Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Construction and Validation of a Diagnostic Model for Gastroesophageal Reflux Disease Based on Pyroptosis-Related Genes

Lijian Liu
The First Affiliated Hospital of Guangxi University of Traditional Chinese Medicine
Wenyu Wu
Guangxi University of Chinese Medicine
Xiaoyun Lei
Guangxi University of Chinese Medicine
Sheng Xie
The First Affiliated Hospital of Guangxi University of Traditional Chinese Medicine
Liqun Li
lilq@gxtcmu.edu.cn

The First Affiliated Hospital of Guangxi University of Traditional Chinese Medicine

Research Article

Keywords: Gastroesophageal reflux disease, Pyroptosis, Machine learning, Diagnostic model, Experimental validation

Posted Date: October 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10885979/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background:

GERD is a common chronic digestive disorder, yet its diagnosis remains challenging due to the lack of objective molecular markers are lacking. Pyroptosis may contribute to esophageal mucosal injury, but a systematic search for pyroptosis-related diagnostic genes in GERD has not been performed. In this study, we used integrated bioinformatics, machine learning, and experimental validation to identify pyroptosis-associated diagnostic biomarkers for GERD and examine their relationship with the esophageal immune microenvironment.

Methods:

Esophageal transcriptome data from patients with GERD were obtained from the GEO database. We intersected differentially expressed genes, WGCNA module genes, and a curated set of pyroptosis-related genes to obtain candidate genes. We constructed a diagnostic model from 113 combinations of machine learning algorithms and validated it in an independent dataset using an artificial neural network. Immune cell infiltration was estimated with CIBERSORT, and the key bioinformatics results were further checked in a rat model of GERD.

Results:

The machine learning pipeline identified five pyroptosis-related genes with diagnostic value: GABARAPL1, CD55, CA9, SLC4A2, and LCN2. Among these, GABARAPL1 and CD55 performed best individually, with AUCs of 0.876 and 0.862. When the two genes were fed into an artificial neural network, the AUC reached 0.919 in the training cohort and 0.900 in the test cohort. Immune profiling showed that GERD esophageal mucosa contained more resting/activated dendritic cells, M2 macrophages, and activated NK cells, but fewer $CD8^+$ T cells and γδ T cells. The animal experiments confirmed that GABARAPL1 was downregulated and CD55 was upregulated in GERD tissues.

Conclusion:

These observations point to GABARAPL1 and CD55 as potential pyroptosis-related diagnostic biomarkers for GERD. They may also influence disease pathogenesis through immunomodulatory mechanisms.

Introduction

Gastroesophageal reflux disease (GERD) occurs when gastroduodenal contents reflux into the esophagus and cause troublesome symptoms or complications[1, 2]. It affects around 14% of the global

population[3], lowers quality of life, and is an established risk factor for esophageal adenocarcinoma[4]. Diagnosis is often difficult in practice: 30%–50% of patients show no mucosal breaks on endoscopy[5], and pH-impedance monitoring, though regarded as the reference standard, is invasive and technically demanding. There is therefore a clear need for objective, noninvasive molecular markers to support early detection and individualized treatment.

The mucosal inflammation seen in GERD is complex and has a strong immune component[6, 7]. Pyroptosis is a regulated cell death pathway activated by caspases such as caspase-1/4/5/11. Cleavage of gasdermin family members creates membrane pores and triggers the release of IL-1β and IL-18, which together drive a potent inflammatory response[8–10]. Several lines of evidence suggest that pyroptosis of esophageal epithelial cells contributes to GERD pathology[11, 12], yet the key pyroptosis-related genes and their underlying mechanisms remain poorly defined. Immune cell-mediated inflammation also plays a central role in esophageal injury[13]; whether pyroptosis and immune responses act independently or are functionally connected during GERD progression is still unclear.

Transcriptomic datasets from GERD patients offer a way to identify pyroptosis-related diagnostic markers, but extracting reliable markers from high-dimensional expression data is not straightforward. Conventional methods, such as differential expression analysis and WGCNA, are often affected by batch effects and multicollinearity and may retain redundant features. Machine learning methods can capture complex patterns in high-dimensional data and are advantageous for building diagnostic models and selecting biomarkers[14]. CIBERSORT estimates the relative abundance of 22 immune cell types in tissue and is widely used to study immune infiltration[15, 16]. The combination of bioinformatics, machine learning, immune profiling, and experimental validation has identified key markers in several diseases, but this approach has not yet been applied to pyroptosis-related biomarkers in GERD. Here, we used this integrated strategy to identify pyroptosis-associated diagnostic markers in GERD, examine their biological functions and relationship to the immune microenvironment, and provide potential tools for more precise diagnosis.

Materials and Methods

Study Design

The overall workflow is summarized in Fig. 1. We obtained four esophageal tissue microarray datasets (GSE13083, GSE39491, GSE34619, and GSE26886) from GEO and corrected them for batch effects. Differentially expressed genes (DEGs) were selected with $|\log_2 FC| \geq 1.25$ and adjusted $P < 0.05$, and key coexpression modules were identified through WGCNA. The intersection of DEGs, key module genes, and a curated pyroptosis gene set was taken as the candidate gene pool. GSE13083 and GSE39491 were combined into a training cohort, and GSE34619 was used as a test set for model selection. A total of 113 combinatorial machine learning algorithms were applied with these candidates as input features to build diagnostic models. Core genes were defined as those shared by the three models achieving the highest

area under the curve (AUC). Single-gene receiver operating characteristic (ROC) analysis was performed in the training set, and targets with AUC ≥ 0.85 were designated as high-priority biomarkers. An artificial neural network was then constructed using these biomarkers, and its joint diagnostic performance was evaluated on the external test set GSE26886. For the high-priority biomarkers, we conducted multivariate logistic regression and nomogram generation, as well as GSEA, GSVA, and CIBERSORT immune infiltration analyses. Subsequently, AI-based drug prediction, molecular docking, and molecular dynamics simulations were conducted. A GERD rat model was established to confirm target expression via HE staining, immunohistochemistry, qPCR, and Western blot. The clinical data analysis was retrospective and exempt from ethics approval; the animal experiment protocol was approved by the Animal Ethics Committee of Guangxi University of Chinese Medicine (approval number: DW20230320-05).

Data Acquisition and Preprocessing

Gene expression profiles were downloaded from the Gene Expression Omnibus (GEO). We selected microarray datasets according to the following criteria: (1) human esophageal tissue; (2) microarray platform; (3) inclusion of GERD patients and healthy controls; and (4) at least three samples per group and no fewer than six samples overall. Four datasets—GSE13083, GSE39491, GSE34619, and GSE26886—met these requirements and were retained for analysis; details are given in Supplementary Table S1. Batch effects were corrected with the “sva” package, and data quality was checked visually using box plots and principal component analysis (PCA).

Pyroptosis-associated genes were retrieved from the GeneCards database. After ranking by relevance score, we selected the top 20% of genes as the candidate pyroptosis-related gene set.

Identification of Differentially Expressed Genes and WGCNA

Differential expression between GERD patients and healthy controls in the training set was assessed with the “limma” package. Genes were considered differentially expressed if $|\log_2$ fold change ($\log_2$FC)$| \geq 1.25$ and adjusted $P < 0.05$. Heatmaps and volcano plots were generated with “pheatmap” and “ggplot2,” respectively. We then built a gene co-expression network using “WGCNA” to identify the module most strongly associated with the clinical trait. The soft-thresholding power was set to 5 based on the pickSoftThreshold function; the adjacency matrix was converted into a topological overlap matrix (TOM). Hierarchical clustering of TOM dissimilarity with dynamic tree cutting (minimum module size of 60 genes) was applied to detect modules, and the module with the highest correlation and statistical significance was designated as the key module.

Construction of the Pyroptosis Gene Diagnostic Model and Evaluation of Model Genes

With the key module defined, we combined GSE13083 and GSE39491 into a training cohort and used GSE34619 as an independent validation set. Candidate pyroptosis-related genes were then obtained by intersecting the DEGs, the key module genes, and the curated pyroptosis gene set using the “ggplot2” and “ggvenn” packages.Taking these candidate genes as input features, 113 combinatorial machine learning algorithms (detailed in Supplementary Table S2) were applied to construct diagnostic models. Model performance was assessed using the area under the receiver operating characteristic curve (AUC) and confusion matrix metrics. The genes common to the three models achieving the highest AUC values were then intersected to derive stable hub genes. Single-gene diagnostic capacity of these hub genes was evaluated by ROC curves plotted with the “pROC” package. Genes with an AUC ≥ 0.7 were model genes, and their expression levels in GERD samples were analyzed with the “ggpubr” package and visualized through box plots.

Independent Validation Using an Artificial Neural Network

To evaluate the joint diagnostic ability of the most promising genes, we selected those with AUC ≥ 0.85 as input features for an artificial neural network (ANN). The merged GSE13083 and GSE39491 datasets served as the training set, and GSE26886 was used as an independent external test set. We built a feedforward neural network with three hidden layers using the “neuralnet” package, with a sigmoid activation function, a maximum of 1×10^6 iterations, and a convergence threshold of 0.01. Discrimination was assessed by ROC curves from the “pROC” package, and AUCs with 95% confidence intervals were calculated. The network structure was visualized with the NeuralNetTools package.

Multivariate Regression Analysis and Nomogram Construction

Multivariate logistic regression was carried out in R 4.2 using the glm function with maximum likelihood estimation, and the results were displayed as a forest plot. Genes with AUC ≥ 0.85 were entered into the logistic model, and a nomogram was constructed with the rms package. We used a calibration curve to examine agreement between predicted and observed outcomes, and decision curve analysis (rmda package) to estimate net clinical benefit. Two-sided P values < 0.05 were considered statistically significant.

Biological Function of Signature Genes and Their Association with Immune Cells

For genes with AUC ≥ 0.85, we generated a protein–protein interaction network with GeneMANIA. Pathway activity was evaluated by GSVA, and GSEA was performed using the clusterProfiler package. Immune cell composition was estimated from the expression data with CIBERSORT for 22 cell types. Differences in infiltration between groups were compared using the Wilcoxon test and shown as box plots. Spearman correlation was used to examine relationships among immune cell subsets and associations between gene expression and immune cell proportions; the results were visualized with corrplot and linkET. A two-tailed P < 0.05 was considered statistically significant.

AI-Based Drug Prediction, Molecular Docking, and Molecular Dynamics Simulation

To look for potential therapeutic candidates, we merged GSE13083 and GSE39491 and applied the AI-driven drug screening framework DrugRefLector to identify small molecules predicted to bind the signature gene targets stably. Molecular docking and molecular dynamics (MD) simulations were then performed to assess binding affinity. Crystal structures of the target proteins were obtained from the RCSB Protein Data Bank (PDB), and ligand structures were retrieved from PubChem. Protein and ligand structures were prepared with AutoDock 4.2.6, and docking was carried out with AutoDock Vina 1.2.7. Conformations with binding energies below − 7.0 kcal/mol were visually inspected in PyMOL, and the best pose was selected for MD simulation. MD simulations were run with GROMACS 2023.2. The ligand was parameterized under the GAFF force field, the protein under amber14SB, and water was modeled as TIP3P. A cubic water box was built with a minimum solute–wall distance of 1.0 nm. After steepest descent energy minimization, the system underwent 100 ps of NVT equilibration at 310 K and 100 ps of NPT equilibration, followed by a 100 ns production run with a 2 fs time step.

Animal Experimental Validation

Animal Model Establishment

Twelve SPF-grade Sprague-Dawley rats (six males and six females; 160–200 g) were obtained from Hunan Slaike Jingda Laboratory Animal Co., Ltd. (license No.: SCXK (Xiang) 2019-0004). After seven days of acclimatization in the SPF animal facility at the First Affiliated Hospital of Guangxi University of Chinese Medicine, the animals were randomly assigned to control and model groups (n = 6 each) using a random number table. The GERD model was induced based on an optimized protocol from reference[17]. Briefly, rats received daily intragastric administration (0.2 mL per animal) of either 0.1% iodoacetamide solution (I8010, Solarbio) or acidic saline (pH 5.0) containing pepsin (10 mg/mL, P7000, Sigma) and deoxycholic acid (10 mg/mL, D8460, Solarbio), with the two solutions alternated every other day for 14 consecutive days. The control group received an equal volume of normal saline on the same schedule. All animals were sacrificed 24 hours after the last gavage, and tissue samples were collected. Rats were deeply anesthetized with 2% isoflurane delivered in 33% oxygen (balanced with air) using a calibrated vaporizer. Depth of anesthesia was confirmed by absence of pedal withdrawal reflex. Under deep anesthesia, exsanguination was performed via transection of the abdominal aorta to ensure death without recovery of consciousness.

HE Staining

Tissue specimens were fixed, dehydrated through graded ethanol, cleared in xylene, and embedded in paraffin. Sections (3 μm) were stained with hematoxylin and eosin (BA-4041, BA-4024; Baso), dehydrated, cleared, and mounted. Stained sections were examined using an Olympus BX43 light microscope.

Immunohistochemistry

Paraffin sections were deparaffinized, rehydrated, and subjected to antigen retrieval in citric acid buffer. Endogenous peroxidase was blocked with 3% methanol–hydrogen peroxide, and non-specific binding was blocked with 5% BSA. Sections were incubated overnight at 4°C with primary antibodies against CD55 (AF5259, 1:100) and GABARAPL1 (83817-1-RR, 1:120), followed by HRP-conjugated secondary antibody (PV-6000) at 37°C for 20 min. DAB was used for visualization, and nuclei were counterstained with hematoxylin. Positive staining was brownish-yellow; mean optical density was quantified using Image-Pro Plus software.

Real-Time Quantitative PCR

Total RNA was isolated with TRIzol (15596-026, Invitrogen) and reverse transcribed using the HiFiScript cDNA First Strand Synthesis Kit (CW2582M, CWBIO). qPCR was conducted with SYBR Green I master mix (MQ00401S, Monad Biotech). GAPDH was the internal reference, and relative expression was calculated using the $2^{\wedge -\Delta\Delta Ct}$ method. Primer sequences are listed in Supplementary Table S3.

Western Blot

Proteins were extracted with RIPA buffer (R0020, Solarbio) containing PMSF. Protein concentration was measured, and 20 μg per lane was separated by SDS-PAGE and transferred to PVDF membranes (IPVH00010, Millipore). Membranes were blocked with 5% non-fat milk for 1 hour at room temperature and then incubated overnight at 4°C with primary antibodies against CD55 (loading control β-tubulin; Affinity, 1:10000) and GABARAPL1 (loading control GAPDH; Affinity, 1:10000). After washing, membranes were incubated with HRP-conjugated secondary antibodies (ZB2301/ZB2305, ZSGB-BIO, 1:3000) for 1 hour at room temperature. Signals were developed with ECL (BMU102-CN, Abbkine) and exposed to film. Band intensities were quantified with ImageJ and normalized to the loading control.

Statistical Analysis

Statistical analysis of the animal experiment data was performed using SPSS 22.0. Quantitative data are expressed as mean ± standard deviation ($\bar{x} \pm s$). For Western blot analysis, six independent biological replicates per group (n = 6 per group) were used. For all other experimental assays, at least three independent biological replicates were performed. The Shapiro-Wilk test was used to assess normality, and Levene's test was employed to evaluate homogeneity of variances. When data met the assumptions of normality and equal variance, group comparisons were made using the independent samples t-test; otherwise, the Mann–Whitney U test was applied. A two-tailed P value < 0.05 was considered statistically significant.

Results

Identification of Differentially Expressed Genes and WGCNA

Both the box plot (Fig. S1A) and PCA plot (Fig. S1B) demonstrated that batch effects have been effectively eliminated. Using the criteria of $|\log_2 FC| \geq 1.25$ and adjusted $P < 0.05$, a total of 258 significant DEGs were identified, including 178 up-regulated and 80 down-regulated genes (Supplementary Table S4). The volcano plot illustrated the differential distribution of the top 20 DEGs (Fig. 2A), and the heatmap revealed specific clustered expression of the top 40 DEGs between the two sample groups (Fig. 2B).

WGCNA was performed with a soft-thresholding power of 5 (Fig. 2C) and identified eight gene co-expression modules (Fig. 2D). As shown in Fig. 2E, the blue module displayed the highest correlation coefficient with the GERD phenotype (r = -0.53, $P = 5\times10^{-8}$), and its module membership exhibited a strong positive correlation with gene significance (r = 0.84, P = 2.8e-200, Fig. 2F).

Construction and Evaluation of the Gene Diagnostic Model

Six pyroptosis-related DEGs were identified by overlapping the blue module, the DEG set, and the pyroptosis gene list: GABARAPL1, CD55, CA9, SLC4A2, LCN2, and AZGP1 (Fig. 3A). Using these six candidates as input features, screening of 113 algorithm combinations yielded 12 diagnostic models (Supplementary Table S5, Fig. 3B). Ranked by AUC, the top three algorithms were Random Forest combined with Gradient Boosting Machine (RF + GBM), Gradient Boosting Machine (GBM), and forward stepwise generalized linear model (Stepglm [forward]) (Fig. 3B). All three models achieved AUCs exceeding 0.85, with favorable sensitivity and specificity (Fig. 3C–E). Extraction of the genes shared by these three models yielded the same six genes: GABARAPL1, CD55, CA9, SLC4A2, LCN2, and AZGP1 (Fig. 3F).

Model Gene Evaluation and Validation of the Artificial Neural Network

We next evaluated the individual diagnostic performance of these six candidate genes by ROC analysis. Five genes—GABARAPL1, CD55, CA9, SLC4A2, and LCN2—had AUCs above 0.7, whereas AZGP1, with an AUC of 0.695, was excluded from further analysis (Fig. 4A). Among the retained genes, GABARAPL1 and CD55 showed the strongest discrimination, both with AUCs above 0.85. In GERD samples, CD55 and LCN2 were upregulated, while GABARAPL1, CA9, and SLC4A2 were downregulated (Fig. 4B). We then used GABARAPL1 and CD55 as inputs for an ANN with three hidden neurons (Fig. 4C, Fig. S2A–B). The model achieved an AUC of 0.919 (95% CI 0.855–0.969) in the training cohort and 0.900 (95% CI 0.771–0.997) in the independent test set (Fig. 4D, E).

Multivariate Regression Analysis and Nomogram Construction

In the multivariate model, CD55 was an independent risk factor for GERD (OR 2.277, P = 0.013), whereas GABARAPL1 was protective (OR 0.283, P = 0.006; Fig. 5A). The calibration curve showed close agreement between predicted and observed probabilities (Fig. 5B), and decision curve analysis indicated

a favorable net benefit across threshold probabilities (Fig. 5C). Using these two genes, we constructed a nomogram that converts expression values into an individual GERD risk score (Fig. 5D).

Biological Functions of Signature Genes and Their Association with Immune Cells

The PPI network linked GABARAPL1 and CD55 to core autophagy proteins (ATG4D, ATG7) and key complement components (C3, CFB), and the two genes themselves were negatively correlated (Fig. 6A). GSEA and GSVA indicated that both genes were negatively enriched in the cell cycle and oxidative phosphorylation pathways. CD55 was positively enriched in immune-related pathways such as NOD-like receptor and cytosolic DNA sensing, while GABARAPL1 showed enrichment in complement cascade and autophagy-related pathways (Fig. 6B–C, Supplementary Tables S6–7). These results point to a joint role for these genes in remodeling the immune-inflammatory microenvironment through pyroptosis–autophagy interactions.

Relative to healthy tissue, GERD samples had higher proportions of resting and activated dendritic cells, M2 macrophages, activated NK cells, and follicular helper T cells, but lower proportions of $CD8^+$ T cells and γδ T cells (Fig. 6D). Resting and activated dendritic cells were positively correlated, whereas resting and activated NK cells were negatively correlated (Fig. 6E). GABARAPL1 expression correlated negatively with resting/activated dendritic cells and positively with plasma cells and resting mast cells; CD55 showed the opposite pattern for dendritic cells and plasma cells (Fig. 6F).

AI-Based Drug Prediction, Molecular Docking, and Molecular Dynamics Simulation

AI-based virtual screening identified isoevodiamine (BRD-A68631409) as a lead candidate compound (Fig. 7A); the full ranked list of candidate compounds is provided in Supplementary Table S8. Molecular docking analysis showed that isoevodiamine could specifically bind to both GABARAPL1 and CD55, with optimal binding free energies of − 8.988 kcal/mol and − 7.382 kcal/mol, respectively. The compound formed a 2.2 Å hydrogen bond with Asn81 at the GABARAPL1 dimer interface and a 2.1 Å hydrogen bond with Arg65 within the extracellular β-sheet region of CD55; a single dominant binding mode was observed for each target (Fig. 7B, C). Full docking scores are provided in Fig. S3A. Based on the optimal docking conformations, 100 ns molecular dynamics simulations were performed. The RMSD profiles demonstrated that the backbones of both complexes reached a stable equilibrium in the later phase of simulation, with the ligand persistently anchored to the key interacting residues (Fig. 7D). The Gibbs free energy landscapes of both systems showed a single low-energy basin, indicating conformational convergence and stable thermodynamics of the complexes (Fig. 7E). Additional dynamic trajectories, including hydrogen bond timelines, radius of gyration (Rg), SASA, and residue-level RMSF, are provided in Fig. S3B–E. Collectively, these results suggest that isoevodiamine can target both GABARAPL1 and CD55, with stronger affinity for GABARAPL1 and multi-target regulatory potential.

Animal Experimental Validation

In the model group, HE staining showed basal cell hyperplasia, disordered epithelial architecture, mucosal erosion, dense inflammatory infiltrates, papillary elongation, and tissue edema, confirming that the GERD model was successfully established (Fig. 8A). Immunohistochemistry revealed stronger CD55 staining and weaker GABARAPL1 staining in the model group than in controls (Fig. 8B, C). Quantitative IOD analysis confirmed that CD55 was upregulated ($P < 0.01$) and GABARAPL1 was downregulated ($P < 0.0001$) (Fig. 8D). qPCR yielded similar results: CD55 mRNA was higher ($P < 0.01$) and GABARAPL1 mRNA was lower ($P < 0.001$) in the model group (Fig. 8E). Western blotting also showed increased CD55 and decreased GABARAPL1 protein levels ($P < 0.0001$; Fig. 8F, G), consistent with the mRNA data.

Discussion

Pyroptosis is an inflammatory form of programmed cell death dependent on the activation of caspase family proteins and is critically involved in various inflammatory diseases[8–10]. Evidence indicates that pyroptosis-mediated esophageal epithelial injury contributes significantly to GERD pathogenesis[11, 12]. Nevertheless, systematic identification of pyroptosis-related diagnostic markers and their connection to the immune microenvironment in GERD is lacking. Therefore, the present study approached GERD from a pyroptosis perspective, screened core pyroptosis diagnostic genes, constructed a diagnostic model, and explored underlying biological mechanisms while also predicting potential targeted therapeutics. By intersecting differentially expressed genes, key WGCNA module genes, and a curated set of pyroptosis-associated genes, we obtained six candidates. Following screening with 113 combinatorial machine learning algorithms and validation through an artificial neural network, GABARAPL1 and CD55 were ultimately identified as key genes, yielding AUCs of 0.876 and 0.862, respectively, suggesting their promise as potential pyroptosis-related diagnostic biomarkers for GERD.

Ample evidence demonstrates that autophagy suppresses excessive NLRP3 inflammasome activation by clearing damaged mitochondria, thereby preserving epithelial barrier integrity[18–20]. GABARAPL1, a member of the ATG8 family, is essential for autophagosome formation[21, 22]. In a myocardial infarction model, silencing GABARAPL1 aggravated hypoxia-triggered endothelial pyroptosis[23]; knockout of the family member GABARAP markedly enhanced NLRP3, caspase-1, and IL-1β pathways in mouse macrophages[24]; moreover, GABARAPL1 restricts the overactivation of caspase-11, a key component of the non-canonical inflammasome, and its loss dramatically exacerbates pyroptosis and IL-1β release[25]. Collectively, these findings indicate that GABARAPL1 exerts a negative regulation on pyroptosis, which is consistent with the downregulation of GABARAPL1 observed in GERD. GSEA and GSVA enrichment analyses further provided pathway-level support, showing that GABARAPL1 was positively enriched in autophagy regulation and complement cascade pathways, whereas it was negatively enriched in oxidative phosphorylation and cell cycle pathways. Based on this evidence, we speculate that GABARAPL1 downregulation may impair the autophagic clearance of damaged mitochondria, leading to the accumulation of mitochondrial reactive oxygen species (ROS) accumulation and sustained NLRP3 inflammasome activation; concurrently, dysregulation of complement cascade control could further amplify inflammatory signals. These hypotheses require confirmation in future investigations.

CD55 acts as a negative regulator of complement by inhibiting C3/C5 convertase activity, which protects host cells from complement-mediated lysis [26]. In our rat model, CD55 was clearly upregulated in esophageal tissue, in line with the bioinformatic findings and earlier reports[27]. Notably, CD55 overexpression has also been linked to a higher risk of esophageal adenocarcinoma[28], suggesting that this molecule may be involved in the transition from GERD to malignancy. At the mechanistic level, complement anaphylatoxins such as C5a and C3a maintain complement activation[29], and complement components can directly prime the NLRP3 inflammasome. This process leads to caspase-1 activation and the release of IL-1β and IL-18, thereby creating a self-amplifying inflammatory loop[30]. Our enrichment analyses support this interpretation, as CD55 showed positive enrichment in the NOD-like receptor signaling pathway. Thus, although elevated CD55 may restrain excessive complement activation, it could also participate in pyroptotic regulation in GERD through the complement–NOD-like receptor axis.

Immune infiltration analysis revealed a distinct cellular profile in GERD esophageal tissues: higher proportions of activated NK cells, activated dendritic cells, M2 macrophages, and follicular helper T cells, together with fewer $CD8^+$ T cells and γδ T cells. This pattern points to a disturbed microenvironment in which pro-inflammatory/pro-fibrotic signals are enhanced while anti-tumor cytotoxic immunity is weakened. The anatomical features of the esophageal mucosa may partly account for these observations. Because the mucosa lacks secondary lymphoid structures, dendritic cells encounter luminal antigens directly and often promote tolerance rather than protective immunity[31]. Under steady-state conditions, mucosal dendritic cells tend to ignore antigens or induce regulatory responses[32]. One earlier study reported fewer dendritic cells in GERD mucosa than in healthy tissue[33], which differs from our results. This discrepancy likely reflects differences in GERD subtypes and patient populations. M2 macrophages contribute to tissue repair and immunosuppression[34]. In active eosinophilic esophagitis, polarization toward an M2 phenotype is considered to complicate treatment[35]. In early reflux esophagitis, M1 cells infiltrate first and activate Stat3 signaling; subsequently, M2 cells cooperate with regulatory T cells to favor esophageal cancer progression[36]. Normal esophageal intraepithelial lymphocytes are predominantly $CD8^+$ [31]; their numbers increase with erosive esophagitis severity but decline as Barrett's esophagus develops, owing to a shift toward humoral immunity[33]. This pattern is consistent with our observations. Immature dendritic cells can also stimulate NK cell proliferation and activation through direct cell contact[37], which may explain the simultaneous increase in resting dendritic cells and activated NK cells. Since M2 macrophages and activated NK cells are important sources of IL-1β and IL-18, the altered proportions of these cells may help shape the pyroptosis-related inflammatory microenvironment in GERD.

We next explored how the two model genes relate to specific immune cell subsets. GABARAPL1 showed a negative correlation with resting and activated dendritic cells, whereas CD55 was positively correlated with the same populations. From a functional perspective, GABARAPL1 can deliver antigens to dendritic cells through autophagy to initiate T cell responses[38], while CD55 helps maintain dendritic cells in an

immature state by suppressing pro-inflammatory signals[39]. These observations suggest that reduced GABARAPL1 may weaken autophagy-dependent antigen presentation and impair immune surveillance, whereas increased CD55 may reinforce local immunosuppression by keeping dendritic cells immature. Although the two changes operate through different mechanisms, they appear to converge on a common outcome: pushing the GERD microenvironment toward immune tolerance. This view is consistent with reports that soluble factors from Barrett's epithelium induce dendritic cell tolerance and support immune escape[40].

GERD diagnosis, particularly for endoscopy-negative cases, remains challenging[41]. Existing serum markers, such as pepsinogen, exhibit unsatisfactory accuracy[42, 43]. The ANN diagnostic model constructed using GABARAPL1 and CD55 reliably discriminated patients with GERD from healthy individuals, achieving an AUC of 0.900 in independent external validation. In addition, we transformed the expression values of these two genes into an individualized risk score via a nomogram, and decision curve analysis confirmed favorable net clinical benefit. Therefore, evaluating GABARAPL1 and CD55 expression in esophageal tissue holds promise as a novel auxiliary diagnostic method for GERD. Considering that the invasive nature of tissue biopsy somewhat restricts its clinical applicability, while liquid biopsy offers non-invasiveness and repeatability, future research could explore the diagnostic potential of these two markers in peripheral blood or circulating exosomes, though this will require confirmation in prospective multicenter clinical cohorts.

Using AI-driven drug prediction, we identified the natural alkaloid isoevodiamine as a candidate compound. Isoevodiamine is a quinazoline carboline alkaloid derived from the traditional Chinese medicine *Evodia rutaecarpa*. Although no studies on its role in GERD have been reported to date, our molecular docking and dynamics simulations demonstrated that isoevodiamine stably binds to both GABARAPL1 and CD55, suggesting its potential to intervene at the autophagy–pyroptosis crosstalk node. Isoevodiamine is a structural analog of evodiamine, which has been shown to suppress NLRP3 inflammasome assembly in colitis models and modulate macrophage polarization in GERD models[44, 45]. This structural relationship provides indirect support for inferring the functions of isoevodiamine, but its actual pharmacological effects still need to be verified through in vitro and in vivo experiments. If subsequent studies confirm its efficacy, isoevodiamine would represent a structurally novel candidate molecule for GERD treatment targeting the autophagy–pyroptosis axis.

Several limitations should be kept in mind. The sample size from public databases is modest, and inter-individual variation may affect generalizability. All human cohorts were retrospective, so prospective multicenter studies are needed to assess clinical utility. Animal experiments confirmed that GABARAPL1 and CD55 are associated with GERD, but the precise mechanisms by which they regulate pyroptosis and the immune microenvironment remain unclear and will require functional experiments such as knockout/overexpression and co-immunoprecipitation. Isoevodiamine offers a structural lead for drug development, but its actual efficacy and direction of action have not yet been tested in cells or animals.

Conclusion

In conclusion, we built a pyroptosis-related diagnostic model for GERD and found that downregulation of GABARAPL1 and upregulation of CD55 are associated with an immunosuppressive esophageal microenvironment. These results offer a new molecular diagnostic tool for GERD and a rationale for therapies aimed at the pyroptosis–autophagy crosstalk.

Abbreviations

GERD: Gastroesophageal reflux disease;

GABARAPL1: GABA type A receptor-associated protein like 1;

CD55: Complement decay-accelerating factor;

AUC: Area under the curveANN: Artificial neural network;

DEG: Differentially expressed gene;

GEO: Gene Expression Omnibus;

WGCNA: Weighted gene co-expression network analysis;

GSEA: Gene Set Enrichment Analysis;

GSVA: Gene Set Variation Analysis;

ROC: Receiver operating characteristic.

Declarations

Ethics approval and consent to participate

The animal experiment protocol was approved by the Animal Ethics Committee of Guangxi University of Chinese Medicine (approval number: DW20230320-05). This study utilized publicly available, de-identified datasets (GSE13083, GSE39491, GSE34619, and GSE26886) from the GEO database. No additional ethical approval or informed consent was required for this retrospective bioinformatics analysis.

Consent for publication

Not applicable.

Availability of data and materials

The datasets GSE13083, GSE39491, GSE34619, and GSE26886 analyzed during the current study are publicly available in the Gene Expression Omnibus (GEO) repository under accession numbers GSE13083, GSE39491, GSE34619, and GSE26886 (https://www.ncbi.nlm.nih.gov/geo/). The pyroptosis-associated gene set was obtained from the GeneCards database (https://www.genecards.org/). All other data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 82560908];the Construction Project of the National Traditional Chinese Medicine Inheritance and Innovation Center (Research direction of predominant diseases of gastroesophageal reflux disease) [grant number 2023019-05];the Natural Science Foundation of Guangxi Province [grant numbers 2026GXNSFAA00641081].

Authors' contributions

LJ Liu: Conceptualization, Data curation, Investigation, Software, Writing original draft, Funding acquisition. WY Wu: Conceptualization, Data curation, Formal analysis, Methodology, Writing original draft. S Xie and XY Lei: Supervision, Validation, Writing original draft. LQ Li: Supervision, Validation, Writing review and editing, Project administration. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the researchers and study participants for their contributions to the GWASs, R packages, and datasets used in this study.

References

1. China International Exchange and Promotive Association for Medical and Health Care, Spleen and Stomach Disease Branch of Integrated Traditional Chinese and Western Medicine, Xie S et al. Consensus on integrated traditional Chinese and Western medicine collaborative diagnosis and treatment of gastroesophageal reflux disease. World J Tradit Chin Med. 2025:1–21. (In Chinese).
2. Dunbar KB. Gastroesophageal reflux disease. Ann Intern Med. 2024;177(8):ITC113–28.
3. Wickramasinghe N, Devanarayana NM. Insight into global burden of gastroesophageal reflux disease: Understanding its reach and impact. World J Gastrointest Pharmacol Ther. 2025;16(1):97918.

4. Digestive System Disease Committee of Chinese Association of Integrative Medicine. Expert consensus on integrated traditional Chinese and Western medicine diagnosis and treatment of gastroesophageal reflux disease (2025). Chin J Integr Tradit West Med Dig. 2025;33(3):217–29. (In Chinese).
5. Wang F, Li G, Ning J, Chen L, Xu H, Kong X, et al. Alcohol accumulation promotes esophagitis via pyroptosis activation. Int J Biol Sci. 2018;14(10):1245–55.
6. Kandulski A, Malfertheiner P. Gastroesophageal reflux disease–from reflux episodes to mucosal inflammation. Nat Rev Gastroenterol Hepatol. 2011;9(1):15–22.
7. Rieder F, Biancani P, Harnett K, Yerian L, Falk GW. Inflammatory mediators in gastroesophageal reflux disease: impact on esophageal motility, fibrosis, and carcinogenesis. Am J Physiol Gastrointest Liver Physiol. 2010;298(5):G571–81.
8. Zhou B, Abbott D. Gasdermins, executors of pyroptosis: a decade in perspective. Annu Rev Pathol. 2026;21(1):163–84.
9. Zhao H, Fu X, Zhang Y, Liu C, Wang J, Li M, et al. The role of pyroptosis and autophagy in the nervous system. Mol Neurobiol. 2024;61(3):1271–81.
10. Kayagaki N, Webster JD, Newton K. Control of cell death in health and disease. Annu Rev Pathol. 2024;19:157–80.
11. Liu J, Xiao Y, Xu Q, Chen L, Zhang W, Yang H, et al. Britannilactone 1-O-acetate induced ubiquitination of NLRP3 inflammasome through TRIM31 as a protective mechanism against reflux esophagitis-induced esophageal injury. Chin Med. 2024;19(1):118.
12. Nadatani Y, Huo X, Zhang X, Yu C, Cheng E, Zhang Q, et al. NOD-like receptor protein 3 inflammasome priming and activation in Barrett's epithelial cells. Cell Mol Gastroenterol Hepatol. 2016;2(4):439–53.
13. Rochman M, Kellerman K, Jankowski MP, Caldwell JM, Rothenberg ME. The oesophagus as an immune organ. Nat Rev Gastroenterol Hepatol. 2025;22(9):657–67.
14. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022;23(1):40–55.
15. Alonso-Moreda N, Berral-González A, De La Rosa E, Sánchez-Santos JM, González-Buitrago JM. Comparative analysis of cell mixtures deconvolution and gene signatures generated for blood, immune and cancer cells. Int J Mol Sci. 2023;24(13):10765.
16. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–7.
17. Xie S, Han BJ, Wang JX, Chen SY, Li XQ, Liu LJ et al. Mechanism of back-shu point acupuncture intervention in GERD rats based on NLRP3/Caspase-1 pathway-mediated esophageal ICC pyroptosis. Liaoning J Tradit Chin Med. 2025:1–13. (In Chinese).
18. Kundu S, Chauhan S, Dhar K, Saha S. Chats and spats of autophagy and innate immune systems. J Mol Biol. 2025;437(21):169387.

19. Klochkova A, Karami AL, Fuller AD, Muir AB, Nakagawa H, Wang DH, et al. Autophagy contributes to homeostasis in esophageal epithelium where high autophagic vesicle level marks basal cells with limited proliferation and enhanced self-renewal potential. Cell Mol Gastroenterol Hepatol. 2024;18(1):15–40.
20. Kong J, Whelan KA, Laczkó D, Dany M, DeMauro G, Duan S, et al. Autophagy levels are elevated in barrett's esophagus and promote cell survival from acid and oxidative stress. Mol Carcinog. 2016;55(11):1526–41.
21. Meng J, Xu L, Ma B, Li Y, Wang H, Zhang Z, et al. GABARAPL1 is essential for ACR-induced autophagic cell death of mouse Leydig cells. Ecotoxicol Environ Saf. 2025;289:117426.
22. Chen J, Zhao H, Liu M, Wang L, Chen Y, Zhang X, et al. A new perspective on the autophagic and non-autophagic functions of the GABARAP protein family: a potential therapeutic target for human diseases. Mol Cell Biochem. 2024;479(6):1415–41.
23. Liu C, Yan C, Liu D, Wang Y, Zhang H, Li S, et al. GABARAPL1 exerts regulatory effects on hypoxia-induced pyroptosis in the pathogenesis of myocardial infarction. J Cell Mol Med. 2025;29(6):e70469.
24. Zhang Z, Xu X, Ma J, Wu J, Wang Y, Zhou R, et al. Gene deletion of Gabarap enhances Nlrp3 inflammasome-dependent inflammatory responses. J Immunol. 2013;190(7):3517–24.
25. Sakaguchi N, Sasai M, Bando H, Lee YJ, Núñez G, Yamamoto M. Role of Gate-16 and Gabarap in prevention of caspase-11-dependent excess inflammation and lethal endotoxic shock. Front Immunol. 2020;11:561948.
26. Luo Q, Mao Y, Yu R, Chen S, Li J, Wang X, et al. CD55 may be a new target for colorectal cancer treatment. Sci Rep. 2025;15(1):23086.
27. Hiraoka S, Mizuno M, Nasu J, Okada H, Inaba T, Yamamoto K, et al. Enhanced expression of decay-accelerating factor, a complement-regulatory protein, in the specialized intestinal metaplasia of Barrett's esophagus. J Lab Clin Med. 2004;143(4):201–6.
28. Murao T, Shiotani A, Fujita Y, Matsumoto H, Hori K, Haruma K, et al. Overexpression of CD55 from Barrett's esophagus is associated with esophageal adenocarcinoma risk. J Gastroenterol Hepatol. 2016;31(1):99–106.
29. Suresh R, Chandrasekaran P, Sutterwala FS, Mosser DM. Complement-mediated 'bystander' damage initiates host NLRP3 inflammasome activation. J Cell Sci. 2016;129(9):1928–39.
30. Janžič L, Kouter K. Complement and inflammasome crosstalk in chronic inflammation. Front Immunol. 2026;17:1778759.
31. Kaymak T, Hruz P, Niess JH. Immune system and microbiome in the esophagus: implications for understanding inflammatory diseases. FEBS J. 2022;289(16):4758–72.
32. Iwasaki A. Mucosal dendritic cells. Annu Rev Immunol. 2007;25:381–418.
33. Zheng J, Tao L. Multidimensional mechanisms and therapies underlying gastroesophageal reflux disease: focus on immunity, signaling pathways, and the microbiota-gut-brain axis. Front Immunol. 2025;16:1629944.

34. Strizova Z, Benesova I, Bartolini R, Novysedlak R, Cerny J, Hruskova Z, et al. M1/M2 macrophages and their overlaps - myth or reality? Clin Sci (Lond). 2023;137(15):1067–93.
35. Ding J, Garber JJ, Uchida A, Hara T, Nakagawa H, Oka M, et al. An esophagus cell atlas reveals dynamic rewiring during active eosinophilic esophagitis and remission. Nat Commun. 2024;15(1):3344.
36. Miyashita T, Tajima H, Shah FA, Okamoto K, Sakuma K, Kitagawa H, et al. Impact of inflammation-metaplasia-adenocarcinoma sequence and inflammatory microenvironment in esophageal carcinogenesis using surgical rat models. Ann Surg Oncol. 2014;21(6):2012–9.
37. Li X, Wu X, Wu Y, Liu C, Wang J, Zhang H, et al. Influence of autologous dendritic cells on the in-vitro expansion and functions of peripheral blood NK cells. Immunopharmacol Immunotoxicol. 2015;37(1):72–80.
38. Fonderflick L, Baudu T, Adotévi O, Roussey-Bihouée T, Borg C, Ferrand C, et al. The ATG8 family proteins GABARAP and GABARAPL1 target antigen to dendritic cells to prime CD4(+) and CD8(+) T cells. Cells. 2022;11(18):2782.
39. Strainic MG, Liu J, An F, Harnack U, Wang Q, Heeger PS, et al. CD55 is essential for CD103(+) dendritic cell tolerogenic responses that protect against autoimmunity. Am J Pathol. 2019;189(7):1386–401.
40. Somja J, Demoulin S, Roncarati P, Herfs M, Hubert P, Delvenne P, et al. Dendritic cells in Barrett's esophagus carcinogenesis: an inadequate microenvironment for antitumor immunity? Am J Pathol. 2013;182(6):2168–79.
41. Goda K, Abe K, Kanamori A, Koyama Y, Yamada K, Matsuura N, et al. Advanced endoscopy for benign esophageal disease: a review focused on non-erosive reflux disease and eosinophilic esophagitis. Healthc (Basel). 2022;10(11):2183.
42. Farooqi MS, Podury S, Crowley G, Kwon S, Lee S, Kim Y, et al. Noninvasive, multiomic, and multicompartmental biomarkers of reflux disease: a systematic review. Gastro Hep Adv. 2023;2(4):608–20.
43. Gyawali CP, Carlson DA, Chen JW, Patel A, Wong RJ, Yadlapati RH, et al. ACG clinical guidelines: clinical use of esophageal physiologic testing. Am J Gastroenterol. 2020;115(9):1412–28.
44. Ding W, Ding Z, Wang Y, Li X, Liu J, Zhang H, et al. Evodiamine attenuates experimental colitis injury via activating autophagy and inhibiting NLRP3 inflammasome assembly. Front Pharmacol. 2020;11:573870.
45. Cui G, Wang M, Li X, Zhang Y, Liu S, Chen H, et al. Berberine in combination with evodiamine ameliorates gastroesophageal reflux disease through TAS2R38/TRPV1-mediated regulation of MAPK/NF-κB signaling pathways and macrophage polarization. Phytomedicine. 2024;135:156251.

Figures

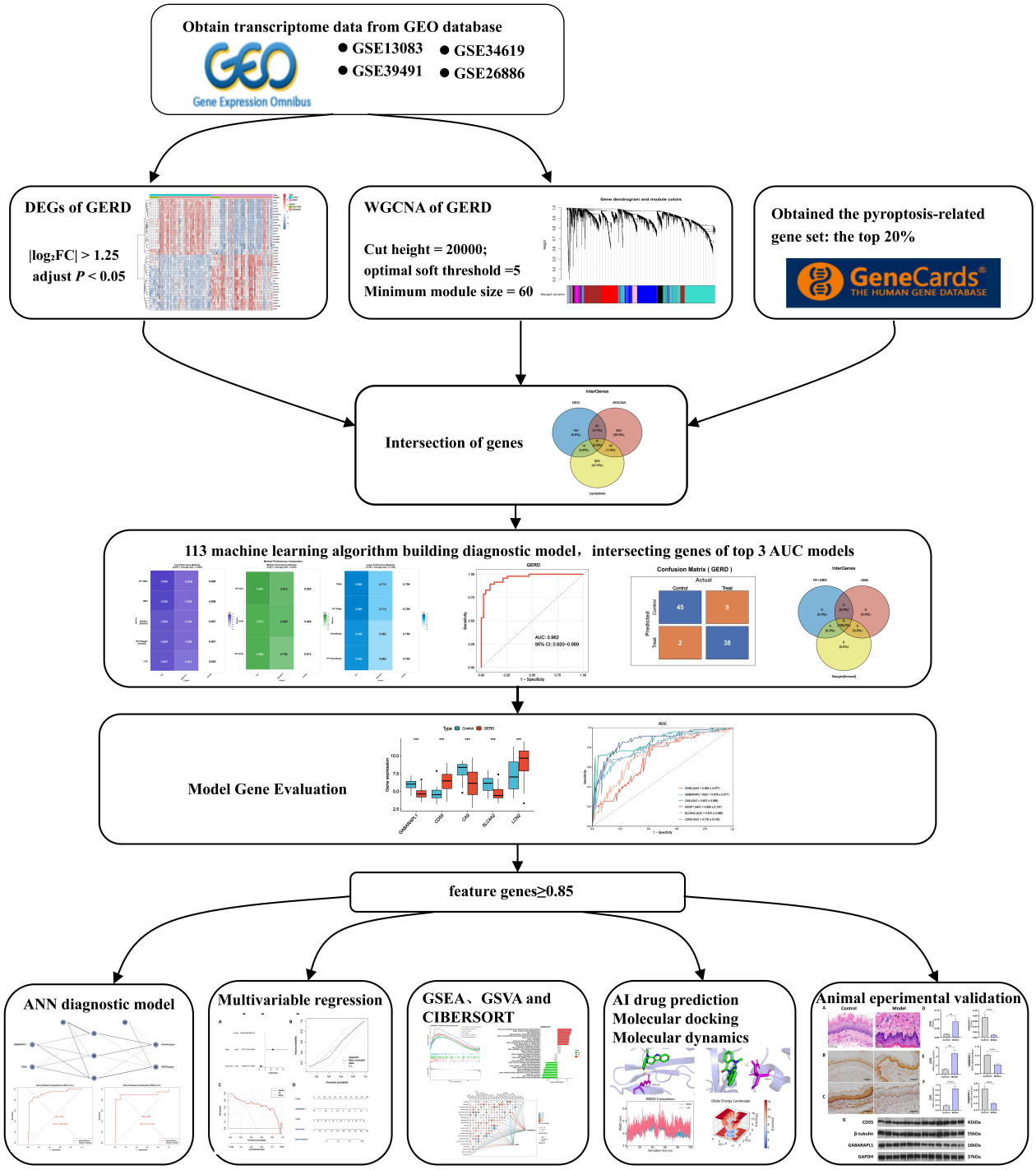


Figure 1

Flow chart of investigation.

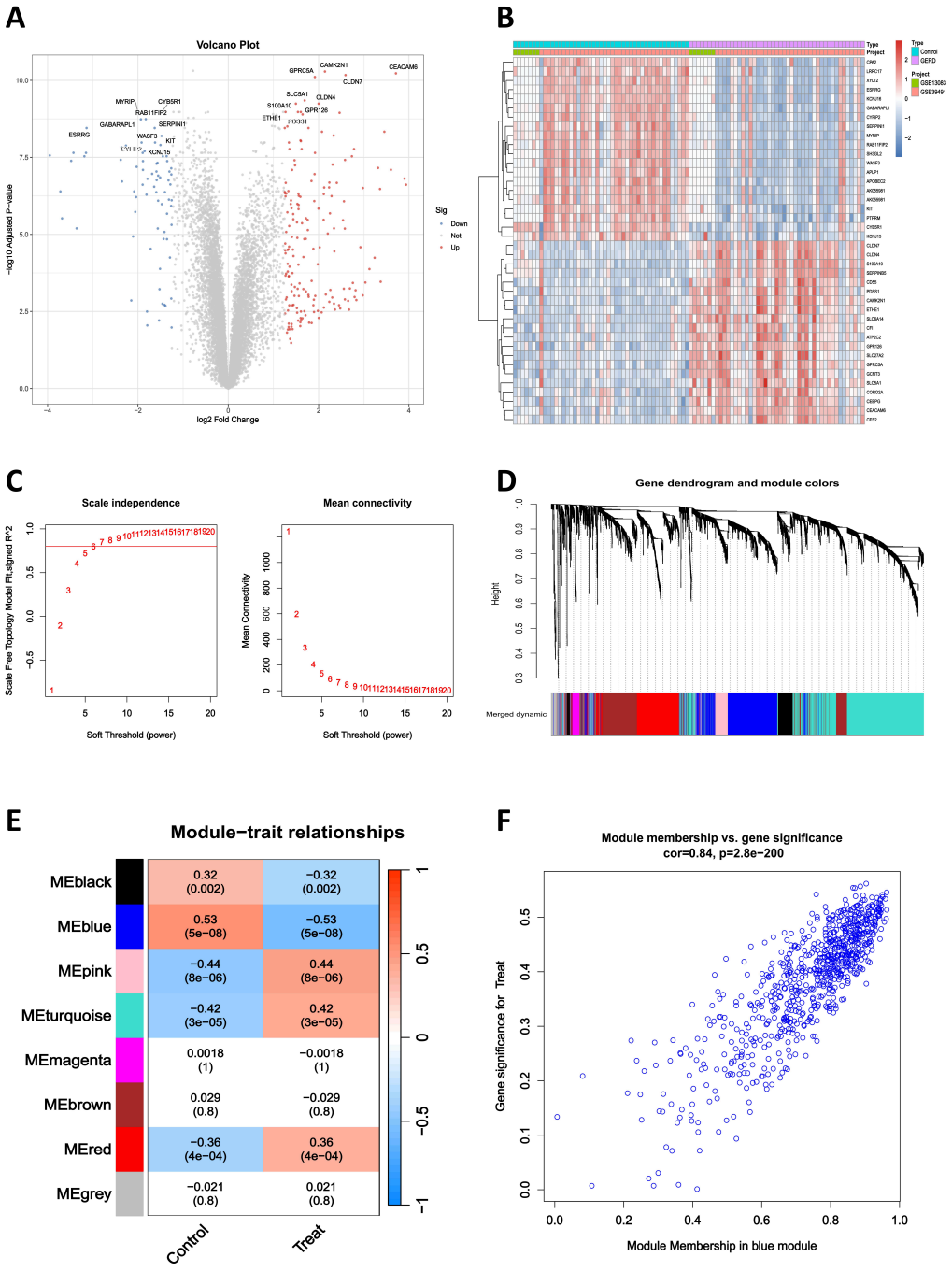


Figure 2

Differentially expressed genes and WGCNA results.
(A) Volcano plot of DEGs in GERD esophageal samples. Red dots indicate upregulated genes, blue dots indicate downregulated genes, and gray dots indicate non-significant genes.

(B) Heatmap of the top 40 DEGs between GERD and control groups; red indicates high expression and blue indicates low expression.

(C) Scale independence (left) and mean connectivity (right) across soft-thresholding powers.

(D) Gene clustering dendrogram with co-expression modules indicated by colors. (E) Module–trait correlation heatmap showing the relationship between each module and the GERD phenotype.

(F) Scatter plot of module membership (MM) versus gene significance (GS) in the blue module, with fitted line ($r = 0.84$, $P = 2.8\times10^{-200}$).

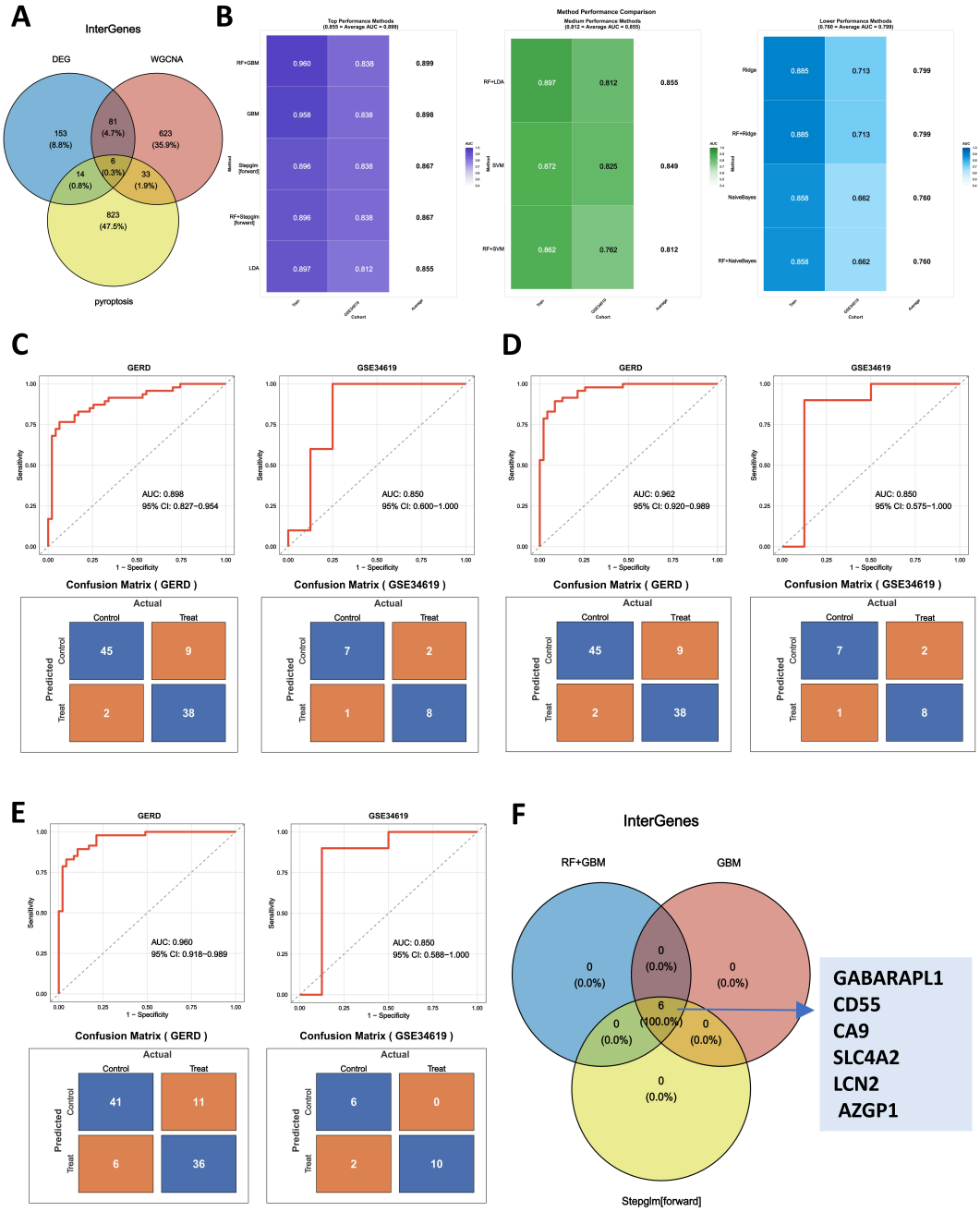


Figure 3

Machine learning-based construction of the gene diagnostic model.
(A) Venn diagram showing the overlap among GERD DEGs, WGCNA key module genes, and pyroptosis-related genes.
(B) Ranking of 113 combinatorial machine learning algorithms according to their diagnostic performance in GERD.

(C–E) ROC curves and confusion matrices of the top three models: RF+GBM, GBM, and Stepglm (forward).
(F) Venn diagram of the genes commonly selected by the top three models.

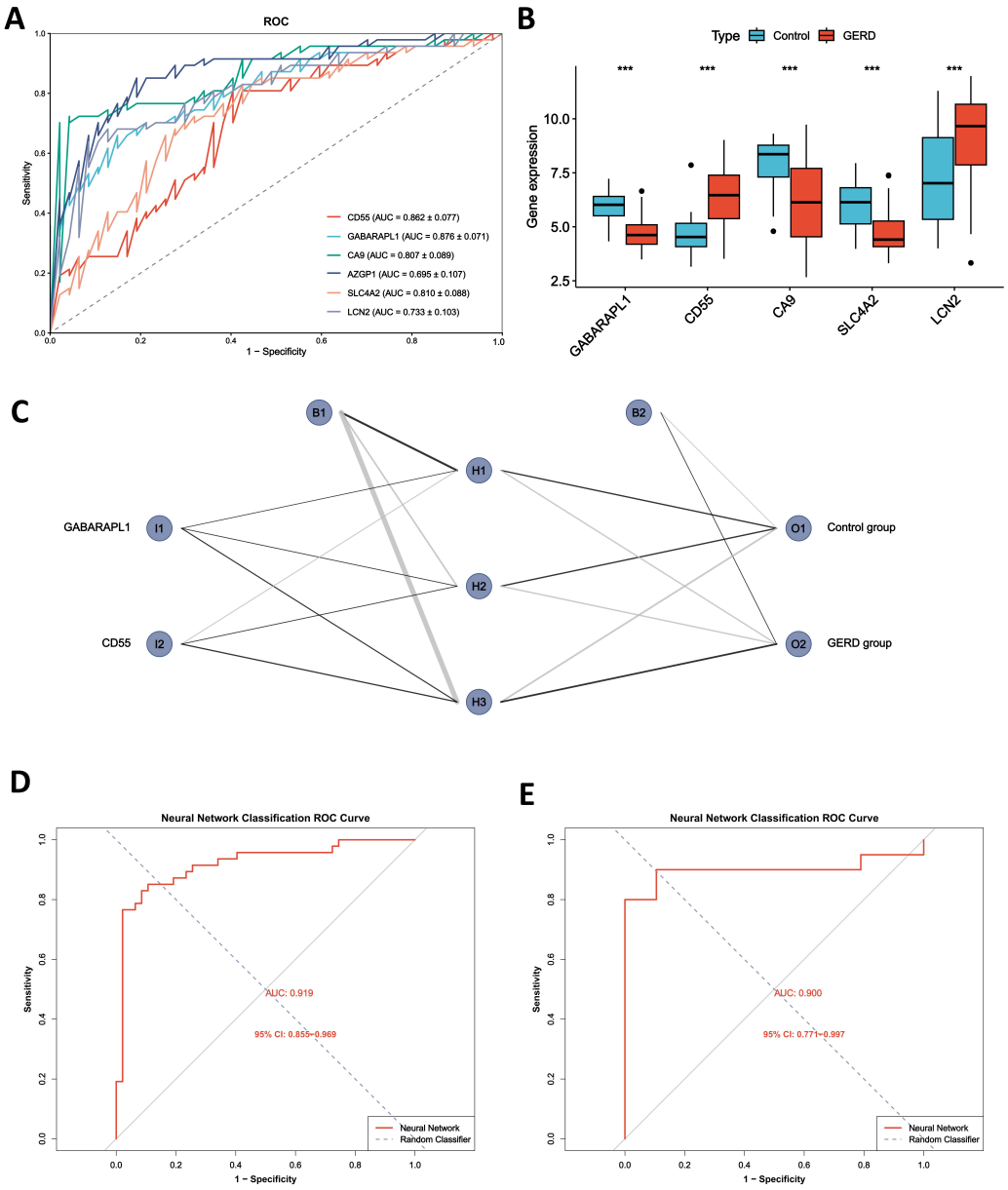


Figure 4

Evaluation of model genes and validation of the artificial neural network.
(A) ROC curves for the discriminatory ability of the six diagnostic model genes.
(B) Expression patterns of the model genes in GERD and control samples.
(C) Schematic of the ANN topology with three hidden neurons.
(D) ROC curve of the ANN model on the training set (AUC = 0.919, 95% CI: 0.855–0.969).
(E) ROC curve of the ANN model on the independent test set GSE26886 (AUC = 0.900, 95% CI: 0.771–0.997).

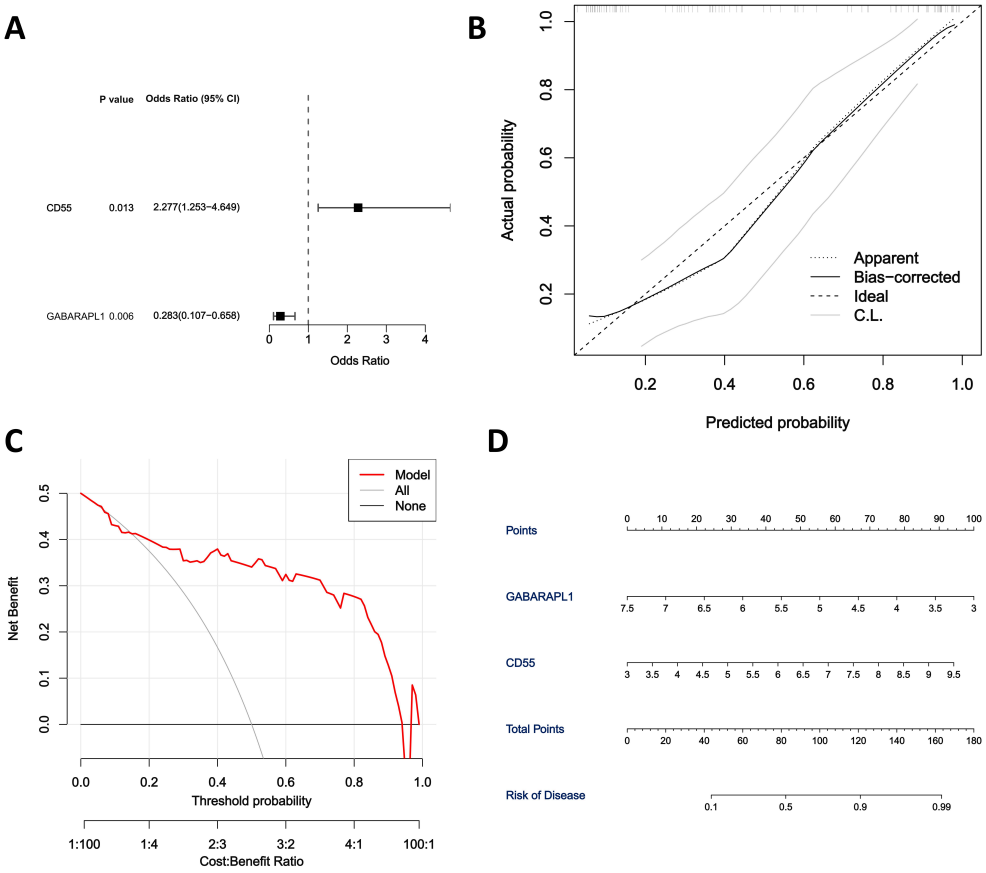


Figure 5

GERD risk prediction nomogram based on CD55 and GABARAPL1.
(A) Forest plot from the multivariable logistic regression, identifying CD55 as an independent risk factor and GABARAPL1 as an independent protective factor.
(B) Calibration curve comparing predicted and observed probabilities.
(C) Decision curve analysis (DCA) of net clinical benefit across threshold probabilities.
(D) Nomogram integrating CD55 and GABARAPL1 for individualized GERD risk prediction.

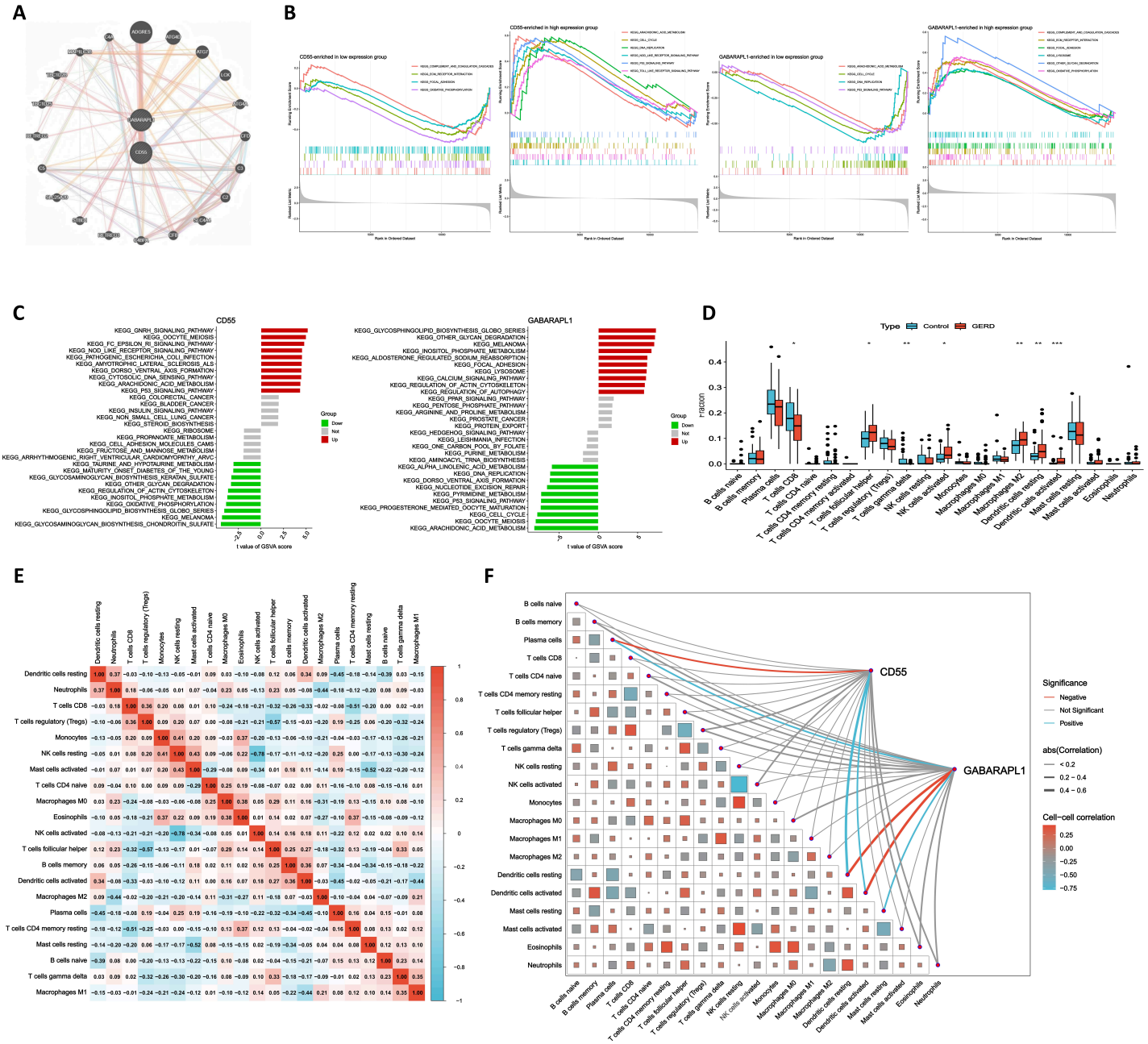


Figure 6

Biological functions of the signature genes and their relationship with immune cells.

(A) Protein–protein interaction network of CD55 and GABARAPL1 generated by GeneMANIA.

(B) GSEA enrichment results.

(C) GSVA enrichment results.

(D) Box plots of immune cell fractions estimated by CIBERSORT, showing cell types with significant

differences between groups.

(E) Correlation heatmap among immune cell subsets.

(F) Correlation heatmap between CD55/GABARAPL1 expression and immune cell fractions.

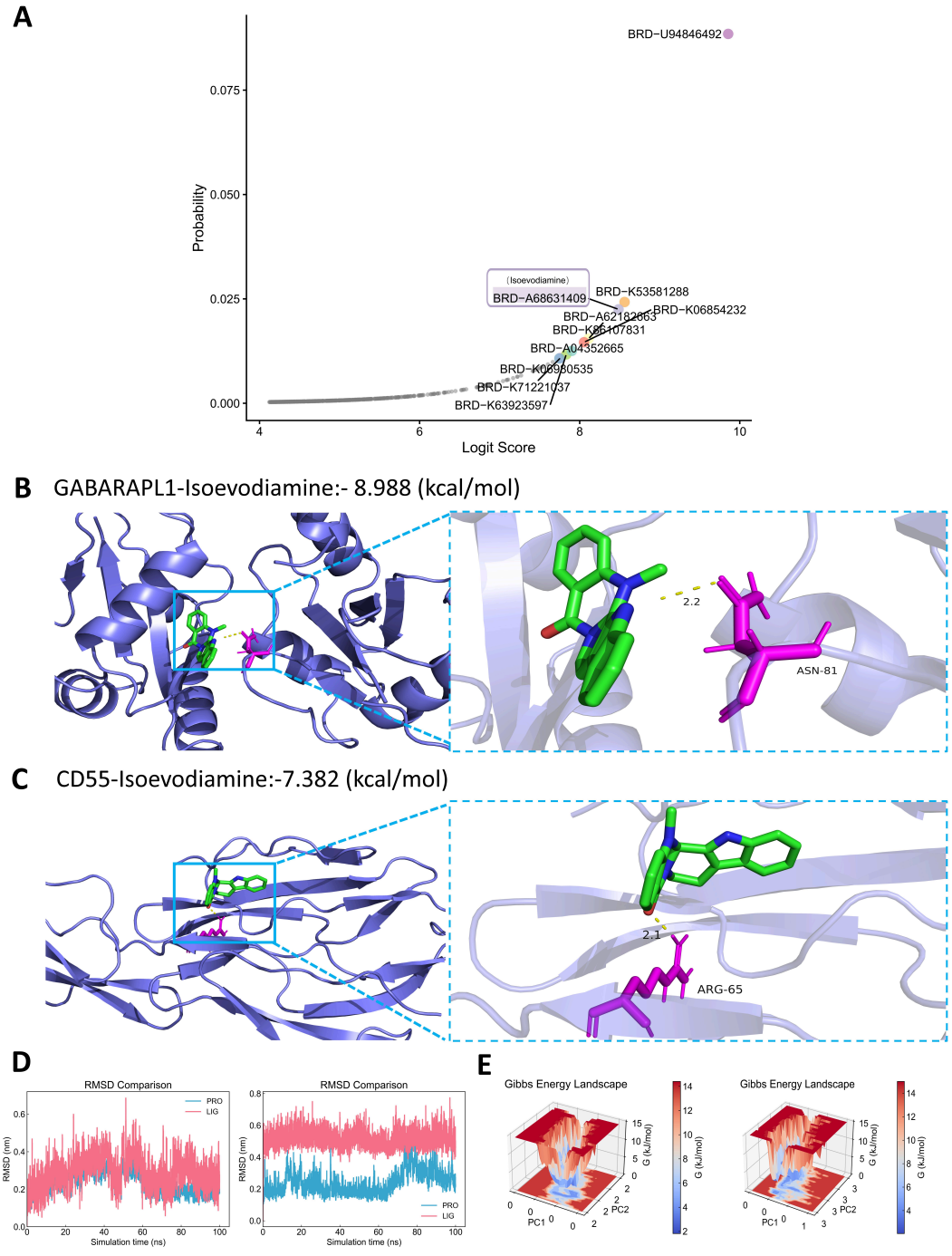


Figure 7

Molecular docking and molecular dynamics simulation of isoevodiamine with the signature proteins.
(A) Scatter plot of the top 10 candidate compounds identified by AI-based virtual screening.
(B) Predicted binding mode of isoevodiamine to GABARAPL1; the 2.2 Å hydrogen bond with Asn81 is indicated.
(C) Predicted binding mode of isoevodiamine to CD55; the 2.1 Å hydrogen bond with Arg65 is indicated.
(D) RMSD profiles of the isoevodiamine–protein complexes during the 100 ns MD simulation.
(E) Gibbs free energy landscapes of the complexes, demonstrating conformational convergence and thermodynamic stability.

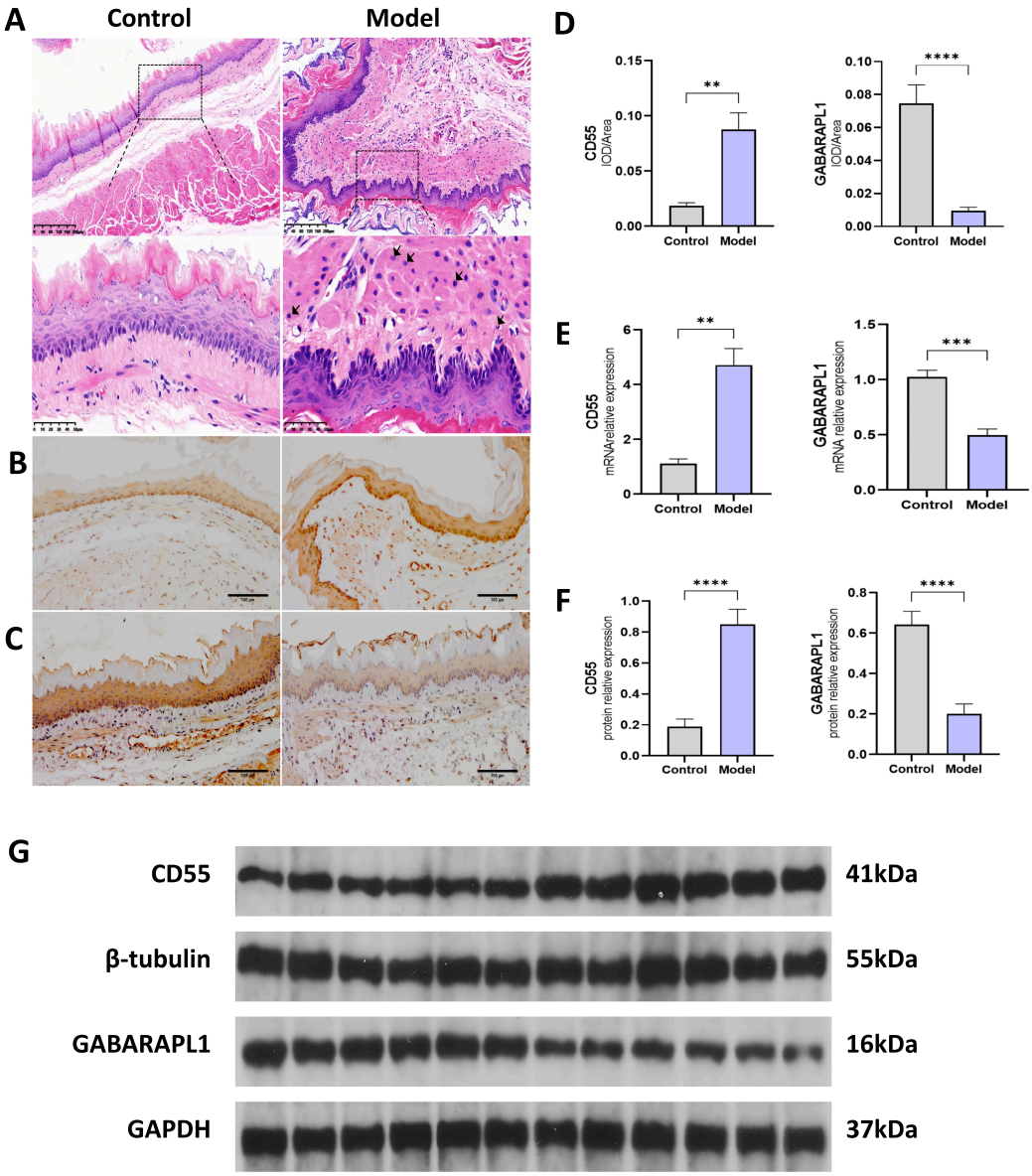


Figure 8

Validation of key findings in a rat GERD model.

(A) Representative HE staining of esophageal mucosa from control and model rats. Scale bar = 50 μm.

(B) Immunohistochemical staining for CD55.

(C) Immunohistochemical staining for GABARAPL1.

(D) Quantification of immunohistochemical staining by integrated optical density (IOD).

(E) qPCR analysis of CD55 and GABARAPL1 mRNA levels.
(F) Western blot quantification of CD55 and GABARAPL1 protein levels.
(G) Representative Western blot bands. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 vs. control group.

Supplementary Files

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

- SupplementaryFiguresS1.jpg
- SupplementaryFiguresS2.jpg
- SupplementaryFiguresS3.jpg
- supplementarytables.xlsx
- SupplementaryMethodsAnimalExperimentalValidation.docx
- Westernblotwithoutclipping.pdf
- Additionalfile.docx