European Review for Medical and Pharmacological Sciences 2023; 27: 8447-8462

Age-related genes affecting the immune cell infiltration in ulcerative colitis revealed by weighted correlation network analysis and machine learning

H.-L. CHEN[1], Y.-H. LIU[1], C.-H. TAN[2]

[1]Department of Anus-Intestines, The Eighth People's Hospital of Hefei, Hefei, Anhui Province, China
[2]Department of Medical Oncology, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui Province, China

H.-L. Chen, Y.-H. Liu, and C.-H. Tan contributed equally to this work

Abstract. – OBJECTIVE: The crosstalk between age and immunity in the context of ulcerative colitis (UC) remains incompletely understood. Our objective is to elucidate the specific age-associated genetic factors that modulate immune cell infiltration in UC, with the aim of identifying innovative therapeutic targets for the treatment of this disease.

MATERIALS AND METHODS: Potential batch effects between samples were removed by R package "inSilicoMerging". Unsupervised clustering analysis *via* the "ConsensusClusterPlus" R package was utilized to perform consensus molecular subtyping of immune subtypes in UC. The construction of a heat map was accomplished through the utilization of the R package "pheatmap", while functional enrichment analysis was executed by means of the Metascape database. The identification of the age-related gene module was achieved by performing weighted gene co-expression network analysis (WGCNA) analysis using the R package "WGCNA". The support vector machine (SVM), least absolute shrinkage and selector operation (LASSO), and random forest algorithms were performed *via* the "e1071", "glmnet" and "randomForest" packages in R, respectively. The diagnostic performance of the parameter was assessed using the receiver operating characteristic (ROC) curve. Correlation analysis was performed by Spearman correlation. The "XSum" package in R was employed to identify potential small-molecule drugs for UC utilizing the Connectivity Map (CMap) database. Molecular docking was performed with Autodock Vina molecular docking software.

RESULTS: A significantly greater frequency of UC patients aged below 40 years was observed in the group with extensive disease extent as compared to those with non-extensive disease extent (70% *vs.* 47%; Chi-square test, $p = 0.02$). The application of unsupervised clustering analysis allowed for the stratification of UC patients into two distinct immune subtypes, namely cluster C1 and cluster C2. The distribution of immune subtypes was significantly different between different age categories (Chi-square test, $p = 0.00219$). The UC samples that were grouped under cluster C1 were distinguished by a higher abundance of macrophages and an elevated number of neutrophils relative to those in cluster C2. Based on both WGCNA and Limma analysis, 146 age-related genes were identified, which exhibited a predominant enrichment in the biological process of cellular senescence. Two age-related genes (MIDN, and PLD6) affecting the immune cell infiltration in UC were identified based on machine learning algorithms (SVM, LASSO, and random forest). The diagnostic performance of MIDN (AUC = 0.93) and PLD6 (AUC = 0.90) in discerning UC patients belonging to cluster C1 was found to be satisfactory, as demonstrated by ROC curve analysis. MIDN demonstrated a positive correlation ($r = 0.50$, $p < 0.0001$) with Neutrophil, while PLD6 exhibited a negative correlation ($r = -0.52$, $p < 0.0001$) with Neutrophil levels. The "XSum" algorithm revealed that Entinostat has therapeutic potential for UC. The docking glide score between Entinostat and MIDN, and PLD6 protein was -8.9 kcal/mol and -6.8 kcal/mol, respectively.

CONCLUSIONS: We have identified two age-related genes, MIDN and PLD6, that are involved in immune cell infiltration in patients with ulcerative colitis. Furthermore, a small molecule drug (Entinostat) with potential therapeutic effects for UC was screened out. This study presented new perspectives on personalized clinical management and therapy research for UC.

Key Words:
Ulcerative Colitis, Aging, Neutrophil, Inflammation, Genomics, Bioinformatics.

Corresponding Author: Hulin Chen, MD; e-mail: chenhulin2019@163.com

Introduction

Ulcerative colitis (UC) is a chronic relapsing bowel disease characterized by a diffuse inflammation that generates an excess of detrimental injury on colonic mucosa with bloody diarrhea as a predominant symptom[1]. UC has become a globally prevalent disease with increasing incidence rates in both developed and developing nations, making UC a major health burden globally[2]. Although UC can manifest at any age, it is most commonly observed in individuals between their second and fourth decades of life[3]. As yet, the exact etiology of UC remains unclear. Genetic and environmental factors, impaired epithelial barrier function, and dysregulated immune response is hypothesized to be an underlying cause of pathogenesis in UC[2,4].

The intestinal epithelium serves as a protective barrier in the gut, preventing luminal microbiota and antigenic material from infiltrating the underlying lamina propria, which harbors the mucosal immune system[5]. However, compromised intestinal barrier function can lead to bacterial translocation, resulting in localized inflammation and immune cell infiltration in the upper regions of the colonic mucosa. The cellular infiltrates include cells from both the innate and the adaptive immune responses, such as neutrophils , dendritic cells , natural killer T cells (NKTs) , macrophages, and T cells[5-12]. Furthermore, the activated immune cells are responsible for the production of several cytokines that regulate cellular function, such as tumor necrosis factor (TNF), interferon gamma (IFNγ), interleukin-1β (IL-1β), IL-6, and IL-23, as well as T helper (Th) 17 cell-associated cytokines. The recruitment of leukocytes is regulated by chemokines, which serve as a hallmark of inflammation[13,14].

Although UC can affect individuals of any age, recent investigations have suggested that early-onset UC exhibits distinct phenotypic characteristics compared to those of older-onset UC. The underlying mechanisms accounting for the age-related phenotypic distinctions in UC are not fully comprehended and could be related to variations in intestinal immunity, intestinal microbiota, and genetic and environmental risk factors[15]. In general, the natural disease course of pediatric-onset UC is considered to be more severe than that of elderly-onset patients[16]. Early-onset UC is featured by widespread location at diagnosis and a high rate of disease extension[17-19]. Previous studies[20-22] have detected obvious differences in both the systemic and mucosal immune systems between young and elderly patients. The elderly have the added challenge of immune aging, which is related to heightened susceptibility to infection, vaccine failure, autoimmunity, and cancer. In the mucosal immune system, more gut-associated lymphoreticular tissues (GALT), were observed in the young compared to the elderly. The decline in the level of MALT (mucosa-associated lymphoid tissue) cells, Peyer's patches, isolated lymphoid follicles, and immunoglobulin occurs more frequently in older adults[23-25].

The interplay between age and immune activity remains incompletely elucidated with regard to UC. Our present investigation endeavors to examine the influence of age-related genes on immune cell infiltration in UC, with the ultimate goal of promoting tailored clinical interventions and treatment strategies. A graphical overview of our study protocol is provided in **Supplementary Figure 1.**

Materials and Methods

Data Acquisition

The microarray data and clinical information of UC patients were acquired from the Gene Expression Omnibus (GEO) database, specifically GSE38713 and GSE87473. GSE38713 and GSE87473 cohorts contained patient age information for all samples. The sample sizes for the two datasets analyzed in this study were as follows: GSE38713 (13 normal samples and 30 UC samples), and GSE87473 (21 normal samples and 106 UC samples). The microarray data was downloaded at https://www.ncbi.nlm.nih.gov/geo/ on November 1, 2022[26]. The R package "inSilicoMerging" was used for dataset normalization[27]. The batch effects across platforms were removed by the "ComBat" algorithm of the R package "inSilicoMerging". As an empirical Bayesian method, the "ComBat" algorithm estimated the parameters representing the batch effect by summarizing the information among genes in each batch, thereby reducing the batch effect parameters to the overall estimated average[28]. By merging the GSE38713 and GSE87473 cohorts, we acquired a unified dataset comprising a total of 170 samples, including 136 UC and 34 normal colon tissues.

Identification of Immune Infiltration Subtype Characterization of UC

We utilized the Immune Cell Abundance Identifier (ImmuCellAI) database to determine the

levels of 24 distinct immune cell types in the merged UC dataset by inputting the microarray data[29,30]. We employed the R package "ConsensusClusterPlus" to conduct an unsupervised hierarchical clustering analysis on the abundance of distinct immune cells in the merged UC dataset, aiming to identify diverse immune infiltration subtypes[31-33].

Weighted Correlation Network Analysis (WGCNA) and Gene Differential Analysis

R package "WGCNA" was utilized to construct the coexpression networks based on the microarray data[34]. As a soft-thresholding power, the primary role of β was to emphasize strong correlations between the genes and penalize weak correlations. The topological overlap matrix (TOM) was transformed from the adjacency after we chose the β of 20 based on the "pickSoftThreshold" algorithm which came with the "WGCNA" R package[35]. Pearson's correlation analysis was conducted to appraise the correlation between module eigengenes (MEs) and age. Subsequently, the gene module with the highest Pearson's coefficient was considered as the module most relevant to the age (age-related module) in UC. Using the criteria of $|MM| > 0.8$ and $|GS| > 0.1$, we identified the unique hub genes in the age-related module[33]. Specific schematic processes of WGCNA can be found in **Supplementary Figure 2**.

Differential gene expression analysis followed the linear models for microarray data (Limma) pipeline performed by R package "limma" (version 3.40.6). Differential expression genes (DEGs) were identified according to the filter criteria (|fold change| > 1.5, false discovery rate < 0.05)[36].

Gene Enrichment Analysis

The Metascape database was utilized to perform enrichment analyses. Terms with a p-value < 0.01, minimum count of 3, and an enrichment factor > 1.5 were utilized in the next step of the analysis[37]. Using screening criteria of kappa scores = 4 and similarity > 0.3. Metascape was utilized to perform hierarchical clustering to partition enrichment terms into distinct clusters, and the terms with the minimum p-value were selected as the representative terms.

Machine Learning for the Age-Immune-Related Key Gene Signatures

The identification of key age-immune-related gene signatures in the merged UC dataset was performed through the implementation of the support vector machine (SVM), least absolute shrinkage and selector operation (LASSO), and random forest algorithms, which were available through the "e1071", "glmnet", and "randomForest" packages in R, respectively[38-40]. The following parameters were set in the LASSO algorithm: family = "binomial"; alpha = 1; type measure = "deviance"; and nfolds = 10. The "randomForest" package in R was used to grow a forest of 500 trees using the default settings. The parameters of the SVM were set to its default value. We computed feature importance scores with the random forest model using the "importance" function in the "randomForest" package in R. The top 10 genes with the highest importance were selected for downstream analysis using the "randomForest" algorithms. The overlapping genes that emerged from the results of SVM, LASSO, and random forest analyses were identified as the key age-immune-related gene signatures.

Small Molecule Drugs Screening and Molecular Docking

A similarity scoring algorithm called eXtreme Sum (XSum) was performed to screen the candidate small molecule drugs based on the connectivity map (CMAP) database[41]. The DEGs between different immune infiltration subtypes were used as input files of "XSum" algorithm. Subsequently, a score was calculated for each small molecule drug of the CMAP database by "XSum" algorithm. A lower score suggests a higher potential to serve as a therapeutic agent to reverse immune infiltration.

The crystal structures of proteins encoded by the hub gene were obtained from the RCSB Protein Data Bank (PDB) database (San Diego, CA, USA). The specific website address used for accessing the PDB was www.rcsb.org/pdb/home/home.do[42]. Furthermore, the 3D structure of the small molecule drugs was downloaded from PubChem (https://www.ncbi.nlm.nih.gov/pccompound) database (NCBI, USA)[43]. The molecular docking process consisted of several steps, including the preparation of proteins and ligands, grid setting, and compound docking. These procedures were executed with the Autodock Vina software (The Scripps Research Institute, La Jolla, CA, USA)[44]. The best pose was chosen based on the docking score and the rationality of molecular conformation.

Statistical Analysis

R software (version 4.0.4, Boston, MA, USA) was utilized for all statistical procedures. Statistical analysis involved the utilization of the Wilcoxon/Kruskal-Wallis test to compare continuous variables, while differences in proportions were assessed using

the Chi-square test. Significance was determined by a *p*-value threshold lower than 0.05. Receiver operating characteristic (ROC) curve analysis was performed to evaluate the diagnostic performance of variables.

Results

Merging Gene Expression Data Sets

Boxplots were utilized to visualize the distribution of global array expression in GSE38713 and GSE87473 datasets before and after integration with the "ComBat" algorithm (Figure 1A, Figure 1B). The density plot of the gene expression distribution (Figure 1C, Figure 1D) suggested that the batch effects between GSE38713, and GSE87473 were all eliminated to some extent. UMAP plots showed identification (Figure 1E) and removal (Figure 1F) of inter-study batch effects through "ComBat" algorithm.

Association Between Age and Immune Cell Infiltration in UC

A cohort of 136 UC patients was stratified into two groups based on the median age value (40). UC patients with extensive disease extent exhibited a lower age compared to other patients ($p = 0.0024$; Figure 2A). The extensive disease extent group had a higher proportion of UC patients aged below 40 years (70% *vs.* 47%; Chi-square test, $p = 0.02$; Figure 2B).

The role of immune infiltration is of utmost importance in the pathogenesis of UC. Hence, our study primarily aimed to investigate the immune infiltration patterns, aiming to elucidate the factors contributing to the observed heterogeneity in disease extent across distinct age groups. The extent of immune cell infiltration of all the UC patients evaluated by ImmuCellAI database was presented in **Supplementary Table I.** We used unsupervised clustering to classify UC patients into diverse molecular subtypes based on the abundance of 24 immune cell types in the merged UC cohort *via* the R package "ConsensusClusterPlus". The optimal cluster number (K = 2) is determined by the area under the cumulative distribution function (CDF) curve, which corresponds to the largest number of clusters that induced the smallest incremental change in the area under the CDF curves (Figure 3A, Figure 3B, Figure 3C). Thus, "ConsensusClusterPlus" algorithm partitioned the UC patients into 2 major immune subtypes (cluster C1, cluster C2; Figure 3D). Significantly notable differences were observed in the distribution of immune subtypes across various age categories (Chi-square test, $p = 0.00219$; Figure 3E). UC patients aged below 40 demonstrated a higher proportion of cluster C1 and a lower proportion of cluster C2 compared to UC patients aged 40 and above.

Heatmap of the abundance of 24 immune cell types for the two immune subtypes was presented in Figure 4. UC samples within cluster C1 exhibited a distinct abundance of macrophages and a notably higher presence of neutrophils when compared to cluster C2. Conversely, cluster C1 exhibited a notably lower abundance of Tfh, NK, CD4+ T, and CD8+ T cells compared to cluster C2.

Age-Related Gene Module Revealed by WGCNA

The soft threshold for network construction was set to 20 (Figure 5A, Figure 5B). Subsequently, 8 gene modules with 9,539 genes in the merged UC cohort were identified by WGCNA (Figure 5C, Figure 5D). The MEs of modules were utilized to evaluate Pearson's correlation coefficients between the modules and age. Subsequently, the brown4 module was determined to exhibit the strongest association with age in UC, indicating its significant correlation with the aging process in this context (Pearson's correlation $r = 0.59$, $p < 0.0001$; Figure 6A). The brown4 module comprised a total of 4,291 genes (Figure 6B). In addition, we illustrated the correlation between module membership (MM) and gene significance (GS) for age in brown4 (Pearson's correlation $r = 0.79$, $p < 0.0001$; Figure 6C). Subsequently, we screened 1,279 distinct hub genes in the brown4 module based on the criteria of $|MM| > 0.8$ and $|GS| > 0.1$ **(Supplementary Table II)**.

Identification of DEGs Between Different Age Categories

Differential gene expression analysis was conducted to compare the transcriptomic profiles between two distinct age categories, namely, individuals younger than 40 and those aged 40 or older. Compared with patients with age ≥ 40, there were 505 DEGs (116 up-regulated and 389 down-regulated) identified in patients younger than 40 (Figure 7A; **Supplementary Table III**). We generated heatmaps to visualize the expression patterns of the top 10 up-regulated and down-regulated DEGs, respectively (Figure 7B).

Functional Enrichment Analysis

The intersection of 505 DEGs and 1,279 hub genes of WGCNA was taken, and then 146 genes in common were identified as age-related genes

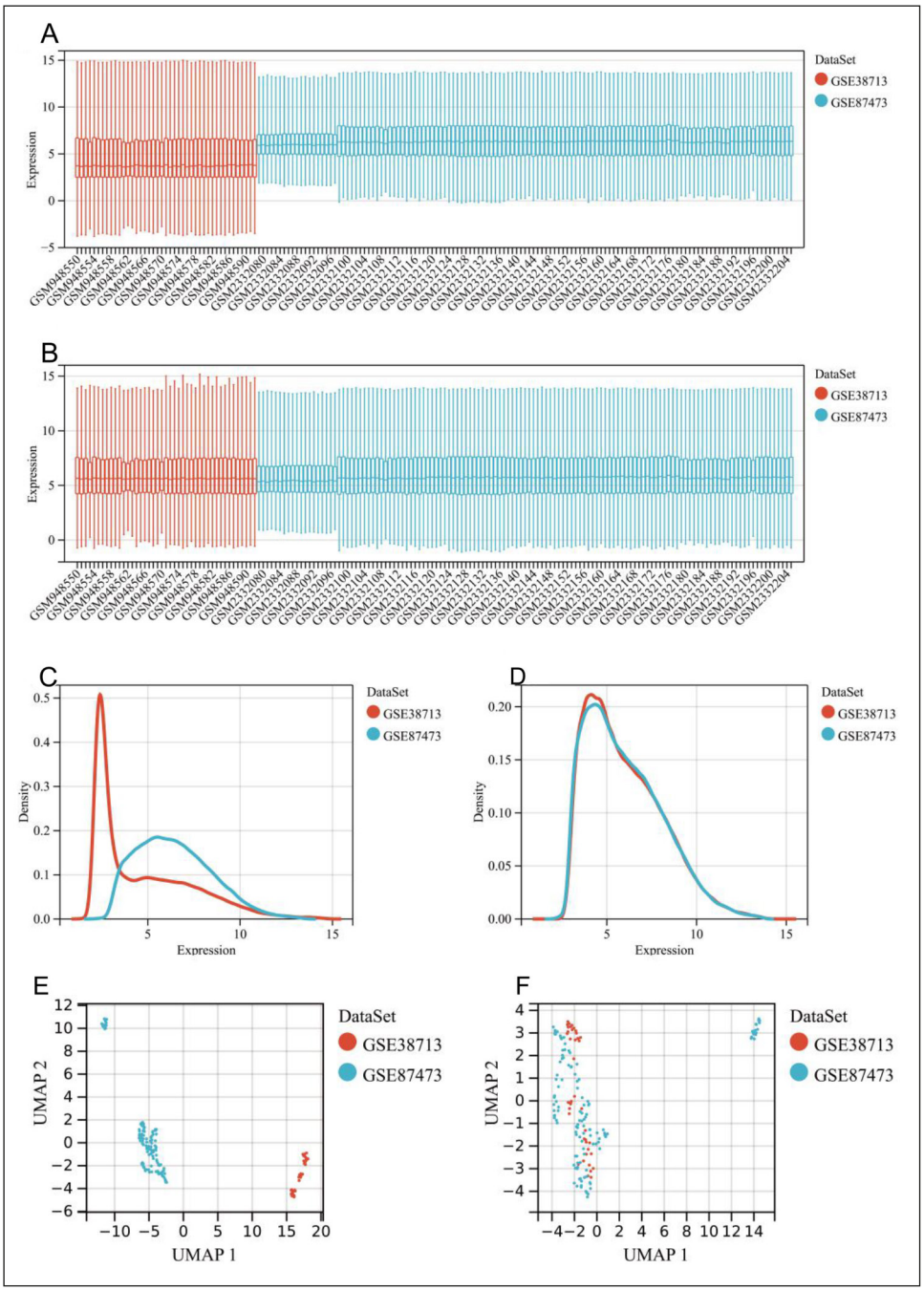

Figure 1. The global array expression distribution in GSE38713 and GSE87473 before (**A**) and after (**B**) merged using "ComBat" algorithm. The density plot of the gene expression distribution of GSE38713, and GSE87473 before (**C**) and after (**D**) merged. UMAP plots for GSE38713, and GSE87473 datasets before (**E**) and after (**F**) batch effect correction.

(Figure 7C). Enrichment analysis of age-related genes was performed to explore their biological functions based on the Metascape database (Figure 7D). The findings indicate that age-related genes exhibit significant enrichment in biological processes such as cellular senescence, regulation of intracellular transport, cell cycle checkpoints, and DNA damage/telomere stress-induced senescence.

Identification of Age-Immune-Related Key Genes Using Machine Learning

The "Limma" algorithm revealed that among the 146 age-related genes, 102 exhibited signifi-

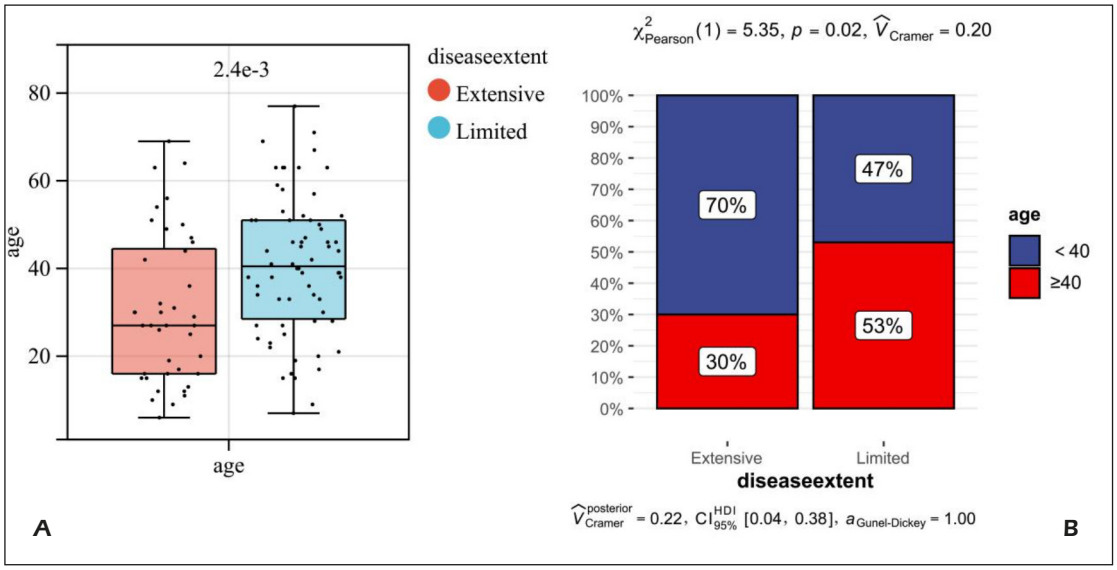

Figure 2. A, Boxplot of the differences of age between UC patients with extensive and limited disease extent. **B**, The proportion of patients with different ages in that with extensive and limited disease extent.

cant differential expression across different immune subtypes. Among these genes, 25 demonstrated elevated expression levels in cluster C1, whereas 77 exhibited higher expression in cluster C2 **(Supplementary Table IV)**. The gene expression matrix of the 102 DEGs was used as the input file of the SVM, LASSO, and random forest algorithms (Figure 8A, Figure 8B, Figure 8C; **Supplementary Table V**). The dependent variable in all algorithms was immune subtypes (cluster C1 *vs*. cluster C2). The key genes strongly associated with both age and immune infiltration (referred to as age-immune-related key genes) in UC were identified as the intersection of *MIDN* and *PLD6* from the three algorithms utilized in this study (Figure 8D). *MIDN* was upregulated in cluster C1, and *PLD6* was upregulated in cluster C2 (Figure 8E). The ROC curve showed satisfactory diagnostic efficacy of both *MIDN* and *PLD6* for cluster C1 (Figure 8F). Spearman's correlation analyses indicated positive correlations between *MIDN* and Neutrophil ($r = 0.50$, $p < 0.0001$; Figure 9A). A significant negative correlation between *PLD6* and Neutrophils was observed ($r = -0.52$, $p < 0.0001$).

Small Molecule Drugs Screening and Molecular docking

The DEGs between different immune subtypes (cluster C1 *vs*. cluster C2) were used as the input file of "XSum" algorithm. The XSum score calculated for each small molecule drug of the CMAP database is shown in **Supplementary Table VI**. Entinostat (MS.275) obtained the lowest score, indicating its potential as a small molecular compound for reversing the transition from cluster C1 to cluster C2. Molecular docking was then performed between the Entinostat and age-immune-related key genes (Figure 9B, Figure 9C). We found that Entinostat showed a good binding affinity for both *MIDN* and *PLD6* protein with the docking glide score of -8.9 kcal/mol and -6.8 kcal/mol, respectively.

Discussion

While the immunopathogenesis of ulcerative colitis (UC) has been comprehended to some extent, a comprehensive exploration of the distinctive immune landscape across various age categories in UC remains incomplete. This unexplored aspect could significantly contribute to the age-related phenotypic variations observed in the disease. This study employed an extensive range of bioinformatics analysis and machine learning techniques to elucidate the underlying mechanisms and identify key genes involved in the interplay between age and the immune landscape in UC.

Two distinct subtypes of UC characterized by diverse immune landscapes, namely cluster C1 and cluster C2, were identified using unsupervised hierarchical clustering. UC samples in cluster C1 were characterized by a higher degree of

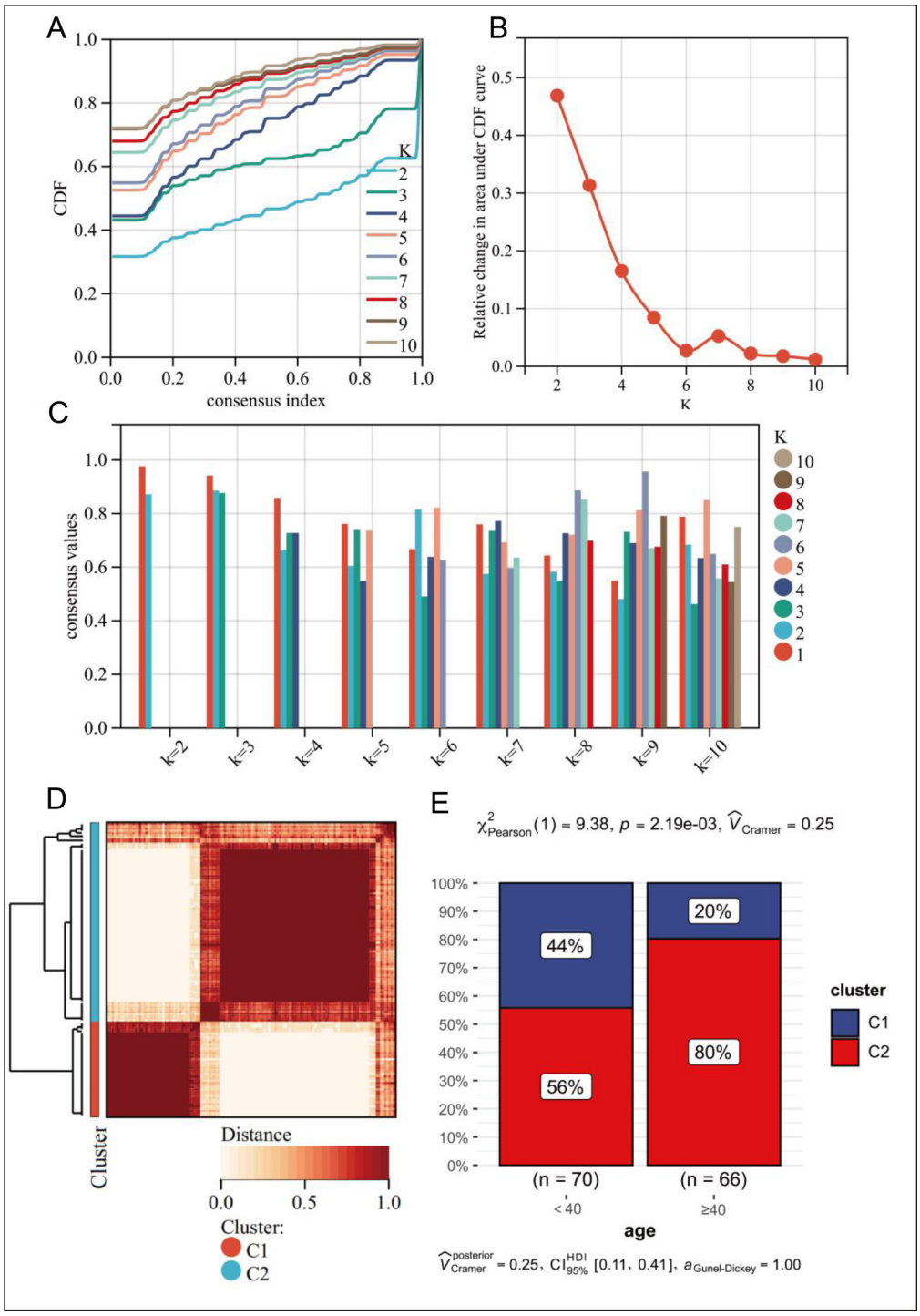

Figure 3. Identify different immune subtypes in UC. The optimal number of clusters was determined using the (**A**) empirical cumulative distribution function plot, (**B**) relative change in area under CDF curve and (**C**) consensus values. **D**, Consensus matrices of the merged dataset for k = 2. **E**, The proportion of immune subtypes in different age categories, *p*-values were from the chi-squared test (cluster C1, blue; cluster C2, red).

infiltration of macrophage and neutrophils compared to cluster C2. As reported previously, the mucosal immune dysregulation in UC is predominantly characterized by intensive infiltration with inflammatory cells, mainly neutrophils, macrophages, and dendritic cells. Uncontrolled inflam-

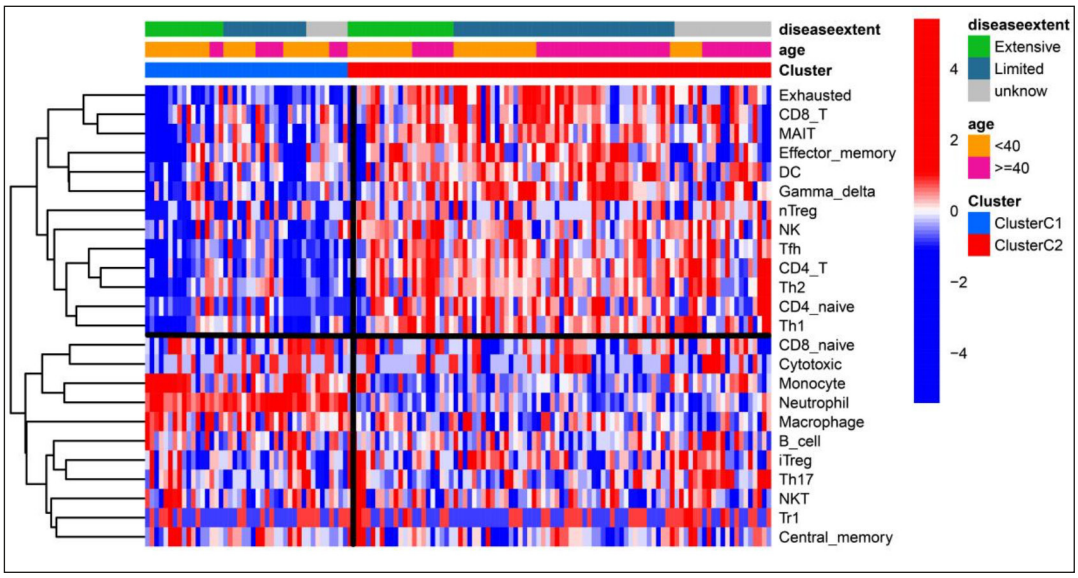

Figure 4. Heatmap of the levels of immune cell infiltrates.

mation will inevitably result in neutrophilic cryptitis, crypt abscesses, and mucosal ulceration[45]. When a damaged intestinal mucosa is stimulated by pathogenic organisms or inflammatory mediators, the activation and aggregation of neutrophils in the lesion region occurs to recognize stimuli and undergo phagocytosis[46-49]. In addition, the consequent neutrophil extracellular traps (NETs) also play a crucial role in the detection and clearance of potential pathogens. The study of Dinallo et al[50] found that NET-related proteins were up-regulated in inflamed regions of the colon of UC patients as compared to Crohn's disease patients and normal controls. Yan et al[51] study uncovered the suppressive effects of SM934 on macrophages, which subsequently contributed to the protective effects observed in a murine model of dextran sulfate sodium (DSS)-induced colitis. As the first-line defense in the lamina propria of mucosa, intestinal macrophages are derived from circulating monocytes, increasing in the active phase of UC[52]. The activated macrophages release a large amount of cytokine [IL-1, IL-6, tumor necrosis factor-alpha (TNF-α), etc.] and reactive metabolites of oxygen and nitrogen. The accumulated proinflammatory factors and chemokines stimulate the neutrophils' recruitment and are involved in the lymphocyte activation[53]. The long-term consequences of macrophage accumulation are exacerbation of intestinal epithelial tissue damage and dissemination of intestinal bacteria[54]. Therefore, it is suggested that targeting the population of macrophages may be a potential therapeutic modality for UC. Taken together, the greater percentage of cluster C1 UC patients in the group aged below 40 years is potentially an important contributor to the more extensive extent of lesions.

In our study, "XSum" algorithm suggested Entinostat as a potential small molecular compound that can reverse cluster C1 to cluster C2. Entinostat, an oral synthetic benzamide-derivative capable of inhibiting HDAC1 and HDAC3 enzymes, has been reported to mediate endocrine resistance of breast cancer through co-repressor proteins in clinical practice[55,56]. *In vitro* and *in vivo* experiments[57] have suggested that Entinostat possesses the ability to ameliorate inflammation, reduce apoptosis, and maintain intestinal barrier health and function. Furthermore, upon lipopolysaccharide (LPS) stimulation of bone marrow-derived macrophages, there was a notable increase observed in the expression levels of *HDAC1*[58]. Previous studies[59] demonstrated that *HDAC1* substantially reduces LPS-induced Cox-2 expression level in RAW264.7 macrophages. Therefore, Entinostat, an inhibitor of HDAC1, might be a potential drug candidate targeting the modulation of macrophage responses to treat UC.

Furthermore, we identified two age-related genes (*MIDN*, and *PLD6*) affecting the immune cell infiltration in UC based on machine learning. *MIDN* was found to modulate the expression of a diverse array of genes, such as α-synuclein, parkin, and *EIF4G1*, which play crucial roles in

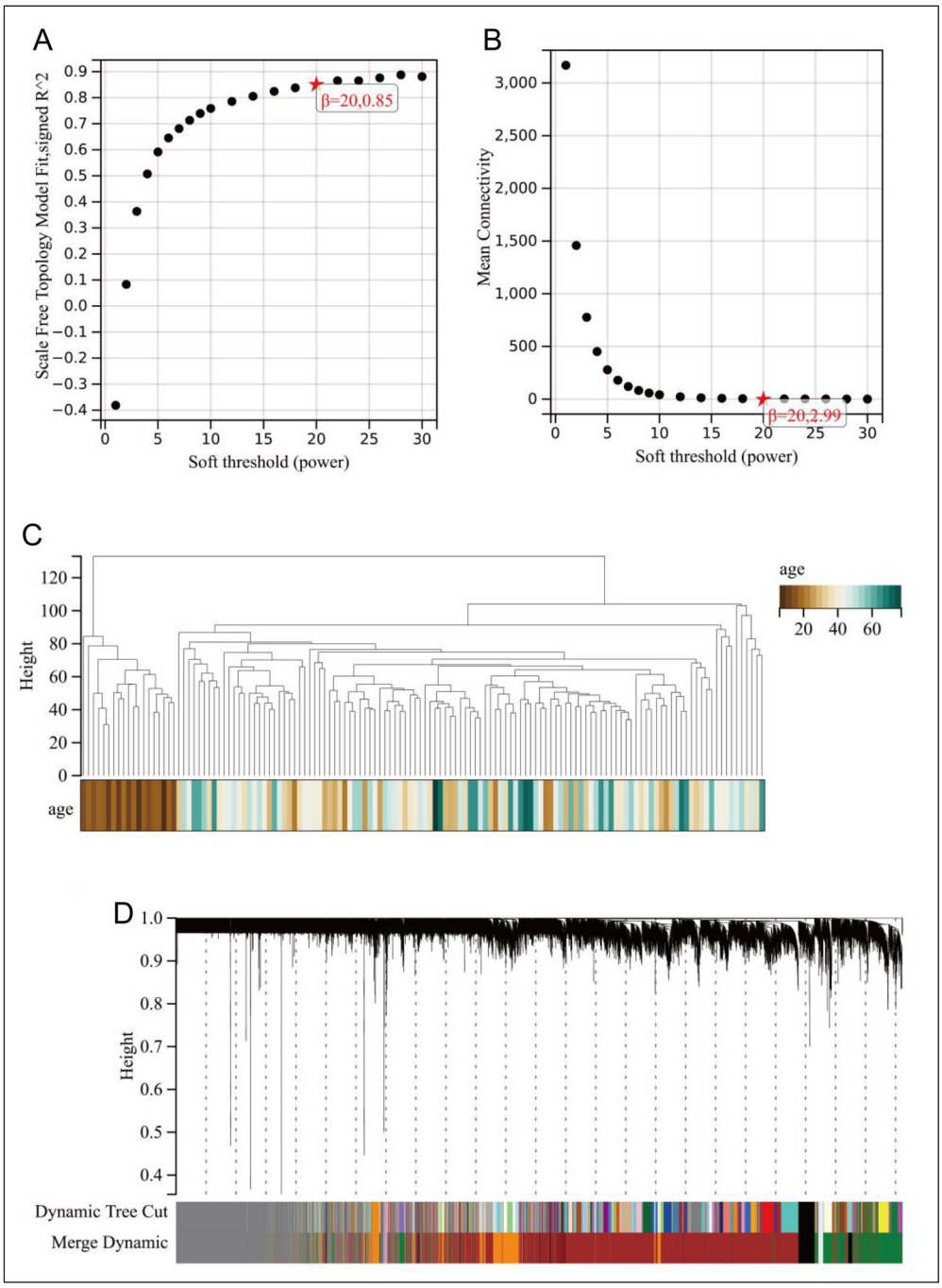

Figure 5. Determination of soft-threshold power in the WGCNA. **A**, Analysis of the scale-free index for various soft-threshold powers (β). **B**, Analysis of the mean connectivity for various soft-threshold powers. **C**, Clustering dendrogram of UC patients in the merged dataset. Identification of modules closely associated with age. **D**, Dendrogram of all differentially expressed genes clustered based on the measurement of dissimilarity (1-TOM). The color band shows the results obtained from the automatic single-block analysis.

the pathogenesis of neurodegenerative diseases associated with aging. This underscores the involvement of *MIDN* in the intricate molecular mechanisms underlying age-related neurodegeneration[60]. Previous studies[61,62] revealed that *PLD6* plays an important role in hydrolyzing cardiolipin in the outer mitochondrial membrane to generate phosphatidic acid. Phosphatidic acid is a crucial

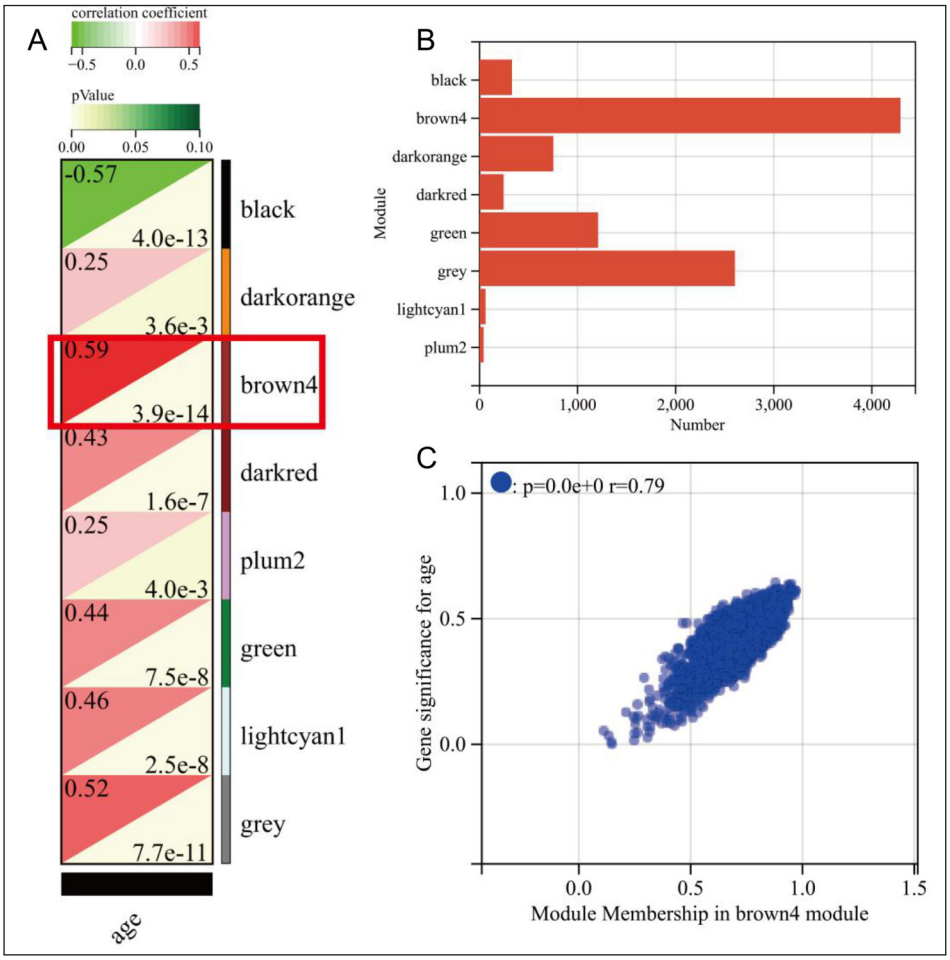

Figure 6. **A**, Heatmap of the correlation between the module eigengenes and age of patients with UC. **B**, Number of assigned genes to the different modules. **C**, A scatterplot of gene significance (GS) for age versus module membership (MM) in the brown4 module.

signaling molecule in controlling mitochondrial dynamics to promote mitochondrial fusion, which is tightly linked to cell growth, proliferation, and differentiation. However, the biological function of *MIDN* and *PLD6* in immune cell infiltration has not been reported in the literature before. In our study, we conducted correlation analyses and observed significant positive correlations between Neutrophil levels and *MIDN*, as well as negative correlations between Neutrophil levels and *PLD6*. These findings provide valuable insights into the interplay between Neutrophil activity and the expression of *MIDN* and *PLD6*, highlighting their potential roles in the context of our investigation. Although further experimental validation is required, *MIDN* and *PLD6* may represent a novel target for therapeutic intervention in an immune-mediated UC. This study offers novel insights and potential resources for the development of personalized clinical treatment strategies for individuals diagnosed with UC.

Limitations

It is important to acknowledge certain limitations that should be taken into account when interpreting the findings of the present study. First of all, this research only included a bioinformatics analysis, lacking further experimental verification as a solid foundation. Secondly, one of the limitations of our study is that this research is a retrospective study rather than a prospective trial. Therefore, future follow-up studies with prospective clinical trials and mechanistic exploration are required for corroboration of our findings.

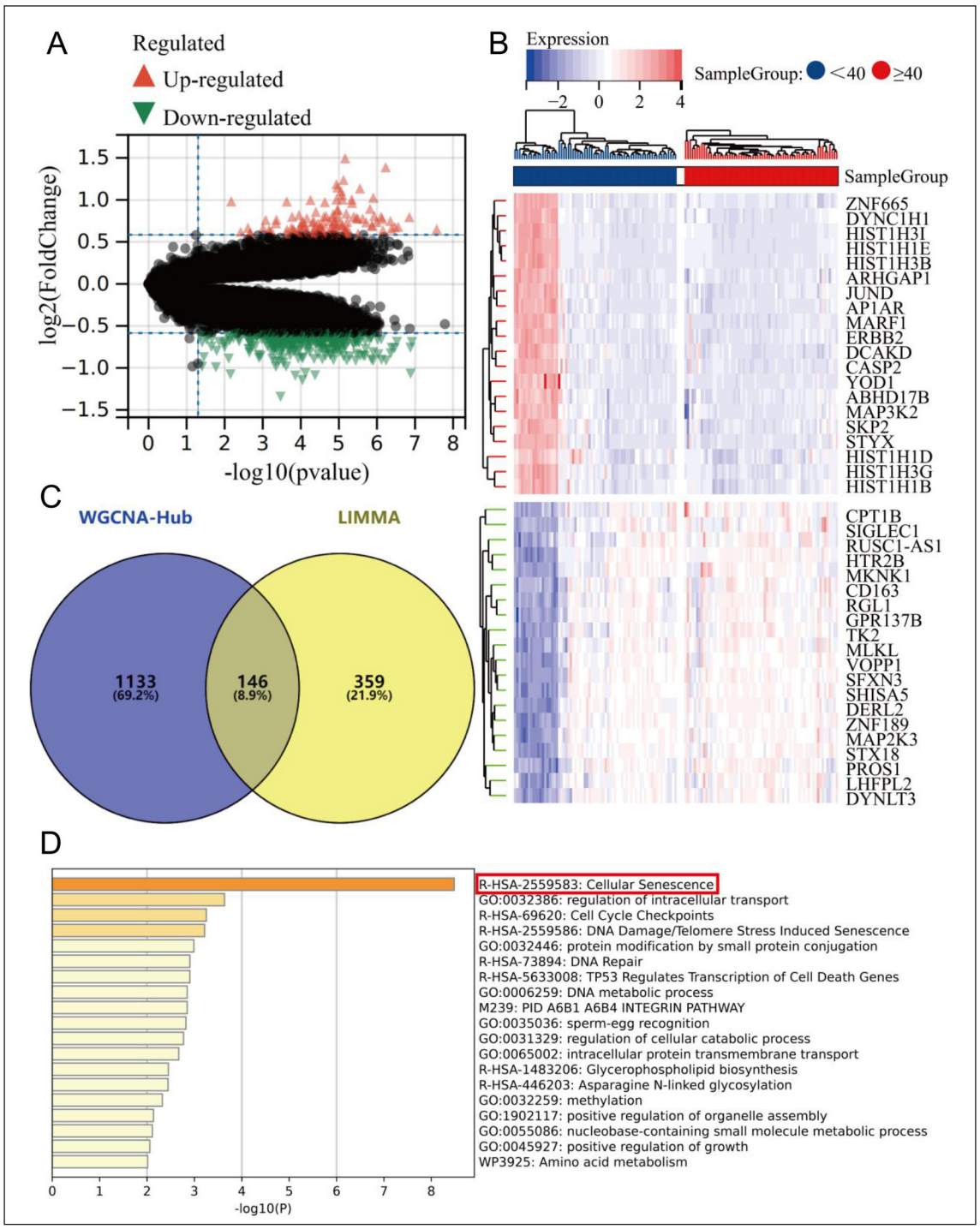

Figure 7. **A**, The differences generated by comparison were reflected in the volcanic map, the black genes were nonsignificant differences, and the red and green genes were the significant difference genes. **B**, Heatmap of the top 20 up-regulated and down-regulated DEGs (Red, up-regulation; blue, down-regulation). **C**, Venn diagram demonstrates overlapping genes of the DEGs of "Limma"algorithm and WGCNA hub genes. **D**, Enrichment analysis of 146 overlapping genes was performed to explore their biological functions based on Metascape database.

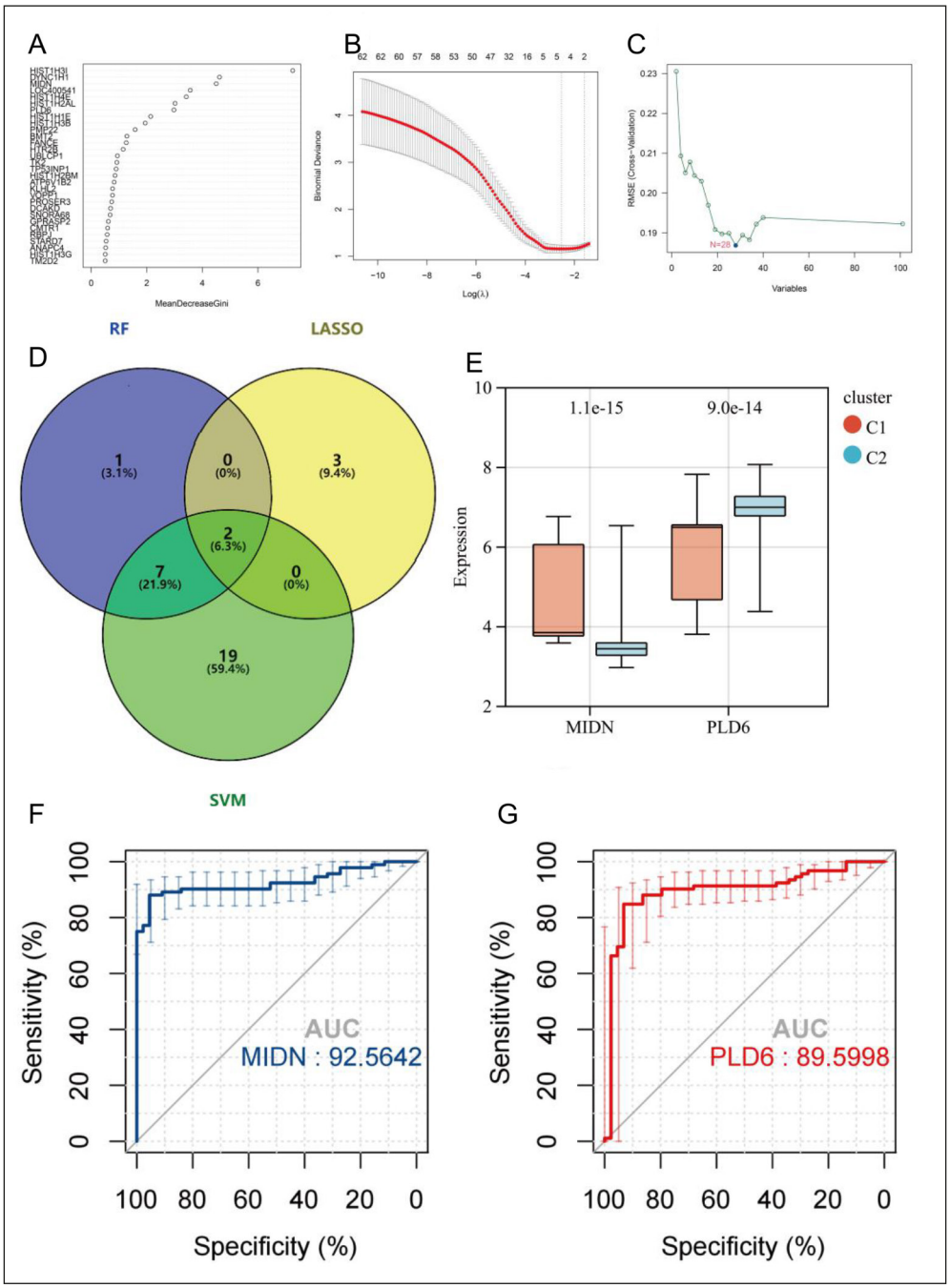

Figure 8. The identification of candidate age-immune-related key genes using random forest (**A**), LASSO (**B**), and SVM (**C**) algorithms. **D**, Venn diagram demonstrates overlapping genes of different machine learning algorithms. **E**, Boxplot shows significant differences in the expression of MIDN and PLD6 between the cluster C1 and C2 subtypes. ROC analysis of the diagnostic ability of MIDN (**F**) and PLD6 (**G**) for cluster C1.

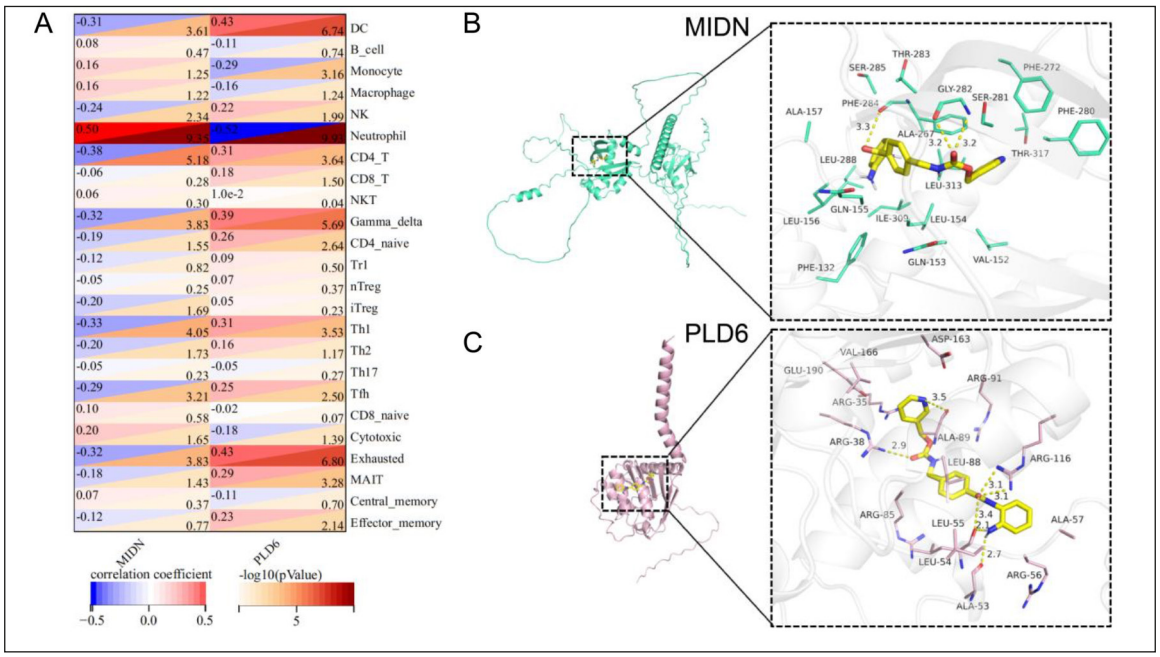

Figure 9. **A**, Heatmap of the correlation between the immunocyte fractions and MIDN and PLD6. **B**, Molecular Docking results and the best pose of Entinostat that fits MIDN (**B**) and PLD6 (**C**) protein.

Conclusions

The present study comprehensively investigated the crosstalk between age and immunity in UC based on bioinformatics analysis. By combining WGCNA and machine learning algorithms, we identified two age-related genes (*MIDN* and *PLD6*) affecting the immune cell infiltration in UC. Furthermore, we identified a small molecule drug (Entinostat) with potential therapeutic effects for UC. Thus, the present study contributed to the development of personalized clinical management and treatment regimens for UC.

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.

Ethics Approval

GEO belongs to public databases. The patients involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles. Our study is based on open-source data, so there are no ethical issues or other conflicts of interest.

Authors' Contributions

HC and YL: collection and analysis of data, and manuscript writing. HC and CT: analysis and interpretation of data. HC: project development and critical revision. HC YL and CT participated in the discussion and editing of the manuscript.

Funding

No funding was received for this study.

Acknowledgments

We thank Dr. Sonia Zebaze Dongmo (Department of Neurosurgery, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui Province, China) for carefully reading the manuscript and editing the language.

ORCID ID

H.-L. Chen: 0000-0002-2367-2868
Y.-H. Liu: 0009-0007-7200-9986
C.-H. Tan: 0009-0000-4259-6381

References

1) Yao D, Zhou Z, Wang P, Zheng L, Huang Y, Duan Y, Liu B, Li Y. MiR-125-5p/IL-6R axis regulates macrophage inflammatory response and intestinal

epithelial cell apoptosis in ulcerative colitis through JAK1/STAT3 and NF-κB pathway. Cell Cycle 2021; 20: 2547-2564.

2) Kobayashi T, Siegmund B, Le Berre C, Wei SC, Ferrante M, Shen B, Bernstein CN, Danese S, Peyrin-Biroulet L, Hibi T. Ulcerative colitis. Nat Rev Dis Primers 2020; 6: 74.
3) Ren PW, Yang WJ, Wang DD, Shan JY, Kang DY, Hong Q, Wen S, Zhang RW. Kangfuxinye Enema Combined with Mesalamine for Ulcerative Colitis: A Systematic Review and GRADE Approach. Evid Based Complement Alternat Med 2017; 2017: 6019530.
4) Ordás I, Eckmann L, Talamini M, Baumgart DC, Sandborn WJ. Ulcerative colitis. Lancet 2012; 380: 1606-1619.
5) Takiishi T, Fenero C, Câmara N. Intestinal barrier and gut microbiota: Shaping our immune responses throughout life. Tissue Barriers 2017; 5: e1373208.
6) Hanai H, Takeuchi K, Iida T, Kashiwagi N, Saniabadi AR, Matsushita I, Sato Y, Kasuga N, Nakamura T. Relationship between fecal calprotectin, intestinal inflammation, and peripheral blood neutrophils in patients with active ulcerative colitis. Dig Dis Sci 2004; 49: 1438-1443.
7) Hart AL, Al-Hassi HO, Rigby RJ, Bell SJ, Emmanuel AV, Knight SC, Kamm MA, Stagg AJ. Characteristics of intestinal dendritic cells in inflammatory bowel diseases. Gastroenterology 2005; 129: 50-65.
8) Buonocore S, Ahern PP, Uhlig HH, Ivanov II, Littman DR, Maloy KJ, Powrie F. Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. Nature 2010; 464: 1371-1375.
9) Fuss IJ, Heller F, Boirivant M, Leon F, Yoshida M, Fichtner-Feigl S, Yang Z, Exley M, Kitani A, Blumberg RS, Mannon P, Strober W. Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. J Clin Invest 2004; 113: 1490-1497.
10) Lissner D, Schumann M, Batra A, Kredel LI, Kühl AA, Erben U, May C, Schulzke JD, Siegmund B. Monocyte and M1 Macrophage-induced Barrier Defect Contributes to Chronic Intestinal Inflammation in IBD. Inflamm Bowel Dis 2015; 21: 1297-1305.
11) Heller F, Florian P, Bojarski C, Richter J, Christ M, Hillenbrand B, Mankertz J, Gitter AH, Bürgel N, Fromm M, Zeitz M, Fuss I, Strober W, Schulzke JD. Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. Gastroenterology 2005; 129: 550-564.
12) Lechner K, Mott S, Al-Saifi R, Knipfer L, Wirtz S, Atreya R, Vieth M, Rath T, Fraass T, Winter Z, August A, Luban J, Zimmermann VS, Weigmann B, Neurath MF. Targeting of the Tec Kinase ITK Drives Resolution of T Cell-Mediated Colitis and Emerges as Potential Therapeutic Option in Ulcerative Colitis. Gastroenterology 2021; 161: 1270-1287.e19.
13) Heiseke AF, Faul AC, Lehr HA, Förster I, Schmid RM, Krug AB, Reindl W. CCL17 promotes intestinal inflammation in mice and counteracts regulatory T cell-mediated protection from colitis. Gastroenterology 2012; 142: 335-345.
14) Torres J, Danese S, Colombel JF. New therapeutic avenues in ulcerative colitis: thinking out of the box. Gut 2013; 62: 1642-1652.
15) Ruel J, Ruane D, Mehandru S, Gower-Rousseau C, Colombel JF. IBD across the age spectrum: is it the same disease. Nat Rev Gastroenterol Hepatol 2014; 11: 88-98.
16) Gower-Rousseau C, Dauchet L, Vernier-Massouille G, Tilloy E, Brazier F, Merle V, Dupas JL, Savoye G, Baldé M, Marti R, Lerebours E, Cortot A, Salomez JL, Turck D, Colombel JF. The natural history of pediatric ulcerative colitis: a population-based cohort study. Am J Gastroenterol 2009; 104: 2080-2088.
17) Langholz E, Munkholm P, Krasilnikoff PA, Binder V. Inflammatory bowel diseases with onset in childhood. Clinical features, morbidity, and mortality in a regional cohort. Scand J Gastroenterol 1997; 32: 139-147.
18) Hyams JS, Davis P, Grancher K, Lerer T, Justinich CJ, Markowitz J. Clinical outcome of ulcerative colitis in children. J Pediatr 1996; 129: 81-88.
19) Malaty HM, Abraham BP, Mehta S, Garnett EA, Ferry GD. The natural history of ulcerative colitis in a pediatric population: a follow-up population-based cohort study. Clin Exp Gastroenterol 2013; 6: 77-83.
20) Fujihashi K, Kiyono H. Mucosal immunosenescence: new developments and vaccines to control infectious diseases. Trends Immunol 2009; 30: 334-343.
21) Yung RL, Julius A. Epigenetics, aging, and autoimmunity. Autoimmunity 2008; 41: 329-335.
22) Pawelec G. Immunosenescence: impact in the young as well as the old. Mech Ageing Dev 1999; 108: 1-7.
23) Koga T, McGhee JR, Kato H, Kato R, Kiyono H, Fujihashi K. Evidence for early aging in the mucosal immune system. J Immunol 2000; 165: 5352-5359.
24) McDonald KG, Leach MR, Huang C, Wang C, Newberry RD. Aging impacts isolated lymphoid follicle development and function. Immun Ageing 2011; 8: 1.
25) Kawanishi H, Kiely J. Immune-related alterations in aged gut-associated lymphoid tissues in mice. Dig Dis Sci 1989; 34: 175-184.
26) Innes CL, Hesse JE, Morales AJ, Helmink BA, Schurman SH, Sleckman BP, Paules RS. DNA damage responses in murine Pre-B cells with genetic deficiencies in damage response genes. Cell Cycle 2020; 19: 67-83.
27) Taminau J, Meganck S, Lazar C, Steenhoff D, Coletta A, Molter C, Duque R, de Schaetzen V, Weiss Solís DY, Bersini H, Nowé A. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. BMC Bioinformatics 2012; 13: 335.

28) Wang Y, Sun J, Yang Y, Zebaze Dongmo S, Qian Y, Wang Z. Identification and Development of Subtypes with Poor Prognosis in Gastric Cancer Based on Both Hypoxia and Immune Cell Infiltration. Int J Gen Med 2021; 14: 9379-9399.

29) Miao YR, Zhang Q, Lei Q, Luo M, Xie GY, Wang H, Guo AY. ImmuCellAI: A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. Adv Sci (Weinh) 2020; 7: 1902880.

30) Miao YR, Xia M, Luo M, Luo T, Yang M, Guo AY. ImmuCellAI-mouse: a tool for comprehensive prediction of mouse immune cell abundance and immune microenvironment depiction. Bioinformatics 2022; 38: 785-791.

31) Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010; 26: 1572-1573.

32) Wang Y, Wang Z, Sun J, Qian Y. Identification of HCC Subtypes With Different Prognosis and Metabolic Patterns Based on Mitophagy. Front Cell Dev Biol 2021; 9: 799507.

33) Chen H, Zhang J, Sun X, Wang Y, Qian Y. Mitophagy-mediated molecular subtypes depict the hallmarks of the tumour metabolism and guide precision chemotherapy in pancreatic adenocarcinoma. Front Cell Dev Biol 2022; 10: 901207.

34) Feng H, Zhang X, Lai W, Wang J. Long non-coding RNA SLC16A1-AS1: its multiple tumorigenesis features and regulatory role in cell cycle in oral squamous cell carcinoma. Cell Cycle 2020; 19: 1641-1653.

35) Wang F, Wang B, Long J, Wang F, Wu P. Identification of candidate target genes for endometrial cancer, such as ANO1, using weighted gene co-expression network analysis. Exp Ther Med 2019; 17: 298-306.

36) Cheng Q, Wang L. LncRNA XIST serves as a ceRNA to regulate the expression of ASF1A, BRWD1M, and PFKFB2 in kidney transplant acute kidney injury via sponging hsa-miR-212-3p and hsa-miR-122-5p. Cell Cycle 2020; 19: 290-299.

37) Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019; 10: 1523.

38) Xu N, Guo H, Li X, Zhao Q, Li J. A Five-Genes Based Diagnostic Signature for Sepsis-Induced ARDS. Pathol Oncol Res 2021; 27: 580801.

39) Wang S, Su W, Zhong C, Yang T, Chen W, Chen G, Liu Z, Wu K, Zhong W, Li B, Mao X, Lu J. An Eight-CircRNA Assessment Model for Predicting Biochemical Recurrence in Prostate Cancer. Front Cell Dev Biol 2020; 8: 599494.

40) Alderden J, Pepper GA, Wilson A, Whitney JD, Richardson S, Butcher R, Jo Y, Cummins MR. Predicting Pressure Injury in Critical Care Patients: A Machine-Learning Model. Am J Crit Care 2018; 27: 461-468.

41) Yang C, Zhang H, Chen M, Wang S, Qian R, Zhang L, Huang X, Wang J, Liu Z, Qin W, Wang C, Hang H, Wang H. A survey of optimal strategy for signature-based drug repositioning and an application to liver cancer. Elife 2022; 11: e71880.

42) Goodsell DS, Burley SK. RCSB Protein Data Bank tools for 3D structure-guided cancer research: human papillomavirus (HPV) case study. Oncogene 2020; 39: 6623-6632.

43) Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res 2021; 49: D1388-D1395.

44) Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model 2021; 61: 3891-3898.

45) Lu H, Lin J, Xu C, Sun M, Zuo K, Zhang X, Li M, Huang H, Li Z, Wu W, Feng B, Liu Z. Cyclosporine modulates neutrophil functions via the SIRT6-HIF-1α-glycolysis axis to alleviate severe ulcerative colitis. Clin Transl Med 2021; 11: e334.

46) Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. Nat Rev Immunol 2013; 13: 159-175.

47) Nicolás-Ávila JÁ, Adrover JM, Hidalgo A. Neutrophils in Homeostasis, Immunity, and Cancer. Immunity 2017; 46: 15-28.

48) Wang W, Zhang J, Zheng N, Li L, Wang X, Zeng Y. The role of neutrophil extracellular traps in cancer metastasis. Clin Transl Med 2020; 10: e126.

49) Zhou G, Yu L, Fang L, Yang W, Yu T, Miao Y, Chen M, Wu K, Chen F, Cong Y, Liu Z. CD177(+) neutrophils as functionally activated neutrophils negatively regulate IBD. Gut 2018; 67: 1052-1063.

50) Dinallo V, Marafini I, Di Fusco D, Laudisi F, Franzè E, Di Grazia A, Figliuzzi MM, Caprioli F, Stolfi C, Monteleone I, Monteleone G. Neutrophil Extracellular Traps Sustain Inflammatory Signals in Ulcerative Colitis. J Crohns Colitis 2019; 13: 772-784.

51) Yan YX, Shao MJ, Qi Q, Xu YS, Yang XQ, Zhu FH, He SJ, He PL, Feng CL, Wu YW, Li H, Tang W, Zuo JP. Artemisinin analogue SM934 ameliorates DSS-induced mouse ulcerative colitis via suppressing neutrophils and macrophages. Acta Pharmacol Sin 2018; 39: 1633-1644.

52) Grimm MC, Pullman WE, Bennett GM, Sullivan PJ, Pavli P, Doe WF. Direct evidence of monocyte recruitment to inflammatory bowel disease mucosa. J Gastroenterol Hepatol 1995; 10: 387-395.

53) Neurath MF. Cytokines in inflammatory bowel disease. Nat Rev Immunol 2014; 14: 329-342.

54) Mahida YR. The key role of macrophages in the immunopathogenesis of inflammatory bowel disease. Inflamm Bowel Dis 2000; 6: 21-33.

55) Wang B, Lyu H, Pei S, Song D, Ni J, Liu B. Cladribine in combination with entinostat synergistically elicits anti-proliferative/anti-survival effects on multiple myeloma cells. Cell Cycle 2018; 17: 985-996.

56) Connolly RM, Zhao F, Miller KD, Lee MJ, Piekarz RL, Smith KL, Brown-Glaberman UA, Winn JS, Faller BA, Onitilo AA, Burkard ME, Budd GT, Levine EG, Royce ME, Kaufman PA, Thomas A, Trepel JB, Wolff AC, Sparano JA. E2112: Randomized Phase III Trial of Endocrine Therapy Plus Entinostat or Placebo in Hormone Receptor-Positive Advanced Breast Cancer. A Trial of the ECOG-ACRIN Cancer Research Group. J Clin Oncol 2021; 39: 3171-3181.

57) Li C, Chen Y, Zhu H, Zhang X, Han L, Zhao Z, Wang J, Ning L, Zhou W, Lu C, Xu L, Sang J, Feng Z, Zhang Y, Lou X, Bo X, Zhu B, Yu C, Zheng M, Li Y, Sun J, Shen Z. Inhibition of Histone Deacetylation by MS-275 Alleviates Colitis by Activating the Vitamin D Receptor. J Crohns Colitis 2020; 14: 1103-1118.

58) Aung HT, Schroder K, Himes SR, Brion K, van Zuylen W, Trieu A, Suzuki H, Hayashizaki Y, Hume DA, Sweet MJ, Ravasi T. LPS regulates proinflammatory gene expression in macrophages by altering histone deacetylase expression. FASEB J 2006; 20: 1315-1327.

59) Deng WG, Zhu Y, Wu KK. Role of p300 and PCAF in regulating cyclooxygenase-2 promoter activation by inflammatory mediators. Blood 2004; 103: 2135-2142.

60) Obara Y, Ishii K. Transcriptome Analysis Reveals That Midnolin Regulates mRNA Expression Levels of Multiple Parkinson's Disease Causative Genes. Biol Pharm Bull 2018; 41: 20-23.

61) Riew TR, Kim S, Jin X, Kim HL, Hwang WC, Kang M, Yang ES, Lee MY, Min DS. Cellular and subcellular localization of endogenous phospholipase D6 in seminiferous tubules of mouse testes. Cell Tissue Res 2021; 385: 191-205.

62) Sun Y, Chen J. mTOR signaling: PLD takes center stage. Cell Cycle 2008; 7: 3118-3113.