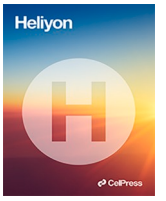
Heliyon 10 (2024) e29849

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Revealing the impact of autophagy-related genes in rheumatoid arthritis: Insights from bioinformatics

Xin Li [a], Shuang Ding [b], Pengcheng Zhang [c], Jing Yan [c], Xingxing Yu [c], Xukai Wang [d], Hongsheng Zhan [e], Zhengyan Wang [f,*]

[a] *Traumatology Hand Surgery Department, Haicheng Orthopedic Hospital, Haicheng, China*
[b] *Department of Orthopedics, The Affiliated Hospital of Changchun University of Chinese Medicine, Changchun, China*
[c] *Changchun University of Chinese Medicine, Changchun, China*
[d] *Department of Orthopedics, The Affiliated Hospital of Changchun University of Chinese Medicine, China*
[e] *ShuGuang Hospital, ShangHai, China*
[f] *Department of Orthopedics, Changchun University of Chinese Medicine, Changchun, China*

ARTICLE INFO

Keywords:
Rheumatoid arthritis
Autophagy
Autophagy-associated genes
Weighted gene co-expression network analysis
Consensus clustering
Machine learning models
Nomogram

ABSTRACT

Background: Rheumatoid arthritis is a systemic inflammatory autoimmune disease that severely impacts physical and mental health. Autophagy is a cellular process involving the degradation of cellular components in lysosomes. However, from a bioinformatics perspective, autophagy-related genes have not been comprehensively elucidated in rheumatoid arthritis.

Methods: In this study, we performed differential analysis of autophagy-related genes in rheumatoid arthritis patients using the GSE93272 dataset from the Gene Expression Omnibus database. Marker genes were screened by least absolute shrinkage and selection operator. Based on marker genes, we used unsupervised cluster analysis to elaborate different autophagy clusters, and further identified modules strongly associated with rheumatoid arthritis by weighted gene co-expression network analysis. In addition, we constructed four machine learning models, random forest model, support vector machine model, generalized linear model and extreme gradient boosting based on marker genes, and based on the optimal machine learning model, a nomogram model was constructed for distinguishing between normal individuals and rheumatoid arthritis patients. Finally, five external independent rheumatoid arthritis datasets were used for the validation of our results.

Results: The results showed that autophagy-related genes had significant expression differences between normal individuals and osteoarthritis patients. Through least absolute shrinkage and selection operator screening, we identified 31 marker genes and found that they exhibited significant synergistic or antagonistic effects in rheumatoid arthritis, and immune cell infiltration analysis revealed significant changes in immune cell abundance. Subsequently, we elaborated different autophagy clusters (cluster 1 and cluster 2) using unsupervised cluster analysis. Next, further by weighted gene co-expression network analysis, we identified a brown module strongly associated with rheumatoid arthritis. In addition, we constructed a nomogram model for five marker genes (CDKN2A, TP53, ATG16L2, FKBP1A, and GABARAPL1) based on a generalized linear model (area under the curve = 1.000), and the predictive efficiency and accuracy of this

* Corresponding author. Department of Orthopedics, Changchun University of Chinese Medicine, Changchun, China.
E-mail addresses: 308409220@qq.com (X. Li), 109383510@qq.com (S. Ding), 942684620@qq.com (P. Zhang), 22102570143@stu.ccucm.edu.cn (J. Yan), fishstar12@163.com (X. Yu), wangxukai_2003@163.com (X. Wang), 13918449223@139.com (H. Zhan), 13636731877@163.com (Z. Wang).

https://doi.org/10.1016/j.heliyon.2024.e29849
Received 18 October 2023; Received in revised form 16 April 2024; Accepted 16 April 2024
Available online 20 April 2024

nomogram model were demonstrated in the calibration curves, the decision curves and the five external independent datasets were validated.
Conclusion: This study identified marker autophagy-related genes in rheumatoid arthritis and analyzed their impact on the disease, providing new perspectives for understanding the role of autophagy-related genes in rheumatoid arthritis and providing new directions for its individualized treatment.

1. Introduction

Rheumatoid arthritis (RA) is a common autoimmune disease that adversely affects both physical and mental health. It arises from a combination of genetic, epigenetic, and environmental factors [1]. RA patients' immune systems exhibit abnormal cellular and humoral responses, leading to autoantibodies production, including **Rheumatoid Factor** (RF) and Anti-Mutated Citrullinated Protein Antibodies (AMPAs). Additionally, T cells and B cells' abnormal activation contributes to synovitis formation [2].

RA's clinical and histological manifestations are diverse. Synovitis results in joint swelling, pain, and cartilage and bone destruction, impairing patients' daily functions [3]. Synovial hyperplasia, a typical RA characteristic, is a significant factor in pannus formation. The extent of synovial hyperplasia correlates with cartilage erosion severity, and osteoclasts activation leads to bone destruction. In RA synovium, synovial cells secrete numerous matrix-degrading enzymes, such as collagenases, matrix metalloproteinases, and gelatinases. RA can potentially harm various organ systems, such as the cardiovascular system, kidneys, lungs, and nervous system, not just the joints [4,5]. Therefore, early diagnosis and management are vital to prevent RA-induced joint damage. Research indicates that severe permanent joint damage can occur within the first 2 years of RA onset, making optimal RA management within the first 3–6 months crucial. Early diagnosis can prevent severe joint damage and improve patient prognosis [6,7]. Modern medicine offers improved management and long-term prognosis for RA patients through a wide range of drugs and personalized treatment plans.

Autophagy, a cellular process involving cellular components degradation in lysosomes, is categorized into macroautophagy, chaperone-mediated autophagy (CMA), and microautophagy [8]. Macroautophagy degrades substances within autophagosomes and is vital for maintaining cellular homeostasis. CMA selectively degrades proteins, affecting cellular metabolism, while micro-autophagy non-selectively degrades cytoplasmic cargo. Autophagy disruption can lead to cellular imbalance and contribute to disease development [9]. Recent research has indicated a strong correlation between autophagy and RA. Inhibiting persistently activated autophagy can reduce synovial inflammation and osteoclastogenesis and prevent structural damage in RA mice [10]. Studies have indicated the involvement of PINK1/Parkin-mediated mitochondrial autophagy in the excessive growth of RA-FLS caused by H2O2, suggesting that PINK1/Parkin-mediated mitochondrial autophagy could potentially serve as a crucial mechanism for treating RA. H2O2-induced oxidative stress leads to mitochondrial dysfunction in RA-FLS. This is characterized by a decreased mitochondrial membrane potential and elevated levels of mitochondrial and intracellular reactive oxygen species (ROS). The disruption of mitophagy results in abnormal cell proliferation and anti-apoptotic behavior. The study highlighted the critical role of the PINK1/Parkin pathway in maintaining mitochondrial quality control under stress conditions. Oxidative stress downregulates PINK1 and Parkin, while *N*-acetylcysteine (NAC) treatment can counteract oxidative stress, indicating that protecting this pathway is crucial for preventing excessive growth of RA-FLS [11]. Recent research has also demonstrated that Wenhua Juanbi Recipe (WJR) inhibits autophagy, apoptosis, and proliferation through the miRNA-146a-mediated PI3K/AKT/mTOR pathway, having a therapeutic impact in a rat model [12]. However, the potential regulatory mechanism of cellular autophagy in RA based on bioinformatics has not been fully elucidated. Therefore, further clarification of the molecular characteristics of **Autophagy Related Genes** (ARGs) may explain the complex pathogenesis of RA and provide new ideas and targets for personalized RA treatment.

In this study, we conducted a systematic bioinformatics study on normal individuals and RA patients for the first time, exploring their differences in ARGs expression and immune characteristics. We performed Consensus Clustering analysis on 232 RA patients based on the expression profiles of 31 significantly differentially expressed characteristic ARGs, dividing them into two autophagy-related clusters, and further studied the differences in immune cells between these two clusters. We also conducted Gene Set Variation Analysis (GSVA) and Gene Set Enrichment Analysis (GSEA) on several clusters. After that, we identified **Differentially Expressed Genes** (DEGs) based on clusters using **Weighted Gene Co-Expression Network Analysis** (WGCNA), which also indicated linked biological functions and pathway enrichment based on these DEGs. We once more separated the 232 RA patients into two clusters based on characteristic DEGs using the Least Absolute Shrinkage and Selection Operator (LASSO). To show the association between the two clusters, we calculated the score of each RA patients based on the expression profiles of 31 characteristic ARGs using the Principal Component Analysis (PCA) algorithm and plotted the difference box plot. Finally, we also established a nomogram model by comparing multiple machine learning algorithms and validated the performance of this model through decision curve analysis (DCA), calibration curves, and 5 external datasets. Therefore, based on autophagy, we provide new insights into the immune characteristics, clustering, and risk prediction of RA from a bioinformatics perspective.

2. Methods

2.1. Dataset acquisition and selection of characteristic differential genes

We obtained six datasets, GSE93272 [13], GSE17755 [14], GSE56649 [15], GSE68689, GSE77298 [16], and GSE71370 [17], from the Gene Expression Omnibus (GEO) database [18]. The GSE93272 dataset (GPL 570 platform) included 43 normal individuals and 232 RA patients for further analysis. The other five datasets were used for validation. ARGs totaling 232 were retrieved from the Human Autophagy Database (http://www.autophagy.lu/index.html). We normalized the original gene expression profile of the GSE93272 dataset using the "limma" package in R 4.2.2 [19], extracted the expression levels of the 232 ARGs in each sample, and conducted differential analysis with a standard of p-value less than 0.05. We used the LASSO to screen out differentially expressed characteristic ARGs and performed correlation analysis of characteristic ARGs using the "corrplot" package in R 4.2.2 [20]. The flow chart of this study was shown in Fig. 1.

2.2. Immune cell infiltration analysis

To assess the relative abundance of 22 different types of immune cells in each sample of the gene expression data, we employed the CIBERSORT method and the LM 22 characteristic matrix [21]. Then, we analyzed the correlation between immune cells and characteristic ARGs using the Spearman correlation coefficient. Correlations were considered significant if the p-value <0.05. We visualized the results using the "corrplot" package in R 4.2.2.

2.3. Consensus Clustering and Immune Cell Infiltration Analysis of Different Clusters in Rheumatoid Arthritis Patients

The 232 RA patients were divided into distinct ARGs clusters using the k-means method with 1000 iterations using the "ConsensusClusterPlus" package in R 4.2.2 [22]. We selected the maximum subtype number k ($k = 9$) and thoroughly assessed the ideal number of clusters using the consistency matrix and **Cumulative Distribution Function** (CDF) curve. Then, we used the PCA algorithm to validate the reliability of the optimal clustering. Finally, using the CIBERSORT method, we conducted immune cell infiltration analyses on several clusters.

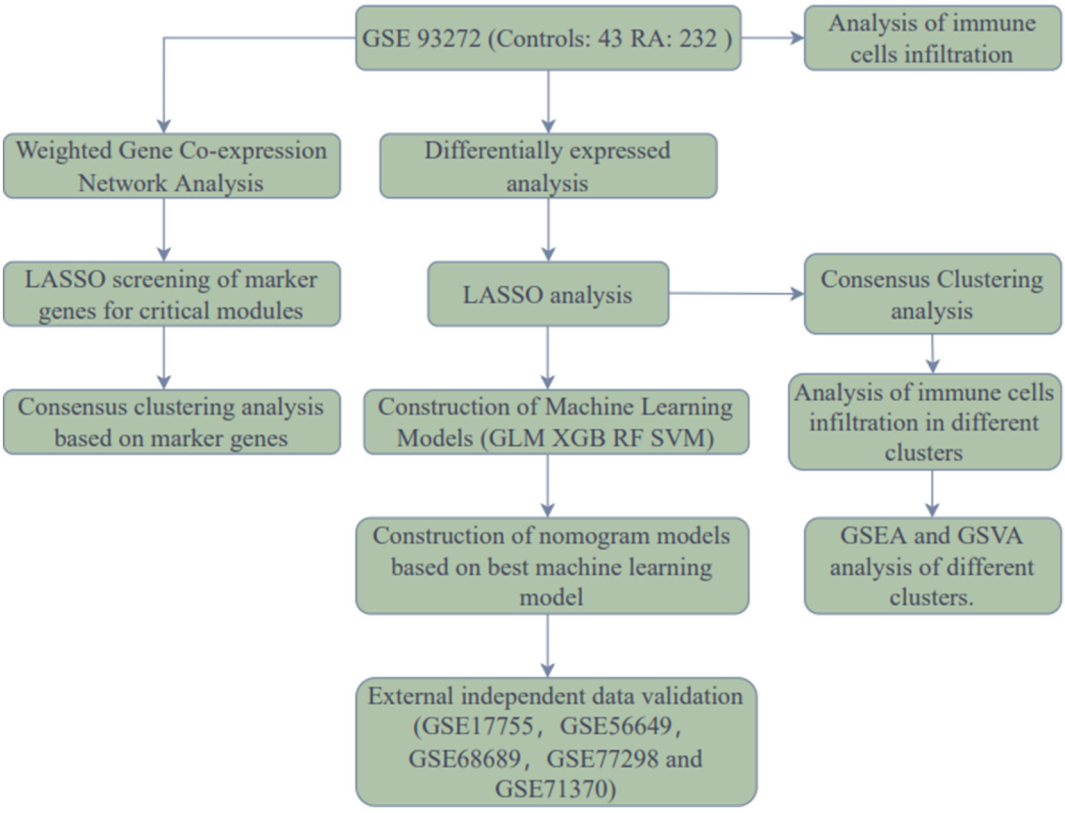


Fig. 1. Flowchart of this study.

2.4. Gene Set Enrichment Analysis and Gene Set Variation Analysis

We used GSEA and GSVA to analyze gene expression data. In order to evaluate associated pathways and biological mechanisms based on various **ARGs cluster gene expression profiles**, we obtained the GSEA program (version 3.0) from the GSEA website [23] as well as the "c2.cp.kegg.v7.4.symbols.gmt" and "c5.go.symbols.gmt" **sub-sets** from the MSigDB website [24]. We ran 1000 repetitions of resampling with a minimum gene set of 5 and a maximum gene set of 5000. Statistical significance was defined as P-values ≤ 0.05 and FDR < 0.25. In GSVA analysis, to clarify the variations in enriched gene sets between various ARGs clusters, we used the "GSVA" package in R 4.2.2 for GSVA enrichment analysis [25]. Specifically, we first used the gene expression profile, the method of David A Barbie [26], and predefined gene sets to evaluate the enrichment scores of different gene sets, set the minimum gene set to 5, the maximum gene set to 5000, calculated the enrichment score of each sample in each gene set, and finally obtained the enrichment score matrix. By evaluating the GSVA scores between different ARGs clusters, we used the "limma" in R 4.2.2 to find biological activities and pathways that were differentially expressed. GSVA is a non-parametric, unsupervised method for estimating pathway and biological process activity changes in expression datasets, suitable for analyzing gene expression data to estimate pathway or gene set activity across different samples or conditions. Within ARG cluster analysis, GSVA aids in identifying the activity changes of biological processes in each cluster. Particularly when ARG clusters represent different biological states or conditions, GSVA can reveal the biological pathways that distinguish these clusters. Utilizing the "limma" package for statistical analysis can enhance such analyses; it identifies significant differences through robust statistical methods, which is crucial for deciphering pathogenesis, identifying biological markers, or exploring therapeutic targets.

2.5. Weighted gene Co-expression network analysis

We utilized the WGCNA method to reveal gene interactions and functionally related gene sets through the construction of gene co-expression networks and modular analysis [27]. We chose the top 25 % of the genes with the highest variance for WGCNA analysis in order to improve the biological relevance, computational feasibility, and interpretive clarity of the analysis, thereby increasing the likelihood of meaningful and actionable in-depth analyses of the biological systems under study. We created a topological overlap matrix (TOM) using the weighted adjacency matrix to select the optimal soft threshold. Using the hierarchical clustering tree and the TOM dissimilarity measure (1-TOM), we produced modules with the minimum module size set to 100. Each module received a random color that represented the genes that make it up. Module significance (MS) was used to assess the relationship between the module and RA. Gene significance (GS) was used to represent the association of genes with clinical phenotypes. In this study, we set GS as 0.5. Through these steps, we were able to accurately identify gene co-expression modules and explore their association with disease.

2.6. Construction and Consensus Clustering Analysis of the Least Absolute Shrinkage and Selection Operator Model

We used the LASSO algorithm to process the most significant gene module in the WGCNA analysis to identify characteristic genes in the module. Then, based on these characteristic genes, we again used the "ConsensusClusterPlus" package in R 4.2.2 for consensus clustering analysis, dividing the 232 RA patients into different gene clusters through the execution of the k-means algorithm with 1000 iterations. The maximum cluster number k ($k = 9$) was chosen, and the consistency matrix and cumulative distribution function (CDF) curve were used to thoroughly examine the ideal number of clusters. Next, we used the PCA algorithm to quantify the two clustering modes and calculated the ARGs-score of each sample based on the expression level of characteristic ARGs in each sample. Finally, we used "limma" package of R 4.2.2 and "ggpubr" to visualize the results [28].

2.7. Construction and selection of four machine learning models

We utilized the "caret" [29], "xgboost" [30], "kernlab" [31], "randomForest" [32], and "ggplot2" [33] packages in R 4.2.2 to construct four machine learning models based on the characteristic ARGs, and visualized them. These four machine learning models were the Extreme Gradient Boosting (XGB), the Generalized Linear Model (GLM), the Support Vector Machine (SVM), and the Random Forest (RF) model. RF is an ensemble learning method based on decision trees, which makes predictions by randomly selecting characteristics and using a voting mechanism. SVM classifies or regresses in the characteristic space by finding the optimal hyperplane, suitable for both linear and non-linear problems. GLM is based on the linear model, modeling and predicting relationships through appropriate link functions and error distributions. XGB builds multiple weak learners iteratively and weights them based on the errors of the previous round to improve the accuracy of predictions. These four machine learning models have wide applications in bioinformatics research. Using characteristic ARGs, we built machine learning models in our studies. A training set (70 %, $N = 162$) and a validation set (30 %, $N = 70$) were created from a total of 232 RA patients. Then, we evaluated them through 10-fold cross-validation. The residual distribution, the significance of the characteristic genes, and the Receiver Operating Characteristic (ROC) curve of the four machine learning models were all visualized using the "DALEX" [34] and "pROC" [35] packages in R 4.2.2. Finally, we selected the top five significant variables (characteristic ARGs) as the most crucial predictor genes associated with RA after determining the most effective machine learning model.

2.8. Development of the nomogram model

We used the "rms" package [36] in R 4.2.2 to develop a nomogram model based on the most powerful machine learning model to

determine the probability that RA would occur. The "total score" is the total of all predictor scores, with each predictor having a corresponding score. In order to assess the nomogram model's capability for prediction, we used calibration curves and DCA.

3. Validation with external independent datasets

We applied five external datasets GSE17755 (GPL 1291 platform, Con: 45, RA: 112), GSE56649 (GPL 570 platform, Con: 9, RA: 13), GSE68689 (GPL 20171 platform, Con: 5, RA: 16), GSE77298 (GPL 570 platform, Con: 7, RA: 16), and GSE71370 dataset (GPL 16268 platform, Con: 8, RA: 18) to validate the predictive model's ability to distinguish RA patients from normal individuals. Subsequently, we visualized the ROC curves of these five validation datasets using the "pROC" package in R 4.2.2.

4. Results

4.1. Differential analysis of autophagy-related genes

In our initial analysis, we examined the expression profiles of 232 ARGs in RA patients and normal individuals using the GSE93272 dataset. We found significant differential expression in 142 ARGs (Fig. 2A). Using a LASSO model, we identified 31 characteristic ARGs, with 13 upregulated and 18 downregulated genes (Fig. 2B and C). In the RA patients, genes such as CDKN2A, FKBP1A, GABARAPL1, APOL1, ATG5, BAX, CASP4, FAS, PIK3C3, RAB24, SAR1A, TNFSF10, and ULK2 showed increased expression compared to the normal individuals. Conversely, genes like TP53, ATG16L2, ARSA, BAG3, CANX, EDEM1, EIF4G1, ERBB2, GAA, GNB2L1, HSPA5, IKBKB, IL24, ITGA6, MAP2K7, MAPK3, SPNS1, and TMEM74 were downregulated in the RA patients (Fig. 2D and E). To understand the role of these ARGs in RA progression, we performed a correlation analysis. FKBP1A displayed a significant negative connection with TP53 and a substantial positive correlation with RAB24. Similarly, ATG5 and GABARAPL1 had a negative correlation whereas MAP2K7 and GABARAPL1 had a positive correlation. We found significant correlations among CDKN2A, FKBP1A, GABARAPL1, TP53, ATG16L2, and other ARGs (Fig. 2F).

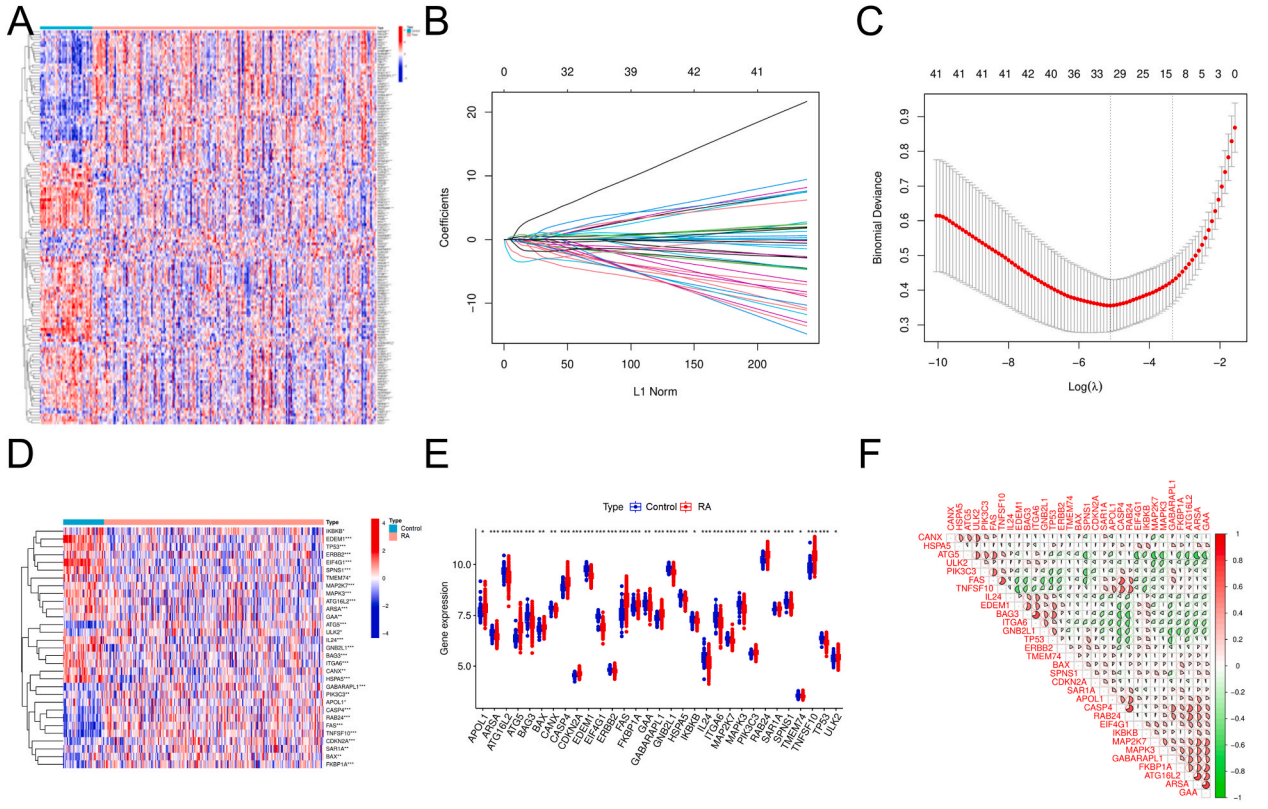


Fig. 2. Differential analysis and filtering of autophagy-related genes (ARGs). (A) Heatmap of 142 ARGs. (B) Coefficient diagram of Least Absolute Shrinkage and Selection Operator (LASSO). (C) Parameter plot of LASSO. (D) Heat map of 31 characteristic ARGs. (E) Boxplot of 31 characteristic ARGs. (F) Correlation analysis of 31 characteristic ARGs.

4.2. Analysis of immune cell infiltration

We then analyzed the proportions of 22 types of immune cell infiltration in the RA patients and normal individuals using the CIBERSORT algorithm. Significant differences were observed in B cells naive, T cells regulatory (Tregs), T cells CD8, T cells CD4 naïve, B cells memory, T cells gamma delta, NK cells resting, Macrophages M2, Dendritic cells resting, and Mast cells resting between the two groups, indicating possible changes in the immune system as a cause of RA (Fig. 3A). Moreover, NK cells resting, Macrophages M0, Neutrophils, B cells naive, and T cells CD4 naive showed significant correlations with most of the ARGs (Fig. 3B), suggesting these ARGs could be crucial in regulating immune infiltration in RA patients.

4.3. Consensus clustering of rheumatoid arthritis patients and immune cell infiltration analysis

To understand the expression patterns related to the ARGs in RA, we performed consensus clustering on 232 RA patients based on the expression profiles of the 31 characteristic ARGs. When k = 2, the clusters were the most stable (Fig. 4A), and the CDF curve ranged between 0.2 and 0.6, the smallest range of shared index (Fig. 4B and C). We grouped the 232 RA patients into two clusters: cluster 1 (n = 111) and cluster 2 (n = 121). In cluster 2, TP53 was significantly upregulated, while FKBP1A, CDKN2A, and GABARAPL1 were downregulated compared to cluster 1. Between the two clusters, the majority of ARGs revealed significant changes (Fig. 4D and E).

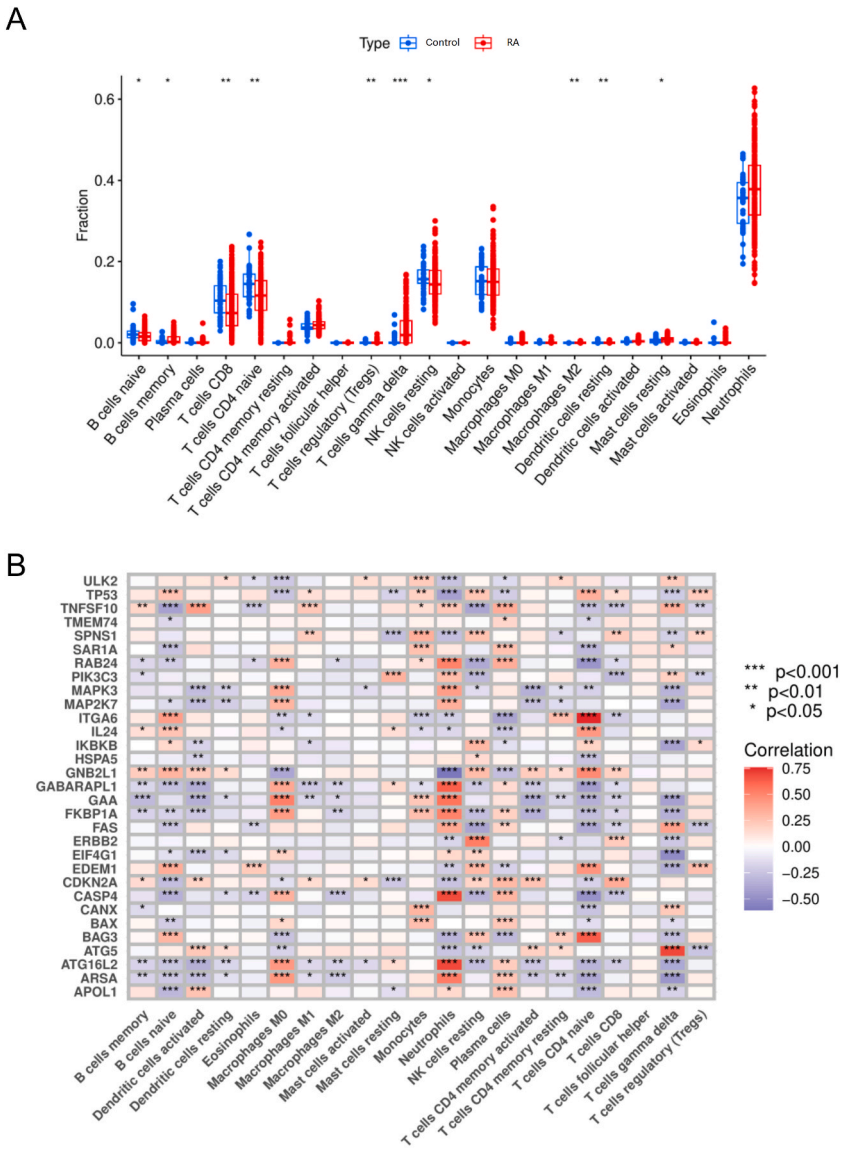


Fig. 3. Analysis of immune infiltration in normal individuals and rheumatoid arthritis patients. (A) Differences in immunity between normal individuals and rheumatoid arthritis patients. (B) Correlation analysis of 31 characteristic ARGs and immune cells.

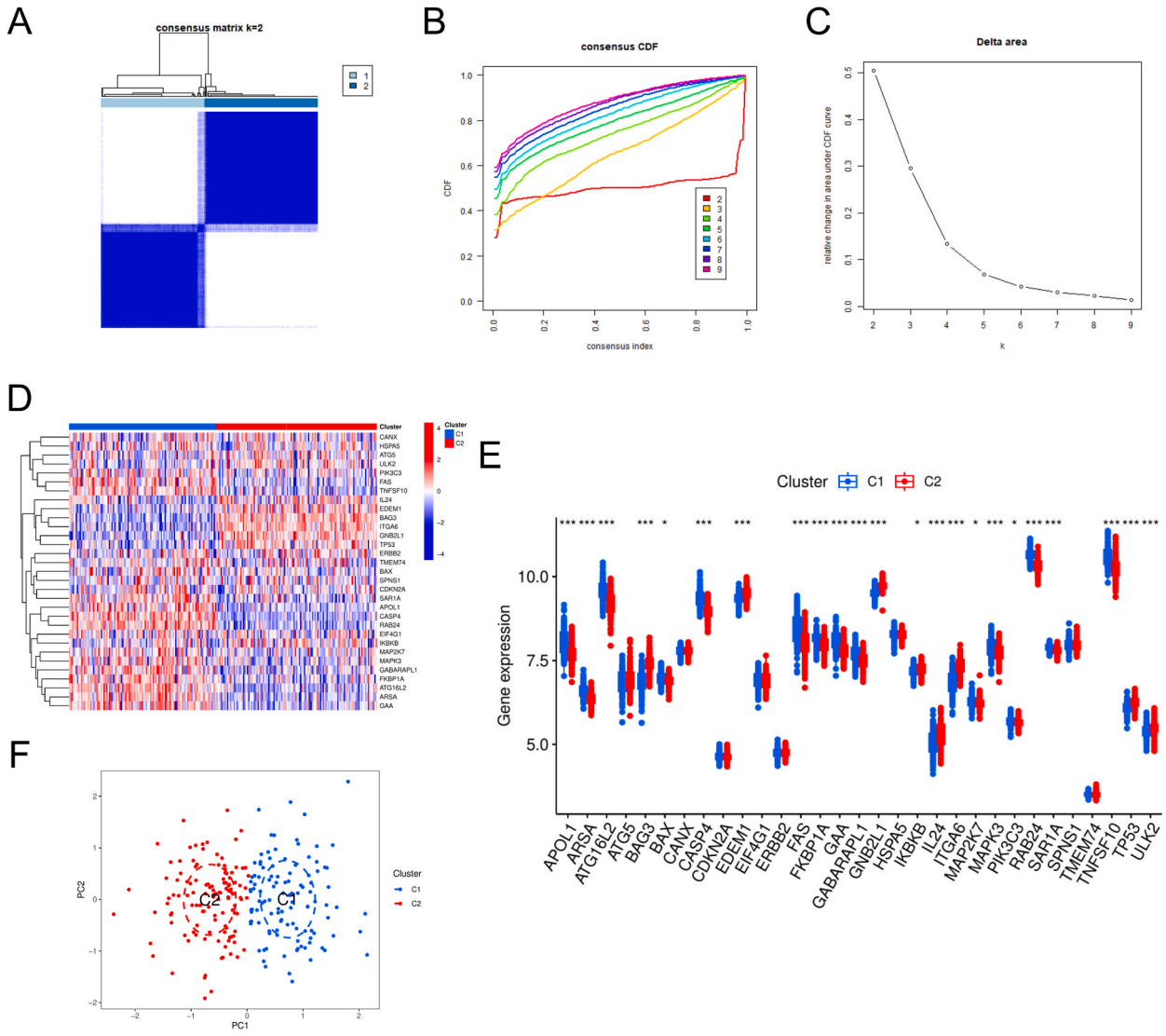


Fig. 4. Consensus clustering analysis. (A) Consensus clustering matrix when k = 2. (B) Representative cumulative distribution function (CDF) curves. (C) CDF delta area curves. (D) Heatmap of 31 characteristic ARGs in cluster 1 and cluster 2. (E) Boxplots of 31 characteristic ARGs in cluster 1 and cluster 2. (F) Principal Component Analysis of cluster 1 and cluster 2.

PCA indicated significant differences between these two clusters (Fig. 4F). Additionally, immune infiltration studies showed that the immunological microenvironment between the two clusters had significantly changed (Fig. 5A). While plasma cells, monocytes, macrophages M0, and neutrophils were more prevalent in cluster 1, cluster 2 contained greater numbers of B cells naive, B cells memory, T cells CD4 naive, T cells CD4 memory resting, and NK cells resting (Fig. 5B).

4.4. Gene Set Enrichment Analysis and Gene Set Variation Analysis

To delve deeper into the functional differences associated with the DEGs in the two autophagy-related clusters, we conducted GSEA and GSVA. In cluster 1, GSEA showed a substantial increase of the tertiary, specialized, and secretory granule membranes (Fig. 6A). Pathway enrichment analysis highlighted the enhancement of immune-related signaling pathways such as Toll-like receptor signaling pathway, Fc gamma R-mediated phagocytosis, and complement and coagulation cascades in cluster 1, suggesting its primary involvement in immune response pathways (Fig. 6B). In cluster 2, GSEA showed a substantial decrease of the tertiary, specialized, and secretory granule membranes (Fig. 6C). Pathway enrichment analysis highlighted the decreased in immune-related signaling pathways such as Toll-like receptor signaling pathway, Fc gamma R-mediated phagocytosis, and complement and coagulation cascades in cluster 2. These results were in contrast to Cluster 1 (Fig. 6D). GSVA results associated cluster 2 with negative regulation of tumor necrosis factor-mediated signaling pathway, interleukin 1 receptor binding, cellular response to oxidized low-density lipoprotein particle

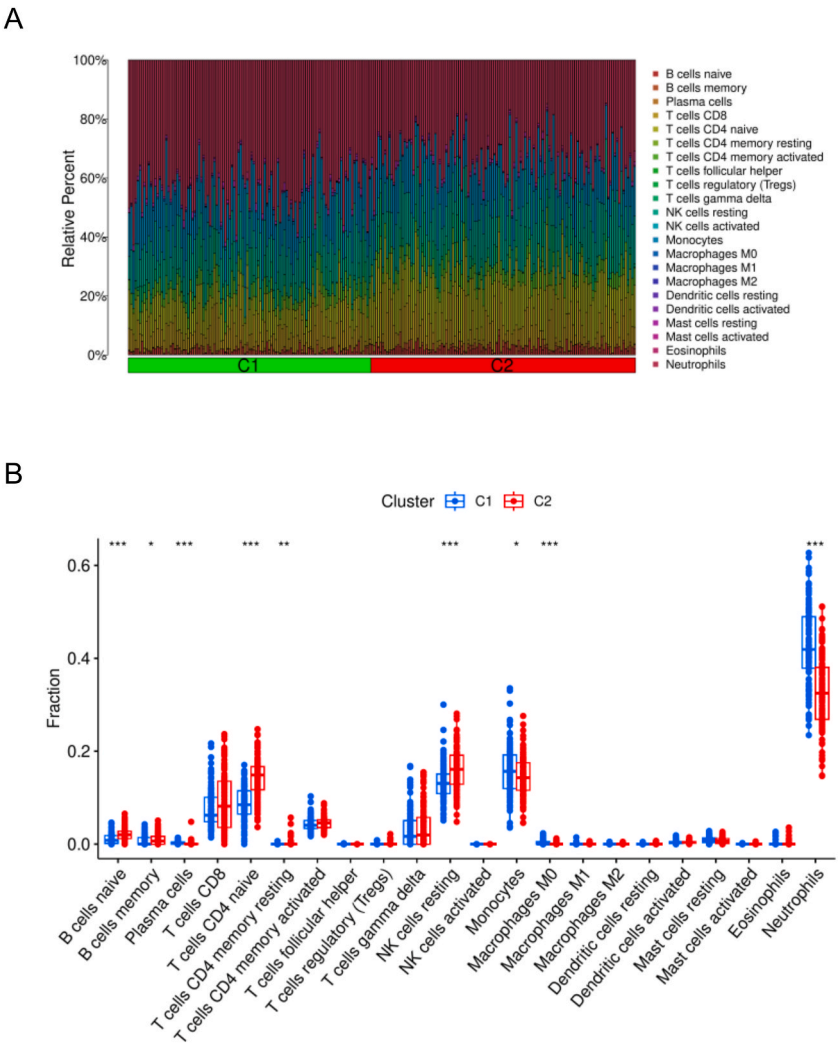


Fig. 5. Immune cell infiltration analysis of cluster 1 and cluster 2. (A) Histogram of immune cells of cluster1 and cluster2. (B) Boxplots of immune cells from cluster 1 and cluster 2.

stimulus, immune response, and response to macrophage colony-stimulating factor. Cluster 1 showed significant correlations with integral component of lysosomal membrane and transcription elongation factor activity, among others (Fig. 6E). Pathway enrichment analysis by GSVA revealed significant enhancement of Fc epsilon RI signaling pathway, SNARE interactions in vesicular transport, Fc gamma R-mediated phagocytosis, RIG-I-like receptor signaling pathway, and other signaling pathways in cluster 2, while Nucleotide excision repair pathways and Aminoacyl-tRNA biosynthesis were significantly enhanced in cluster 1 (Fig. 6F). These findings suggest that both clusters are closely associated with different immune pathways.

4.5. Weighted gene Co-expression network analysis

In order to locate significant gene modules linked to RA, we used the WGCNA algorithm. In the GSE93272 dataset, we measured the variation of gene expression and chose the top 25 % of the most variable genes for further study. We discovered co-expression gene modules by setting the threshold to 9 and the scale-free R^2 to 0.9 (Fig. 7A). We generated 11 unique co-expression gene modules using the dynamic cutting method, which we depicted by different colors on a heatmap of the TOM (Fig. 7B–D). We then analyzed the co-expression similarity and adjacency of module-clinical characteristics using these genes from the 11 color modules. The brown module, which included 646 genes, exhibited the strongest correlation with RA (Fig. 7E). A positive correlation was also observed between the module-related genes and brown module (Fig. 7F).

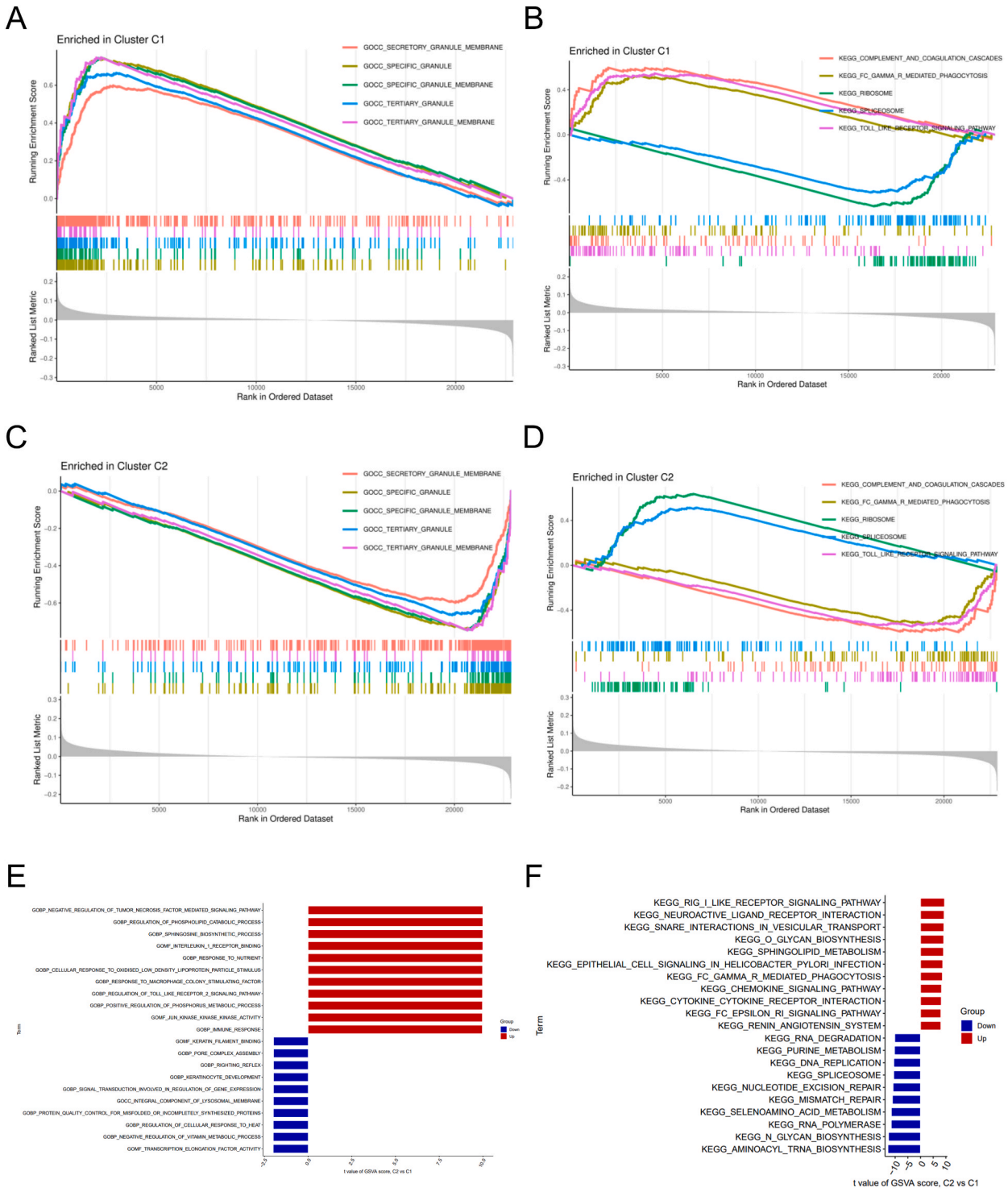


Fig. 6. Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA). (A) GSEA functional enrichment analysis in cluster 1. (B) GSEA pathway enrichment analysis in cluster 1. (C) GSEA functional enrichment analysis in cluster 2. (D) GSEA pathway enrichment analysis in cluster 2. (E) Functional enrichment analysis of GSVA. (F) Pathway enrichment analysis of GSVA.

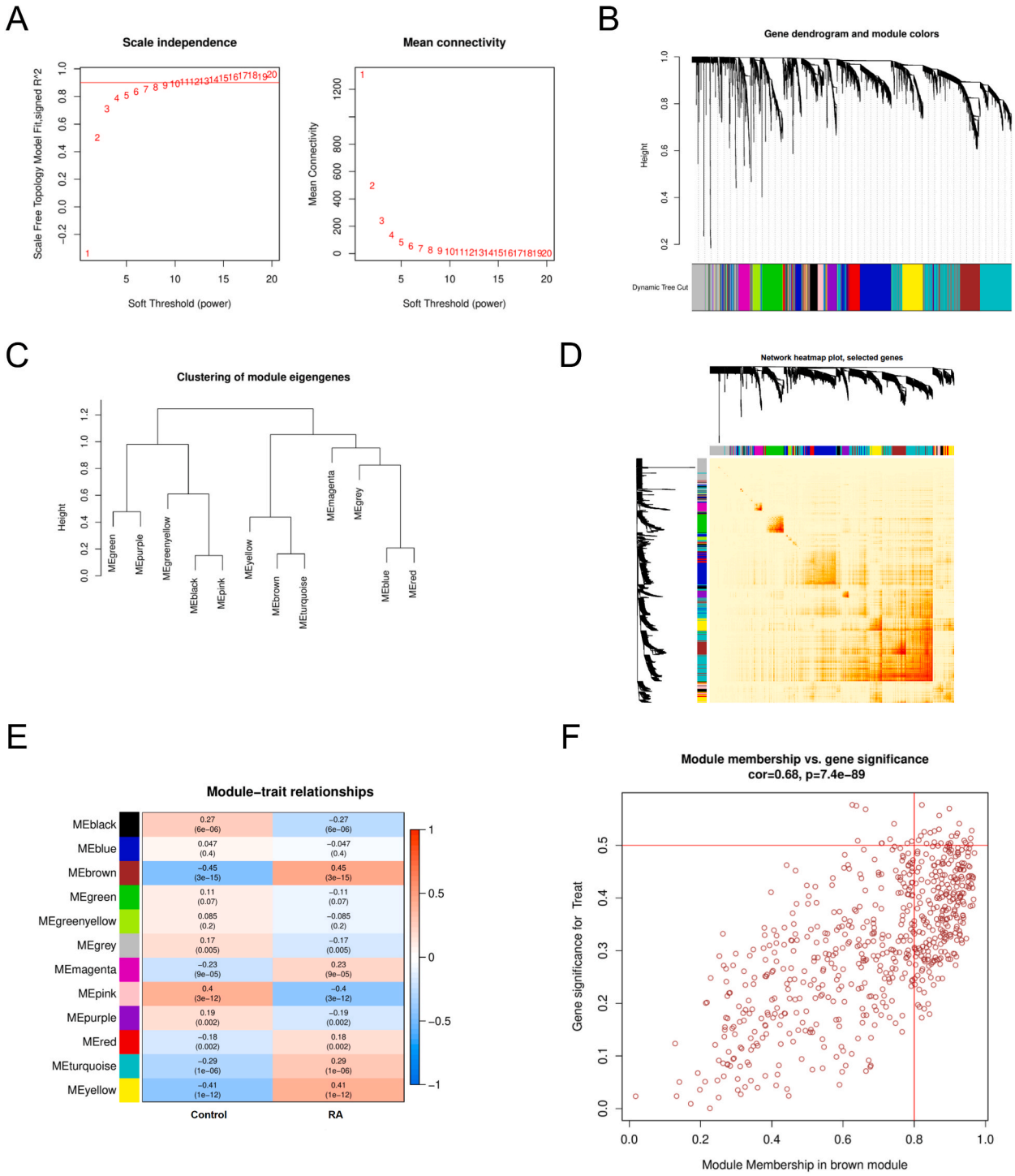


Fig. 7. Co-expression network of differentially expressed genes between the two clusters. (A) The selection of soft threshold power. (B) Cluster tree dendrogram of co-expression modules. Different colors represent distinct co-expression modules. (C) Representative of clustering of module eigengenes. (D) Representative heatmap of the correlations among 11 modules. (E) Correlation analysis between module and clinical status. Each row represents a module; each column represents a clinical status. (F) Scatter plot between module membership in brown module and the gene significance for cluster 2. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

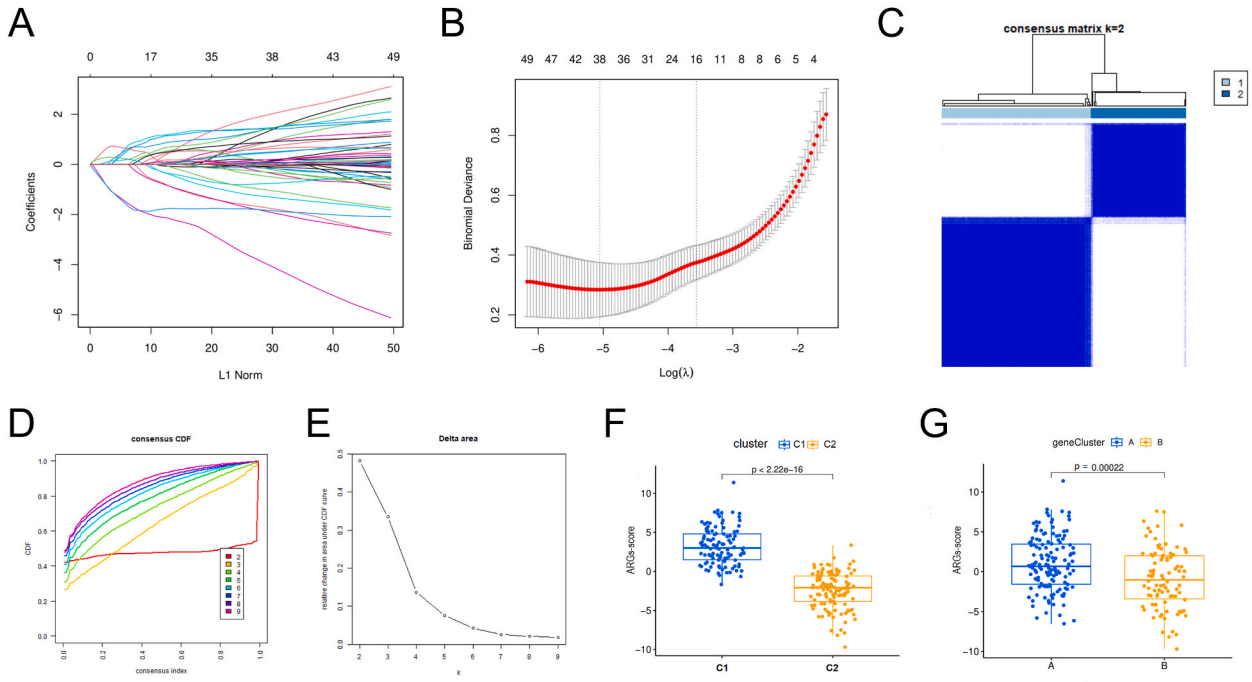


Fig. 8. The brown module characterizes differentially expressed genes (DEGs) filtering and consensus clustering analysis. (A) Coefficient diagram of LASSO for brown module DEGs. (B) Parametric diagram of LASSO for brown module DEGs. (C) Consensus clustering matrix when k = 2. (D) Representative CDF curves. (E) CDF delta area curves. (F) Differences in ARGs-score between cluster 1 and cluster 2. (G) Differences in ARGs-score between gene cluster A and gene cluster B. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4.6. Construction of Least Absolute Shrinkage and Selection Operator Model and Consensus Clustering Analysis

To explore the relationship between ARGs clusters (Cluster 1 and Cluster 2) and WGCNA, we used the 646 genes from the brown module to construct a LASSO model. This model helped us select 38 characteristic genes from the brown module (Fig. 8A and B). We then performed consensus clustering analysis on 232 RA patients based on the expression profiles of these 38 characteristic genes. When the value of k was set to 2 (k = 2), the number of clusters remained the most steady (Fig. 8C), while the CDF curve fluctuated within the shared index's lowest range of 0.2–0.6 (Fig. 8D and E). We divided the 232 RA patients into two clusters based on these 38 characteristic genes: Gene cluster A (A = 142) and Gene cluster B (B = 90). Using a PCA technique, we determined the ARGs-score for each sample based on the expression levels of the characteristic ARGs. In comparison to cluster 2 or gene cluster B, cluster 1 or gene cluster A had higher scores (Fig. 8F and G). Cluster 1 primarily corresponded to gene cluster A, exhibiting high ARGs-score, while cluster 2 mainly corresponded to gene cluster B, exhibiting low ARGs-score.

4.7. Construction and selection of four machine learning models

Based on the expression profiles of the 31 characteristic ARGs, we constructed XGB, SVM, GLM and RF machine learning models to find characteristic ARGs with high diagnostic value. The GLM model exhibited the lowest residual values (Fig. 9A and B). We computed the ROC curve based on 10-fold cross-validation to assess the discriminatory performance of the four machine learning algorithms in the test set. Among the models (GLM, AUC = 1.000; XGB, AUC = 0.948; SVM, AUC = 0.998; RF, AUC = 0.967) (Fig. 9C), the GLM model had the largest area under the curve (AUC). Based on root mean square error, the importance of each variable's attributes was ranked (Fig. 9D). The GLM model was found to best distinguish RA patients. The top five most important variables (CDKN2A, TP53, ATG16L2, FKBP1A, and GABARAPL1) selected by the GLM model were used to construct the nomogram model.

4.8. Construction and validation of nomogram model

We constructed a column chart based on the five ARGs (CDKN2A, TP53, ATG16L2, FKBP1A, and GABARAPL1) to evaluate the predictive efficiency of the GLM model (Fig. 10A). To evaluate the nomogram model's capability for prediction, calibration curves and DCA were employed. The calibration curve demonstrated that there were few discrepancies between the real and predicted risks of RA (Fig. 10B), and DCA results suggested that our nomogram had excellent accuracy and would be useful for clinical patients (Fig. 10C). We validated the nomogram prediction model using five external datasets: GSE17755, GSE56649, GSE68689, GSE77298, and

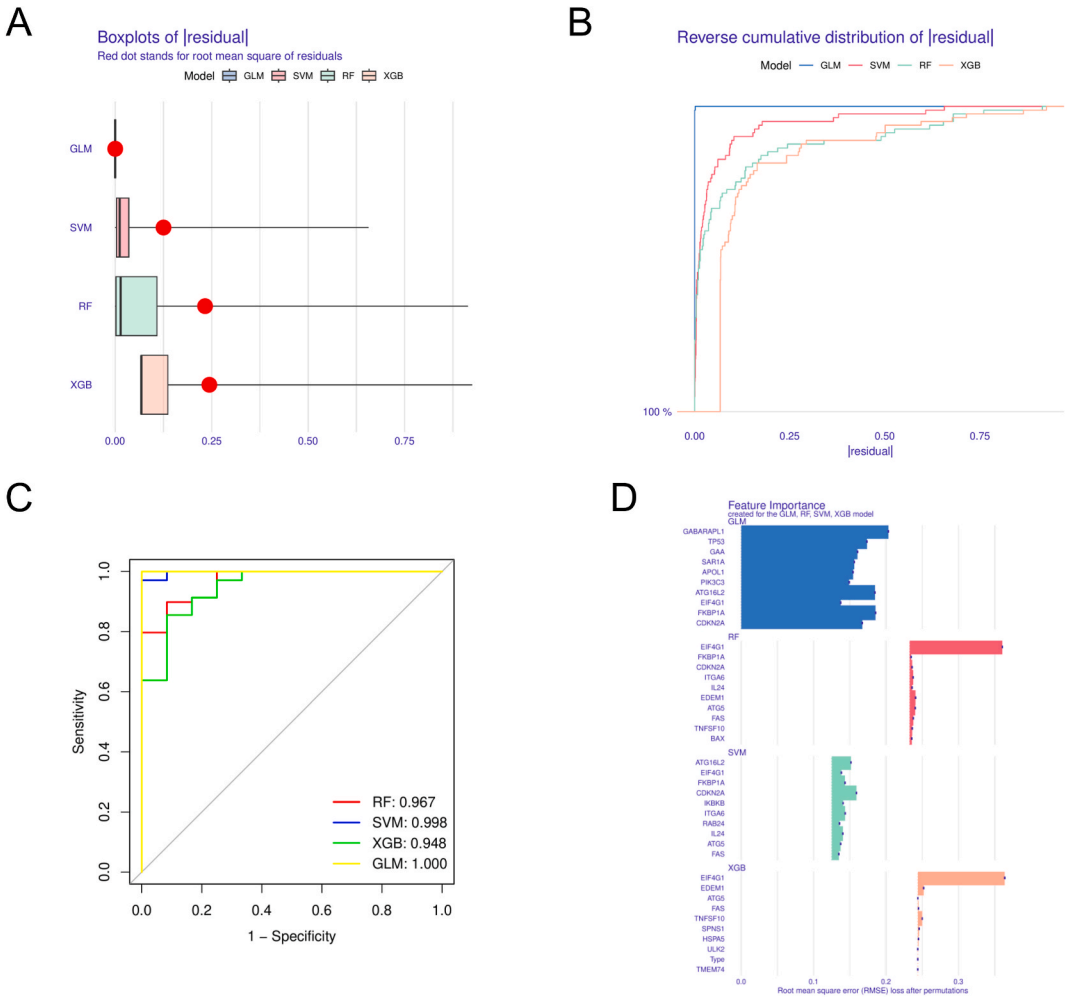


Fig. 9. Construction and evaluation of GLM, SVM, RF and XGB machine models. (A) Boxplots showed the residuals of each machine learning model. Red dot represented the root mean square of residuals. (B) Cumulative residual distribution of each machine learning model. (C) Receiver Operating Characteristic (ROC) analysis of four machine learning models based on 10-fold cross-validation in the testing cohort. (D) The important features in GLM, SVM, RF and XGB machine models. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

GSE71370. The ROC curves showed AUC values of 0.911 for GSE17755 (Fig. 11A), 0.915 for GSE56649 (Figs. 11B), 0.850 for GSE68689 (Figs. 11C), 0.893 for GSE77298 (Figs. 11D), and 0.910 for GSE71370 (Fig. 11E). The nomogram prediction model based on the five ARGs exhibited AUC values greater than 0.850 in all five external independent datasets, indicating excellent performance.

5. Discussion

RA is a systemic inflammatory autoimmune disease that affects not only the joints but also the heart, blood vessels, kidneys, lungs, and nervous system, impacting both physical and mental health. Early diagnosis and management are crucial to prevent joint damage caused by RA [37,38]. However, current RA treatments lack sufficient efficacy, underscoring the need for more suitable clusters to guide personalized RA therapy. Numerous studies have identified autophagy as a crucial factor in RA, but the role of autophagy in RA from a bioinformatics perspective remains underexplored [39]. Therefore, our study aims to elucidate the specific role of ARGs in RA clusters and the immune microenvironment. Further research on ARGs may reveal the complex pathogenesis of RA and provide new directions for treatment.

In our study, we conducted a comprehensive analysis of ARGs expression profiles in normal individuals and RA patients. We found significantly higher abnormal expression of ARGs in RA patients, suggesting the potential role of ARGs in RA onset. We used the LASSO to select characteristic ARGs and calculated the correlation between these ARGs to understand their association with RA. Some characteristic ARGs exhibited significant synergistic or antagonistic effects. Our subsequent analysis of immune cell infiltration revealed notable differences in immune cell abundance between normal individuals and RA patients. Furthermore, we used consensus

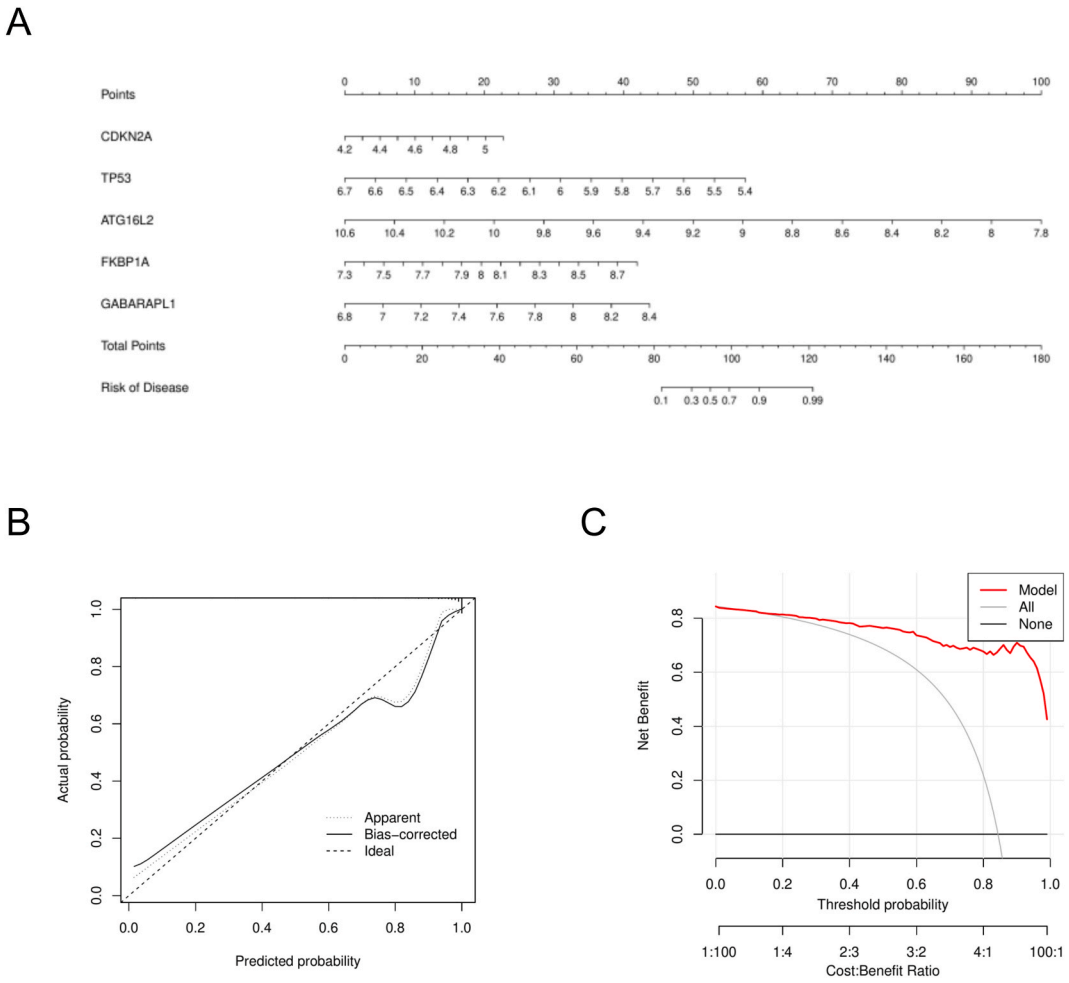


Fig. 10. Construction of the nomogram model. (A) Construction of a nomogram. (B and C) Construction of calibration curve (B) and DCA (C) for assessing the predictive efficiency of the nomogram model.

clustering analysis based on the expression profiles of characteristic ARGs in RA patients to illustrate different autophagy regulation patterns. We identified two distinct ARGs clusters (Cluster 1 and Cluster 2) with significantly different levels of immune infiltration. Through the WGCNA algorithm, we identified a brown module containing 646 genes strongly associated with RA. Using the LASSO model and consensus clustering algorithm, we selected 38 characteristic genes from the brown module and classified the 232 RA patients into two clusters (Gene cluster A and B), which showed significant differences in ARGs expression.

With reduced error rates and more dependable outcomes when compared to univariate analysis, various machine learning models have been used more frequently in recent years to predict the incidence of various diseases. In our study, we compared four machine learning models based on 31 characteristic ARGs (XGB, SVM, GLM, and RF). The GLM model exhibited the best diagnostic capability for RA with the highest AUC (AUC = 1.000). Finally, we constructed a nomogram model using five characteristic ARGs (CDKN2A, TP53, ATG16L2, FKBP1A, and GABARAPL1), which demonstrated excellent predictive efficiency and accuracy. This model was validated in five external independent datasets (GSE17755, GSE56649, GSE68689, GSE77298, and GSE71370) with AUC values greater than 0.850.

Cyclin Dependent Kinase Inhibitor 2A (CDKN2A) is a protein-encoding gene that can induce cell cycle arrest in the G1 and G2 phases, acting as a tumor suppressor [40,41]. Recent studies have linked the CDKN2A gene with poor prognosis in a variety of cancers such as pancreatic cancer, meningioma, and melanoma. CDKN2A encodes the P16 gene, which is involved in a series of cellular pathways, including promoting tumor cell proliferation, inhibiting tumor cell apoptosis, inducing tumor stromal angiogenesis, and reducing cancer cell sensitivity to chemotherapy [42–44]. By closely interacting with CDK4 and CDK6, which inhibits their capacity to bind with cyclin D and phosphorylate the retinoblastoma protein, CDKN2A functions as a negative regulator of normal cell growth [45]. Furthermore, CDKN2A plays a crucial role in controlling cell proliferation by inhibiting the activity of CDK4 and CDK6, which are essential for cell cycle progression. Regarding immune cells, this regulatory mechanism can influence the proliferation rates of lymphocytes or other types of immune cells, particularly under conditions of immune activation or during immune responses. Aberrations in CDKN2A expression or function could lead to dysregulation of immune cell proliferation, thereby impacting immune

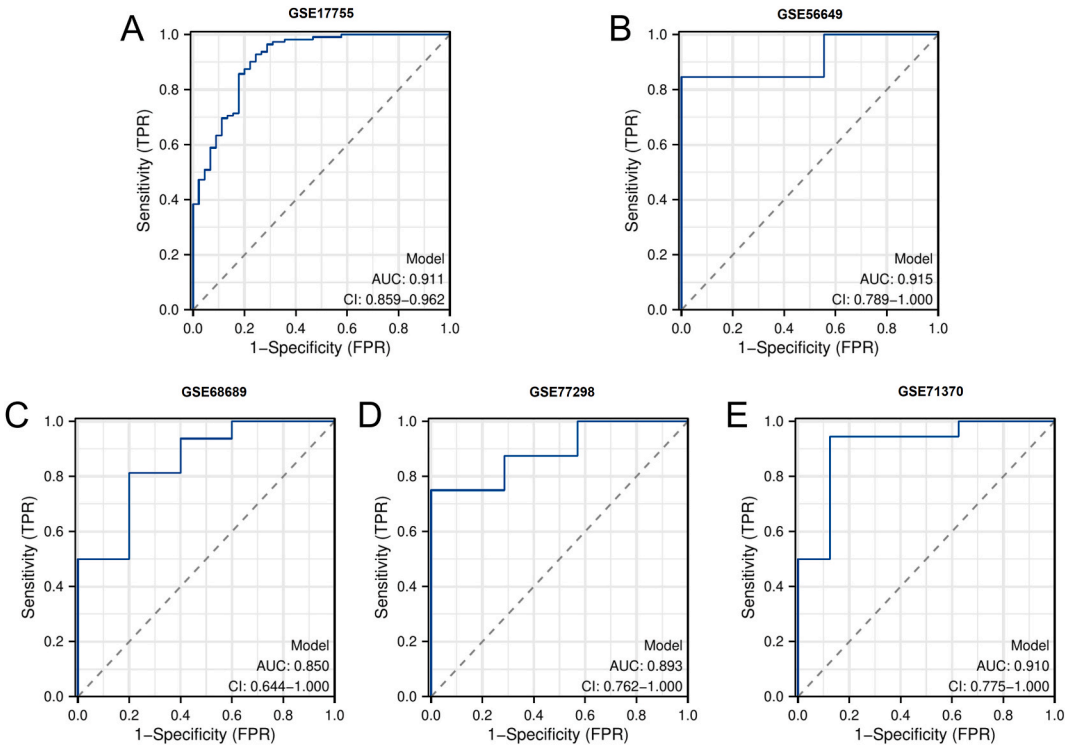


Fig. 11. External independent dataset validation. (A) Area under the ROC (AUC) of GSE17755 (AUC = 0.911). (B) AUC of GSE56649 (AUC = 0.915). (C) AUC of GSE68689 (AUC = 0.850). (D) AUC of GSE77298 (AUC = 0.893). (E) AUC of GSE71370 (AUC = 0.910).

responses [46]. While the primary role of CDKN2A is to inhibit the cell cycle, it also impacts cell survival and apoptosis. For instance, one of the proteins encoded by the CDKN2A locus, p16INK4a, can initiate cellular senescence, a stable form of cell cycle arrest that is associated with the promotion of apoptosis under certain conditions. In immune cells, dysregulation of this pathway could affect cell survival, potentially leading to premature death of immune cells [47].

Tumor protein P53 (TP53), a protein-encoding gene, serves as a tumor suppressor in various cancer types. Depending on the physiological conditions and cell type, it can either cause growth arrest or apoptosis [48,49]. Specifically, P53 inhibits **Epithelial-Mesenchymal Transition** (EMT) transcription factors, such as Zeb1, Snail1, and Twist2, by activating miR-200 and miR-34. This action suppresses the self-renewal of **Cancer Stem Cells** (CSCs) [50]. Additionally, research indicates that autophagy can enhance the stemness of lung CSCs by degrading ubiquitinated P53, highlighting autophagy's dual role in promoting and inhibiting tumor development in lung cancer [51]. High-grade diffuse intrinsic pontine gliomas and non-brainstem gliomas frequently have mutations in genes like TP53. These mutations affect the control of the cell cycle, histone modification or chromatin remodeling, and the receptor tyrosine kinase-RAS-PI3K signaling pathway [52]. In cancer, TP53 mutations lead to multiple changes. For instance, PDGF receptor-mediated autonomous TP53 protein signaling promotes pancreatic cancer metastasis. It can also increase genetic instability, thereby promoting cancer development [53]. Different mutant forms of TP53 protein can regulate the Ras signaling pathway in various ways, inducing the expression of cancer-related genes [54]. Furthermore, TP53 loss can promote the invasion of multiple myeloma cells by upregulating miR19a/CXCR5 [55]. Studies have shown that TP53 plays a key role in regulating cell proliferation. In immune cells, TP53 helps maintain appropriate proliferation rates, preventing excessive cell division to avert autoimmune or other immune response imbalances. By inducing cell cycle arrest under conditions of stress or DNA damage, TP53 ensures that only healthy, intact immune cells are able to proliferate [56].

Autophagy Related 16 Like 2 (ATG16L2) is another protein-encoding gene. It can be regulated by microRNA Mir885-3p, which binds to ATG16L2, inhibiting its translation, reducing ATG16L2 levels, and inducing autophagy [57,58]. Research showed that ATG16L2 participates in autophagy by forming a complex with ATG5, contributing to autophagosome formation and the degradation and recycling of cellular components [59]. Moreover, studies have found that ATG16L2 expression is lower in non-small cell lung cancer, and functional genetic variations in ATG16L2 may affect its mRNA expression level. The expression level of ATG16L2 is also associated with resistance to platinum-based chemotherapy, suggesting that ATG16L2 may serve as a biomarker or potential therapeutic target for tumor cell resistance to platinum-based drugs [60].

A member of the immunophilin family, FKBP Prolyl Isomerase 1A (FKBP1A) encodes a protein that is involved in basic cellular activities such as protein folding and transport as well as immunological modulation [61]. Studies have found that upregulated FKBP1A can inhibit the growth of glioblastoma through the apoptosis pathway [62]. Recently, it has also been shown that FKBP1A overexpression can rescue the enhancement of sensitivity to paclitaxel in paclitaxel-resistant Prostate Adenocarcinoma (PCa) cells

mediated by miR-195-5p [63]. According to a mechanism investigation, SNHG15 was found in the cytoplasm of PCa cells and functioned as a molecular sponge for miR-338-3p. Additionally, miR-338-3p has FKBP1A as a target. Therefore, the LncRNA SNHG15 controls the miR-338-3p/FKBP1A axis, acting as an oncogene in prostate cancer [64].

The development of autophagosomal vesicles is aided by GABA Type A Receptor Associated Protein Like 1 (GABARAPL1) [65]. Studies have found that low expression of GABARAPL1 leads to an increased proliferation rate in the lung adenocarcinoma cell line A549. By interacting with SMAD2/3 proteins, GABARAPL1 can block the **EMT** signaling pathway that is activated by transforming growth factor-beta (TGF-B)/tumor necrosis factor-alpha (TNF-A). GABARAPL1 regulates autophagy to degrade SMAD and GABARAPL1 proteins, further inhibiting the EMT signaling pathway [66]. Furthermore, knockdown of GABARAPL1 inhibits the growth of androgen receptor (AR)-positive prostate cancer cells LNCaP and CWR22rv1, as well as reduces the transcriptional activity and nuclear translocation of AR/**Androgen Receptor Variants** (ARV). This suggests that GABARAPL1 inhibits the growth of testosterone receptor-positive prostate cancer by reducing the transcriptional activity and nuclear translocation of AR/ARVN [67]. Recent research has also revealed the important role of GABARAPL1 in breast cancer, controlling cell proliferation, invasion, and resistance to cell death by regulating autophagy pathways, maintaining mitochondrial homeostasis, and modulating cellular metabolic programs. Loss of GABARAPL1 leads to disrupted autophagy pathways, abnormal mitochondrial function, cellular metabolic dysregulation, and tumor cell deterioration [68]. While there are no extensive records of the direct impact of GABARAPL1 on immune cell migration, its role in autophagy may be indirectly related to this process. Autophagy can regulate the turnover of cellular components and provide substrates for energy production, which could affect the dynamics of the actin cytoskeleton and cell adhesion, both crucial for cell migration. Thus, GABARAPL1, through its role in autophagy, may influence the cellular energy and cytoskeletal rearrangements necessary for immune cell migration [69]. Studies have shown that GABARAPL1, by modulating the autophagic response, may indirectly influence the threshold for immune cell apoptosis. For instance, the enhancement of autophagic activity mediated by GABARAPL1 can promote cell survival under stress, thereby reducing the propensity for cell apoptosis [70,71]. Although these five characteristic ARGs are closely related to the occurrence and development of various diseases, their role in the mechanism of RA has not been extensively studied.

5.1. Limitations

Despite the excellent diagnostic predictive ability for RA patients shown by the nomogram model we constructed based on the five characteristic ARGs, there are limitations to this study. Firstly, all RA patients used in this study were obtained from public databases. Therefore, we plan to expand the recruitment of clinical patients in the future to further validate the clinical application value of our nomogram model in diagnosing RA patients. In addition, comprehensive functional experiments are essential to gain a deeper understanding of the complex roles of the five characteristic ARGs in the pathogenesis of RA.

6. Conclusion

In this study, we conducted an extensive analysis of ARG expression profiles in both normal individuals and RA patients. Our research showed that RA patients had considerably higher levels of aberrant ARGs expression than normal individuals did. Using the LASSO algorithm, we selected characteristic ARGs and identified significant synergistic or antagonistic effects in RA. We also uncovered significant changes in immune cell infiltration and elucidated different autophagy regulation patterns through consensus clustering analysis. Furthermore, we constructed an optimal machine learning model, GLM, based on 31 characteristic ARGs. In addition, we developed a nomogram model based on five characteristic ARGs (CDKN2A, TP53, ATG16L2, FKBP1A, and GABARAPL1). The predictive efficiency and accuracy of this model were validated in five external independent datasets, with AUC values exceeding 0.850. This study provides a fresh perspective on the role of ARGs in RA and offers new directions for personalized RA treatment. We anticipate that this ARGs-based model will offer valuable insights for further RA research and assist clinicians in formulating more personalized and precise treatment strategies.

Funding

None.

Data availability statement

The datasets generated and/or analyzed during the current study are available in the GEO repository (GSE93272, GSE17755, GSE56649, GSE68689, GSE77298, and GSE71370) (https://www.ncbi.nlm.nih.gov/geo/).

Ethics approval and consent to participate

This study does not involve ethical approval.

Consent for publication

Not applicable.

CRediT authorship contribution statement

Xin Li: Writing – review & editing, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Shuang Ding:** Writing – review & editing, Visualization, Validation, Methodology, Investigation. **Pengcheng Zhang:** Visualization, Validation, Conceptualization. **Jing Yan:** Project administration, Methodology, Investigation. **Xingxing Yu:** Supervision, Software. **Xukai Wang:** Writing – review & editing, Writing – original draft, Validation, Data curation, Conceptualization. **Hongsheng Zhan:** Writing – original draft, Visualization. **Zhengyan Wang:** Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the researchers who provided open access to the raw data.

References

[1] H.U. Scherer, T. Haupl, G.R. Burmester, The etiology of rheumatoid arthritis, Epub 2020/01/26, J. Autoimmun. 110 (2020) 102400, https://doi.org/10.1016/j.jaut.2019.102400. PubMed PMID: 31980337.
[2] J.A. Sparks, Rheumatoid arthritis, Epub 2019/01/01, Ann. Intern. Med. 170 (1) (2019) ITC1–ITC16, https://doi.org/10.7326/AITC201901010. PubMed PMID: 30596879.
[3] A.F. Radu, S.G. Bungau, Management of rheumatoid arthritis: an overview, Epub 2021/11/28, Cells 10 (11) (2021), https://doi.org/10.3390/cells10112857. PubMed PMID: 34831081; PubMed Central PMCID: PMCPMC8616326.
[4] A.L. Mueller, Z. Payandeh, N. Mohammadkhani, S.M.H. Mubarak, A. Zakeri, A. Alagheband Bahrami, et al., Recent advances in understanding the pathogenesis of rheumatoid arthritis: new treatment strategies, Epub 2021/11/28, Cells 10 (11) (2021), https://doi.org/10.3390/cells10113017. PubMed PMID: 34831240; PubMed Central PMCID: PMCPMC8616543.
[5] J.J. Cush, Rheumatoid arthritis: early diagnosis and treatment, Epub 2021/02/17, Med Clin North Am 105 (2) (2021) 355–365, https://doi.org/10.1016/j.mcna.2020.10.006. PubMed PMID: 33589108.
[6] G.R. Burmester, J.E. Pope, Novel treatment strategies in rheumatoid arthritis, Epub 2017/06/15, Lancet (London, England) 389 (10086) (2017) 2338–2348, https://doi.org/10.1016/s0140-6736(17)31491-5. PubMed PMID: 28612748.
[7] J. Zhao, S. Guo, S.J. Schrodi, D. He, Molecular and cellular heterogeneity in rheumatoid arthritis: mechanisms and clinical implications, Epub 2021/12/14, Front. Immunol. 12 (2021) 790122, https://doi.org/10.3389/fimmu.2021.790122. PubMed PMID: 34899757; PubMed Central PMCID: PMCPMC8660630.
[8] T. Ichimiya, T. Yamakawa, T. Hirano, Y. Yokoyama, Y. Hayashi, D. Hirayama, et al., Autophagy and autophagy-related diseases: a review, Epub 2020/12/02, Int. J. Mol. Sci. 21 (23) (2020), https://doi.org/10.3390/ijms21238974. PubMed PMID: 33255983; PubMed Central PMCID: PMCPMC7729615.
[9] D.J. Klionsky, G. Petroni, R.K. Amaravadi, E.H. Baehrecke, A. Ballabio, P. Boya, et al., Autophagy in major human diseases, Epub 2021/08/31, The EMBO journal 40 (19) (2021) e108863, https://doi.org/10.15252/embj.2021108863. PubMed PMID: 34459017; PubMed Central PMCID: PMCPMC8488577.
[10] N.Y. Lin, C. Beyer, A. Giessl, T. Kireva, C. Scholtysek, S. Uderhardt, et al., Autophagy regulates TNFα-mediated joint destruction in experimental arthritis, Epub 2012/09/15, Ann. Rheum. Dis. 72 (5) (2013) 761–768, https://doi.org/10.1136/annrheumdis-2012-201671. PubMed PMID: 22975756.
[11] G. Wang, X. Chen, Y. Shao, B. Xu, PINK1/Parkin-Mediated mitochondrial autophagy participates in H(2)O(2)-induced abnormal proliferation of fibroblast-like synoviocytes in rheumatoid arthritis, Epub 2023/03/31, J. Inflamm. Res. 16 (2023) 1271–1282, https://doi.org/10.2147/JIR.S398690. PubMed PMID: 36993991; PubMed Central PMCID: PMCPMC10042253.
[12] H. Zhou, L. Huang, K. Zhan, X. Liu, Wenhua Juanbi Recipe attenuates rheumatoid arthritis via inhibiting miRNA-146a-mediated autophagy, Epub 2022/11/29, BioMed Res. Int. 2022 (2022) 1768052, https://doi.org/10.1155/2022/1768052. PubMed PMID: 36440364; PubMed Central PMCID: PMCPMC9683957.
[13] S. Tasaki, K. Suzuki, Y. Kassai, M. Takeshita, A. Murota, Y. Kondo, et al., Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission, Epub 2018/07/18, Nat. Commun. 9 (1) (2018) 2755, https://doi.org/10.1038/s41467-018-05044-4. PubMed PMID: 30013029; PubMed Central PMCID: PMCPMC6048065 Yo.K., T.A., Y.O., Mai.T. and R.K. are employed by Takeda Pharmaceutical Company Limited. Y.N. is employed by ONO Pharmaceutical. T.M. is employed by Nektar Therapeutics. H.T. is employed by FRONTEO. K.S. has received research grants from Eisai, Bristol-Myers Squibb, Kissei Pharmaceutical, and Daiichi Sankyo, and speaking fees from Abbie Japan, Astellas Pharma, Bristol-Myers Squibb, Chugai Pharmaceutical, Eisai, Fuji Film Limited, Janssen Pharmaceutical, Kissei Pharmaceutical, Mitsubishi Tanabe Pharmaceutical, Pfizer Japan, Shionogi, Takeda Pharmaceutical, and UCB Japan, consulting fees from Abbie, and Pfizer Japan. A.Y. has received speaking fees from Chugai Pharmaceutical, Mitsubishi Tanabe Pharmaceutical, Pfizer Japan, Ono Pharmaceutical, Maruho, and Novartis, and consulting fees from GSK Japan. Ku.Y. has received consultant fees from Pfizer, Chugai Pharma, Mitsubishi Tanabe Pharma, Abbvie, received honoraria from Pfizer, Chugai Pharma, Mitsubishi Tanabe Pharma, Bristol-Myers Squibb, Takeda Industrial PharmaGlaxoSmithkline, Nippon Shinyaku, Eli lilly, Janssen Pharma, Eisai Pharma, Astellas Pharma, Actelion Pharmaceuticals and received research grants from Chugai Pharma, Mitsubishi Tanabe Pharma., and Glaxo Smith Kline. H.Y. has received research grants from Daiichi Sankyo, Takeda Pharmaceutical, Eisai, and Japan Blood Products Organization, and speaking fees from Abbie Japan, Bristol-Myers Squibb, Takeda Pharmaceutical, Chugai Pharmaceutical, and Eisai. T.T. has received research grants from Astellas Pharma Inc, Bristol-Myers K.K., Chugai Pharmaceutical Co. Ltd., Daiichi Sankyo Co. Ltd., Takeda Pharmaceutical Co. Ltd., Teijin Pharma Ltd., AbbVie GK, Asahikasei Pharma Corp., Mitsubishi Tanabe Pharma Co., Pfizer Japan Inc., and Taisho Toyama Pharmaceutical Co. Ltd., Eisai Co. Ltd., AYUMI Pharmaceutical Corporation, and Nipponkayaku Co. Ltd, and speaking fees from AbbVie GK., Bristol-Myers K.K., Chugai Pharmaceutical Co. Ltd., Mitsubishi Tanabe Pharma Co., Pfizer Japan Inc., and Astellas Pharma Inc., and Diaichi Sankyo Co. Ltd., and consultant fees from Astra Zeneca K.K., Eli Lilly Japan K.K., Novartis Pharma K.K., Mitsubishi Tanabe Pharma Co., Abbivie GK, Nipponkayaku Co. Ltd, Janssen Pharmaceutical K.K., Astellas Pharma Inc., and Taiho Pharmaceutical Co. Ltd. The remaining authors declare no competing interests..
[14] H.M. Lee, H. Sugino, C. Aoki, N. Nishimoto, Underexpression of mitochondrial-DNA encoded ATP synthesis-related genes and DNA repair genes in systemic lupus erythematosus, Epub 2011/04/19, Arthritis Res. Ther. 13 (2) (2011) R63, https://doi.org/10.1186/ar3317. PubMed PMID: 21496236; PubMed Central PMCID: PMCPMC3132058.
[15] H. Ye, J. Zhang, J. Wang, Y. Gao, Y. Du, C. Li, et al., CD4 T-cell transcriptome analysis reveals aberrant regulation of STAT3 and Wnt signaling pathways in rheumatoid arthritis: evidence from a case-control study, Epub 2015/04/17, Arthritis Res. Ther. 17 (1) (2015) 76, https://doi.org/10.1186/s13075-015-0590-9. PubMed PMID: 25880754; PubMed Central PMCID: PMCPMC4392874.
[16] M.G. Broeren, M. de Vries, M.B. Bennink, O.J. Arntz, A.B. Blom, M.I. Koenders, et al., Disease-regulated gene therapy with anti-inflammatory interleukin-10 under the control of the CXCL10 promoter for the treatment of rheumatoid arthritis, Epub 2015/12/30, Hum. Gene Ther. 27 (3) (2016) 244–254, https://doi.org/10.1089/hum.2015.127. PubMed PMID: 26711533.

[17] M. Rajasekhar, A.M. Olsson, K.J. Steel, M. Georgouli, U. Ranasinghe, Read C. Brender, et al., MicroRNA-155 contributes to enhanced resistance to apoptosis in monocytes from patients with rheumatoid arthritis, Epub 2017/01/26, J. Autoimmun. 79 (2017) 53–62, https://doi.org/10.1016/j.jaut.2017.01.002. PubMed PMID: 28118944; PubMed Central PMCID: PMCPMC5397583.

[18] E. Clough, T. Barrett, The gene expression Omnibus database, Epub 2016/03/24, Methods Mol. Biol. 1418 (2016) 93–110, https://doi.org/10.1007/978-1-4939-3578-9_5. PubMed PMID: 27008011; PubMed Central PMCID: PMCPMC4944384.

[19] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Epub 2015/01/22, Nucleic acids research 43 (7) (2015) e47, https://doi.org/10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMCPMC4402510.

[20] T.S. Wei, Viliam, R Package "corrplot": Visualization of a Correlation Matrix, 2017. Available from, https://github.com/taiyun/corrplot.

[21] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, et al., Robust enumeration of cell subsets from tissue expression profiles, Epub 2015/03/31, Nat. Methods 12 (5) (2015) 453–457, https://doi.org/10.1038/nmeth.3337. PubMed PMID: 25822800; PubMed Central PMCID: PMCPMC4739640.

[22] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Epub 2010/04/30, Bioinformatics 26 (12) (2010) 1572–1573, https://doi.org/10.1093/bioinformatics/btq170. PubMed PMID: 20427518; PubMed Central PMCID: PMCPMC2881355.

[23] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles 102 (43) (2005) 15545–15550, https://doi.org/10.1073/pnas.0506580102.

[24] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J.P. Mesirov, Molecular signatures database (MSigDB) 3.0, Epub 2011/05/07, Bioinformatics 27 (12) (2011) 1739–1740, https://doi.org/10.1093/bioinformatics/btr260. PubMed PMID: 21546393; PubMed Central PMCID: PMCPMC3106198.

[25] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, Epub 2013/01/18, BMC Bioinf. 14 (2013) 7, https://doi.org/10.1186/1471-2105-14-7. PubMed PMID: 23323831; PubMed Central PMCID: PMCPMC3618321.

[26] D.A. Barbie, P. Tamayo, J.S. Boehm, S.Y. Kim, S.E. Moody, I.F. Dunn, et al., Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1, Epub 2009/10/23, Nature 462 (7269) (2009) 108–112, https://doi.org/10.1038/nature08460. PubMed PMID: 19847166; PubMed Central PMCID: PMCPMC2783335.

[27] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, Epub 2008/12/31, BMC Bioinf. 9 (2008) 559, https://doi.org/10.1186/1471-2105-9-559. PubMed PMID: 19114008; PubMed Central PMCID: PMCPMC2631488.

[28] A. Kassambara, Ggpubr: 'ggplot2' Based Publication Ready Plots, 2023. https://CRAN.R-project.org/package=ggpubr.

[29] M. Kuhn, Building predictive models in R using the caret package, J. Stat. Software 28 (5) (2008) 1–26, https://doi.org/10.18637/jss.v028.i05.

[30] T.H. Chen, Benesty Tong, Khotilovich Michael, Vadim;. Xgboost: Extreme Gradient Boosting, 2022. https://CRAN.R-project.org/package=xgboost.

[31] A.S. Karatzoglou, Hornik Alex, Kurt kernlab, Kernel-Based Machine Learning Lab, 2023. https://CRAN.R-project.org/package=kernlab.

[32] A. Liaw, M.C. Wiener (Eds.), Classification and Regression by randomForest, 2007.

[33] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, 2016.

[34] P. Biecek, DALEX: explainers for complex predictive models in R, J. Mach. Learn. Res. 19 (84) (2018) 1–5.

[35] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, Epub 2011/03/19, BMC Bioinf. 12 (2011) 77, https://doi.org/10.1186/1471-2105-12-77. PubMed PMID: 21414208; PubMed Central PMCID: PMCPMC3068975.

[36] F.E. Harrell Jr., Rms: Regression Modeling Strategies, 2023. https://CRAN.R-project.org/package=rms.

[37] D. DeCock, K. Hyrich, Malignancy and rheumatoid arthritis: epidemiology, risk factors and management, Epub 2019/08/21, Best Pract. Res. Clin. Rheumatol. 32 (6) (2018) 869–886, https://doi.org/10.1016/j.berh.2019.03.011. PubMed PMID: 31427060.

[38] M.A.M. van Delft, T.W.J. Huizinga, An overview of autoantibodies in rheumatoid arthritis, Epub 2020/01/09, J. Autoimmun. 110 (2020) 102392, https://doi.org/10.1016/j.jaut.2019.102392. PubMed PMID: 31911013.

[39] J. Zhao, P. Jiang, S. Guo, S.J. Schrodi, D. He, Apoptosis, autophagy, NETosis, necroptosis, and pyroptosis mediated programmed cell death as targets for innovative therapy in rheumatoid arthritis, Epub 2022/01/11, Front. Immunol. 12 (2021) 809806, https://doi.org/10.3389/fimmu.2021.809806. PubMed PMID: 35003139; PubMed Central PMCID: PMCPMC8739882.

[40] L. Zhang, M. Zeng, B.M. Fu, Inhibition of endothelial nitric oxide synthase decreases breast cancer cell MDA-MB-231 adhesion to intact microvessels under physiological flows, Epub 2016/04/10, Am. J. Physiol. Heart Circ. Physiol. 310 (11) (2016) H1735–H1747, https://doi.org/10.1152/ajpheart.00109.2016. PubMed PMID: 27059076; PubMed Central PMCID: PMCPMC4935524.

[41] F.J. Stott, S. Bates, M.C. James, B.B. McConnell, M. Starborg, S. Brookes, et al., The alternative product from the human CDKN2A locus, p14(ARF), participates in a regulatory feedback loop with p53 and MDM2, Epub 1998/09/02, The EMBO journal 17 (17) (1998) 5001–5014, https://doi.org/10.1093/emboj/17.17.5001. PubMed PMID: 9724636; PubMed Central PMCID: PMCPMC1170828.

[42] H. Wang, X. Wang, L. Xu, Y. Lin, J. Zhang, H. Cao, Identification of genomic alterations and associated transcriptomic profiling reveal the prognostic significance of MMP14 and PKM2 in patients with pancreatic cancer, Epub 2020/09/21, Aging 12 (18) (2020) 18676–18692, https://doi.org/10.18632/aging.103958. PubMed PMID: 32950968; PubMed Central PMCID: PMCPMC7585111.

[43] P. Sievers, T. Hielscher, D. Schrimpf, D. Stichel, D.E. Reuss, A.S. Berghoff, et al., CDKN2A/B homozygous deletion is associated with early recurrence in meningiomas, Epub 2020/07/10, Acta Neuropathol. 140 (3) (2020) 409–413, https://doi.org/10.1007/s00401-020-02188-w. PubMed PMID: 32642869; PubMed Central PMCID: PMCPMC7423850.

[44] N.A. Ipenburg, M.A. El Sharouni, R. van Doorn, P.J. van Diest, M.E. van Leerdam, J.I. van der Rhee, et al., Lack of association between CDKN2A germline mutations and survival in patients with melanoma: a retrospective cohort study, Epub 2021/10/26, J. Am. Acad. Dermatol. 87 (2) (2022) 479–482, https://doi.org/10.1016/j.jaad.2021.10.024. PubMed PMID: 34695526.

[45] A. Okamoto, D.J. Demetrick, E.A. Spillare, K. Hagiwara, S.P. Hussain, W.P. Bennett, et al., Mutations and altered expression of p16INK4 in human cancer, Epub 1994/11/08, Proc. Natl. Acad. Sci. U.S.A. 91 (23) (1994) 11045–11049, https://doi.org/10.1073/pnas.91.23.11045. PubMed PMID: 7972006; PubMed Central PMCID: PMCPMC45163.

[46] W.Y. Kim, N.E. Sharpless, The regulation of INK4/ARF in cancer and aging, Epub 2006/10/24, Cell 127 (2) (2006) 265–275, https://doi.org/10.1016/j.cell.2006.10.003. PubMed PMID: 17055429.

[47] M. Collado, M. Serrano, Senescence in tumours: evidence from mice and humans, Epub 2009/12/24, Nat. Rev. Cancer 10 (1) (2010) 51–57, https://doi.org/10.1038/nrc2772. PubMed PMID: 20029423; PubMed Central PMCID: PMCPMC3672965.

[48] L.A. Donehower, T. Soussi, A. Korkut, Y. Liu, A. Schultz, M. Cardenas, et al., Integrated analysis of TP53 gene and pathway alterations in the cancer genome atlas, 1370-84.e5. Epub 2019/08/01, Cell Rep. 28 (5) (2019), https://doi.org/10.1016/j.celrep.2019.07.001. PubMed PMID: 31365877; PubMed Central PMCID: PMCPMC7546539.

[49] T. Grob, A.S.A. Al Hinai, M.A. Sanders, F.G. Kavelaars, M. Rijken, P.L. Gradowska, et al., Molecular characterization of mutant TP53 acute myeloid leukemia and high-risk myelodysplastic syndrome, Epub 2022/02/03, Blood 139 (15) (2022) 2347–2354, https://doi.org/10.1182/blood.2021014472. PubMed PMID: 35108372.

[50] S.M. Akula, P.P. Ruvolo, J.A. McCubrey, TP53/miR-34a-associated signaling targets SERPINE1 expression in human pancreatic cancer, Epub 2020/01/28, Aging 12 (3) (2020) 2777–2797, https://doi.org/10.18632/aging.102776. PubMed PMID: 31986125; PubMed Central PMCID: PMCPMC7041729.

[51] J. Wang, D. Liu, Z. Sun, T. Ye, J. Li, B. Zeng, et al., Autophagy augments the self-renewal of lung cancer stem cells by the degradation of ubiquitinated p53, Epub 2021/01/21, Cell Death Dis. 12 (1) (2021) 98, https://doi.org/10.1038/s41419-021-03392-6. PubMed PMID: 33468994; PubMed Central PMCID: PMCPMC7815724.

[52] G. Wu, A.K. Diaz, B.S. Paugh, S.L. Rankin, B. Ju, Y. Li, et al., The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma, Epub 2014/04/08, Nat. Genet. 46 (5) (2014) 444–450, https://doi.org/10.1038/ng.2938. PubMed PMID: 24705251; PubMed Central PMCID: PMCPMC4056452.

[53] S. Weissmueller, E. Manchado, M. Saborowski, JPt Morris, E. Wagenblast, C.A. Davis, et al., Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor β signaling, Epub 2014/04/15, Cell 157 (2) (2014) 382–394, https://doi.org/10.1016/j.cell.2014.01.066. PubMed PMID: 24725405; PubMed Central PMCID: PMCPMC4001090.
[54] D.P. Liu, H. Song, Y. Xu, A common gain of function of p53 cancer mutants in inducing genetic instability, Epub 2009/11/03, Oncogene 29 (7) (2010) 949–956, https://doi.org/10.1038/onc.2009.376. PubMed PMID: 19881536; PubMed Central PMCID: PMCPMC2837937.
[55] Z. Yue, Y. Zhou, P. Zhao, Y. Chen, Y. Yuan, Y. Jing, et al., p53 Deletion promotes myeloma cells invasion by upregulating miR19a/CXCR5, Epub 2017/08/08, Leuk. Res. 60 (2017) 115–122, https://doi.org/10.1016/j.leukres.2017.07.003. PubMed PMID: 28783539.
[56] F. Mantovani, L. Collavin, G. Del Sal, Mutant p53 as a guardian of the cancer cell, Epub 2018/12/13, Cell Death Differ. 26 (2) (2019) 199–212, https://doi.org/10.1038/s41418-018-0246-9. PubMed PMID: 30538286; PubMed Central PMCID: PMCPMC6329812.
[57] L. Don Wai Luu, N.O. Kaakoush, N. Castano-Rodriguez, The role of ATG16L2 in autophagy and disease, Epub 2022/03/04, Autophagy 18 (11) (2022) 2537–2546, https://doi.org/10.1080/15548627.2022.2042783. PubMed PMID: 35239457; PubMed Central PMCID: PMCPMC9629082.
[58] Y. Cao, D. Zhao, P. Li, L. Wang, B. Qiao, X. Qin, et al., MicroRNA-181a-5p impedes IL-17-induced nonsmall cell lung cancer proliferation and migration through targeting VCAM-1, Epub 2017/05/24, Cell. Physiol. Biochem. : international journal of experimental cellular physiology, biochemistry, and pharmacology 42 (1) (2017) 346–356, https://doi.org/10.1159/000477389. PubMed PMID: 28535543.
[59] N. Mizushima, M. Komatsu, Autophagy: renovation of cells and tissues, Epub 2011/11/15, Cell 147 (4) (2011) 728–741, https://doi.org/10.1016/j.cell.2011.10.026. PubMed PMID: 22078875.
[60] J. Wen, H. Liu, L. Wang, X. Wang, N. Gu, Z. Liu, et al., Potentially functional Variants of ATG16L2 predict radiation pneumonitis and outcomes in patients with non-small cell lung cancer after definitive radiotherapy, Epub 2018/02/20, J. Thorac. Oncol. 13 (5) (2018) 660–675, https://doi.org/10.1016/j.jtho.2018.01.028. PubMed PMID: 29454863; PubMed Central PMCID: PMCPMC6124306.
[61] T. Yamaguchi, A. Kurisaki, N. Yamakawa, K. Minakuchi, H. Sugino, FKBP12 functions as an adaptor of the Smad7-Smurf1 complex on activin type I receptor, Epub 2006/05/25, J. Mol. Endocrinol. 36 (3) (2006) 569–579, https://doi.org/10.1677/jme.1.01966. PubMed PMID: 16720724.
[62] S. Cai, Z. Chen, H. Tang, S. Meng, L. Tao, Q. Wang, Upregulated FKBP1A suppresses glioblastoma cell growth via apoptosis pathway, Epub 2022/12/12, Int. J. Mol. Sci. 23 (23) (2022), https://doi.org/10.3390/ijms232314935. PubMed PMID: 36499275; PubMed Central PMCID: PMCPMC9739687.
[63] W. Leng, Q. Liu, S. Zhang, D. Sun, Y. Guo, LncRNA AFAP1-AS1 modulates the sensitivity of paclitaxel-resistant prostate cancer cells to paclitaxel via miR-195-5p/FKBP1A axis, Epub 2020/11/04, Cancer Biol. Ther. 21 (11) (2020) 1072–1080, https://doi.org/10.1080/15384047.2020.1829266. PubMed PMID: 33138677; PubMed Central PMCID: PMCPMC7678925.
[64] Y. Zhang, D. Zhang, J. Lv, S. Wang, Q. Zhang, LncRNA SNHG15 acts as an oncogene in prostate cancer by regulating miR-338-3p/FKBP1A axis, Epub 2019/04/15, Gene 705 (2019) 44–50, https://doi.org/10.1016/j.gene.2019.04.033. PubMed PMID: 30981837.
[65] F.Z. Chakrama, S. Seguin-Py, J.N. Le Grand, A. Fraichard, R. Delage-Mourroux, G. Despouy, et al., GABARAPL1 (GEC1) associates with autophagic vesicles, Epub 2010/04/21, Autophagy 6 (4) (2010) 495–505, https://doi.org/10.4161/auto.6.4.11819. PubMed PMID: 20404487.
[66] M. Jacquet, E. Hervouet, T. Baudu, M. Herfs, C. Parratte, J.P. Feugeas, et al., GABARAPL1 inhibits EMT signaling through SMAD-tageted negative feedback, Epub 2021/10/24, Biology 10 (10) (2021), https://doi.org/10.3390/biology10100956. PubMed PMID: 34681055; PubMed Central PMCID: PMCPMC8533302.
[67] B. Su, L. Zhang, S. Liu, X. Chen, W. Zhang, GABARAPL1 promotes AR+ prostate cancer growth by increasing FL-AR/AR-V transcription activity and nuclear translocation, Epub 2019/12/06, Front. Oncol. 9 (2019) 1254, https://doi.org/10.3389/fonc.2019.01254. PubMed PMID: 31803623; PubMed Central PMCID: PMCPMC6872515.
[68] M. Boyer-Guittaut, L. Poillet, Q. Liang, E. Bole-Richard, X. Ouyang, G.A. Benavides, et al., The role of GABARAPL1/GEC1 in autophagic flux and mitochondrial quality control in MDA-MB-436 breast cancer cells, Epub 2014/06/01, Autophagy 10 (6) (2014) 986–1003, https://doi.org/10.4161/auto.28390. PubMed PMID: 24879149; PubMed Central PMCID: PMCPMC4091181.
[69] Y. Choi, J.W. Bowman, J.U. Jung, Autophagy during viral infection - a double-edged sword, Epub 2018/03/21, Nat. Rev. Microbiol. 16 (6) (2018) 341–354, https://doi.org/10.1038/s41579-018-0003-6. PubMed PMID: 29556036; PubMed Central PMCID: PMCPMC6907743.
[70] S. Saha, D.P. Panigrahi, S. Patil, S.K. Bhutia, Autophagy in health and disease: a comprehensive review, Epub 2018/05/26, Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 104 (2018) 485–495, https://doi.org/10.1016/j.biopha.2018.05.007. PubMed PMID: 29800913.
[71] V. Deretic, T. Saitoh, S. Akira, Autophagy in infection, inflammation and immunity, Epub 2013/09/26, Nat. Rev. Immunol. 13 (10) (2013) 722–737, https://doi.org/10.1038/nri3532. PubMed PMID: 24064518; PubMed Central PMCID: PMCPMC5340150.