JBUON 2018; 23(4): 891-901
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
E-mail: editorial_office@jbuon.com

ORIGINAL ARTICLE

Identification of key candidate genes, pathways and related prognostic values in ER-negative/HER2-negative breast cancer by bioinformatics analysis

Nan Shao[1]*, Kaitao Yuan[2]*, Yunjian Zhang[1], Tuck Yun Cheang[1], Jie Li[1], Ying Lin[1]

[1]Breast Disease Center and [2]Center of Gastrointestinal Surgery, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, 510080, China

*These authors contributed equally to this work.

Summary

Purpose: *Breast cancer possesses different molecular expressions and biological behaviors. The purpose of this study was to identify the key genes, pathways, and related prognostic values in estrogen receptor (ER)-negative/human epidermal growth factor 2 (HER2)-negative breast cancer by bioinformatics analysis.*

Methods: *The mRNA expression profiles of GSE20194 and GSE23988 were obtained from the Gene Expression Omnibus (GEO) database. Differently expressed genes (DEGs) were analyzed by GEO2R. A functional and pathway enrichment analysis of DEGs was conducted using DAVID. A protein-protein interaction (PPI) network was constructed using STRING and a module analysis of the PPI network was conducted using Cytoscape software. Survival analysis of hub genes was analyzed using the Kaplan- Meier plotter online tool.*

Results: *108 ER-negative/HER2-negative and 172 ER-positive/HER2-negative breast cancer samples were collected from the datasets GSE20194 and GSE23988. A total of 355 DEGs were identified in the ER-negative/HER2-negative samples, including 140 up-regulated and 215 down-regulated genes. The PPI network of DEGs consisted of 265 nodes and 648 edges. A significant module (12 nodes and 56 edges) was acquired from the PPI network of DEGs. Geneontology (GO) and pathway enrichment analysis demonstrated that this module was mainly related with transcription, cell proliferation, binding, and pathways in the PI3K-Akt signaling pathway. The high expression of CCNE1, KRT16, and MYBL2 was associated with worse relapse-free survival (RFS) and overall survival (OS) in ER-negative/HER2-negative breast cancer.*

Conclusions: *An integrated bioinformatics analysis was utilized to discover key candidate genes and pathways in ER-negative/HER2-negative breast cancer. This can improve the understanding of molecular mechanisms and provide potential candidate genes for diagnosis, prognosis, and individualized therapy.*

Key words: *bioinformatics analysis, breast cancer, differently expression genes, survival*

Introduction

Breast cancer is widely considered a heterogeneous disease that has differing molecular expressions and biological behaviors. Gene microarray technology and immunohistochemical techniques have classified breast cancers into hormone receptor positive (luminal A and luminal B), human epidermal growth receptor 2 (HER2)-positive, and basal-like type [1]. The differences of expression

Correspondence to: Ying Lin MD, PhD. Breast Disease Center, The First Affiliated Hospital, Sun Yat-Sen University, No.57 Zhong Shan Er Lu, Guangzhou, 510080, China.
Tel: +86 20-87755766, ext 8198, E-mail: linying3@mail.sysu.edu.cn
Received: 06/11/2017; Accepted: 24/11/2017

of genes and patterns of mutation may result in biological and clinical variations between ER-positive and ER-negative breast cancers. ER-negative breast cancers correlate with increased histological grade, sensitivity to chemotherapy, and metastasize to visceral organs [2,3]. Therefore, understanding the molecular mechanism of ER-negative breast cancer is critically demanded.

To date, high-throughput platforms used to analyze gene expression can screen out the hundreds of DEGs that mediate different biological and molecular processes, as well as pathways during tumorigenesis [4]. However, results were always limited or contradictory between breast cancer studies [5]. Most microarray data has been saved in public databases, such as GEO and ArrayExpress Archive of Functional Genomics Data. The integrated bioinformatics approach combined with expression profiling techniques may offer valuable evidence for additional exploration.

In this study, two original gene expression datasets, namely GSE20194 and GSE23988, were downloaded from the GEO repository. Using the web-based tool GEO2R, DEGs were screened between ER-negative/HER2-negative and ER-positive/HER2-negative breast cancer samples. Subsequently, GO and pathway enrichment analysis of the DEGs were screened using DAVID Bioinformatics Resources, PANTHER Classification System, and KEGG Pathway. The PPI network of the DEGs and modular analysis were used to identify hub genes in ER-negative/HER2-negative breast cancer samples. The genes were analyzed and visualized using the online STRING database and Cytoscape software. The Kaplan Meier plotter was utilized to combine the survival analysis of the hub genes. Finally, the series of bioinformatics analysis may provide possible genes for diagnosis, prognosis, and individualized breast cancer therapy.

Methods

Microarray data

The gene expression profiles of GSE20194 and GSE23988 were accessed from the GEO repository (available at: https://www.ncbi.nlm.nih.gov/geo/). Both gene profiles were built on the Affymetrix GPL96 platform (Affymetrix Human Genome U133A Array). Excluding 59 HER2-positive breast cancer samples, the GSE20194 dataset contained 219 samples, including 79 ER-negative/HER2-negative breast cancer samples and 140 ER-positive/HER2-negative breast cancer samples [6]. GSE23988 consisted of 29 ER-negative/HER2-negative breast cancer samples and 32 ER+/HER2- breast cancer samples [7].

Identification of EDGs

The identification of DEGs was carried out using the web-based tool GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/). GEO2R enables comparison between at least two groups of samples in a GEO series to categorize differentially expressed genes in several experimental settings. The adjusted p values (adj.P) are frequently utilized for microarray data and offer balance between the detection of statistically significant genes and the limits of false positives. The adj. $p<0.05$ and $|logFC|>1$ were defined as cut-off conditions to identify the statistically significant DEGs.

Gene ontology and pathway enrichment analysis

The DAVID Bioinformatics Resources [8] (available at: https://david.ncifcrf.gov/) and PANTHER Classification System [9] (available at: http://www.pantherdb.org/) are web-accessible programs that provide a detailed set of functional annotation tools to determine the function and utility of biological systems using molecular-level information, for example, extensive molecular datasets produced by genome sequencing and investigational technologies. GO enrichment and KEGG pathway analysis were conducted with the DAVID online tool. A p value <0.05 was chosen as the cut-off criterion to analyze the DEGs at the functional level.

Protein-protein interaction network and modules analysis

The PPI network of DEGs was constructed by the online STRING database [10] (available at: http://string-db.org/). Cytoscape software was utilized to explore the relationship of the key DEGs in ER-negative/HER2-negative breast cancer [11]. Apps of Cytoscape software, NetworkAnalyzer, and the Molecular Complex Detection (MCODE) were utilized to measure node degree. This is the number of inter-connections to filter hub genes of PPI and screen modules of the PPI network, with cut-off= 2, node score cut-off= 0.2, k-core= 2, and max. depth= 100. The functional enrichment analysis of hub genes in each module was performed using DAVID.

Survival analysis of up-regulated hub genes

RFS and OS of breast cancer patients were investigated by the Kaplan Meier plotter (available at: http://kmplot.com/analysis/), which assesses the influence of 54,675 genes on survival using breast, ovarian, lung, and gastric cancer patients, with a mean follow-up of 69, 40, 49, and 33 months [12]. The hazard ratio (HR) with 95% confidence intervals and log rank p values were quantified based on the online database.

Results

Identification of DEGs in ER-negative/HER2-negative breast cancer

A total of 108 ER-negative/HER2-negative and 172 ER-positive/HER2-negative breast cancer sam-

ples were collected from GSE20194 and GSE23988. Based on the GEO2R analysis, 588 and 626 DEGs were extracted from the expression profile datasets GSE20194 and GSE23988, respectively. After integrated bioinformatics analysis, 355 consistently expressed genes were screened out in datasets using the web-based tool 'calculate and draw custom Venn diagrams' (available at: http://bioinformatics.psb.ugent.be/webtools/Venn/) (Figure 1), including 140 up-regulated and 215 down-regulated genes in ER-negative/HER2-negative breast cancer samples compared to ER-positive/HER2-negative breast cancer samples (Table 1).

Figure 1. Identification of consistently DEGs from the datasets GSE20194 and GSE23988. 355 consistently expressed genes were screened out in datasets by the web-based tool "Calculate and draw custom Venn diagrams" (available at: http://bioinformatics.psb.ugent.be/webtools/Venn/).

GO and pathway enrichment analysis

To investigate the function and pathway of candidate DEGs, DAVID and PANTHER online database were used to perform the functional enrichment analysis. The GO analyses of DEGs were categorized into three functional groups: biological process (BP), molecular function (MF), and cellular component (CC) (Figures 2 and 3). Biological process is associated with cellular, metabolic, and development processes. For molecular function, DEGs possessed enriched catalytic, binding, and

Table 1. Identification of 355 consistently expressed genes from the datasets GSE20194 and GSE23988

Expression	*DEGs*
Up-regulated (140)	HACD1, CDH3, MICALL1, C1orf106, CAPN6, BBOX1, PDE9A, TTYH1, KCNG1, GLS, GSTA1, HSD17B2, MPZL2, CYP39A1, CRABP1, ACTG2, KLK6, MMP7, KRT17, FZD7, APBA2, SERPINB5, TMEM158, ZIC1, WWTR1, PKP1, ST8SIA1, NXN, FABP7, FERMT1, SMOX, ANP32E, ARHGEF4, UBE2E3, ELF5, RARRES1, SOX10, FAT1, PTK7, CSRP2, SERPINE2, MFGE8, FOLH1B, SNORA11E, ART3, NUDT11, CLDN8, ST3GAL6, PHGDH, UCHL1, PAX6, GABRE, NFIB, SLPI, GPM6B, RRAGD, CBS, CLDN1, EGFR, CAV2, SOX11, KRT5, TMSB15B, CCNE1, GABBR2, MAP7D3, HLA-DOB, FSCN1, DZIP1, GPR161, CHI3L2, SMCO4, CALD1, GATA6, BCL11A, ITM2C, KRT6B, PSAT1, FOXC1, TRIM2, CALB2, LAMP3, FABP5, ZNF750, ROPN1B, LY6D, MCM5, PLIN2, CRYAB, MAP4K4, GBP1, TSPYL5, PTX3, KCNK5, LPIN1, TRIM29, CDC20, MPP6, KRT16, CDKN2A, EN1, GABRP, KHDRBS3, CX3CL1, LDHB, VSNL1, PLS3, TTLL4, ROR1, PRKD3, UGT8, MIA, PI3, SLC6A14, GAL, ADGRL2, CHODL, JRKL, CHI3L1, SFRP1, RNF144A, PRNP, SLC16A1, DSC2, HRASLS, MYBL2, MYO10, GSTP1, PROM1, GJB3, SCRG1, BTG3, CYP1B1, CHST3, SIX3, VGLL1, MTMR2, FAM171A1, PLAGL1, CBR1
Down-regulated (215)	NPDC1, CREB3L1, TJP3, TSPAN1, ERBB4, HDGFRP3, KRT19, UGCG, SH3BGRL, ARMT1, APBB2, CRIP1, HSPB1, TP53TG1, EPS8L1, SIGIRR, HSPB8, SYT1, GDF15, SIDT1, NPRL2, FAM134B, RTN1, SIAH2, RARA, CLGN, SLC2A10, DCXR, CEACAM6, MREG, COX6C, CYP2B7P, KCND3, GUSBP3, GATA3, MED13L, PLK2, GJA1, ELOVL5, SCUBE2, PIP, ENPP1, SCNN1A, ACADSB, HSPA2, INPP4B, HGD, PDZK1, PNPLA4, UGDH, FAM198B, TIMP3, CIRBP, SMA4, CUEDC1, DACH1, SSH3, SLC16A6, PLA2G16, ITPR1, PLAT, C16orf45, TSPAN13, BLVRA, GAMT, GPRC5A, GALNT6, GFRA1, REPS2, SLC27A2, DUSP4, ABLIM3, CPE, COQ4, CCND1, IGFBP2, MYO6, STC2, EVA1B, BTD, TTC39A, REEP5, FOXA1, NPY1R, GREB1, ESR1, KITLG, SCCPDH, SLC1A1, SLC1A4, CHST15, LGALS8, GALNT7, TMCO6, CERS6, MUC1, RHOB, BCL2, IGFBP4, C8orf4, LRIG1, SPDEF, ABHD2, CYB5A, PTP4A2, TMC5, LMF1, CPB1, XYLT2, GLUL, NME3, TOX3, DNAJC12, AREG, IGF1R, MAST4, ABCC8, ELOVL2, MAGED2, CAMK2N1, ACOX2, EFHC1, IRS1, SERPINA3, SCGB2A2, ABAT, C4B_2, ECI1, NAT1, MLPH, RAI2, TMBIM6, TBC1D9, TFF1, CELSR1, NRIP1, AR, TBX3, ZNF552, HOXB2, PLEKHF2, SCGB2A1, IL6ST, FGFR3, PRSS23, METRN, CLSTN2, LAMA3, SCGB1D2, TFF3, KDM4B, KCNK15, ADCY9, CST3, GPD1L, SERPINA5, AMIGO2, G6PC3, VAV3, RABEP1, FBP1, FXYD3, INPP5J, KRT18, FUT8, SLC7A8, ITGB5, EVL, WWP1, CA12, BHLHE40, KCNE4, CFD, MISP, CHAD, TPBG, RET, MAPT, IGFBP5, SEMA3B, SLC9A3R1, SLC24A1, POLD4, SEMA3C, CYP2B7P, MZF1, GSTM3, CYP4B1, MCCC2, REEP1, GALNT10, ADIRF, AKR7A3, ADCY1, XBP1, CANT1, MSX2, SLC19A2, AGR2, PGR, CERS4, TNNT1, FAM174B, ANXA9, DHRS2, ASPN, AFF3, KIAA0040, SLC7A2, EEF1A2, DNALI1, SYBU, MYB, SLC39A6, SLC44A4

140 up-regulated genes and 215 down-regulated genes in ER-negative/HER2-negative breast cancer samples compared to ER-positive/HER2-negative breast cancer samples.

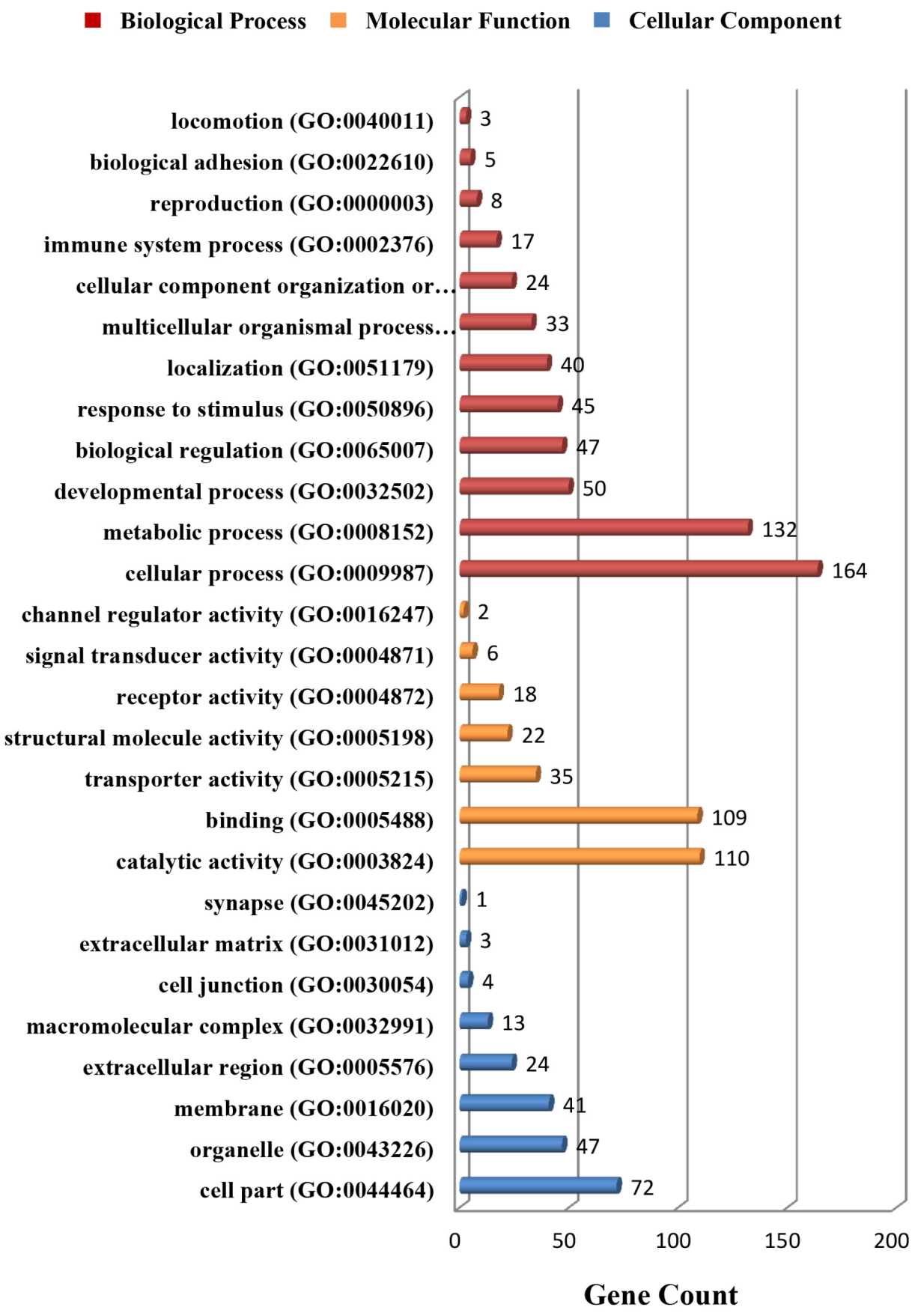

Figure 2. Gene ontology analysis of DEGs into 3 groups (Enrichment analysis by PANTHER). To investigate the function and pathway of candidate DEGs, the PANTHER online database was used to perform the functional enrichment analysis. The GO analyses of DEGs were grouped into three functional groups: biological process (BP), molecular function (MF), and cellular component (CC).

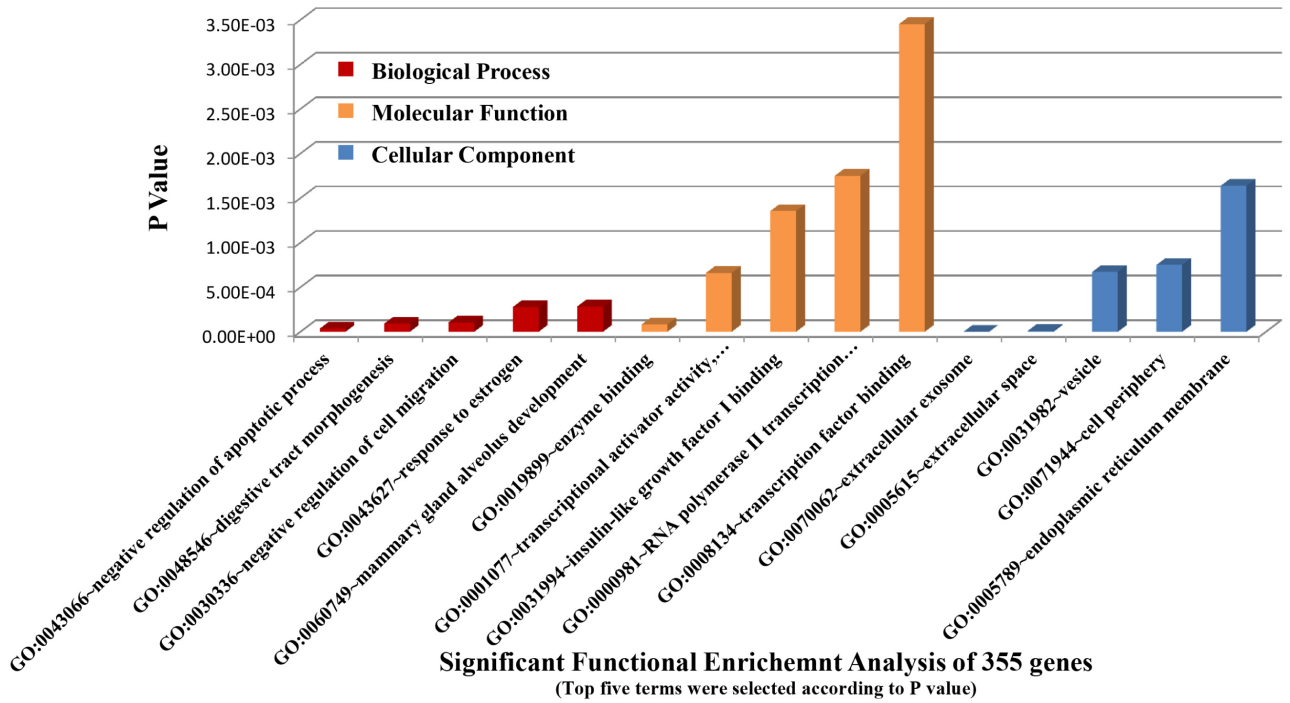

Figure 3. Functional enrichment analysis of 355 consistently expressed genes in ER-/HER2- breast cancer (Enrichment analysis by DAVID). To investigate the function and pathway of candidate DEGs, the DAVID online database was used to perform the functional enrichment analysis. The GO analyses of DEGs were classified into three functional groups: biological process (BP), molecular function (MF), and cellular component (CC).

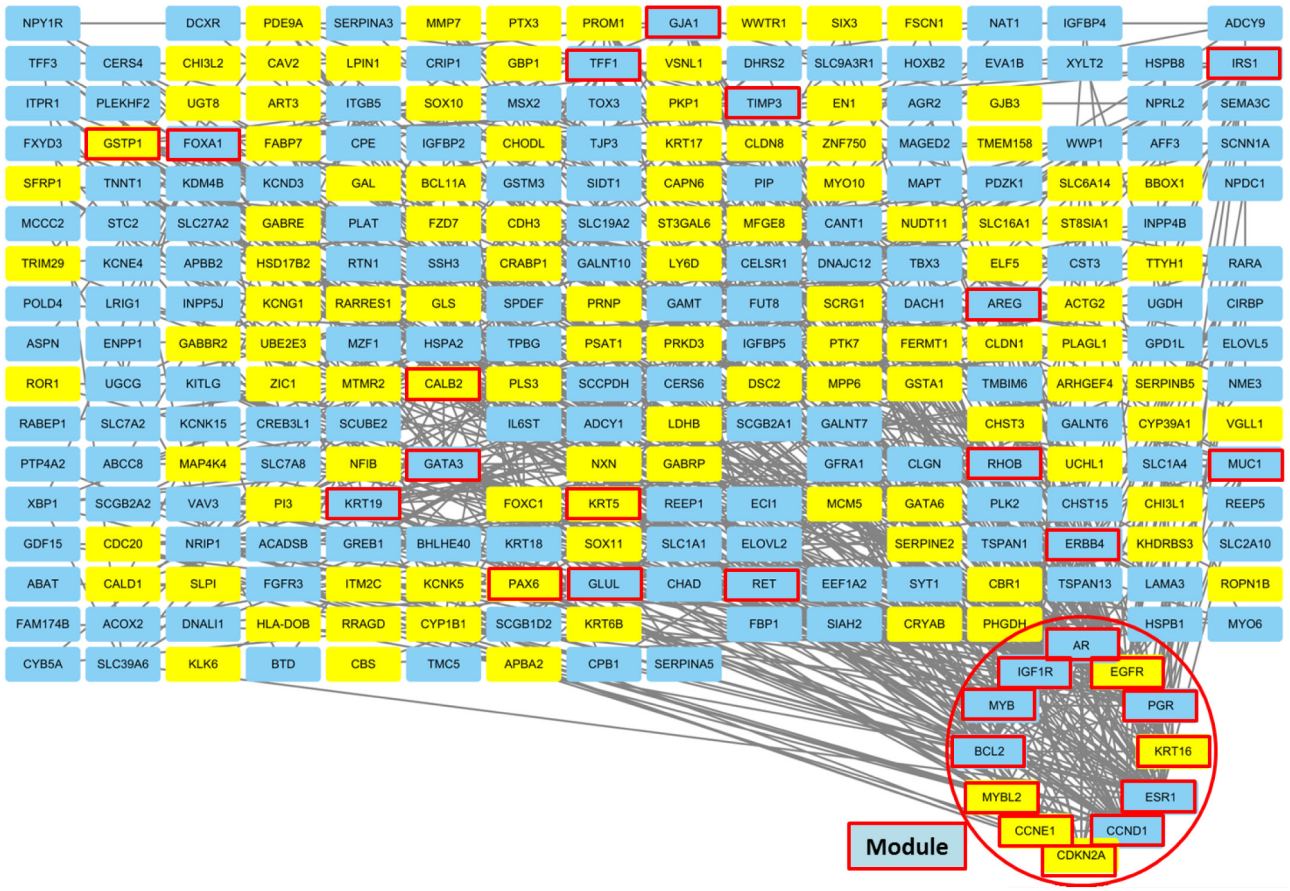

Figure 4. PPI network of DEGs and modular analysis. A total of 265 DEGs (110 up-regulated genes in yellow nodes and 155 down-regulated genes in blue nodes) were analyzed by STRING and Cytoscape. A significant module (12 nodes and 56 edges) was selected from protein-protein interaction network. Red frames: 29 hub genes.

Table 2. Gene ontology analysis of DEGs in ER-negative/HER2-negative breast cancer (Enrichment analysis by DAVID)

Category	*Term*
Up-regulated	
GOTERM_BP_DIRECT	GO:0016337~single organismal cell-cell adhesion
GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter
GOTERM_BP_DIRECT	GO:0006366~transcription from RNA polymerase II promoter
GOTERM_MF_DIRECT	GO:0000981~RNA polymerase II transcription factor activity, sequence-specific DNA binding
GOTERM_MF_DIRECT	GO:0042802~identical protein binding
GOTERM_MF_DIRECT	GO:0003705~transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding
GOTERM_CC_DIRECT	GO:0070062~extracellular exosome
GOTERM_CC_DIRECT	GO:0005829~cytosol
GOTERM_CC_DIRECT	GO:0005737~cytoplasm
Down-regulated	
GOTERM_BP_DIRECT	GO:0060749~mammary gland alveolus development
GOTERM_BP_DIRECT	GO:0043066~negative regulation of apoptotic process
GOTERM_BP_DIRECT	GO:0032355~response to estradiol
GOTERM_MF_DIRECT	GO:0031994~insulin-like growth factor I binding
GOTERM_MF_DIRECT	GO:0019899~enzyme binding
GOTERM_MF_DIRECT	GO:0005102~receptor binding
GOTERM_CC_DIRECT	GO:0070062~extracellular exosome
GOTERM_CC_DIRECT	GO:0005615~extracellular space
GOTERM_CC_DIRECT	GO:0005789~endoplasmic reticulum membrane

Top three terms were selected according to p value. Gene count: the number of enriched genes in each term. GO: Gene ontology, BP: Biological process, MF: Molecular function, CC: Cellular component

transporter activities. As shown in Table 2, GO analysis revealed that up-regulated genes formed part of the biological development related with the regulation of single organism cell-cell adhesion, positive regulation of transcription from RNA polymerase II promoter, and transcription from the RNA polymerase II promoter, whereas down-regulated genes were mainly enriched in mammary gland alveolus development and negative in the regulation of apoptotic process and response to estradiol. As for molecular function, up-regulated genes were mainly enriched in RNA polymerase II transcription factor activity, sequence-specific DNA binding, identical protein binding and transcription factor activity, and RNA polymerase II distal enhancer sequence-specific binding, whereas, down-regulated genes were mainly enriched in insulin-like growth factor I enzyme, and receptor binding. The DAVID and PANTHER analysis demonstrated that most of the DEGs were significantly enriched in cellular and metabolic process, binding and extracellular exosome.

DEGs functional and signaling pathway enrichment were analyzed by the KEGG PATHWAY. As shown in Table 3, signaling pathway analysis demonstrated that DEGs were enriched in metabolic pathways, HTLV-I infection, proteoglycans in cancer, and the estrogen signaling pathway.

PPI network and modules analysis

To detect the key candidate genes and pathways, the STRING online database and Cytoscape software were used to construct the PPI network and modules. The PPI network of DEGs comprised of 265 nodes and 648 edges, and included 110 up-regulated and 155 down-regulated genes (Figure 4). 29 hub genes (red frames in Figure 4) were identified from the 265 nodes (NetworkAnalyzer, Degree≥10). A significant module (12 nodes and 56 edges) was also attained from the PPI network of DEGs using Cytoscape MCODE. GO and pathway enrichment analysis showed that this module was connected with transcription, cell proliferation, binding, pathways in cancer, and the PI3K-Akt signaling pathway (Table 4).

The Kaplan-Meier plotter

The prognostic values of 5 up-regulated hub genes (CCNE1, CDKN2A, EGFR, KRT16 and MBL2)

Table 3. KEGG pathway analysis of differentially expressed genes

Pathway ID	*Name*	*Count*	*p value*	*Genes*
hsa04727	GABAergic synapse	8	0.00	GABRE, ADCY1, GLUL, ADCY9, GLS, ABAT, GABBR2, GABRP
hsa04915	Estrogen signaling pathway	8	0.01	EGFR, ADCY1, HSPA2, ADCY9, ESR1, CREB3L1, GABBR2, ITPR1
hsa04114	Oocyte meiosis	8	0.01	PGR, CCNE1, IGF1R, ADCY1, AR, ADCY9, CDC20, ITPR1
hsa05215	Prostate cancer	7	0.02	EGFR, CCNE1, IGF1R, AR, CCND1, BCL2, CREB3L1
hsa01100	Metabolic pathways	40	0.02	ACOX2, LDHB, ACADSB, FUT8, HSD17B2, ENPP1, BTD, GALNT7, GALNT6, ST8SIA1, CERS6, UGDH, CERS4, G6PC3, MTMR2, MCCC2, CBR1, GALNT10, XYLT2, INPP5J, ST3GAL6, UGT8, PLA2G16, NAT1, UGCG, FBP1, HGD, LPIN1, COX6C, POLD4, GLUL, NME3, GLS, PHGDH, ABAT, INPP4B, GAMT, PSAT1, DCXR, CBS
hsa04913	Ovarian steroidogenesis	5	0.03	IGF1R, ADCY1, CYP1B1, HSD17B2, ADCY9
hsa00980	Metabolism of xenobiotics by cytochrome P450	6	0.03	GSTA1, GSTM3, CBR1, CYP1B1, AKR7A3, GSTP1
hsa05166	HTLV-I infection	12	0.03	MSX2, POLD4, ADCY1, CCND1, CDKN2A, ADCY9, XBP1, CDC20, MYB, MYBL2, HLA-DOB, FZD7
hsa05204	Chemical carcinogenesis	6	0.04	GSTA1, GSTM3, CBR1, CYP1B1, NAT1, GSTP1
hsa04923	Regulation of lipolysis in adipocytes	5	0.04	ADCY1, PLA2G16, ADCY9, NPY1R, IRS1
hsa05205	Proteoglycans in cancer	10	0.04	EGFR, IGF1R, CAV2, CCND1, ERBB4, ESR1, ITGB5, TIMP3, FZD7, ITPR1

Table 4 Functional and pathway enrichment analysis of the hub genes in module

Category	*Term*	*Count*	*p value*	*Genes*
GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	7	1.4E-05	PGR, EGFR, AR, CDKN2A, ESR1, MYB, MYBL2
GOTERM_BP_DIRECT	GO:0060749~mammary gland alveolus development	3	5.3E-05	AR, CCND1, ESR1
GOTERM_BP_DIRECT	GO:0045893~positive regulation of transcription, DNA-templated	5	2.4E-04	CCNE1, AR, CDKN2A, ESR1,MYB
GOTERM_BP_DIRECT	GO:0008283~cell proliferation	4	1.5E-03	EGFR, AR, KRT16, BCL2
GOTERM_BP_DIRECT	GO:0000082~G1/S transition of mitotic cell cycle	3	1.9E-03	CCNE1, CCND1, CDKN2A
GOTERM_MF_DIRECT	GO:0001077~transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	5	1.1E-05	PGR, AR, ESR1, MYB, MYBL2
GOTERM_MF_DIRECT	GO:0008134~transcription factor binding	5	2.4E-05	AR, CCND1, CDKN2A, BCL2, ESR1
GOTERM_MF_DIRECT	GO:0019899~enzyme binding	5	4.4E-05	PGR, EGFR, AR, CCND1, ESR1
GOTERM_MF_DIRECT	GO:0000978~RNA polymerase II core promoter proximal region sequence-specific DNA binding	5	5.6E-05	PGR, AR, ESR1, MYB, MYBL2
GOTERM_MF_DIRECT	GO:0005496~steroid binding	3	1.3E-04	PGR, AR, ESR1
KEGG_PATHWAY	hsa05215:Prostate cancer	6	7.2E-08	EGFR, CCNE1, IGF1R, AR, CCND1, BCL2
KEGG_PATHWAY	hsa05200:Pathways in cancer	7	5.6E-06	EGFR, CCNE1, IGF1R, AR, CCND1, CDKN2A, BCL2
KEGG_PATHWAY	hsa04151:PI3K-Akt signaling pathway	6	6.2E-05	EGFR, CCNE1, IGF1R, CCND1, BCL2, MYB

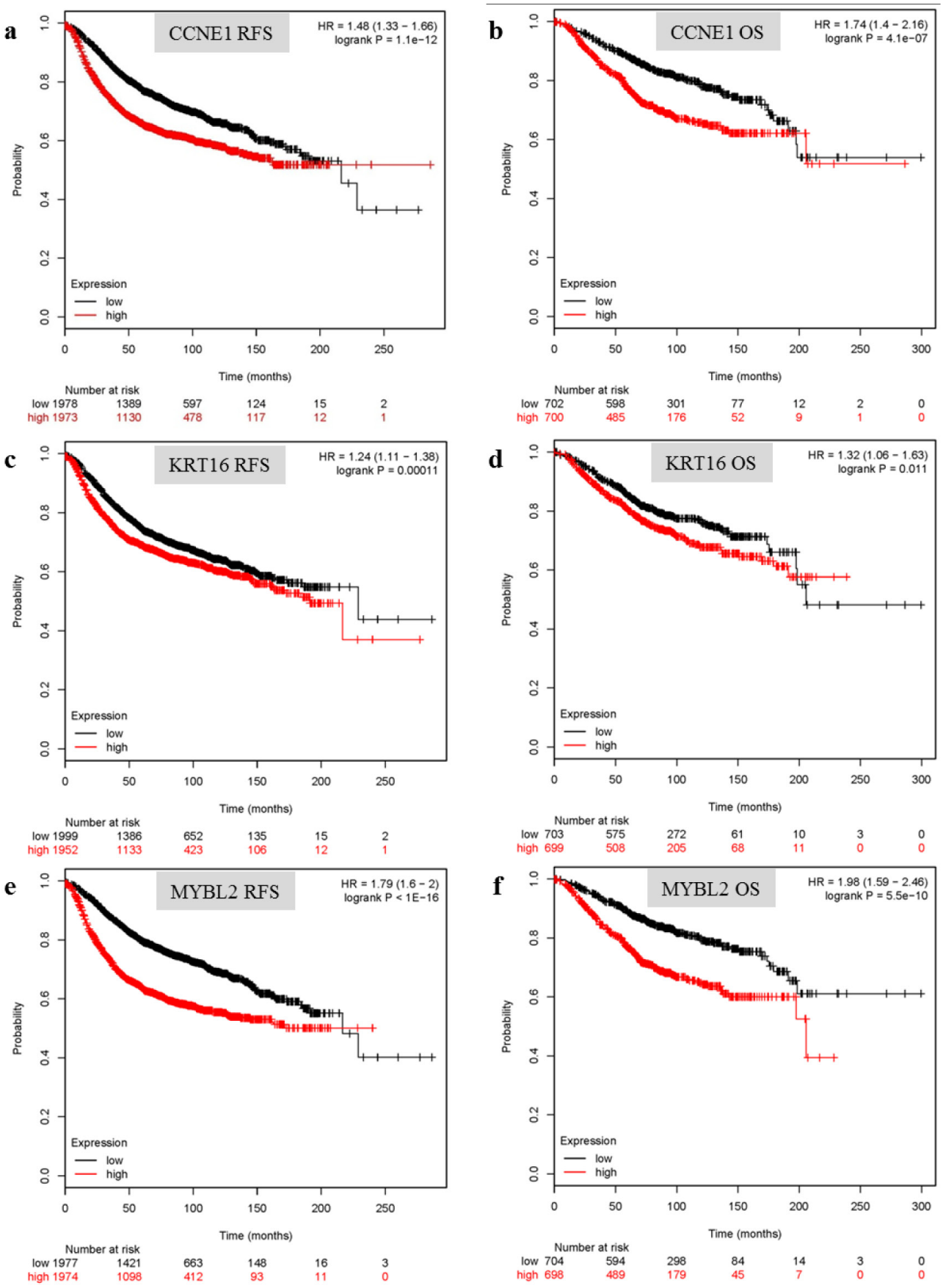

Figure 5. Prognostic value of up-regulated hub genes in breast cancer patients (Kaplan-Meier Plotter). The high expression of CCNE1, KRT16, and MYBL2 was associated with worse RFS and OS analyzed by the Kaplan-Meier plotter (available at: http://kmplot.com/analysis).

in PPI network were assessed in the Kaplan- Meier Plotter. As shown in Figure 5, the high expression of CCNE1, KRT16 and MYBL2 was associated with worse RFS and OS. There was no association between the survival and the expression of EGFR and CDKN2A (data not shown).

Discussion

Microarray and high-throughput sequencing have been widely used in clinical practice to reveal general genetic alteration and to identify targets for diagnosis, therapy, and prognosis in ER-positive breast cancer, such as Oncotype DX Recurrence Score (Genomic Health, CA) and MammaPrint (Agendia, Inc, the Netherlands) [5]. However, no comparable and validated microarray exists for ER-negative breast cancer, suggesting different gene expression, biological processes and signaling pathways in different subtypes of breast cancer.

In the present study, two cohort profiles datasets were deeply analyzed by bioinformatics methods to identify 355 commonly changed DEGs, including 140 up-regulated and 215 down-regulated genes in ER-negative/HER2-negative compared to ER-positive/HER2-negative breast cancer samples. The up and down-regulated DEGs were grouped based on functions and signaling pathways with marked enrichment analysis. As shown in Table 2, 140 up-regulated genes were mainly enriched in cell-cell adhesion, protein binding, and positive regulation of transcription from the RNA polymerase II promoter, which were all linked to cancer progression. Specifically, the regulation of transcription from the RNA polymerase II promoter is necessary to properly regulate the development, growth and existence of eukaryotic organisms [13]. The TATA box, initiator (Inr), TFIIB recognition element (BRE), and downstream core promoter element (DPE) all play critical roles in cancer progression, and are necessary to accurately initiate transcription by RNA polymerase II machinery [13,14], whereas the 215 down-regulated genes were mainly enriched in the negative regulation of apoptotic processes, enzyme, and receptor binding. The signaling pathway analysis demonstrated that the down-regulated genes were associated with metabolic pathways and PI3K-Akt signaling pathways.

Subsequently, the PPI network of DEGs was created and 265 nodes/DEGs and 648 edges were identified. Meanwhile, one significant module was filtered from the PPI network complex. The module consisted of 12 genes, including EGFR, ESR1, CCND1, BCL2, AR, PGR, CDKN2A, IFG1R, MYB, KRT16, MYBL2 and CCNE1, and is recorded at the top of many degree hub genes. GO and pathway enrichment analysis showed that this module is also connected with transcription, cell proliferation, binding and the PI3K-Akt signaling pathway. The PI3K-Akt signaling pathway plays a critical role in tumorigenesis and mediates critical cellular functions, including survival, proliferation, and metabolism. The pathway is connected with better outcomes in ER-positive early-stage breast cancer and tends to a worse prognosis in ER-negative subtype [15,16]. In the present analysis, EGFR, CCNE1, IGF1R, CCND1, BCL2 and MYB were enriched in the PI3K-Akt signaling pathway.

EGFR, KRT16, MYBL2, CCNE1 and CDKN2A were up-regulated hub genes. Survival analysis of these 5 up-regulated hub genes revealed that CCNE1, KRT16, and MYBL2 were associated with worse RFS and OS with breast cancer. CCNE1 (cyclin E1) operates as a subunit of CDK2 and is essential for cell cycle G1/S transition [17]. Another gene, MYBL2, is also connected with cell cycle progression. The nuclear protein is phosphorylated by cyclin A/CDK2 during the S-phase of the cell cycle, has both activator and repressor activities and triggers cell division cycle 2, cyclin D1, and insulin-like growth factor-binding protein 5 genes. CCNE1 and MYBL2 overexpression can cause chromosome instability and possibly tumorigenesis [17-20]. Meta-analyses have shown that the overexpression of CCNE1 is an autonomous prognostic feature for both overall and breast cancer-specific survival [21,22]. Another meta-analysis reported that CCNE1 might be a prognostic marker for gastrointestinal cancer in clinical practice [23]. MYBL2 mediates snail expression to induce epithelial-to-mesenchymal transition (EMT) and invasion of breast cancer cells [24]. To date, the link between MYBL2 and survival in breast cancer has not been known. KRT16 is a keratin gene family member and is expressed in the cytoskeleton of epithelial cells [25]. During the EMT process, epithelial markers, including E-cadherin, claudins, and keratins are down-regulated [26]. Two earlier studies of breast cancer patients showed that the overexpression of KRT16 in the primary tumor had a shorter RFS in comparison to patients with KRT16 low expression [27,28]. A recent study of circulating tumor cells (CTCs) reported that KRT16 was up-regulated in basal-like breast cancer cell lines and that KRT16 expressions of CTCs were associated with shorter RFS in metastatic breast cancer patients [25]. The expression of KRT16 in primary breast cancer and CTCs may be further examined for potential clinical application.

In conclusion, this study has provided a comprehensive bioinformatics analysis of DEGs to discover the key candidate genes, pathways, and related prognostic values in ER-negative/HER2-negative breast cancer. A total of 12 hub genes (EGFR, ESR1, CCND1, BCL2, AR, PGR, CDKN2A, IFG1R, MYB, KRT16, MYBL2, and CCNE1) were screened out and were significantly enriched in the positive regulation of transcription from RNA polymerase II promoters, cell proliferation, binding, and the PI3K-Akt signaling pathway. Among these hub genes, the high expression of CCNE1, KRT16 and MYBL2 was associated with worse RFS and OS of breast cancer patients. Even though these findings provide the potential genes for diagnosis, prognosis and individualized therapy of breast cancer, additional molecular biological experiments are mandatory to investigate the identified genes in ER-negative/HER2-negative breast cancer.

Acknowledgements

This work was supported by Natural Science Foundation of Guangdong Province, China (2016A030310134).

Conflict of interests

The authors declare no conflict of interests.

References

1. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61-70.
2. Andre F, Job B, Dessen P et al. Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. Clin Cancer Res 2009;15:441-51.
3. Hess KR, Pusztai L, Buzdar AU, Hortobagyi GN. Estrogen receptors and distinct patterns of breast cancer relapse. Breast Cancer Res Treat 2003;78:105-18.
4. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LJ, Kinzler KW. Cancer genome landscapes. Science 2013;339:1546-58.
5. Harris LN, Ismaila N, McShane LM et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 2016;34:1134-50.
6. Popovici V, Chen W, Gallas BG et al. Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. Breast Cancer Res 2010;12:R5.
7. Iwamoto T, Bianchini G, Booser D et al. Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer. J Natl Cancer Inst 2011;103:264-72.
8. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.
9. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res 2013;41:D377-86.
10. Szklarczyk D, Franceschini A, Wyder S et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.
11. Shannon P, Markiel A, Ozier O et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-2504.
12. Gyorffy B, Lanczky A, Eklund AC et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat 2010;123:725-31.
13. Butler JE, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev 2002;16:2583-92.
14. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 2006;7:29-59.
15. Yang SX, Polley E, Lipkowitz S. New insights on PI3K/AKT pathway alterations and clinical outcomes in breast cancer. Cancer Treat Rev 2016;45:87-96.
16. Mukohara T. PI3K mutations in breast cancer: prognostic and therapeutic implications. Breast Cancer (Dove Med Press) 2015;7:111-23.
17. Mazumder S, DuPree EL, Almasan A. A dual role of cyclin E in cell proliferation and apoptosis may provide a target for cancer therapy. Curr Cancer Drug Targets 2004;4:65-75.
18. Huang L, Ren F, Tang R, Feng Z, Chen G. Prognostic Value of Expression of Cyclin E in Gastrointestinal Cancer: A Systematic Review and Meta-Analysis. Technol Cancer Res Treat 2016;15:12-9.
19. Nakayama N, Nakayama K, Shamima Y et al. Gene amplification CCNE1 is related to poor survival and potential therapeutic target in ovarian cancer. Cancer 2010;116:2621-34.
20. Tao D, Pan Y, Lu H et al. B-myb is a gene implicated in cell cycle and proliferation of breast cancer. Int J Clin Exp Pathol 2014;7:5819-27.
21. Wang L, Shao ZM. Cyclin E expression and prognosis in breast cancer patients: a meta-analysis of published studies. Cancer Invest 2006;24:581-7.
22. Gao S, Ma JJ, Lu C. Prognostic value of cyclin E expression in breast cancer: a meta-analysis. Tumour Biol 2013;34:3423-30.

23. Huang L, Ren F, Tang R, Feng Z, Chen G. Prognostic Value of Expression of Cyclin E in Gastrointestinal Cancer: A Systematic Review and Meta-Analysis. Technol Cancer Res Treat 2016;15:12-9.
24. Tao D, Pan Y, Jiang G et al. B-Myb regulates snail expression to promote epithelial-to-mesenchymal transition and invasion of breast cancer cell. Med Oncol 2015;32:412.
25. Joosse SA, Hannemann J, Spotter J et al. Changes in keratin expression during metastatic progression of breast cancer: impact on the detection of circulating tumor cells. Clin Cancer Res 2012;18:993-1003.
26. Willipinski-Stapelfeldt B, Riethdorf S, Assmann V et al. Changes in cytoskeletal protein composition indicative of an epithelial-mesenchymal transition in human micrometastatic and primary breast carcinoma cells. Clin Cancer Res 2005;11:8006-14.
27. Bos PD, Zhang XH, Nadal C et al. Genes that mediate breast cancer metastasis to the brain. Nature 2009;459:1005-9.
28. Wang Y, Klijn JG, Zhang Y et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365: 671-9.