Identification of potential crucial genes associated with breast cancer using bioinformatics analysis and experimental verification

Xiaoyu Ni  
Zhuzhou Central Hospital

Haibing Yang  
Zhuzhou Central Hospital

Chao Liu (✉ xqwdbb@163.com)  
Zhuzhou Central Hospital

Research Article

Keywords: Breast cancer, Integrated bioinformatics, Differentially expressed genes, Biological pathways

Posted Date: January 12th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2457642/v1

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

In this study, we identified a total of 492 DEGs, including 176 up-regulated and 316 down-regulated DEGs. GO analysis showed that the up-regulated DEGs are mainly involved in cell division, nucleus and protein binding. The down-regulated DEGs mainly involve immune response, extracellular exosome and calcium ion binding. Top five enriched pathways obtained in the KEGG pathway analysis are pathways in cancer, cytokine-cytokine receptor interaction, focal adhesion, the PI3K-akt signaling pathway and ECM-receptor interaction. Top 10 up-regulated hub genes identified from the PPI network are AURKA, CDC6, CCNA2, CDCA8, NUSAP1, CDK1, CCNB1, CCNB2, UBE2C, HMMR. The top 10 down-regulated hub genes are IGF1, JUN, FGF2, CXCL12, KIT, PTGS2, LEP, EGF, EGR1, FOS. Survival analysis showed that the expression levels of WIF1 (P = 0.019) and HMMR (P = 0.027) were correlated with the prognosis of patients with breast cancer. In addition, gene expression and methylation analysis showed that COL11A1 is highly expressed and hyper-methylation. MMP1 is highly expressed and hypo-methylation. SFRP1, WIF1 is low expressed and hyper-methylation in breast cancer. In terms of tumor purity and immune cell infiltration analysis, Interestingly, it is found that HMMR makes a strong connection with B Cell, CD8+ T Cell, neutrophil, dendritic cell (P <0.05). MMP1 was negatively associated with tumor purity. The use of bioinformatics can effectively analyze the data of the gene chip, obtain the inherent information of the organism, and provide the basis for the next experiment. This study identifies key genes and pathways in breast cancer that will advance our understanding of molecular mechanisms.

Introduction

Worldwide, breast cancer is the most common cancer affecting women, and its morbidity and mortality are expected to increase significantly in the coming years. Despite tremendous advances in human cancer research, breast cancer is still a major health issue and represents the highest priority of biomedical research(1). Various forms are currently used in the diagnosis and treatment of breast cancer, such as introducing precision medicine to the challenges associated with cancer care(2). There is ample evidence that lifestyle (high-fat diet, drinking, lack of physical exercise) and environmental factors have an impact on the development of breast cancer. Eliminating these factors (primary prevention) may help reduce morbidity and mortality. Secondary prevention including diagnostic tests (such as mammography, ultrasound, magnetic resonance imaging, breast screening, and modern and more accurate imaging methods) can help early detection of tumors or lesions susceptible to tumors(3). Previous studies have shown that its onset may be linked to genetic, environmental and other factors. In recent years, research has focused on the molecular mechanism of its onset, but the specific etiology is still unclear. Recently, some updates on breast cancer screening recommendations have been released internationally. On the other hand, advances in genomics have made it possible to establish new molecular classifications of breast cancer(4).

Microarray analysis is a novel method to study tumor genes, find molecular targets for tumor drug therapy and monitor prognosis. However, due to the heterogeneity of the experimental samples, the use of different detection platforms and data processing methods will result in inconsistent results. The Robust
Rank Aggregation (RRA) method is suitable for comparing multiple sequence gene lists(5), because this method checks the ranking of each gene in each list and is based on the idea that each gene identified in each experiment is randomly arranged. Therefore, RRA compares the ranking of the randomly ordered list with the baseline situation, while a higher gene ranking is linked to a lower P value. RRA integrates the results of multiple gene expression data sets, thereby enhancing the understanding of the molecular mechanism of tumor genes. Our research will provide reliable molecular markers and effective therapeutic targets for breast cancer.

**Methods And Materials**

**Gained microarray data**

The GSE20711(6), GSE61304(7), GSE139038(8), GSE124646(9), GSE33447(10) and GSE5764(11) gene expression profile matrix files were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The platform of the GSE20711 dataset is the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, and this dataset contains 2 normal breast tissue and 88 breast cancer tissues. The platform of the GSE61304 dataset is the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, and this dataset contains 4 normal breast tissue and 58 breast cancer tissues. The platform of the GSE139038 dataset is the GPL27630 Print_1437, and this dataset contains 24 normal breast tissue and 41 breast cancer tissues. The platform of the GSE124646 dataset is the GPL96, and this dataset contains 10 normal breast tissue and 10 breast cancer tissues. The platform of the GSE33447 dataset is the GPL14550, and this dataset contains 8 normal breast tissue and 8 breast cancer tissues. The platform of the GSE5764 dataset is the GPL570, and this dataset contains 20 normal breast tissue and 12 breast cancer tissues.

**Identification of robust DEGs**

We downloaded series matrix files of datasets from GEO. The R package “limma”(12) was utilized to normalize the data and find DEGs. We then used RRA to integrate the results of those 6 datasets to find the most significant DEGs(13). The P value of each gene indicated its ranking in the final gene list, and genes with adjusted P < 0.05 were regarded as significant DEGs in the RRA analysis.

**Function enrichment analyses**

The commonly used bioinformatics analysis database, DAVID 6.8 database (https://david.ncifcrf.gov/) is a commonly used database for gene enrichment and functional annotation analysis. The database integrates biological data and analysis tools to provide systematic and comprehensive annotations of biological functions for large-scale gene or protein lists. Use DAVID to perform and KEGG pathway enrichment analysis on the identified DEG, and download GO and KEGG pathway enrichment analysis results for subsequent. Then utilize Cytoscape 3.6.1 software to conduct a visual network analysis of the KEGG analysis results. If P <0.05, the result is considered statistically significant.
**PPI network analysis**

Studying the interaction network between proteins helps to mine the core regulatory genes. What we are interested in being actually "gene interaction". The search tool for searching interacting genes/proteins is a search tool that can analyze the interaction between proteins (https://string-db.org/). Using STRING to analyze DEG's PPI network can help us understand this relationship between different genes. Cytoscape software was utilized to screen hub genes according to degree.

**Prognosis analysis and Methylation analyses**

UALCAN is a comprehensive, user-friendly and interactive web resource for analyzing cancer OMICS data. UALCAN provides easy access to published cancer OMICS data (TCGA and MET500) and enables users to identify biomarkers or perform computer verification of potential genes of interest. It provides graphs and graphs describing gene expression and patient survival information based on gene expression, evaluate gene expression in molecular subtypes of breast and prostate cancer, and evaluate epigenetic regulation of gene expression by promoter methylation, and correlate with gene expression. UALCAN conducts a full-oncogene expression analysis. These resources allow researchers to collect valuable information and data about genes/targets of interest(14). We utilized this website to compare methylation levels of hub genes between the breast cancer and paracancerous normal tissues.

**Analysis of gene expression and tumor-infiltrating immune cells**

To investigate the correlation between the expression of selected hub genes and tumor infiltrating immune cells (B cells, CD4$^+$ T cells, CD8$^+$ T cells, neutrophils, macrophages, and dendritic cells), we applied the online tool TIMER (https://cistrome.shinyapps.io/timer/)(15, 16) which contains 10,897 samples from diverse cancer types available in the TCGA database.

**Ethical statement**

The study was approved by the Ethics Committee of Zhuzhou Central Hospital and conducted in accordance with the Declaration of Helsinki. Prior to the start of the study, all participants gave written informed consent.

**Tissue samples and clinical data**

51 breast cancer tissues (age, 45±0.26 years; male/female patient ratio, 1/60) and 32 non-tumor breast tissues (age, 47±0.73 years; n=31 female patients) were collected from the Zhuzhou Central Hospital (Hunan, China) between February in 2015 and July 2019. Patients with diabetes, nephritis, or cardiovascular disease were excluded. Patient information was obtained from medical records. The present study was approved by the Ethics Committee of Zhuzhou Central Hospital. Written informed consent was obtained from all of the participants.

**Cell culture**
The human breast cancer cell lines MCF-7 and breast cell lines MCF10A were obtained from American Type Culture Collection (Manassas, VA, USA). Cells cultured in high glucose Dulbecco's modified Eagle's medium (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA) containing 10% fetal bovine serum (FBS; Gibco; Thermo Fisher Scientific, Inc.) and maintained in a humidified atmosphere of 5% CO2 in air at 37°C.

**RNA extraction, real-time PCR and RT-PCR**

Total RNA was isolated from tissue samples using TRIzol reagent (Invitrogen) according to the manufacturer’s protocol. The cDNA was synthesized from the total RNA using a Reverse Transcription System (Fermentas, Glen Burnie, MD, USA) according to the manufacturer’s instructions. GAPDH was amplified in parallel as an internal control. The expression level of each gene was quantified by measuring the cycle threshold (Ct) values and normalized relative to that of GAPDH using the 2-ΔΔCt method. The primers used in the reaction were as follows:

- COL11A1, (forward, 5'-TAACATCGCTGACGGGAAGTG-3', reverse, 5'-CCGTGATTCCATTGGTATCAACA-3').
- SFRP1, (forward, 5'-ACGTGGGCTACAAGAAGATGG-3', reverse, 5'-CAGCGACACGGGTAGATGG-3').
- MMP1, (forward, 5'-CTCTGGAGTAATGTCACCTCT-3', reverse, 5'-TGTTGGTCCACCTTTTCATCTTC-3').
- WIF1, (forward, 5'-CTGATGGGTTCCACGGACC-3', reverse, 5'-AGAAACCAGGAGTCACACAAG-3').

**Western blot**

Protein was extracted from indicated cells by using RIPA lysis buffer. Protein concentrations were determined using a BCA Protein Assay kit (Thermo Fisher Scientific, Rockford, IL, USA). A total of 60 μg of protein was separated on a 10% SDS-PAGE gel and transferred onto polyvinylidene difluoride membranes (Millipore, Billerica, MA, USA), which were soaked in 5% nonfat milk for 1 h and then incubated with corresponding primary antibodies overnight at 4°C. Antibodies used in this study: rabbit polyclonal anti-COL11A1(ab64883)(1:500 dilution), rabbit polyclonal anti-SFRP1(ab4193)(1:500 dilution), rabbit polyclonal anti-MMP1(ab137332)(1:500 dilution), rabbit polyclonal anti-WIF1(ab186845) (1:500 dilution), and rabbit polyclonal anti-β-Tubulin (1:3000 dilution) from Proteintech (Wuhan, China). After washing with 1×TBST three times for 8 min each, the membranes were incubated with the corresponding secondary antibodies for 1 h at 37°C, and then washed with 1×TBST for three times again, and finally the bands were visualized using an ECL kit (Millipore). Signals were quantified by Image-J software and normalized to β-tubulin.

**Results**

3.1 Identification of differentially expressed genes in breast cancer
The dataset information was presented in Table 1. Owing to the dataset information was chaotic at first, so it must be standardized. The breast cancer chip expression datasets GSE20711, GSE61304, GSE139038, GSE124646, GSE33447, and GSE5764 were normalized, and the results were shown in Fig.S1(Supplementary figure1). We screened DEG using the limma R package (adjusted P <0.05 and |fold change (FC)|>1).

Table 1. Details of the GEO breast cancer data.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Dataset ID</th>
<th>Number of samples</th>
<th>GPL ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>GSE20711</td>
<td>88T 2N</td>
<td>GPL570</td>
</tr>
<tr>
<td>Breast</td>
<td>GSE61304</td>
<td>58T 4N</td>
<td>GPL570</td>
</tr>
<tr>
<td>Breast</td>
<td>GSE139038</td>
<td>41T 24N</td>
<td>GPL27630</td>
</tr>
<tr>
<td>Breast</td>
<td>GSE124646</td>
<td>10T 10N</td>
<td>GPL96</td>
</tr>
<tr>
<td>Breast</td>
<td>GSE33447</td>
<td>8T 8N</td>
<td>GPL14550</td>
</tr>
<tr>
<td>Breast</td>
<td>GSE5764</td>
<td>12T 20N</td>
<td>GPL570</td>
</tr>
</tbody>
</table>

Note: GSE, Gene Expression Omnibus Series; T, tumor samples; N, paracancerous normal samples

3.2 Identification of robust DEGs by RRA analysis

The RRA method is based on the assumption that each gene in each data set is randomly arranged. If genes rank higher in all data sets, the associated P-values will decrease and the probability of differential gene expression will be greater. We identify DEGs using integrated bioinformatics in breast cancer, the top 20 up-regulated and down-regulated DEGs are shown in heatmap (Fig1). Build on the results of RRA analysis, a total of 176 up-regulated and 316 down-regulated significant DEGs were identified.

3.3 GO term analysis of DEGs

GO functional analysis is divided into the following three parts: biological processes (BP), molecular functions (MF), and cellular components (CC). We use the DAVID database and its online analysis tools to annotate GO functions for the integrated DEG. The results were considered statistically significant if P < 0.05, and the three parts of the GO consequences are shown in Figs2 and 3. The up-regulated genes were mainly enriched in cell division (ontology: BP), nucleus (ontology: CC), and protein binding (ontology: MF) and the down-regulated genes were mainly enriched in immune response (ontology: BP), extracellular exosome (ontology: CC) and calcium ion binding (ontology: MF).

KEGG is a database of systematic analysis of gene function, genome information, which serves helps to analyze genes as a whole network. KEGG pathway analysis of the integrated DEGs was performed using the DAVID database, and the results of the analysis are shown in Fig4. The integrated DEGs were mainly enriched in pathways in cancer, cytokine-cytokine receptor interaction, focal adhesion, the PI3K-akt signaling pathway, ECM-receptor interaction.
3.5 Integration of protein-protein interaction (PPI) network analysis

The STRING online database was utilized to analyze the 492 integrated DEGs and to construct a PPI network. The PPI network has a guiding role in the study of breast cancer target genes and proteins, and has a prominent function in the future study of breast cancer. The results were downloaded and analyzed using Cytoscape software. Based on STRING database, we chose the top 10 interacting proteins from up-regulated genes, the protein-protein interactional (PPI) network of DEGs included AURKA, CDC6, CCNA2, CDCA8, NUSAP1, CDK1, CCNB1, CCNB2, UBE2C, HMMR in Fig5. We chose the top 10 interacting proteins from down-regulated genes, the PPI network of DEGs included IGF1, JUN, FGF2, CXCL12, KIT, PTGS2, LEP, EGF, EGR1, FOS in Fig6.

3.6 Association between prognostic significance and methylation of hub genes

Prognostic analysis, Among the hub genes, the following two genes were considered to be associated with the prognosis of breast cancer patients: (Figure7). WIF1(P=0.019), HMMR(P=0.027). We explored the correlation between the expression levels of the four hub genes and their methylation status to elucidate the underlying mechanism of abnormal upregulation in breast cancer tissues. COL11A1 is highly expressed and hyper-methylation in breast cancer. MMP1 is highly expressed and hypo-methylation in breast cancer. SFRP1, WIF1 is low expressed and hyper-methylation in breast cancer in Figs.8 and 9.

3.7 Analysis of tumor-infiltrating immune cells

To investigate the correlation between the expression of selected hub genes and tumor infiltrating immune cells (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells), we applied the online tool TIMER (https://cistrome.shinyapps.io/timer/)(17, 18), which contains 10,897 samples from diverse cancer types available in the TCGA database. Interestingly, HMMR were very positively associated with tumor purity. Simultaneously, HMMR makes a strong connection with B Cell, CD8+ T Cell, neutrophil, dendritic cell(P<0.05). MMP1 were negatively associated with tumor purity. MMP1 has a strong connection with B Cell, CD8+ T Cell, neutrophil, CD4+ T Cell macrophage, and dendritic cells(P<0.05). COL11A1 was negatively associated with tumor purity in Fig10.

3.8 The validation of differential genes by q-PCR and Western blot

To verify the results, we analyzed the expression levels of COL11A1, SFRP1, MMP1 and WIF1 in breast cancer tissues and non-tumor breast tissues samples by qRT-PCR assay. COL11A1 and MMP1 was found to be upregulated, SFRP1 and WIF1 was found to be downregulated in breast cancer tissues when compared to non-tumor breast. However, the results of western blot showed that WIF1 and SFRP1 were downregulated, COL11A1 and MMP1 was upregulated in MCF-7 when compared to MCF10A in Fig11.

Discussion
Breast cancer is the most common malignant tumor in women worldwide, and about 70-80% of patients with early non-metastatic disease can be cured. Advanced breast cancer with distant organ metastases is considered incurable by currently available treatments(19). In 2018, an estimated 2.1 million women were newly diagnosed with breast cancer, and a new case was diagnosed approximately every 18 seconds, in addition, 626,679 women with breast cancer died(20).

In this study, we identified important DEGs between cancerous and normal samples, and conducted a series of bioinformatics analyses to screen for key genes and pathways closely related to breast cancer. The results showed that the GSE20711, GSE61304, GSE139038, GSE124646, GSE33447 and GSE5764 datasets were analyzed using the RRA method, and 492 integrated DEGs were found, including 176 up-regulated and 316 down-regulated significant DEGs. The up-regulated genes were mainly enriched in cell division (ontology: BP), nucleus (ontology: CC), and protein binding (ontology: MF) and the down-regulated genes were mainly enriched in immune response (ontology: BP), extracellular exosome (ontology: CC) and calcium ion binding (ontology: MF). These results suggest that these DEGs are involved in the proliferation and migration of breast cancer cells. KEGG pathway analysis found five significantly enriched pathways. The integrated DEGs were mainly enriched in pathways in cancer, cytokine-cytokine receptor interaction, focal adhesion, the PI3K-akt signaling pathway, ECM-receptor interaction. Defective components in DNA damage and repair mechanisms are the root causes of the occurrence and development of different types of cancer, and breast cancer is no exception(21, 22). Cytokine-cytokine receptor interaction pathway appeared to be a key factor in triple-negative breast cancer drug resistance(23). Interestingly, Wang L's research found that calcium signaling were significantly enriched in breast cancer(24). Gkretsi V study found that the silencing of growth differentiation factor-15 promotes breast cancer cell invasion by down-regulating focal adhesion genes(25). Costa RLB believes that the development of PI3K/AKT/ mTOR pathway drugs for the treatment of breast cancer is an evolving field. In addition to their interaction with altered cancer pathways, the efficacy and toxicity of new drugs should also be considered(26).

The top 10 hub genes were HMMR, AURKA, CDC6, CCNA2, CDCA8, NUSAP1, CDK1, CCNB1, CCNB2, and UBE2C in up-regulated DEGs. HMMR, MDM2 and PALB2 genes' polymorphic site combinations appear to be candidate markers of genetic predisposition with breast cancer in the Kyrgyz population(27). Incomplete inhibition of AURKA was a common source of therapy failure, and combinations of PI3K, AKT or mTOR inhibitors with the AURKA inhibitor MLN8237 were highly synergistic and durable suppressed mTOR signaling, resulting in apoptosis and tumor regression in vivo(28). Cdc6 is a potential prognostic marker and therapeutic target in breast cancer patients(29, 30). The protein encoded by the gene CCNA2 belongs to the highly conserved cyclin family, and its members act as cell cycle regulators. The protein binds and activates cyclin-dependent kinase 2, thereby facilitating the transition through G1/S and G2/M(31, 32). When comparing normal tissue and tumor samples by microarray analysis, the biggest difference most often occurs in the expression level of genes that control cell proliferation(33). The top 10 hub genes were IGF1, JUN, FGF2, CXCL12, KIT, PTGS2, LEP, EGF, EGR1 and FOS in down-regulated DEGs. Paracrine recruitment and activation of fibroblasts by c-Myc expressing breast epithelial cells through the IGFs/IGF-1R axis(34). Sahores shows that HMW-FGF2 isoforms are PRB targets which confer
endocrine resistance and were localized in the nuclei of breast cancer samples(35). JUN is a putative transformation gene for avian sarcoma virus 17. JUN encodes a protein highly similar to viral proteins, which directly interacts with specific target DNA sequences to regulate gene expression. The gene is intron-free and is located at 1p32-p31, a chromosomal region involved in the translocation and deletion of human malignancies(36).

We performed a prognostic analysis of those hub genes using the UALCAN(37). The following two genes were found to be associated with the prognosis of breast cancer patients: WIF1(P=0.019), HMMR(P=0.027). Liu S study found that the expression of WIF1, DKK2, SFRP2 and AXIN2 were positively correlated with the survival of patients(38). HMMR was significantly associated with metastasis and overall survival in patients with lung adenocarcinoma(39). The protein encoded by gene HMMR is involved in cell motility. It is expressed in breast tissue and is expressed together with other proteins, forming a complex with BRCA1 and BRCA2, so it may increase the risk of breast cancer. It has been noted that this gene encodes splice transcript variants of different isoforms(40-42). Silencing of tumor suppressor genes RASSF1A, SLIT2, and WIF1 by promoter hypermethylation in hereditary breast cancer(43). The protein encoded by gene WIF1 acts to inhibit the WNT protein, which is an extracellular signaling molecule that plays a part in embryonic development. Promoter methylation of WNT inhibitory factor-1 may be associated with the pathogenesis of multiple human tumors(44, 45). Moreover, Veeck J's research found have found prognostic relevance of WIF1 and Dickkopf-3 (DKK3) promoter methylation in human breast cancer(46). We also referred to UALCAN to explore DNA methylation patterns that could account for the abnormal expression of the above hub genes in breast cancer. COL11A1 was highly expressed and hyper-methylation in breast cancer. We found that MMP1 was highly expressed and hypomethylation in breast cancer. SFRP1 and WIF1 were low expressed and hyper-methylation in breast cancer. Such as coordinated over-expression of particular collagens, mainly COL11A1. The composition of the overexpressed genes indicates invasion-facilitating altered proteolysis in the extracellular matrix(47).

Ameku T identification of MMP1 as a novel risk factor for intracranial aneurysms in ADPKD using iPSC models(48). Lim JP found that silencing Y-box binding protein-1 inhibits triple-negative breast cancer cell invasiveness via regulation of MMP1 and beta-catenin expression(49). SFRP1 expression is strongly correlated with triple-negative breast cancer on the protein level. Associations with age and tumor grade support the role of SFRP1 as a biomarker for chemotherapy response in triple-negative breast cancer(50). Loss of SFRP1 expression was a significant regulator of typical breast hyperplasias transcriptional profiles driving previously unidentified changes affecting responses to estrogen and possibly other pathways(51).

**Conclusion**

In summary, the purpose of this study is to improve our understanding of the molecular mechanisms of breast cancer development through comprehensive bioinformatics analysis aimed at identifying DEGs and breast cancer progression related pathways. Our research also identified some key candidate genes
and biological pathways that can help find biomarkers and therapeutic targets for breast cancer. However, further molecular biology experiments are needed to verify the results of this study.

**Declarations**

**Author Contributions**

Chao Liu conceived and supervised the study; Chao Liu designed experiments; Xiaoyu Ni Haibing Yang performed experiments; Xiaoyu Ni Haibing Yang analysed data; Xiaoyu Ni Haibing Yang wrote the manuscript; Xiaoyu Ni Haibing Yang made manuscript revisions. All authors reviewed the manuscript.

**Compliance with ethical standards**

Conflict of interests The authors declare that they have no conflict of interests.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

**References**


Figure 1

**Log FC heatmap of each expression microarray.** The abscissa represent the GEO IDs, the ordinate represents the gene name, the red represents log FC > 0, the green represents log FC < 0 and the value in the box represents the log FC value.
Figure 2

**Top 15 enriched GO terms.** (A) Upregulated DEGs with the top 15 enriched GO terms. (B) Downregulated DEGs with the top 15 enriched GO terms.

Figure 3

**Distribution of integrated DEGs in breast cancer for different GO-enriched functions.** (A) Up-regulated DEGs. (B) Down-regulated DEGs.
Figure 4

KEGG pathway enrichment analysis of the integrated DEGs
Figure 5

The up-regulation DEGs with highest connectivity degree of PPI network. $P < 0.05$ was considered statistically significant.
Figure 6

The down-regulation DEGs with highest connectivity degree of PPI network. P < 0.05 was considered statistically significant.
Figure 7

Kaplan-Meier analysis results of hub genes (P < 0.05). Four hub genes were found to be associated with the prognosis of breast cancer patients. (A) WIF1 (P=0.019), (B) HMMR (P=0.027).
Figure 8

Validation of hub genes in the TCGA dataset. (A) COL11A1, (B) SFRP1, (C) MMP1, and (D) WIF1. Gene expression differences between breast cancer and adjacent normal tissues.
Figure 9

Methylation analyses of breast cancer hub genes. The methylation levels of (A) COL11A1, (B) SFRP1, (C) MMP1, and (D) WIF1.
Figure 10

Association of hub genes’ expression with immune infiltration in breast cancer. (A) COL11A1. (B) SFRP1. (C) MMP1. (D) WIF1. (E) HMMR. P<0.05 denotes significance. Each dot represents a sample in the TCGA dataset.
The gene COL11A1, SFRP1, MMP1, WIF1, HMMR validated with qPCR and western blot. P < 0.05 was considered statistically significant.

Supplementary Files

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

- FigS1.tif