Abstract. Breast cancer is the most prevalent malignancy among females, but the molecular mechanisms involved in its pathogenesis and progression have remained to be fully elucidated. The aim of the present study was to identify novel potential therapeutic targets for breast cancer. The dataset GSE76275 was downloaded from the Gene Expression Omnibus database and weighted gene co-expression network analysis (WGCNA) was performed to identify hub genes. Furthermore, the dataset GSE25055, containing gene expression data and clinical information, was downloaded to validate the expression and survival association of these hub genes. In addition, the datasets GSE25065 and GSE42568 were used to validate the association between hub gene expression levels and clinical features. Immunohistochemistry (IHC), reverse transcription-quantitative PCR, as well as proliferation, migration, invasion and apoptosis assays, were used to verify gene expression and function. A total of 4,052 genes were selected for WGCNA and 18 modules were established; the red module was identified as the key module, as it had a strong positive correlation with the tumor grade. Survival analyses of hub genes [S-adenosylmethionine decarboxylase proenzyme (AMD1), homeobox protein engrailed-1 (EN1) and vestigial-like protein 1 (VGLL1)] indicated that higher levels of gene expression were associated with poor prognosis of patients with breast cancer. This association was based on survival analysis of GSE25055 using the Kaplan-Meier plotter tool. Expression validation revealed that the upregulation of hub genes was associated with advanced tumor grade and malignant molecular subtype (basal-like). IHC results from the Human Protein Atlas also demonstrated that protein expression levels of the hub genes were higher in tumor tissues compared with those in adjacent normal tissues. Furthermore, the expression levels of AMD1, EN1 and VGLL1 were strongly correlated with each other. These results demonstrated that AMD1 is highly expressed in breast cancer tissues and cells and AMD1 knockdown decreased the proliferation and metastatic potential, while increasing apoptosis of breast cancer cells. These results suggested that AMD1, EN1 and VGLL1 are likely to contribute to breast cancer progression and unfavorable prognosis.

Introduction

According to Global Cancer Statistics 2018, newly diagnosed cases of breast cancer account for ~25% of all cancers in females. Female breast cancer is the most frequently diagnosed malignancy worldwide (in 154 of 185 countries) and is the primary cause of cancer-associated death in over 100 countries (1). Marusyk and Polyak (2) revealed that breast cancer is a heterogeneous disease on a clinicopathological, cellular and molecular level. According to a molecular study, breast cancer may be classified into ≥4 subtypes, including luminal, human epidermal growth factor receptor 2 (HER2)-enriched, basal-like and normal-like breast cancer (3). Among these subtypes, basal-like breast cancer is the most highly malignant type, accounting for 75% of triple-negative breast cancers (TNBCs) that lack estrogen receptor (ER), progesterone receptor and HER2 expression (4). TNBC is a highly aggressive and heterogeneous disease with an earlier age of onset and greater metastatic potential than non-TNBC (5). Evidence suggests that patients with non-TNBC have improved survival
rates compared with those with TNBC and that these patients benefit from targeted therapy. Due to a lack of available targeted therapies, chemotherapy is currently the standard treatment for TNBC (6,7). However, patients frequently experience drug resistance, which results in tumor recurrence and disease progression (8). Therefore, it is critical to identify novel potential therapeutic targets for breast cancer, particularly TNBC.

Weighted gene co-expression network analysis (WGCNA) uses systems biology to identify modules of highly related genes and associate these modules with clinical traits. Therefore, WGCNA is widely used to identify and screen for biomarkers (9), and has been successfully used to discover therapeutic targets for a variety of cancer types, including, but not limited to, laryngeal cancer (10), leiomyosarcoma (11) and advanced gastric cancer (12).

In the present study, the breast cancer microarray dataset GSE76275 was downloaded from the Gene Expression Omnibus (GEO) database and WGCNA was used to select target gene candidates, which were then validated using alternative datasets and in vitro experimentation.

Materials and methods

Data preprocessing. Gene expression data and clinical information from patients with breast cancer were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The gene expression profiles included GSE76275 (13) and GSE42568 (14) (platform, GPL570 (HG-U133_Plus_2); Affymetrix Human Genome U133 Plus 2.0 Array; Thermo Fisher Scientific, Inc.), as well as GSE25055 (15) and GSE25065 (15) (platform, GPL96 (HG-U133A); Affymetrix Human Genome U133A Array; Thermo Fisher Scientific, Inc.). The inclusion criteria were as follows: i) Patients with a diagnosis of breast cancer; ii) in GSE76275, patients with complete clinical data, including age, tumor stage, tumor size, lymph node status, metastasis and tumor grade; iii) in GSE25055, patients with complete clinical data on tumor size, lymph node status, tumor stage, breast cancer subtype, status (dead or alive) and specific follow-up time; iv) in GSE25065, patients with complete clinical data on tumor size, lymph node status, tumor stage, tumor grade and breast cancer subtype; and v) in GSE42568, patients with complete clinical information on tumor size, lymph node status, tumor grade and ER status.

Weighted gene co-expression network construction. The top 25% most variable genes in GSE76275 were selected for further analysis using the WGCNA package (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/; version 1.69). First, the samples were clustered to construct the sampleTree and detected outliers were selected based on cut height. ‘Sample dendrogram’ and ‘trait heatmap’ were used to develop each network in order to investigate the relationship between the corresponding sample gene expression data and clinical phenotypes. The value of the soft-thresholding parameter used to construct the adjacency matrix was set as $\beta=6$. Furthermore, the adjacency matrix was transformed into the topological overlap matrix (TOM). According to the TOM-based dissimilarity measure, genes with absolute correlation values were clustered into the same module to generate a cluster dendrogram (deep-split, 2; minimum cluster size, 30; cut height, 0.25). In an effort to visually represent the relationships between modules and the clinical features of breast cancer, Pearson's correlation coefficient was calculated and plotted in a heatmap. Modules were determined to have a significant correlation to clinical traits when $P<0.05$. The highest correlating module was selected as the key module for further analysis.

Identification and validation of hub genes. In the present study, hub genes were screened out based on the cut-off criteria of module membership (MM), gene significance (GS) and survival analysis. MM was defined as the Pearson's correlation coefficient between each gene in the key module and the module eigengene, where MM reflects the module connectivity of each gene. GS was defined as the correlation coefficient between each gene in the key module and its corresponding clinical trait, where GS represents the biological significance of each gene. MM>0.7 and GS>0.2 were set as cut-off criteria to screen genes in the key module with high functional significance. Gene expression and clinical prognostic information (vital status and follow-up time) from patients with breast cancer were integrated based on GSE25055. The survival package in R (https://github.com/therneau/survival; version 3.2-7) was used to perform survival analysis. In order to assess the prognostic value of these genes in patients with breast cancer, the Kaplan-Meier plotter (http://kmplot.com/) database was used to generate relapse-free survival (RFS) and overall survival (OS) curves. Genes that were indicated to be associated with RFS through both methods were designated as hub genes for deeper validation. In order to validate their reliability, the expression of the hub genes in relation to clinicopathological characteristics (such as pathological T stage, pathological N stage, tumor stage, tumor grade and breast cancer subtype) were analyzed based on the GSE25055, GSE25065 and GSE42568 datasets. The R packages 'ggplot' (https://ggplot2.tidyverse.org/; version 3.3.0), ‘ggpubr’ (https://rpkg.s.datanovia.com/ggpubr/; version 0.2.5) and 'ggseed' (https://github.com/const-ac/ggseed; version 0.6.0) were used to perform correlation analyses between gene expression and clinical traits. To verify the protein expression levels of the hub genes in breast cancer and normal tissues, immunohistochemistry (IHC) data were downloaded from the Human Protein Atlas (HPA; http://www.proteinatlas.org). The R package ‘corrplot’ (https://github.com/taiyun/corrplot; version 0.84) was used to assess the correlation between the expression levels of each hub gene.

Gene set enrichment analysis (GSEA). GSEA was used to predict the potential function of each hub gene. For each hub gene, a total of 267 breast cancer samples in the GSE25055 dataset were divided into high-risk and low-risk groups; c2.cp.kegg.v7.1.symbols.gmt was selected as the reference gene set. The number of permutations was set at 1,000 times for each analysis. Nominal $P<0.05$, false discovery rate $<25$% and gene size $\geq50$ were selected as the thresholds.

IHC. Breast cancer and adjacent normal tissues were collected from patients (age range, 52-67 years; mean age, 60 years) undergoing mastectomy and with a postoperative pathology diagnosis of breast cancer at Zhongnan Hospital (Wuhan,
China) between May and September 2019. Written informed consent was obtained from each patient prior to surgery and the patient protocols were approved by the hospital's ethics committee (approval no. 2015073). IHC was used to detect the expression levels of S-adenosylmethionine decarboxylase proenzyme (AMD1) in both sets of tissues. The tissue samples were excised, fixed with formalin, dehydrated and embedded in paraffin, and subsequently cut into 3-μm sections. For IHC, the sections were incubated with primary antibodies against AMD1 (1:500 dilution; cat. no. ab65820; Abcam) at 4°C overnight. After washing three times with PBS, the sections were incubated with an HRP-conjugated secondary antibody (1:400 dilution; cat. no. AS061; ABeclonal Biotech Co., Ltd.) at room temperature for 1.5 h. After further washing, the peroxidase activity was visualized using freshly prepared diaminobenzidine (OriGene Technologies, Inc.) and the slides were then lightly counterstained with Harris' hematoxylin. The negative controls were processed in the same way, but with PBS in place of the primary antibody. Finally, the slides were observed under a light microscope (Nikon, Inc.; x200).

Cell culture. Human breast cancer cell lines (MCF-7, MDA-MB-231, MDA-MB-468 and MDA-MB-157) and a mammary epithelial cell line (MCF-10A) were purchased from the Cell Bank of the Chinese Academy of Sciences. The cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco; Thermo Fisher Scientific, Inc.) supplemented with PBS in place of the primary antibody. The tissue samples were excised, fixed with formalin, dehydrated and embedded in paraffin, and subsequently cut into 3-μm sections. For IHC, the sections were incubated with primary antibodies against AMD1 (1:500 dilution; cat. no. ab65820; Abcam) at 4°C overnight. After washing three times with PBS, the sections were incubated with an HRP-conjugated secondary antibody (1:400 dilution; cat. no. AS061; ABeclonal Biotech Co., Ltd.) at room temperature for 1.5 h. After further washing, the peroxidase activity was visualized using freshly prepared diaminobenzidine (OriGene Technologies, Inc.) and the slides were then lightly counterstained with Harris' hematoxylin. The negative controls were processed in the same way, but with PBS in place of the primary antibody. Finally, the slides were observed under a light microscope (Nikon, Inc.; x200).

Reverse transcription-quantitative (RT-q)PCR. Total RNA was extracted from each cell type using TRIzol® reagent (Invitrogen; Thermo Fisher Scientific, Inc.). The miScript Reverse Transcription kit (Qiagen GmbH) was used according to the manufacturer's instructions for RT with 3 μg total RNA. qPCR was performed using the SYBR®-Green Master Mix (Takara Biotechnology Co., Ltd.) supplemented with 10% fetal bovine serum (FBS; Gibco; Thermo Fisher Scientific, Inc.) at 37°C with 5% CO₂.

Migration and invasion assays. Cellular invasion and migration assays were performed using Transwell inserts (Corning, Inc.) coated with or without Matrigel, respectively. MDA-MB-231 cells (2×10⁵ cells/ml) were seeded into the upper chamber, and DMEM containing 20% FBS was added to the lower chamber. After culturing at 37°C for 24 h, cells that had migrated to the lower chambers were fixed with methanol, stained with 0.5% crystal violet and counted in three randomly selected fields using ImageJ software (version 1.53; National Institutes of Health).

Apoptosis analysis. Propidium Iodide/Annexin V-APC staining and flow cytometric analysis were performed to estimate the apoptotic rates of breast cancer cells. The Annexin V/PI Cell Apoptosis kit (Sungene Biotech Co., Ltd.) was used to detect apoptotic cells according to the manufacturer's protocol. Flow cytometric analysis was performed using a BD Accuri C6 flow cytometer (BD Biosciences) and results were evaluated with FlowJo software (version 7.6.1; FlowJo LLC).

Statistical analysis. The data were analyzed using R (version 3.6.3) and GraphPad Prism 7 software (GraphPad Software, Inc.). All experiments were performed in triplicate and the data are expressed as the mean ± standard error of the mean. Differences between two groups were assessed using unpaired Student's t-test and those among multiple groups were assessed by one-way ANOVA and Bonferroni's post-hoc test. A two-tailed P<0.05 was considered to indicate a statistically significant difference.

Results

Clinicopathological characteristics of patients with breast cancer. The present study included 120 patients with breast cancer from the GSE76275 dataset, 260 from the GSE25055 dataset, 183 from the GSE25065 dataset and 101 from the GSE42568 dataset, all with complete clinicopathological data for the following analyses. Detailed clinicopathological information for each cohort is displayed in Table I.

Weighted co-expression network construction and key module identification. The top 25% most variable genes (n=4,052) were selected for co-expression analysis using the WGCNA package. A sample dendrogram and trait heatmap were used to split the selected samples into the appropriate clusters; the distribution map of clinical trait data is provided in Fig. 1A. When the power was equal to 6, the R² scale was equal to 0.9 (Fig. 1B and C). Therefore, β=6 was selected as the soft threshold for breast cancer co-expression analysis. The 15 original co-expression modules were obtained using the dynamic tree cut method. After setting the cut height to 0.25, thereby merging highly similar modules (Fig. 2A), 12 modules
were screened out (Fig. 2B). The association between the
modules and clinical traits was then analyzed, allowing for
the selection of key modules for further investigation. The
red module, which contained 273 genes, was identified as the
key module. The heatmap and histogram indicated that the
red module was positively correlated with the tumor grade.
In a scatter plot of GS vs. MM, a significant correlation was
evident in the red module. The plot revealed that MM in the
red module demonstrated a significant correlation with the
tumor grade (r=0.62, P=2.2x10^{-30}) (Fig. 3A-C).

Hub gene screening and validation. Hub genes were screened
out using the red module. MM>0.7 and GS>0.2 were set as
the cut-off criteria to screen 36 genes with high functional
significance. Among them, 9 (prominin 1, γ-butyrobetaine
hydroxylase 1, BAF chromatin remodeling complex subunit
BCL11A, AMD1, rhophilin associated tail protein 1B,
vestigial-like protein (VGLL1), tripartite motif containing 2,
homeobox protein engrailed-1 (EN1) and keratin 6B) and
6 genes (Kruppel like factor 5, AMD1, EN1, desmocollin 2,
VGLL1 and allograft inflammatory factor 1 like) were
negatively correlated with RFS of patients with breast
cancer, based on the validation dataset GSE25055 and the
Kaplan-Meier plotter tool, respectively. AMD1, EN1 and
VGLL1 were associated with poor RFS in both analyses
and only VGLL1 was associated with a worse OS prognosis
(according to Kaplan-Meier survival curves; Fig. 4A-I).
Consequently, AMD1, EN1 and VGLL1 were identified as hub
genes. Upon WGCNA of the dataset GSE76275, the red module
was determined to be highly associated with tumor grade.

Table I. Clinicopathological characteristics of breast cancer patients in the different datasets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GSE76275</th>
<th>GSE25055</th>
<th>GSE25065</th>
<th>GSE42568</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;65</td>
<td>88 (73.3)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>≥65</td>
<td>32 (26.7)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Tumor size (cm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤2</td>
<td>26 (21.7)</td>
<td>19 (7.3)</td>
<td>10 (5.5)</td>
<td>18 (17.8)</td>
</tr>
<tr>
<td>2-5</td>
<td>79 (65.8)</td>
<td>145 (55.8)</td>
<td>87 (47.5)</td>
<td>80 (79.2)</td>
</tr>
<tr>
<td>&gt;5</td>
<td>9 (7.5)</td>
<td>56 (21.5)</td>
<td>67 (36.6)</td>
<td>3 (3.0)</td>
</tr>
<tr>
<td>Any size with direct extension</td>
<td>6 (5.0)</td>
<td>40 (15.4)</td>
<td>19 (10.4)</td>
<td>NA</td>
</tr>
<tr>
<td>Metastatic lymph nodes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>57 (47.5)</td>
<td>76 (29.2)</td>
<td>66 (36.1)</td>
<td>44 (43.6)</td>
</tr>
<tr>
<td>Positive</td>
<td>63 (52.5)</td>
<td>184 (70.1)</td>
<td>117 (63.9)</td>
<td>57 (56.4)</td>
</tr>
<tr>
<td>Distant metastasis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>118 (98.3)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Positive</td>
<td>2 (1.7)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Tumor stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>17 (14.2)</td>
<td>5 (1.9)</td>
<td>2 (1.1)</td>
<td>NA</td>
</tr>
<tr>
<td>II</td>
<td>72 (60.0)</td>
<td>143 (55.0)</td>
<td>104 (56.8)</td>
<td>NA</td>
</tr>
<tr>
<td>III</td>
<td>29 (24.1)</td>
<td>112 (43.1)</td>
<td>77 (42.1)</td>
<td>NA</td>
</tr>
<tr>
<td>IV</td>
<td>2 (1.7)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Tumor grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>2 (1.7)</td>
<td>13 (5.0)</td>
<td>13 (7.1)</td>
<td>18 (17.8)</td>
</tr>
<tr>
<td>Moderate</td>
<td>43 (35.8)</td>
<td>108 (41.5)</td>
<td>63 (34.4)</td>
<td>80 (79.2)</td>
</tr>
<tr>
<td>Poor</td>
<td>75 (62.5)</td>
<td>139 (53.5)</td>
<td>107 (58.5)</td>
<td>3 (3.0)</td>
</tr>
<tr>
<td>Breast cancer subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basal</td>
<td>120 (100)</td>
<td>109 (41.9)</td>
<td>64 (35.0)</td>
<td>NA</td>
</tr>
<tr>
<td>Her2</td>
<td>NA</td>
<td>19 (7.3)</td>
<td>12 (6.6)</td>
<td>NA</td>
</tr>
<tr>
<td>Luminal</td>
<td>NA</td>
<td>132 (50.8)</td>
<td>88 (48.0)</td>
<td>NA</td>
</tr>
<tr>
<td>Normal</td>
<td>NA</td>
<td>NA</td>
<td>19 (10.4)</td>
<td>NA</td>
</tr>
<tr>
<td>Survival status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>NA</td>
<td>205 (78.8)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Dead</td>
<td>NA</td>
<td>55 (21.2)</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Values are expressed as n (%); NA means relevant information is not available. Her2, human epidermal growth factor receptor 2.
Figure 1. Clustering dendrogram and determination of soft-thresholding power for the weighted gene co-expression network analysis. (A) Sample dendrogram and trait heatmap based on gene expression data and clinical data. (B) Soft-thresholding power analysis of scale independence and mean connectivity. (C) Checking scale free topology for $\beta=6$.

Figure 2. (A) Cluster dendrogram of MEs. The value corresponding to the red line indicates the merge threshold. The clustering height is the value of the criterion associated with the clustering method for the particular agglomeration. (B) Clustering dendrogram of genes by hierarchical clustering based on the dissimilarity topological overlap matrix. ME, module eigengene.
Subsequently, three validation datasets were used to determine the relationship between tumor grade and the expression levels of AMD1, EN1 and VGLL1. The results revealed that higher hub gene expression levels were associated with advanced tumor grade in dataset GSE25055 (Fig. 5A-C). From dataset GSE25055, the expression levels of these genes were also determined to be increased in basal breast tumors compared to luminal and HER2-enriched breast tumors (Fig. 5D-F). However, with regard to tumor size, lymph node status and tumor stage, no significant association was observed between these clinicopathological parameters and the expression levels of the hub genes (Fig. S1). In datasets GSE25065 and GSE42568, hub gene upregulation also corresponded with advanced tumor grade and a more malignant cancer subtype (Figs. S2 and S3). IHC data from the HPA online database also demonstrated that the protein levels of AMD1 and EN1 were higher in tumor tissues than in normal tissues (Fig. 6A-F) and that the expression of each individual protein was strongly

Figure 3. Identification of modules associated with the clinical traits of patients with breast cancer. (A) Heatmap of the correlation between MEs and the clinical traits of patients with breast cancer. The top row is the R value, the bottom row is the P-value. (B) Scatter plot of module membership vs. gene significance for tumor grade in the red module; correlation coefficient r=0.62 and P=2.2x10^{-30}. (C) Distribution of average gene significance and standard errors in modules associated with breast cancer tumor grade. ME, module eigengene.
correlated with that of the other two (in both the GSE76275 and GSE25055 datasets) (Fig. 7).

GSEA. To investigate potential signaling pathways associated with the three hub genes, GSEA was used to identify Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched in breast cancer samples with high AMD1, EN1 and VGLL1 expression. Based on the cut-off criteria, the top 5 KEGG pathways enriched in the samples with high AMD1, EN1 and VGLL1 expression are displayed in Fig. 8.

These hub genes were commonly enriched in ‘cell cycle’, ‘oocyte meiosis’, ‘pathogenic Escherichia coli infection’ and ‘pyrimidine metabolism’.

**AMD1 knockdown inhibits proliferation and metastatic capacity, while promoting apoptosis in breast cancer cells.**

A series of additional experiments were performed to further investigate the expression levels and functions of AMD1 in breast cancer. Based on the IHC results, AMD1 was indicated to be upregulated in breast cancer tissues (Fig. 9A), which was
also consistent with the results from the HPA. Compared with that in MCF-10A cells, AMD1 expression was significantly upregulated in all breast cancer cell lines, but most notably in MDA-MB-231 cells (Fig. 9B). Therefore, the MDA-MB-231 cell line was selected for further analyses. The results of the CCK-8 assay indicated that AMD1-knockdown significantly inhibited the proliferation of MDA-MB-231 cells (Fig. 9C). The number of migratory and invasive MDA-MB-231 cells transfected with si-AMD1 was also significantly reduced compared with those of the NC-transfected group (Fig. 9D and E). Furthermore, flow cytometric analysis demonstrated that the apoptotic rate of MDA-MB-231 cells was significantly increased in the si-AMD1 group as compared with that in the si-NC group (Fig. 9F).

Discussion

In the present study, the top 25% most variable genes in the GSE76275 dataset were used for co-expression analysis, from which 12 modules were identified. Among these modules, the red module was highly correlated with tumor grade. Using survival analysis, AMD1, EN1 and VGLL1 were subsequently identified as hub genes within the red module and their upregulation was associated with a poorer prognosis in patients with breast cancer in both the validation dataset GSE25055 and Kaplan-Meier plotter. The expression levels of hub genes were further validated and were indicated to be highly expressed in samples with advanced tumor grade and basal-like breast cancer. IHC staining demonstrated that the protein levels of AMD1 and EN1 were higher in breast cancer tissues than in normal tissues. In addition, the expression levels of these genes were strongly correlated with each other. According to the GSEA, the hub genes were confirmed to be commonly enriched in ‘cell cycle’, ‘oocyte meiosis’, ‘pathogenic Escherichia coli infection’ and ‘pyrimidine metabolism’. Finally, in vitro experiments were used to validate the expression and function of AMD1. As relevant experimental studies of the effect of EN1 and VGLL1 on breast cancer progression have previously been published (17,18), in vitro experimentation was not performed for these genes. The present results suggested that AMD1 is upregulated in breast cancer tissues and cells, and that AMD1 knockdown decreased the proliferation, invasion and migration abilities, whilst increasing apoptosis in breast cancer cells.

AMD1 encodes an important enzyme involved in polyamine biosynthesis, in which various aliphatic amine-associated polyamines are essential for promoting cellular proliferation and tumorigenesis. AMD1 has been demonstrated to promote epidermal wound healing by regulating cellular migration (19), and has been reported to have a significant role in the pathogenesis of multiple tumor types, such as prostate (20,21), non-small cell lung (22) and gastric cancer (23). A study revealed that mammalian target of rapamycin complex 1 regulates AMD1 to sustain polyamine metabolism in prostate cancer (20). AMD1 was also indicated to be upregulated in gastric cancer samples and patients with high AMD1 expression levels exhibited
Figure 6. Immunohistochemical staining of the three hub genes from the Human Protein Atlas. (A) Protein levels of AMD1 in normal tissue (staining, low; intensity, weak; quantity, 75-25%). (B) Protein levels of AMD1 in tumor tissue (staining, medium; intensity, moderate; quantity, >75%). (C) Protein levels of EN1 in normal tissue (staining, not detected; intensity, negative; quantity, none). (D) Protein levels of EN1 in tumor tissue (staining, low; intensity, weak; quantity, >75%). (E) Protein levels of VGLL1 in normal tissue (staining, not detected; intensity, negative; quantity, none). (F) Protein levels of VGLL1 in tumor tissue (staining, not detected; intensity, negative; quantity, none). AMD1, S-adenosylmethionine decarboxylase proenzyme; EN1, homeobox protein engrailed-1; VGLL1, vestigial-like protein 1.

Figure 7. Correlations between hub genes in the (A) GSE76275 and (B) GSE25055 datasets. The numbers in the circles are the R values. AMD1, S-adenosylmethionine decarboxylase proenzyme; EN1, homeobox protein engrailed-1; VGLL1, vestigial-like protein 1.
Figure 8. Gene Set Enrichment Analysis. Top 5 enriched pathways in samples with high (A) AMD1, (B) EN1 and (C) VGLL1 expression. KEGG, Kyoto Encyclopedia of Genes and Genomes; AMD1, S-adenosylmethionine decarboxylase proenzyme; EN1, homeobox protein engrailed-1; VGLL1, vestigial-like protein 1.
poorer OS rates; furthermore, inhibiting AMD1 suppressed cellular proliferation and migration in vitro, as well as tumor growth in vivo (23). The results of the present study provided results on the carcinogenic effects of AMD1 in breast cancer, which are consistent with its effect on the progression of gastric cancer.

The EN1 gene encodes a homeodomain-containing protein that regulates pattern formation during central
nervous system development. EN1 expression was reported to be significantly higher in TNBC than in other breast cancer subtypes (17). Studies have indicated that upregulation of EN1 is correlated with significantly shorter OS times and increased risk of brain metastases in patients with TNBC (24), and that EN1 protein expression is increased in adenoïd cystic carcinoma, with the higher expression of EN1 being associated with a lower survival rate (25). EN1 was also specifically expressed in normal eccrine glands and focally expressed in skin tumors and sweat gland neoplasms (26). The results of the present study are consistent with those of previous studies on the carcinogenicity of EN1.

The VGLL1 gene encodes a transcriptional co-activator involved in regulating the Hippo pathway in Drosophila (27). A study revealed that VGLL1 is predominantly expressed in BRCA1-associated TNBC and serves an oncogenic role in breast cancer (18). VGLL1 is also reportedly involved in human papillomavirus (HPV) gene expression via transcriptional enhancer factor 1, and thus, is crucial to the growth of HPV-associated malignancies such as cervical cancer (28). In gastric cancer, VGLL1 promoted cancer cell proliferation and metastasis, which was regulated by PI3K/AKT/β-catenin signaling (29). Furthermore, VGLL1 has been indicated to possess oncogenic functions in pediatric neuroepithelial neoplasms (30).

In conclusion, the present study aimed to identify hub genes involved in the pathogenesis of breast cancer using WGCNA. The results indicated that the upregulation of AMD1, EN1 and VGLL1 are correlated, and potentially detrimentally associated, with progression and prognosis in breast cancer. Therefore, inhibiting the expression of AMD1, EN1 and VGLL1 may be a potential therapeutic strategy for breast cancer.

Acknowledgements

Not applicable.

Funding

The present study was supported by the National Natural Science Foundation of China (grant nos. 81770283 and 82070302) and the Clinical Medical Research Center of Peritoneal Cancer of Wuhan (grant no. 2015060911020462). The funding agencies did not participate in the design of the study, the collection, analysis and interpretation of data or manuscript writing.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

LY and WX conceived the study. XL, TY and HW performed data mining, acquisition and analysis. LS collected tissue samples. MF and YL performed the experiments. YL and XL drafted the manuscript and confirm the authenticity of all the raw data. WX and MF revised the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Written informed consent was obtained from each patient prior to surgery and the patient protocols were approved by the ethics committee of the Zhongnan Hospital of Wuhan University (approval no. 2015073).

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.