RESEARCH ARTICLE

Analysis of disulfidoptosis-related genes in rheumatoid arthritis using bioinformatics and Mendelian randomization

Zhichen Zou[†], Feng Li[†], Xiajun Liang, Wei Li[*]

Department of Orthopedics, Tangdu Hospital, Fourth Military Medical University, Xi'an, Shaanxi, China.

Received: June 17, 2026; **accepted:** October 7, 2026.

Rheumatoid arthritis (RA) is a chronic autoimmune disease with elusive pathogenesis. The role of disulfidoptosis, a new programmed cell death form, in this disease remains unexplored. This research explored the relationship between disulfidoptosis-related genes (DRGs) and RA. Using the GSE93272 dataset, DRGs with diagnostic value were identified through optimal machine learning selection, Extreme Gradient Boosting (XGB). A predictive nomogram was constructed, which was subsequently rigorously validated using an independent external dataset (GSE45291) and calibration curves. RA samples were categorized into distinct subtypes using unsupervised consensus clustering. Immune infiltration analysis was performed to link DRGs with microenvironment heterogeneity. Crucially, causality was established using Mendelian randomization (MR) integrating expression quantitative trait loci (eQTL), genotype-tissue expression (GTEx), and FinnGen genome-wide association studies (GWAS) data. The XGB-derived nomogram presented exceptional discriminatory power, providing the highest area under the curve (AUC) and lowest root mean square error (RMSE). Two disulfidoptosis-related molecular clusters with significant disparities in immune cell infiltration profiles were identified. MR analysis proved filamin B (FLNB) as a causal risk factor for RA with inverse variance weighted [IVW]:odds ratio > 1 ($P < 0.05$) and robust results showing no pleiotropy or heterogeneity. This research delineated disulfidoptosis subtypes in RA, providing novel genetic insights and potential therapeutic targets for precision medicine.

Keywords: rheumatoid arthritis; disulfidoptosis; immune cell infiltration; machine learning model; Mendelian randomization.

[*]**Corresponding author:** Wei Li, Department of Orthopedics, Tangdu Hospital, Fourth Military Medical University, Xi'an, Shaanxi, China. Email: allenliweizq@163.com.

[†]These authors contributed equally to this work.

Introduction

Rheumatoid arthritis (RA) is a common and refractory autoimmune disease categorized in rheumatology and orthopedics, severely endangering human joint health and quality of life. Epidemiological data have demonstrated that global RA incidence ranges from 0.5% to 1%, affecting about five million individuals worldwide with an increasing prevalence in recent years [1]. Clinically, RA is predominantly characterized by persistent articular swelling, irreversible joint deformity, synovial inflammation, and even long-term physical incapacitation [2]. These debilitating symptoms not only cause severe functional impairment and physical pain in patients but also place heavy mental and economic burdens on individual families and constitute a substantial public health burden for society [3]. Considering its high disabling

potential and morbidity, RA has always been a hot research topic in the fields of clinical rheumatology and autoimmune disease.

Over the past decades, extensive basic and clinical research has been conducted to investigate the diagnosis, pathogenesis, and treatment of RA. Previous studies have preliminarily identified the main pathological characteristics of RA including immune cell infiltration, synovial hyperplasia, and inflammatory factor overexpression, giving rise to the establishment of a series of conventional diagnostic indicators and clinical intervention strategies [4]. Meanwhile, with the rapid development of bioinformatics and molecular biology, an increasing number of disease-related genes and regulatory pathways have been identified, providing novel theoretical insights for the mechanistic exploration and targeted treatment of RA. Furthermore, multiple regulated cell death forms have been confirmed to participate in the occurrence and progression of autoimmune inflammatory diseases, providing new research perspectives for the investigation of RA heterogeneous pathogenesis [5]. Despite continuous progress in research on RA, there still remain critical research deficiencies and unresolved problems limiting the accurate diagnosis and effective treatment of RA. The precise molecular pathophysiological mechanism of RA is not fully understood yet and there is still no radical cure for the disease, giving rise to long-term recurrent attacks in most patients. The existing diagnostic biomarkers and therapeutic targets cannot fully explain RA clinical heterogeneity, resulting in inconsistent treatment efficacy among different patients. Most importantly, as a newly identified regulated cellular death form triggered by intracellular cystine accumulation, disulfide stress, and abnormal actin cytoskeletal protein bonding [6], disulfidoptosis has been proven to play vital regulatory roles in a variety of immune and inflammatory diseases. However, the regulatory mechanisms, specific expression patterns, and immune-related functions of disulfidoptosis and disulfidoptosis-related genes (DRGs) in RA remain unclear and poorly systematically studied. The correlation between RA heterogeneity and DRGs has not been explored, greatly limiting the comprehensive understanding of RA pathogenesis and development of new targeted therapeutic strategies.

This research primarily focused on evaluating the potential functional roles and clinical value of DRGs in RA to systematically clarify the expression characteristics of DRGs in RA patients, determine the correlation between RA immune heterogeneity and DRG expression, and identify reliable disulfidoptosis-related susceptibility genes and effective diagnostic biomarkers for RA. The research integrated machine learning modeling, comprehensive bioinformatics analysis, and Mendelian randomization (MR) analysis to screen differentially expressed disulfidoptosis-related genes (DEGs), develop a machine learning-based predictive model, and evaluate model accuracy and reliability by using multiple validation methods. In addition, unsupervised clustering analysis was performed to identify disulfidoptosis-related RA molecular subtypes and analyze differences in immune cell infiltration, signaling pathways, and biological functions among different subtypes. MR analysis was performed to screen credible DRGs with causal susceptibility to RA. This research provided a new molecular basis to elucidate the pathogenesis of RA and improve its diagnostic accuracy, which not only systematically revealed internal correlations among DRG expression, disease heterogeneity, and immune microenvironment variations in RA, but also identified potential novel diagnostic biomarkers and therapeutic targets for RA.

Materials and methods

Data acquisition and pre-processing

Two microarray datasets were obtained from the expansive Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). GSE93272 (https://www.ncbi.nlm.nih.gov/geo/

query/acc.cgi?acc=GSE93272), as the primary dataset, encompassed a cohort of 43 healthy control subjects and 232 RA peripheral blood samples, all annotated by GPL13158 platform (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL13158). GSE45291 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45291) was incorporated to confirm the authenticity of findings, which was consisted of peripheral blood samples collected from 493 RA patients and 20 healthy control individuals, all annotated by GPL13158 platform, thus enabling an extensive validation analysis. To ensure maximum reliability and precision, both datasets underwent meticulous pre-processing and normalization using the advanced "normalizeBetweenArrays" function from the "limma" package (v3.54.2) (https://bioconductor.org/packages/3.16/bioc/html/limma.html) in R (https://cran.r-project.org/). These crucial steps played pivotal roles in the identification of pertinent gene expression patterns intricately linked to RA. Additional 15 distinctively expressed DRGs were acquired including SLC7A11, INF2, CD2AP, ACTN4, PDLIM1, MYH10, MYH9, IQGAP1, FLNA, FLNB, TLN1, ACTB, MYL6, DSTN, and CAPZB [7]. The differential gene expression analyses were conducted to compare the transcriptional profiles of these DRGs between individuals with RA and healthy control subjects. The results were visualized by utilizing "pheatmap" (https://cran.r-project.org/web/packages/pheatmap/index.html) and "ggplot2" R packages (https://cran.r-project.org/web/packages/ggplot2/index.html). The Expression Quantitative Trait Loci (eQTL) dataset (https://gtexportal.org/) was obtained from whole blood samples by GTEx Consortium (https://gtexportal.org/). Data pertaining to RA were obtained from the FennGenn consortium's GWAS summary including 13,621 cases and 262,844 participants (https://www.finngen.fi/en) [8]. To improve the reliability and precision of causal deductions, genetic marker selection adhered to the tripartite criteria essential for MR. Initially, single nucleotide polymorphisms (SNPs) were selected based on their adherence to a genome-wide significance threshold, coupled with F-statistics > 10 [9]. Then, SNPs devoid of linkage disequilibrium as indicated by $r^2 < 0.1$ and a clumping distance of 100 kb were retained. Finally, SNPs linked with either the outcome or potential confounders ($P < 5\times10^{-6}$) were excluded, and palindromic SNPs exhibiting intermediate allele frequencies were discarded during data harmonization between exposure and outcome datasets [10].

Building accurate predictive models using multiple machine learning methods

Multiple machine learning algorithms were applied *via* the "caret" package (v6.0.91) in R, which included random forest (RF) (https://cran.r-project.org/web/packages/randomForest/index.html), support vector machine (SVM) (https://cran.r-project.org/web/packages/e1071/index.html), generalized linear model (GLM) (https://stat.ethz.ch/R-manual/R-patched/library/stats/html/glm.html), and extreme gradient boosting (XGB) (https://cran.r-project.org/web/packages/xgboost/index.html). RF constructed an ensemble of decision trees to perform regression or classification tasks [11], SVM generated a hyperplane in feature space to differentiate negative and positive samples [12], GLM established associations among dependent variables and independent covariates [13], and XGB efficiently managed the trade-off between model complexity and classification error [14]. To realize maximum precision, four distinct models predicated upon differentially expressed genes implicated in disulfidoptosis were meticulously computed. The exceptional ca·pabilities of "caret" package in R were employed for automatic parameter tuning, using grid search method to fine-tune proposed models. In addition, a comprehensive 5-fold cross-validation approach was applied to effectively explore the robustness and efficacy of these models. To interpret the results of these machine learning models, the prodigious potential of "DALEX" R package (v2.4.0) (https://cran.r-project.org/web/packages/DALEX/index.html) was employed, which enabled to visually depict residual distributions and ascertain the relative

importance of various features. Furthermore, "pROC" package (v1.18.0) (https://cran.r-project.org/web/packages/pROC/index.html) was applied in R to draw enlightening plots illustrating the area under the ROC curves. Remarkably, the quest for the ultimate machine learning model resulted in the identification of the top five predictive genes inextricably linked to RA, thereby enhancing the understanding of this intricate disorder.

Construction and validation of a nomogram model

The versatile "rms" package (v6.2.0) (https://cran.r-project.org/web/packages/pROC/index.html) in R was applied with the main goal of precisely evaluating the presence of RA. To ensure a comprehensive assessment of the effectiveness of the model, two robust and well-established performance metrics including DCA and calibration curve were adopted to explore model performance. Calibration curve served in ensuring model accuracy and reliability, using a meticulous evaluation of predicted probabilities versus observed outcomes. Additionally, DCA emerged as a crucial evaluation metric to appraise the clinical application of the model. A comparative analysis was performed to juxtapose the net benefit of proposed model against alternative clinical strategies or predictive models. The integration of the calibration curve, nomogram model, and DCA proved to be an indomitable approach, which allowed personalized predictions pertaining to the incidence of RA clusters.

Evaluating immune cell infiltration

A total of 22 distinct immune cell types were evaluated in this research. The formidable CIBERSORT algorithm (https://cibersort.stanford.edu/) was developed, and LM22 signature matrix was used, which applied the power of gene expression data to accurately estimate the proportions of various immune cell types present within any certain sample [15]. By using Monte Carlo sampling, an inverse fold product (IFP) *P* value was obtained for each sample. By applying this cutting-edge algorithm, a more nuanced comprehension of immune cell landscape in RA was unraveled, unearthing promising targets for potential therapeutic interventions. Rigorous measures were taken to guarantee the coherence of the data by ensuring that the cumulative sum of the proportions of the 22 distinct immune cell types within each sample was equal to unity. This meticulous approach discerned the relative proportions of each type of immune cell and made meaningful comparisons among various samples. Application of CIBERSORT algorithm in conjunction with LM22 signature matrix provided a profound understanding of the intricate cellular components within the immune system implicated in RA context.

Correlation analysis between drgs and infiltrated immune cells

To fortify the body of evidence regarding the conceivable correlation between DRGs and RA-related immunological attributes, additional analyses were conducted to investigate the correlation coefficients of DRG expression levels and relative proportions of immune cells. Using Spearman's rank correlation coefficient, the potential relationship between these variables was meticulously explored. Versatile "corrplot" R package (v0.92) (https://cran.r-project.org/web/packages/corrplot/index.html) was employed to enhance the clarity of the obtained correlation results, which generated visually appealing and customizable plots, making it possible to provide a detailed representation of the correlation matrix. The resulting plot effectively portrayed the strength and direction of the obtained correlation coefficients, aided by a color scheme to facilitate the identification of negative and positive correlations.

Unsupervised clustering of RA patients and evaluation of immune cell infiltration

Unsupervised clustering analyses were performed on the expression profiles of DRGs using "ConsensusClusterPlus" R package (v2.60) (https://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html). By using k-means algorithm with 1,000 iterations, 232 RA

samples were successfully classified into distinct subtypes. The optimal number of subtypes was determined by performing a comprehensive assessment using cumulative distribution function (CDF) matrix and consensus curve. A consistent cluster score exceeding 0.9 was sought. Subsequently, a parallel analysis comparing differences in immune cell infiltration among various clusters was performed.

Gene set variation analysis (GSVA)

To investigate enriched gene sets among different DRGs clusters, an enrichment analysis was conducted using "GSVA" R package (v2.11) (https://bioconductor.org/packages/release/bioc/html/GSVA.html). "c2.cp.kegg.v7.4.symbols" and "c5.go.bp.v7.5.1.symbols" files from the MSigDB database (https://www.gsea-msigdb.org/gsea/msigdb/) were utilized. Differential pathways and biological functions were determined using limma R package (v3.52.1), which involved comparing the GSVA scores across different DRGs clusters. Scores showing significant changes corresponded to GSVA scores with t-values exceeding 2. By using GSVA enrichment analysis, valuable insights were gained into the gene pathways and functions that contributed to DRGs development and progression.

Mendelian randomization analysis

"TwoSampleMR" R package (https://mrcieu.github.io/TwoSampleMR/) facilitated the two-sample MR analysis, contrasting exposures with outcomes. Predominantly, analyses used inverse-variance weighted (IVW) methods under a random effects model. Supplementary methodologies included MR-Egger, simple mode, weighted median, and weighted mode analyses. IVW approach provided precise estimations assuming that all included SNPs served as effective instrumental variables [16]. Conversely, while capable of identifying and adjusting pleiotropy, MR-Egger regression typically resulted in lower accuracy of estimation [17]. Weighted median method provided reliable estimates contingent upon at least half of the instrumental variables being valid [18]. Although simple mode lacked the statistical power of IVW, it increased robustness against pleiotropy [19]. Weighted mode was particularly sensitive to the selection of bandwidth necessary for mode estimation [20].

Sensitivity analysis

Heterogeneity among SNP effect estimates for FLNB-related markers on RA was explored using Cochran's Q statistic in the IVW framework. Rucker's Q statistic within MR-Egger regression was used to test the significant heterogeneity [21]. MR-Egger intercept testing was further applied to screen for horizontal pleiotropy and assess its potential effect on causal estimates [17]. Leave-one-SNP-out analyses were performed to pinpoint SNPs that might significantly affect the outcomes and verified the robustness of the obtained results [22].

Results

Dysregulation of disulfidoptosis modulators accompanies enhanced immunological activity in RA

The results revealed 13 DRGs exhibiting significant differential expression between RA and non-RA control groups within peripheral blood samples. Among them, SLC7A11, CD2AP, and MYL6 presented increased expression in RA, while ACTB, INF2, PDLIM1, IQGAP1, FLNA, TLN1, FLNB, CAPZB, ACTN4, and LIAS exhibited significant downregulation in RA compared to non-RA controls (Figures 1A and 1B). The chromosomal locations of the 15 DRGs were further analyzed (Figure 1C). The correlation analysis identified strong synergistic interactions among certain disulfidoptosis modulators including ACTN4 and MYH9 with coefficient of 0.83, as well as antagonistic relationships such as CDKN2A and DLAT with coefficient of 0.68 (Figure 1D). Furthermore, gene relationship network diagram depicted the interconnected roles of these DRGs in the pathogenesis of RA (Figure 1E). To evaluate possible immunological disparities between healthy controls and RA patients, immune cell infiltration was analyzed to explore

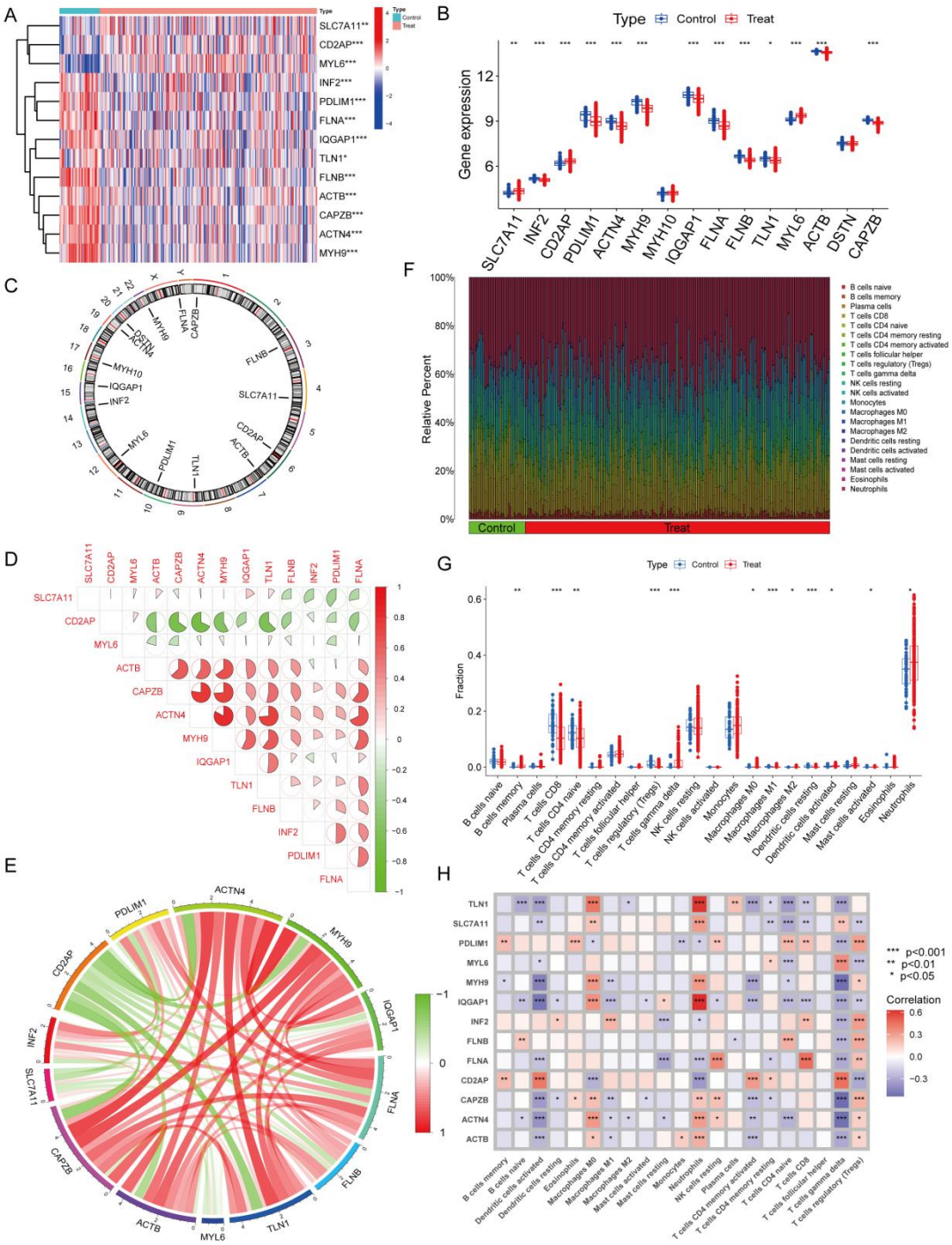


Figure 1. Identification of dysregulated DRGs in RA. **A.** The expression patterns of 13 DRGs. **B.** The expression of 13 DRGs between RA and non-RA controls. **C.** The location of 13 DRGs on chromosomes. **D.** Correlation analysis of 13 differentially expressed DRGs. **E.** Gene relationship network diagram of 13 differentially expressed DRGs. **F.** The relative abundances of 22 infiltrated immune cells between RA patients and healthy controls. **G.** The differences in immune infiltrating between RA patients and healthy controls. **H.** Correlation analysis of differentially expressed DRGs and infiltrated immune cells. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$.

variations in the distribution of 22 types of infiltrated immune cell using CIBERSORT algorithm (Figure 1F). The results revealed that RA presented elevated infiltration levels by B cell

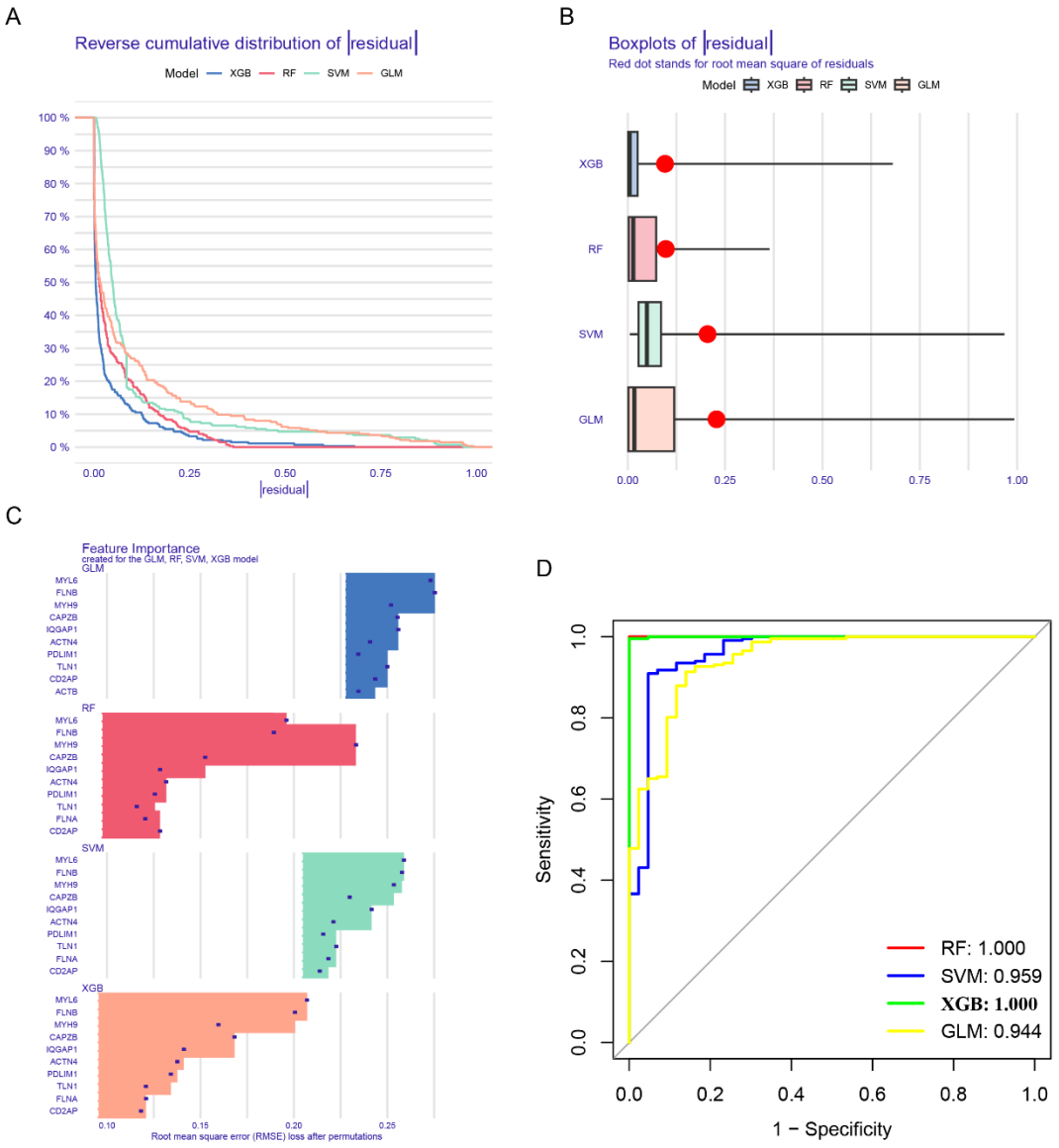


Figure 2. Performance evaluation of RF, SVM, GLM, and XGB models. **A.** Cumulative residual distribution of each model. **B.** Boxplots displaying the residuals of each model, with red dots indicating RMSE. **C.** Feature importance in the four models. **D.** ROC curve analysis of the models based on 5-fold cross-validation in the testing cohort.

memory, macrophages M0, macrophages M1, macrophages M2, dendritic cells resting, mast cells activated, and neutrophils (Figure 1G). The results suggested that modifications in the immune system might serve as a principal etiological factor in RA. The correlation analysis revealed associations among all disulfidoptosis modulators and various immune cell infiltrates (Figure 1H), underscoring the potential significance of DRGs in regulating the molecular and immune infiltration profiles of RA.

Construction and assessment of machine learning models

The results demonstrated that RF and XGB models presented lower residual distributions, indicating a more accurate fit to the data (Figures 2A and 2B). To further explore the predictive importance of variables, root mean square error (RMSE) was applied to rank the top 10 influential features for each model (Figure 2C). 5-fold cross-validation was applied to generate receiver operating characteristic (ROC) curves to evaluate

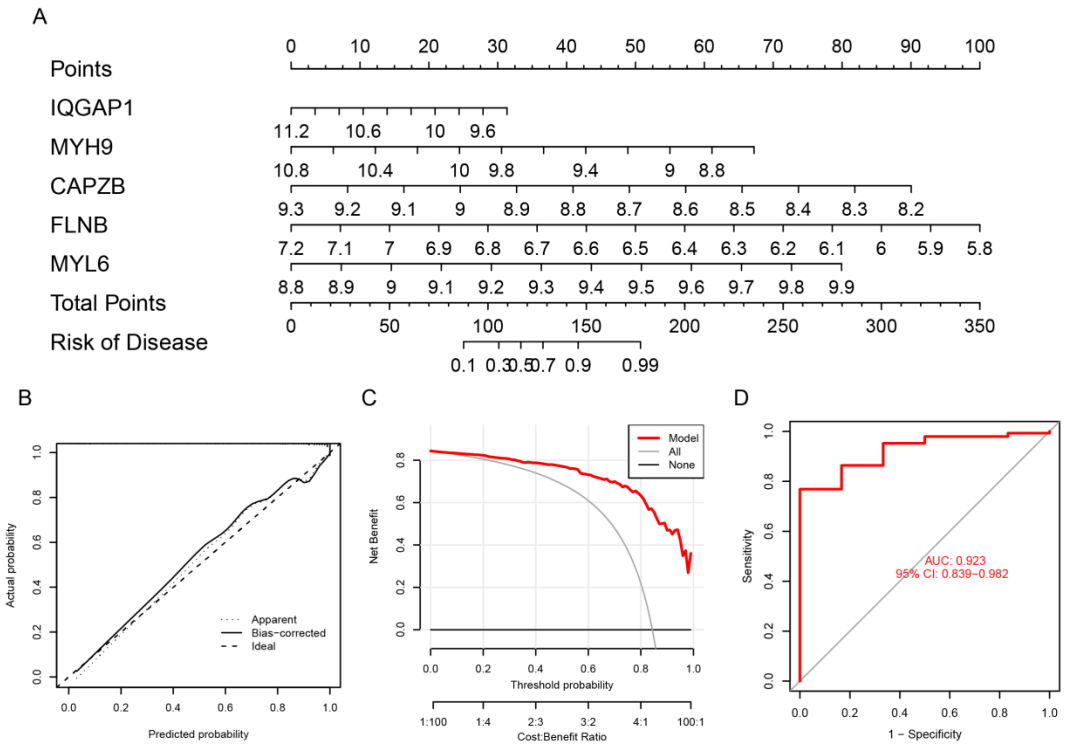


Figure 3. Validation and analysis of the 5-gene-based XGB model. **A.** A nomogram was constructed to predict the risk of RA using the 5-gene-based XGB model. **B.** The calibration curve showed the agreement between predicted and observed risks of RA. **C.** Decision curve analysis (DCA) plot was used to evaluate the clinical usefulness of the nomogram model. **D.** Receiver operating characteristic (ROC) curve analysis of the 5-gene-based XGB model based on 5-fold cross-validation in the GSE45291 dataset. The area under the curve (AUC) indicated the accuracy of the classification model.

the discriminatory performance of these machine learning algorithms in the testing set. Remarkably, both RF and XGB models exhibited higher area under the curve (AUC) scores compared to SVM and GLM models with RF AUC of 1.000, XGB AUC of 1.000, SVM AUC of 0.959, GLM AUC of 0.944 (Figure 3D), signifying their superior predictive accuracy. XGB model showed the highest capability in effectively distinguishing between RA patients and healthy controls. Therefore, the top five predictor genes of IQGAP1, MYH9, CAPZB, FLNB, and MYL6 were adopted from XGB model for further analyses. To further evaluate the predictive efficacy of XGB model, a nomogram model was developed to estimate disease risk (Figure 3A). The predictive accuracy of nomogram model was explored by using calibration curves and DCA. The obtained calibration curve demonstrated minimal discrepancy between the actual and predicted risks of RA (Figure 3B). In addition, DCA substantiated the high precision of the nomogram model, highlighting its potential application in clinical decision-making (Figure 3C). Further, an independent dataset comprising peripheral blood samples from both normal individuals and RA patients was applied to validate resulting 5-gene prediction model. In GSE45291 dataset, ROC analysis revealed robust predictive performance for the 5-gene model, achieving an AUC of 0.923 (Figure 3D).

Identification of disulfidoptosis clusters in RA

To explore disulfidoptosis-related transcriptional patterns in RA, a consensus clustering approach was applied to stratify 232 RA cases using expression data from 13 DRGs. The results showed that the highest stability of cluster numbers was obtained when k value was set to 2. CDF curves remained consistently stable within a narrow range with consensus index values ranging from 0.2 to 0.6 (Figures 4A and 4B). The

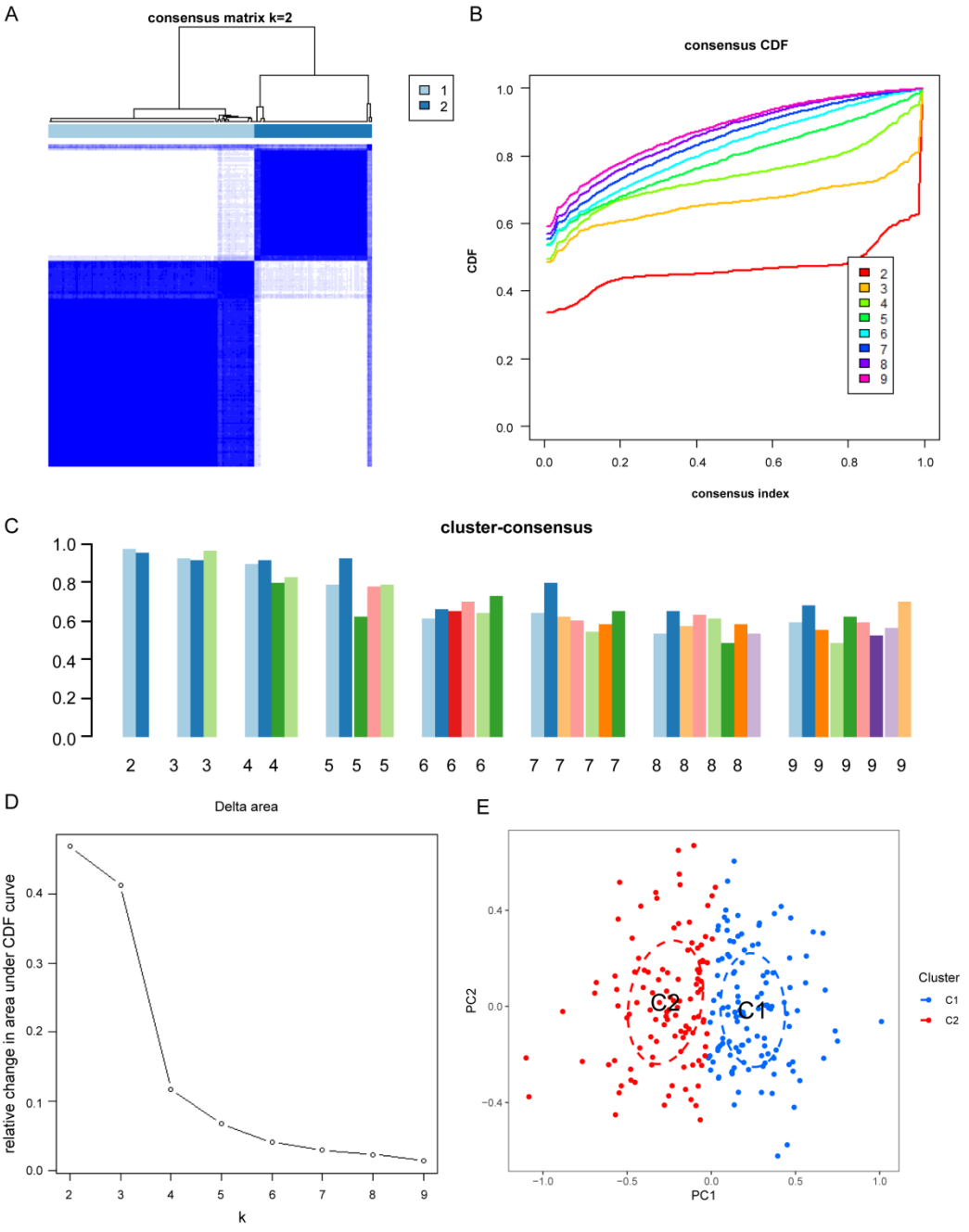


Figure 4. Discovery of disulfidptosis-related molecular clusters in RA. **A.** Consensus clustering matrix with two molecular clusters identified. **B.** Representative cumulative distribution function (CDF) curves. **C.** CDF delta area curves. **D.** Consensus clustering score. **E.** Non-negative matrix heatmap.

consistency score of each subtype significantly surpassed 0.9, revealing a strong classification efficacy at k = 2 (Figure 4C). A noticeable discrepancy was observed in the AUC of CDF curves when comparing k and k-1, ranging from 2 to 6 (Figure 4D). This classification was further validated through t-distributed stochastic neighbor embedding (t-SNE) analyses and resulted in a distinct separation between these two clusters (Figure 4E). These analyses confirmed the presence of two distinct molecular clusters associated with disulfidptosis in RA.

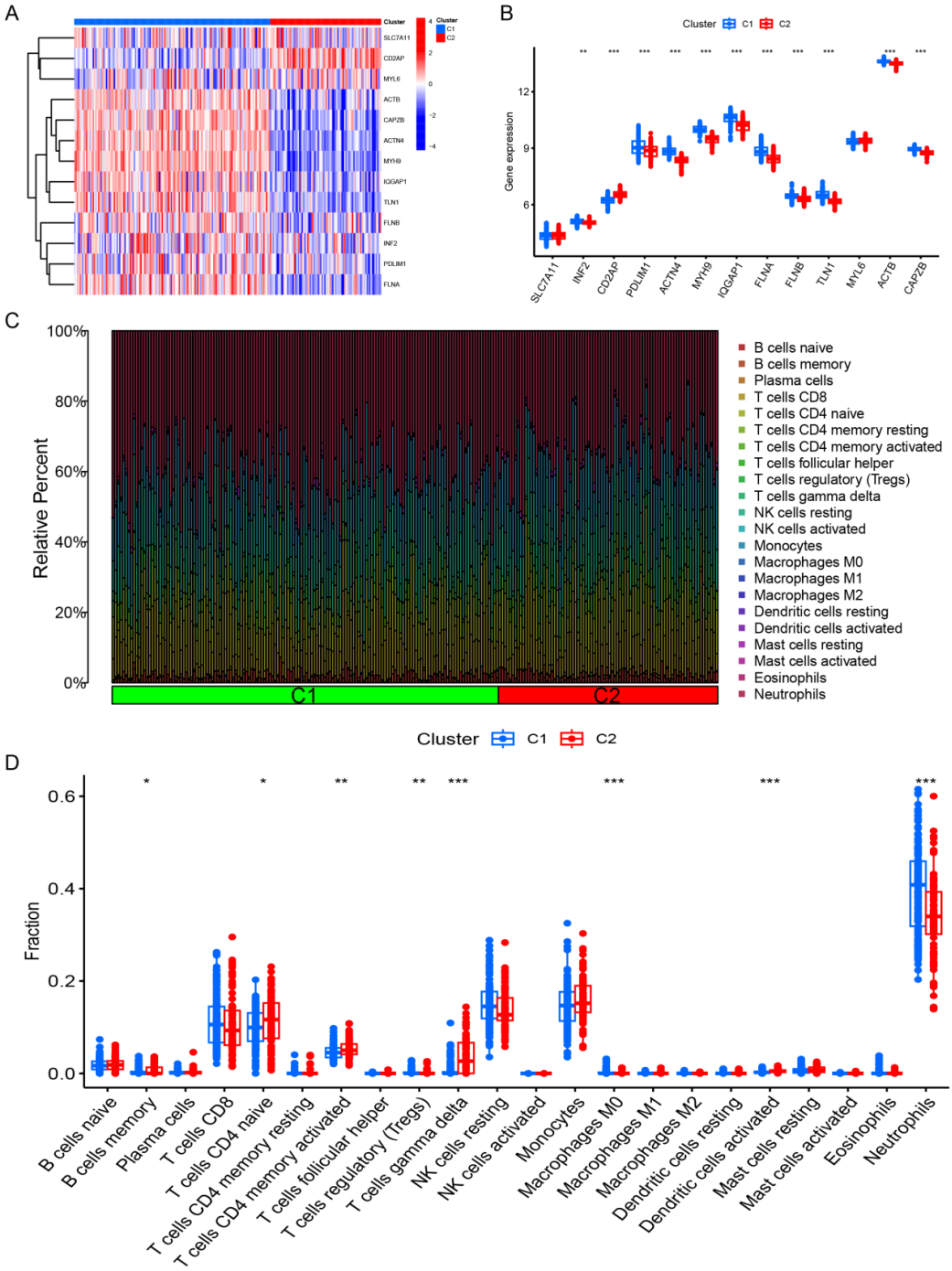


Figure 5. The molecular and immune profiles of the two distinct clusters in disulfidptosis. **A.** The expression patterns of DRGs and their corresponding clinical features. **B.** The expression level differences of DRGs between the two clusters. **C.** The relative abundances of 22 immune cell types between the two disulfidptosis clusters. **D.** The differences in immune cell infiltration between the two clusters. *: $P < 0.05$. **: $P < 0.01$. ***: $P < 0.001$.

Differentiation of disulfidoptosis regulators and immune infiltration characteristics between disulfidoptosis clusters

To explore the molecular characteristics distinguishing the two clusters, a comprehensive analysis was performed on the expression

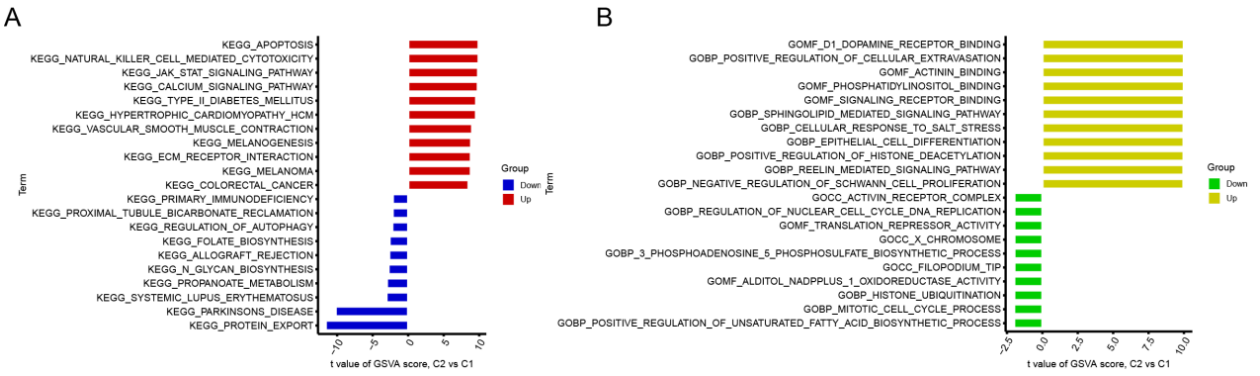


Figure 6. Comparison of pathway activities (**A**) and biological functions (**B**) between Cluster 1 and Cluster 2 disulfidoptosis clusters based on t-values obtained from GSVA method.

differences among the 13 DRGs between Cluster 1 and Cluster 2. Distinct DRGs expression patterns between the two disulfidoptosis subtypes were noticed (Figure 5A). The results showed that disulfidoptosis Cluster 1 presented elevated expression levels of INF2, PDLIM1, MYH9, IQGAP1, ACTN4, FLNB, TLN1, ACTB, and CAPZB, whereas CD2AP expression was increased in disulfidoptosis Cluster 2 (Figure 5B). In addition, discernible differences in immune microenvironments between disulfidoptosis Cluster 1 and Cluster 2 were found (Figure 5C). Notably, Cluster 1 exhibited higher proportions of regulatory T cells (T_{regs}), M0 macrophages, and neutrophils, while Cluster 2 presented relatively greater abundance of memory B cells, gamma delta T cells, naive CD4 T cells, activated memory CD4 T cells, and activated dendritic cells (Figure 5D).

Functional and pathway annotation between clusters

Gene set variation analysis (GSVA) was applied to explore functional disparities between the two clusters. The results found significant variations in functional pathways with Parkinson's disease and protein export pathways being enhanced in Cluster 1 and natural killer cell mediated cytotoxicity, apoptosis, JAK STAT signaling, calcium signaling, type II diabetes mellitus, and hypertrophic cardiomyopathy (HCM) pathways being upregulated in Cluster 2 (Figure 6A). Functional enrichment analyses unveiled distinct associations for each cluster with Cluster 1 presenting a significant involvement in the positive regulation of unsaturated fatty acid biosynthetic process, histone ubiquitination, mitotic cell cycle process, and alditol NADP plus 1 oxidoreductase activity. On the other hand, Cluster 2 exhibited enrichment in the positive regulation of cellular extravasation, D1 dopamine receptor binding, and actinin binding (Figure 6B).

MR analysis

Two-sample MR analysis showed that genetic predisposition to RA was causally related with an increased risk of FLNB. By using IVW method, the odds ratio of 1.091 (95% CI: 1.018 - 1.169, P = 0.013) was obtained. Weighted-median method provided consistent estimates (OR = 1.090, 95% CI: 1.008 - 1.179, P = 0.031). No significant causal effect was witnessed in MR-Egger, weighted-mode and simple mode analyses (Figure 7A). Sensitivity analyses revealed no significant horizontal pleiotropy. No prominent heterogeneity was observed for IVW estimates (Figure 7B). Taken together, genetically determined RA might confer an increased risk of FLNB.

Discussion

RA manifests with a wide range of pathological features and current therapeutic interventions have proven inadequate [23]. In recent decades,

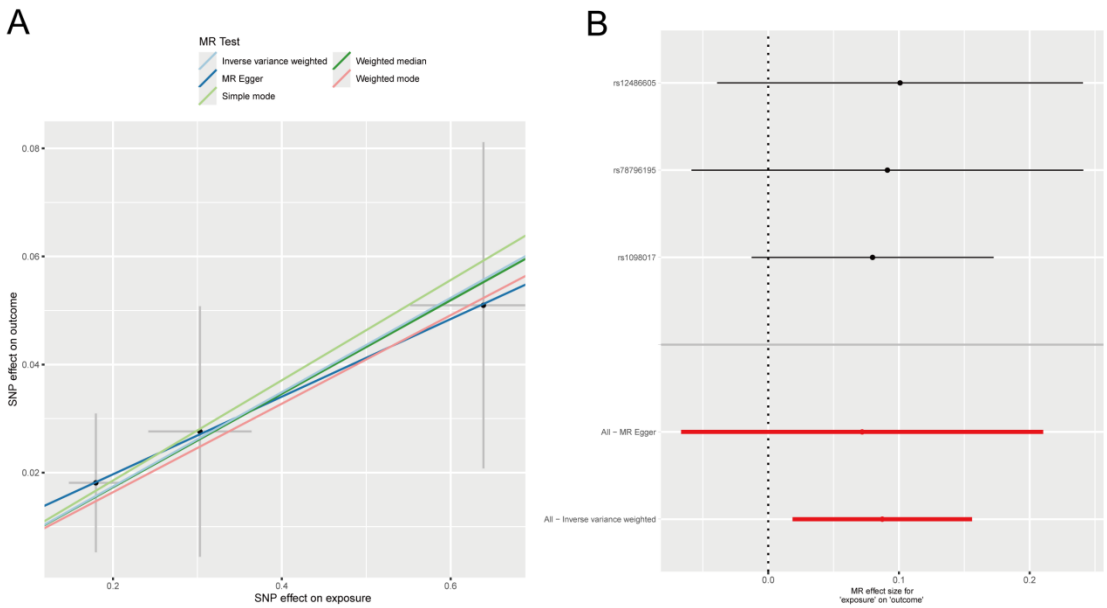


Figure 7. MR results of FLNB and RA. **A.** Genetic correlations of FLNB and RA using different MR methods. **B.** The causal effects of FLNB associated SNPs on RA. The red and black dot/bar indicated the cause estimate of FLNB level on risk of patients with RA.

the emergence of resistance to conventional RA treatments has posed a significant challenge [24]. Therefore, identification of pertinent molecular clusters is of significant importance in devising personalized approaches for RA management. Characterized by a programmed cell demise arising from the accumulation of cystine within cellular compartments, disulfidptosis has been implicated in the progression of various ailments. However, there remains a dearth of research investigating the precise regulatory functions and mechanisms of disulfidoptosis across different diseases. Therefore, this research aimed to evaluate the effects of disulfidoptosis-related genes on the phenotypic characterization of RA and intricate immune microenvironments and apply gene signatures associated with disulfidoptosis to prognosticate distinct RA subtypes. Through meticulous analysis of expression profiles of disulfidoptosis-related regulators in blood samples obtained from individuals with RA and healthy subjects, this research uncovered compelling evidence indicating aberrant expression of these regulators in RA pathogenesis. Further, this study explored intricate correlations among these regulators, revealing significant synergistic or antagonistic interactions among certain regulators, as witnessed by the presence of interactions involving DRGs within RA patients. Furthermore, substantial disparities in the composition of immune cell populations between healthy individuals and those with RA were identified. The latter presented elevated levels of infiltration by B cell memory, macrophages M0, macrophages M1, macrophages M2, activated mast cells, resting dendritic cells, and neutrophils. Importantly, the findings of this study aligned with earlier studies in terms of immune cell discrepancies observed in either blood or synovial tissues, highlighting the consistency of the conclusions from this research [25–28].

Recently, application of machine learning models to prognosticate RA occurrence in patients has become significantly popular [29, 30]. It has been shown that multifactorial analyses, considering intricate relationships among variables, offer more accurate outcomes with lower errors than those obtained from univariate analyses. This research compared the performance of four machine learning models including RF, GLM, SVM, and XGB utilizing DEGs and found that the

XGB-based model showed the highest predictive efficacy in the testing cohort, obtaining an impressive AUC of 1.000. This outcome emphasized the potential value of using XGB-based machine learning approaches in prognosticating RA. IQGAP1, MYH9, FLNB, CAPZB, and MYL6 were identified as crucial genes implicated in RA pathological processes. A nomogram model was then developed to diagnose RA using these five identified genes, providing exceptional predictive efficacy and potential clinical applicability. This developed nomogram model accurately predicted the presence of RA in an external validation dataset, exhibiting an AUC of 0.923, thereby providing valuable insights into the diagnosis of this intricate disease. IQGAP1 serves as a pivotal mediator contributing to cellular morphology, inflammation, and movement. Suppression of IQGAP1 expression has been shown to impede cytokine production by macrophages, thereby decreasing the severity of arthritis [31]. Notably, IQGAP1 has emerged as a novel susceptibility gene for RA in a comprehensive genome-wide association study [32]. Upregulated expression of IQGAP1 in monocytes from RA patients suggests its crucial role in the pathogenesis and advancement of this disease [33]. However, the remaining four genes have not yet been directly associated with RA. MYH9 encodes a non-muscle myosin that plays key roles in cellular motility, cytokinesis, and preservation of cellular morphology [34]. Proteomic research has unveiled elevated levels of MYH9 in apoptotic T cells. In RA, intensified inflammatory pathways contribute to an abundance of apoptotic cells due to the rapid turnover of effector T cells, undergoing apoptosis upon fulfilling their function [35]. Therefore, MYH9 may be involved in the immunopathological dynamics of RA. CAPZB encodes the beta subunit of the barbed-end actin binding protein, contributing to tissue repair and neuronal survival in spinal cord [36]. FLNB participates in vascular injury repair, and its mutations give rise to rare skeletal disorders [37, 38]. On the other hand, MYL6 encodes a myosin alkali light chain expressed in smooth muscle and non-muscle tissues and it markedly regulates the formation of neutrophil extracellular traps [39], which have been shown to drive inflammation when generated by the neutrophils present in the synovial fluid of RA patients [40]. More studies are required to fully reveal the roles that these genes play in RA development and progression. During clustering analysis, the subtypes exhibited significant variations in terms of disulfidoptosis-related gene expression and immune cell infiltration. Cluster 1 manifested prominently elevated DRG expression levels, along with a more frequent occurrence of immune cell infiltration. These findings strongly suggested the presence of diverse pathogenic mechanisms underlying the two distinct RA subtypes, implying the necessity for tailored treatment strategies for each subtype. To strengthen confounding control and enhance causal inference, MR was applied to explore potential causal links between DRGs and RA. Genetic predisposition to RA was found to be related to specific DRGs.

This research offered notable strengths. The joint application of bioinformatics and MR analyses facilitated a thorough examination of relationships between disulfidoptosis and RA. Such an integrative design enhanced overall interpretability and ensured robustness through triangulation of findings across multiple inquiry lines. However, this research faced some limitations. The reliance on extensive bioinformatic analyses necessitated the supplementation of additional experimental or clinical studies to validate DRGs expression levels. The inclusion of more comprehensive clinical information is essential to confirm the accuracy of the proposed predictive model. In addition, a larger sample size of RA patients would be advantageous in the validation of disulfidoptosis associated with clusters and further investigation of the potential correlations among DRGs and immune responses. Due to a lack of data, only eQTL data for FLNB were obtained. eQTL data for the remaining four predicted genes were unavailable, making it impossible to validate the causal relationship between these four genes and RA *via* Mendelian

randomization (MR) analysis. This research confirmed a robust association between DRGs and RA and developed a new 5-gene-based XGB model, which demonstrated excellent performance in the prediction of the pathological outcomes of RA. Importantly, this study confirmed that disulfidoptosis acted as a key regulator in RA pathogenesis and further deciphered the molecular mechanisms underlying disease heterogeneity. Additionally, a causal relationship was identified between FLNB and RA.

References

1. Delft MAMV, Huizinga TWJ. 2020. An overview of autoantibodies in rheumatoid arthritis. J Autoimmun. 110:102392.
2. Li X, Ding S, Zhang P, Yan J, Yu X, Wang X, *et al*. 2024. Revealing the impact of autophagy-related genes in rheumatoid arthritis: Insights from bioinformatics. Heliyon. 10(9):e29849.
3. Li Z, Wang J, Lin Y, Fang J, Xie K, Guan Z, *et al*. 2022. Newly discovered circRNAs in rheumatoid arthritis, with special emphasis on functional roles in inflammatory immunity. Front Pharmacol. 13:983744.
4. Aletaha D. 2020. Precision medicine and management of rheumatoid arthritis. J Autoimmun. 110:102405.
5. Huang T, Takashima Y, Kumar J, Morales J, Le K, Hu Z, *et al*. 2026. Macrophage metabolic exhaustion and PANoptotic cell death drive chronic tissue inflammation in rheumatoid arthritis. Immunity. 59(4):970–987.
6. Liu X, Nie L, Zhang Y, Yan Y, Wang C, Colic M, *et al*. 2023. Actin cytoskeleton vulnerability to disulfide stress mediates disulfidptosis. Nat Cell Biol. 25(3):404–414.
7. Liu H, Tang T. 2023. Pan-cancer genetic analysis of disulfidptosis-related gene set. Cancer Genet. 278–279:91–103.
8. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, *et al*. 2023. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 613(7944):508–518.
9. Kuppa A, Tripathi H, Al-Darraji A, Tarhuni WM, Abdel-Latif A. 2023. C-reactive protein levels and risk of cardiovascular diseases: A two-sample bidirectional Mendelian randomization study. Int J Mol Sci. 24(11):9129.
10. Yu Z, Coresh J, Qi G, Grams M, Boerwinkle E, Snieder H, *et al*. 2020. A bidirectional Mendelian randomization study supports causal effects of kidney function on blood pressure. Kidney Int. 98(3):708–716.
11. Acharjee A, Larkman J, Xu Y, Cardoso VR, Gkoutos GV. 2020. A random forest based biomarker discovery and power analysis framework for diagnostics research. BMC Med Genomics. 13(1):178.
12. Gold C, Sollich P. 2003. Model selection for support vector machine classification. Neurocomputing. 55(1–2):221–249.
13. Nelder JA, Wedderburn RWM. 1972. Generalized linear models. J R Stat Soc. 135(3):370–384.
14. Wang YF, Hsu MH, Wang MYF, Lin JW. 2025. Estimating hair density with XGBoost. Int J Cosmet Sci. 47(2):336–342.
15. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, *et al*. 2015. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 12(5):453–457.
16. Yuan S, Larsson S. 2020. Causal associations of iron status with gout and rheumatoid arthritis, but not with inflammatory bowel disease. Clin Nutr. 39(10):3119–3124.
17. Burgess S, Thompson SG. 2017. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 32(5):377–389.
18. Bowden J, Davey Smith G, Haycock PC, Burgess S. 2016. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 40(4):304–314.
19. Milne RL, Kuchenbaecker KB, Michailidou K, Beesley J, Kar S, Lindström S, *et al*. 2017. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet. 49(12):1767–1778.
20. Hartwig FP, Davey Smith G, Bowden J. 2017. Robust inference in summary data Mendelian randomization *via* the zero modal pleiotropy assumption. Int J Epidemiol. 46(6):1985–1998.
21. Shu MJ, Li JR, Zhu YC, Shen H. 2022. Migraine and ischemic stroke: A Mendelian randomization study. Neurol Ther. 11(1):237–246.
22. Zhang J. 2022. Mendelian randomization study implies causal linkage between telomere length and juvenile idiopathic arthritis in a European population. J Inflamm Res. 15:977–986.
23. Yu MB, Firek A, Langridge WHR. 2018. Predicting methotrexate resistance in rheumatoid arthritis patients. Inflammopharmacology. 26(3):699–708.
24. Zhang KX, Ip CK, Chung SK, Lei KK, Zhang YQ, Liu L, *et al*. 2020. Drug-resistance in rheumatoid arthritis: The role of p53 gene mutations, ABC family transporters and personal factors. Curr Opin Pharmacol. 54:59–71.
25. van Duivenvoorde LM, van Mierlo GJD, Boonman ZFHM, Toes REM. 2006. Dendritic cells: Vehicles for tolerance induction and prevention of autoimmune diseases. Immunobiology. 211(6–8):627–632.
26. Fedele AL, Tolusso B, Gremese E, Bosello SL, Carbonella A, Canestri S, *et al*. 2014. Memory B cell subsets and plasmablasts are lower in early than in long-standing rheumatoid arthritis. BMC Immunol. 15:28.
27. Cheng L, Wang Y, Wu R, Ding T, Xue H, Gao C, *et al*. 2021. New insights from single-cell sequencing data: Synovial fibroblasts and synovial macrophages in rheumatoid arthritis. Front Immunol. 12:709178.
28. Mariola KS, Stefano A. 2022. Synovial tissue macrophages in joint homeostasis, rheumatoid arthritis and disease remission. Nat Rev Rheumatol. 18(7):384–397.
29. Vincent B, Julien D, Signe H, Paul-Henry C, Bruno F, Francis G, *et al*. 2022. Machine learning predicts response to TNF inhibitors in rheumatoid arthritis: Results on the ESPOIR and ABIRISK cohorts. RMD open. 8(2):e002442.

30. Weiyang T, Arno NC, Marieke V, Anne CAM, Floris PGJL, Timothy RDJR, *et al*. 2021. Multiomics and machine learning accurately predict clinical response to adalimumab and etanercept therapy in patients with rheumatoid arthritis. Arthritis & rheumatology (Hoboken, NJ). 73(2):212–222.
31. Akula MK, Ibrahim MX, Ivarsson EG, Khan OM, Kumar IT, Erlandsson M, *et al*. 2019. Protein prenylation restrains innate immunity by inhibiting Rac1 effector interactions. Nat Commun. 10(1):3975.
32. Leng RX, Di DS, Ni J, Wu XX, Zhang LL, Wang XF, *et al*. 2020. Identification of new susceptibility loci associated with rheumatoid arthritis. Ann Rheum Dis. 79(12):1565–1571.
33. Roszkowski L, Jaszczyk B, Plebańczyk M, Ciechomska M. 2022. S100A8 and S100A12 proteins as biomarkers of high disease activity in patients with rheumatoid arthritis that can be regulated by epigenetic drugs. Int J Mol Sci. 24(1):710.
34. Ye G, Yang Q, Lei X, Zhu X, Li F, He J, *et al*. 2020. Nuclear MYH9-induced CTNNB1 transcription, targeted by staurosporin, promotes gastric cancer cell anoikis resistance and metastasis. Theranostics. 10(17):7545–7560.
35. Alessanfra C, Rossana S, Helene M, Carmela M, Paolo D, Anna RV, *et al*. 2015. CD8+ T cells specific to apoptosis-associated antigens predict the response to tumor necrosis factor inhibitor therapy in rheumatoid arthritis. PloS One. 10(6):e0128607.
36. Wang CY, Chen JK, Wu YT, Tsai MJ, Shyue SK, Yang CS, *et al*. 2011. Reduction in antioxidant enzyme expression and sustained inflammation enhance tissue damage in the subacute phase of spinal cord contusive injury. J Biomed Sci. 18(1):13.
37. Prachi S, Yamini C, Lamya AA-K, Mohd S, Nawaf A, Sachidanand S. 2022. Classifying integrated signature molecules in macrophages of rheumatoid arthritis, osteoarthritis, and periodontal disease: An omics-based study. Curr Issues Mol Biol. 44(8):3496–3517.
38. Wilson SG, Jones MR, Mullin BH, Dick IM, Richards JB, Pastinen TM, *et al*. 2009. Common sequence variation in FLNB regulates bone structure in women in the general population and FLNB mRNA expression in osteoblasts *in vitro*. J Bone Miner Res. 24(12):1989–1997.
39. Yoshinari M, Nishibata Y, Masuda S, Nakazawa D, Tomaru U, Arimura Y, *et al*. 2022. Low disease activity of microscopic polyangiitis in patients with anti-myosin light chain 6 antibody that disrupts actin rearrangement necessary for neutrophil extracellular trap formation. Arthritis Res Ther. 24(1):274.
40. Wright HL, Lyon M, Chapman EA, Moots RJ, Edwards SW. 2020. Rheumatoid arthritis synovial fluid neutrophils drive inflammation through production of chemokines, reactive oxygen species, and neutrophil extracellular traps. Front Immunol. 11:584116.