Research article

Construction and validation of a diagnostic model for rheumatoid arthritis based on mitochondrial autophagy-related genes

Iong Iok In [a,*], Weiming Deng [b]

[a] Department of Rheumatology, Kiang Wu Hospital, Rua de Coelho do Amaral 60–70, Macao, SAR, China
[b] Department of Rheumatology, Guangdong Second Provincial General Hospital, No. 466 Xingang Middle Road, Guangdong, China

ARTICLE INFO

ABSTRACT

Rheumatoid arthritis (RA) is an autoimmune disease associated with an increased risk of disability. Due to its slow progression, timely diagnosis and treatment during the early stages can effectively decelerate disease advancement. Consequently, there is a pressing need to investigate additional biomarkers and therapeutic targets relevant to RA diagnosis. Mitochondrial autophagy, a biological process that regulates the quantity of mitochondria, is intricately linked to the development of tumor diseases. However, the role of autophagy in RA remains unclear. To address this, transcriptome data from the GEO database were collected for RA and its controls and subjected to differential expression analysis. The differentially expressed genes obtained were then intersected with mitochondrial autophagy-related genes. Subsequently, the overlapping genes were further intersected with genes from critical modules obtained through the weighted co-expression network algorithm. Diagnostic genes were identified, and diagnostic models were constructed for the resulting genes using the random forest and LASSO algorithms. The model achieved an AUC of 0.916 in the GSE93272 dataset and 0.951 in the GSE17755 dataset. Additionally, qPCR experiments were conducted on the diagnostic genes. Finally, we explored the correlation between the abundance of immune cell infiltration and diagnostic genes, constructing a drug-gene interaction network. The diagnostic genes identified in this study can serve as a reference for early diagnosis and the discovery of therapeutic targets in RA.

1. Introduction

Rheumatoid arthritis (RA) is an autoimmune disease primarily affecting connective tissues, leading to synovitis, bone erosion, and synovial hyperplasia [1,2]. Given its relatively low prevalence (approximately 1 % of the global population suffers from RA), the progression of this disease poses a potential risk of disability to affected individuals [3]. Due to its gradual development, early detection and intervention play a crucial role in effectively managing the inflammatory response and alleviating joint pain [4]. The decreasing cost of sequencing has facilitated easy access to substantial amounts of publicly available RA transcriptome data. Leveraging machine learning algorithms for the identification of biomarkers associated with the onset and progression of RA holds promise for aiding in early diagnosis and precise treatment [5]. In the treatment of RA, targeted inflammatory cytokines such as IL-1β, IL-6, and TNF-α have been extensively utilized [6–8].

Autophagy entails the translocation of eukaryotic components to lysosomes for degradation, a process demonstrated to be crucial

---

for energy metabolism, development, and other cellular functions [9,10]. Mitochondrial autophagy specifically targets mitochondria, serving as a mechanism to regulate the number of mitochondria in cases of mutations or protein misfolding [11]. This biological process is indispensable for preserving mitochondrial homeostasis and quality. Prior investigations have elucidated the role of metabolic reprogramming in mitochondria within tumors, contributing to the prolonged survival of cancer cells [12]. Nevertheless, the involvement of mitochondrial autophagy in rheumatoid arthritis (RA) remains unclear.

To investigate the role of mitochondrial autophagy-related genes (MRGs) in (RA), we obtained two transcriptomic datasets (GSE93272 and GSE17755) for RA and its corresponding controls from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). These datasets were utilized for training and testing the model, respectively. Differentially expressed analysis was conducted on the GSE93272 dataset by identifying differentially expressed genes (DEGs) and intersecting them with MRGs previously documented in the literature. A weighted co-expression network analysis (WGCNA) was then performed on the transcriptome data from the GSE93272 dataset. This involved intersecting genes in the key modules selected by WGCNA with the previously identified intersecting genes. Subsequently, the new intersecting genes were ranked for feature importance using the random forest algorithm. Additionally, the Lasso algorithm was employed to screen for genes associated with RA diagnosis and to construct diagnostic models. To validate the diagnostic genes and models, an external dataset was utilized. Finally, diagnostic genes exhibiting a significant correlation with immune cells were explored, along with drugs interacting with these diagnostic genes, through immune infiltration analysis and drug-gene network construction, respectively.

2. Methods

2.1. Data set acquisition

In this paper, transcriptomic data from two sets of RA and their control samples (GSE93272 and GSE17755) were downloaded from the GEO database. Among them, GSE93272 includes transcriptomic data from 232 RAs to 43 controls. This data set is used for model training. GSE17755 included transcriptomic data from 112 RAs to 53 controls. This dataset is used for model validation. 137 mitochondrial autophagy related data were obtained from Ref. [13], which can be obtained in supplementary material (MRGs.xlsx).

2.2. Batch correction and differential expression analysis

The ComBat method in the R package "sva" was applied to batch correct the two datasets, which were then normalised using the scale function of the R program (version 4.2.0). For the GSE93272 data set, we used the R package "limma" to extract the DEGs for the RA and control groups. Ultimately, 10677 DEGs with $p < 0.05$ and $|logFC|>0$ were retained.

The ComBat method from the "sva" R package was employed for batch correction on the two datasets, followed by normalization using the scale function in the R program (version 4.2.0). In the case of the GSE93272 dataset, we utilized the "limma" R package to identify the DEGs between the RA and control groups. Ultimately, 10,677 DEGs meeting the criteria of $p < 0.05$ and $|logFC|>0$ were retained.

2.3. Enrichment analysis

We compared 137 MRGs identified from previous literature with 10,677 Differentially DEGs, resulting in 69 shared genes. Subsequently, we conducted Gene Ontology (GO) enrichment analysis, Metascape enrichment analysis, and GSEA for these shared genes. In particular, GO enrichment analysis of the intersecting genes was performed using the R package 'clusterProfiler,' with the retention of significant pathways at $p < 0.05$. The results were then visualized using the R package 'ggplot 2.' The Metascape database (http://metascape.org) was employed for the enrichment analysis of intersecting genes, and pathways with a significance level of $p < 0.05$ were retained. Lastly, gene set enrichment scores were computed for both the RA and control groups using the GSEA method, facilitating the comparison of biological pathways between the two groups. The h.all.v7.2.symbols.gmt gene set was downloaded and used as a reference for this analysis.

2.4. Weighted co-expression network analysis and diagnostic model construction

The Weighted Co-expression Network Analysis (WGCNA) algorithm can be implemented by the R package "WGCNA" for further screening of intersecting genes. The algorithm consists of three steps: outlier removal, pathway/gene module formation, and module screening. Specifically, all genes from the GSE93272 data set and their labels were input to the WGCNA algorithm to construct a weighted correlation network. To calculate the strength of the connections between nodes in the network, this paper calculated the Pearson correlation coefficient between nodes $corr(i.j)$. Next, it calculated the weighted adjacency matrix $a_{ij} = (0.5*(1 + corr(ij)))$. In addition, the algorithm has several performance metrics that β represent soft threshold power, which has the ability to represent module independence, and average connectivity. The topological overlap metric (TOM) can represent the overlap of network neighbors, and (1-TOM) retrieves pairwise distances to identify hierarchically clustered nodes and modules. Finally, the clustering was performed by applying the dynamic tree-cutting technique of the mapping package in R software. In the end, we set the minimum number of modules to 100, the optimal power value to 8, and the threshold for module similarity to 0.25.

We selected the genes in the module most relevant to RA and intersected them with the previously intersected genes to obtain the new intersected genes. In this paper, we performed a feature importance ranking of newly intersected genes by applying the R package

"randomForest." The R package "glmnet" was then applied to implement the LASSO algorithm, which screens for genes associated with RA diagnosis in the new intersection of genes. Specifically, the GSE93272 data set uses gene expression values as features and diagnostic labels as response variables. During the training phase, we utilize part of the data set (training data) to build our diagnostic model. Apply LASSO regression to the subset data to select a subset of relevant genes (features) and estimate their coefficients. To optimize LASSO, we used cross-validation. Cross-validation allows us to evaluate the model's performance across different lambda values. Typically, 10-fold cross-validation is used, where the training set is divided into ten subsets (folds). LASSO regression fitting was performed on 10-1 folds, and the remaining folds were verified. This process is repeated ten times as a validation set, and the average performance metric (e.g., mean squared error or classification accuracy) is calculated. Once the optimal λ value was determined by cross-validation on the training data, we applied the selected model (with the selected set of genes and coefficients) to our GSE17755 dataset.

The ROC curves for the diagnostic genes and diagnostic models in the GSE93272 and GSE17755 data sets were subsequently plotted using the R package "glmnet." In addition, the "glm" function of R language is used to implement logistic regression.

2.5. Immuno-infiltration analysis and drug-gene network construction

This article applied the CIBERSORT algorithm (implemented by R software) to explore the differences in the infiltration abundance of 22 immune cells in RA and control groups. Then the correlation and significance of the abundance of immune cell infiltration with diagnostic genes were also calculated. Finally, the paper uses the DGIdb database (https://dgidb.org/) to construct a network of interactions between diagnostic genes and drugs that have interactions with them.

2.6. RT-qPCR

2.6.1. Sample collection and processing

Five patients with rheumatoid arthritis diagnosed in the outpatient clinic of the Department of Rheumatology and Immunology of the Second People's Hospital of Guangdong Province from April 2018 to May 2020 were selected, and five healthy persons undergoing physical examination were selected as the control group. Patient inclusion criteria: Patients with rheumatoid arthritis included in this study were chosen by relevant clinical physicians based on the applicable standards revised by the American College of Rheumatology in 2010. Patient exclusion criteria: 1) Patients with systemic lupus erythematosus, Sjogren's syndrome, and other autoimmune diseases. 2) Patients with malignant tumors and systemic infections. 3)Patients with severe heart, liver, kidney, etc. failure. 4) Who are unwilling to participate in this study.

Inclusion criteria for healthy control groups: The healthy control groups selected for this study are all healthy volunteers who underwent physical examinations in our hospital from April 2018 to May 2020. The healthy control groups for this study need to meet the following requirements: 1) Do not meet any of the criteria for the diagnosis of rheumatoid arthritis; 2) The body does not suffer from systemic lupus erythematosus, Sjogren's syndrome, and other autoimmune system diseases: 3) People with malignant tumors and auto-infections: the latest one Be in good health within the month and have not used any hormones or immunosuppressive drugs: 4) Those who have immediate family members who do not suffer from autoimmune system diseases: 5) Subjects who agree to participate in this study.

2.6.2. Extraction of serum/plasma total mRNA

In order to detect the expression of related genes in patients with RA by RT-qPCR, we used five serum samples from RA patients and five samples from normal people. The experimental steps were as follows: at first, 750 μl RLS was added to every 250 μl serum, vortexed, and incubated at 15–30 °C for 5 min. 200 μl chloroform was added to every 750 μl Lysis buffer, which was shaken violently for 15 s and left at room temperature for 2 min, then centrifuged at 4 °C and 12,000 rpm for 10 min. The content of the water phase layer was about 60 % of the RLS volume. We carefully transferred the upper water phase into a new centrifuge tube, added 1.5 times the volume of anhydrous ethanol (room temperature), and mixed it upside down. Then we added the mixture to an adsorption column RA, put the adsorption column into a collection tube, centrifuged at 12,000 rpm for 30 s, and discarded the waste liquid. We added 500 μl deproteinized liquid RE, centrifuged at 12,000 rpm for 30 s, and discarded the waste liquid. Next, we added 500 μl of rinsing liquid RW, centrifuged at 12,000 rpm for 15 s, and discarded the waste liquid. And we repeated it again. Then we put the adsorption column RA back into the empty collection tube, centrifuged at 13,000 rpm for 2 min, and removed the rinsing liquid. Then we took out the adsorption column RA, put it into an RNase free centrifuge tube, added 30 μl RNase free water to the adsorption membrane, left it at room temperature for 2 min, and centrifuged at 12,000 rpm for 1 min. At last, we added the eluent back to the adsorption column and repeated the elution once.

2.6.3. Experimental consumables

qPCR: enzyme-free pipette tip, 1.5 ml enzyme-free centrifuge tube (KeyGEN, China)); pipette (Ebende); nitrile latex gloves (Loctite); CPE gloves (Guangming); 0.2 ml PCR single tube (Aisigen); 0.1 ml qPCR eight-strip tube (BIO-RAD); fluorescence quantitative 96-well PCR plate, fluorescence quantitative nuclease-free sealing membrane (Sangon, China); whole blood total RNA extraction kit (RN23)/Serum/plasma microRNA rapid extraction kit (Aidlab, China); HiScript lll 1st Strand cDNA Synthesis Kit, Taq Pro Universal SYBR qPCR Master Mix (Vazyme, China).

Table 1
qPCR reaction.

Steps	Reaction	Cycle times	Temperature	Time
Stage 1	Predenaturation	Reps:1	95 °C	5 min
Stage 2	Cyclic reaction	Reps:40	95 °C	10 s
			60 °C	30 s
Stage 3	Melting curve	Reps:1	95 °C	15 s
			60 °C	60 s
			95 °C	15 s

Table 2
Primer sequence.

Gene	Sense/antisense primer	Sequence (5′-3′)
ATG5	forward	AAAGCGAGAGTCTAGTCGCA
	reverse	GAGGACATACCCACCGAGTT
MFN1	forward	ACGCCTTATCCTCTAGTTCCTGTG
	reverse	GGTCTGTCAGCATCATCTTCCAAG
NRAS	forward	CTCCTTGGTGGGGGCTGTTC
	reverse	TCTCCCCACGTAGGCACCAA
SLC25A46	forward	TTCTTTCACCCCCGACGTGC
	reverse	TCCCGGCCAGAAATACCCCA
RELA	forward	CGCCAGGCTTCAGAGACGAG
	reverse	CCCTCTCCAGCTAAAGCGCC
TSC2	forward	TGGTCCACCATGGCCAAACC
	reverse	AGTTGCCAGCAACCCGTCTC
GSK3A	forward	CGGAAAGGCATCTGTCGGGG
	reverse	GGGAGGAAGAGGGCTCGGAT
RRAS2	forward	TTGTTGCCCAGGCTGGAGTG
	reverse	GGATCACCTGCGGGTTGGGAG
GAPDH	forward	GGTCTCCTCTGACTTCAACA
	reverse	GTGAGGGTCTCTCTCTTCCT

2.6.4. cDNA synthesis

CDNA was synthesized in a 20 μl system, and RNA 500 ng, 5×gDNA wiper mix 2 μl and RNase-free water to10 μl were added into a 200 μl eppendorf tube. After reacting at 42 °C for 2 min, we added 10x RT Mix 2 μl, Hiscriptl enzyme mix 2 μ l, Stem-loop primer (2 m) 1 μ l, and RNase-free water 5 μl to the above mixed solution, mix them evenly, react at 25 °C for 5 min, at 50 °C for 15 min, and at 85 °C for 5 min to obtain cDNA, and store them at −20 °C.

2.6.5. Real-time quantitative PCR

We used 20 μl system, added 3 times diluted cDNA 5 μl, Nuclease-free water 4.2 μl, 2 × mirna universal sybr qpcr master mix 10 μ l, Forward primer 0.4 μl(10 μM), and Reverse primer 0.4 μl(10 μM). Then we added the premixed reagent into the 8-row tube, sealed it, put it into the instrument, set the reaction parameters, and amplified it. The result was calculated by2-ΔΔCt. In addition, the reference gene in this paper is GAPDH. The qPCR reaction stages and primer sequences for each gene are presented in Tables 1 and 2, respectively.

2.7. Statistical analysis

This article applied R software (version 4.2.0) and SPSS (version 26.0) for statistical analysis. Subject operating characteristic (ROC) curves (glmnet package in R software) were performed to estimate the specificity and sensitivity of potential biomarkers. $p < 0.05$ should be significant. Real-time quantitative PCR data were analyzed and plotted using GraphPad Prism 9 (Version 9.5.0). Photoshop was used to organize the images. All plots are presented as means ± SD, statistical differences between groups were tested by T-Test, and P values less than 0.05 were considered significant differences.

3. Results

3.1. The expression landscape of MRGs in RA

The flowchart of the paper is given in Fig. 1. First of all, two transcriptomic datasets of RA and its controls (GSE93272 and GSE17755) were downloaded from the GEO database for this paper. The batch correction was then performed on both datasets (Fig. 2A and B). Differential expression analysis was performed on the GSE93272 data set samples using the limma algorithm (screening criteria: $p < 0.05$ and |logFC|>0), and 10677 DEGs were retained. Fig. 2C provides a heat map of the volcanoes during differential
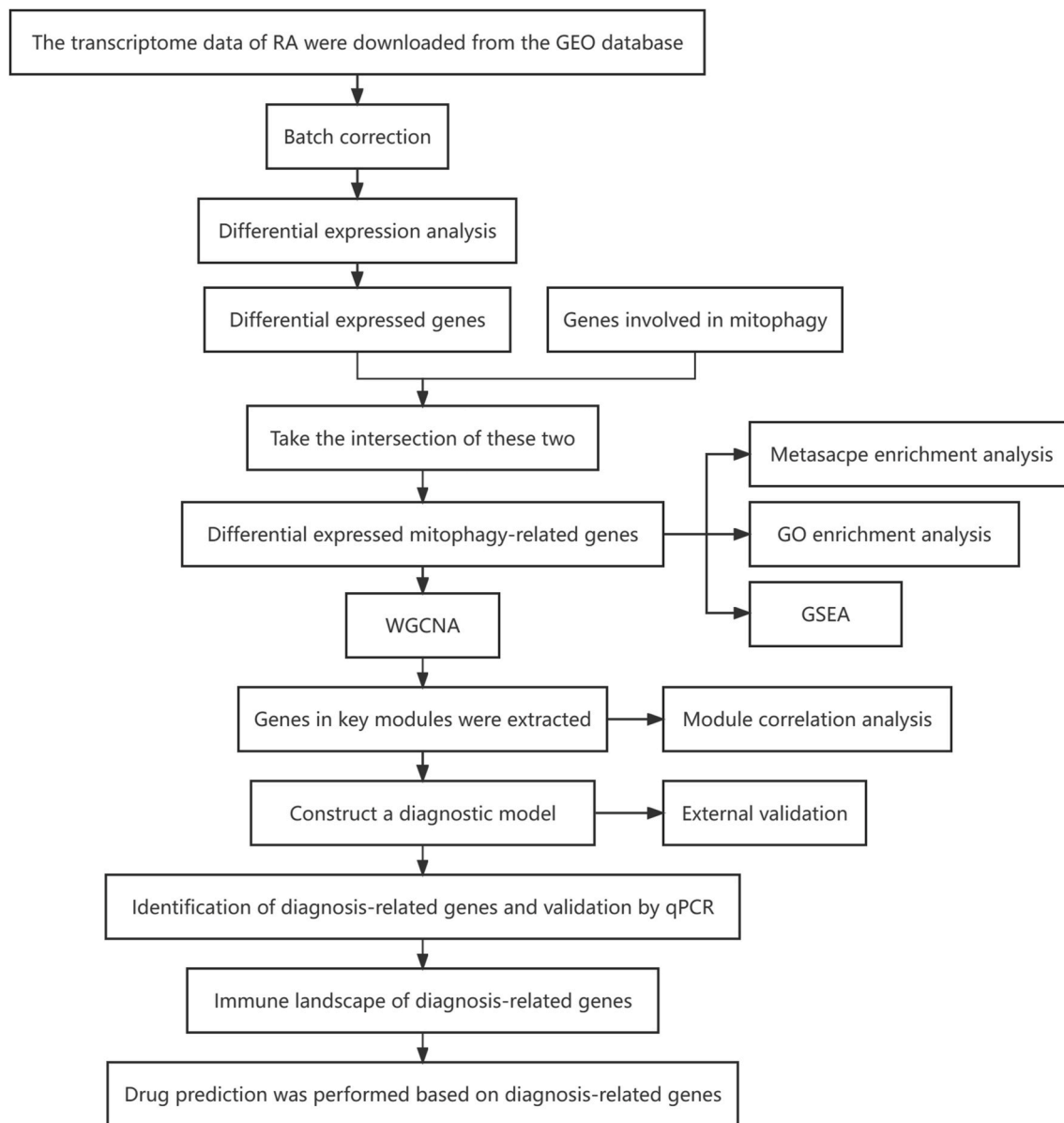
Fig. 1. The flowchart of the paper.

expression analysis. Fig. 2D gives a heat map of the expression of the top 40 DEGs in the RA and control groups. The expression of these genes was significantly different in the two groups. We collected 137 MRGs from the previous literature and intersected them with DEGs, and the intersected Venn diagram is given in Fig. 2E. Details of the DEGs, MRGs, and intersecting genes are shown in the file "DEGs&MRGs.xlsx" in our supplementary materials.

For the intersecting genes, we employed various enrichment analyses to extract information regarding the pathways in which they participate. The results of the GO enrichment analysis are presented in Fig. 3A and B. Fig. 3C and D illustrate the outcomes of the Metascape analysis. Furthermore, the GSEA provided insights into the two highest-ranked pathways (Fig. 4A–I), with additional pathways identified by GSEA available in Fig. S1 of the Supplementary Material. Dysfunctional autophagy plays a significant role in the development and pathogenesis of autoimmune diseases such as RA [14]. Sugawara et al. investigated the involvement of macrophage/autophagy in RA autoimmunity. They discovered that SFs may enhance RA autoimmunity through the citrullination of citrullinated vimentin (VIM) and its interaction with major histocompatibility complex (MHC) class II, facilitated by autophagy [15]. Additionally, excessive reactive oxygen species (ROS) generated by abnormal mitochondria can contribute to RA. Mitochondrial autophagy, mediated by the PINK1/Parkin pathway, effectively eliminates damaged mitochondria, preventing the overproduction of ROS [16]. Researchers also observed that oxidative stress influences the proliferation of RA fibroblast-like synoviocytes (FLS cells) by regulating mitochondrial autophagy [17]. Mitochondrial autophagy, a selective process targeting damaged mitochondria through
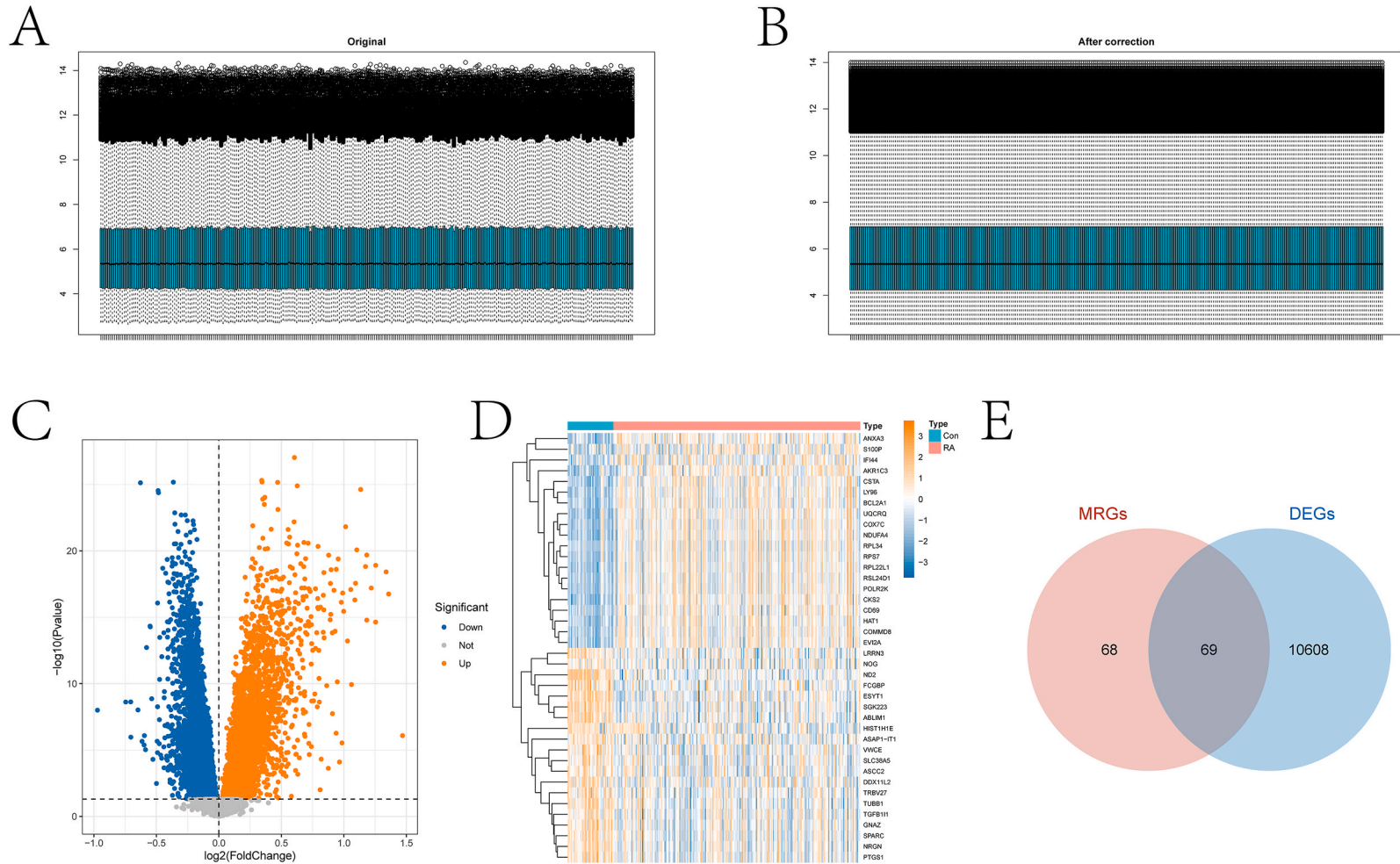
Fig. 2. Acquisition and expression of intersecting genes. A and B are box plots of the data before and after correction, respectively. The horizontal coordinate is the sample, and the vertical coordinate is the chip signal intensity. C shows a volcano plot of the transcriptome data in the GSE93272 data set during differential expression analysis. D is the expression of Top DEGs in the GSE93272 data set in the RA and control groups. E is a Venn diagram of the DEGs and MRGs after taking the intersection.

A
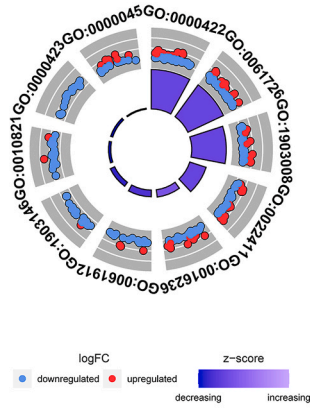
| ID	Description
GO:0000422	autophagy of mitochondrion
GO:0061726	mitochondrion disassembly
GO:1903008	organelle disassembly
GO:0022411	cellular component disassembly
GO:0016236	macroautophagy
GO:0061912	selective autophagy
GO:1903146	regulation of autophagy of mitochondrion
GO:0010821	regulation of mitochondrion organization
GO:0000423	mitophagy
GO:0000045	autophagosome assembly

B
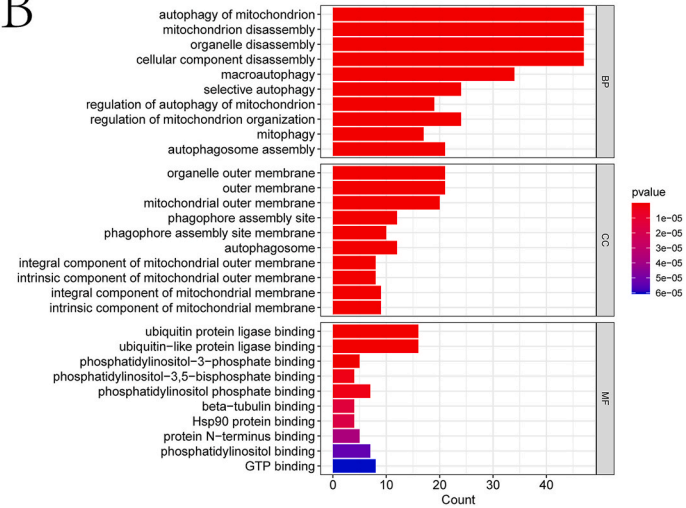
C
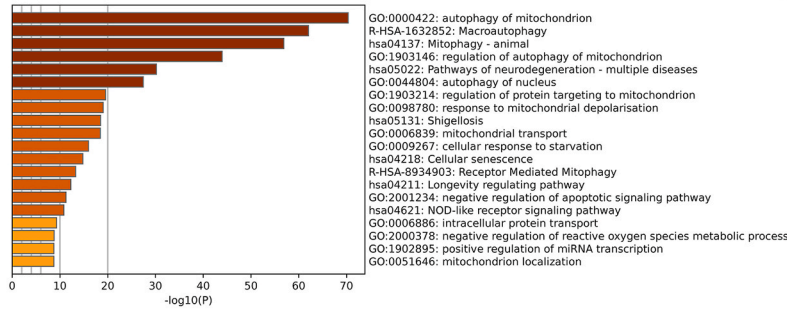
D
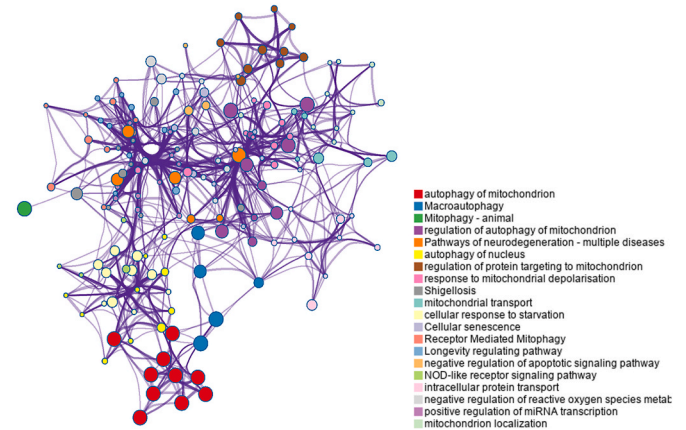
Fig. 3. Enrichment analysis of intersectional genes. A and B are the results of GO enrichment analysis. C and D are histograms and pathway interaction network diagrams obtained from metascape enrichment analysis.
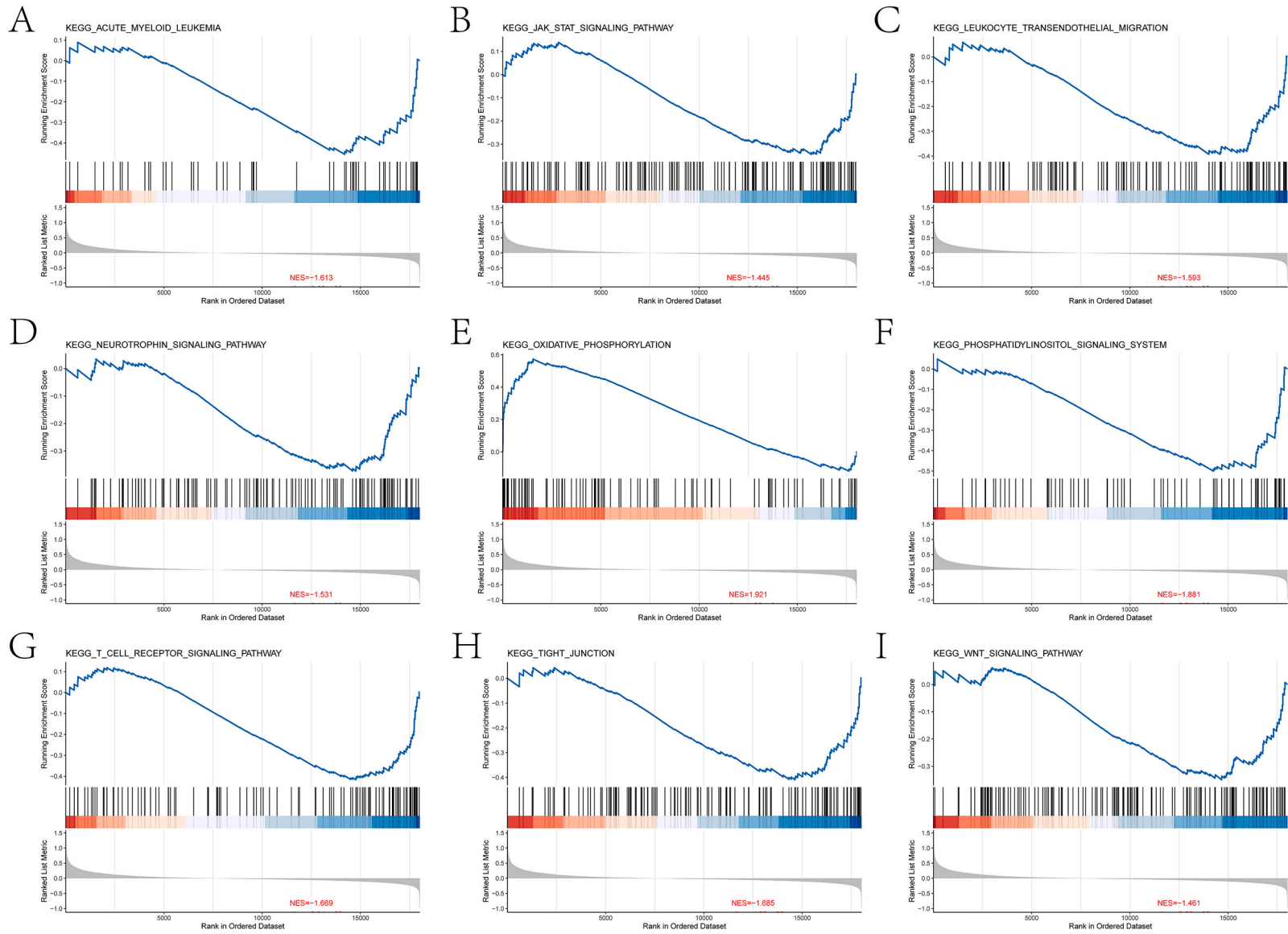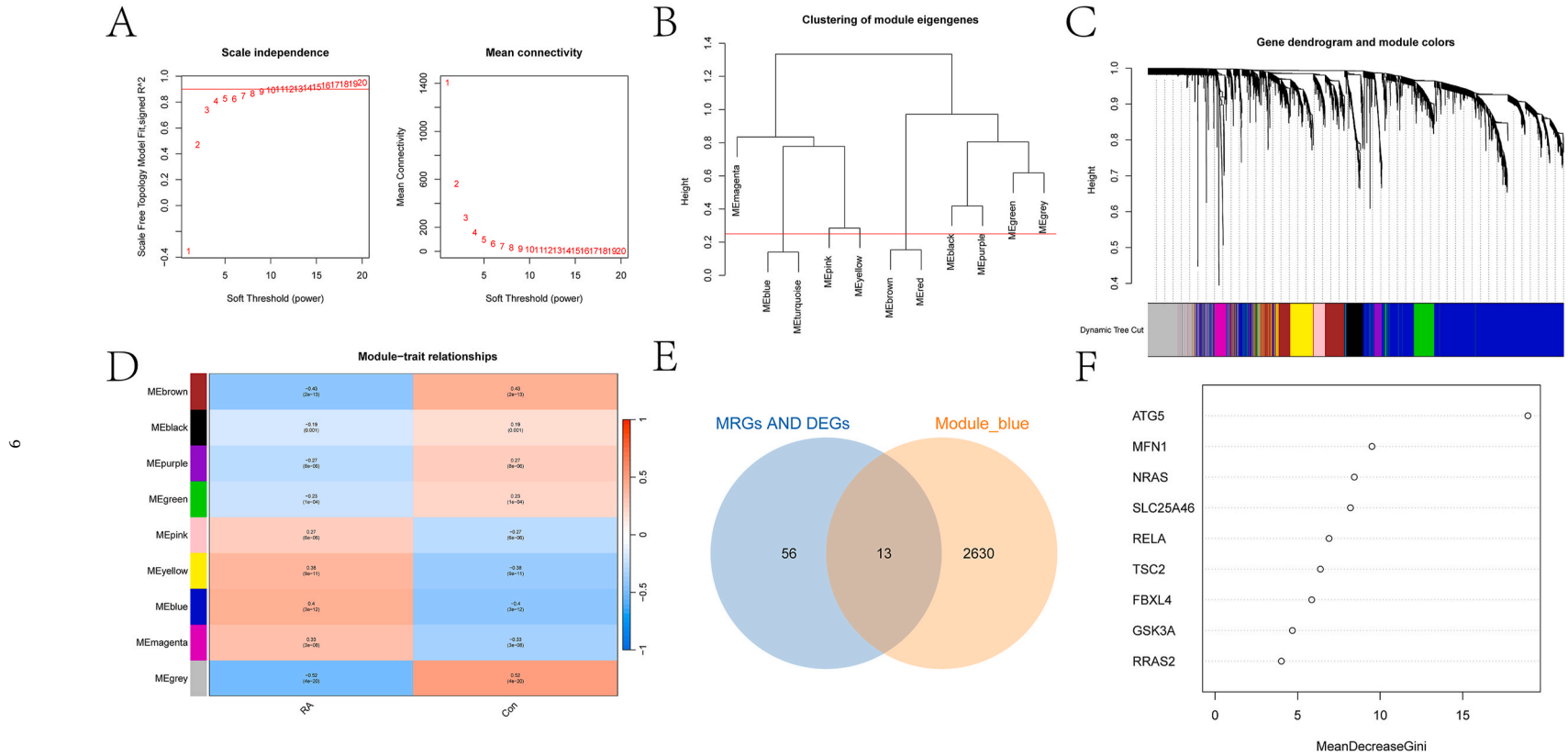
Fig. 4. The results based on GSEA analysis of the intersecting genes in RA and its controls. A-I are partially significant pathways obtained from GSEA analysis.

Fig. 5. Results of the WGCNA analysis of the GSE93272 data set. A shows scale independence and mean connectivity analyses for various soft threshold powers. B is the characteristic gene dendrogram. C shows the clustering dendrogram of genes. The different colors below indicate the different co-expression modules. D is the heat map of module-group relationships. Each row represents a module signature gene, and each column represents a subgroup (RA or control group). Each cell contains the corresponding correlation and p-value. E is a Venn diagram of the intersection of the blue module and the differential mitochondrial autophagy-associated genes taken. F indicates the random forest algorithm screened diagnostic genes with their corresponding mean decrease accuracy bubble plots. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
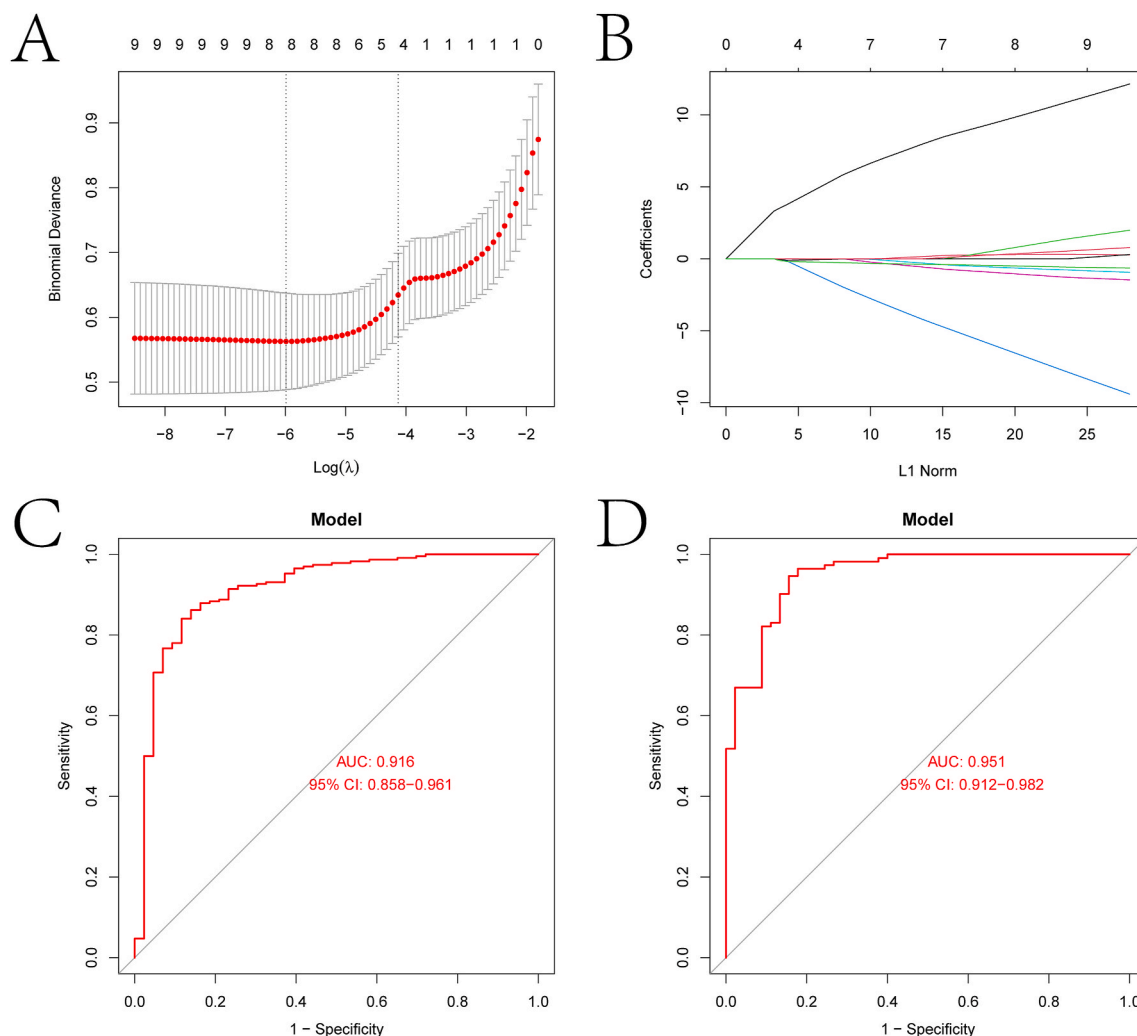
Fig. 6. LASSO analysis and the construction of diagnostic models. A and B are the LASSO algorithm screening for the most relevant genes for RA diagnosis and determining the best penalty value in the LASSO regression show the ROC curves for a diagnostic model consisting of diagnostic genes in the GSE93272 and GSE93272 data sets, respectively.

receptor-mediated mechanisms, plays a pivotal role in maintaining the healthy and stable state of damaged mitochondria in cells [18]. Recent research indicates that ultrasound (US) piezoelectric catalytic therapy can alleviate RA symptoms by activating mitochondrial autophagy under two-dimensional piezoelectric nanoplates [19]. Pointud et al. reported the first case of acute myeloid leukemia caused by methotrexate (MTX) in RA. They suggested that reporting all suspected cases could aid in evaluating the safety of MTX in rheumatology [20]. The JAK/STAT pathway is crucial in RA treatment, and inhibiting this pathway with specific JAK inhibitors can alleviate patient pain [20]. Lin et al. investigated the DNA methylation characteristics in fibroblast-like synovial cells (FLS) of RA patients, revealing that genes with hypomethylation sites were enriched in the neurotrophic factor signaling pathway [21] (see Fig. 4).

3.2. Screening and diagnostic model construction of critical mitochondrial autophagy genes

In this article, WGCNA analysis was applied to the full gene expression data of the GSE93272 data set. From Fig. 5A, it can be seen that the soft threshold power should be set to 14. We presented the generated dendrograms of the characteristic genes in Fig. 5B. The clustered dendrogram of genes is given in Fig. 5C. Nine modules were identified and shown in different colors by mean linkage clustering. Fig. 5D shows the correlation and significance heat map for the RA and the control two traits in the different modules. We eventually selected the genes in the blue module for subsequent analysis. This paper intersected the 2643 genes in the blue module with the intersected genes to obtain 13 new intersected genes (Fig. 5E). Furthermore, the RF algorithm was applied to obtain the nine genes for which the mean decrease accuracy was not 0 (Fig. 5F). The larger the mean decrease in accuracy, the more important the gene was in the RA diagnosis process.

Moreover, we further screened the nine genes based on the LASSO algorithm. The selection of the best penalty values during the
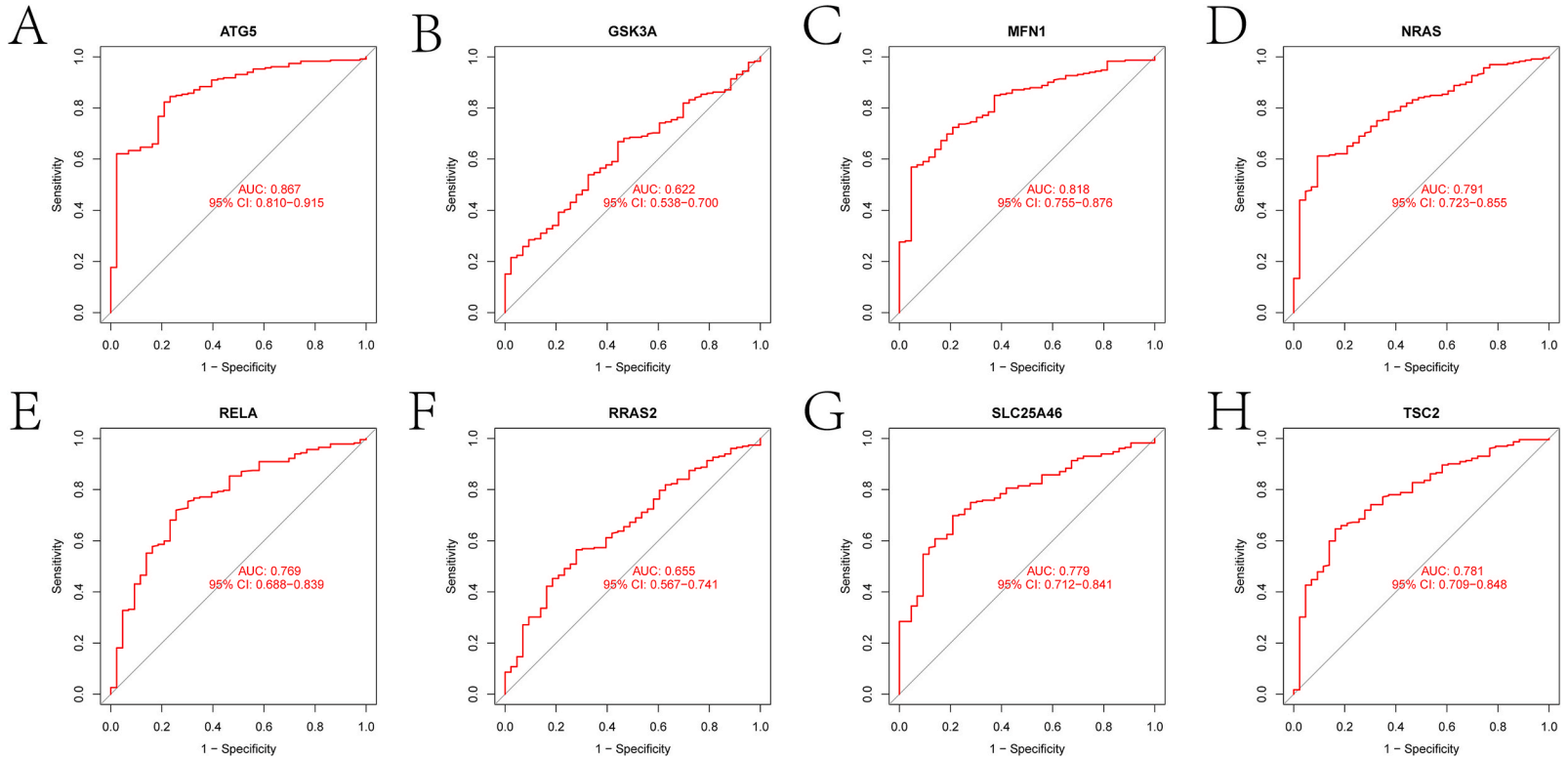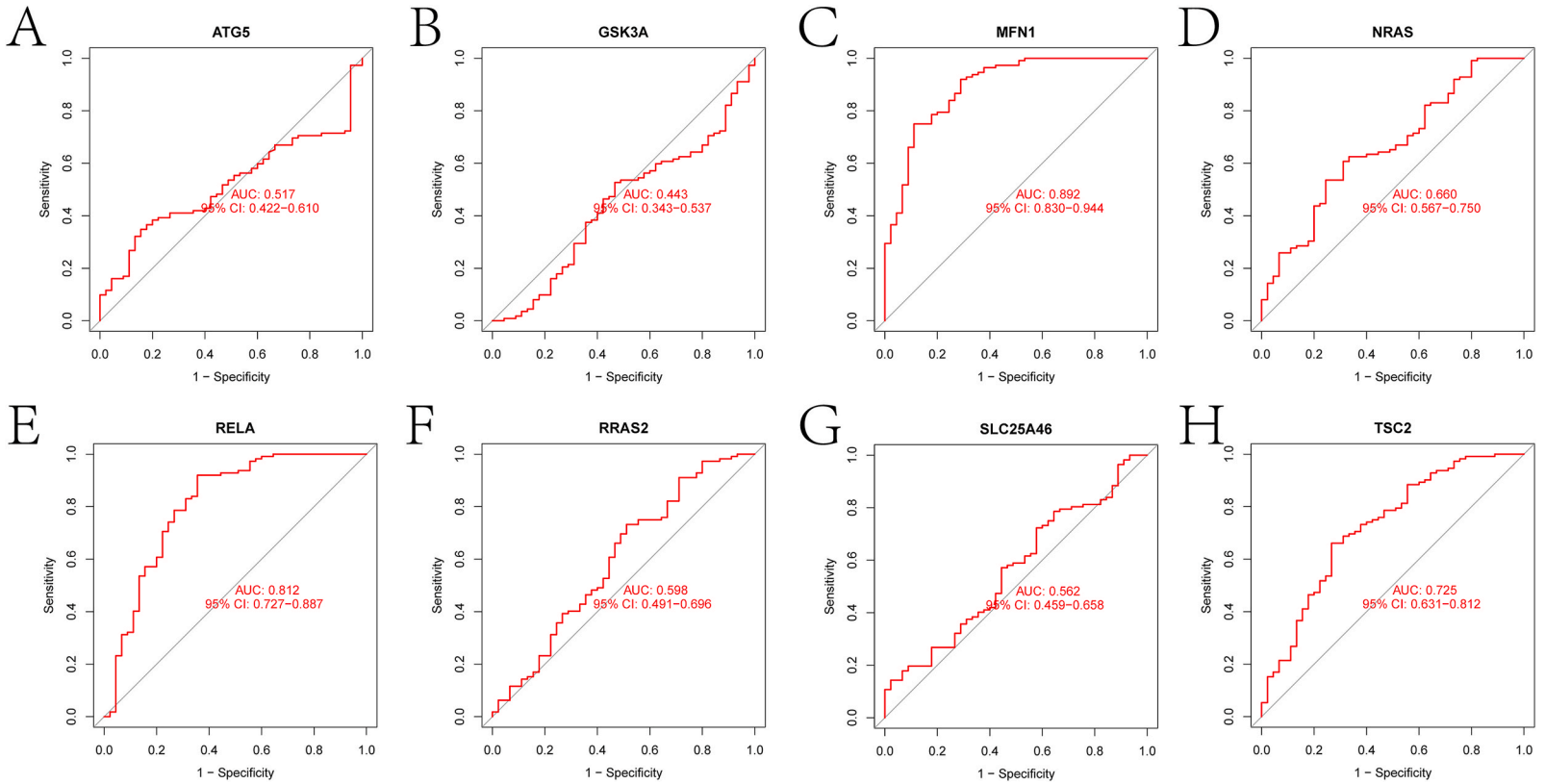
Fig. 7. ROC analysis of genes in the GSE93272 data set. A-H are ROC curves for ATG5, GSK3A, MFN1, NRAS, RELA, RRAS2, SLC25A46 and TSCA, respectively.
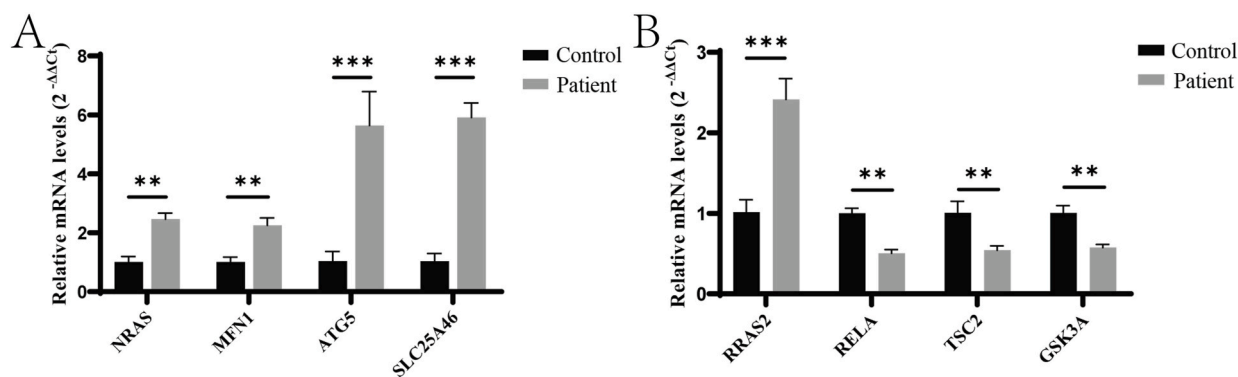
Fig. 8. ROC analysis of genes in the GSE17755 data set. A-H are ROC curves for ATG5, GSK3A, MFN1, NRAS, RELA, RRAS2, SLC25A46 and TSCA, respectively.s.

Fig. 9. The results of RT-qPCR (**P < 0.01, ***P < 0.001). A and B are bar graphs of hub gene expression in the control and diseased groups, respectively.

screening process using the LASSO algorithm is given in Fig. 6A and B. Ultimately, eight diagnostic genes (ATG5, GSK3A, MFN1, NRAS, RELA, RRAS2, SLC25A46 and TSCA) were retained. ATG5 is involved in the formation of autophagic vesicles, which may play an important role in the process of apoptosis. Petrasca et al. proved that CD4 T cells and FLS cooperated in initiating and maintaining the function of joint inflammation. They found that when CD4 T cells were activated, the expressions of six glycolytic enzymes, such as GSK3A in RA FLS were up-regulated [22]. RELA-related diseases include autoimmune diseases. The protein encoded by the MFN1 gene is the medium of mitochondrial fusion. Its related pathways include selective autophagy and mitochondrial autophagy. Previous literature has proved that mitochondrial autophagy is closely associated with RA [17–19]. RRAS2 participates in the regulation of the MAPK signal pathway [23]. 3′ 3′3-Diindolylmethane (DIM) can affect RA-FLS by blocking MAPK and AKT/mTOR pathways, which has great potential in the treatment of RA [24].

We constructed a diagnostic model for RA based on eight diagnostic genes and used the logistic regression algorithm. The ROC curves of the model in the GSE93272 and GSE93272 data sets are presented in Fig. 6C and D, respectively. The ROC in the GSE93272 data set was 0.916, and the ROC in the GSE93272 data set was 0.951, which indicated the high diagnostic accuracy of the constructed model. AUC validation of the eight diagnostic genes was performed on the GSE93272 and GSE93272 data sets in this paper. The validation results for the GSE93272 data set are given in Fig. 7A–H. All genes had an AUC greater than 0.6, and the validation results in the GSE17755 data set are given in Fig. 8A–H. Among them, the AUCs of MFN1, NRAS, RELA, and TSCA were greater than 0.6. In the end, the expression of the eight genes in both groups was further validated in this paper based on qPCR. We collected 5 serum samples from RA patients and 5 samples from normal people. As shown in Fig. 9, compared with the normal population, the expression levels of NRAS, MFN1, ATG5, SLC25A46, and RRAS2 genes in the serum of patients with RA increased significantly (P < 0.01). In contrast, the expression levels of RELA, TSC2, and GSK3A decreased (P < 0.01).

3.3. Immune microenvironment exploration and drug interactions network

In this paper, the CIBERSORT algorithm was applied to assess the infiltration abundance of multiple immune cells in the RA and control groups. Fig. 10A gives a plot of the percentage of infiltration abundance of each immune cell in all samples. A box plot of the difference in infiltration abundance of each immune cell in the RA and control groups is shown in Fig. 10B. As can be seen from the diagram, B cells memory, T cells CD8, T cells CD4 naïve, T cells regulatory (Tregs), T cells gamma delta, macrophages M0, macrophages M1, dendritic cells resting, dendritic cells activated and neutrophils were significantly different in both groups. This suggested that there was a large difference in the immune landscape between the RA and control groups. Recent studies have shown that the autophagy of neutrophils is enhanced. The autophagy rate of CD4 T cells is significantly increased in RA patients compared with healthy controls [25,26]. Autophagy is a key regulator of memory CD8$^+$ T cell formation [27]. Kumar et al. found that naïve and memory CD4+T cells in RA patients showed enhanced autophagy. They identified MYC as the critical regulator of autophagy memory, while inhibition of MYC was the core of autophagy memory and long-term persistence [28]. The dysfunction of Tregs is one of the mechanisms leading to the collapse of self-tolerance in RA progress [29]. Smith et al. found that, compared with the control group, the blood level of T gamma delta cells in RA patients was significantly decreased [30].

In addition, the correlation analysis allowed us to obtain scatter plots of the correlation between ATG5, GSK3A, MFN1, NRAS, RELA, RRAS2, SLC25A46, TSCA, and the abundance of immune cell infiltration. For each diagnostic gene, we presented in Fig. 10C-R the infiltration abundance of the two immune cells that were the most significantly associated with the diagnostic gene. Interestingly, T cell regulatory (Tregs) and T cell gamma delta were both significantly associated with multiple diagnostic genes. The rest of the results of the analysis of immune cell infiltration abundance and diagnostic genes are given in Figs. S2–S9 in the supplementary material.

The drugs that have interactions with diagnostic genes are predicted and visualized in this paper based on the DGIdb database (Fig. 11). Among them, NRAS, RELA, and GSK3A have interactions with a variety of drugs. Among the drugs in Fig. 11, VEMURAFENIB is a potential ligand against IFI44L. The analysis of the gene interaction network by Yadalam et al. reveals that IFI44L is the drug target of RA [31,32]. GEFITINIB is an epidermal growth factor receptor kinase inhibitor that has great potential in the treatment of RA [33].
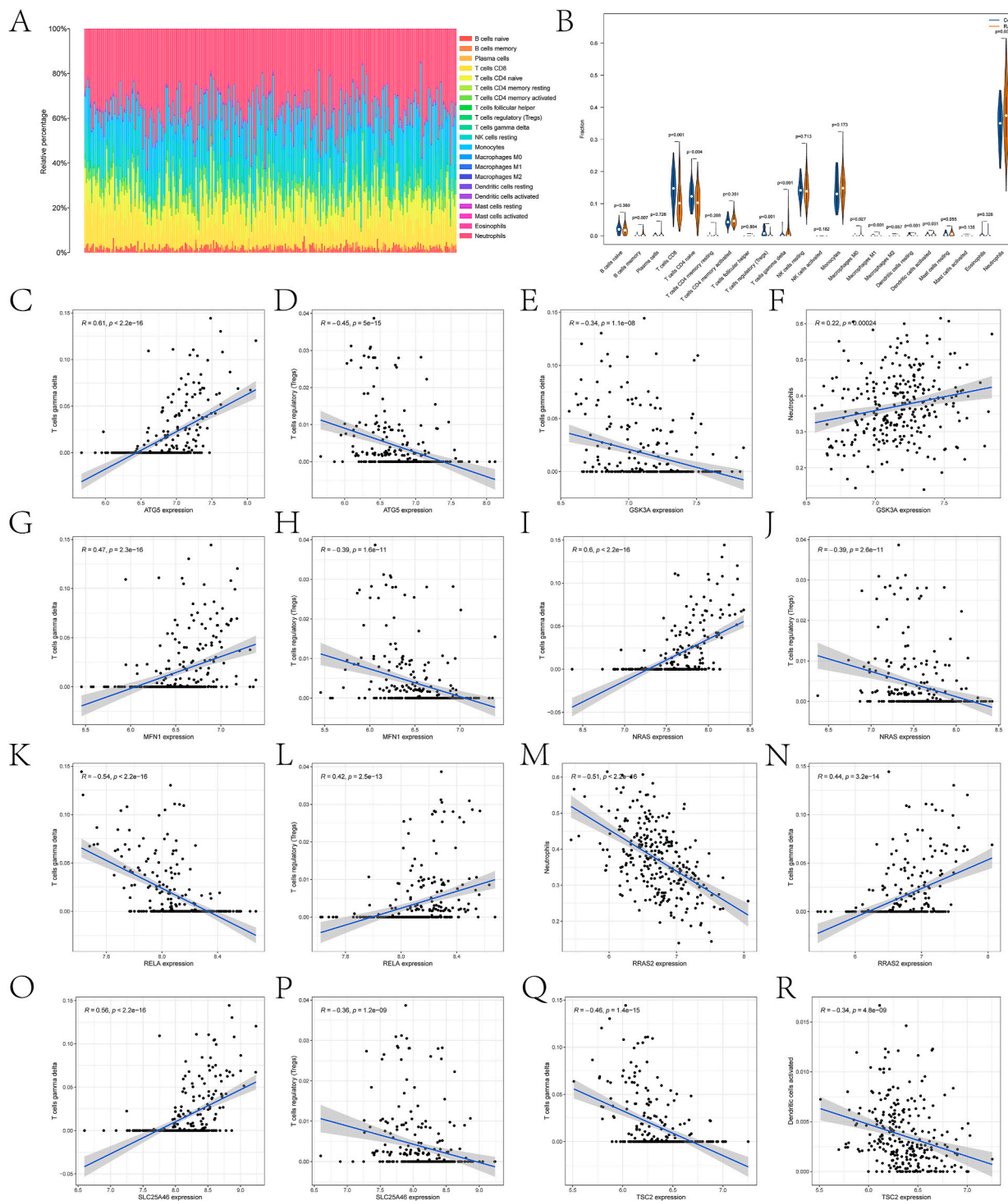
Fig. 10. Immune landscape of the RA and control groups. A shows the percentage of different immune cell infiltration abundance in all samples. B is a box plot of the difference in infiltration abundance of 22 types of immune in the RA and control groups. C–R are scatter plots obtained from correlation analysis of the abundance of ATG5, GSK3A, MFN1, NRAS, RELA, RRAS2, SLC25A46, TSCA, and immune cell infiltration. For each diagnostic gene, this paper shows the two immune cell infiltration abundances that are most significantly associated with it.
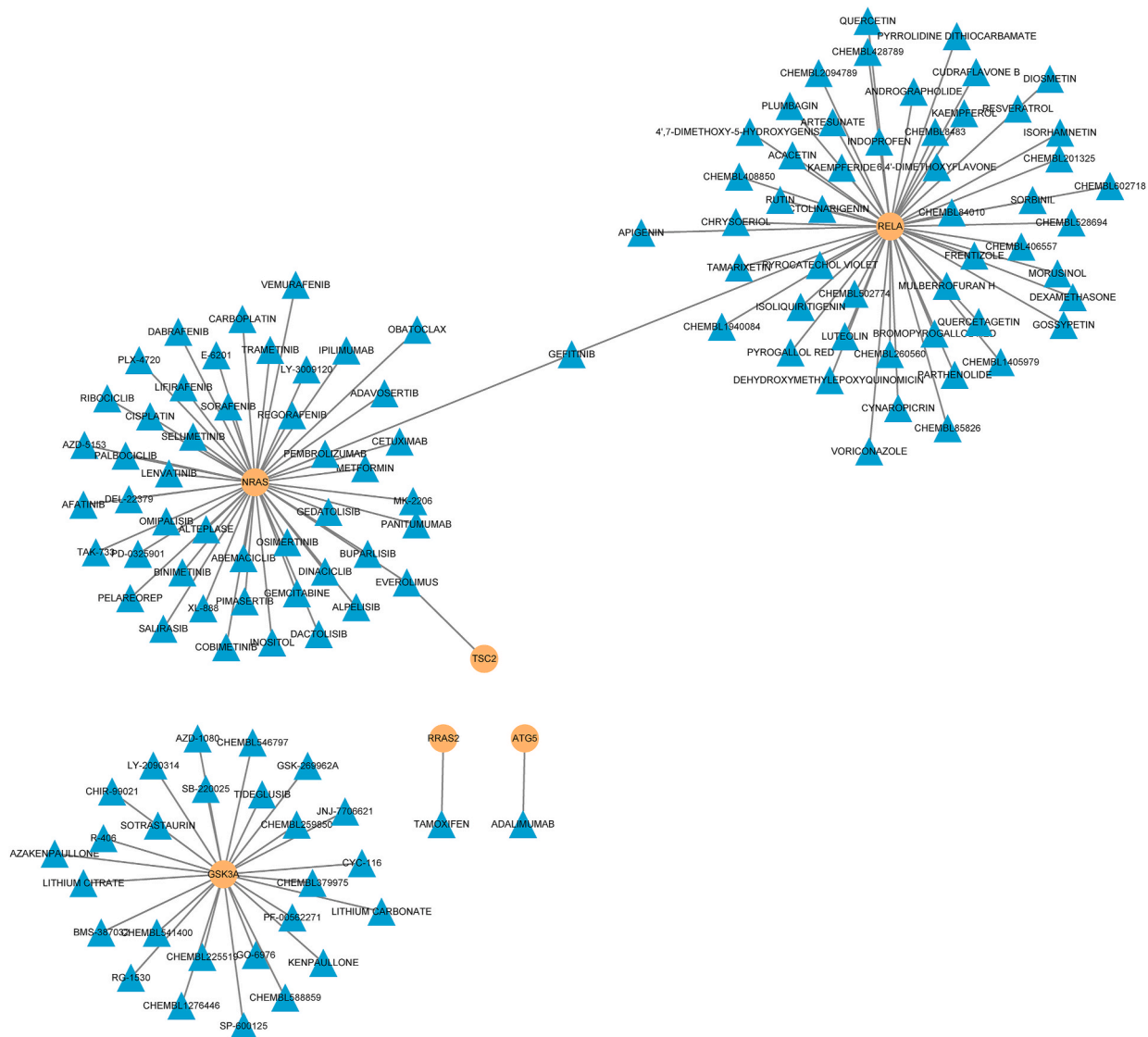
Fig. 11. Drug prediction results for diagnostic genes.

Inflammation and proliferation of synovial tissue are one of the characteristics of RA. DIOSMETIN shows anti-proliferation and anti-inflammatory effects by inhibiting Akt and NF-κB pathways in MH7A cells, thus achieving the purpose of treating RA [34].

4. Discussion

RA is a progressive inflammatory autoimmune disease that can lead to synovial hyperplasia and disability if the disease is not treated promptly. Mitochondrial autophagy is an essential biological process by which mitochondria control their own quantity and quality. This process has been shown to be closely related to tumorigenesis and progression. However, the role it plays in RA is unclear. Therefore, this article explores in detail the role of differentially expressed MRGs in RA from a bioinformatics perspective. We first performed GO enrichment analysis, metascape enrichment analysis, and GSEA analysis. Macroautophagy, hereinafter referred to as autophagy, is the main intracellular degradation system, which is characterized by delivering cytoplasmic substances to lysosomes [35].

Additionally, genes in the critical modules obtained by WGCNA analysis screening were intersected with differentially expressed MRGs, and genes associated with RA diagnosis were screened using the RF algorithm and LASSO algorithm for the intersected genes, respectively. Eight diagnostic genes (ATG5, GSK3A, MFN1, NRAS, RELA, RRAS2, SLC25A46, and TSCA) were eventually obtained.

We explored the abundance of multiple immune cell infiltrates in the RA and control groups with the help of the CIBERSORT algorithm. We found that the infiltration abundance of B cells memory, T cells CD8, T cells CD4 naïve, T cells regulatory (Tregs), T cells

gamma delta, macrophages M0, macrophages M1, dendritic cells resting, dendritic cells activated and neutrophils were significantly different in both groups. In conclusion, we explored drugs that have reciprocal relationships with diagnostic genes.

5. Conclusion

In conclusion, this study used multiple bioinformatics approaches to identify the role of mitochondrial autophagy-related genes in rheumatoid arthritis. The results suggest that eight genes (ATG5, GSK3A, MFN1, NRAS, RELA, RRAS2, SLC25A46, and TSCA) have an essential role in the diagnosis of RA and may be potential biomarkers for RA.

Data availability

The data in this paper come from GEO database (https://www.ncbi.nlm.nih.gov/geo/) GSE93272 and GSE17755.

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the University of Kiang Wu Hospital, Macau (approval number:2021–015). All participants were provided with a written informed consent form that was signed before participation in the trial.

CRediT authorship contribution statement

Iong Iok In: Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Weiming Deng:** Writing – review & editing, Validation, Supervision, Methodology.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24818.

References

[1] J.S. Smolen, D. Aletaha, I.B. McInnes, Rheumatoid arthritis, Lancet (London, England) 388 (2016) 2023–2038.
[2] D. van der Woude, A.H.M. van der Helm-van Mil, Update on the epidemiology, risk factors, and disease outcomes of rheumatoid arthritis, Best Pract. Res. Clin. Rheumatol. 32 (2018) 174–187.
[3] G.S. Firestein, I.B. McInnes, Immunopathogenesis of rheumatoid arthritis, Immunity 46 (2017) 183–196.
[4] P. Won, Y. Kim, H. Jung, Y.A. Rim, D.H. Sohn, W.H. Robinson, S.J. Moon, J.H. Ju, Pathogenic role of circulating citrullinated antigens and anti-cyclic monoclonal citrullinated peptide antibodies in rheumatoid arthritis, Front. Immunol. 12 (2021) 692242.
[5] Y. Aihaiti, Y. Song Cai, X. Tuerhong, Y. Ni Yang, Y. Ma, H. Shi Zheng, K. Xu, P. Xu, Therapeutic effects of naringin in rheumatoid arthritis: network pharmacology and experimental validation, Front. Pharmacol. 12 (2021) 672054.
[6] A. Levescot, M.H. Chang, J. Schnell, N. Nelson-Maney, J. Yan, M. Martínez-Bonet, R. Grieshaber-Bouyer, P.Y. Lee, K. Wei, R.B. Blaustein, et al., IL-1β-driven osteoclastogenic Tregs accelerate bone erosion in arthritis, J. Clin. Invest. (2021) 131.
[7] T. Kishimoto, S. Kang, IL-6 revisited: from rheumatoid arthritis to CAR T cell therapy and COVID-19, Annu. Rev. Immunol. 40 (2022) 323–348.
[8] H. Radner, D. Aletaha, Anti-TNF in rheumatoid arthritis: an overview, Wien Med. Wochenschr. 165 (1946) (2015) 3–9.
[9] Y. Gong, N. Tang, P. Liu, Y. Sun, S. Lu, W. Liu, L. Tan, C. Song, X. Qiu, Y. Liao, et al., Newcastle disease virus degrades SIRT3 via PINK1-PRKN-dependent mitophagy to reprogram energy metabolism in infected cells, Autophagy 18 (2022) 1503–1521.
[10] C. Yan, X. Duanmu, L. Zeng, B. Liu, Z. Song, Mitochondrial DNA: distribution, mutations, and elimination, Cells 8 (2019).
[11] K. Palikaras, E. Lionaki, N. Tavernarakis, Mechanisms of mitophagy in cellular homeostasis, physiology and pathology, Nat. Cell Biol. 20 (2018) 1013–1022.
[12] Y. Li, W. Li, A.R. Hoffman, J. Cui, J.F. Hu, The nucleus/mitochondria-shuttling LncRNAs function as new epigenetic regulators of mitophagy in cancer, Front. Cell Dev. Biol. 9 (2021) 699621.
[13] Y. Pei, S. Chen, F. Zhou, T. Xie, H. Cao, Construction and evaluation of Alzheimer's disease diagnostic prediction model based on genes involved in mitophagy, Front. Aging Neurosci. 15 (2023) 1146660.
[14] S. Chadha, T. Behl, S. Bungau, A. Kumar, R. Kaur, T. Venkatachalam, A. Gupta, M. Kandhwal, D. Chandel, Focus on the multimodal role of autophagy in rheumatoid arthritis, Inflammation 44 (2021) 1–12.
[15] E. Sugawara, M. Kato, Y. Kudo, W. Lee, R. Hisada, Y. Fujieda, K. Oku, T. Bohgaki, O. Amengual, S. Yasuda, et al., Autophagy promotes citrullination of VIM (vimentin) and its interaction with major histocompatibility complex class II in synovial fibroblasts, Autophagy 16 (2020) 946–955.
[16] H.R. Yun, Y.H. Jo, J. Kim, Y. Shin, S.S. Kim, T.G. Choi, Roles of autophagy in oxidative stress, Int. J. Mol. Sci. 21 (2020).
[17] 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, J. Inflamm. Res. 16 (2023) 1271–1282.
[18] C. Ploumi, I. Daskalaki, N. Tavernarakis, Mitochondrial biogenesis and clearance: a balancing act, FEBS J. 284 (2017) 183–195.
[19] B. Li, C. Yang, M. Guo, S. Wang, W. Peng, Q. Guo, D. Ming, Y. Teng, B. Zheng, Ultrasound-remote selected activation mitophagy for precise treatment of rheumatoid arthritis by two-dimensional piezoelectric nanosheets, ACS Nano 17 (2023) 621–635.

[20] N. Crispino, F. Ciccia, JAK/STAT pathway and nociceptive cytokine signalling in rheumatoid arthritis and psoriatic arthritis, Clin. Exp. Rheumatol. 39 (2021) 668–675.

[21] Y. Lin, Z. Luo, Aberrant methylation patterns affect the molecular pathogenesis of rheumatoid arthritis, Int. Immunopharm. 46 (2017) 141–145.

[22] A. Petrasca, J.J. Phelan, S. Ansboro, D.J. Veale, U. Fearon, J.M. Fletcher, Targeting bioenergetics prevents CD4 T cell-mediated activation of synovial fibroblasts in rheumatoid arthritis, Rheumatology 59 (2020) 2816–2828.

[23] Y. Capri, E. Flex, O.H.F. Krumbach, G. Carpentieri, S. Cecchetti, C. Lißewski, S. Rezaei Adariani, D. Schanze, J. Brinkmann, J. Piard, et al., Activating mutations of RRAS2 are a rare cause of noonan syndrome, Am. J. Hum. Genet. 104 (2019) 1223–1232.

[24] H. Du, X. Zhang, Y. Zeng, X. Huang, H. Chen, S. Wang, J. Wu, Q. Li, W. Zhu, H. Li, et al., A novel phytochemical, DIM, inhibits proliferation, migration, invasion and TNF-α induced inflammatory cytokine production of synovial fibroblasts from rheumatoid arthritis patients by targeting MAPK and AKT/mTOR signal pathway, Front. Immunol. 10 (2019) 1620.

[25] Q. An, W. Yan, Y. Zhao, K. Yu, Enhanced neutrophil autophagy and increased concentrations of IL-6, IL-8, IL-10 and MCP-1 in rheumatoid arthritis, Int. Immunopharm. 65 (2018) 119–128.

[26] J. van Loosdregt, M. Rossetti, R. Spreafico, M. Moshref, M. Olmer, G.W. Williams, P. Kumar, D. Copeland, K. Pischel, M. Lotz, et al., Increased autophagy in CD4(+) T cells of rheumatoid arthritis patients results in T-cell hyperactivation and apoptosis resistance, Eur. J. Immunol. 46 (2016) 2862–2870.

[27] X. Xu, K. Araki, S. Li, J.H. Han, L. Ye, W.G. Tan, B.T. Konieczny, M.W. Bruinsma, J. Martinez, E.L. Pearce, et al., Autophagy is essential for effector CD8(+) T cell survival and memory formation, Nat. Immunol. 15 (2014) 1152–1161.

[28] P. Kumar, L.J. Yao, S. Saidin, B. Paleja, J. van Loosdregt, C. Chua, T. Arkachaisri, A. Consolaro, M. Gattorno, A. Martini, et al., Molecular mechanisms of autophagic memory in pathogenic T cells in human arthritis, J. Autoimmun. 94 (2018) 90–98.

[29] Q. Jiang, G. Yang, Q. Liu, S. Wang, D. Cui, Function and role of regulatory T cells in rheumatoid arthritis, Front. Immunol. 12 (2021) 626193.

[30] M.D. Smith, B. Bröker, L. Moretta, E. Ciccone, C.E. Grossi, J.C. Edwards, F. Yüksel, B. Colaco, C. Worman, L. Mackenzie, et al., T gamma delta cells and their subsets in blood and synovial tissue from rheumatoid arthritis patients, Scand. J. Immunol. 32 (1990) 585–593.

[31] H. Zhou, H. Cao, J. Skolnick, FINDSITE(comb 2.0): a new approach for virtual ligand screening of proteins and virtual target screening of biomolecules, J. Chem. Inf. Model. 58 (2018) 2343–2354.

[32] P.K. Yadalam, T. Sivasankari, S. Rengaraj, M.H. Mugri, M. Sayed, S.S. Khan, M.A. Kamil, S. Bhandi, A.T. Raj, S. Patil, et al., Gene Interaction Network Analysis Reveals IFI44L as a Drug Target in Rheumatoid Arthritis and Periodontitis, vol. 27, Molecules, Basel, Switzerland, 2022.

[33] M.B. Brooks, Erlotinib and gefitinib, epidermal growth factor receptor kinase inhibitors, may treat non-cancer-related tumor necrosis factor-α mediated inflammatory diseases, Oncol. 18 (2013) e3–e5.

[34] Y. Chen, Y. Wang, M. Liu, B. Zhou, G. Yang, Diosmetin exhibits anti-proliferative and anti-inflammatory effects on TNF-α-stimulated human rheumatoid arthritis fibroblast-like synoviocytes through regulating the Akt and NF-κB signaling pathways, Phytother Res. : PTR 34 (2020) 1310–1319.

[35] N. Mizushima, M. Komatsu, Autophagy: renovation of cells and tissues, Cell 147 (2011) 728–741.