Bioinformatics Identification of Prognostic Factors Associated With Breast Cancer

wei ying  
Sichuan University  https://orcid.org/0000-0001-8178-4705

Shipeng Zhang  
North Sichuan Medical College [Search North Sichuan Medical College]. North Sichuan Medical University

Li Xiao  
Sichuan University

Jing Zou  
Sichuan University

Yingqing Fu  
Sichuan University

Yi Ye  
Sichuan University

Linchuan Liao (linchuanliao@scu.edu.cn)  
Department of Forensic Toxicological Analysis, West China School of Basic Medical Sciences& Forensic Medicine, Sichuan University, 8th Floor, Fa Yi Building, No.16, Section 3, Renmin Nan Road, Chengdu 610016, Sichuan, China  https://orcid.org/0000-0003-3700-8471

Primary research

Keywords: Breast cancer, Differentially expressed genes, miRNAs, Transcription factors, Bioinformatic analysis

Posted Date: October 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-86580/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

**Background:** Breast cancer (BRCA) remains one of the most common forms of cancer and is the most prominent driver of cancer-related death among women. The mechanistic basis for BRCA, however, remains incompletely understood. In particular, the relationships between driver mutations and signaling pathways in BRCA are poorly characterized, making it difficult to identify reliable clinical biomarkers that can be employed in diagnostic, therapeutic, or prognostic contexts.

**Methods:** First, we downloaded publicly available BRCA datasets (GSE45827, GSE42568, and GSE61304) from the Gene Expression Omnibus (GEO) database. We then compared gene expression profiles between tumor and control tissues in these datasets using Venn diagrams and the GEO2R analytical tool. We further explore the functional relevance of BRCA-associated differentially expressed genes (DEGs) via functional and pathway enrichment analyses using the DAVID tool, and we then constructed a protein-protein interaction network incorporating DEGs of interest using the Search Tool for the Retrieval of Interacting Genes (STRING) database. Modules within this PPI network were then identified using Cytoscape, leading to the identification of key candidate genes. The prognostic relevance of these candidate genes was then established through Kaplan-Meier survival analyses and further Gene Expression Profiling Interactive Analysis (GEPIA) validation. Then, key gene-target miRNA regulatory network and transcription factor-key gene regulatory relationships were established using the online miRWalk2.0, TargetScan7.2, miRDB and TRRUST tools. Moreover, four representative key molecules (AURKA, RRM2, BIRC5, and E2F1) were optionally chosen for verification by using quantitative real-time polymerase chain reaction (RT-PCR) and western blot.

**Results:** We identified 85 BRCA-related DEGs across these three datasets. The 31 upregulated DEGs were found to be enriched for pathways and functions including mitotic nuclear division, cell division, G2/M transition of mitotic cell cycle, collagen catabolic process, endodermal cell differentiation, oocyte meiosis, ECM-receptor interactions, and p53 signaling pathway. The 54 downregulated DEGs were, in contrast, enriched in pathways and functions such as lipid metabolic processes, lipid transport, regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ions, positive regulation of cell proliferation, positive regulation of cell-matrix adhesion, tyrosine metabolism, cytochrome P450 drug metabolism, protein digestion and absorption, and PPAR signaling. We were further able to select 16 upregulated candidate genes of interest from our PPI network, and in subsequent Kaplan-Meier analyses we were able to determine that elevated expression of 14 of these genes was associated with a poorer BRCA patient prognosis. We then employed GEPIA to validate these 14 gene candidates, confirming them to all be expressed at elevated levels in BRCA relative to normal tissue controls. In addition, a regulatory network consisting of 9 genes, 10 miRNAs and 3 TFs was constructed, enabling the identification of potential biomarkers of BRCA, including AURKA, RRM2, BIRC5, and E2F1. RT-PCR results suggested that significantly elevated AURKA, RRM2 and BIRC5 mRNAs expressed in the breast cancer cells than in the normal cells. Western blot results shown that E2F1 protein was highly expressed in breast cancer cells compared to normal cells. In conclusion, these candidate molecules may offer insight regarding the underlying pathogenesis of BRCA and highlight a number of potential therapeutic avenues for the treatment of breast cancer patients.

1. Background

Breast cancer (BRCA) remains the leading cancer-related cause of death among women and accounts for almost 25% of all female cancer diagnoses, with 2.1 million new cases in 2018 alone[1]. The incidence of breast cancer has increased drastically in the past few decades and is projected over the next few years[2]. Recent research has led to the development of many novel screening and therapeutic strategies that have significantly improved the survival outcomes for many breast cancer patients. Currently, the five-year survival rate of women diagnosed with stage I BRCA is nearly 100%, but this percentage falls dramatically to 26% when women are diagnosed with stage IV disease (~ 5% of all cases)[3]. Advanced BRCA is thus associated with a very poor prognosis. Therefore, it is particularly important to identify biomarkers capable of enabling the early diagnosis of breast cancer. Additionally, significant heterogeneity between BRCA patients also means that individual responsiveness to a given therapeutic intervention can be highly variable[4, 5]. A number of different genetic and phenotypic biomarkers have been leveraged to date to guide BRCA treatment and diagnosis, including wound-response gene signatures[6], hypoxia-response signatures[7], 21-gene recurrence scores[8], and a 70-gene signature[9]. Despite the promise of these biomarker panels, they are relatively non-specific and can necessitate costly testing that may impose a significant financial burden on patients. This, coupled with the poorly characterized mechanistic basis of BRCA heterogeneity, underscores the importance of identifying more reliable and economical therapeutic, diagnostic, and prognostic biomarkers of BRCA.

Advances in high-throughput microarray and sequencing approaches have allowed for the development of more efficient screening approaches aimed at clarifying the genetic and epigenetic changes associated with tumor development and progression. Such microarray datasets can thus be readily leveraged to identify diagnostic and prognostic biomarkers of specific cancer types[10, 11]. Microarrays can allow for the rapid detection of all genes that are differentially expressed between tumors and normal tissues[12]. In recent years, some progress has been made in screening for potential biomarkers related to the occurrence, development and prognosis of breast cancer at the
whole-genome level. For instance, DNAPTP3 was first found to play an oncogenic role in breast cancer initiation and development by Hou et al[13]. A recent study also conducted WGCNA and found that immune response- and ubiquitin-mediated proteolysis can be used as prognostic and predictive biomarkers in BRCA[14]. Furthermore, studies have identified numerous candidate biomarkers associated with the prognosis and metastasis of breast cancer[15, 16]. Simultaneously, several studies have demonstrated that some microRNAs (miRNAs) and transcription factors (TFs) play critical regulatory roles in BRCA[17–20]. While many BRCA-related differentially expressed genes, miRNAs, and TFs have been studied in recent years, the detailed pathogenesis of BRCA remains unclear due to limited understanding of mRNA-miRNA-TF relationships in this oncogenic context.

In the present study, we downloaded four publicly available microarray datasets (GSE45827, GSE42568, and GSE61304 from the Gene Expression Omnibus (GEO) database. Using Venn diagrams, the GEO2R analysis tool, and the DAVID platform we were able to identify BRCA-related DEGs and to conduct gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of these genes. We further generated a DEG protein-protein interaction (PPI) network and leveraged this network to identify core BRCA genes that were found to be significantly associated with BRCA patient prognosis in survival analyses conducted using the Kaplan-Meier plotter database. Subsequently, a prognosis-related gene-miRNA targets interaction network analysis was performed, and TF-mediated regulation of prognosis-related genes was analyzed. Furthermore, experimental verification was conducted. This work will enable the identification of core genes, miRNAs, and TFs associated with BRCA, thereby facilitating the construction of a prognosis-related gene-target miRNA-TF regulatory network that can clarify the mechanisms of breast cancer pathogenesis and highlight novel candidate molecular targets.

2. Materials And Methods

2.1 Microarray datasets and DEG identification

We downloaded the gene expression data from the GSE45827, GSE42568, and GSE61304 microarray datasets. All three datasets compared gene expression profiles in BRCA tumors and normal breast tissue samples from human patients, and all three microarrays were conducted using GPL570 Platforms([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). In total, these datasets incorporated 142 BRCA tissues and 11 control tissues, 104 BRCA tissues and 17 control tissues, and 58 BRCA tissues and 4 control tissues, respectively. Subsequently, we used the GEO2R online tool[21] to identify DEGs by comparing BRCA and normal control tissue gene expression levels, using |log FC| >2 and adjusted P-value < 0.05 as criteria for differential expression. We then loaded the resultant DEG lists into an online Venn diagram software in order to identify those DEGs that were common to all three of these DEG datasets.

2.2 Functional enrichment analyses

In order to explore the biological significance of identified DEGs, we conducted both GO and KEGG enrichment analyses on DEG gene lists. GO analyses offer direct insight into the annotated characteristics and functions of genes contained within genomic or transcriptomic datasets, offering a high-level view of processes or properties that may be impacted in a given biological context[22]. KEGG analyses, in contrast, provide insight into the signaling and biological pathways that are impacted in a given context, as in response to a particular treatment or in cells impacted by a particular disease state[23, 24]. We utilized the DAVID tool (https://david.ncifcrf.gov/) to conduct GO and KEGG analyses[25], with P < 0.05 as the significance threshold.

2.3 PPI network construction and analysis

We constructed our PPI network using the online STRING tool (https://string-db.org)[26] in an effort to further understand the mechanistic basis for BRCA development and/or progression. We employed the Cytoscape tool[27] to visualize this network and to examine the relationships between the DEGs found therein, with network edges representing confidence and with medium confidence cut-off (0.400) being used for network construction. Modules within this network were identified using the Molecular Complex Detection (MCODE) application within Cytoscape using the following criteria: degree cutoff = 2, max. Depth = 100, k-core = 2, and node score cutoff = 0.2.

2.4 DEG validation and prognostic analysis

We employed the Kaplan–Meier Plotter (Online: http://kmplot.com/) database in order to explore the prognostic relevance of key hub genes identified in our PPI network module analysis, as this database compiles cancer patient survival data from sources including GEO, TCGA, and EGA[28].

For this analysis we conducted overall survival analyses for genes of interest, with P < 0.05 as the significance threshold. We additionally utilized GEPIA (http://gepia.cancer-pku.cn/) to explore RNA-seq datasets derived from thousands of TCGA and GTEx samples[29] in an effort to validate the relevance of specific DEGs to BRCA progression.
2.5 Prediction of target miRNAs associated with prognosis-related genes and construction of gene-miRNA target interaction network

miRNA targets of core genes were predicted using miRWalk 2.0 (http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/) [30], TargetScan7.2 (http://www.targetscan.org/vert_72/) [31] and miRDB (http://www.mirdb.org/) [32]. To ensure the accuracy of these results, the miRNA targets were selected based on the intersection of results from these three databases. A regulatory network incorporating these core genes and putative targeting miRNAs was then constructed using the Cytoscape software.

2.6 Construction of a prognosis-related gene-target miRNA-transcription factor (TF) regulatory network

miRNAs and TFs function as trans-regulatory factors that can control broad gene regulatory networks. To explore TFs that can regulate prognosis-related genes in BRCA, we screened the TRRUST v2 database (https://www.grnpedia.org/trrust/), which is a curated database of human and mouse transcriptional regulatory network that provides insight into the regulation of TF-target relationships [33]. For these analyses, the species was set to 'human', with \( p < 0.05 \) as a threshold for significance. A mRNA-miRNA-TF regulatory network was then constructed to assess the functional roles of identified genes.

2.7 Quantitative real-time PCR and Western blot analysis

Human breast cancer cell line MDA-MB-231 (CL-0150) and MCF-10A (CL-0525) cell line were purchased from the Procell Life Science & Technology Co., Ltd. (Wuhan, China). Total RNA was isolated from cultured cells using the Trizol reagent (Aidlab, China), subsequently, 5 µg RNA was converted to cDNA by using the HiScript Reverse Transcriptase (VAZYME, China) following the manufacturer's protocol. Real-time PCR was performed using SYBR Green Master Mix (VAZYME, China) and analyzed with QuantStudio 6 Real-time PCR System (Applied Biosystems, USA). GADPH was applied as an endogenous control. Relative gene expression level was calculated by the \( 2^{-\Delta\Delta CT} \) method between case and controls. Primer sequences used in the study are shown in Supplementary Table 1.

The MDA-MB-231 and MCF-10A cells were lysed with PMSF lysis (Aladdin, China) buffer. The concentrations of protein were determined by the BCA protein assay (Beyotime, China) according to the manufacturer's protocol. Equal amounts of protein were loaded and analyzed by 10% SDS-polyacrylamide gel electrophoresis (Bioharp, China) and transferred to a PVDF membrane (Millipore, USA). The size-cut membrane was blocked in TBST with 5% non-fat milk for 2 h at room temperature and then incubated overnight at 4 °C with the primary antibody. After that, the membrane was incubated with HRP conjugated secondary antibody at room temperature for 2 h. Finally, ECL reagent was performed to determine the antigen-antibody complexes. An rabbit anti-human GAPDH polyclonal antibody (Goodhere, China), mouse anti-human E2F1 monoclonal antibody (Proteintech Group, Inc, China) and HRP conjugated goat anti-rabbit/mouse antibody (Boster, China) were used.

3. Results

3.1 BRCA-related DEG identification

By pooling data from three GEO microarray datasets, we were collectively able to analyze 304 BRCA tumor tissue samples and 32 normal breast tissue control samples. With the aid of the GEO2R tool, we were able to extract 2384, 1176, and 260 BRCA-related DEGs from the GSE45827, GSE42568, and GSE61304 datasets, respectively. Using a Venn diagram approach, we then identified 85 DEGs that were common to all three datasets. Of these DEGs, 54 were downregulated (logFC<0) and 31 were upregulated (logFC>0) in BRCA tumor tissues (Supplementary Table 2 & Figure 1).

3.2 Functional enrichment analyses of identified DEGs

In order to understand the functional roles of these 85 BRCA-associated DEGs, we next conducted GO and KEGG analyses thereof. GO analyses examined the biological processes (BPs), cellular components (CCs), and molecular functions (MFs) with which these DEGs were affiliated. The GO BP analysis revealed the upregulated DEGs to be primarily enriched for processes including mitotic nuclear division, cell division, G2/M transition of mitotic cell cycle, collagen catabolic process, and endodermal cell differentiation, whereas downregulated DEGs were enriched lipid metabolic processes, lipid transport, regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ions, positive regulation of cell proliferation, and positive regulation of cell-matrix adhesion. The GO CC analysis revealed upregulated DEGs to be primarily enriched in the proteinaceous extracellular matrix, midbody, condensed chromosome kinetochore, centriole, and microtubule cytoskeleton, whereas downregulated DEGs were enriched in the extracellular space, extracellular region, lipid particles, proteinaceous extracellular matrix, and the apical plasma membrane. With respect to the GO MF analysis, upregulated DEGs were
enriched for functions including collagen binding, heparin binding, ATP binding, and metalloendopeptidase activity, with downregulated DEGs being enriched lipid binding, heparin binding, lipoprotein particle binding, transporter activity, zinc-dependent alcohol dehydrogenase activity (Table 1).

We further conducted a KEGG pathway analysis of these DEGs. This approach revealed the upregulated DEGs to primarily be associated with the p53 signaling pathway, ECM-receptor interactions, and oocyte meiosis, whereas downregulated DEGs were primarily associated with the PPAR signaling pathway, tyrosine metabolism, cytochrome P450 drug metabolism, and protein digestion and absorption. KEGG pathway enrichment analysis results for upregulated and downregulated DEGs are shown in Figure 2.

**Table 1** GO enrichment analysis of BRCA-related DEGs (The terms with the top 5 P-values are shown)
<table>
<thead>
<tr>
<th>Expression</th>
<th>Category</th>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
<th>Fold enrichment</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0007067~mitotic nuclear division</td>
<td>11</td>
<td>9.55E-12</td>
<td>24.02601457</td>
<td>1.31E-08</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0051301~cell division</td>
<td>9</td>
<td>1.29E-07</td>
<td>13.92884793</td>
<td>1.77E-04</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0000086~G2/M transition of mitotic cell cycle</td>
<td>6</td>
<td>4.06E-06</td>
<td>23.72309866</td>
<td>0.00556875</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0030574~collagen catabolic process</td>
<td>5</td>
<td>4.88E-06</td>
<td>42.31854839</td>
<td>0.006689601</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0035987~endodermal cell differentiation</td>
<td>4</td>
<td>1.46E-05</td>
<td>80.24850657</td>
<td>0.020038146</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005578~proteinaceous extracellular matrix</td>
<td>8</td>
<td>2.10E-07</td>
<td>17.5483871</td>
<td>2.23E-04</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0030496~midbody</td>
<td>6</td>
<td>2.03E-06</td>
<td>27.34283571</td>
<td>0.002166272</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0000777~condensed chromosome kinetochore</td>
<td>4</td>
<td>3.89E-04</td>
<td>27.02855024</td>
<td>0.413450196</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005814~centriole</td>
<td>4</td>
<td>8.34E-04</td>
<td>20.80959178</td>
<td>0.885431205</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0015630~microtubule cytoskeleton</td>
<td>4</td>
<td>0.001454401</td>
<td>17.16411585</td>
<td>1.539162614</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0005518~collagen binding</td>
<td>4</td>
<td>1.62E-04</td>
<td>36.30322581</td>
<td>0.17777753</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0008289~lipid binding</td>
<td>4</td>
<td>0.005831972</td>
<td>10.64711447</td>
<td>6.708154954</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0004024~alcohol dehydrogenase activity</td>
<td>2</td>
<td>0.014486531</td>
<td>133.9761905</td>
<td>15.90609305</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:00006629~lipid metabolic process</td>
<td>6</td>
<td>5.42E-05</td>
<td>14.26072187</td>
<td>0.080640059</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0006869~lipid transport</td>
<td>4</td>
<td>0.00103261</td>
<td>19.63976608</td>
<td>1.52497516</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0010881~regulation of cardiac muscle contraction by regulation of</td>
<td>3</td>
<td>0.001115403</td>
<td>58.91929825</td>
<td>1.646301724</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
<td>the release of sequestered calcium ion</td>
<td>7</td>
<td>0.001276576</td>
<td>5.605340963</td>
<td>1.882087273</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GO:0008284~positive regulation of cell proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0001954~positive regulation of cell-matrix adhesion</td>
<td>3</td>
<td>0.001499269</td>
<td>50.88484848</td>
<td>2.207004361</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005615~extracellular space</td>
<td>14</td>
<td>4.96E-05</td>
<td>3.713932195</td>
<td>0.055590194</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005576~extracellular region</td>
<td>14</td>
<td>3.01E-04</td>
<td>3.107246377</td>
<td>0.33641381</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005811~lipid particle</td>
<td>4</td>
<td>7.87E-04</td>
<td>21.65656566</td>
<td>0.878449732</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005578~proteinaceous extracellular matrix</td>
<td>6</td>
<td>8.18E-04</td>
<td>8</td>
<td>0.912134466</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0016324~apical plasma membrane</td>
<td>6</td>
<td>0.001180873</td>
<td>7.367697595</td>
<td>1.314786411</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0008289~lipid binding</td>
<td>4</td>
<td>0.005831972</td>
<td>10.64711447</td>
<td>6.708154954</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0004222~metalloendopeptidase activity</td>
<td>3</td>
<td>0.017098388</td>
<td>14.45703683</td>
<td>17.26865537</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0001954~positive regulation of cell-matrix adhesion</td>
<td>3</td>
<td>0.001499269</td>
<td>50.88484848</td>
<td>2.207004361</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0008289~lipid binding</td>
<td>4</td>
<td>0.005831972</td>
<td>10.64711447</td>
<td>6.708154954</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0008201~heparin binding</td>
<td>4</td>
<td>0.00684227</td>
<td>10.04821429</td>
<td>7.827449644</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:00071813~lipoprotein particle binding</td>
<td>2</td>
<td>0.012086409</td>
<td>160.7714286</td>
<td>13.44231459</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0005215~transporter activity</td>
<td>4</td>
<td>0.012887447</td>
<td>7.958981612</td>
<td>14.27185096</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0004024~alcohol dehydrogenase activity</td>
<td>2</td>
<td>0.014486531</td>
<td>133.9761905</td>
<td>15.90609305</td>
</tr>
</tbody>
</table>
Abbreviations: BP, biological process; CC, cellular component; Count, the number of enriched genes in each term; FDR, false discovery rate; MF, molecular function; BRCA, breast cancer.

3.3 PPI network construction and analysis

We next generated a PPI network in an effort to better understand the relationships between the 85 BRCA-related DEGs identified above. To that end, we leveraged the STRING database to construct a network that successfully incorporated 63 of these 85 DEGs (28 and 35 up- and down-regulated, respectively). The final network contained 63 nodes and 215 edges (Figure 3A). Using the MCODE plugin, we then identified 16 nodes corresponding to hub genes within this network using filtering degree ≥ 14 as a criterion for hub identification (Figure 3B).

3.4 Assessment of the prognostic relevance of core genes

Next, we tested the prognostic relevance of these 16 core genes using the Kaplan-Meier Plotter database. This analysis revealed that the upregulation of 14 of these genes was significantly associated with poorer patient survival (P < 0.05, Supplementary Table 3 & Figure 4). We then further validated the expression of these 14 prognostic hub genes using the GEPIA tool, confirming that all 14 of these genes were expressed at higher levels in BRCA tumor tissue samples relative to normal breast tissue (P < 0.01, Figure 5). Importantly, patients with higher expression of these genes had poorer outcomes than did patients with lower expression thereof. The gene symbols, full names, and functions of these prognosis-related genes are shown in Table 2.

**Tab 2** Functions of key prognosis-related genes
<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Full name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP2A</td>
<td>Topoisomerase (DNA) II α</td>
<td>TOP2A can be therapeutic targets for anticancer, and mutations in this gene is associated with the development of drug resistance[34].</td>
</tr>
<tr>
<td>ASPM</td>
<td>Abnormal spindle microcephaly associated</td>
<td>ASPM plays a role in the regulation of mitotic spindle, with a priority role in regulating neurogenesis. Overexpression of ASPM links with tumor progression and poor prognosis in various tumors, including breast cancer[35]. Furthermore, it has been reported that ASPM promotes tumor development by regulating Wnt-β-catenin signaling[36, 37].</td>
</tr>
<tr>
<td>NEK2</td>
<td>Never in Mitosis (NIMA) Related Kinase 2</td>
<td>NEK2 plays a critical role in multiple aspects of mitotic processes, such as centrosome duplication and separation, microtubule stabilization, kinetochore attachment, and spindle assembly checkpoint[38]. NEK2 is frequently overexpressed in a variety of human malignancies, has been implicated in tumorigenesis, tumor progression and drug resistance[39, 40].</td>
</tr>
<tr>
<td>RRM2</td>
<td>ribonucleotide reductase M2</td>
<td>RRM2, a rate-limiting enzyme for DNA synthesis and repair, overexpression exhibited enhanced cellular invasiveness, serving as a potential prognostic marker and a possible therapeutic target for cancer treatment.</td>
</tr>
<tr>
<td>UBE2T</td>
<td>Ubiquitin-conjugating enzyme E2T</td>
<td>UBE2T plays a key role in cellular processes, for example, signal transduction, cell cycle control and tumorigenesis[41].</td>
</tr>
<tr>
<td>CEP55</td>
<td>Centrosomal protein 55</td>
<td>CEP55 localizes to the centrosome throughout the cell cycle, and to regulate both the mitotic spindle and microtubule organization[42].</td>
</tr>
<tr>
<td>ANLN</td>
<td>Anillin, actin binding protein</td>
<td>ANLN plays an important role in cytokinesis, highly express in numerous cancer types[43, 44].</td>
</tr>
<tr>
<td>DLGAP5</td>
<td>discs large associated protein 5</td>
<td>DLGAP5 is a regulator of cell cycle involved in carcinogenesis. Elevated DLGAP5 expression correlates with poor prognosis in breast cancer patients[45].</td>
</tr>
<tr>
<td>CCNB2</td>
<td>Cyclin B2</td>
<td>Cyclin B2 may regulate the cell cycle with transforming growth factor beta-mediated. Aberrant expression of CCNB2 is correlated with tumor invasion, metastasis, poor prognosis of various human cancers[46].</td>
</tr>
<tr>
<td>BIRC5</td>
<td>Survivin, Baculoviral IAP repeat containing 5</td>
<td>BIRC5 may inhibit apoptosis and is usually overexpressed in most tumors[47].</td>
</tr>
<tr>
<td>TPX2</td>
<td>Targeting protein for Xenopus kinesin-like protein 2</td>
<td>TPX2 is a microtubule-associated protein required for mitosis and spindle assembly.</td>
</tr>
<tr>
<td>AURKA</td>
<td>Aurora kinase A</td>
<td>AURKA, a serine/threonine kinase, has important functions in regulating cell cycle and mitosis.</td>
</tr>
<tr>
<td>KIF2C</td>
<td>Kinesin family member 2C</td>
<td>KIF2C, a crucial mitotic regulator, is highly expressed in many cancer types[48].</td>
</tr>
<tr>
<td>UBE2C</td>
<td>Ubiquitin-conjugating enzyme E2C</td>
<td>UBE2C is overexpressed in many human tumor types[49].</td>
</tr>
</tbody>
</table>

### 3.5 Integrated mRNA-miRNA interaction network analysis

Subsequently, Candidate target miRNAs associated with fourteen prognosis-related genes were predicted using miRWalk, TargetScan, and miRDB. Overlapping miRNAs predicted by all three databases were then selected, with a specific focus on miRNAs capable of binding to 3'UTR regulatory regions. An interaction network incorporating these miRNAs and core genes was then constructed using Cytoscape (Figure 6). The network contained 217 nodes and 214 edges. By comparing the targets of core genes, RRM2, AURKA, NEK2, and BIRC5 were found to be potential targets of 53 miRNAs, 41 miRNAs, 30 miRNAs and 28 miRNAs, respectively. Those miRNAs that simultaneously regulated a high number of gene cross-links (≥2) were selected (Supplementary Table 4).

### 3.6 Construction of a mRNA-target miRNA-TF regulatory network
The selected miRNAs and corresponding genes may play a crucial role in the pathogenesis of BRCA. Hence, RRM2, CEP55, ANLN, UBE2C, NEK2, TOP2A, AURKA, KIF2C, and BIRC5 were considered to be key genes of interest in this context. In order to further understand the function of these genes, a miRNA-target gene-TF regulatory network for these key genes was constructed containing 9 genes, 10 miRNAs, and 3 TFs (Figure 7).

3.7 Verification of candidate molecules expression by RT-PCR and Western Blot

For the purpose of enhancing the credibility of results, the mRNA levels of the three hub genes (AURKA, RRM2 and BIRC5) were measured by RT-PCR and E2F1 protein expression was determined with western blot in MDA-MB-231 and MCF-10A cells. As illustrated in Figure 8A, all three genes were significantly upregulated in breast cancer cells (P < 0.01), as predicted by the bioinformatics analysis. E2F1 protein was highly expressed in breast cancer cells as shown in Figure 8B.

4. Discussion

Despite many recent therapeutic innovations, BRCA remains the leading cancer-related cause of death among women. BRCA remains a highlight heterogeneous disease, with four major BRCA subtypes having been defined based upon tumor expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2, and the proliferative marker Ki67 (MKi67) [50–52]. This classification scheme is typically used to predict BRCA patient prognosis and to define treatment strategies together with information pertaining to tumor histological grade, type, and TNM stage. Given the high degree of BRCA tumor heterogeneity, however, it is difficult to predict tumor response to treatment or associated patient outcomes. Additionally, the cooperative effects of TFs and miRNAs in gene regulation remain largely unclear. Recent molecular studies have highlighted a number of genes and signaling pathways that govern BRCA development and progression[53–56], but further work is needed to fully elucidate the mechanistic basis for this complex disease. This study represents the first attempt to construct a trans-regulatory mRNA-miRNA-TF network in order to identify genes, miRNAs, and TFs related to BRCA, thereby clarifying the pathogenesis of BRCA at the molecular level and improving the detection, treatment, and prognostic assessment of this disease.

In the present study, we conducted an integrated bioinformatics analysis of three microarray datasets, leading us to identify 85 BRCA-related DEGs (54 and 31 down- and up-regulated, respectively). We found that our upregulated DEGs were mainly enriched in the cell cycle, cell differentiation, oocyte meiosis, and the p53 signaling pathway, which plays a significant role in the occurrence and progression of tumors. Mitotic catastrophe/error can also contribute to tumorigenesis [57, 58]. As such, mitosis has become an attractive anti-tumor target, with paclitaxel being one of the most successful drugs targeting this pathway that is commonly used to treat BRCA[59, 60]. Previous studies have shown that the p53 signaling pathway controls cell cycle regulation, mitosis, reproduction, and the inhibition of neoplastic transformation and tumor progression[61, 62]. The down-regulated genes were enriched in the lipid metabolic process, lipid transport, regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ions, positive regulation of cell proliferation, positive regulation of cell-matrix adhesion, tyrosine metabolism, cytochrome P450 drug metabolism, protein digestion and absorption, and PPAR signaling pathway. Previous studies have reported that PPAR, as a ligand-activated transcription factor, regulates lipid metabolism and inflammation and contributes to the survival of breast cancer cells[63, 64]. Six DEGs were enriched in these two signaling pathways, including CDK1, CCNB2, RRM2, and AURKA, which have been proven to be involved in the mitotic process[65]. All in all, our results are consistent with the above theories. BRCA is mostly correlated with cell cycle and cell differentiation, mainly for mitotic cell cycle and enriching p53 signaling pathway.

We then constructed PPI network from identified DEGs and identified 16 DEGs as key genes (degree ≥ 14), of which 14 were found to be significantly associated with poorer BRCA patient survival in the Kaplan-Meier Plotter database. We further confirmed that all 14 of these were significantly upregulated in BRCA tissue samples relative to normal tissue controls in a GEPIA analysis (P < 0.05). Mitotic error is not the only cause of BRCA, but it is an important pathogenic factor and can explain one possible mechanism whereby BRCA can develop. Most of the core genes screened out in this study were related to mitosis, including ASPM[66], NEK2[67], RRM2, CEP55[42], CCNB2, BIRC5[47], KIF2C, AURKA, GLGAP5, and TPX2[68]. Of these genes, several were previously correlated with the occurrence, development, and prognosis of BRCA. Accumulating evidence indicates that overexpression of AURKA causes drug resistance, promotes progression, and predicts poor prognosis for breast cancer. Suppression AURKA can accelerate PI3K-pathway inhibitor efficacy in breast cancer resulting in apoptosis and tumor regression[69]. Zheng et al. have also shown that AURKA translocates to the nucleus and causes carcinogenic activities in malignant tumor cells by enhancing breast cancer stem cell phenotypes[70]. Furthermore, UBE2C has been shown to regulate the activity of AURKA by controlling the activity of APC/C[49]. Overexpression of UBE2C in breast microcalcification lesions suggests that UBE2C is also related to the occurrence and development of breast cancer[71]. One study found that AURKA, TPX2, and DLGAP5 were highly upregulated in non-small cell lung cancer and correlated with one another. TPX2 and DLGAP5 are also phosphorylated by AURKA during the cell cycle[68]. Chen et al. found TPX2 silencing negatively regulates PI3K / Akt, activates the p53 signaling pathway, inhibits breast cancer cell proliferation, and
promotes cell apoptosis, indicating that TPX2 may be a potential target for the treatment of breast cancer[72]. Another study has also found that TPX2 may be a novel prognostic marker of breast cancer[73].

In one prior analysis, RRM2 was found to be highly expressed in diverse cancers, and inhibiting RRM2 overexpression in retroperitoneal liposarcoma can impede tumor progression via downregulating the Akt/mTOR/4EBP1 pathway[74]. Meanwhile, Koppenhafer et al. found that CDK2-mediated RRM2 downregulation promotes DNA damage and cell apoptosis in Ewing sarcoma via inhibiting the ATR-CHK1 pathway[75]. Increasing evidence suggests that miRNAs and lncRNAs related to RRM2 can modulate tumor progression by regulating the expression of RRM2[76]. For example, in BRCA tissues, lncRNA DSCAM-AS1 directly targets miR-204-5p, stimulates proliferation and invasion, and impedes BRCA cell apoptosis by inhibiting miR-204-5p and facilitating RRM2 expression[77]. Consistent with these reports, we found that RRM2 upregulation was associated with a poorer BRCA patient prognosis. According to a previous study, CEP55 is a cell fate determinant in the context of perturbed mitosis in breast cancer. Loss of CEP55 renders breast cancer cells sensitive to anti-mitotic drugs by prematurely activating CDK1/ cyclin B and CDK1 caspase-dependent mitotic cell death. Furthermore, blocking MEK1/2-PLK1 signaling may reduce the outgrowth of MYC-CEP55-dependent basal-like, triple-negative breast cancers[78]. Our results were consistent with these findings, indicating that patients with higher levels of these genes expression may have poorer prognosis.

miRNAs are endogenous non-coding RNA molecules, that can modulate gene expression at the post-transcriptional level by reducing or inhibiting the translation of target genes[79]. This study screened out mRNA-miRNA target pairs using online prediction tools. RRM2, AURKA, NEK2, and BIRC5 were the top four potential miRNA targets in this study (≥ 28). Recent studies have found that some miRNAs are involved in the development of cancer, including BRCA[17–19]. TFs also serve as trans-regulatory factors, much as do miRNAs. To gain a further understanding of key gene functions in this oncogenic context, an mRNA-target miRNA-TF regulatory network was therefore constructed. E2F1 was identified as a co-regulator of KIF2C, TOP2A, RRM2, AURKA, and BIRC5. MED1 was found to regulate AURKA and UBE2C, while SP1 was predicted to regulate KIF2C and BIRC5. Several studies have shown that E2F1 plays a critical role in the regulation of both cell proliferation and apoptosis[80, 81]. AURKA overexpression has also been observed in many cancers. He et al. [82] found that the transcriptional activity of E2F1 was increased by AURKA overexpression. We also found that AURKA was associated with miRNA cluster expression by regulating E2F1. AURKA may thus be a putative therapeutic target in BRCA. In addition, a previous study has reported that silencing RRM2 to inhibit the up-regulation of RRM2 can enhance DNA damage, which could in turn improve the efficacy of topoisomerase I inhibitors[83]. ASA has also been found to promote TRAIL-induced apoptosis by down-regulating BIRC5 gene expression, which was mediated by inhibiting E2F-1 binding to the BIRC5 promoter region[84]. The experiments in vitro further conduced to validate. These results suggest that AURKA, RRM2, BIRC5, and related regulatory factors including E2F1 and miRNAs may be key molecules involved in controlling the onset and progression of BRCA.

5. Conclusions

In conclusion, we were able to leverage three microarray datasets in an integrated bioinformatics analysis in order to identify 85 BRCA-related DEGs, including 16 key candidate genes that were detected through a PPI network analysis. Additionally, a regulatory network consisting of 9 genes, 10 miRNAs and 3 TFs was constructed, revealing the co-regulatory associations among these key genes, corresponding targeted miRNAs, and TFs in BRCA. These key molecules may be regarded as potential biomarkers of BRCA, with AURKA, RRM2, BIRC5, and E2F1 being particularly relevant targets of interest. Nevertheless, further prospective studies will be required in order to validate our findings. Even so, the results of this analysis may offer valuable insight into the molecular basis for BRCA carcinogenesis and progression, while also identifying reliable candidate biomarkers that can guide the precision diagnosis and individualized treatment of BRCA.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data used to support the findings of this study are included within the article. The data and materials in this study are available from the corresponding author on reasonable request.
Conflict of interest

All the authors declared that there are no conflicts of interest relevant to this study.

Funding

Not applicable.

Authors' contributions

Ying Wei contributed to conceive and design the experiments, authored or reviewed drafts of the paper. Shipeng Zhang performed the experiments, analyzed the data. Li Xiao conducted the literature search and contributed reagents/materials/analysis tools. Ying Wei gave the final approval of the version to be submitted. Runfang Xie prepared figures and/or tables. Jing Zhou wrote drafts of the article. Yi Ye and Linchuan Liao revised and reviewed the article. All authors read and approved the final draft.

Acknowledgments

The authors would like to express their sincere gratitude to all those who have given their assistance in the preparation of this paper.

References


Identification of shared BRCA-associated DEGs. A total of 85 shared DEGs in the GSE45827, GSE42568, and GSE61304 datasets were identified using a Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/). Different colors correspond to different microarray datasets. A 31 shared upregulated DEGs (logFC>0) and B 54 shared downregulated DEGs (logFC < 0) were identified via this approach.

Pathway enrichment analysis of BRCA-related DEGs. A upregulated DEGs B downregulated DEGs. Bubble color represent p-value (-log10 p-value) while the size represent the number of DEGs in the corresponding pathway.
Figure 3

PPI network analysis of BRCA-related DEGs. A We used the STRING database and Cytoscape tools to construct a PPI network from 85 BRCA-related DEGs. B Cytoscape was used to conduct a module analysis, with up- and down-regulated genes being shown in red and blue, respectively, and with lines corresponding to gene-gene interactions.
Figure 4

Prognostic analysis of 16 BRCA-related hub genes. The Kaplan-Meier Plotter tool was used to assess the prognostic relevance of the 16 identified core genes, with Logrank $P < 0.05$ as the significance threshold.
Figure 5

A GEPIA approach was used to confirm the differential expression of the 14 identified prognostic core genes in BRCA tumors relative to control tissues. The tumor and control tissues are indicated in red and gray, respectively, with dots corresponding to expression in individual samples. *P < 0.01.
Figure 6

An interaction network of prognosis-related genes and targeted miRNAs. The prognosis-related genes are blue, while corresponding targeted miRNAs are red, and miRNAs associated with more than two genes are green.
Figure 7


A

<table>
<thead>
<tr>
<th></th>
<th>MCF-10A</th>
<th>MDA-MB-231</th>
</tr>
</thead>
<tbody>
<tr>
<td>AURKA</td>
<td><img src="chart1.png" alt="Bar chart" /></td>
<td><img src="chart2.png" alt="Bar chart" /></td>
</tr>
<tr>
<td>RRM2</td>
<td><img src="chart3.png" alt="Bar chart" /></td>
<td><img src="chart4.png" alt="Bar chart" /></td>
</tr>
<tr>
<td>BIRC5</td>
<td><img src="chart5.png" alt="Bar chart" /></td>
<td><img src="chart6.png" alt="Bar chart" /></td>
</tr>
</tbody>
</table>

B

[Western blot images for E2F1 and GAPDH]
Figure 8

The results of experimental verification. A The relative expression of AURKA, RRM2 and BIRC5 in MDA-MB-231 and MCF-10A cells by RT-PCR. B The expression of E2F1 protein in MDA-MB-231 and MCF-10A cells by Western blot. Data were calculated by paired Student’s t-test. *P < 0.01

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable4.docx
- SupplementaryTable3.docx
- SupplementaryTable2.docx
- SupplementaryTable1.docx