Using Weighted Gene Co-Expression Network Analysis to Identify Increased MND1 Expression as a Predictor of Poor Breast Cancer Survival

Zhaokang Bao1,*, Jiale Cheng1,*, Jiahao Zhu2,*, Shengjun Ji3, Ke Gu1, Yutian Zhao2, Shiyou Yu1, You Meng1

1Department of Oncology Surgery, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, Jiangsu, People’s Republic of China; 2Department of Radiotherapy and Oncology, Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu, People’s Republic of China; 3Department of Radiotherapy and Oncology, The affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, Jiangsu, People’s Republic of China

*These authors contributed equally to this work

Correspondence: You Meng, Department of Oncology Surgery, The affiliated Suzhou Hospital of Nanjing Medical University, 16 West Baita Road, Suzhou, Jiangsu, People’s Republic of China, Email drmengyou@163.com

Objective: We used bioinformatics analysis to identify potential biomarker genes and their relationship with breast cancer (BC).

Materials and Methods: We used a weighted gene co-expression network analysis (WGCNA) to create a co-expression network based on the top 25% genes in the GSE24124, GSE33926, and GSE86166 datasets obtained from the Gene Expression Omnibus. We used the DAVID online platform to perform GO and KEGG pathway enrichment analyses and the Cytoscape CytoHubba plug-in to screen the potential genes. Then, we related the genes to prognostic values in BC using the Oncomine, GEPIA, and Kaplan–Meier Plotter databases. Findings were validated by immunohistochemical (IHC) staining in the Human Protein Atlas and the TCGA-BRCA cohort. LinkedOmics identified the interactive expressions of hub genes. We used UALCAN to evaluate the methylation levels of these hub genes. MethSurv and SurvivalMeth were used to assess the multilevel prognostic value. Finally, we assessed hub gene association with immune cell infiltration using TIMER.

Results: The mRNA levels of MKI67, UBE2C, GTSE1, CCNA2, and MND1 were significantly upregulated in BC, whereas ESR1, THSD4, TFF1, AGR2, and FOXA1 were significantly downregulated. The DNA methylation signature analysis showed a better prognosis in the low-risk group. Further subgroup analyses revealed that MND1 might serve as an independent risk factor for unfavorable BC prognosis. Additionally, MND1 expression levels positively correlate with the immune infiltration statuses of CD4+ T cells, CD8+ T cells, B cells, neutrophils, dendritic cells, and macrophages.

Conclusion: Our results indicate that the ten hub genes may be involved in BC’s carcinogenesis, development, or metastasis, and MND1 may be a potential biomarker and therapeutic target for BC.

Keywords: breast neoplasms, immunohistochemistry, polymerase chain reaction, biomarker

Introduction
Breast cancer (BC) is the most commonly diagnosed cancer in women, and it is the main cause of cancer mortality in women around the world.1 Although standard treatments for breast cancer, including surgery, radiotherapy, chemotherapy, endocrine therapy, targeted therapy, and immunotherapy, have greatly improved during the last few decades,2,3 however, BC remains a serious clinical challenge. The outcomes of patients with different molecular BC subtypes are very different. For example, the triple-negative breast cancer (TNBC) subtype usually has poorer progression-free survival (PFS) and overall survival (OS) outcomes, and it has higher risks of recurrence and metastasis compared to other molecular BC subtypes.3 There is no effective treatment for TNBC. However, prognostic biomarkers are a valuable tool to develop treatments. For example, the expression level of human epidermal growth factor receptor 2 (HER2) is one...
of the most effective prognostic biomarkers for breast cancer, and anti-HER2 drugs, including trastuzumab and lapatinib, are highly beneficial for BC patients with HER2 overexpression. To better treat BC, novel prognostic biomarkers with high sensitivity and specificity and effective therapeutic targets need development.

Weighted gene co-expression network analysis (WGCNA) is now widely used to explore potential genes associated with clinical characteristics of cancer progression, which include tumor grade, stage, and metastasis. Furthermore, due to easy access to public databases, such as The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO), bioinformatics analysis has become a popular tool to screen tumor target genes. For example, Tang et al found that the hub genes RSAD2, HERC5, and CCL8 are potential prognostic biomarkers for BC. Moreover, miR-124 regulates CDK6 via the Wnt signaling pathway in BC. This study analyzed BC expression profile data from the GEO public database using WGCNA and identified BC-associated gene modules. Five hub genes, namely MKI67, UBE2C, GTSE1, CCNA2, and MND1, were eventually selected from two modules. By combining the IHC assays of the Human Protein Atlas and the TCGA-BRCA cohort, we confirmed the differential expression of MND1 in BC tissues and its association with BC prognosis. Moreover, we investigated the molecular mechanisms involving MND1 during carcinogenesis and the development of BC and the possible connection between MND1 and tumor immune infiltration.

**Materials and Methods**

**Dataset Collection and Processing**

Supplementary Figure 1 illustrates the study’s workflow. Hsieh et al, Chang et al, and Cheng et al provided the GSE24124, GSE33926, GSE86166 gene expression profile datasets we retrieved and analyzed these data with the GEO (https://www.ncbi.nlm.nih.gov/geo). The gene expression data based on 20, 51, and 366 BC samples were generated using the GPL887 (Agilent-012097 Human 1A Microarray V2 G4110B, Feature Name version), GPL7264 (Agilent-012097 Human 1A Microarray V2 G4110B Feature Name version), and GPL15048 (Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray HuRSTA_2a520709.CDF) platforms, respectively. The datasets used in this study were normalized by batch, and the top 25% of variant genes were selected to build weighted co-expression networks.

The gene expression dates (WorkflowType: HTSeqFPKM) for 1096 BC samples and 119 normal breast samples were downloaded from the TCGA database (http://cancergenome.nih.gov/). The relationship between MND1 and the prognosis of BC patients was evaluated using clinical data.

The genes that were differentially expressed (DEGs) between tumor and normal samples were visualized using the R package. Here, we define a fold-change that is correlated with significant DEGs as a $|\log_2$ fold-change$|$ with a threshold $>1.5$ and false discovery rate (FDR) $< 0.05$.

**Weighted Gene Co-Expression Network Construction**

We constructed a scale-free gene co-expression network using the “WGCNA” package in the R software environment. To ensure the network structure’s stability, we removed cases with incomplete clinical information. We also transformed the gene expression profile matrix by x. The similarity between genes “m” and “n” was defined as $S_{mn} = |\text{cor}(m, n)|$. A weighted adjacency matrix was constructed by the power function, $w_{mn} = \text{power}(S_{mn}, \beta) = |S_{mn}|^\beta$. The “soft thresholding” parameter highlighted similarities between genes, while poor correlations were penalized. The WGCNA package’s “pickSoftThreshold” x was used to find an acceptable value for a scale-free network. The resulting adjacency matrix was transformed to a topological overlap matrix and used to determine dissimilarity values (1-TOM). Topological overlap assesses the biological similarity between two genes based on their expression patterns. A dissimilarity-based cluster analysis was then performed to build a hierarchically clustered tree, with each leaf in the branch representing a gene and each branch representing a gene module. Gene modules were extracted via the complex tree cut approach, with the minimum size cutoff set to 30 and the deep break parameter set to 2. With a height cutoff of 0.25, only extremely related modules were grouped together.
Identification of Modules with Clinical Significance
We used two approaches to screen modules linked to specific clinical features of BC. First, principal component analysis was used to generate the module eigengene (ME), which captures the bulk of the module’s knowledge. Second, Pearson’s correlation coefficients were used to calculate the ME and different clinical characteristics’ correlations. Modules that were highly clustered (P < 0.05) were eliminated. Clinically relevant modules were identified by examining the associations between gene significance and module membership. Using these two approaches, we could select the most strongly associated modules with BC phenotypes for further study.

Function and Pathway Enrichment Analysis
To better understand the biological processes involved in the main modules, we used the Database for Annotation, Visualization, and Integrated Discovery (DAVID) to conduct Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment studies (https://david.ncifcrf.gov). A P value of <0.05 indicated significant enrichment.

Identification and Validation of Hub Genes in Pivotal Modules
We identified hub genes using the CytoHubba plug-in of Cytoscape 3.5.1. The candidate genes were selected from the top ten nodes with the highest degree ranking. The Kaplan–Meier plot (http://kmplot.com) was then used to assess the association between hub gene expression and BC prognosis. GEPIA (http://gepia.cancerpku.cn/) was used to evaluate the differential expression levels of x. We used the Oncomine database (https://www.oncomine.org/) to analyze the differential expression of x between cancer and normal tissue. The expression of each hub gene was validated in the Human Protein Atlas (http://www.proteinatlas.org). Differences in x were considered statistically significant at P < 0.05.

LinkedOmics Database Analysis
The LinkedOmics function module (http://www.linkedomics.org/login.php) was used to analyze GO categories and KEGG pathways through gene set enrichment analysis. The rank criterion was based on a FDR of <0.05. A total of 500 simulations were performed.

Expression and Methylation Analysis
UALCAN (http://ualcan.path.uab.edu/) was used to access the hub genes expression and methylation. MethSurv (https://biit.cs.ut.ee/methsurv/) to evaluate the prognostic significance of different CpG methylation patterns of hub genes. SurvivalMeth (http://bio-bigdata.hrbmu.edu.cn/survivalmeth/) was used to access the effect of DNA methylation of hub genes signature on BC prognosis.

International Journal of General Medicine 2022:15
https://doi.org/10.2147/IJGM.S354826
Dovepress
Bao et al

Powered by TCPDF (www.tcpdf.org)
TIMER Analysis
The Spearman correlations between hub gene expression and immune infiltration abundance were determined using the “Gene” module and the TIMER database (https://cistrome.shinyapps.io/timer/). The “Survival” module determined the relationships between clinical outcomes and immune cell infiltration or hub gene expression.26

Potential Drugs Screening
Connectivity Map tools (https://clue.io/) was used to analyze the mechanism of actions and potential drugs.27–29 Top upregulated and downregulated hub genes was input. “ComplexHeatmap” package in R was used to visualized the results.

Statistical Analysis
All statistical analyses used SPSS software (version 24; SPSS Inc., Chicago, IL, USA). Chi-squared and Fisher’s exact tests, Wilcox test, Kaplan–Meier method, and the Cox proportional regression model analyzed differences in the expression of hub genes. And Two-side P values < 0.05 were considered significant.

Results
Gene Screening
The gene expression profile dataset containing 417 BC and 20 normal breast samples was downloaded from the GEO database (Supplement of batch correction was shown in Supplementary Figure 2). The expression matrix files and clinical data (.txt format) were obtained and sorted from the TCGA database before WGCNA analysis. The “edgeR” package in R Studio was used to process the data, resulting in a total of 484 differentially expressed genes (DEGs) for further WGCNA analysis (Supplementary Figure 3).

Construction of the Co-Expression Network Using WGCNA
The selected genes were clustered hierarchically using the “FlashClust” package in R. Based on the results, 51 outlier samples were omitted, and 366 samples were used to construct the co-expression network (Figure 1A and B). Scale independence and mean connectedness were calculated with soft thresholding settings (β) ranging from 1 to 20. The degree of independence was more than 0.8, when the value of β reached 6, and the associated mean connectivity was near to zero, implying that the network complies with the scale-free network distribution criteria. As a result, the co-expression network was built with a value of 6 (Figure 1C). The dynamic tree cut approach was used to identify and merge gene modules that were close in proximity, resulting in a total of seven gene modules illustrated in Figure 1D (the gray module includes genes that were not co-expressed with other genes). Correlations among the seven modules were analyzed and plotted in the heat map shown in Figure 1E.

Modules with Clinical Significance
Figure 1E shows the gene modules significantly correlated to certain clinical features. The blue module is negatively correlated to ER (estrogen receptor) and PR (progesterone receptor) positively correlated to grade. The brown module positively correlated to ER and PR, and negatively correlated to grade. Furthermore, GS (gene significance) and MM (Module membership) are significantly correlated within the blue module (cor = 0.37, P = 6.3e-06; cor = 0.62, P = 2.4e-16; cor = 0.83, P = 4.6e-37), indicating that this module is closely related to BC (Figure 1F–H). In addition, GS and MM in the brown module are also significantly correlated (cor = 0.88, P = 4.8e-22; cor = 0.88, P = 4.8e-22; cor = 0.74, P = 1.9e-12), indicating that the brown module is also closely related to BC (Figure 1I–K). These results are supported by the clustering heat maps of ER, PR, and grade (Figure 1L–N).

Screening of Hub Genes in Key Modules and Functional Enrichment Analysis of Genes
We used Cytoscape to calculate each node’s degree in modules X and Y, and the top five nodes were selected as candidate genes. Figure 2A shows the blue module network. The hub nodes are: Marker of Proliferation Ki-67 (MKI67), Ubiquitin Conjugating Enzyme E2C (UBE2C), G2, and S-Phase Expressed 1 (GTSE1), Cyclin A2 (CCNA2), and...
Meiotic Nuclear Divisions 1 (MND1). The brown module network (Figure 2B) shows the hub nodes Estrogen Receptor 1 (ESR1), Thrombospondin Type 1 Domain Containing 4 (THSD4), Trefoil Factor 1 (TFF1), Anterior Gradient 2 (AGR2), and Forkhead Box A1 (FOXA1).

GO enrichment analysis on the blue module indicates its involvement in cell cycle, cell division, cell cycle phase transition, regulation of chromosomal segregation, PID Plk1 pathway, and cyclin-dependent protein serine/threonine regulation kinase activity, and positive regulation of the cell cycle process (Figure 2C). Similarly, the pathways involved in the brown module include evolution-dependent gene expression, response to iron ion, regulation of fibroblast promotion, cellular amino acid biosynthetic process, post-translational modification, synthesis of GPI-anchored proteins, etc (Figure 2D).

Figure 1 Co-expression network constructed with WGCNA (weighted gene co-expression network analysis). (A) No outlier samples were identified based on hierarchical cluster analysis. (B) Hierarchical clustering of the top 25% variant genes in the GSE24124, GSE33926, and GSE86166 dataset and associated clinical information. Color intensity indicates a positive correlation with ER (estrogen receptor), PR (progesterone receptor), Her2, and the grade, status, and stage of tumors. (C) Determination of the optimal soft thresholding value (β) of the sample. (D) Dendrogram of the top 25% variant genes clustered based on dissimilarity values (1-TOM), where TOM is the topological overlap matrix. The upper panel shows the gene dendrogram, and the lower panel shows seven gene modules displayed in different colors, including the gray module. (E) Heat map of the correlation between modules and clinical features. The number above each cell represents the correlation between the module and the feature, and the number below each cell represents the corresponding P value. Red and green represent positive and negative correlations, respectively. The scatter plots show the correlations between the blue module gene and ER (F), PR (G), and grade (H); the correlations between the brown module gene and ER (I), PR (J), and grade (K); and the correlations between each module and ER (L), PR (M), and grade (N).
Expression and Clinical Significance of the Hub Genes

The ten candidate genes identified above were analyzed using the KM Plotter online tool. Their expressions were significantly upregulated in BC tissues compared with normal tissues (P < 0.05, Figure 3A). And as shown in Figure 3B, overexpression of MKI67, UBE2C, GTSE1, CCNA2, and MND1 in the blue module significantly correlated with the poor OS of breast cancer (P < 0.05).

The different expression levels in BC and normal breast tissues of the four genes in the brown module were validated using the GEPIA database. In BC tissues, ESR1, THSD4, TFF1, AGR2, and FOXA1 were significantly upregulated (P < 0.05, Supplementary Figure 4A). Survival analysis showed that ESR1, THSD4, and TFF1 significantly correlated...
Figure 3 Survival analysis and differential expression of hub genes in blue modules. (A) Differential expression of hub genes analyzed using the GEPIA database. Box plots show mRNA expression of hub genes in breast cancer (red plot) and the corresponding normal tissues (gray plot). * represents P < 0.05. (B) Survival analysis of the hub genes in breast cancer (Kaplan–Meier plot). (C) Relationship analysis between the expression of hub genes and stage of breast cancer. (D) Differential expression of hub genes analyzed using the Oncomine database. Differential expression of hub genes in different tumors. The intensity of the color (red or blue) indicates positive correlation with the degree of significance of upregulation or downregulation, respectively. (E) Expression of CCNA2, MKI67, MND1, and UBE2C in breast cancer and normal breast tissues validated in the Human Protein Atlas.
with BC prognosis (P < 0.05), while AGR2 was only marginally correlated with BC prognosis (P = 0.01). Moreover, high expression levels of ESR1, THSD4, TFF1 and AGR2 are related to better BC prognosis compared to those in patients with low-expression levels of these genes (Supplementary Figure 4B). Thus, high expression levels of these genes may inhibit BC development.

We chose the blue gene module due to its cancer-promoting function to conduct follow-up research. Analysis of the association between tumor stage and hub gene expression showed a significant relationship (P < 0.05) (Figure 3C). Using the Oncomine database, we analyzed the expression levels of hub genes in different types of cancer and found that MKI67, UBE2C, GTSE1, CCNA2, and MND1 were upregulated in most cancers, although they were also downregulated in a few types of cancers (Figure 3D). In the human protein atlas database, the hub gene was upregulated in BC tissues (Figure 3E).

Expression Validation of the Five Hub Genes by qRT-PCR
15 normal breast samples and 15 BC samples were collected. As shown in Figure 4A–E, MKI67, UBE2C, GTSE1, CCNA2, and MND1 expression levels in BC samples were significantly higher compared with normal breast samples (P < 0.01).

Prognostic Value of Single CpG of Hub Genes in BC
MethSurv was used to investigate the prognosis value of DNA methylation of these five hub genes. The heat map of DNA methylation results is displayed in Supplementary Figure 5. We observed that five CpGs of UBE2C, five CpGs of CCNA2, two CpGs of GTSE1, one CpG of MKI67 and six CpGs of MND1 were significantly associated with prognosis in breast cancer patients (Table 1).

Validation of MND1 in the TCGA Database
MND1 expression data and related clinical information were downloaded from the TCGA-BRCA database. As shown in Figure 5A–F, higher expression of MND1 in luminal A, triple negative and the whole BC has close association with worse OS and PFS. The expression of MND1 in luminal A, luminal B, HER2 overexpression, and triple negative BC tissue was significantly higher than that in normal tissue. We considered the median value of MND1 expression as the threshold, thus the upper 50% was the high expression group, and the lower 50% was the low-expression group of MND1 (Figure 5G). Univariate Cox analysis showed that high expression of MND1, age, and tumor stage are independent risk factors for the prognosis of BC patients, with hazard ratio values of 1.128, 1.035, and 1.626, respectively (all P < 0.05; Figure 5H). Similarly, multivariate Cox analysis showed that high expression of MND1, age, and M stage are independent risk factors for BC patients’ prognoses, with hazard ratio values of 1.213, 1.019, and 1.640 (all P < 0.05; Figure 5I).

Methylation Analysis
Methylation level in BC and normal tissues was investigated (Figure 6A–J). Almost all of the ten hub genes have the significant hypermethylation. SurvivalMeth was used to analyze the prognostic value of ten signature (Figure 6K–M). An optimal cutoff (prognostic index = −26.057) was calculated and all BC cases were divided into high/low risk groups. Significant expression patterns of all ten hub genes were found between the low- and high-risk groups (Figure 6K). The heat map showed that DNA methylation level (Figure 6L), and the high-risk group demonstrated a worse OS compared with the low-risk group (P=0.012) (Figure 6M).

Correlation Analysis of MND1 Expression and Immune Infiltration Status
To explore the potential correlation between MND1 and the immune microenvironment of BC, we evaluated the relationship between MND1 expression and immune infiltration level in the TCGA-BRCA cohort using the TIMER database. The expression of MND1 was positively correlated with the infiltration level of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells in patients with BC (r = 0.255, P = 3.35e-16; r = 0.172, P < 0.05) = 30e-08; r = 0.769, P = 5.58e-10; r = −0.319, P = 7.14e-25; r = 0.266, P = 1.38e-17; r = 0.308, P = 2.84e-23, respectively)
MND1 is also upregulated in a variety of cancers besides BC, including lung adenocarcinoma, colon adenocarcinoma, and head and neck squamous cell carcinoma (Figure 7B).

Potential Mechanism of Action of Small Molecules Screening

CMap mode-of-action analysis of the 42 compounds revealed 35 mechanisms of action shared by the above compounds (Figure 8).

Discussion

We screened for potential biomarkers associated with BC’s progression and prognosis using the GSE24124, GSE33926, and GSE86166 gene expression profiles from the GEO database. These profiles include 417 BC and 20 normal breast samples. Using WGCNA, we constructed a co-expression network of 484 of the most variable genes, resulting in the identification of seven independent modules. We then used the DAVID database to perform enrichment analyses on two selected modules. In the blue module, we found mainly genes involved in the cell cycle, cell division, cell cycle phase transition, regulation of chromosome segregation, PID Plk1 pathway, and cyclin-dependent protein serine/threonine regulation...
kinase activity, and positive regulation of the cell cycle process. In the brown module, we found genes involved in biological processes, such as evolution-dependent gene expression, response to iron ion, regulation of fibroblast promotion, cellular amino acid biosynthetic process, post-translational modification, synthesis of GPI-anchored proteins, etc. Cell cycle-related biological functions are involved in developing many tumors, including BC, while the PID Plk1 pathway is associated with tumor cell proliferation and recurrence. Therefore, these modules may influence the development of BC by regulating the above biological processes and pathways.

The five genes with the highest degree of x in the blue and brown modules were selected using the CytoHubba plug-in in x. Further analysis demonstrated that the expression levels of MKI67, UBE2C, GTSE1, CCNA2, and MND1 in blue modules were negatively associated with prognosis and were significantly upregulated in BC. Thus, these five genes may promote the development of BC. Meanwhile, ESR1, THSD4, and TFF1 in brown modules were positively correlated with favorable patient prognosis and were significantly downregulated in BC. High levels of expression of these genes may thus serve as a tumor suppressor. The five hub genes (MKI67, UBE2C, GTSE1, CCNA2, and MND1), that likely affect BC development were selected for further analysis. MKI67 is also known as Antigen Ki-67, and it is a well-known cell cycle-related protein. Ki-67/MKI67 has both therapeutic and prognostic value in various cancers, including breast, cervical, colorectal, and gastric cancers. Ki67 has been designated a therapeutic target for cancer therapy due to its important cell proliferation role. Recently, Zhang et al achieved favorable outcomes after using a novel siRNA to inhibit the proliferation and induce the apoptosis of 786-O human renal carcinoma cells. In another study, replicative adenovirus vectors combined with Ki67 contributed to the destruction of melanoma cells. Thus, the application of Ki67-targeted cancer therapy is expected to treat BC. UBE2C, a member of the E2 ubiquitin conjugating enzyme family, is involved in mitotic cyclin B degradation, and it promotes the transition from the M to the G1 phase of the cell cycle. Dastsooz et al showed that UBE2C is upregulated in all 27 different types of cancer, a finding that agrees with our results showing increased somatic expression of UBE2C in BC. Moreover, Qin et al found that UBE2C upregulation is associated with poor prognosis in BC. Rawat et al demonstrated that inhibiting UBE2C sensitizes BC cells to radiation, doxorubicin, and hormone-blocking agents. Therefore, UBE2C is a potential therapeutic target for the enhancement of radiation and chemo-sensitization treatments against BC. CCNA2, also known as CyclinA2, belongs to the highly conserved cyclin

| Table 1 The Prognostic Value of CpG in the Upregulated Five Hub Genes by MethSurv (P<0.05) |
|---------------------------------------------|---------|----------------|----------------|
| Gene-CpG | HR     | LR Test P value |
| UBE2C-TSS1500-N_Shore-cg00242976 | 0.538   | 0.008          |
| UBE2C-1stExon; 5'TR; Body-S_Shore-cg02405476 | 0.649 | 0.038          |
| UBE2C-5'TR; 1stExon; Body-S_Shore-cg03969725 | 1.596 | 0.021          |
| UBE2C-5'TR; TSS1500; sExon-Island-cg212563604 | 1.614 | 0.036          |
| UBE2C-1stExon; 5'TR; TSS200; Body-cg13853156 | 0.399 | 0.000          |
| CCNA2-TSS200-Island-cg01854459 | 0.520   | 0.002          |
| CCNA2-TSS200-Island-cg20391652 | 0.530 | 0.003          |
| CCNA2-TSS200-Island-cg22219489 | 1.692 | 0.008          |
| CCNA2-1stExon; 5'TR; Island-cg10002561 | 0.637 | 0.027          |
| CCNA2-TSS1500-S_Shore-cg18913527 | 0.621 | 0.042          |
| GTSE1-Body-S_Shel-cg10084251 | 1.662 | 0.018          |
| GTSE1-3'UTR-S_Shore-cg21542574 | 0.658 | 0.048          |
| MKI67-TSS1500-Island-cg26235537 | 0.603 | 0.012          |
| MND1-Body-Island-cg13825058 | 0.636 | 0.028          |
| MND1-Body-Island-cg18350702 | 0.662 | 0.037          |
| MND1-Body-S_Shore-cg06303282 | 1.751 | 0.005          |
| MND1-Body-S_Shore-cg14104036 | 1.650 | 0.011          |
| MND1-TSS1500-N_Shore-cg13326581 | 1.631 | 0.014          |
| MND1-TSS1500-N_Shore-cg14320796 | 1.633 | 0.014          |
Figure 5 Survival analysis and differential expression of MND1 (Meiotic Nuclear Divisions 1) in the molecular subtypes of BC with TCGA-BRCA database. (A, C, E) Correlation of MND1 expression level with overall survival in luminal A, triple negative, and the whole cases of BC. (B, D, F) Correlation of MND1 expression level with progression free survival in luminal A, triple negative, and the whole cases of BC. (G) Box plot validating mRNA expression of MND1 among luminal A, luminal B, HER2 overexpression, triple negative BC and normal tissues. (H and I) Forest plot based on univariate and multivariate Cox regression analyses of overall survival. Significant differences are defined by ***: P < 0.001, ****: P < 0.0001.
family and is found in almost all human body tissues. Results of several studies have shown that CCNA2 is overexpressed in various malignancies and may be involved in the processes of epithelial-mesenchymal transitions (EMT) and metastasis. Gao et al showed that CCNA2 might serve as a biomarker for the prognosis of ER+ BC and monitor the efficacy of tamoxifen. In another study, cyclinA2 proteins were suppressed by aspirin and salicylic acid in multiple cancer cell lines, indicating that they may serve as a novel target and mechanism of chemoresistance. Previous studies have demonstrated that GTSE1, an x that is upregulated in dozens of cancers, plays an essential role in cell cycle regulation and is involved in chemoresistance, chromosome instability (CIN), and metastasis. BC patients with high GTSE1 levels always have poor OS and are unlikely to survive disease-free, indicating that GTSE1 overexpression correlates with unfavorable outcomes. The silencing of GTSE1 results in cancer cell death by increasing its sensitivity to paclitaxel and cisplatin. Thus, GTSE1 can be considered a new target for the treatment of cancers, including BC. MND1, along with HOP2, is part of a highly conserved heterodimeric protein complex that plays important roles in binding DNA and stimulating the recombinase activities of RAD51 and DMC1, which are indispensable for meiotic recombination.

Figure 6 Methylaction analysis in BC with UALCAN and SurvivalMeth online tool. (A–J) Methylation level of ten hub genes in BC cases. (K) The methylation level of CpGs in the high- and low-risk group. (L) The heatmap of CpG methylation level. (M) The Kaplan–Meier curve of OS. Significant differences are defined by “*”: P < 0.05.
Another study demonstrated that mRNA patterns combining AKT3, PTTG1, MND1, and CENPF can be associated with the prognosis of BC patients. The Cox multivariate regression model based on the TCGA-BRCA cohort demonstrated that high MND1 expression, metastasis, and age are independent risk factors affecting BC’s prognosis. MND1 expression is both positive and significantly greater in the BC group than in the normal breast groups, according to IHC staining results. The Cox multivariate regression model based on the TCGA-BRCA cohort demonstrated that high MND1 expression, metastasis, and age are independent risk factors affecting BC’s prognosis. MND1 upregulation may play an important role in BC, and MND1 could serve as a useful biomarker for screening early BC and assessing the prognosis of BC patients.

While evidence for the association between MND1 and cancers is sparse, Wu et al found that this gene is included in a panel of four genes (MND1, FOXM1, CDC6, and POLE2) that are significantly associated with low survival against squamous cell lung cancer. Another study demonstrated that mRNA patterns combining AKT3, PTTG1, MND1, and CENPF can potentially serve as a strong biomarker for monitoring the growth and metastasis of ovarian cancer. A recent study conducted by Dastsooz et al showed that the expression of MND1 is associated with BC prognosis. However, no study has yet explored the biological functions of MND1 and how its role in tumors. As a result, we selected MND1 as a major hub gene for further investigation, analyzing its expression in BC as well as its relationship to clinical features and prognosis.

MND1 expression is both positive and significantly greater in the BC group than in the normal breast groups, according to IHC staining results. The Cox multivariate regression model based on the TCGA-BRCA cohort demonstrated that high MND1 expression, metastasis, and age are independent risk factors affecting BC’s prognosis. MND1 upregulation may play an important role in BC, and MND1 could serve as a useful biomarker for screening early BC and assessing the prognosis of BC patients.

The presence of immune cells in tumors is closely related to a patient’s clinical prognosis. Immune cell infiltration is the movement of immunological cells from the bloodstream to tumor tissue, which is a part of cancer’s immune escape strategy. Anti-cancer medicines also target immune cell infiltration. TIMER analysis showed the expression of MND1 is positively correlated with the infiltration level of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells in BC patients. Immune cell infiltration deficiency or malfunctioning in malignancies can lead to immune escape and increase tumor development. Novel treatments targeting tumor immune escape, notably inhibitory receptors involved in T cell activity control, have been proposed. However, further experimental verification is needed to...
identify reliable BC immunotherapy targets and understand the relationship between MND1 and immune infiltrating cells and the effect of MND1 on the BC immune system. MND1 is upregulated in a variety of cancers besides BC; specifically, it is upregulated in lung adenocarcinoma, colon adenocarcinoma, and head and neck squamous cell carcinoma.

**Conclusions**

Bioinformatics analysis based on the construction of a co-expression network by WGCNA has identified five BC hub genes: MKI67, UBE2C, GTSE1, CCNA2, and MND1. These genes may participate in BC and impact BC prognosis. Moreover, we speculate that MND1, via immune-related biological processes and pathways, may act as a tumor suppressor gene in BC development. Determining the clinical value of our findings and the molecular mechanisms involved will require further exploration in future studies.

**Ethics Approval**

Ethical approval for this study was obtained from the Clinical Research Ethics Committee of Suzhou Municipal Hospital. This study was conducted in accordance with the declaration of Helsinki. Written informed consent was obtained from all participants.

**Acknowledgments**

This study was supported by Suzhou Applied Basic Research Fund (SKJY2021126); Beijing Xisk Clinical Oncology Research Foundation (Y-Young2021-0087); Wuxi Taihu Lake Talent Plan, Supports for Leading Talents in Medical and Health Profession; Maternal and Child Health Research Project of Jiangsu Commission of Health (F202009); Scientific Research Project of Jiangsu Maternal and Child Health Association (FYX202016); and Scientific Research Project of Wuxi Commission of Health (M202041).
Disclosure

The authors have no conflicts of interest to declare.

References


