our KEGG pathways also involve those pathways. Several immune related pathways were up and downregulated in the breast cancer patients. Complement and coagulation cascades was also among the reported pathways. Our results

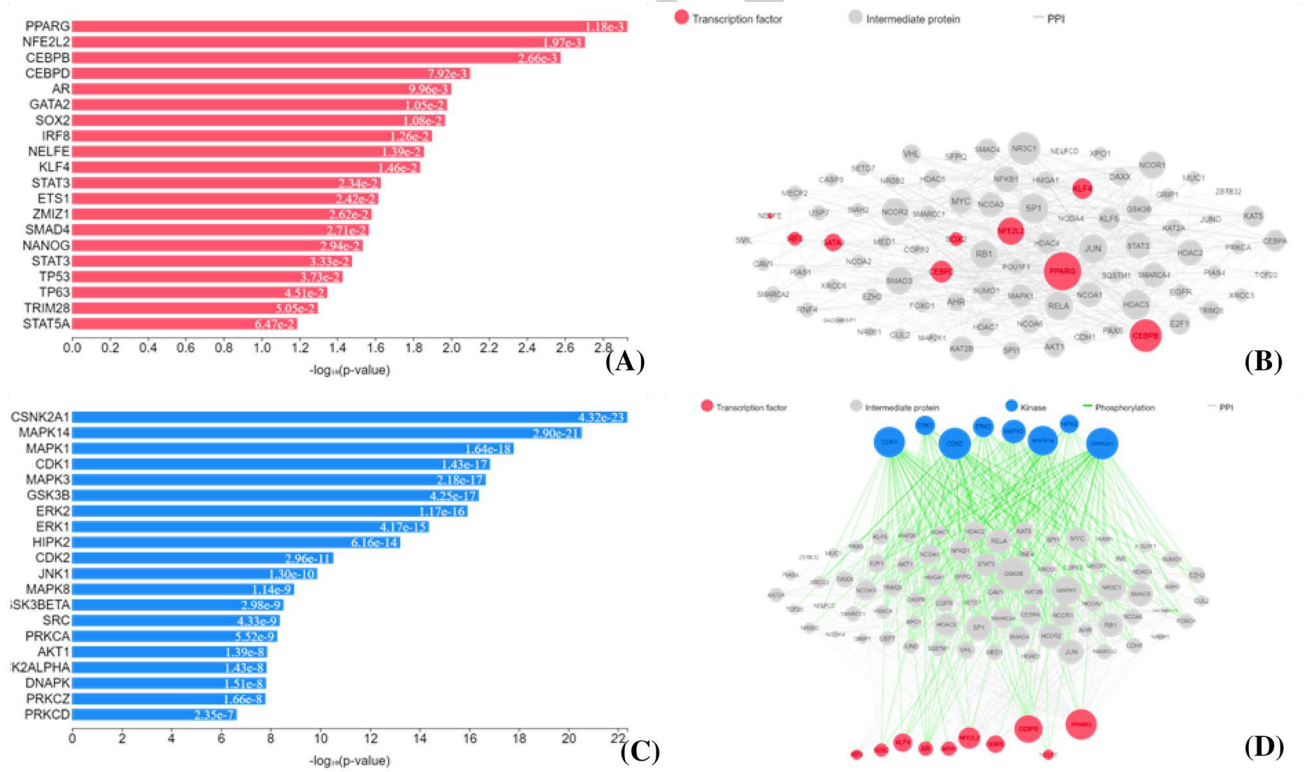


Fig. 5 **a** showing the predicted transcription factors in the DEGs list. The bar is showing the scores, hypergeometric p-value, of each identified transcription factors. **b** A subnetwork of connected transcription factors and their interacting proteins is visualized as a ball-and-stick diagram. Transcription factors are the pink nodes, while the proteins that connect them are in grey. The size of the nodes in the network is proportional to their degree. **c** A ranked list of the top predicted kinases is displayed as a bar graph showing the score (hypergeometric *p*-value)

Journal : Large 12539 | Article No : 360 | Pages : 14 | MS Code : 360 | Dispatch : 1-2-2020

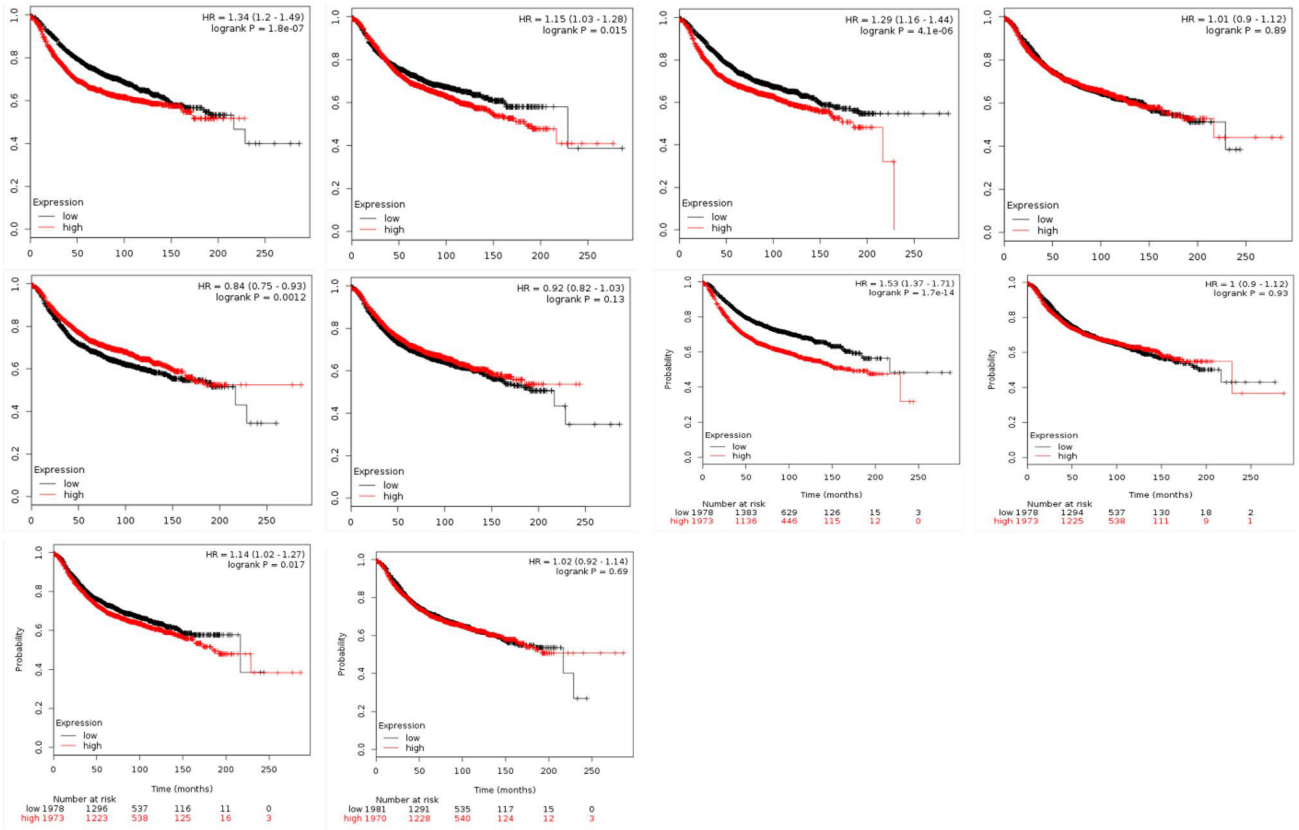


Fig. 6 Survival plots of all the hub genes (VEGFA, CCND1, CXCL8, CCL2, CDKN1A, CDK4, CDK2, IFIT1, OASL and DDX58)

also suggests that immune related pathways are significantly altered in breast cancer [38–41].

Furthermore, the drugs identified for these targets added valuable information regarding the inhibition of these targets and the discovery of novel FDA approved drugs. Furthermore, survival analysis extensively cleared the role of these hub genes in the progression and breast cancer. Because the compliance of patients to adjuvant treatment is different, this may definitely influence the treatment result. This bioinformatics model suggested a catalogue of candidate cellular proteins that could be the targets for breast cancer therapy that are recognized as key genes in breast cancer.

Table 4 Showing the survival values of each of the hub genes in different clinical conditions of breast cancer

S. No	Gene name	*p*-value			
		Luminal A	Luminal B	HEr2	Basal
1	VEGFA	0.6659	0.0464	0.002	0.0462
2	CCND1	0.0003	0.0024	0.0002	0.3991
3	CXCL8	0.0348	0.8219	0.7441	0.0035
4	CCL2	0.4091	0.1116	0.0025	0.3413
5	CDKN1A	0.1168	0.9111	0.2543	0.79
6	CDK4	0.0012	2.1e–6	0.8943	0.0763
7	CDK2	0.1533	0.7382	0.1103	0.9991
8	IFIT1	0.0014	0.4857	0.9991	0.3998
9	OASL	0.9941	0.0059	2.3e–5	0.0011
10	DDX58	0.4638	0.2361	0.0997	0.6295

5 Conclusion

In conclusion, this research offers some views on future biomarkers associated with breast cancer patient prognosis. This research further emphasizes the significance of PPI and TF network assessment as a powerful structure for gaining understanding into the main hub nodes affecting breast cancer's prognosis and recognizing future breast cancer biomarkers.

Author contributions AK, AAK, ZR and HFH conceptualized the methodology. AK, AAK and HFH did the analysis. AK, AS, SSA, FH and DQW wrote the manuscript. DQW supervised the study. All the authors approved the manuscript.

Funding Dong-Qing Wei is supported by the grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science

and Technology of China, the National Natural Science Foundation of China (Contract no. 61832019, 61503244), the Natural Science Foundation of Henan Province (162300410060) and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The computations were partially performed at the Center for High-Performance Computing, Shanghai Jiao Tong University.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

1. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, Carter A, Casey DC, Charlson FJ, Chen AZ (2016) Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. The Lancet 388(10053):1545–1602. https://doi.org/10.1016/S0140-6736(16)31678-6
2. Sarma H, Mattaparthi VSK (2019) Structure-based virtual screening of high-affinity ATP-competitive inhibitors against human lemur tyrosine Kinase-3 (LMTK3) domain: a novel therapeutic target for breast cancer. Interdiscip Sci 11(3):527–541. https://doi.org/10.1007/s12539-018-0302-7
3. Su L, Meng X, Ma Q, Bai T, Liu G (2018) LPRP: a gene–gene interaction network construction algorithm and its application in breast cancer data analysis. Interdiscip Sci 10(1):131–142. https://doi.org/10.1007/s12539-016-0185-4
4. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 6(7):e1000100. https://doi.org/10.1016/j.jclinepi.2009.06.006.
5. Stender JD, Kim K, Charn TH, Komm B, Chang KC, Kraus WL, Benner C, Glass CK, Katzenellenbogen BS (2010) Genome-wide analysis of estrogen receptor α DNA binding and tethering mechanisms identifies Runx1 as a novel tethering factor in receptor-mediated transcriptional activation. Mol Cell Biol 30(16):3943–3955. https://doi.org/10.1128/MCB.00118-10
6. Xia J, Gill EE, Hancock RE (2015) NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc 10(6):823. https://doi.org/10.1038/nprot.2015.052
7. Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods—a bioconductor package providing PCA methods for incomplete data. Bioinformatics 23(9):1164–1167
8. Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S (2006) Bioinformatics and computational biology solutions using R and Bioconductor. Springer Science & Business Media, (https://doi.org/10.1093/bioinformatics/btm069)
9. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
10. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461. https://doi.org/10.1002/jcc.21334
11. Chen Y-C, Hsiao C-C, Chen K-D, Hung Y-C, Wu C-Y, Lie C-H, Liu S-F, Sung M-T, Chen C-J, Wang T-Y (2013) Peripheral immune cell gene expression changes in advanced non-small cell lung cancer patients treated with first line combination chemotherapy. PLoS ONE One 8(2):e57053. https://doi.org/10.1371/journal.pone.0057053
12. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P (2016) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res gkw937 (doi: 10.1093/nar/gkw937)
13. Ma L, Huang Y, Zhu W, Zhou S, Zhou J, Zeng F, Liu X, Zhang Y, Yu J (2011) An integrated analysis of miRNA and mRNA expressions in non-small cell lung cancers. PLoS ONE One 6(10):e26502. https://doi.org/10.1371/journal.pone.0026502)
14. Galvan A, Frullanti E, Anderlini M, Manenti G, Noci S, Dugo M, Ambrogi F, De Cecco L, Spinelli R, Piazza R (2013) Gene expression signature of non-involved lung tissue associated with survival in lung adenocarcinoma patients. Carcinogenesis 34(12):2767–2773. https://doi.org/10.1093/carcin/bgt294
15. Frullanti E, Colombo F, Falvella FS, Galvan A, Noci S, De Cecco L, Incarbone M, Alloisio M, Santambrogio L, Nosotti M (2012) Association of lung adenocarcinoma clinical stage with gene expression pattern in noninvolved lung tissue. Int J Cancer 131(5):E643–E648. https://doi.org/10.1002/ijc.27426
16. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool BMC Bioinformatics 14(1): 128. Doi: 10.1186/1471-2105-14-128
17. Brandes U, Eiglsperger M, Herman I, Himsolt M, Marshall MS (2001) GraphML progress report structural layer proposal. In: International symposium on graph drawing, Springer, pp 501–512
18. Győrffy B, Surowiak P, Budczies J, Lánczky A (2013) Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS ONE One 8(12):e82241. https://doi.org/10.1371/journal.pone.0082241
19. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B (2011) survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics 27(22):3206–3208. https://doi.org/10.1093/bioinformatics/btr511
20. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z (2017) DrugBank 50: a major update to the DrugBank database for 2018. Nucleic Acids Res 46(D1):D1074–D1082. https://doi.org/10.1093/nar/gkx1037
21. Arora D, Chaudhary R, Singh A (2019) System biology approach to identify potential receptor for targeting cancer and biomolecular interaction studies of indole [2, 1-a] isoquinoline derivative as anticancerous drug candidate against it. Interdiscip Sci 11(1):125–134. https://doi.org/10.1007/s12539-017-0249-0
22. Wang Y, Cai Y, Miao Y (2015) Evolving-pattern analysis of transient and long-term biomarkers for cancers: hepatocellular carcinoma as a case. Interdiscip Sci 7(4):414–422. https://doi.org/10.1007/s12539-015-0276-7
23. Li G-G, Wang Z-Z (2009) Evaluation of similarity measures for gene expression data and their correspondent combined measures. Interdiscip Sci 1(1):72–80. https://doi.org/10.1007/s12539-008-0005-3
24. Hu B, Chang X, Liu X (2019) predicting functional modules of liver cancer based on differential network analysis. Interdiscip Sci 1–9. Doi: 10.1007/s12539–018–0314–3
25. Liu S, Wang X, Qin W, Genchev GZ, Lu H (2018) Transcription factors contribute to differential expression in cellular pathways in lung adenocarcinoma and lung squamous cell carcinoma. Interdiscip Sci 10(4):836–847. https://doi.org/10.1007/s12539-018-0300-9
26. Ni Q-S, Wang Z-Z, Li G-G, Wang G-Y, Zhao Y-J (2009) Effect of the quality of the interaction data on predicting protein function

from protein-protein interactions. Interdiscip Sci 1(1):40–45. https://doi.org/10.1007/s12539-008-0015-4
27. Ma R, Wang C, Wang J, Wang D, Xu J (2016) miRNA–mRNA Interaction Network in Non-small Cell Lung Cancer. Interdiscip Sci 8(3):209–219. https://doi.org/10.1007/s12539-015-0117-8
28. Khan A, Ali A, Junaid M, Liu C, Kaushik AC, Cho WC, Wei D-Q (2018) Identification of novel drug targets for diamond-blackfan anemia based on RPS19 gene mutation using protein-protein interaction network. BMC Syst Biol 12(4):39. https://doi.org/10.1186/s12918-018-0563-0
29. Selvaraj G, Kaliamurthi S, Kaushik AC, Khan A, Wei Y-K, Cho WC, Gu K, Wei D-Q (2018) Identification of target gene and prognostic evaluation for lung adenocarcinoma using gene expression meta-analysis, network analysis and neural network algorithms. J Biomed Inform 86:120–134. https://doi.org/10.1016/j.jbi.2018.09.004
30. Malhotra AG, Jha M, Singh S, Pandey KM (2018) Construction of a comprehensive protein-protein interaction map for Vitiligo disease to identify key regulatory elements: a systemic approach. Interdiscip Sci 10(3):500–514. https://doi.org/10.1007/s12539-017-0213-z
31. Nayarisseri A, Yadav M, Wishard R (2013) Computational evaluation of new homologous down regulators of translationally controlled tumor protein (TCTP) targeted for tumor reversion. Interdiscip Sci 5(4):274–279. https://doi.org/10.1007/s12539-013-0183-8
32. Liu Y, Tamimi RM, Collins LC, Schnitt SJ, Gilmore HL, Connolly JL, Colditz GA (2011) The association between vascular endothelial growth factor expression in invasive breast cancer and survival varies with intrinsic subtypes and use of adjuvant systemic therapy: results from the Nurses' Health Study. Breast Cancer Res Treat 129(1):175–184. https://doi.org/10.1007/s10549-011-1432-3
33. Roy PG, Pratt N, Purdie CA, Baker L, Ashfield A, Quinlan P, Thompson AM (2010) High CCND1 amplification identifies a group of poor prognosis women with estrogen receptor positive breast cancer. Int J Cancer 127(2):355–360. https://doi.org/10.1002/ijc.25034
34. Afzali F, Salimi M (2019) Unearthing regulatory axes of breast cancer circRNAs networks to find novel targets and fathom pivotal mechanisms. Interdiscip Sci:1–12. Doi: 10.1007/s12539–019–00339–6
35. PA Ruffini 2019 The CXCL8-CXCR1/2 Axis as a therapeutic target in breast cancer stem-like cells. Front Oncol 9 10.3389/fonc.2019.00040
36. Sun X, Glynn DJ, Hodson LJ, Huo C, Britt K, Thompson EW, Woolford L, Evdokiou A, Pollard JW, Robertson SA (2017) CCL2-driven inflammation increases mammary gland stromal density and cancer susceptibility in a transgenic mouse model. Breast Cancer Res 19(1):4. https://doi.org/10.1186/s13058-016-0796-z
37. Wei C-Y, Tan Q-X, Zhu X, Qin Q-H, Zhu F-B, Mo Q-G, Yang W-P (2015) Expression of CDKN1A/p21 and TGFBR2 in breast cancer and their prognostic significance. Int J Clin Exp Pathol 8(11):14619
38. Barupal DK, Gao B, Budczies J, Phinney BS, Perroud B, Denkert C, Fiehn O (2019) Prioritization of metabolic genes as novel therapeutic targets in estrogen-receptor negative breast tumors using multi-omics data and text mining. Oncotarget 10(39):3894. https://doi.org/10.18632/oncotarget.26995
39. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, Delorenzi M, Piccart M, Sotiriou C (2008) Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res 14(16):5158–5165. https://doi.org/10.1158/1078-0432.CCR-07-4756
40. Fasoulakis Z, Kolios G, Papamanolis V, Kontomanolis EN (2018) Interleukins associated with breast cancer. Cureus 10(11). Doi: 10.7759/cureus.3549
41. Nicolini A, Ferrari P, Diodati L, Carpi A (2018) Alterations of signaling pathways related to the immune system in breast cancer: new perspectives in patient management. Int J Mol Sci 19(9):2733. https://doi.org/10.3390/ijms19092733