



OPEN ACCESS

EDITED BY

Tsvetelina Velikova,
Sofia University, Bulgaria

REVIEWED BY

Joaquim Carreras,
Tokai University, Japan
Mudasir Rashid,
Howard University Hospital, United States
Hua Huang,
Tianjin Medical University General Hospital,
China

*CORRESPONDENCE

Yanbing Liu

✉ violetee@163.com

Xin Zeng

✉ zengxinmd1978@163.com

[†]These authors have contributed equally to this work

RECEIVED 19 September 2024

ACCEPTED 21 February 2025

PUBLISHED 06 March 2025

CITATION

Liu P, Sun C, Wang X, Han B, Sun Y, Liu Y and Zeng X (2025) Comprehensive analysis of anoikis-related gene signature in ulcerative colitis using machine learning algorithms. *Front. Med.* 12:1498864. doi: 10.3389/fmed.2025.1498864

COPYRIGHT

© 2025 Liu, Sun, Wang, Han, Sun, Liu and Zeng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comprehensive analysis of anoikis-related gene signature in ulcerative colitis using machine learning algorithms

Peng Liu^{1†}, Chunyan Sun^{1†}, Xiaojuan Wang^{2†}, Bing Han², Yuhao Sun¹, Yanbing Liu^{1*} and Xin Zeng^{1*}

¹Department of Gastroenterology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, ²Department of Pharmacy, Minhang Hospital, Fudan University, Shanghai, China

Ulcerative colitis (UC) is a chronic inflammatory bowel disease with an idiopathic origin, characterized by persistent mucosal inflammation. Anoikis is a programmed cell death mechanism activated during carcinogenesis to eliminate undetected isolated cells from the extracellular matrix. Although existing evidence indicates that anoikis contributes to the modulation of immune response, the involvement of anoikis-related genes (ARGs) in UC pathogenesis and their interaction with infiltrating immune cells has not been thoroughly explored. The GSE75214, GSE92415, and GSE16879 datasets were acquired and integrated from the GEO database. Additionally, 58 ARGs were identified through the GSEA database. Key anoikis-DEGs in UC were identified using three machine learning algorithms, including least absolute shrinkage and selection operator (LASSO) Cox regression, random forest (RF), and support vector machine (SVM). Receiver operating characteristic (ROC) analysis was utilized to evaluate the diagnostic accuracy of each gene. Subsequently, Single sample GSEA (ssGSEA) was executed to explore the relationships within immune cell infiltration, UC subtypes, and key anoikis-DEGs. Besides, unsupervised cluster analysis was conducted to categorize the UC samples into distinct subgroups, followed by comparing subtype differences. Finally, the upstream regulatory network was constructed and visualized. A comprehensive analysis of the involvement of ARGs in UC was performed, revealing their expression profile, correlation with infiltrating immune cells, and enrichment analyses. We identified five key anoikis-DEGs (*PDK4*, *CEACAM6*, *CFB*, *CX3CL1*, and *HLA-DMA*) and demonstrated their high diagnostic accuracy for UC. Moreover, *CEACAM6*, *CFB*, *CX3CL1*, and *HLA-DMA* exhibited positive associations with infiltrating immune cells in UC, whereas *PDK4* displayed a negative correlation with all immune cells. Unsupervised cluster analysis enabled the classification of UC patients into two clusters, both of which exhibited distinct gene expression profiles and immune signaling pathways. Further, based upon the upstream regulatory network, TP53, RARB, RXRB, and CTCF potentially exerted regulatory functions. Our analysis identified five key anoikis-DEGs as characteristic biomarkers of UC. These genes were strongly associated with the infiltration of both innate and adaptive immune cells, as well as immune pathways. This study highlights the role of anoikis genes in UC pathophysiology and offers valuable insights for further elucidating UC pathogenesis and individualized therapy.

KEYWORDS

ulcerative colitis, anoikis, diagnostic marker, machine learning algorithm, immune cells

1 Introduction

Ulcerative colitis (UC), a primary subtype of inflammatory bowel disease (IBD), exhibits continuous colonic mucosal inflammation that extends proximally from the rectum (1, 2). This disease affects nearly 5 million individuals worldwide in 2023, with its global disease burden on the rise (3). In China, the reported incidence of UC is 3.35 per 100,000 males and 2.65 per 100,000 females (4). Early diagnosis is critical for effective UC treatment, yet specific diagnostic signatures are lacking. At present, the biomarkers for UC include serum anti- $\alpha\text{v}\beta 6$ antibodies and serum oncostatin M in diagnosing, fecal calprotectin and serum trefoil for disease activity assessment, as well as whole blood transcriptomic panels and CLEC5A/CDH2 ratio for determining the need for escalated treatment (5). In addition, despite recent advancements in diagnostic and therapeutic methods, including surgical and immunotherapeutic strategies, the prognosis for UC remains unsatisfactory. A growing body of evidence has identified that UC is frequently driven by abnormalities in a wide range of immune cell types, encompassing leukocytes, macrophages, B cells, dendritic cells, and both regulatory and effector T cells (6, 7). However, the multifaceted pathophysiology and immune mechanisms underlying UC are not completely understood. Consequently, the identification of novel characteristic genes has the potential to accelerate the unveiling of prospective therapeutic targets, innovative intervention approaches, and a deeper understanding of the pathogenic mechanisms and etiology of UC. This, in turn, can enable the implementation of more successful strategies for diagnosis and therapy.

Anoikis, a form of programmed cell death, occurs upon the detachment of cells from their designated extracellular matrix (ECM). Its primary role lies in preventing dysplastic cell growth or improper matrix attachment (8, 9). Anoikis regulates anchorage-dependent growth and epithelial-mesenchymal transition, thus being indispensable for preserving tissue homeostasis and proper development. Additionally, it serves as a significant regulatory factor in the context of metastatic cancers, cardiovascular diseases, and diabetes. Molecular pathways and mechanisms that affect anoikis resistance have been revealed in recent years (8, 10). These pathways include a range of factors, such as growth factors, cell adhesion molecules (CAMs), and pathways with the capacity to induce epithelial-mesenchymal transitions. The complex network of downstream molecules includes focal adhesion kinase (11), Src kinase (12), mitogen-activated protein kinase (MAPK) (13), ERK1/2 (14), Bcl-2 family (15), PI3K/Akt (16, 17), and insulin-like growth factor receptors (18). Each of these pathways is crucial for preventing cell death and promoting cell survival. In recent years, studies have shown various factors influencing anoikis, comprising integrins, IGFR, EGFR, TGF- β , Trk, NF- κ B, E-calmodulin, the Hippo pathway, eEF-2 kinase, ROS, acidosis, hypoxia, and protective autophagy (8, 10). Cell death has the potential to directly or indirectly compromise barrier function and impede epithelial restitution in IBD. The primary pathological damage of UC results from dysregulated immune responses induced by commensal microflora. This dysregulation leads to the production of inflammatory cytokines, the infiltration of lymphocytes from the bloodstream into the inflamed bowel, and innate immune cell signaling (19).

Currently, therapeutic strategies for UC include anti-adhesion therapy (20), anti-TNF monoclonal antibodies (21, 22), anti-IL-12/

IL-23p40 antibodies (23), and JAK inhibitors (24). Therefore, anoikis may play a crucial role in the onset and progression of UC. Multiple scientific studies have robustly confirmed the integral role of anoikis in the pathogenesis of diverse diseases, with a particular emphasis on its significance in tumor immunity. These studies have explored a wide range of conditions, including glioblastoma (25), head and neck squamous cell carcinoma (26), and lung adenocarcinoma (27). A previous study demonstrated that nuclear MYH9 bound to the CTNNB1 promoter and promote CTNNB1 transcription, thereby conferring resistance to anoikis in gastric cancer (28). Moreover, the formation of a protein complex comprising Bim-EL, LC8, and Beclin-1 might contribute to the evasion of anoikis in inflammatory breast cancer (29). In another study by Jin et al., it has been verified that GDH1-mediated metabolic reprogramming of glutaminolysis promotes resistance to anoikis and tumor metastasis in LKB1-deficient lung cancer (30). In addition, a study revealed that *Aloe vera* gel polysaccharide (AGP) induced Nrf2 activation, reduced ROS levels, ameliorated mitochondrial dysfunction, and alleviated anoikis caused by impaired-mitochondrial function of colonic epithelial cells to maintain intestinal barrier integrity in DSS-induced colitis mice (31). Despite increasing evidence indicating that anoikis contributed to UC progression, its predictive value and association with immune response regulation in UC remains unexplored. Similarly, the underlying mechanisms associated with anoikis in the manifestation and progression of UC remain unknown. Thus, a comprehensive study is warranted to investigate differences in immune characteristics between normal tissues, UC specimens, and different UC subtypes. This research is vital for elucidating the cellular features and molecular mechanisms underlying anoikis and its associated genes. Additionally, establishing characteristics related to anoikis can offer valuable insights for the individualized treatment of UC patients.

A thorough and multiscale bioinformatics analysis was executed to determine the differentially expressed genes (DEGs) in samples from UC patients and normal individuals, focusing on anoikis-related genes (ARGs) and immune infiltration profiles. Moreover, the enrichment analyses were performed to investigate the differences between UC and normal samples. Subsequently, anoikis-DEGs were identified to explore the association between anoikis and UC. Using three machine learning algorithms, we finally obtained five risk signature genes for their potential to predict disease onset. Immune infiltration and functional enrichment analyses were also performed using these five key anoikis-DEGs. Simultaneously, an unsupervised clustering analysis and functional enrichment analysis was executed to distinguish differences between various clusters. Further, gene-miRNA and gene-transcription factor regulatory network associated with the five risk genes were assessed. In summary, our findings offer a comprehensive insight into the connection between anoikis and UC, establishing a foundation for individualized diagnosis and management of UC.

2 Materials and methods

2.1 Data sources and processing

The gene expression profiling data for UC, encompassing GSE75214, GSE92415, and GSE16879 datasets, were retrieved from

the gene expression omnibus (GEO) database.¹ The GSE75214 dataset included gene expression profiles from 97 individuals with UC and 11 controls, while the GSE92415 dataset comprised gene expression profiles from 87 pre-treatment UC patients and 21 healthy individuals. The GSE16879 dataset featured gene expression data from 24 pre-treatment UC patients and six control healthy volunteers.

Samples unrelated to this study were excluded, resulting in 246 samples included: 208 UC mucosal biopsy samples and 38 normal samples. The merged dataset was subjected to batch correction utilizing the “limma” and “sva” packages (32). Principal component analysis (PCA) was employed to evaluate the effectiveness of batch effect removal and to visualize the distribution of samples from UC and healthy individuals. Furthermore, the merged data underwent normalization utilizing the R package “preprocessCore.” Subsequently, 58 ARGs were extracted from the molecular signatures database,² as illustrated in [Supplementary Table 1](#).

2.2 Identification of DEGs and functional enrichment analysis

Identification of DEGs was carried out by the “limma” R package, ensuring the criteria of $|\log_2 \text{fold change (FC)}| > 0.5$ and adjusted p -value < 0.05 (33). Volcano plots were created using the R packages “limma” and “ggplot2” to show the distribution of DEGs. The “ComplexHeatmap” R package was employed to visualize the top 20 upregulated and downregulated genes. The Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were determined on the merged DEGs utilizing the R package “clusterProfiler” to explore differential biological pathways related to the signature genes. Statistically significance was defined as $p < 0.05$.

2.3 Identification of anoikis-DEGs in UC and functional enrichment analysis

The overlapping genes between the upregulated and downregulated DEGs of UC and ARGs were assessed *via* the “limma” R package. Additionally, the GO and KEGG pathway enrichment analysis was conducted to further analyze the signaling pathways of anoikis-DEGs. Statistically significance was defined as $p < 0.05$.

2.4 Differential gene expression analysis of ARGs between UC and healthy individuals

To present a thorough overview of the differential expression of ARGs in UC and healthy individuals, the differential expression levels of ARGs were visualized in volcano plots and heatmap using the R packages “limma,” “ggplot2,” and “ComplexHeatmap.” Subsequently, we explored the expression of 21 anoikis-DEGs in UC and normal samples. Simultaneously, a protein–protein interaction (PPI) network analysis was conducted on 21 anoikis-DEGs utilizing the STRING database (<https://string-db.org/>) and displayed using the Cytoscape 3.9.0 software.

2.5 Identification of key Anoikis-DEGs in UC using machine learning

Three machine learning algorithms were utilized in this study to identify the key anoikis-DEGs as characteristic biomarkers in UC. The first algorithm employed was the least absolute shrinkage and selection operator (LASSO) logistic regression by using the R package “glmnet” (34). LASSO was selected for its ability to perform both variable selection and regularization, which helps identify important genes while reducing overfitting, especially in high-dimensional datasets. To ensure robustness and avoid overfitting of the model, we employed 10-fold cross-validations to determine the parameter λ . According to the optimal λ value and the corresponding coefficients, nine genes were obtained. The second algorithm was the random forest (RF) algorithm, assessed utilizing the R package “randomForest” to grow a forest consisting of 500 trees with default settings (35). RF algorithm, an ensemble learning method, was chosen for its robustness and ability to construct multiple decision trees and aggregate their predictions to improve accuracy and prevent overfitting. Using the RF algorithm, the top 10 genes with the highest importance were selected for downstream analysis. Finally, the support vector machine (SVM) algorithm was utilized, with the analysis conducted using the R package “kernlab” (36). The SVM algorithm can rank features based on recursion to avoid overfitting, and 13 genes were obtained through this approach. The five characteristic biomarkers for diagnosis were finally identified as the common core genes among the three algorithms. Furthermore, the five key anoikis-DEGs’ interactions were illustrated using the “chordDiagram” R package. Subsequently, the ROC curves and the area under the curve (AUC) were obtained utilizing the R package “pROC.”

2.6 Validation of the key anoikis-DEGs in colitis animal model

Animal experiments were conducted in accordance with the ethical requirements and approved by the Tongji University Animal Ethics Committee. The C57BL/6 male mice were randomly allocated into normal control and colitis groups. The Control healthy mice were provided normal water. The colitis model was induced by administering 3.5% dextran sulfate sodium (DSS, Yeasen Biotechnology, 60316ES76, China) for 7 days, followed by 3 days of regular drinking water. Then, the mice colon tissues were obtained for RNA extraction to quantify the gene expression of the key anoikis-DEGs using Quantitative Real-time PCR (qPCR). Tissue total RNA was extracted using Trizol reagent (Invitrogen, 15,596,018, United States) and reverse to cDNA using an Evo M-MLV kit (Accurate Biology, AG11705, China). The Primer sequences were obtained from PrimerBank³ and synthesized at Sangon Biotech (Shanghai, China). qPCR was performed using the SYBR Green Pro Taq HS premix qPCR (Accurate Biology, AG11701, China) on a SLAN-96S Real-Time PCR System. The results were replicated three times, and the levels of target genes were normalized to GAPDH and analyzed utilizing the $2^{-(\Delta\Delta Ct)}$ method. *CEACAM6* is not present in rodents, and *H2-DMA* is orthologous to human *HLA-DMA*. Thus,

1 <https://www.ncbi.nlm.nih.gov/geo/>

2 <https://www.gsea-msigdb.org/gsea>

3 <https://pga.mgh.harvard.edu/primerbank/>

we analyzed the expression of four key anoikis genes: *PDK4*, *H2-DMa*, *CX3CL1*, and *CFB*. The following is a list of the primer sequences (5'-3'):

PDK4-F AGGGAGGTCGAGCTGTTCTC.
 PDK4-R GGAGTGTTCACCTAAGCGGTCA.
 H2-DMa-F CTCGAAGCATCTACACCAGTG.
 H2-DMa-R TCCGAGAGCCCTATGTTGGG.
 CX3CL1-F ACGAAATGCGAAATCATGTGC.
 CX3CL1-R CTGTGTCGTCTCCAGGACAA.
 CFB-F GAGCGCAACTCCAGTGCTT.
 CFB-R GAGGGACATAGGTACTCCAGG.
 GAPDH-F TGGCCTTCCGTGTTCTTAC.
 GAPDH-R GAGTTGCTGTTGAAGTCGCA.

2.7 The immune infiltration characteristics in UC samples and the association between key anoikis-DEGs and infiltrating immune cells

Single sample GSEA (ssGSEA) is an extension of the GSEA method that enables the assessment of infiltrated immune cells and the activity of specific immune factors (37). ssGSEA was performed to measure the correlation coefficients between immune infiltration cells using the “GSVA” R package. Pearson’s correlation coefficient analysis was utilized to investigate the association between immune infiltrating cells. Heatmaps were generated using the R package “ggplot2.” Simultaneously, the “ggplot2” package was utilized to display immune cell infiltration as box plots in UC and healthy volunteers, with Pearson’s correlation coefficient evaluating relationships between different infiltrating immune cells. The association between the key anoikis-DEGs and immune infiltrating cells was assessed utilizing Pearson’s analysis and visualized in Lollipop plots using the R package “ggplot2.”

2.8 Correlation of key anoikis-DEGs with all UC genes and GSEA analysis

Next, the correlation of the key anoikis-DEGs with all UC genes was assessed, and a heatmap was utilized to display the 50 leading positively correlated genes. Additionally, Reactome pathway analysis was performed using GSEA for each diagnostic marker utilizing the R package “clusterProfiler” to explore the top 20 associated biological pathways.

2.9 Identification of anoikis-related subtypes in UC and functional enrichment analysis

To explore the functional role of ARGs in UC, unsupervised cluster analysis was performed utilizing the R package “ConsensusClusterPlus” as per the key anoikis-DEGs (38). Moreover, the differential expression of key anoikis-DEGs among different subtypes were explored. Moreover, a heatmap was generated utilizing the R package “pheatmap” to demonstrate the association between clinical features, gene expression, and subtypes.

Three gene sets (“h.all.v7.5.1.symbols.gmt,” “c2.cp.kegg.v7.5.1.symbols,” and “c2.cp.reactome.v7.5.1.symbols”) were acquired from the MSigDB database⁴ as input files for gene set variation analysis (GSVA) (39). Further, pathway enrichment analysis was performed to identify distinct pathways associated with different subtypes.

2.10 Differential gene expression analysis and functional annotation of UC subtypes

To assess the reproducibility of the data, PCA was utilized to examine the overall distribution and determine the reproducibility of the data across subtypes. Additionally, a volcano plot was generated for visualizing the DEGs in distant subtypes. Enriched GO terms and KEGG pathways within the differential subtypes were analyzed. Furthermore, the correlation between the top five KEGG pathways and the DEGs in the distinct subtypes was established.

2.11 Regulatory network of TFs and miRNAs associated with the key anoikis-DEGs

Transcription factors (TFs) and microRNAs (miRNAs) are pivotal regulators of gene regulations, functioning at the transcriptional and posttranscriptional levels, respectively. To explore TF- and miRNA-mediated gene interactions, the RegNetwork database⁵ was utilized to predict potential upstream miRNAs and TFs (40). Regulatory network of miRNAs and TFs were constructed based on key Anoikis-DEGs using Cytoscape 3.9.0 software.

2.12 Statistical analysis

Data analyses were performed utilizing R 4.3.3 and related R packages. GraphPad Prism 9.3.0 (GraphPad Software, Inc., La Jolla, CA, United States) was used for statistical analysis. All data are shown as mean ± standard deviation (SD). Differences between two groups were assessed using the *t*-test. The correlation between the two variables was determined by using the Pearson product-moment correlation coefficient. *p* < 0.05 indicated statistical significance.

3 Results

3.1 Data processing and identification of DEGs in UC individuals

To examine the biological role of ARGs in UC progression, the expression data of GSE75214, GSE92415, and GSE16879 datasets was firstly integrated. Tissues from different platforms exhibited distinct clustering patterns prior to batch effect removal (Figure 1A). The batch effects were effectively removed from the

⁴ <https://www.gsea-msigdb.org/gsea/msigdb/>

⁵ <https://regnetworkweb.org/>

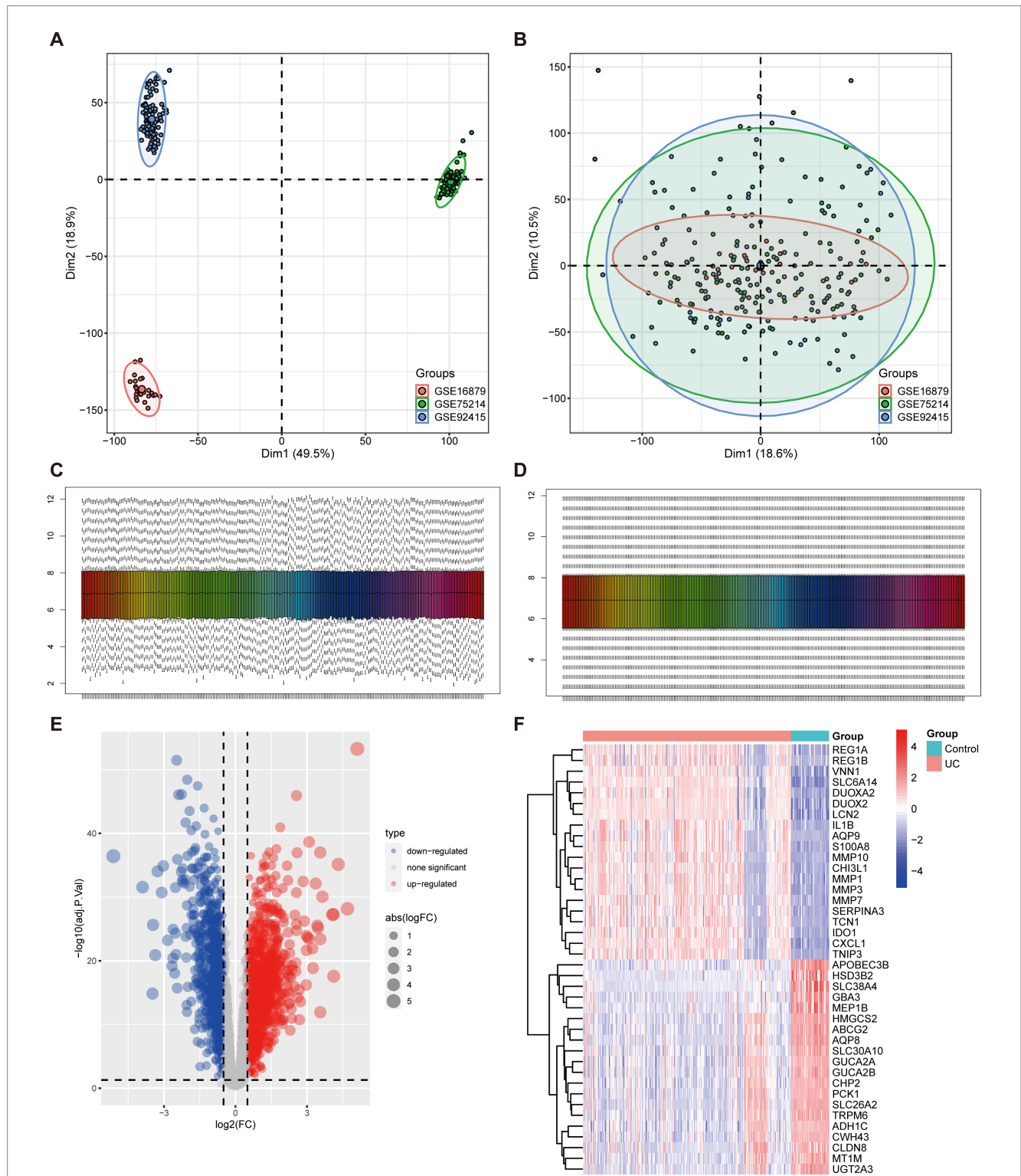


FIGURE 1 Identification of DEGs associated with UC. **(A,B)** Principal component analysis (PCA) of three datasets before **(A)** and after **(B)** batch effect removal. **(C,D)** Combined data normalized before **(C)** and after **(D)**. **(E)** The volcano plot illustrates the upregulated and downregulated DEGs associated with UC. Horizontal coordinates represent \log_2FC , and vertical coordinates represent $-\log_{10}(\text{adj.}P\text{.Val})$. Red nodes represent upregulated DEGs, blue nodes represent downregulated DEGs, and gray nodes denote genes that do not exhibit significant differential expression. **(F)** Heatmap displays the top 20 upregulated and downregulated DEGs in UC. Red signifies upregulated gene expression, and blue signifies downregulated gene expression.

GEO datasets, resulting in an integrated dataset comprising 208 UC samples and 38 control samples (Figure 1B). Next, the data underwent normalization using the R software package

“preprocessCore” (Figures 1C,D). We obtained 2,744 DEGs differentiating UC from normal tissues, comprising 1,577 upregulated and 1,167 downregulated genes, as illustrated by the

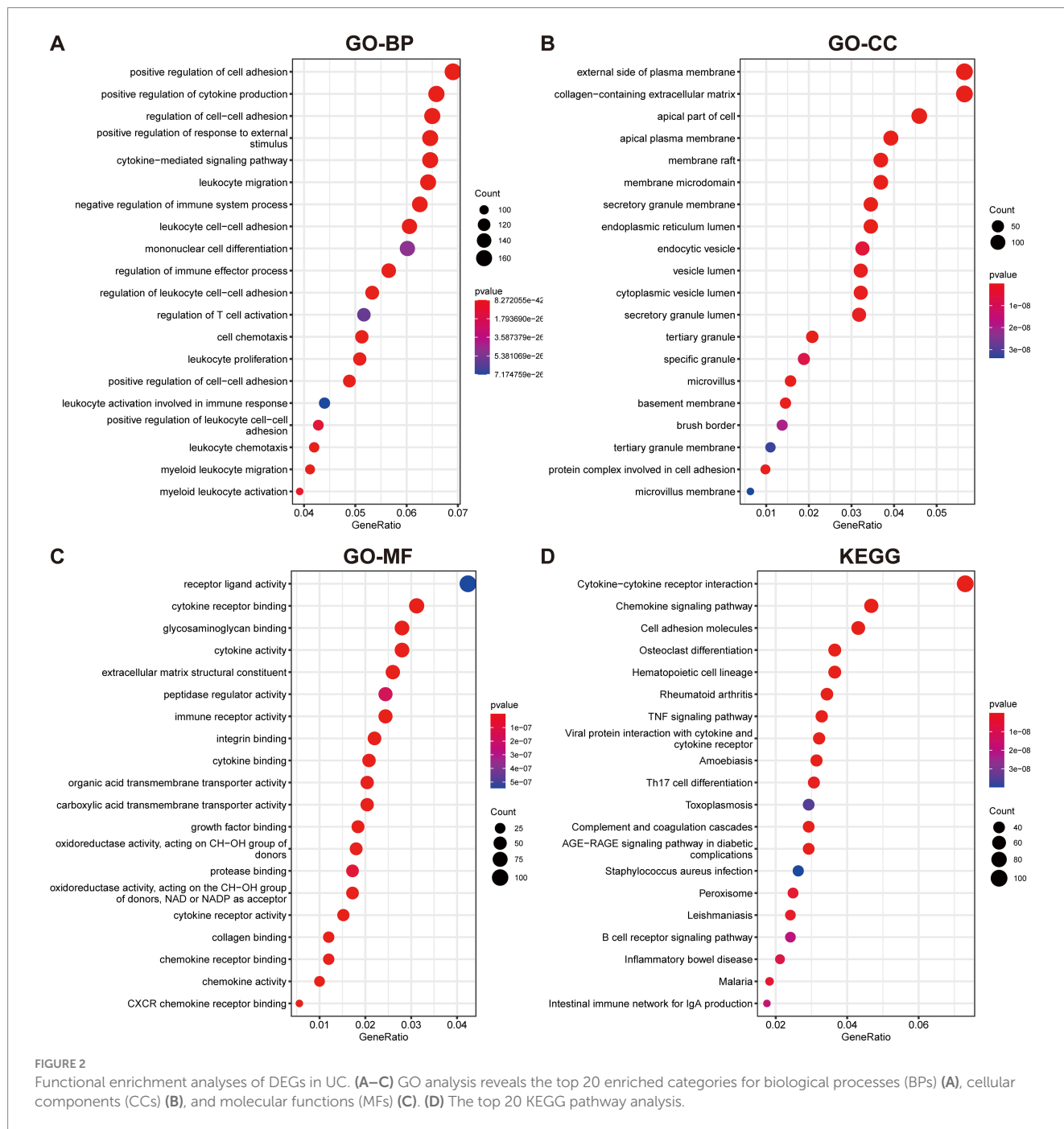


FIGURE 2 Functional enrichment analyses of DEGs in UC. (A–C) GO analysis reveals the top 20 enriched categories for biological processes (BPs) (A), cellular components (CCs) (B), and molecular functions (MFs) (C). (D) The top 20 KEGG pathway analysis.

volcano plot (Figure 1E) and the top 20 genes that have been upregulated and downregulated illustrated in a heatmap (Figure 1F).

3.2 Functional enrichment analysis of DEGs

To elucidate the underlying mechanisms and roles of these DEGs in UC, we carried out GO and KEGG pathway analyses. The GO enrichment analysis indicated these genes significantly involved in pathways such as cell adhesion, cytokine production, plasma membrane, and extracellular matrix (Figures 2A–C). Furthermore, the KEGG enrichment analysis demonstrated that the DEGs were

predominantly enriched in multiple inflammatory response-related pathways, including the cytokine-cytokine receptor interaction, TNF signaling pathway, Th17 cell differentiation, and B cell receptor signaling pathway (Figure 2D).

3.3 Identification of anoikis-DEGs in UC and functional enrichment analysis

To ascertain the role of ARGs in UC, the overlap between anoikis genes and DEGs in UC was initially identified. By intersecting the upregulated and downregulated DEGs with anoikis, 19 upregulated anoikis-DEGs (*HLA-DMA*, *CFB*, *CD44*, *ITGA5*, *VCAM1*, *STAT1*,

PIK3R3, *CCL20*, *NOS2*, *IRF1*, *CX3CL1*, *SIK1*, *CXCL10*, *CAV1*, *CEACAM6*, *MMP2*, *MX1*, *LTA*, and *SNAI2*) and two downregulated genes (*PDK4* and *DAPK2*) in UC were identified (Figures 3A,B). Subsequently, we conducted GO and KEGG enrichment analyses utilizing the 21 anoikis-DEGs. The GO term analysis highlighted that anoikis-DEGs exhibited enrichment in BPs, including cytokine-mediated signaling pathway, anoikis, regulation of cell–cell adhesion, and regulation of anoikis (Figure 3C). For CCs, enrichment was observed in the secretory granule membrane, focal adhesion, external side of the plasma membrane, and cell-substrate junction (Figure 3D). For MFs, the enriched functions included cytokine receptor binding, signaling receptor activator activity, receptor ligand activity, and G protein-coupled receptor binding (Figure 3E). KEGG analysis demonstrated anoikis-DEGs exhibited enrichment in the TNF signaling pathway, Chemokine signaling pathway, and Toll-like receptor signaling pathway (Figure 3F).

3.4 Differential gene expression analysis of ARGs between UC and healthy individuals

To evaluate the expression profiles of ARGs in UC and healthy individuals, we performed differential gene expression analysis. The volcano plot in Figure 4A illustrated the upregulation of 19 anoikis-DEGs and the downregulation of two genes in UC patients. Moreover, Figure 4B presented a heatmap displaying the differential expression of ARGs between UC groups and healthy control. Next, we compared the expression of these 21 anoikis-DEGs between UC patients and healthy individuals. Specifically, *PDK4* and *DAPK2* demonstrated relatively low expression in UC. In contrast, the other 19 genes (*CEACAM6*, *CFB*, *HLA-DMA*, *CX3CL1*, etc.) showed significant elevation in UC compared with normal samples (Figure 4C). Moreover, a PPI network was created to explore the relationship between anoikis-DEGs (Supplementary Figure 1). Among these, *CXCL10*, *VCAM1*, *CCL20*, *STAT1*, and *CX3CL1* highlighted their interaction with other anoikis-related proteins. The PPI network highlights critical immune and inflammatory pathways, particularly involving *VCAM1* and *CXCL10*, providing insights into the molecular mechanisms that may potentially contribute to UC pathogenesis.

3.5 Key anoikis-DEGs identification through machine-learning

Three machine-learning algorithms were employed to identify key anoikis-DEGs as characteristic biomarkers of UC from a set of 21 anoikis-DEGs. The LASSO Cox regression analysis demonstrated that nine genes were associated with UC, namely *HLA-DMA*, *CFB*, *CCL20*, *CX3CL1*, *SIK1*, *CAV1*, *CEACAM6*, *PDK4*, and *DAPK2* (Figure 5A). The RF algorithm ranked the importance scores of signature genes and identified 10 genes, including *CFB*, *HLA-DMA*, *CD44*, *CEACAM6*, *PDK4*, *VCAM1*, *CX3CL1*, *CAV1*, *STAT1*, and *IRF1* (Figure 5B). In the case of SVM outcomes, 13 signature genes were identified, namely *ITGA5*, *HLA-DMA*, *CX3CL1*, *CD44*, *PDK4*, *VCAM1*, *STAT1*, *CFB*, *SIK1*, *CEACAM6*, *DAPK2*, *PIK3R3*, and *CCL20* (Figure 5C). Subsequently, after intersecting of the three machine learning results, five key anoikis-DEGs, namely *PDK4*, *HLA-DMA*,

CEACAM6, *CX3CL1*, and *CFB* were identified (Figure 5D). Additionally, a correlation analysis was conducted on these five key anoikis-DEGs. Except for *PDK4*, which exhibited a negative correlation with the other four genes, all the other genes showed positive correlations with each other (Figure 5E). Subsequently, ROC analysis was carried out to assess the predictive value of these genes. ROC curves are commonly used to assess the performance of risk prediction models and the AUC quantifies this discriminative ability, with higher AUC values indicating better model performance. Moreover, an AUC curve that closely follows the top-left corner of the plot indicates higher accuracy, reflecting a model with strong predictive capability. In Figure 5F, the results indicated that each of the five genes exhibited relatively high predictive values (AUC values ≥ 0.921). Thus, these genes can serve as biomarkers for early diagnosis, disease monitoring, or evaluation the therapeutic effects. By conducting further research on the functions and mechanisms of these genes, we can gain a deeper understanding of the pathophysiological process of UC and provide clues for the development of new treatment approaches. By using the DSS-induced mouse colitis model, we aimed to substantiate the critical role of key anoikis genes in active UC. Because the *CEACAM6* gene is not present in rodents, and *H2-DMA* is orthologous to human *HLA-DMA*, we ultimately analyzed the expression of four key anoikis genes: *PDK4*, *H2-DMA*, *CX3CL1*, and *CFB* in UC and normal samples (Figure 5G). The qPCR results showed an increase in *H2-DMA*, *CX3CL1*, and *CFB* in the DSS group, while *PDK4* was decreased, consistent with the results from UC patients in Figure 4C.

3.6 The immune infiltration characteristics in UC samples and the association between key anoikis-DEGs and infiltrating immune cells

To determine if UC patients exhibited altered immune system activity, an assessment of the infiltration and correlations of 23 distinct immune cell subtypes in UC was conducted. Figure 6A depicts multiple correlations among the infiltrating immune cells in UC. A significant synergistic effect was observed between activated B cell, activated CD4 cell, activated CD8 cell, activated dendritic cell, eosinophil, gamma delta T cell, macrophage, mast cell, monocyte, natural killer T (NKT) cell, natural killer (NK) cell, neutrophil, plasmacytoid dendritic cell (pDC), regulatory T cell (Treg), type 1 T helper (Th1) cell, and type 2 T helper (Th2) cell. Moreover, activated dendritic cell, macrophage, monocyte, neutrophil, pDC, and Th1 cell exhibited synergism with Th17 cell. In contrast, a competitive effect was observed between CD56^{dim} NK cells and most other immune cells. Subsequently, the immune infiltration analysis was performed and compared between the UC and normal groups. A significant elevation of 22 immune cell subtypes was observed in the immune cell infiltration within UC group (Figure 6B). These immune cell types included activated B cell, activated CD4 cell, activated CD8 cell, activated dendritic cell, CD56^{bright} NK cell, eosinophil, gamma delta T cell, immature B cell, immature dendritic cell, MDSC, macrophage, mast cell, monocyte, NKT cell, NK cell, neutrophil, plasmacytoid dendritic cell, regulatory T cell, T follicular helper cell, Th1 cell, Th17 cell, and Th2 cell. These cell types were remarkably abundant in the UC samples.

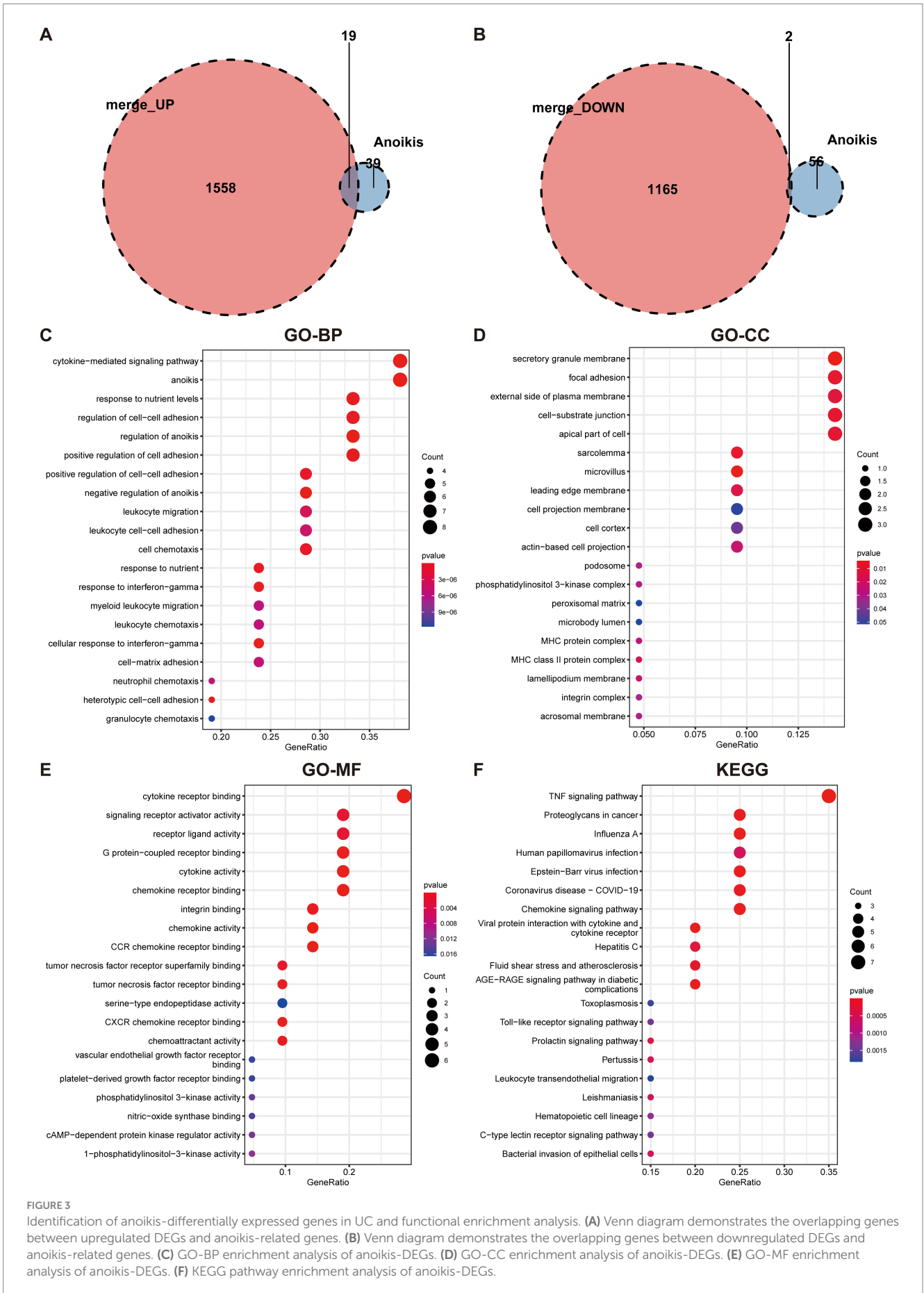
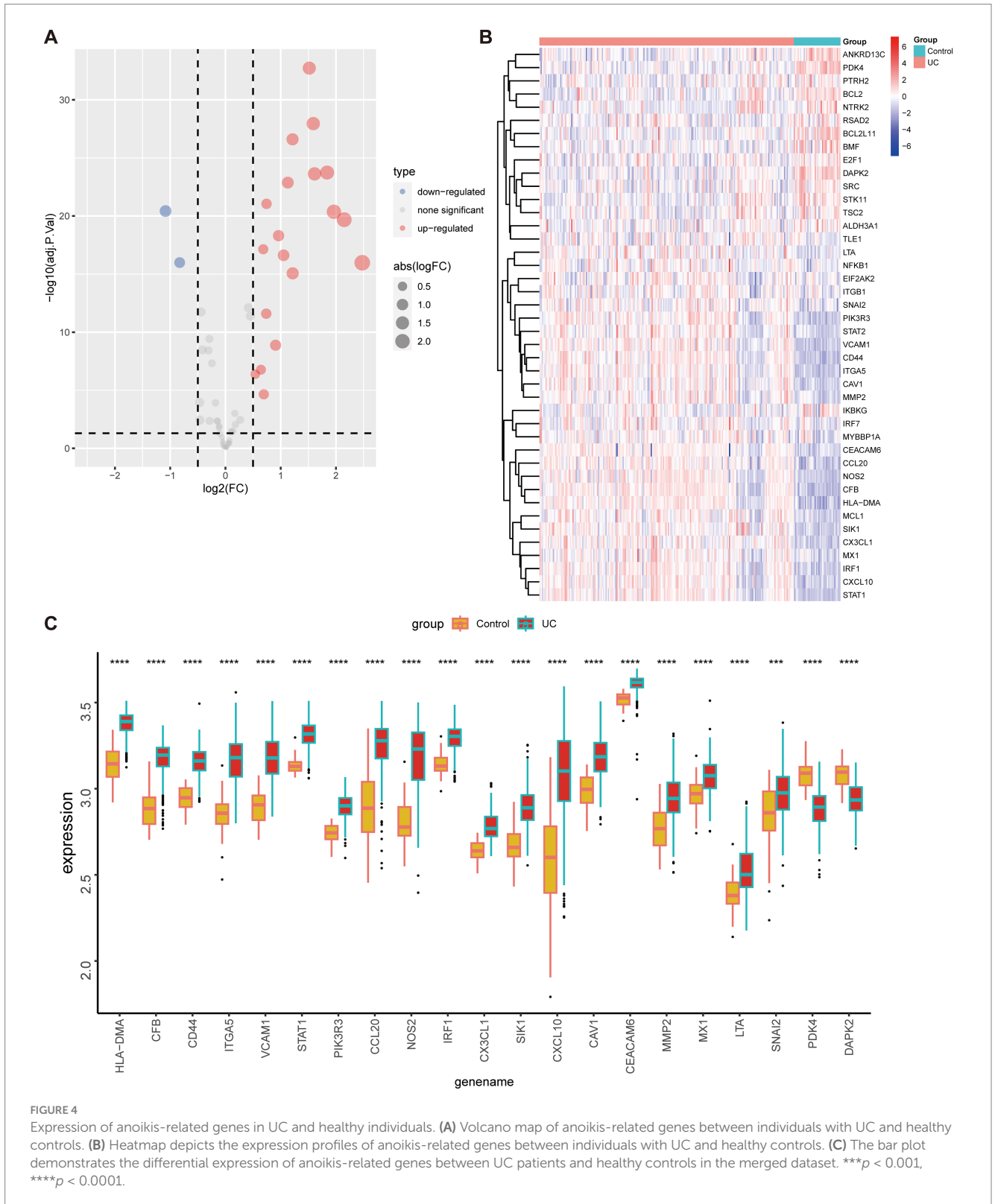


FIGURE 3 Identification of anoikis-differentially expressed genes in UC and functional enrichment analysis. **(A)** Venn diagram demonstrates the overlapping genes between upregulated DEGs and anoikis-related genes. **(B)** Venn diagram demonstrates the overlapping genes between downregulated DEGs and anoikis-related genes. **(C)** GO-BP enrichment analysis of anoikis-DEGs. **(D)** GO-CC enrichment analysis of anoikis-DEGs. **(E)** GO-MF enrichment analysis of anoikis-DEGs. **(F)** KEGG pathway enrichment analysis of anoikis-DEGs.



Furthermore, lollipop charts present the associations between the five key anoikis-DEGs and immune cell infiltration (Figure 6C). *CEACAM6*, *CFB*, *CX3CL1*, and *HLA-DMA* exhibited positive associations with most of the immune cells. In contrast, *PDK4* demonstrated a negative correlation with all immune cells.

3.7 Correlation of five key anoikis-DEGs with all UC genes and GSEA analysis

The correlations between five key anoikis-DEGs and all merged genes were investigated, and the top 50 positively correlated genes

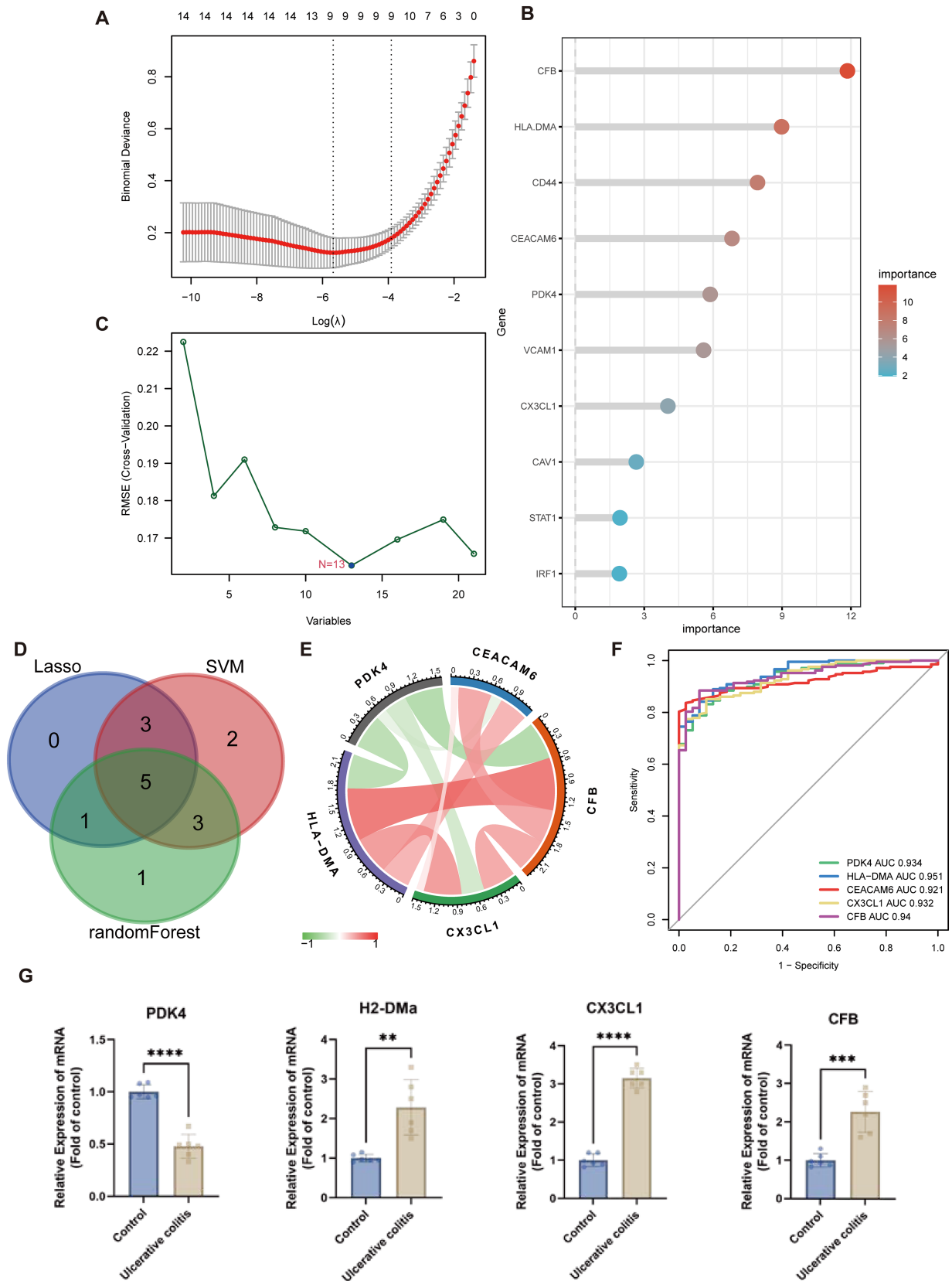


FIGURE 5
 Machine learning algorithms for identifying the key anoikis-DEGs. **(A)** LASSO Cox regression analysis. Vertical dashed lines are plotted at the optimal lambda value. **(B)** Random forest algorithm ranks the importance of anoikis-DEGs. **(C)** SVM-RFE algorithm for signature gene selection. **(D)** Venn diagram displays the overlapping genes between LASSO Cox regression, SVM-RFE algorithms, and random forest method. **(E)** The chord diagram shows the correlations between the five key anoikis-DEGs. **(F)** ROC curves of the five key anoikis-DEGs display their diagnostic value. **(G)** qPCR analysis validation of the key anoikis-DEGs in the mouse model of UC ($n = 6$ for each group). $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

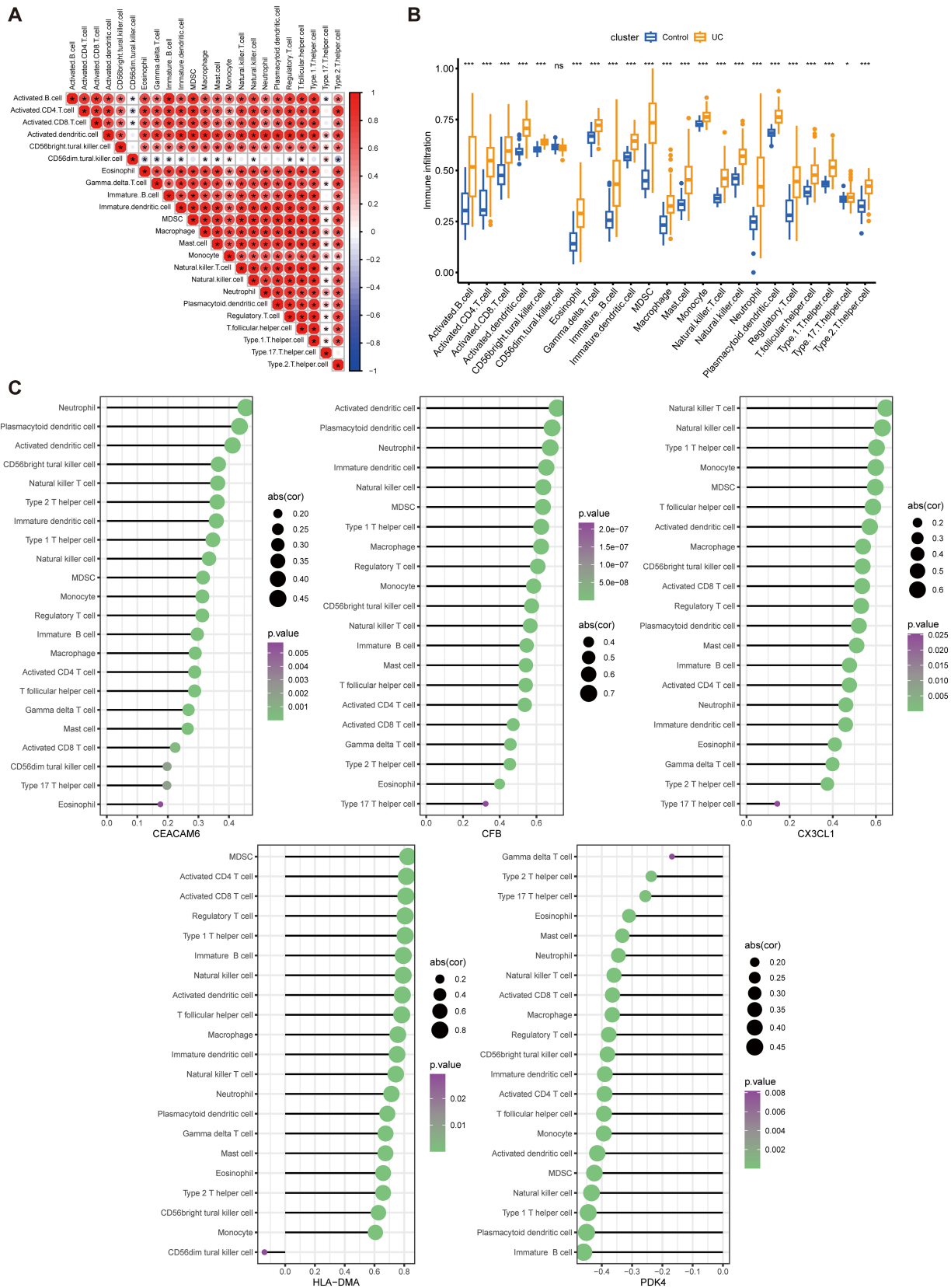
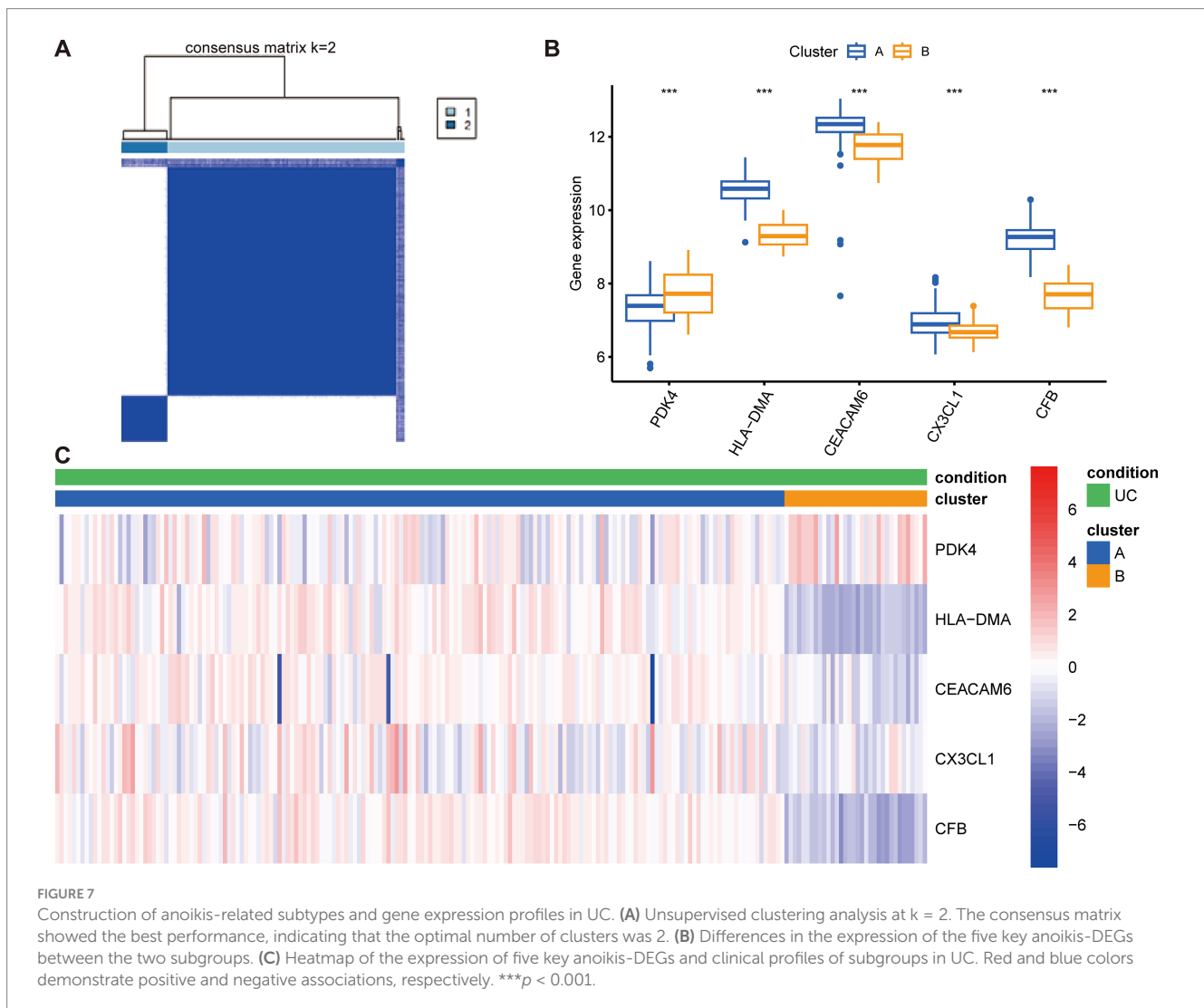


FIGURE 6
 Analysis of immune infiltration characteristics in UC samples and the correlation between key Anokis-DEGs and infiltrating immune cells. **(A)** Correlation analysis among 23 immune cells. The size and color of the circles indicate the correlation coefficient, with blue and red colors representing positive and negative correlation, respectively. The circle size is proportional to the correlation coefficient value. **(B)** Boxplots illustrate the variations in infiltrated immune cells between UC and healthy individuals. **(C)** Lollipop charts present the correlations between the five key anokis-DEGs genes and immune cell infiltration. The size and color of the circles demonstrate the correlation coefficient, with purple and green colors representing the *p*-value and the circle size indicating the correlation coefficient value. **p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001, ns: no significance.



were visualized in [Supplementary Figure 2](#). Moreover, the Reactome pathway enrichment analyses of the five key anoikis-DEGs were conducted utilizing GSEA to identify significant biological signatures, and the top 20 pathways were presented ([Supplementary Figure 3](#)). For each key anoikis-DEGs, the top 20 pathways had the same adjusted p -value, which indicates that, statistically, they are equally important. In UC, the Reactome pathways most positively associated with *CEACAM6* included the ER-phagosome pathway, antigen processing-cross presentation, asparagine N-linked glycosylation, and transport to the Golgi and subsequent modification. The Reactome pathways most positively associated with *CFB* were immunoregulatory interactions between a lymphoid and a non-lymphoid cell, interferon alpha/beta signaling, integrin cell surface interactions, and antigen processing-cross presentation. The top positively affected Reactome pathways of *CX3CL1* included immunoregulatory interactions between lymphoid and non-lymphoid cells, interleukin-4 and interleukin-13 signaling, interferon alpha/beta signaling, and integrin cell surface interactions. The significantly enriched pathways of *HLA-DMA* included immunoregulatory interactions between a lymphoid and a non-lymphoid cell, cytokine signaling in immune system, signaling by interleukins, and interleukin-4 and interleukin-13

signaling. Most pathways positively related to *PDK4* included the citric acid (TCA) cycle and respiratory electron transport, peroxisomal lipid metabolism, pyruvate metabolism and TCA cycle, and other pathways.

3.8 Construction of anoikis-related subtypes and gene expression profiles in UC

The 208 UC samples were classified as per the expression profiles of the five key anoikis DEGs utilizing an unsupervised clustering analysis. Various values of k were tested, varying from 2 to 9, and it was found that when $k = 2$, the consensus matrices showed the best performance, indicating that the optimal number of clusters was 2 ([Figure 7A](#)). Differences in the five key anoikis-DEGs expression between the two groups were observed. Specifically, higher expression of *HLA-DMA*, *CEACAM6*, *CX3CL1*, and *CFB* and lower expression of *PDK4* were observed in cluster A ([Figure 7B](#)). Furthermore, a heatmap demonstrated the relationships among the clinical features, expression of the five key anoikis-DEGs, and subtypes ([Figure 7C](#)).

3.9 Functional enrichment analysis between UC different subtypes

We subsequently conducted pathway enrichment analysis to investigate whether distinct pathways were perturbed in patients with different subtypes. The KEGG, Hallmark, and Reactome pathway analyses were performed to uncover the underlying pathways that differentiate cluster A from cluster B (Figure 8). Our findings revealed that these subtypes enriched in distinctive KEGG pathways. Cluster A displayed enhanced activation in antigen processing and presentation, intestinal immune network for IgA production, cytokine-cytokine receptor interaction, cell adhesion molecules, NOD-like receptor signaling pathway, Toll-like receptor signaling pathway, chemokine signaling pathway, and NK cell-mediated cytotoxicity. Furthermore, specific enrichment of tyrosine metabolism was observed in cluster B. In cluster A, Hallmark gene sets associated with the following pathways were upregulated: IL-6/JAK/STAT3 signaling, IL2/STAT5 signaling, apoptosis, TNF- α signaling via NF- κ B, reactive oxygen species pathway, and mTORC1 signaling. In contrast, bile acid metabolism, oxidative phosphorylation, and adipogenesis were upregulated in cluster B. Reactome analyses indicated that antigen presentation: folding assembly and peptide loading of class I MHC, chemokine receptors bind chemokines, interferon signaling, and TNFR2 non-canonical NF- κ B pathway were activated in cluster A. Taken together, these results suggest that different UC subtypes are regulated by distinct immune pathways.

3.10 Differential gene expression analysis and functional enrichment between UC subtypes

The distinct distribution of subsets was visualized through PCA to account for the differences between cluster A and B subtypes (Figure 9A). Moreover, a differential gene analysis comprising 579 genes of the two subtypes was conducted, presented as a volcano plot (Figure 9B). The GO enrichment analysis highlighted that the DEGs were primarily linked to pathways such as cytokine-mediated signaling pathway, leukocyte cell-cell adhesion, cell chemotaxis in BP, external side of plasma membrane, secretory granule membrane, collagen-containing extracellular matrix in CC, receptor ligand activity, signaling receptor activator activity, immune receptor activity, cytokine activity, and cytokine receptor activity in MF (Figure 9C). KEGG analysis indicated enrichment in cytokine-cytokine receptor interaction, cell adhesion molecules, chemokine signaling pathway, NOD-like receptor signaling pathway, TNF signaling pathway, IL-17 signaling pathway, and inflammatory bowel disease (Figure 9D). Moreover, Figure 9E illustrated the relationship between the top five KEGG pathways and differential genes.

3.11 Regulatory network of TFs and miRNAs associated with the key anoikis-DEGs

To examine the potential upstream regulators of the five key genes, the RegNetwork database was utilized to extract relationships

between upstream miRNAs and TFs likely to bind to these five key genes (Supplementary Figure 4). The results revealed a multitude of miRNAs and TFs implicated in the modulation of these diagnostic genes. TP53, RARB, RXRB, and CTCF were identified as regulators of multiple genes. Especially, CTCF is critically involved in regulating *CFB*, *CXCL3*, *HLA-DMA*, and *CEACAM6*.

4 Discussion

UC has emerged as a global health concern due to its high prevalence in developed nations and a significant rise in occurrence in developing countries (3, 41). Pathogenesis of UC is associated with disrupted intestinal barriers, imbalance of the gut microbiome, and subsequent dysregulated mucosal immune responses to gut commensal bacteria. At present, the primary treatment for moderate and severe UC is anti-TNF therapy, leading to a substantial improvement in treatment outcomes (41). Nevertheless, many patients remain unresponsive to anti-TNF therapy and develop colitis-associated colorectal dysplasia or cancer, warranting restorative proctocolectomy. Therefore, thoroughly exploring the potential mechanisms of UC and discovering new biomarkers that can aid in the development of novel UC treatment strategies are imperative. Considerable efforts from researchers have been directed toward investigating novel diagnostic techniques and therapeutic approaches with the aim of improving early diagnosis and treatment of UC. Anoikis fundamentally represents a form of programmed cell death that shares similarities with classical apoptosis (9). This process can manifest through two distinct pathways akin to classic apoptosis. One is the extrinsic pathway, initiated by death receptors located on the cell surface, while the other is the intrinsic pathway, which involves mitochondria-mediated mechanisms (Supplementary Figure 5). Nevertheless, the underlying mechanisms through which anoikis regulates UC remain in need of further exploration.

We thoroughly investigated the differential gene profiles between 208 UC samples and 38 control samples obtained from the GEO datasets. Moreover, three machine learning algorithms, including LASSO, RF and SVM, were employed to examine the involvement of ARGs in UC. The 2,744 DEGs between UC and healthy tissues exhibited enrichment in cytokine-cytokine receptor interaction, TNF signaling pathway, and Th17 cell differentiation pathways (Figure 2D). Furthermore, 21 dysfunctional ARGs were identified among UC individuals. This discovery underscores the plausible significance of anoikis in the progression of UC. Among them, *PDK4* and *DAPK2* exhibited relatively low expression levels in UC. In contrast, the other 19 genes, such as *CEACAM6*, *CFB*, *HLA-DMA*, *CX3CL1*, and others were notably enhanced in UC compared to normal samples (Figure 4C).

Recently, utilization of machine learning in diagnosing UC through the screening of key genes and immune cells has been widely used. This is attributed to its superior predictive performance, reduced error rates, and enhanced reliability (42, 43). In this study, five key anoikis-DEGs, *PDK4*, *HLA-DMA*, *CEACAM6*, *CX3CL1*, and *CFB* were identified as characteristic biomarkers of UC using LASSO, RF, and SVM algorithms (Figure 5). These five signature genes displayed strong diagnostic values (all AUC values ≥ 0.921). Therefore, these five specific genes possess predictive value for the occurrence of UC.

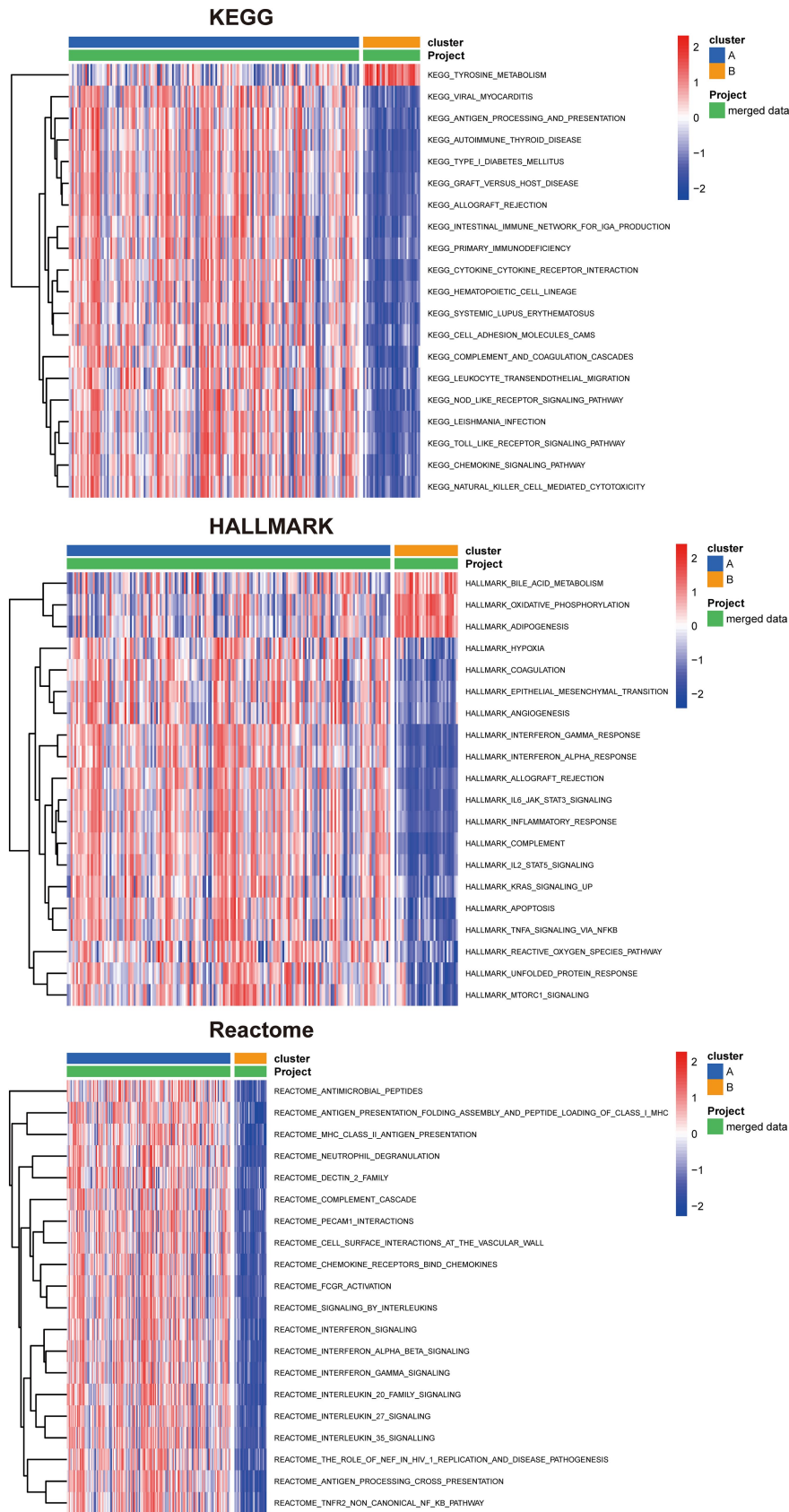


FIGURE 8 Identification of biological functional characteristics in different clusters of UC. The heatmap shows the top 20 significant KEGG, HALLMARK, and Reactome pathways between two clusters of UC separately.

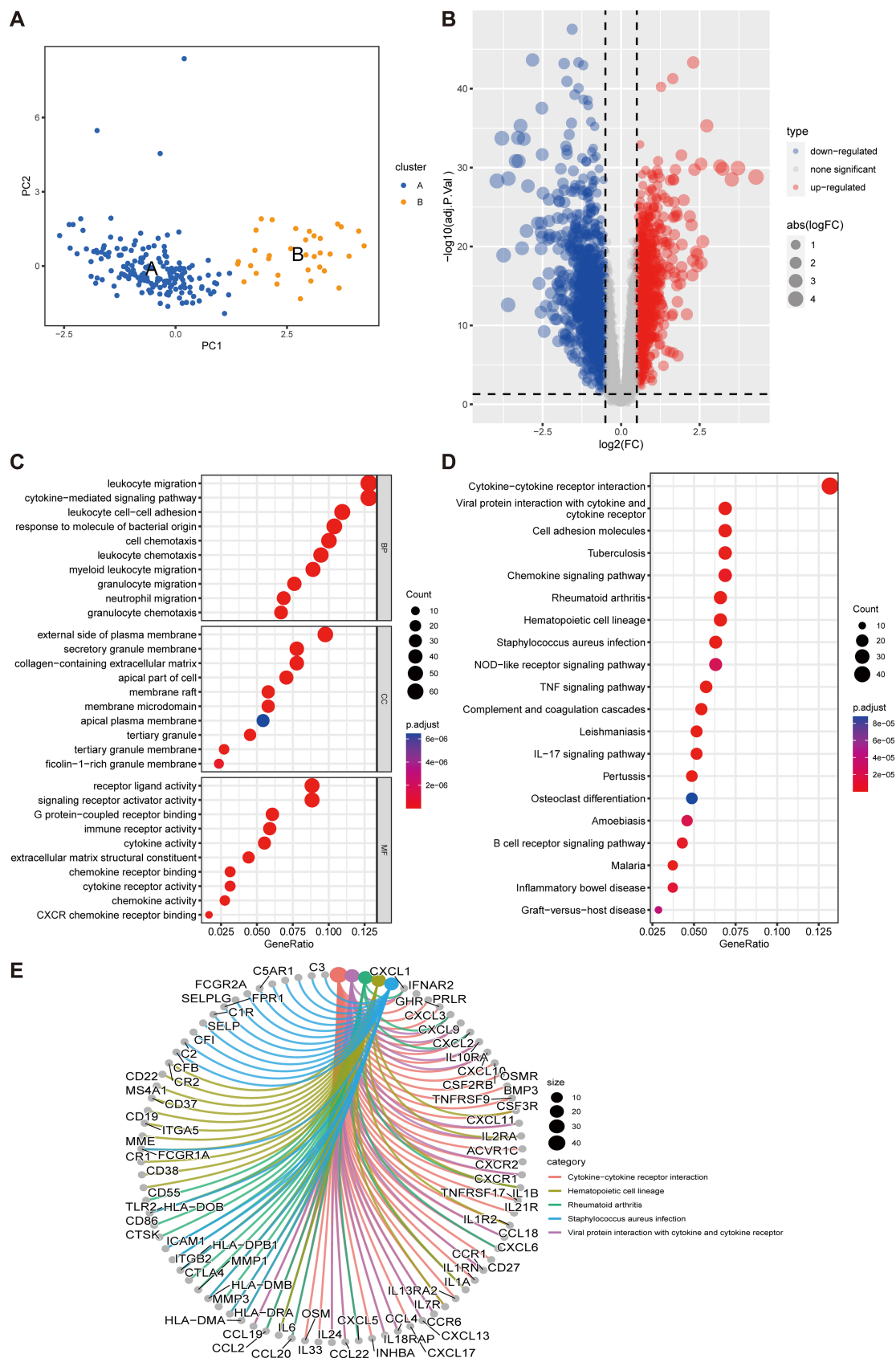


FIGURE 9 Differential gene expression analysis and functional annotation of UC subtypes. **(A)** Principal component analysis (PCA) to visually represent the distribution of the two subtypes. **(B)** Volcano diagram of 579 differentially expressed genes of the two subtypes, with red, blue, and gray dots representing upregulated, downregulated, and no significant difference, respectively. **(C,D)** GO and KEGG enrichment analyses. **(E)** Relationship between the top five KEGG pathways and differential genes.

Several researches have shown the participation of some of these key diagnostic genes in the pathogenesis of UC. For example, pyruvate dehydrogenase kinase isozyme 4 (PDK4) is crucial for regulating both glucose and fatty acid metabolism as well as maintaining homeostasis. Cyclosporine A is reported to facilitate neutrophil glycolysis and the TCA cycle by suppressing Sirtuin 6 (SIRT6) and promoting PDK4, ultimately alleviating clinical symptoms in severe UC (44). Our results demonstrate a negative association between PDK4 and neutrophil count in UC (Figure 6). The carcinoembryonic antigen-related adhesion molecules (CEACAM) are expressed on various cell types, comprising epithelial cells, neutrophils, and T cells (45). It is implicated in several processes, including proliferation, cell adhesion, differentiation, and tumor suppression. CEACAM6 is expressed on granulocytes and monocytes, and its expression is highly enhanced in individuals with Crohn's disease (CD), another chronic intestinal inflammatory disorder (46–48). A transgenic CEABAC10 mouse model infected with adherent-invasive *Escherichia coli* (AIEC) and expressing human CEACAM6 demonstrated increased proinflammatory cytokine IL-6 and IL-17 levels, reduced anti-inflammatory cytokine IL-10 levels, and histopathological damage to the gut mucosa (46). Meanwhile, blocking CEACAM6 with monoclonal antibodies reduced AIEC colonization and the inflammatory response. This suggests that CEACAM6 could be a potential treatment target for CD. The chemokine fractalkine (CX3CL1) is synthesized as a type I transmembrane protein (49). Vascular endothelial cells express CX3CL1, which is upregulated in response to proinflammatory cytokine stimulation. The CX3CR1, a receptor for fractalkine, is highly expressed on monocytes/macrophages, cytotoxic lymphocytes, and dendritic cells. The CX3CL1-CX3CR1 pathway is critically involved in gastrointestinal mucosal immunity. CX3CL1 is significantly upregulated in the mucosal epithelia and vascular endothelium of patients with active CD (50). Moreover, in a transfer IBD model, the anti-CX3CL1 monoclonal antibodies effectively mitigated the reduction in body weight, alleviated diarrhea, and reduced colon thickness. The gene Complement Factor B (CFB) encodes a secreted protein that plays a role in activating the alternative pathway of complement and is expressed primarily by the liver and mononuclear phagocytes (51, 52). The complement system is essential for pathogen lysis, opsonization, inflammation, and immune clearance. Previous research revealed that both CFB mRNA and protein exhibited elevated expression levels in the colonic mucosa of UC patients, indicating its involvement in the inappropriate activation of complement system (53). This activation contributes to chronic inflammation, resulting in active UC. Hence, the key anoikis-DEGs, *PDK4*, *CEACAM6*, *CX3CL1*, and *CFB* are closely related to immune cell infiltration and inflammatory pathways.

The primary pathological damage in UC arises from dysregulated immune responses, which leads to the production of inflammatory cytokines and the activation of immune cell signaling. These key anoikis genes are deeply involved in inflammatory processes and associated with the pathology of UC, particularly *CEACAM6* and *CX3CL1*. In the future, therapies targeting key anoikis-DEGs, especially blocking *CEACAM6* and *CX3CL1-CX3CR1* pathway with monoclonal antibodies could

effectively improve UC pathology by ameliorating the inflammatory response. Additionally, the current biomarkers for UC included serum anti- $\alpha\beta 6$ antibodies and serum oncostatin M for diagnosis, fecal calprotectin and serum trefoil for assessing disease activity, as well as whole blood transcriptomic panels and *CLEC5A/CDH2* ratio for determining the need for escalated treatment (5). The diagnostic and prognostic accuracy is often limited by factors such as variability in disease stage or response to treatment. However, the ARGs have considerable advantages as biomarkers for UC. *CEACAM6* has been associated with the activation of inflammatory pathways and immune cell migration (45). Elevated levels of *CEACAM6* have been associated with chronic inflammation and could serve as a more specific marker for immune dysregulation in UC compared to serum oncostatin M. *CX3CL1*, as a chemokine involved in immune cell trafficking, plays a crucial role in gastrointestinal mucosal immunity (50). It is significantly upregulated in active UC and could serve as a novel marker for UC and immune response regulation, complementing biomarkers like fecal calprotectin. *CFB* has been identified as a key component in the complement system (51, 52). Elevated expression of *CFB* in the colonic mucosa of UC patients suggests its involvement in immune activation and could serve as an early indicator of UC onset or exacerbation, in contrast to the more general serum trefoil factors used for disease activity assessment. *PDK4* plays a role in metabolic regulation and cell survival (44). Its downregulation is associated with inflammation and immune cell infiltration, suggesting that it may serve as a marker for UC progression and tissue damage, potentially complementing whole blood transcriptomic panels used for determining UC severity. By focusing on these key ARGs, we highlighted their potential roles as specific and reliable biomarkers for UC, particularly in identifying disease activity, immune dysregulation, and the need for targeted therapeutic interventions. These ARGs can provide a more targeted approach to managing UC, improving diagnostic accuracy, and guiding clinical decisions. Of note, DSS-induced colitis mouse model was constructed to demonstrate that the expression of key anoikis genes, namely *H2-DMA*, *CX3CL1*, and *CFB* were significantly upregulated, while *PDK4* was evidently downregulated. These results are in accordance with the previous results. However, additional researches are necessary to reveal the underlying mechanisms of *HLA-DMA* in UC. In summary, the studies involving these key anoikis-DEGs validate the reliability of our screening outcomes.

UC is characterized by a persistent chronic inflammatory triggered by an overactive immune response against gut bacteria or dietary antigens (3). This abnormal immune response involves both innate and adaptive immune cells, playing pivotal roles in the development and escalation of UC. Among these, innate immune cells such as dendritic cells, neutrophils, and macrophages have a substantial effect in the pathogenesis of UC. Additionally, innate lymphoid cells, which have garnered increased attention recently, also contribute to the disease. Simultaneously, the adaptive immune response mechanisms encompass various types of cells, including cytotoxic T lymphocytes, Tregs, and different subsets of helper T lymphocytes (Th), such as Th1, Th2, Th9, Th17, and Th22 (7). A recent study revealed increased abundances of $IL1B^+$ macrophages and

monocytes, IL17A⁺ CD161⁺ effector memory T cells, and IL17A⁺ regulatory T cells in colonic mucosa samples from UC patients (54). Additionally, a significant accumulation of CD11b⁺ B cells has been observed in the intestinal lamina propria and Peyer's patches in both mouse models with colitis induced by dextran sulfate sodium and individuals with UC. The adoptive transfer of CD11b⁺ B cells successfully ameliorated colitis symptoms and showed therapeutic benefits (55). In this research, we analyzed the immune cell infiltration and found higher enrichment of activated dendritic cells, macrophages, neutrophils, multiple T cells (activated CD4, activated CD8, gamma delta T, NKT, Treg, Th1, Th2, and Th17 cells), and B cells (activated and immature B cells) in UC compared to normal group. Further, our results showed that among the key anoikis-DEGs, *CEACAM6*, *CFB*, *CX3CL1*, and *HLA-DMA* exhibited a positive association with immune cells, while *PDK4* exhibited a negative association with immune cells (Figure 6C). These findings deepen our understanding of immune dysregulation and immune cell involvement in UC.

The enrichment analysis revealed the involvement of five key anoikis genes in immunoregulatory interactions between lymphoid and non-lymphoid cells, interferon alpha/beta signaling, interleukin-4 and interleukin-13 signaling, and antigen processing-cross presentation, among others (Supplementary Figure 3). Using the clustering algorithm, patients were classified into two clusters according to the five key anoikis genes (Figure 7A). *HLA-DMA*, *CEACAM6*, *CX3CL1*, and *CFB* exhibited high expression levels in cluster A, whereas *PDK4* displayed relatively low expression in cluster A (Figure 7B). Moreover, the relationship among the clinical profiles, five key anoikis gene expression, and subtypes was visualized using a heatmap (Figure 7C). To better understand the differences between UC subtypes, the underlying pathways were analyzed, revealing that the two subtypes displayed distinct pathways (Figure 8). Additionally, the interaction network among these characteristic genes, miRNA, and TFs were explored (Supplementary Figure 4), providing insights into the upstream signaling pathways for our subsequent research. In the future, we will persist in exploring the potential mechanisms underlying UC by conducting molecular biology experiments.

Nevertheless, this research has some limitations. Initially, this research relied on a publicly available dataset, and it is crucial to emphasize that these findings generated using bioinformatics techniques require further validation to ensure their reliability. Furthermore, it is essential to highlight that the expression levels of the above genes among individuals from distinct regions or racial backgrounds remain unclear. More *in vitro* and *in vivo* studies are needed to fill these knowledge gaps and understand how these crucial anoikis genes are linked to different immune signaling pathways in UC. These experiments will provide deeper insights into the potential mechanisms underlying the correlation of *PDK4*, *HLA-DMA*, *CEACAM6*, *CX3CL1*, and *CFB* with the infiltration of immune cells in UC, helping to establish a more comprehensive understanding of the disease.

5 Conclusion

This research offers a thorough examination of the involvement of ARGs in UC, marking the first instance of revealing the expression

profiles of ARGs in UC and their association with immune cells infiltration. Machine learning algorithms and unsupervised clustering analysis based on ARGs were employed to identify five signature genes: *PDK4*, *HLA-DMA*, *CEACAM6*, *CX3CL1*, and *CFB*. These genes are critically involved in immune cell infiltration and immune signaling pathways, showing a considerable diagnostic value. The results offer insights into the classification of UC patients into two clusters, each regulated by distinct pathways. Our findings have the potential to be a valuable reference, offering deeper insights into the underlying mechanisms of anoikis in UC. They could serve as a foundation for the development of innovative strategies in drug screening, personalized treatment, and immunotherapy for individuals with UC.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements. The animal study was approved by Tongji University Animal Ethics Committee. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

PL: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. CS: Conceptualization, Methodology, Validation, Writing – review & editing. XW: Conceptualization, Funding acquisition, Methodology, Resources, Software, Writing – original draft, Writing – review & editing. BH: Writing – review & editing. YS: Writing – review & editing. YL: Supervision, Validation, Writing – review & editing. XZ: Conceptualization, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding for this study was provided by the National Natural Science Foundation of China grant nos. (82103095 and 82104242), the China Postdoctoral Foundation (2024 M762430), and the Minhang District Healthcare System Program for Outstanding Young Pharmacology Scholars (mwjyxx01).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2025.1498864/full#supplementary-material>

References

1. Bisgaard TH, Allin KH, Keefer L, Ananthakrishnan AN, Jess T. Depression and anxiety in inflammatory bowel disease: epidemiology, mechanisms and treatment. *Nat Rev Gastroenterol Hepatol.* (2022) 19:717–26. doi: 10.1038/s41575-022-00634-6
2. Nakase H, Uchino M, Shinzaki S, Matsuura M, Matsuoka K, Kobayashi T, et al. Evidence-based clinical practice guidelines for inflammatory bowel disease 2020. *J Gastroenterol.* (2021) 56:489–526. doi: 10.1007/s00535-021-01784-1
3. Le Berre C, Honap S, Peyrin-Biroulet L. Ulcerative colitis. *Lancet.* (2023) 402:571–84. doi: 10.1016/s0140-6736(23)00966-2
4. Shao B, Yang W, Cao Q. Landscape and predictions of inflammatory bowel disease in China: China will enter the compounding prevalence stage around 2030. *Front Public Health.* (2022) 10:1032679. doi: 10.3389/fpubh.2022.1032679
5. Nowak JK, Kalla R, Satsangi J. Current and emerging biomarkers for ulcerative colitis. *Expert Rev Mol Diagn.* (2023) 23:1107–19. doi: 10.1080/14737159.2023.2279611
6. Hong D, Kim HK, Yang W, Yoon C, Kim M, Yang CS, et al. Integrative analysis of single-cell rna-seq and gut microbiome metabarcoding data elucidates macrophage dysfunction in mice with dss-induced ulcerative colitis. *Commun Biol.* (2024) 7:731. doi: 10.1038/s42003-024-06409-w
7. Kalužna A, Olczyk P, Komosińska-Vashev K. The role of innate and adaptive immune cells in the pathogenesis and development of the inflammatory response in ulcerative colitis. *J Clin Med.* (2022) 11:400. doi: 10.3390/jcm11020400
8. Adeshakin FO, Adeshakin AO, Afolabi LO, Yan D, Zhang G, Wan X. Mechanisms for modulating anoikis resistance in cancer and the relevance of metabolic reprogramming. *Front Oncol.* (2021) 11:626577. doi: 10.3389/fonc.2021.626577
9. Taddei ML, Giannoni E, Fiaschi T, Chiarugi P. Anoikis: an emerging hallmark in health and diseases. *J Pathol.* (2012) 226:380–93. doi: 10.1002/path.3000
10. Zhu Z, Fang C, Xu H, Yuan L, Du Y, Ni Y, et al. Anoikis resistance in diffuse glioma: the potential therapeutic targets in the future. *Front Oncol.* (2022) 12:976557. doi: 10.3389/fonc.2022.976557
11. Alanko J, Ivaska J. Endosomes: emerging platforms for integrin-mediated fak signaling. *Trends Cell Biol.* (2016) 26:391–8. doi: 10.1016/j.tcb.2016.02.001
12. Li K, Zhao G, Ao J, Gong D, Zhang J, Chen Y, et al. Znf 32 induces anoikis resistance through maintaining redox homeostasis and activating src/fak signaling in hepatocellular carcinoma. *Cancer Lett.* (2019) 442:271–8. doi: 10.1016/j.canlet.2018.09.033
13. Mason JA, Davison-Versagli CA, Leliaert AK, Pape DJ, McCallister C, Zuo J, et al. Oncogenic ras differentially regulates metabolism and anoikis in extracellular matrix-detached cells. *Cell Death Differ.* (2016) 23:1271–82. doi: 10.1038/cdd.2016.15
14. Song J, Liu Y, Liu F, Zhang L, Li G, Yuan C, et al. The 14-3-3 σ protein promotes hcc anoikis resistance by inhibiting egfr degradation and thereby activating the egfr-dependent erk 1/2 signaling pathway. *Theranostics.* (2021) 11:996–1015. doi: 10.7150/thno.51646
15. Sharma R, Gogoi G, Saikia S, Sharma A, Kalita DJ, Sarma A, et al. Bmp4 enhances anoikis resistance and chemoresistance of breast cancer cells through canonical bmp signaling. *J Cell Commun Signal.* (2022) 16:191–205. doi: 10.1007/s12079-021-00649-9
16. Wang LN, Zhang ZT, Wang L, Wei HX, Zhang T, Zhang LM, et al. Tgf- β 1/sh2b3 axis regulates anoikis resistance and emt of lung cancer cells by modulating jak2/stat3 and shp2/grb2 signaling pathways. *Cell Death Dis.* (2022) 13:472. doi: 10.1038/s41419-022-04890-x
17. Moro L, Arbini AA, Yao JL, di Sant'Agnese PA, Marra E, Greco M. Mitochondrial DNA depletion in prostate epithelial cells promotes anoikis resistance and invasion through activation of pi3k/akt2. *Cell Death Differ.* (2009) 16:571–83. doi: 10.1038/cdd.2008.178
18. Luey BC, May FE. Insulin-like growth factors are essential to prevent anoikis in oestrogen-responsive breast cancer cells: importance of the type i igf receptor and pi3-kinase/akt pathway. *Mol Cancer.* (2016) 15:8. doi: 10.1186/s12943-015-0482-2
19. Kaur A, Goggolidou P. Ulcerative colitis: understanding its cellular pathology could provide insights into novel therapies. *J Inflamm (Lond).* (2020) 17:15. doi: 10.1186/s12950-020-00246-4
20. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med.* (2013) 369:699–710. doi: 10.1056/NEJMoa1215734
21. Panaccione R, Ghosh S, Middleton S, Márquez JR, Scott BB, Flint L, et al. Combination therapy with infliximab and azathioprine is superior to monotherapy with either agent in ulcerative colitis. *Gastroenterology.* (2014) 146:392–400.e3. doi: 10.1053/j.gastro.2013.10.052
22. Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, Johanns J, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med.* (2005) 353:2462–76. doi: 10.1056/NEJMoa050516
23. Sands BE, Sandborn WJ, Panaccione R, O'Brien CD, Zhang H, Johanns J, et al. Ustekinumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med.* (2019) 381:1201–14. doi: 10.1056/NEJMoa1900750
24. Sandborn WJ, Su C, Sands BE, D'Haens GR, Vermeire S, Schreiber S, et al. Tofacitinib as induction and maintenance therapy for ulcerative colitis. *N Engl J Med.* (2017) 376:1723–36. doi: 10.1056/NEJMoa1606910
25. Sun Z, Zhao Y, Wei Y, Ding X, Tan C, Wang C. Identification and validation of an anoikis-associated gene signature to predict clinical character, stemness, idh mutation, and immune infiltration in glioblastoma. *Front Immunol.* (2022) 13:939523. doi: 10.3389/fimmu.2022.939523
26. Chi H, Jiang P, Xu K, Zhao Y, Song B, Peng G, et al. A novel anoikis-related gene signature predicts prognosis in patients with head and neck squamous cell carcinoma and reveals immune infiltration. *Front Genet.* (2022) 13:984273. doi: 10.3389/fgene.2022.984273
27. Diao X, Guo C, Li S. Identification of a novel anoikis-related gene signature to predict prognosis and tumor microenvironment in lung adenocarcinoma. *Thorac Cancer.* (2023) 14:320–30. doi: 10.1111/1759-7714.14766
28. Ye G, Yang Q, Lei X, Zhu X, Li F, He J, et al. Nuclear myh9-induced cttnb1 transcription, targeted by staurosporin, promotes gastric cancer cell anoikis resistance and metastasis. *Theranostics.* (2020) 10:7545–60. doi: 10.7150/thno.46001
29. Buchheit CL, Angarola BL, Steiner A, Weigel KJ, Schafer ZT. Anoikis evasion in inflammatory breast cancer cells is mediated by bim-el sequestration. *Cell Death Differ.* (2015) 22:1275–86. doi: 10.1038/cdd.2014.209
30. Jin L, Chun J, Pan C, Kumar A, Zhang G, Ha Y, et al. The plg1-gdh1 axis promotes anoikis resistance and tumor metastasis through camkk2-ampk signaling in lkb1-deficient lung cancer. *Mol Cell.* (2018) 69:87–99.e7. doi: 10.1016/j.molcel.2017.11.025
31. Zhang D, Zhou X, Zhang K, Yu Y, Cui SW, Nie S. Glucosaminan from *Aloe vera* gel maintains intestinal barrier integrity via mitigating anoikis mediated by nrf2-mitochondria axis. *Int J Biol Macromol.* (2023) 235:123803. doi: 10.1016/j.ijbiomac.2023.123803
32. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034
33. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
34. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics.* (2019) 11:123. doi: 10.1186/s13148-019-0730-1
35. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* (2019) 37:773–82. doi: 10.1038/s41587-019-0114-2
36. Van Belle V, Van Calster B, Van Huffel S, Suykens JA, Lisboa P. Explaining support vector machines: a color based nomogram. *PLoS One.* (2016) 11:e0164568. doi: 10.1371/journal.pone.0164568
37. Yi M, Nissley DV, McCormick F, Stephens RM. Ssgsea score-based ras dependency indexes derived from gene expression data reveal potential ras addiction mechanisms with possible clinical implications. *Sci Rep.* (2020) 10:10258. doi: 10.1038/s41598-020-66986-8

38. Markovich Gordon M, Moser AM, Rubin E. Unsupervised analysis of classical biomedical markers: robustness and medical relevance of patient clustering using bioinformatics tools. *PLoS One*. (2012) 7:e29578. doi: 10.1371/journal.pone.0029578
39. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (msigdb) hallmark gene set collection. *Cell Syst*. (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004
40. Liu ZP, Wu C, Miao H, Wu H. Regnetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database (Oxford)*. (2015) 2015:95. doi: 10.1093/database/bav095
41. Chudy-Onwugaje KO, Christian KE, Farraye FA, Cross RK. A state-of-the-art review of new and emerging therapies for the treatment of ibd. *Inflamm Bowel Dis*. (2019) 25:820–30. doi: 10.1093/ibd/izy327
42. 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
43. Brugnara G, Neuberger U, Mahmutoglu MA, Foltyn M, Herweh C, Nagel S, et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke*. (2020) 51:3541–51. doi: 10.1161/strokeaha.120.030287
44. Lu H, Lin J, Xu C, Sun M, Zuo K, Zhang X, et al. Cyclosporine modulates neutrophil functions via the sirt6-hif-1 α -glycolysis axis to alleviate severe ulcerative colitis. *Clin Transl Med*. (2021) 11:e334. doi: 10.1002/ctm2.334
45. Kelleher M, Singh R, O'Driscoll CM, Melgar S. Carcinoembryonic antigen (ceacam) family members and inflammatory bowel disease. *Cytokine Growth Factor Rev*. (2019) 47:21–31. doi: 10.1016/j.cytogfr.2019.05.008
46. Carvalho FA, Barnich N, Sivignon A, Darcha C, Chan CH, Stanners CP, et al. Crohn's disease adherent-invasive *Escherichia coli* colonize and induce strong gut inflammation in transgenic mice expressing human ceacam. *J Exp Med*. (2009) 206:2179–89. doi: 10.1084/jem.20090741
47. Barnich N, Carvalho FA, Glasser AL, Darcha C, Jantscheff P, Allez M, et al. Ceacam6 acts as a receptor for adherent-invasive e. coli, supporting ileal mucosa colonization in crohn disease. *J Clin Invest*. (2007) 117:1566–74. doi: 10.1172/jci30504
48. Hammarström S. The carcinoembryonic antigen (Cea) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol*. (1999) 9:67–81. doi: 10.1006/scbi.1998.0119
49. Lucas AD, Chadwick N, Warren BF, Jewell DP, Gordon S, Powrie F, et al. The transmembrane form of the cx3cl1 chemokine fractalkine is expressed predominantly by epithelial cells *in vivo*. *Am J Pathol*. (2001) 158:855–66. doi: 10.1016/s0002-9440(10)64034-5
50. Nishimura M, Kuboi Y, Muramoto K, Kawano T, Imai T. Chemokines as novel therapeutic targets for inflammatory bowel disease. *Ann N Y Acad Sci*. (2009) 1173:350–6. doi: 10.1111/j.1749-6632.2009.04738.x
51. Garnier G, Ault B, Kramer M, Colten HR. Cis and trans elements differ among mouse strains with high and low extrahepatic complement factor b gene expression. *J Exp Med*. (1992) 175:471–9. doi: 10.1084/jem.175.2.471
52. Wu LC, Morley BJ, Campbell RD. Cell-specific expression of the human complement protein factor b gene: evidence for the role of two distinct 5'-flanking elements. *Cell*. (1987) 48:331–42. doi: 10.1016/0092-8674(87)90436-3
53. Ostvik AE, Granlund A, Gustafsson BI, Torp SH, Espevik T, Mollnes TE, et al. Mucosal toll-like receptor 3-dependent synthesis of complement factor b and systemic complement activation in inflammatory bowel disease. *Inflamm Bowel Dis*. (2014) 20:995–1003. doi: 10.1097/mib.0000000000000035
54. Mitsialis V, Wall S, Liu P, Ordovas-Montanes J, Parmet T, Vukovic M, et al. Single-cell analyses of colon and blood reveal distinct immune cell signatures of ulcerative colitis and crohn's disease. *Gastroenterology*. (2020) 159:591–608.e10. doi: 10.1053/j.gastro.2020.04.074
55. Fu Y, Wang Z, Yu B, Lin Y, Huang E, Liu R, et al. Intestinal cd11b(+) b cells ameliorate colitis by secreting immunoglobulin a. *Front Immunol*. (2021) 12:697725. doi: 10.3389/fimmu.2021.697725