



## Original Article

# An Ensemble Learning-based Framework from Multicohort Reveals the Molecular Mechanism of GREM1 in Endometrial Cancer



Yihao Zhu<sup>1,2</sup>, Hui Wang<sup>3\*</sup> and Yao Zu<sup>1,2\*</sup>

<sup>1</sup>International Research Center for Marine Biosciences, Shanghai Ocean University, Shanghai, China; <sup>2</sup>Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Shanghai Ocean University, Shanghai, China; <sup>3</sup>Laboratory of Tumor Molecular Biology, School of Basic Medical Sciences, Shanghai University of Medicine and Health Sciences, Shanghai, China

Received: August 16, 2023 | Revised: September 29, 2023 | Accepted: October 31, 2023 | Published online: November 30, 2023

## Abstract

**Background and objectives:** Gremlin-1 (GREM1) was recently reported to maintain the cellular heterogeneity of pancreatic cancer. However, the function of GREM1 in endometrial cancer remains elusive. The purpose of this study was to investigate the underlying mechanisms of GREM1 in endometrial cancer using an ensemble learning-based framework.

**Methods:** The training and test cohorts were gathered from The Cancer Genome Atlas and Genome Expression Omnibus databases, respectively. The cohorts were divided into low-GREM1 and high-GREM1 groups. Differentially expressed gene analysis, weighted gene co-expression network analysis, and Mfuzz clustering were implemented in the training cohort to screen genes with GREM1. The genes were subjected to machine learning-based integration for selecting key genes with GREM1. Together with the Bayesian network inference and Kyoto Encyclopedia of Genes and Genome enrichment analysis on key genes and GREM1, the potential pathway of GREM1 in endometrial cancer was illustrated. Leveraging the CIBERSORT analysis tool and single sample gene set enrichment analysis, the immune landscape of endometrial cancer was investigated to identify the immune cells with GREM1 and key genes.

**Results:** A set of 10 key genes (*FAP*, *THBS1*, *POSTN*, *INHBA*, *ASPN*, *COL3A1*, *IGFBP5*, *COL8A1*, *FN1*, and *COL5A2*) highly linked to GREM1 were obtained. Moreover, GREM1 may regulate extracellular matrix-related pathways in endometrial cancer, affecting extracellular matrix degradation involving collagen-related key genes. Finally, we found increased infiltration of mast cells in the high-GREM1 group, accompanied by their positive correlations.

**Conclusions:** GREM1 regulated extracellular matrix modulation in endometrial cancer by interacting with key genes, with mast cells serving as a signature.

## Introduction

Global cancer statistics released in 2020 indicated that endometrial

cancer has become the sixth most common gynecological tumor, with an estimated 417,367 new cases and 97,370 deaths worldwide.<sup>1</sup> The American Cancer Society also reported an increasing incidence and mortality trend for endometrial cancer in recent years, which underscores the urgent demand to develop more effective strategies for diagnosis and treatment.<sup>2</sup> As endometrial cancer is frequently symptomatic and detectable at an early stage, such as abnormal vaginal bleeding, most cases are diagnosed at Stage I and have favorable outcomes after prompt surgically curable.<sup>3,4</sup> However, worse clinical outcomes often occurred in patients with advanced-stage, recurrence, or metastatic endometrial cancers, partly direct to the abnormal molecular features of the highly aggressive tumor cells gained in the disease progress.<sup>5,6</sup>

Gremlin-1 (GREM1), a bone morphogenic protein (BMP) an-

**Keywords:** Gremlin-1; Endometrial cancer; Bioinformatics analysis; Machine learning; Extracellular matrix.

\***Correspondence to:** Yao Zu, International Research Center for Marine Biosciences, Shanghai Ocean University, Shanghai 200090, China. ORCID: <https://orcid.org/0000-0001-8081-9136>. Tel: 86-021-61900405, Fax: 86-021-6190040, E-mail: [yzu@shou.edu.cn](mailto:yzu@shou.edu.cn); Hui Wang, Laboratory of Tumor Molecular Biology, School of Basic Medical Sciences, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China. ORCID: <https://orcid.org/0000-0002-7117-7140>. Tel: +86-21-33759000, Fax: +86-21-33759000, E-mail: [wangh\\_19@sumhs.edu.cn](mailto:wangh_19@sumhs.edu.cn)

**How to cite this article:** Zhu Y, Wang H, Zu Y. An Ensemble Learning-based Framework from Multicohort Reveals the Molecular Mechanism of GREM1 in Endometrial Cancer. *Gene Expr* 2024;23(3):167–185. doi: 10.14218/GE.2023.00095.

tagonist, has been known to induce carcinogenesis through multiple mechanisms, such as BMP dependent pathway or vascular endothelial growth factor receptor (VEGFR) signaling pathway.<sup>7</sup> Previous studies have indicated elevated expression levels and carcinogenic roles of GREM1 in various malignancies, such as carcinomas of lung, ovary, kidney.<sup>8</sup> Recently, GREM1 was identified as a critical regulator of cellular heterogeneity in pancreatic cancer whose over-expression could restrict epithelial-to-mesenchymal transition by interacting with BMP, inhibiting tumor metastasis.<sup>9</sup> Cheng *et al.* showed that GREM1 promotes lineage plasticity and castration resistance by targeting fibroblast growth factor receptor 1, which is a candidate therapeutic target for prostate cancer.<sup>10</sup> These inspiring findings provide a new perspective for understanding the function of GREM1 in cancer progression and treatment. Nevertheless, little evidence has illustrated the certain underlying role of GREM1 in endometrial cancer. Accordingly, more efforts should be urgently dedicated to elucidating the mechanism of GREM1 in endometrial cancer.

Herein, we conducted a comprehensive bioinformatics workflow, accompanied by machine learning-based integration, to explore the underlying mechanism of GREM1 in endometrial cancer. First, the gene expression profiles of endometrial cancer were obtained through The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, which were used as the training cohort and test cohort respectively. The mRNA and protein expression patterns of GREM1 were then explored at the pan-cancer level, including endometrial cancer. Second, we classified the training cohort into low-GREM1 and high-GREM1 groups and identified the differentially expressed genes (DEGs) between the two groups. The Weighted Gene Co-expression Network Analysis (WGCNA) and Mfuzz clustering were conducted to discern gene expression patterns highly related to the high-GREM1 group. Third, we took the intersection of genes associated with high-GREM1 generated by the DEGs analysis, WGCNA, and Mfuzz. Subsequently, these genes were subjected to a machine learning-based integration procedure for further selection.<sup>11</sup> Fourth, we assessed the classification performance of the selected key gene on low- and high-GREM1 groups. We then investigated the potential mechanism of GREM1 through the combination of pathway enrichment analysis and Bayesian network (BN) inference on key genes. To deepen the illustration of the underlying mechanism of GREM1 in immune regulation, we further explored the association of the key genes (GREM1 was also included) with immune-cell infiltration in endometrial cancer. Overall, our study revealed the potential pathways of GREM1 in endometrial cancer from integrative bioinformatics methods, which might expand our understanding of its function in tumor development.

## Material and methods

### Data gathering and processing

The gene expression profiles of endometrial cancer were collected from the TCGA database (<https://portal.gdc.cancer.gov/>) and the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). The training cohort [Cancer Genome Atlas-Uterine Corpus Endometrial Carcinoma (TCGA-UCEC)] was acquired from the TCGA database using TCGA biolinks R package, including 35 normal endometrial tissues and 554 endometrial cancer tissues.<sup>12</sup> The test cohort was obtained from five GEO datasets using the GEOquery R package, including GSE2109, GSE17025, GSE36389, GSE106191, and GSE115810.<sup>13-16</sup> The probes were then transformed to official gene symbols through the platform annotations. The raw count

**Table 1. Sample size of datasets used in this study**

Dataset	Sample size
TCGA-UCEC	589 samples (35 controls; 554 tumors)
GSE2109	17 tumors
GSE17025	91 tumors
GSE36389	20 tumors
GSE106191	64 tumors
GSE115810	24 tumors

TCGA-UCEC, The Cancer Genome Atlas-Uterine Corpus Endometrial Carcinoma.

data from TCGA was converted into transcripts per kilobase million format and subjected to the log<sub>2</sub> transformation process. The raw count data from GEO were processed with the limma and sva R packages for log<sub>2</sub> transformation, normalization, and batch effect removal. The five GEO datasets were merged into the test cohort, which contained 216 endometrial cancer samples. The sample size of the included analysis of GREM1 expression in pan-cancer and endometrial cancer datasets is shown in Table 1.

Leveraging the TCGA, genotype-tissue expression (GTEx), and clinical proteomic tumor analysis consortium (CPTAC) databases, we explored the mRNA and protein expression distribution of GREM1 across extensive cancer types. The datasets used in this part can be found in Table 2. The tumor immune estimation resource (TIMER) online tool (<https://cistrome.shinyapps.io/timer/>) was used to investigate the GREM1 expression pattern in the TCGA database.<sup>17</sup> The R package ggplot2 was used to characterize the GREM1 expression pattern of TCGA target GTEx datasets from The University of California Santa Cruz (UCSC) Genome Browser database (<https://xenabrowser.net/>).<sup>18</sup> Subsequently, the protein expression distributions of GREM1 on the CPTAC database were evaluated by The University of Alabama at Birmingham CANcer (UALCAN) online tool (<http://ualcan.path.uab.edu>).<sup>19</sup> Additionally, the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>) was used to demonstrate the differential protein expression of GREM1 in normal endometrial tissue versus endometrial cancer tissue.

### Analysis of GREM1 expression on prognosis and clinicopathological characteristics in different cancers

Herein, we focused on the prognosis and clinicopathological feature analyses of 12 cancer types: breast invasive carcinoma (BRCA), cholangiocarcinoma, head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma, lung squamous cell carcinoma (LUSC), stomach adenocarcinoma (STAD), glioblastoma multiforme (GBM), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC), which displayed the differential expression of GREM1 compared with the normal controls, in the TCGA database. To gain insight into the effect of GREM1 expression on the prognosis and clinicopathological characteristics of tumor patients, the gene expression profiling interactive analysis tool was first used to perform the prognosis analysis. Kaplan-Meier (K-M) curves for overall survival (OS) were performed to investigate the prognostic differences between the low-GREM1 and high-GREM1 expression groups. Clinicopathological characteristics of 12 cancer types, containing age, tumor grade, T/N/M stage, and tumor stage, were downloaded from the UCSC database. Samples with missing or incomplete clinicopathological information were excluded

**Table 2. Datasets used for pan-cancer analysis**

Tool/database for pan-cancer analysis	Dataset list
TIMER	TCGA-ACC, TCGA-BLCA, TCGA-BRCA, TCGA-CESC, TCGA-CHOL, TCGA-COAD, TCGA-DLBC, TCGA-ESCA, TCGA-GBM, TCGA-HNSC, TCGA-KICH, TCGA-KIRC, TCGA-KIRP, TCGA-LAML, TCGA-LGG, TCGA-LIHC, TCGA-LUAD, TCGA-LUSC, TCGA-MESO, TCGA-OV, TCGA-PAAD, TCGA-PCPG, TCGA-PRAD, TCGA-READ, TCGA-SARC, TCGA-SKCM, TCGA-STAD, TCGA-TGCT, TCGA-THCA, TCGA-THYM, TCGA-UCEC, TCGA-UCS, TCGA-UVM.
UCSC	GBM, GBMLGG, LGG, UCEC, BRCA, CESC, LUAD, ESCA, STES, KIRP, KIPAN, COAD, PRAD, STAD, HNSC, KIRC, LUSC, LIHC, WT, SKCM, BLCA, THCA, READ, OV, PAAD, TGCT, UCS, LAML, PCPG, ACC, KICH, CHOL.
UALCAN	Breast cancer, colon cancer, uterine corpus endometrial carcinoma, lung adenocarcinoma, pancreatic adenocarcinoma, head and neck squamous carcinoma, glioblastoma.

ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; DLBC, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; GBMLGG, glioma; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIPAN, Pan-kidney; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; STES, stomach and esophageal carcinoma; TCGA, The Cancer Genome Atlas; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, Thymoma; UALCAN, The University of Alabama at Birmingham CANcer; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UCSC, The University of California Santa Cruz; UVM, uveal melanoma; WT, high-risk Wilms tumor.

from the analysis. Then, we delineated and explored the GREM1 expression distributions in different clinicopathological characteristic groups using the ComplexHeatmap and ggpubr R packages.

#### **Identification of DEGs in the low-GREM1 and high-GREM1 expression groups**

The normal endometrial samples of the training cohort were excluded. The endometrial cancer samples were classified into low-GREM1 group and high-GREM1 group according to the medium value of GREM1 expression. Leveraging the limma R package, the differential expression analysis was performed to screen out the DEGs between the low-GREM1 and high-GREM1 groups. The screening criteria for DEGs were set as  $|\log_2 \text{Fold Change}| > 1$  and adjusted  $p < 0.05$ .<sup>20</sup> Then, the pheatmap R package was used to showcase the differential expression pattern of DEGs among the low-GREM1 and high-GREM1 groups. Moreover, we conducted Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis on these screened DEGs, and the enriched pathways with  $q < 0.05$  were considered significant.<sup>20</sup> The CBNplot R package was used to construct a regulatory network of the 10 most enriched KEGG pathways with the Bayesian network and gene expression profile.<sup>21</sup>

#### **Weighted gene co-expression network analysis**

The WGCNA R package was used to generate a gene co-expression network using the expression profile of the training cohort.<sup>22</sup> Sample labels in the cohort were previously grouped into low-GREM1 and high-GREM1 traits. The optimal soft threshold power ( $\beta = 1-20$ ) was initially calculated to determine the scale-free topology of the network. Then, we constructed a weighted adjacency matrix and transformed it into a topological overlap matrix (TOM). Moreover, the dissTOM was obtained for hierarchical clustering. The dynamic tree-cutting method was adopted to identify various modules clustered by gene similarity. Herein, we set minModuleSize to 60 and MEDissThres to 0.3. Subsequently, we related the recognized modules to two traits (low-GREM1 and high-GREM1). Genes in the module that had high relevance with high-GREM1 were retained for subsequent analysis.

#### **Mfuzz clustering analysis**

The Mfuzz R package uses the fuzzy c-means clustering algorithm

to cluster GREM1 expression patterns.<sup>23</sup> We first obtained distinct clustering patterns according to the ascending GREM1 expression. To investigate the relationship of clustering patterns with GREM1, each sample of the expression profile was subjected to single sample Gene Set Enrichment Analysis (ssGSEA) to assign the scores of clustering patterns.<sup>24</sup> We further explored the clustering characteristics between low-GREM1 and high-GREM1 expression. The correlation of clustering patterns and GREM1 was investigated, with the patterns closely related to GREM1 selected for further analysis.

#### **Potential biological functions, pathways, and diseases of the hub genes**

We intersected the genes in the WGCNA module and Mfuzz clustering patterns highly related to GREM1 with DEGs, thus gaining the hub genes linked to GREM1 expression. To explore the potential biological functions, pathways, and associated diseases of hub genes, gene ontology (GO), KEGG, and disease ontology (DO) enrichment analyses were conducted using the clusterProfiler and DOSE R packages.

#### **Key genes generated from the machine learning-based combinations and protein-protein interaction network**

To perform feature selection on hub genes, we integrated 12 machine-learning algorithms and 113 algorithm combinations. The machine learning algorithms for integration included absolute shrinkage and selection operator, ridge, elastic network, stepwise multiple generalized linear model, support vector machine, a generalized linear model with likelihood-based boosting (GLMBoost), linear discriminant analysis, partial least squares regression for generalized linear models, RandomForest, gradient boosting, eXtreme gradient boosting, and NaiveBayes. One algorithm was used to select features and another was used to construct a classification model using the leave-one-out cross-validation framework.<sup>11</sup> For each model integration, area under curve (AUC) scores within the training and test cohort were calculated under cross-validation. The classification model performed with the highest AUC score was considered optimal, thus the selected features for the model establishment were then obtained. Furthermore, the String database was used to construct the feature gene-dominated Protein-Protein Interaction (PPI) network, and the top 10 key genes with

high betweenness with GREM1 in the network were obtained by the Cytoscape software (version 3.9.1).<sup>25,26</sup> We then visualized the chromosomal locations of the 10 key genes. Moreover, we assessed the principal component analysis (PCA) scores, expression distributions in low-GREM1 and high-GREM1 groups, and correlations of key genes within the training and test cohorts.

#### **Validation of key genes for distinguishing from the low-GREM1 and high-GREM1 groups**

To assess whether the screened key genes could accurately distinguish from the low-GREM1 and high-GREM1 groups, the glmnet R package was first used to construct a logistic regression model using key genes within the training cohort. The model was illustrated in a nomogram developed with the rms R package. The ability of the model to classify low-GREM1 and high-GREM1 groups was then systemically assessed by AUC scores, calibration curves, and decision curve analysis (DCA) in the training and test cohorts.<sup>27</sup>

#### **Potential pathways and regulatory networks of key genes with GREM1**

The regulatory relationship between GREM1 and key genes was inferred through the CBNplot R package. To further investigate the key gene-regulated pathways in endometrial cancer, the clusterProfiler R package was used to conduct KEGG enrichment analysis of the key genes. The enriched KEGG pathways with  $q$  values  $< 0.05$  were considered significant differences.<sup>20</sup> Additionally, we depicted the regulatory network of pathways, and the genes enriched in the core pathways with more edges and interaction strengths using the CBNplot R package. To verify the *in-silico* results, the experimental evidence at the protein level was retrieved from the HPA database to show the expression of the key genes in endometrial cancer tissues.

#### **Correlation of key genes with GREM1 with immune infiltration cells**

Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) and ssGSEA were applied to estimate infiltration degrees of different immune cell types, in which CIBERSORT performs deconvolution and ssGSEA marks enriched scores on bulk RNA-seq data using immune-cell signature.<sup>28,29</sup> The permutations of CIBERSORT were set as 1,000 and the calculated results with  $p < 0.05$  were retained for subsequent analysis.<sup>30</sup> We explored the infiltration patterns of immune cells in low-GREM1 and high-GREM1 groups. The relationship between key genes and immune-cell infiltration was then examined by Spearman's correlation approach.

#### **Statistical analysis**

Statistical analysis was performed with R programming language (version 4.2.0). Wilcoxon test was performed to evaluate the mRNA expression of GREM1 in tumor tissue compared with normal tissues. Wilcoxon test was also carried out to investigate the expression distribution of GREM1 in different groups based on clinicopathological characteristics, including age, tumor grade, T/N/M stage, and tumor stage. An unpaired  $t$ -test was used to estimate the differential expression of GREM1 protein associated with key genes, and immune-cell infiltration levels in the two groups. Log-rank test was used to analyze differences in OS between the low-GREM1 and high-GREM1 expression groups. Spearman's correlation method was adopted to analyze the association of GREM1 and Mfuzz clustering patterns and Pearson's method for hub gene co-expression relationships. Spearman's correlation

analysis was performed to show whether there was a close relationship between extracellular matrix (ECM) signatures and the expression of GREM1 in different cancers, and  $p < 0.05$  was considered statistically significant.

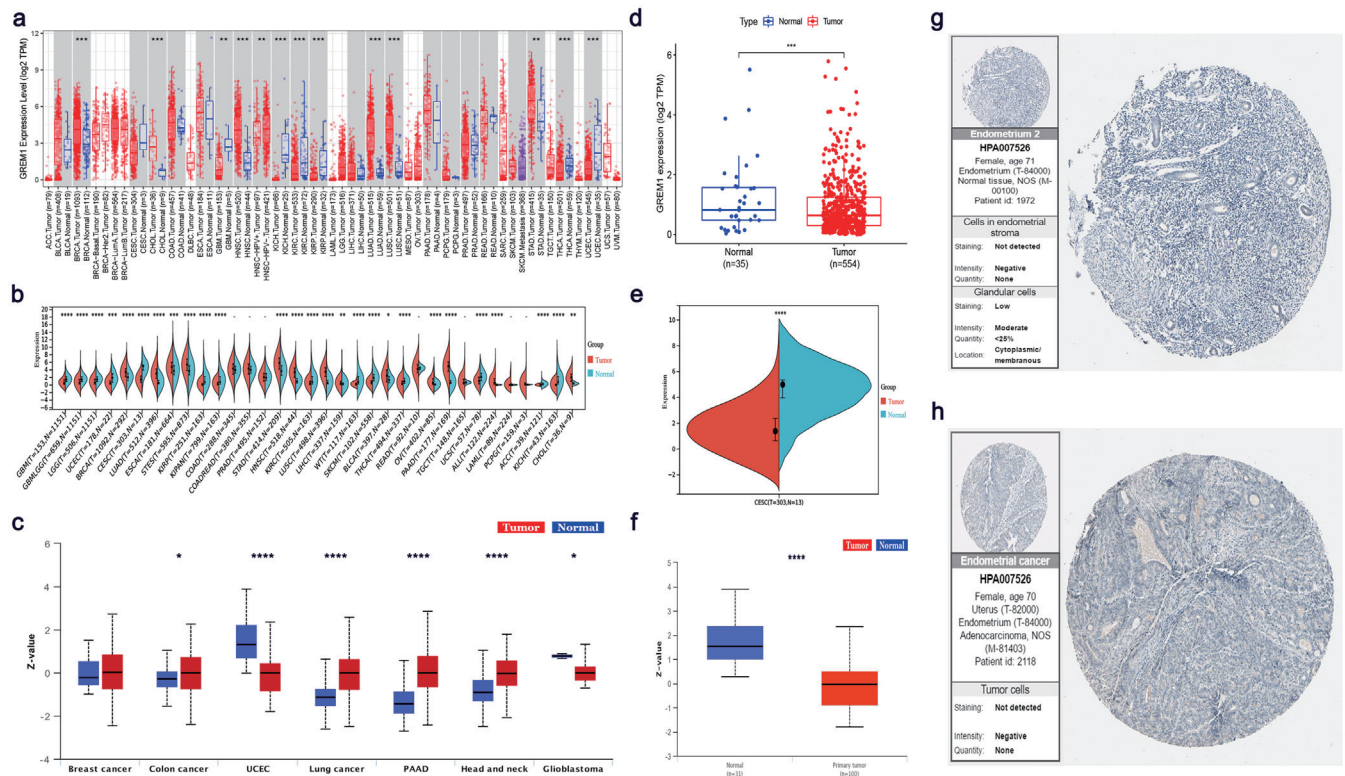
## **Results**

### **GREM1 was downregulated in different types of cancer and endometrial cancer**

First, the expression profiles of GREM1 across different cancer types were investigated via TIMER. Figure 1a shows that lower expression levels of GREM1 were observed in glioblastoma multiforme (GBM), KICH, KIRC, KIRP, THCA, and UCEC (all  $< 0.05$ ). Next, we found that GREM1 was significantly under-expressed in 14 types of cancer compared with normal tissue in the TCGA target GTEx datasets, including GBM, glioma (GBM-LGG), low-grade glioma (LGG), UCEC, cervical squamous cell carcinoma and endocervical adenocarcinoma, KIRP, pan-kidney, KIRC, high-risk Wilms tumor, skin cutaneous melanoma, THCA, uterine carcinosarcoma, adrenocortical carcinoma, and KICH (Fig. 1b). UALCAN was used to explore the protein expression levels of GREM1 in seven cancer types, which showed significant down-regulation of GREM1 in UCEC and glioblastoma (all  $p < 0.05$ ; Fig. 1c). We also observed significant downregulation of GREM1 in endometrial cancer compared with control tissue (all  $p < 0.05$ ; Fig. 1d and f). Immunohistochemical staining patterns of GREM1 in normal endometrial tissue and endometrial cancer tissue were retrieved from the HPA database. The staining scores shown in Figure 1g and h show that GREM1 expression was significantly lower in endometrial cancer than in normal endometrial tissue. These results show that GREM1 was downregulated in endometrial cancer.

### **High expression of GREM is related to poor prognosis and clinicopathological characteristics**

Leveraging the 12 tumors with differential expression patterns compared with normal controls in the TCGA database, we analyzed the effect of GREM1 expression on tumor prognosis and clinicopathological characteristics. To evaluate whether the expression of GREM1 is related to the prognosis of tumor patients, K-M survival curves and log-rank tests compared differences in OS in groups with low-GREM1 and high-GREM1 expression (Fig. 2). As shown in Figure 2e (LUSC), Figure 2i (KIRC), Figure 2j (KIRP), and Figure 2l (UCEC), OS was shorter in groups with high-GREM1 expression than it was in groups with low-GREM1 expression (all  $p < 0.05$ ). These results indicate that the prognostic outcome of tumor patients with high GREM1 expression was poor. Additionally, we investigated the relationship between GREM1 expression and clinicopathological characteristics in 11 tumor types (excluding GBM datasets with insufficient clinicopathological features). Figure 3, shows the expression distribution of GREM1 in groups with diverse clinicopathological characteristics. Interestingly, we observed that the expression of GREM1 was closely linked to at least one characteristic in different tumor types, including BRCA (Fig. 3a), HNSC (Fig. 3c), LUSC (Fig. 3e), STAD (Fig. 3f), KIRC (Fig. 3h), KIRP (Fig. 3i), THCA (Fig. 3j), and UCEC (Fig. 3k). Compared with patients with lower N-stage tumors, those with advanced N-stage had significantly different GREM1 expression levels in BRCA (Fig. 3a), HNSC (Fig. 3c), LUSC (Fig. 3e), KIRC (Fig. 3h), and THCA (Fig. 3j). Additionally, the expression of GREM1 was upregu-



**Fig. 1. Expression pattern of GREM1 at the pan-cancer level.** (a) mRNA expression level of GREM1 in pan-cancer from the TCGA database. (b) mRNA expression level of GREM1 in pan-cancer from the TCGA target GTEx datasets. (c) Protein expression level of GREM1 in pan-cancer from the CPTAC database. (d) Downregulation of GREM1 at mRNA level in endometrial cancer tissues from the TCGA database. (e) Downregulation of GREM1 at mRNA level in endometrial cancer tissues from the TCGA target GTEx datasets. (f) Downregulation of GREM1 at the protein level in endometrial cancer tissue from the CPTAC database. (g) Immunohistochemical staining of GREM1 in normal endometrial tissue from the HPA database. (h) Immunohistochemical staining of GREM1 in endometrial cancer tissue from the HPA database. (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ). ACC, adrenocortical carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CPTAC, Clinical Proteomic Tumor Analysis Consortium; GBM, glioblastoma multiforme; GREM1, Gremlin-1; GTEx, genotype-tissue expression; HPA, Human Protein Atlas; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LGG, low-grade glioma; PAAD, pancreatic adenocarcinoma; TCGA, The Cancer Genome Atlas; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.

lated in advanced T-stage patients with STAD (Fig. 3f), KIRC (Fig. 3h), KIRP (Fig. 3i), and THCA (Fig. 3j). Regarding tumor grading and staging, the over-expression patterns of GREM1 in advanced grade and stages were observed in STAD (Fig. 3f), KIRC (Fig. 3h), KIRP (Fig. 3i), THCA (Fig. 3j), and UCEC (Fig. 3k). It is noteworthy that endometrial cancer young patients with advanced tumor stages had higher GREM1 expression. Taken together, these results show that the high expression of GREM1 is relevant to poor prognosis and worse clinicopathological features of LUSC, KIRC, KIRP, and UCEC, which indicate its pivotal function in tumor development.

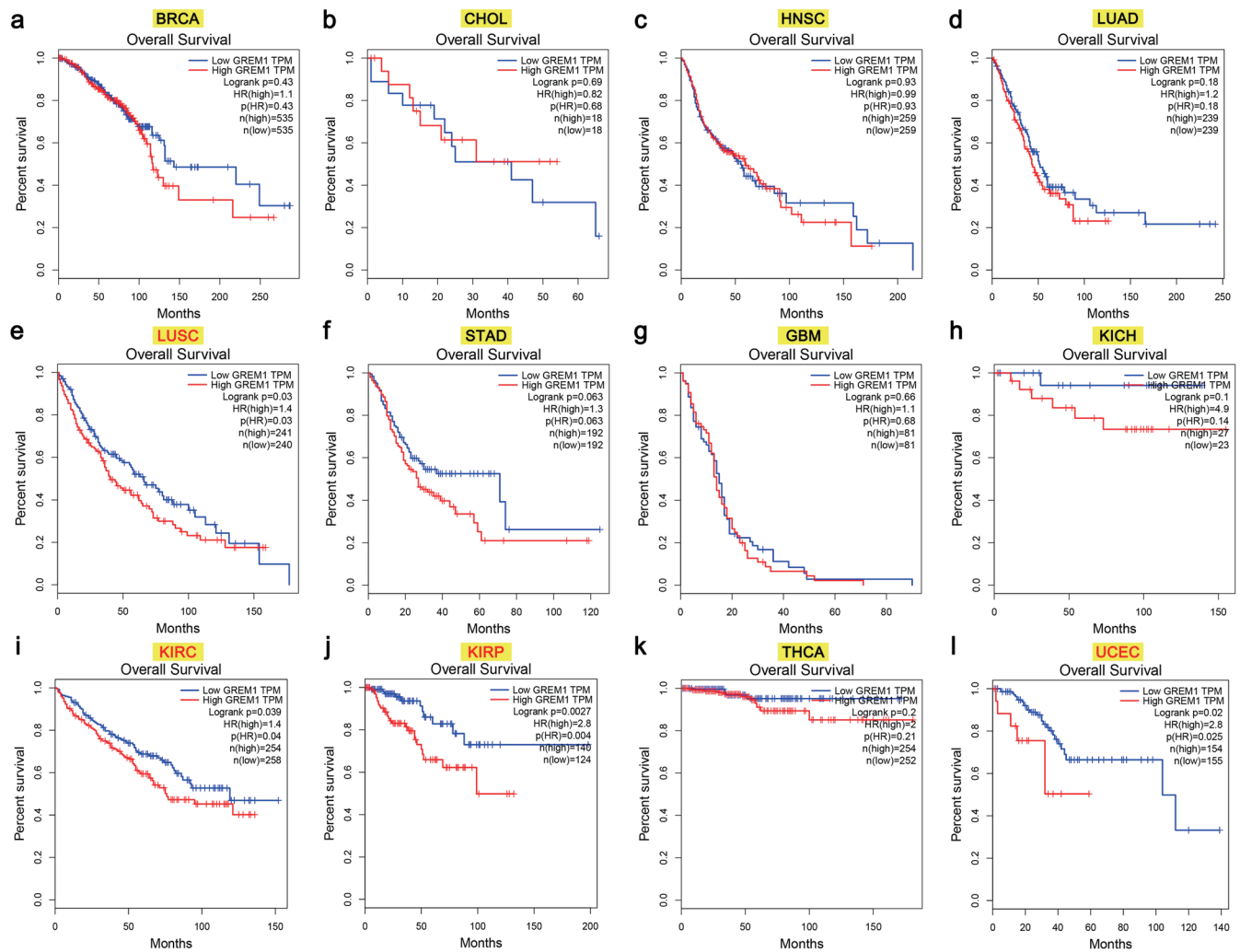
**Identification of DEGs in low and high GREM1 groups**

To identify the high GREM1-related genes, we first divided 554 endometrial cancer samples into low-GREM1 group ( $n = 277$ ) and high-GREM1 group ( $n = 277$ ) according to the medium expression value of GREM1 and performed DEGs analysis. We screened out a total of 65 DEGs, of which 64 genes were up-regulated and one gene was downregulated in the high-GREM1 group (Fig. 4a). The heatmap shows a distinct expression pattern of 65 DEGs between the low-GREM1 and high-GREM1 groups (Fig. 4b). KEGG enrichment analysis showed that upregulated DEGs participated in the organization of the extracellular ma-

trix, integrin cell-surface interactions, ECM proteoglycans, collagen chain trimerization, and degradation of the extracellular matrix (Fig. 4c). We further depicted the interaction network of the top 10 KEGG enriched pathways through the Bayesian inference (Fig. 4d), indicating the core roles of extracellular matrix organization, integrin cell-surface interactions, ECM proteoglycans, and degradation of the extracellular matrix with more edges and gene counts. These results showed that upregulation DEGs may highly cooperate with the extracellular matrix in endometrial cancer.

**Crucial module with GREM1 by WGCNA**

To further identify the high GREM1-related genes, WGCNA was carried out to develop a gene co-expression network and identify genes from the high GREM1-correlated module. First, the optimal soft threshold power ( $\beta = 2$ ) was determined to generate a gene co-expression network (Fig. 5a). Then, seven modules were identified after the dynamic tree cut and merge (Fig. 5b). We then related the traits (low-GREM1 and high-GREM1 groups) to the modules, obtaining a brown module with the highest positive association to the high-GREM1 group ( $R = 0.4$ ,  $p < 0.001$ ; Fig. 5c). A significant positive correlation of gene significance (GS) and module membership (MM) was observed in the ME brown module



**Fig. 2. K-M curves of OS in low-GREM1 and high-GREM1 expression groups in 12 cancer types.** (a) BRCA. (b) CHOL. (c) HNSC. (d) LUAD. (e) LUSC. (f) STAD. (g) GBM. (h) KICH. (i) KIRC. (j) KIRP. (k) THCA. (l) UCEC. Red marks indicate a statistical difference. BRCA, breast invasive carcinoma; CHOL, cholangiocarcinoma; HNSC, head and neck squamous cell carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; STAD, stomach adenocarcinoma; GBM, glioblastoma multiforme; GREM1, Gremlin-1; K-M, Kaplan-Meier; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; OS, overall survival; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.

( $R = 0.51, p < 0.001$ ; Fig. 5d), which suggested the genes from the ME brown module were associated with GREM1 and the most significant elements of the modules correlated with GREM1. Finally, we obtained 1,043 genes from the ME brown module for further analysis.

#### Critical patterns with GREM1 by Mfuzz clustering analysis

Mfuzz clustering analysis was first conducted according to the ascending GREM1 expression levels, combined with ssGSEA marked clustering scores for expression characteristics, to screen out GREM1 a highly positive expression pattern. A total of 50 expression patterns clustering on GREM1 expression were obtained (Fig. 6a and b). We then carried out ssGSEA to classify tumor samples into distinct patterns with the identified Mfuzz clustering results, and each sample was assigned to a certain clustering pattern. Subsequently, Spearman's correlation between distinct clustering patterns and GREM1 was investigated (Fig. 6c). Three critical patterns significantly positive for GREM1 were then identified ( $R >$

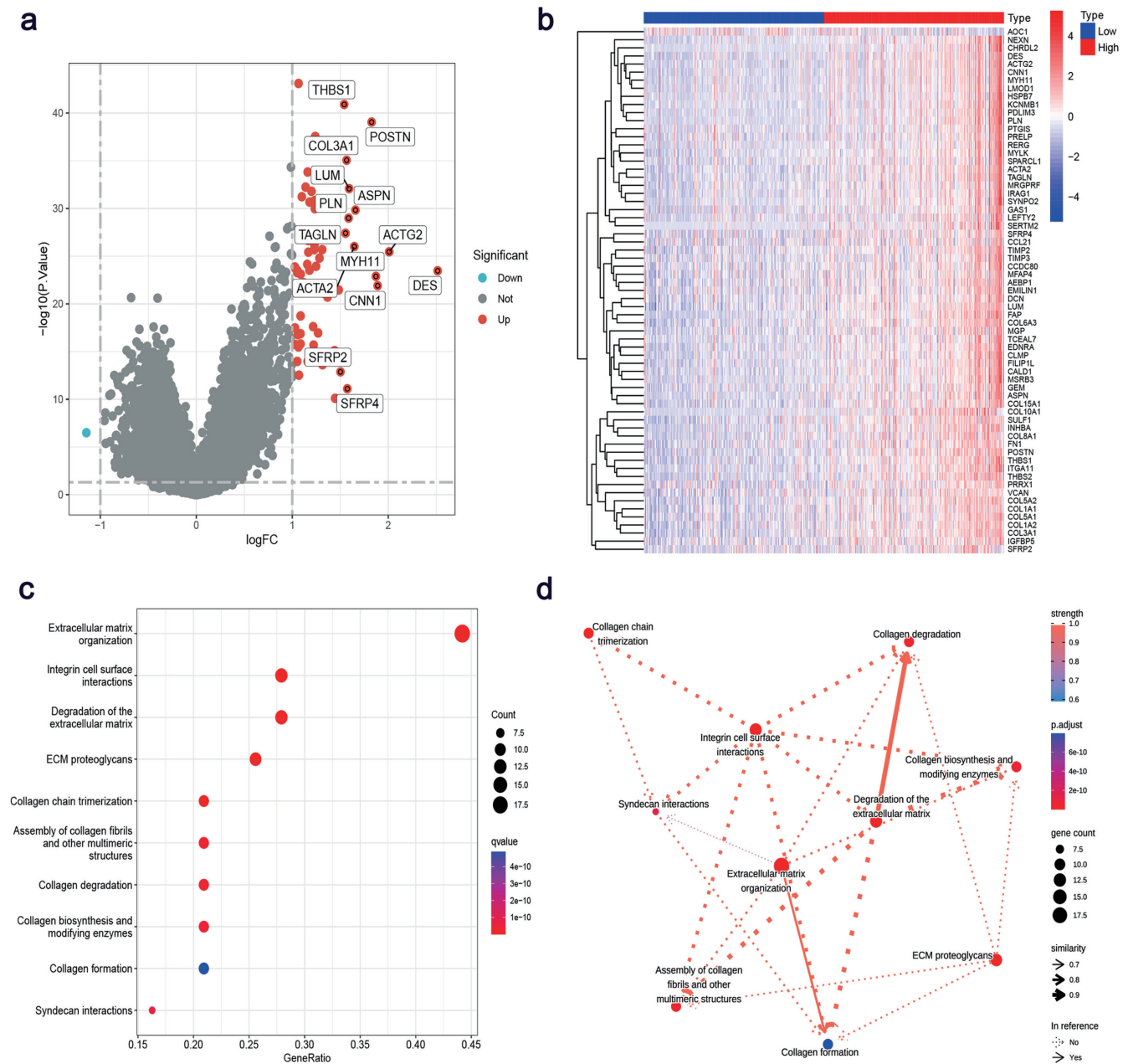
$0.4, p < 0.001$ ; Fig. 6d, e, and f), in which 901 genes obtained from these patterns were obtained for further selection.

#### Inference of biological functions, pathways, and diseases of the hub genes

We took the intersection of 64 upregulated DEGs, 1,043 genes in the ME brown module, and 901 genes in three critical clustering patterns as 57 hub genes with GREM1 (Fig. 7a). GO, DO, and KEGG enrichment analyses were then performed to explore the potential biological functions, pathways, and diseases of the selected hub genes. As shown in Figure 7b, the majority of the hub gene-enriched GO terms were related to extracellular matrix and collagen, mainly including extracellular matrix organization, collagen-containing extracellular matrix, and extracellular matrix structural constituents. The top 20 significant terms of DO enrichment analysis are shown in Figure 7c. Notably, the hub genes were enriched in several DO terms involving gynecologic tumors, such as uterine benign neoplasm, uterine fibroid, female



**Fig. 3. Expression of GREM1 in different clinicopathological characteristics.** (a) BRCA. (b) CHOL. (c) HNSC. (d) LUAD. (e) LUSC. (f) STAD. (g) KICH. (h) KIRC. (i) KIRP. (j) THCA. (k) UCEC. Red marks indicate a statistical difference in any one characteristic. GREM1, Gremlin-1; BRCA, breast invasive carcinoma; CHOL, cholangiocarcinoma; HNSC, head and neck squamous cell carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; STAD, stomach adenocarcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.



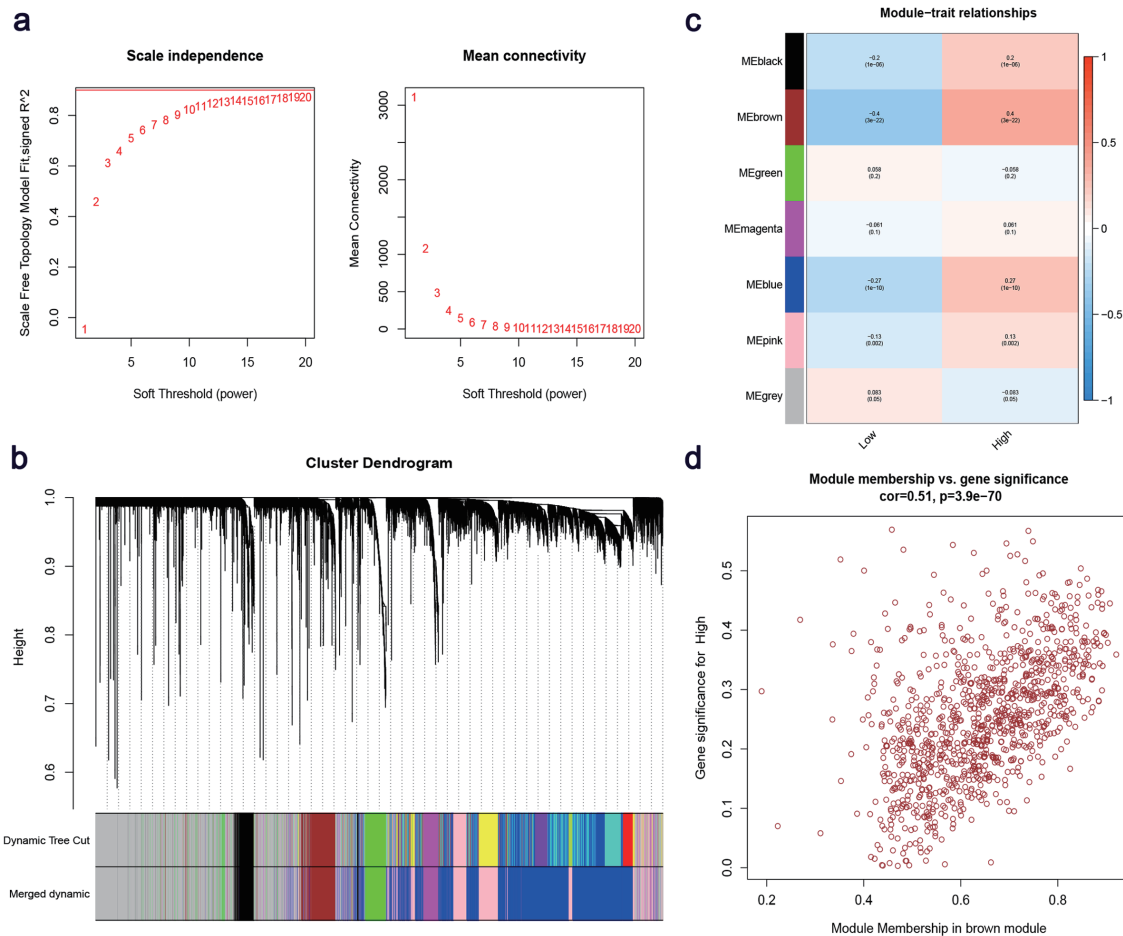
**Fig. 4. Identification of DEGs in low-GREM1 and high-GREM1 expression groups.** (a) Volcano diagram of 65 DEGs, where the blue dot represents downregulated DEGs and red dots represent upregulated DEGs. (b) Heatmap of the expression of 65 DEGs across the samples. (c) Top 10 enrichment KEGG pathways of 65 DEGs. (d) BN plot of the interactions of the top 10 KEGG enrichment pathways. BN, Bayesian network; DEGs, differentially expressed genes; GREM1, Gremlin-1; KEGG, Kyoto Encyclopedia of Genes and Genomes.

reproductive organ benign neoplasm, and reproductive organ benign neoplasm, which indicated the linkage of hub genes and uterine tumors. Moreover, the KEGG enrichment analysis suggested that the upregulation of hub genes may participate in extracellular matrix-related pathways, including ECM-receptor interaction, focal adhesion, and proteoglycans in cancer (Fig. 7d). Figure 7e further depicts the interaction of enriched genes and the top five significant pathways, in which FN1 and THBS1 were the common genes enriched in ECM-receptor interaction, focal adhesion, and proteoglycans in cancer.

**Machine learning-based combination and PPI network identified key genes**

A total of 113 machine learning-based integration was fitted to the training and test cohorts via the leave-one-out cross-validation framework. As shown in Figure 8a, the combination of GLM-Boost and RF was the optimal model performed with the highest average AUC score (0.698) in the training and test cohorts. Subsequently, we obtained 12 genes (FAP, THBS1, POSTN, INHBA, ASPN, COL3A1, DES, IGFBP5, COL8A1, FN1, COL5A2,





**Fig. 5. Construction of a weighted gene co-expression network.** (a) Determination of optimal soft threshold power for scale-free networks. (b) Dendrogram of genes with dissimilarity clustered by dynamic tree shearing. (c) Relationship of seven merged modules and traits (low-GREM1 and high-GREM1 expression groups). A brown module associated the high-GREM1 expression group was observed ( $R = 0.4, p < 0.001$ ). (d) Intramodular analysis shows the correlation of GS and MM in the high-GREM1 expression group. GREM1, Gremlin-1; GS, gene significance; MM, module membership.

*EMILIN1*) screened by the model. We then selected the 10 genes with the highest betweenness in the PPI network of GREM1 (Fig. 8b), thus identifying a final set of 10 key genes highly associated with GREM1 (*FAP, THBS1, POSTN, INHBA, ASPN, COL3A1, IGFBP5, COL8A1, FNI, COL5A2*). The chromosomal locations of GREM1 and 10 genes are shown in Figure 8c. The PCA score plots showing the 10 genes clearly distinguish the low-GREM1 and high-GREM1 groups within the training and test cohorts (Fig. 8d and e). The significant expression differences of *THBS1, ASPN, COL3A1, COL5A2, COL8A1, INHBA, POSTN*, and *FAP* between low-GREM1 and high-GREM1 groups in the training and test cohorts are shown in Figure 8f and g. Additionally, the positive correlation patterns of GREM1 and the 10 genes within the training and test cohorts are shown in Figures 8h and i.

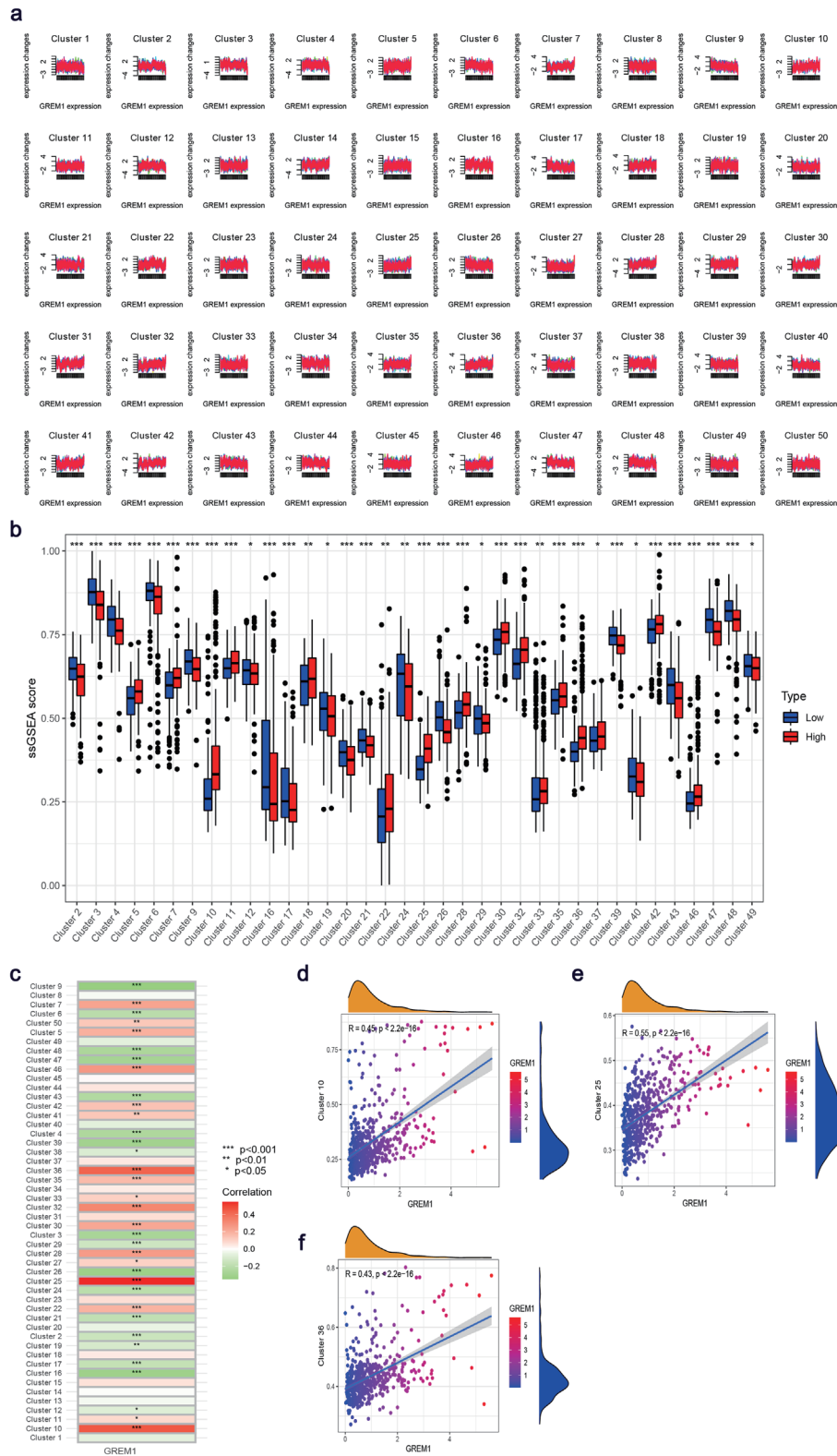
**Assessment of key genes for distinguishing the low-GREM1 and high-GREM1 groups**

We constructed logistic regression models within the training and test cohorts to assess the classification performance of the 10 genes in low- and high-GREM1 groups. The nomograms intuitively visualized the logistic regression model established in the 10 genes within the training cohort (Fig. 9a). The excellent differentiation

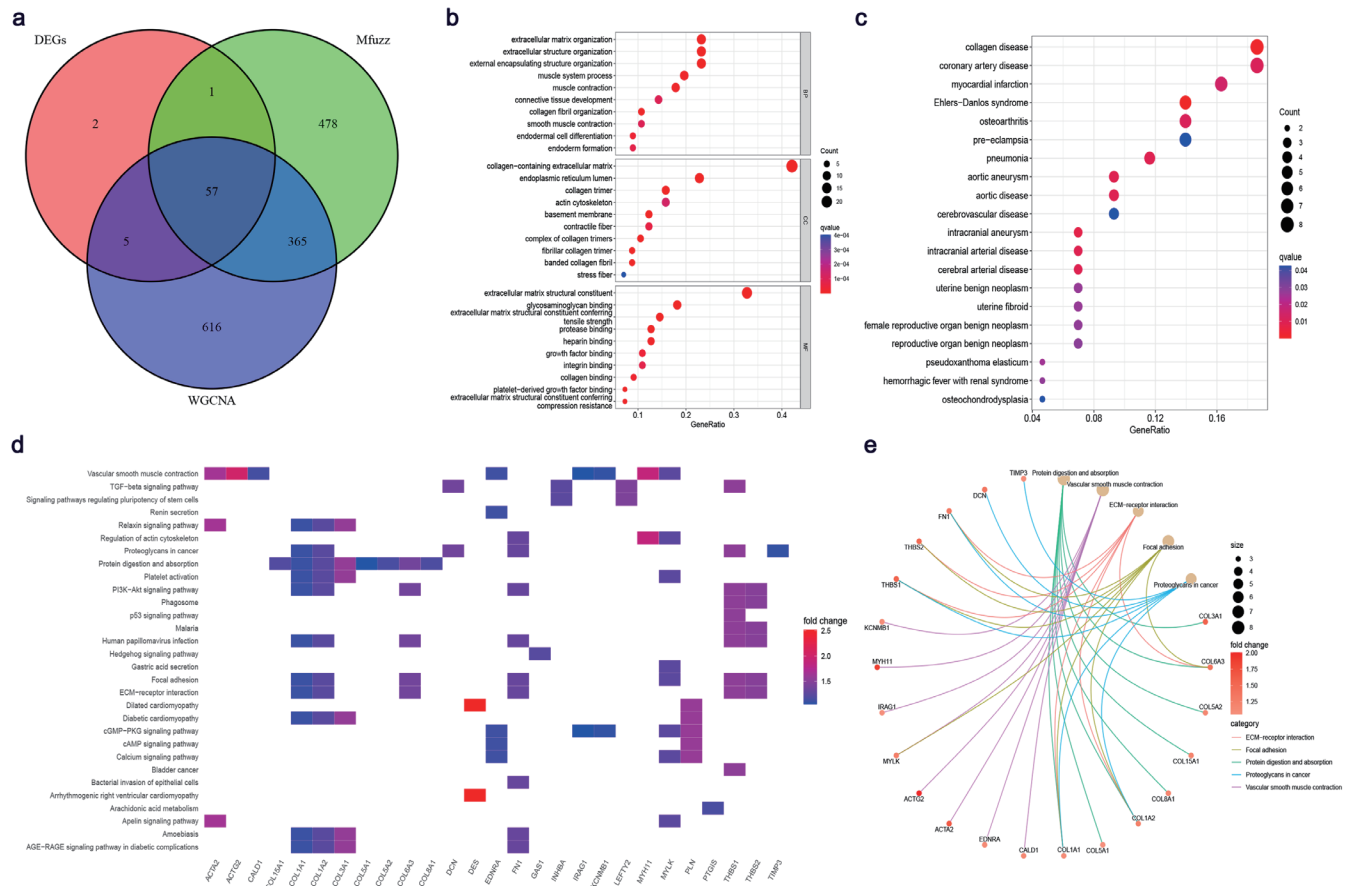
on low- and high-GREM1 groups of the models was evaluated by the AUC scores (all AUC > 0.8; Fig. 9b and c). In addition, the calibration curves showed a minor error between the predicted probability and actual probability, showing the high classification accuracies of the models (Fig. 9d and e). The DCA curves showed that the model deviated from the none and all baselines below the high-risk threshold, which ranged from 0 to 1 and further demonstrated the classification ability and certain application of the models in the training and test cohorts (Fig. 9f and g). These results suggest the critical role of 10 genes closely associated with GREM1 in endometrial cancer.

**Inference of the pathways regulated by key genes and GREM1**

The Bayesian inference-generated regulation network between GREM1 and key genes is shown in Figure 10a. The network demonstrated the intricate crosstalk between key genes that dominated through GREM1 as the upstream regulator. To reveal the potential pathways regulated by 10 genes and GREM1, we carried out KEGG enrichment analysis. Figure 10b shows the top 12 significant pathways of the 10 genes and GREM1. Among the enriched pathways, many extracellular matrix and collagen-related pathways were observed, such as integrin cell-surface interactions,



**Fig. 6. Screening of Mfuzz expression-clustering modules related to GREM1.** (a) A total of 50 Mfuzz clustering modules on the ascending expression of GREM1. (b) Significant distribution of ssGSEA score on clustering modules across the low-GREM1 and high-GREM1 expression groups. (c) Correlation heatmap of clustering modules and GREM1. (d) Significant correlation between GREM1 and clustering module 10. (e) Significant correlation of GREM1 and clustering module 25. (f) Significant correlation of GREM1 and clustering module 36. GREM1, Gremlin-1; ssGSEA, single sample gene set enrichment analysis.



**Fig. 7. Potential biological functions, associated diseases, and signaling pathways of 57 hub genes.** (a) Venn diagram of 57 overlapping genes among 65 DEGs, 1043 genes in ME brown module by WGCNA, and 901 genes obtained from three clustering modules. (b) GO enrichment analysis of hub genes. (c) DO enrichment analysis of hub genes. (d) Heatmap depicting the enrichment pathways including hub genes. (e) Chord diagram of the top five KEGG pathways interacting with the enriched genes. AGE-RAGE, advanced glycation end products; cGMP-PKG, cyclic guanosine monophosphate-protein kinase G; cAMP, cyclic adenosine monophosphate; DEGs, differentially expressed genes; DO, disease ontology; ECM, extracellular matrix; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PI3k-Akt, phosphatidylinositol 3 kinase protein kinase B; TGF- $\beta$ , transforming growth factor-beta; WGCNA, weighted gene co-expression network analysis.

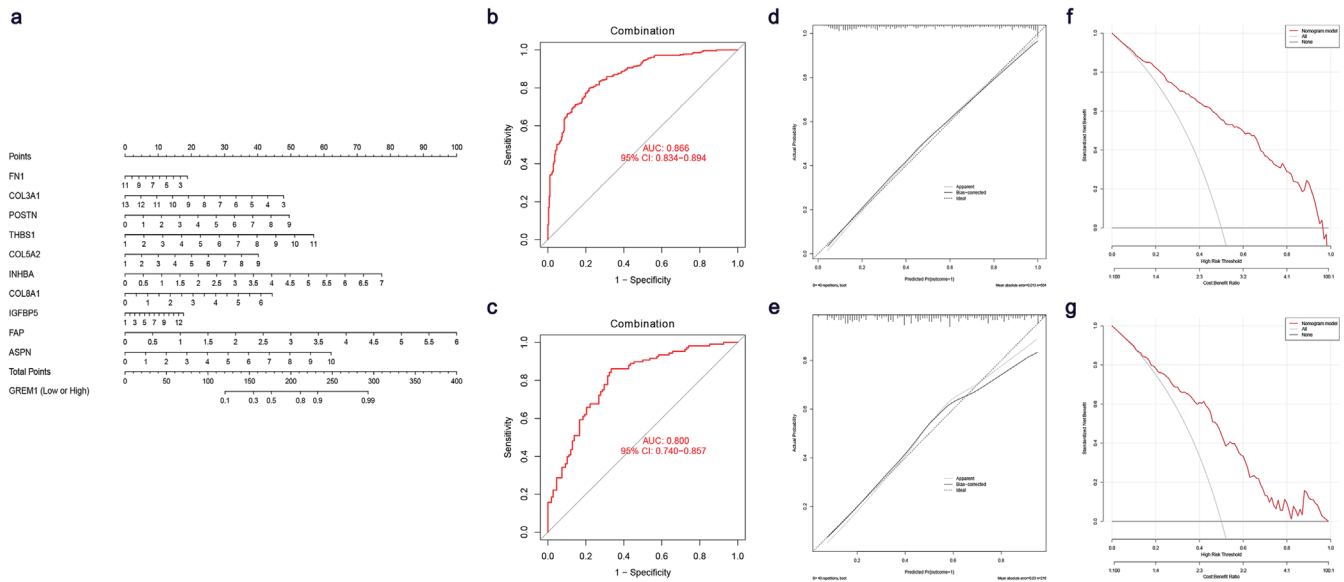
extracellular matrix organization, nonintegrin membrane-ECM interactions, ECM proteoglycans, degradation of the extracellular matrix, collagen chain trimerization, assembly of collagen fibrils and other multimeric structures, and collagen degradation. We performed Bayesian inference of the enrichment pathways and gene expression profile to elucidate the pathway interactions. We obtained an interaction network of the enriched 12 pathways (Fig. 10c) in which integrin cell-surface interactions (Fig. 10d), extracellular matrix organization (Fig. 10e), ECM proteoglycans (Fig. 10f), degradation of the extracellular matrix (Fig. 10g), and collagen degradation (Fig. 10h) were cores with more edges and similarities in the network. The enriched gene interactions in the core pathway are also shown. Generally, five enriched genes (*FN1*, *COL3A1*, *THBS1*, *COL5A2*, and *COL8A1*) were observed in the networks. Of these genes, *FN1* and *THBS1* proved to be significant in the previous KEGG enrichment analysis, suggesting they have potential crucial roles along with *GREM1* in endometrial cancer. The HPA database was searched to obtain experimental evidence of the expressions of these five common enriched key genes at the protein level. Immunohistochemical analysis of four genes, *COL3A1*, *COL8A1*, *FN1*, and *THBS1*, were obtained (Fig.

10i-l). According to the staining and intensity levels (Fig. 10i, k and l), upregulation of *COL3A1*, *FN1*, and *THBS1* was observed in endometrial cancer tissue and *COL8A1* protein was not detected (Fig. 10j). The experimental evidence verified the *in-silico* results and further indicate transactivation of these key genes in endometrial cancer, particularly over-expression of *COL3A1*, *FN1*, and *THBS1* at the protein level. To gain insight into the *GREM1*-regulated mechanism, we also investigated the relationship of fibrotic status and *GREM1* expression in 12 types of cancer. Leveraging the ECM signatures reflecting the fibrotic status, we observed that it was positively correlated with *GREM1* expression of (all  $R > 0$ ,  $p < 0.001$ ) in these cancer types (Fig. 11). These results further showcased the role of *GREM1* in tumor fibrosis of various cancers.

**Expression profiles of immune infiltration cells and correlation analysis**

To investigate whether the immune infiltration cells varied in the low-*GREM1* and high-*GREM1* groups, we performed CIBERSORT and ssGSEA of the gene expression profiles. The results showed the expression profiles of immune infiltration cells (Fig. 12a and b), in which CD8+ T cells, CD4+ memory resting T cells,





**Fig. 9. Classification performance of 10 key genes in low-GREM1 and high-GREM expression groups within the training and test cohorts.** (a) Nomograph of the logistic regression model of 10 key genes within the training cohort. (b) ROC curve of model classification accuracy in low-GREM1 and high-GREM expression groups in the training cohort. (c) ROC curve of model classification accuracy of low- and high-GREM expression groups in the test cohort. (d) Calibration plot of the model predictive and actual classified probabilities in the test cohort. (e) Calibration plot of the model predictive and actual classified probabilities in the training cohort. (f) DCA curve of the application of the model distinguishing the low-GREM1 and high-GREM1 groups in the training cohort; (g) DCA curve of the application of the model distinguishing the low- and high-GREM1 groups in the test cohort. AUC, area under curve; DCA, decision curve analysis; GREM1, Gremlin-1; ROC, receiver operating characteristic.

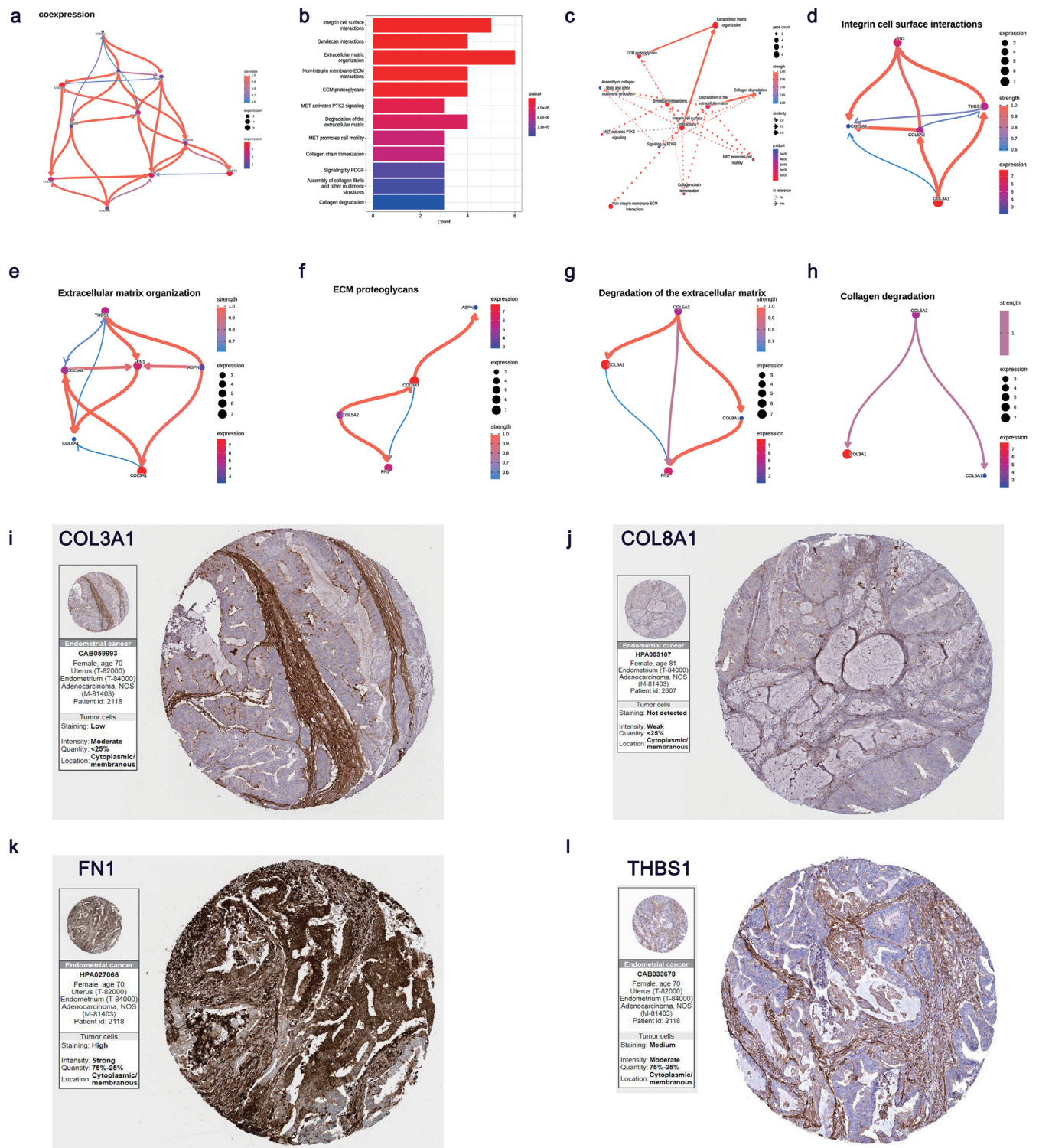
upregulated in the high-GREM1 groups and was positively correlation with ASPN, COL3A1, FAP, GREM1, IGFBP5, POSTN, THBS1, which suggested the mast cells were highly influenced by GREM1 in endometrial cancer.

**Discussion**

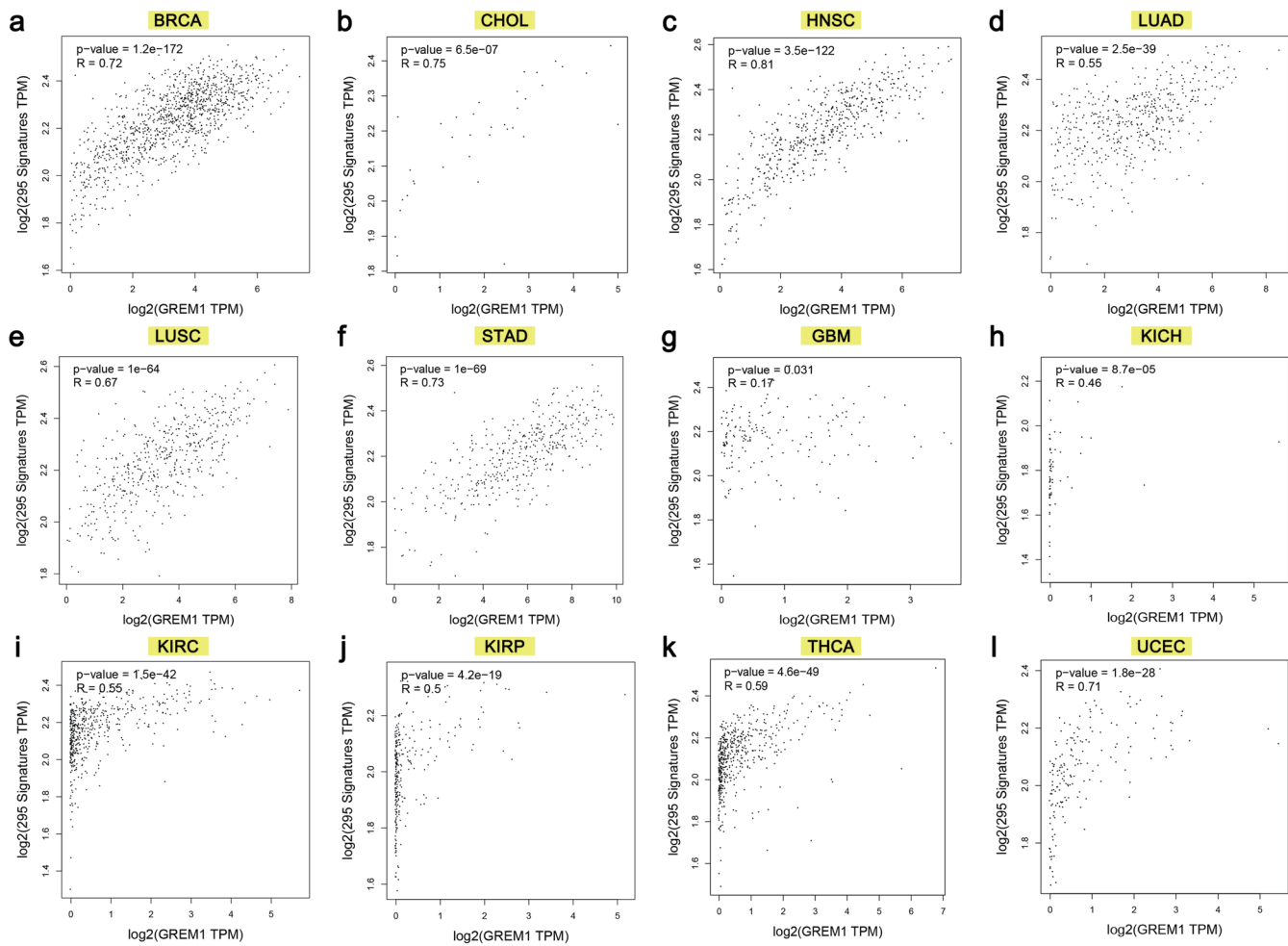
GREM1 involvement in carcinogenesis has been investigated and is seen as a prospective therapeutic target in various malignancies.<sup>7-10,31</sup> Nevertheless, the role of GREM1 in endometrial cancer and its association with immune cells is not well understood. Herein, a comprehensive bioinformatics analysis combined with a machine learning-based combination was applied to reveal the potential mechanism of GREM1 in endometrial cancer. Inspired by evidence that GREM1 over-expression inhibits epithelial-to-mesenchymal transition in pancreatic cancer,<sup>9</sup> we wondered whether a similar mechanism occurred in endometrial cancer. We focused the analysis on the expression profiles of cancer tissues with high-GREM1 expression in the TCGA and GEO databases. We first found downregulation of GREM1 mRNA and protein in endometrial cancer, which was consistent with the expression pattern previously reported in pancreatic cancer.<sup>9</sup> Downregulation of GREM1 expression was observed in glioblastoma. Although we found overall downregulation of GREM1 in endometrial cancer, the heterogeneity of cancer exists, suggesting a certain cell population with a higher GREM1 expression pattern. Additionally, we performed prognosis and clinicopathological analysis of GREM1 in 12 cancer types in the TCGA database. The expression of GREM1 was significantly different from that in normal controls (up/downregulation pattern). The results indicated that high expression of GREM1 predicted poor prognosis (shorter OS) and clinicopathological features (advanced tumor grade/stage) in lung squamous cell carcinoma, kidney renal

clear cell carcinoma, kidney renal papillary cell carcinoma, and endometrial carcinoma. We thus suggest that the patients with tumors that have high GREM1 expression have poor clinical outcomes, indicating that GREM1 contributes to tumor progression.

Then, we classified the endometrial cancer samples into low-GREM1 and high-GREM1 groups. We screened out a total of 65 DEGs (64 upregulated genes and one downregulated gene) between these two groups. Subsequently, WGCNA was performed to discern the ME brown module highly positively linked to high-GREM1 trait, of which 1,043 genes were included. Together with ssGSEA scoring and correlation analysis, Mfuzz clustering was also conducted to screen out three high-GREM1-related gene expression patterns containing 901 genes. We took the intersection of the 64 upregulated DEGs, 1,043 genes identified by WGCNA, and Mfuzz-selected 901 genes. We observed a set of 57 overlapping genes, indicating that the prefeature selection procedure using multiple bioinformatics methods was reliable and accessible. In addition, we conducted DO, GO, and KEGG enrichment analysis on the selected 57 hub genes to investigate the biological processes and pathways in endometrial cancer. Interestingly, the hub genes were enriched in uterine benign neoplasms, uterine fibroids, female reproductive organ benign neoplasms, and reproductive organ benign neoplasms, suggesting an association between the hub genes and uterine carcinomas. Furthermore, GO and KEGG analysis indicated that the hub genes were enriched in extracellular matrix-related terms, including extracellular matrix organization, collagen-containing extracellular matrix, ECM-receptor interaction, focal adhesion, and proteoglycans in cancer. We speculate that the hub GREM1 genes participate in ECM-related pathways. Deregulated ECM dynamics, such as collagen deposition or an increase in ECM stiffness,<sup>32</sup> are signatures of tumor proliferation and metastasis. It is well known that ECM is a complex and dynamic environment mainly composed of collagen, proteoglycans,



**Fig. 10. Exploration of the potential pathways and pathway interactions regulated by 10 core genes.** (a) Gene regulatory networks of GREM1 and key genes. (b) Top 12 KEGG enrichment pathways of 10 core genes. (c) Interaction network of 12 enrichment pathways. (d) Gene regulatory networks in integrin cell-surface interactions. (e) Gene regulatory networks in extracellular matrix organization. (f) Gene regulatory networks in ECM proteoglycans. (g) Gene regulatory networks in degradation of the extracellular matrix. (h) Gene regulatory networks in collagen degradation. (i) Immunohistochemical staining of COL3A1 in endometrial cancer tissue. (j) Immunohistochemical staining of COL8A1 in endometrial cancer tissue. (k) Immunohistochemical staining of FN1 in endometrial cancer tissue. (l) Immunohistochemical staining of THBS1 in endometrial cancer tissue. ECM, extracellular matrix; GREM1, Gremlin-1; KEGG, Kyoto Encyclopedia of Genes and Genomes.



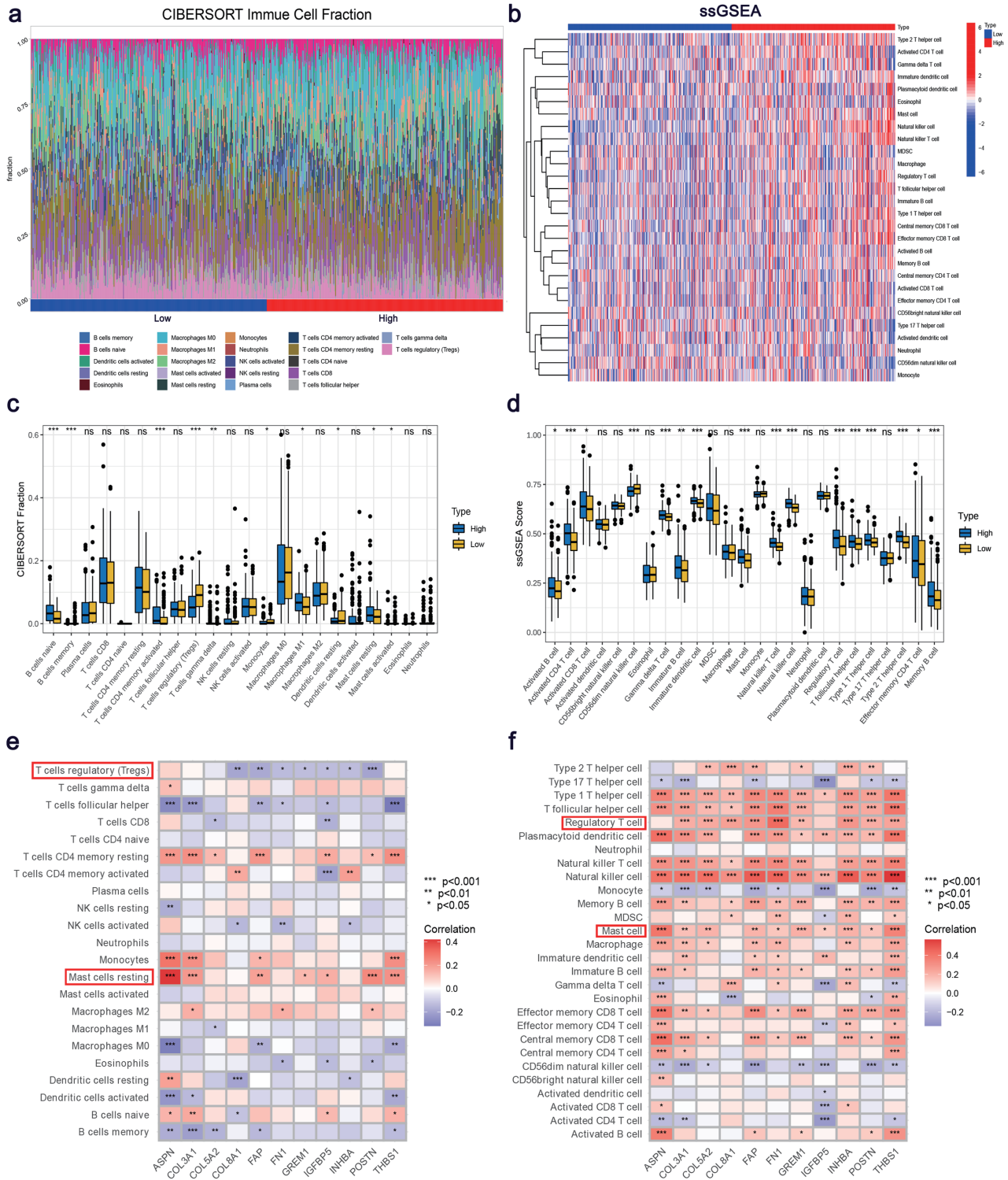
**Fig. 11. Correlation of fibrotic status and GREM1 expression in 12 cancer types.** (a) BRCA. (b) CHOL. (c) HNSC. (d) LUAD. (e) LUSC. (f) STAD. (g) GBM. (h) KICH. (i) KIRC. (j) KIRP. (k) THCA. (l) UCEC. Red marks indicate a statistical difference. BRCA, breast invasive carcinoma; CHOL, cholangiocarcinoma; GBM, glioblastoma multiforme; GREM1, Gremlin-1; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.

laminin, and fibronectin.<sup>33</sup> Thus, we also wondered which specific ECM composition was a great contribution or highly correlated to the higher level of GREM1, further exploring the underlying regulatory relationship with GREM1 in endometrial cancer.

To define the underlying molecular mechanism of GREM1 more precisely in endometrial cancer, feature selection on the hub genes was still needed. Herein, the machine learning-based integrative procedure was taken for feature selection. We obtained a total of 12 genes (*FAP*, *THBS1*, *POSTN*, *INHBA*, *ASP*, *COL3A1*, *DES*, *IGFBP5*, *COL8A1*, *FNI*, *COL5A2*, and *EMILIN1*) via best-performance algorithm integration (GLMBoost target RF) with the highest AUC scores. To explore whether the functional associations between the 12 genes and *GREM1* exist, we constructed a PPI network interacting with 12 genes and *GREM1*. Using Cytoscape software, we calculated the betweenness of 12 genes and *GREM1* in the network. *DES* and *EMILIN1*, with zero betweenness were excluded, and a PPI network composed of 10 genes and *GREM1* was then visualized. We thus obtained a final set of 10 genes (*FAP*, *THBS1*, *POSTN*, *INHBA*, *ASP*, *COL3A1*, *IGFBP5*, *COL8A1*, *FNI*, and *COL5A2*), which are considered as key *GREM1* genes.

Using three-dimensional PCA score plots, we observed the classification performances of key genes in low-*GREM1* and high-*GREM1* groups. The elevated expression of key genes in the high-*GREM1* groups was then demonstrated, following their positive expression correlations with *GREM1*, which implies that the co-expressions of key genes and *GREM1* exist in endometrial cancer.

To systematically assess the classification abilities of key genes from the low-*GREM1* and high-*GREM1* groups, we established logistic regression models, following ROC curves, calibration curves, and DCA curves as evaluation indexes. We noted robust AUC scores, strong calibrations, and distinct utility of the discrimination models within the training and test cohorts. These results show that the key genes were clearly distinguished in the low-*GREM1* and high-*GREM1* groups, indicating their expression was similar to that of *GREM1*. The key genes were subjected to KEGG enrichment analysis, together with conducting BN inference on the enriched pathways and gene expression patterns. Interestingly, the key genes were also enriched in extracellular matrix-related pathways, especially those related to ECM structural constituent and dynamic changes. Through the pathway interaction based on BN



**Fig. 12. Inference of immune-cell composition on the expression profile through CIBERSORT and ssGSEA algorithms.** (a) Bar plot showing the relative composition of 22 kinds of immune cells in the low-GREM1 and high-GREM1 expression groups through CIBERSORT. (b) Heatmap of the relative infiltration abundance of 48 kinds of immune cells in the low-GREM1 and high-GREM1 expression groups compared via CIBERSORT. (c) Levels of immune-cell infiltration in the low-GREM1 and high-GREM1 expression groups compared via CIBERSORT. (d) Levels of immune-cell infiltration in the low-GREM1 and high-GREM1 expression groups compared via ssGSEA. (e) Correlation heatmap of 10 key genes and the infiltration of immune cells based on the CIBERSORT results. (f) Correlation heatmap of 10 core genes and the infiltration levels of immune cells based on the ssGSEA results. CIBERSORT, Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts; GREM1, Gremlin-1; ssGSEA, single sample gene set enrichment analysis. GREM1, Gremlin-1; ssGSEA, single sample gene set enrichment analysis.



inference, we found integrin cell-surface interactions, extracellular matrix organization, ECM proteoglycans, degradation of the extracellular matrix, and collagen degradation were core pathways. We then investigated the gene interactions of the core pathways and observed five common genes (*FNI*, *COL3A1*, *THBS1*, *COL5A2*, and *COL8A1*) enriched in the interactions. Notably, the collagen genes *COL3A1*, *COL5A2*, and *COL8A1* almost occurred in the enriched pathways, suggesting the critical role of collagen in ECM-related pathways. Moreover, we obtained immunohistochemical staining results of *FNI*, *COL3A1*, *THBS1*, and *COL8A1* in endometrial cancer from the HPA database, which was as external experimental evidence supporting our analysis. These results showcased the elevated expression of *FNI*, *COL3A1*, and *THBS1* proteins in endometrial cancer, which strengthened the *in-silico* conclusions.

Recent studies have demonstrated that *COL3A1* over-expression leads to poor prognosis or promotes cancer cell progression in several malignancies, including bladder cancer, colorectal cancer, and triple-negative breast cancer.<sup>34–37</sup> Similarly, recent evidence has revealed that *COL5A2* or *COL8A1* expression was related to poor prognosis or tumor progression,<sup>38–45</sup> also highlighting their potential diagnostic values for different malignancies.<sup>38,46,47</sup> *FNI*, acting as an upstream or downstream regulator, was previously shown to promote or inhibit tumor progression in various malignancies.<sup>48–51</sup> *THBS1* was found to positively associate with aggressive tumor development and progression.<sup>52–54</sup> Additionally, elevated microvessel counts were detected in endometrial cancer patients with high *THBS1* expression, suggesting a role of *THBS1* in tumor angiogenesis.<sup>55</sup> But few studies reported roles of *FNI*, *COL3A1*, *THBS1*, *COL5A2*, and *COL8A1* in endometrial cancer, more studies are warranted. Taken together, we hypothesize that *GREM1* is involved in the modulation of ECM-related pathways, especially those mainly regulated by collagen genes. Modulation of the ECM may allow for tumor invasion or metastasis, which has been shown in studies of estrogen receptor-negative breast cancer.<sup>56</sup>

The tumor microenvironment, especially immune-cell infiltration, was shown to be linked to ECM modulation in tumor progression.<sup>57,58</sup> Leveraging the CIBERSOT and ssGSEA algorithms, the infiltration of immune cells in low-*GREM1* and high-*GREM1* groups was measured to explore the linkage of *GREM1* and the tumor microenvironment of endometrial cancer. We aimed to identify the specific immune cells correlated with *GREM1* and its interactive key genes, with the upregulation of infiltration in high-*GREM1* groups. Differential expression analysis showed that the infiltration of mast cells was elevated in high-*GREM1* groups. Analysis of correlations with immune-cell subtypes revealed that mast cell infiltration levels were positively associated with *GREM1* and key genes. Previous evidence supports the linkage between increased mast cell count and the progression of endometrial cancer, such as angiogenesis.<sup>59</sup> Furthermore, the high density of mast cells is significantly associated with myometrial invasion in endometrial cancer, which shows the function of mast cells in tumor progression.<sup>60</sup> However, because of the limited number of reports on how mast cells contribute to the progression of endometrial cancer, more functional studies are required to clarify their role. Overall, our data indicate that mast cell infiltration in endometrial cancer may be influenced by ECM modulation, which is correlated with *GREM1* and key genes.

## Conclusions

Our results suggest the regulation of ECM-related signal pathways in endometrial cancer by *GREM1* expression, most likely

with ECM degradation regulated by collagen genes (*COL3A1*, *COL5A2*, and *COL8A1*). Additionally, we found that the expression of *GREM1* and its key genes were positively correlated with mast cells infiltrated in endometrial cancer. The infiltration levels of mast cells were upregulated in higher-*GREM1* expression groups. These results suggested that mast cells could be used as a marker of immune infiltration in aggressive endometrial cancer having higher *GREM1* expression. Overall, our study shows that *GREM1* expression correlated with endometrial cancer progression by regulating ECM modulation, which adds to our understanding of the molecular effects of *GREM1* and is a reference for further functional studies in endometrial cancer.

## Acknowledgments

There is nothing to declare.

## Funding

This work was supported by the National Natural Science Foundation of China under Grant numbers 32170423 and 31501166; the Chenguang Program from the Shanghai Education Committee under Grant number 14CG49; the Shanghai Sailing Program of Science and Technology Commission of Shanghai Municipality under Grant number 15YF1405000; and SciTech Funding by CSPFTZ Lingang Special Area Marine Biomedical Innovation Platform.

## Conflict of interest

The authors have no conflict of interests related to this publication.

## Author contributions

Study concept and design (YHZ, YZ), acquisition of data (YHZ), analysis and interpretation of data (YHZ), drafting of the manuscript (YHZ), critical revision of the manuscript for important intellectual content (HW, YZ), administrative, technical, or material support (YHZ, YZ), and study supervision (HW, YZ). All authors made significant contributions to this study and approved the final manuscript.

## Data sharing statement

Data used in this research were available in the TCGA (<https://portal.gdc.cancer.gov/>) and GEO databases (<http://www.ncbi.nlm.nih.gov/geo>).

## References

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71(3):209–249. doi:10.3322/caac.21660, PMID:33538338.
- [2] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* 2022;72(1):7–33. doi:10.3322/caac.21708, PMID:35020204.
- [3] Dou Y, Kawaler EA, Cui Zhou D, Gritsenko MA, Huang C, Blumenberg L, *et al*. Roteogenomic characterization of endometrial carcinoma. *Cell* 2020;180(4):729–748.e26. doi:10.1016/j.cell.2020.01.026, PMID:32059776.
- [4] Morice P, Leary A, Creutzberg C, Abu-Rustum N, Darai E. Endometrial cancer. *Lancet* 2016;387(10023):1094–1108. doi:10.1016/s0140-

- 6736(15)00130-0, PMID:26354523.
- [5] Dizon DS. Treatment options for advanced endometrial carcinoma. *Gynecol Oncol* 2010;117(2):373–381. doi:10.1016/j.ygyno.2010.02.007, PMID:20223510.
  - [6] Urlick ME, Bell DW. Clinical actionability of molecular targets in endometrial cancer. *Nat Rev Cancer* 2019;19(9):510–521. doi:10.1038/s41568-019-0177-x, PMID:31388127.
  - [7] Elemam NM, Malek AI, Mahmoud EE, El-Huneidi W, Talaat IM. Insights into the role of gremlin-1, a bone morphogenic protein antagonist, in cancer initiation and progression. *Biomedicines* 2022;10(2):301. doi:10.3390/biomedicines10020301, PMID:35203511.
  - [8] Kim M, Yoon S, Lee S, Ha SA, Kim HK, Kim JW, *et al.* Gremlin-1 induces BMP-independent tumor cell proliferation, migration, and invasion. *PLoS One* 2012;7(4):e35100. doi:10.1371/journal.pone.0035100, PMID:22514712.
  - [9] Lan L, Evan T, Li H, Hussain A, Ruiz EJ, Zaw Thin M, *et al.* GREM1 is required to maintain cellular heterogeneity in pancreatic cancer. *Nature* 2022;607(7917):163–168. doi:10.1038/s41586-022-04888-7, PMID:35768509.
  - [10] Cheng C, Wang J, Xu P, Zhang K, Xin Z, Zhao H, *et al.* Gremlin1 is a therapeutically targetable FGFR1 ligand that regulates lineage plasticity and castration resistance in prostate cancer. *Nat Cancer* 2022;3(5):565–580. doi:10.1038/s43018-022-00380-3, PMID:35624341.
  - [11] Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, *et al.* Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun* 2022;13(1):816. doi:10.1038/s41467-022-28421-6, PMID:35145098.
  - [12] Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, *et al.* TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016;44(8):e71. doi:10.1093/nar/gkv1507, PMID:26704973.
  - [13] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 2007;23(14):1846–1847. doi:10.1093/bioinformatics/btm254, PMID:17496320.
  - [14] Day RS, McDade KK. A decision theory paradigm for evaluating identifier mapping and filtering methods using data integration. *BMC Bioinformatics* 2013;14:223. doi:10.1186/1471-2105-14-223, PMID:23855655.
  - [15] Hermyt E, Zmarzły N, Grabarek B, Kruszniewska-Rajs C, Gola J, Jęda-Golonka A, *et al.* Interplay between miRNAs and Genes Associated with Cell Proliferation in Endometrial Cancer. *Int J Mol Sci* 2019;20(23):6011. doi:10.3390/ijms20236011, PMID:31795319.
  - [16] Sugiyama Y, Gotoh O, Fukui N, Tanaka N, Hasumi K, Takazawa Y, *et al.* Two distinct tumorigenic processes in endometrial endometrioid adenocarcinoma. *Am J Pathol* 2020;190(1):234–251. doi:10.1016/j.ajpath.2019.09.022, PMID:31733184.
  - [17] Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, *et al.* TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* 2017;77(21):e108–e110. doi:10.1158/0008-5472.CAN-17-0307, PMID:29092952.
  - [18] Lee BT, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, *et al.* The UCSC genome browser database: 2022 update. *Nucleic Acids Res* 2022;50(D1):D1115–D1122. doi:10.1093/nar/gkab959, PMID:34718705.
  - [19] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, *et al.* UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 2017;19(8):649–658. doi:10.1016/j.neo.2017.05.002, PMID:28732212.
  - [20] Shen Y, Liu J, Zhang L, Dong S, Zhang J, Liu Y, *et al.* Identification of potential biomarkers and survival analysis for head and neck squamous cell carcinoma using bioinformatics strategy: a study based on TCGA and GEO datasets. *Biomed Res Int* 2019;2019:7376034. doi:10.1155/2019/7376034, PMID:31485443.
  - [21] Sato N, Tamada Y, Yu G, Okuno Y. CBNplot: Bayesian network plots for enrichment analysis. *Bioinformatics* 2022;38(10):2959–2960. doi:10.1093/bioinformatics/btac175, PMID:35561164.
  - [22] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559. doi:10.1186/1471-2105-9-559, PMID:19114008.
  - [23] Kumar L, E Futschik M. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2007;2(1):5–7. doi:10.6026/97320630002005, PMID:18084642.
  - [24] Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 2009;462(7269):108–112. doi:10.1038/nature08460, PMID:19847166.
  - [25] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13(11):2498–2504. doi:10.1101/gr.1239303, PMID:14597658.
  - [26] von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003;31(1):258–261. doi:10.1093/nar/gkg034, PMID:12519996.
  - [27] Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, *et al.* Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 2017;318(14):1377–1384. doi:10.1001/jama.2017.12126, PMID:29049590.
  - [28] Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol* 2018;1711:243–259. doi:10.1007/978-1-4939-7493-1\_12, PMID:29344893.
  - [29] Xue Y, Tong L, Liu A, Liu F, Liu A, Zeng S, Xiong Q, *et al.* Tumor-infiltrating M2 macrophages driven by specific genomic alterations are associated with prognosis in bladder cancer. *Oncol Rep* 2019;42(2):581–594. doi:10.3892/or.2019.7196, PMID:31233191.
  - [30] Wang J, Kang Z, Liu Y, Li Z, Liu Y, Liu J. Identification of immune cell infiltration and diagnostic biomarkers in unstable atherosclerotic plaques by integrated bioinformatics analysis and machine learning. *Front Immunol* 2022;13:956078. doi:10.3389/fimmu.2022.956078, PMID:36211422.
  - [31] Bach DH, Park HJ, Lee SK. The dual role of bone morphogenetic proteins in cancer. *Mol Ther Oncolytics* 2018;8:1–13. doi:10.1016/j.omto.2017.10.002, PMID:29234727.
  - [32] Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep* 2014;15(12):1243–1253. doi:10.15252/embr.201439246, PMID:25381661.
  - [33] Walker C, Mojares E, Del Río Hernández A. Role of extracellular matrix in development and cancer progression. *Int J Mol Sci* 2018;19(10):3028. doi:10.3390/ijms19103028, PMID:30287763.
  - [34] Tang L, Lei YY, Liu YJ, Tang B, Yang SM. The expression of seven key genes can predict distant metastasis of colorectal cancer to the liver or lung. *J Dig Dis* 2020;21(11):639–649. doi:10.1111/1751-2980.12936, PMID:32896975.
  - [35] Wang XQ, Tang ZX, Yu D, Cui SJ, Jiang YH, Zhang Q, *et al.* Epithelial but not stromal expression of collagen alpha-1(III) is a diagnostic and prognostic indicator of colorectal carcinoma. *Oncotarget* 2016;7(8):8823–8838. doi:10.18632/oncotarget.6815, PMID:26741506.
  - [36] Yang F, Lin L, Li X, Wen R, Zhang X. Silencing of COL3A1 represses proliferation, migration, invasion, and immune escape of triple negative breast cancer cells via down-regulating PD-L1 expression. *Cell Biol Int* 2022;46(11):1959–1969. doi:10.1002/cbin.11875, PMID:35930601.
  - [37] Yuan L, Shu B, Chen L, Qian K, Wang Y, Qian G, *et al.* Overexpression of COL3A1 confers a poor prognosis in human bladder cancer identified by co-expression analysis. *Oncotarget* 2017;8(41):70508–70520. doi:10.18632/oncotarget.19733, PMID:29050298.
  - [38] Ding YL, Sun SF, Zhao GL. COL5A2 as a potential clinical biomarker for gastric cancer and renal metastasis. *Medicine (Baltimore)* 2021;100(7):e24561. doi:10.1097/md.00000000000024561, PMID:33607786.
  - [39] Fischer H, Stenling R, Rubio C, Lindblom A. Colorectal carcinogenesis is associated with stromal expression of COL11A1 and COL5A2. *Carcinogenesis* 2001;22(6):875–878. doi:10.1093/carcin/22.6.875, PMID:11375892.
  - [40] Ren X, Chen X, Fang K, Zhang X, Wei X, Zhang T, *et al.* COL5A2 promotes proliferation and invasion in prostate cancer and is one of seven gleason-related genes that predict recurrence-free survival. *Front Oncol* 2021;11:583083. doi:10.3389/fonc.2021.583083, PMID:33816226.
  - [41] Sagara A, Miura S, Kobinata A, Naganawa R, Yaginuma S, Saito S, *et al.*

- al*. COL8A1 enhances the invasion/metastasis in MDA-MB-231 cells via the induction of IL1B and MMP1 expression. *Biochem Biophys Res Commun* 2023;642:145–153. doi:10.1016/j.bbrc.2022.12.046, PMID:36577251.
- [42] Sato F, Sagara A, Tajima K, Miura S, Inaba K, Ando Y, *et al*. COL8A1 facilitates the growth of triple-negative breast cancer via FAK/Src activation. *Breast Cancer Res Treat* 2022;194(2):243–256. doi:10.1007/s10549-022-06635-y, PMID:35624176.
- [43] Shang J, Wang F, Chen P, Wang X, Ding F, Liu S, *et al*. Co-expression network analysis identified COL8A1 is associated with the progression and prognosis in human colon adenocarcinoma. *Dig Dis Sci* 2018;63(5):1219–1228. doi:10.1007/s10620-018-4996-5, PMID:29497907.
- [44] Tan Y, Chen Q, Xing Y, Zhang C, Pan S, An W, *et al*. High expression of COL5A2, a member of COL5 family, indicates the poor survival and facilitates cell migration in gastric cancer. *Biosci Rep* 2021;41(4):BSR20204293. doi:10.1042/bsr20204293, PMID:33739392.
- [45] Wang J, Jiang YH, Yang PY, Liu F. Increased collagen type V  $\alpha$ 2 (COL5A2) in colorectal cancer is associated with poor prognosis and tumor progression. *Onco Targets Ther* 2021;14:2991–3002. doi:10.2147/ott.S288422, PMID:33981148.
- [46] Li P, Ji W, Wei Z, Wang X, Qiao G, Gao C, *et al*. Comprehensive analysis to identify pseudogenes/lncRNAs-hsa-miR-200b-3p-COL5A2 network as a prognostic biomarker in gastric cancer. *Hereditas* 2022;159(1):43. doi:10.1186/s41065-022-00257-6, PMID:36447214.
- [47] Peng W, Li JD, Zeng JJ, Zou XP, Tang D, Tang W, *et al*. Clinical value and potential mechanisms of COL8A1 upregulation in breast cancer: a comprehensive analysis. *Cancer Cell Int* 2020;20:392. doi:10.1186/s12935-020-01465-8, PMID:32818022.
- [48] Cai X, Liu C, Zhang TN, Zhu YW, Dong X, Xue P. Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. *J Cell Biochem* 2018;119(6):4717–4728. doi:10.1002/jcb.26651, PMID:29274284.
- [49] Ifon ET, Pang AL, Johnson W, Cashman K, Zimmerman S, Muralidhar S, *et al*. U94 alters FN1 and ANGPTL4 gene expression and inhibits tumorigenesis of prostate cancer cell line PC3. *Cancer Cell Int* 2005;5:19. doi:10.1186/1475-2867-5-19, PMID:15972109.
- [50] Sponziello M, Rosignolo F, Celano M, Maggiano V, Pecce V, De Rose RF, *et al*. Fibronectin-1 expression is increased in aggressive thyroid cancer and favors the migration and invasion of cancer cells. *Mol Cell Endocrinol* 2016;431:123–132. doi:10.1016/j.mce.2016.05.007, PMID:27173027.
- [51] Waalkes S, Atschekzei F, Kramer MW, Hennenlotter J, Vetter G, Becker JU, *et al*. Fibronectin 1 mRNA expression correlates with advanced disease in renal cancer. *BMC Cancer* 2010;10:503. doi:10.1186/1471-2407-10-503, PMID:20860816.
- [52] Liu X, Xu D, Liu Z, Li Y, Zhang C, Gong Y, *et al*. THBS1 facilitates colorectal liver metastasis through enhancing epithelial-mesenchymal transition. *Clin Transl Oncol* 2020;22(10):1730–1740. doi:10.1007/s12094-020-02308-8, PMID:32052380.
- [53] Pal SK, Nguyen CT, Morita KI, Miki Y, Kayamori K, Yamaguchi A, *et al*. THBS1 is induced by TGF $\beta$ 1 in the cancer stroma and promotes invasion of oral squamous cell carcinoma. *J Oral Pathol Med* 2016;45(10):730–739. doi:10.1111/jop.12430, PMID:26850833.
- [54] Shen J, Cao B, Wang Y, Ma C, Zeng Z, Liu L, *et al*. Hippo component YAP promotes focal adhesion and tumour aggressiveness via transcriptionally activating THBS1/FAK signalling in breast cancer. *J Exp Clin Cancer Res* 2018;37(1):175. doi:10.1186/s13046-018-0850-z, PMID:30055645.
- [55] Seki N, Kodama J, Hashimoto I, Hongo A, Yoshinouchi M, Kudo T. Thrombospondin-1 and -2 messenger RNA expression in normal and neoplastic endometrial tissues: correlation with angiogenesis and prognosis. *Int J Oncol* 2001;19(2):305–310. doi:10.3892/ijo.19.2.305, PMID:11445843.
- [56] Neckmann U, Wolowczyk C, Hall M, Almaas E, Ren J, Zhao S, *et al*. GREM1 is associated with metastasis and predicts poor prognosis in ER-negative breast cancer patients. *Cell Commun Signal* 2019;17(1):140. doi:10.1186/s12964-019-0467-7, PMID:31694641.
- [57] Acerbi I, Cassereau L, Dean I, Shi Q, Au A, Park C, *et al*. Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration. *Integr Biol (Camb)* 2015;7(10):1120–1134. doi:10.1039/c5ib00040h, PMID:25959051.
- [58] Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, *et al*. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Mol Cancer* 2021;20(1):131. doi:10.1186/s12943-021-01428-1, PMID:34635121.
- [59] Ribatti D, Nico B, Finato N, Crivellato E. Tryptase-positive mast cells and CD8-positive T cells in human endometrial cancer. *Pathol Int* 2011;61(7):442–444. doi:10.1111/j.1440-1827.2011.02680.x, PMID:21707849.
- [60] Cinel L, Aban M, Basturk M, Ertunc D, Arpacı R, Dilek S, *et al*. The association of mast cell density with myometrial invasion in endometrial carcinoma: a preliminary report. *Pathol Res Pract* 2009;205(4):255–258. doi:10.1016/j.prp.2008.10.011, PMID:19091488.