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

Nan Shao1*, Kaitao Yuan2*, Yunjian Zhang1, Tuck Yun Cheang1, Jie Li1, Ying Lin1

1Breast Disease Center and 2Center 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. Gene ontology (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
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 [2,5]. 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 20 ER-negative/HER2-negative breast cancer samples and 32 ER+/HER2- breast cancer samples [7].

Identification of DEGs

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-
samples 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).

**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...

**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/).

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

<table>
<thead>
<tr>
<th>Expression</th>
<th>DEGs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Up-regulated (140)</strong></td>
<td>HACD1, CDH3, MICAL1, C1orf106, CAPN6, BBOX1, PDE9A, TTYH1, KCNG1, GLS, GSTA1, HSD17B2, MPZL2, CYP51A1, CRABP1, ACTG2, KLK6, MMP7, KRT17, FZD7, APBA2, SERPINB5, TMEM158, ZIC1, WWTR1, PKP1, ST8SIA1, NXN, FABP7, FERMT1, SMOX, ANP32E, ARHGEF4, UBE2E3, ELF5, RARRES1, SOX10, FAT1, PT7, CSRP2, SERPINE2, MFGF8, FOLH1B, SNORA11E, ART3, NUDTI1, CLDN8, ST3GAL6, PHGDH, UCHL1, PACX, GABRE, NFI8, SLPI, GPM6B, RRAGD, CBS, CLDN1, EGF, CAVE2, SOX11, KRT5, TMSB15B, CCNE1, GABBR2, MAP7D3, HLA-DOB, FSCN1, DZIP1, GPR161, CHI3L2, SMCO4, CALD1, GATA6, BCL11A, ITM2C, KRT6B, PSA1, FOXI1, TRIM2, CALB2, LAMP5, FABP5, NPSF750, ROP11B, L6YD, MCM5, PLIN2, CRYAB, MAP4K4, GB1, TSPYL5, PTX3, KCNK5, LPIN1, TRIM29, CDC20, MMP6, KRT16, CKDN2A, EN1, GABBP, KHDRBS5, C3XCL1, LDH8, VSNL1, PL53, TTTL4, ROR1, PRK3D, UGT8, MIA, P13, SLC6A14, GAL, ADGRL2, CHODL, JRKL, CHI3L1, SFRP1, RNF144A, PRNP, SLC16A16A1, DSC2, HRASLS, MYB, MYO10, GSTP1, PROM1, JG3, SCRGl, BTG5, CYP1B1, CHST3, SIX3, VGLL1, MTMM2, FAM171A1, PLAGL1, CBR1</td>
</tr>
<tr>
<td><strong>Down-regulated (215)</strong></td>
<td>NPDC1, CREB3L1, TJP3, TSPAN1, ERBB4, HGDFRP3, KRT19, UGCG, SSH3GRL, ARMT1, APBB2, CRIP1, HSPB1, TPST571G, EPS8L1, SIGIR, HSPB8, SYT1, GDF15, SIDT1, NPLR2, FAM134B, RTN1, SIAH2, RARA, CLGN, SLC2A10, DCRX, CEACAM6, MREG, COX6C, CYP2B7P, KCND3, GUSBP3, GATA3, MED11L, PLK2, GJAI, ELOVL5, SCUBE2, FIP, ENPPI1, SCNN1A, ACADSB, HSPA8, INPP4B, HGD, PDZK1, PNPLA4, GDF15, FAM199B, TIMPS2, CIRBP, SNA4, CEUDC1, DAC8, S3H3, SLC16A6, PLAG2, 16PFR1, PLAT, C16orf45, TSPAN13, BLVRA, GAMT, GPRCSA, GALNT6, GFRAX, REP5S, SLC27A2, DUSP4, ABLIM3, CPE, COQ4, CCND1, IGFBP2, MYO6, STC2, EVA1B, BTD, TTC59A, REEPS, FOXA1, NPY1R, GREB1, ER1, KITLG, SCCPDH, SLC1A1, SLC1A4, CHST15, LGALS8, GALNT7, TMOC6, CERS6, MUC1, RHOB, BCL2, IGBP4B, C8orf4, LRG1, SPDEF, ABHD2, CYB5A, FPT4A2, TM5C, LMF1, CBP1, XTLYT2, GLUL, NME3, TOX3, DNAJC12, AREG, IGF1R, MAST4, ABCG8, ELOVL2, MAGED2, CAMK2N1, ACOX2, EFHC1, IRI, SERPINA3, SCGB2A2, ABAT, C4B, 2, ECN1, NAT1, MLPH, RA12, TMBIM6, TBCD19, TFF1, CELSR1, NRP1, AR, TXB5, ZNF552, HOXB2, PLEKHF2, SCGB2A1, IL6ST, FGRF3, PRSS23, METRN, CLSTN2, LAMA3, SCGB1D2, TFF3, KDM4B, KCNK15, ADCY9, TCT3, GDP11, SERPINA5, AMIG02, G6PC3, VAV3, RABEP1, FBPI, FYX3D, INPP5J, KRT18, FUT8, SLC7A8, ITGB5, Euvw, WWPI1, CA12, BHLHE40, KCNE4, CFD, MISP, CHAD, TPBG, RET, MAPT, IGFBP5, SREBF2, SLC9A5R1, SLC24A1, POLD4, SLC43A6, CYP1B2, ZZ5F, GSTM3, CYP4B1, MCCCR2, REEP1, GALNT10, ADIRF, AKR7A3, ADC1, XBP1, CANT1, MSX2, SLC19A2, AGR2, PGR, CERS4, TNNT1, FAM174B, ANX9A, DHR82, ASPN, AFF3, KIAA0040, SLC7A2, EEF1A2, DNDL11, SYBU, MYB, SLC5A9A6, SLC44A4</td>
</tr>
</tbody>
</table>

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.

JBUON 2018; 23(4): 893
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.
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, 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 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

<table>
<thead>
<tr>
<th>Pathway ID</th>
<th>Name</th>
<th>Count</th>
<th>p value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04727</td>
<td>GABAergic synapse</td>
<td>8</td>
<td>0.00</td>
<td>GABRE, ADCY1, GLUL, ADCY9, GLS, ABAT, GABB2, GABRP</td>
</tr>
<tr>
<td>hsa04915</td>
<td>Estrogen signaling pathway</td>
<td>8</td>
<td>0.01</td>
<td>EGFR, ADCY1, HSPA2, ADCY9, ESR1, CREB5L1, GABB2, ITPR1</td>
</tr>
<tr>
<td>hsa04114</td>
<td>Oocyte meiosis</td>
<td>8</td>
<td>0.01</td>
<td>PGR, CCNE1, IGF1R, ADCY1, AR, ADCY9, CDC20, ITPR1</td>
</tr>
<tr>
<td>hsa05215</td>
<td>Prostate cancer</td>
<td>7</td>
<td>0.02</td>
<td>EGFR, CCNE1, IGF1R, AR, CCND1, BCL2, CREB5L1</td>
</tr>
<tr>
<td>hsa01100</td>
<td>Metabolic pathways</td>
<td>40</td>
<td>0.02</td>
<td>ACOX2, LDHB, ACADS, FUT8, HSD17B2, ENPP1, BTD, GALNT7, GALNT6, ST8SIA1, CERS6, UGDH, CERS4, G6PC5, MTMR2, MCCC2, CBR1, GALNT10, XYLTL2, INPP5j, ST3GAL6, UGT8, PLAZ2G16, NAT1, UGCG, FBP1, HGD, LPN1, COX6C, POLD4, GLUL, NME3, GLS, PHGDH, ABAT, INPP4B, GAMT, PSAT1, DCXR, CBS</td>
</tr>
<tr>
<td>hsa04913</td>
<td>Ovarian steroidogenesis</td>
<td>5</td>
<td>0.03</td>
<td>IGF1R, ADCY1, CYP1B1, HSD17B2, ADCY9</td>
</tr>
<tr>
<td>hsa00980</td>
<td>Metabolism of xenobiotics by cytochrome P450</td>
<td>6</td>
<td>0.05</td>
<td>GSTA1, GSTM3, CBR1, CYP1B1, AKR7A3, GSTP1</td>
</tr>
<tr>
<td>hsa05166</td>
<td>HTLV-I infection</td>
<td>12</td>
<td>0.05</td>
<td>MSX2, POLD4, ADCY1, CCND1, CDKN2A, ADCY9, XBP1, CDC20, MYB, MYBL2, HLA-DOB, FZD7</td>
</tr>
<tr>
<td>hsa05204</td>
<td>Chemical carcinogenesis</td>
<td>6</td>
<td>0.04</td>
<td>GSTA1, GSTM3, CBR1, CYP1B1, NAT1, GSTP1</td>
</tr>
<tr>
<td>hsa04923</td>
<td>Regulation of lipolysis in adipocytes</td>
<td>5</td>
<td>0.04</td>
<td>ADCY1, PLAZ2G16, ADCY9, NPY1R, IRS1</td>
</tr>
<tr>
<td>hsa05205</td>
<td>Proteoglycans in cancer</td>
<td>10</td>
<td>0.04</td>
<td>EGFR, IGF1R, CAV2, CCND1, ERBB4, ESR1, ITGB5, TIMP3, FZD7, ITPR1</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Count</th>
<th>p value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0045944—positive regulation of transcription from RNA polymerase II promoter</td>
<td>7</td>
<td>1.4E-05</td>
<td>PGR, EGFR, AR, CDKN2A, ESR1, MYB, MYBL2</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0060749—mammary gland alveolus development</td>
<td>3</td>
<td>5.3E-05</td>
<td>AR, CCND1, ESR1</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0045893—positive regulation of transcription, DNA-templated</td>
<td>5</td>
<td>2.4E-04</td>
<td>CCNE1, AR, CDKN2A, ESR1,MYB</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0008283—cell proliferation</td>
<td>4</td>
<td>1.5E-03</td>
<td>EGFR, AR, KRT16, BCL2</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0000082—G1/S transition of mitotic cell cycle</td>
<td>3</td>
<td>1.9E-03</td>
<td>CCNE1, CCND1, CDKN2A</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0001077—transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding</td>
<td>5</td>
<td>1.1E-05</td>
<td>PGR, AR, ESR1, MYB, MYBL2</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0008154—transcription factor binding</td>
<td>5</td>
<td>2.4E-05</td>
<td>AR, CCND1, CDKN2A, BCL2, ESR1</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0019899—enzyme binding</td>
<td>5</td>
<td>4.4E-05</td>
<td>PGR, EGFR, AR, CCND1, ESR1</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0000978—RNA polymerase II core promoter proximal region sequence-specific DNA binding</td>
<td>5</td>
<td>5.6E-05</td>
<td>PGR, AR, ESR1, MYB, MYBL2</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0005496—steroid binding</td>
<td>3</td>
<td>1.3E-04</td>
<td>PGR, AR, ESR1</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa05215:Prostate cancer</td>
<td>6</td>
<td>7.2E-08</td>
<td>EGFR, CCNE1, IGF1R, AR, CCND1, BCL2</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa05200:Pathways in cancer</td>
<td>7</td>
<td>5.6E-06</td>
<td>EGFR, CCNE1, IGF1R, AR, CCND1, CDKN2A, BCL2</td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>hsa04151:PI3K-Akt signaling pathway</td>
<td>6</td>
<td>6.2E-05</td>
<td>EGFR, CCNE1, IGF1R, CCND1, BCL2, MYB</td>
</tr>
</tbody>
</table>
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 [15]. The TATA box, initiator (Inr), TFIIB recognition element (BRE), and downstream core promoter element (DPE) all play critical roles in cancer 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 9 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 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, IGF1R, 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 (2016A050530154).

Conflict of interests

The authors declare no conflict of interests.

References


