ORIGINAL RESEARCH

Identification of Mitophagy-Associated Genes for the Prediction of Metabolic Dysfunction-Associated Steatohepatitis Based on Interpretable Machine Learning Models

Beiying Deng[1,2,*], Ying Chen[1,2,*], Pengzhan He[1,2,*], Yinghui Liu[3], Yangbo Li[1,2], Yuli Cai[4], Weiguo Dong [1]

[1]Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, People's Republic of China; [2]Central Laboratory, Renmin Hospital of Wuhan University, Wuhan, People's Republic of China; [3]Department of Geriatric, Renmin Hospital of Wuhan University, Wuhan, People's Republic of China; [4]Department of Endocrinology, Renmin Hospital of Wuhan University, Wuhan, People's Republic of China

*These authors contributed equally to this work

Correspondence: Weiguo Dong, Department of Gastroenterology, Renmin Hospital of Wuhan University, No. 99 Zhangzhidong Road, Wuhan, Hubei Province, 430060, People's Republic of China, Email dongweiguo@whu.edu.cn

Background: This study aims to elucidate the role of mitochondrial autophagy in metabolic dysfunction-associated steatohepatitis (MASH) by identifying and validating key mitophagy-related genes and diagnostic models with diagnostic potential.

Methods: The gene expression profiles and clinical information of MASH patients and healthy controls were obtained from the Gene Expression Omnibus database (GEO). Limma and functional enrichment analysis were used to identify the mitophagy-related differentially expressed genes (mito-DEGs) in MASH patients. Machine learning models were used to select key mito-DEGs and evaluate their efficacy in the early diagnosis of MASH. The expression levels of the key mito-DEGs were validated using datasets and cell models. A nomogram was constructed to assess the risk of MASH progression based on the expression of the key mito-DEGs. The mitophagy-related molecular subtypes of MASH were evaluated.

Results: Four mito-DEGs, namely MRAS, RAB7B, RETREG1, and TIGAR were identified. Among the machine learning models employed, the Support Vector Machine demonstrated the highest AUC value of 0.935, while the Light Gradient Boosting model exhibited the highest accuracy (0.9189), kappa (0.7204), and F1-score (0.9508) values. Based on these models, MRAS, RAB7B, and RETREG1 were selected for further analysis. The logistic regression model based on these genes could accurately predict MASH diagnosis. The nomogram model based on these DEGs exhibited excellent prediction performance. The expression levels of the three mito-DEGs were validated in the independent datasets and cell models, and the results were found to be consistent with the findings obtained through bioinformatics analysis. Furthermore, our findings revealed significant differences in gene expression patterns, immune characteristics, biological functions, and enrichment pathways between the mitophagy-related molecular subtypes of MASH. Subtype-specific small-molecule drugs were identified using the CMap database.

Conclusion: Our research provides novel insights into the role of mitophagy in MASH and uncovers novel targets for predictive and personalized MASH treatments.

Keywords: metabolic dysfunction-associated steatohepatitis, mitophagy, biomarkers, diagnostic model, machine learning

Introduction

Metabolic-associated fatty liver disease (MAFLD), previously known as non-alcoholic fatty liver disease, is characterized by an excessive buildup of triacylglycerol (TG) in the liver. It affects approximately 25% of the adult population and has an estimated global prevalence of 32.4%.[1,2] Metabolic dysfunction-associated steatohepatitis (MASH), also called non-alcoholic steatosis (NASH), is characterized by hepatocyte ballooning, necrotizing inflammation, and steatosis.[3–5] Approximately 30%

of MAFLD patients develop MASH, and prolonged MASH can lead to fibrosis, cirrhosis, hepatocellular cancers, and liver failure.[6] Therefore, a thorough comprehension of the pathological and molecular attributes of MASH is imperative for advancing research on this complex ailment and improving its clinical management.

Mitochondria is an essential organelle involved in maintaining cellular homeostasis by playing crucial roles in signaling, biosynthesis, bioenergetics, and dynamics.[7] However, impaired or dysfunctional mitochondria can generate deleterious reactive oxygen species (ROS), which can lead to the development of various diseases, including neurode-generative disorders, metabolic maladies, and cancer.[8] Mitophagy eliminates dysfunctional mitochondria by selective autophagy, thereby ensuring mitochondrial quality and contributing to cellular homeostasis.[9] However, disruption in the mitophagy pathway can lead to aberrant mitochondrial function. Mitochondrial dysfunction can lead to disease progression through various processes, including high and low oxidation of fatty acids (FAs), accumulation of lipotoxic intermediates, and NADH-reductive stress.[10,11] Mitochondrial dysfunction is associated with MAFLD pathophysiology and may be responsible for the "second hit" phenomenon that is often associated with the progression of MAFLD to MASH.[12] Genetic investigations in mice have indicated that the selective absence of mitophagy within the liver is associated with an accelerated advancement of hepatic disorder.[13,14] Therefore, determining the key mitophagy-related regulatory genes holds promise for uncovering diagnostic indicators for MASH. Furthermore, the development of corresponding diagnostic models can effectively guide clinical diagnosis and broaden the range of treatments for MASH. However, research on the interaction between mitophagy-related genes and MASH is limited.

In this study, we aimed to identify key mitophagy-related genes with diagnostic potential for MASH and establish a diagnostic model, thus offering insights into the role of mitophagy in MASH progression.

Materials and Methods

Data Acquisition and Pre-Processing

The gene expression data and associated clinical data were obtained from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo). In this study, we utilized a collection of 137 mitophagy-related genes, sourced from previously published literature.[15] Details of these genes have been provided in Table S1.

Differential Gene Expression and Gene Enrichment Analyses

The Wilcoxon's rank-sum test was used to identify the differentially expressed genes (DEGs) between MASH and healthy control samples. The "ggplot2" and "pheatmap" R packages were used to visualize the results of the DEG analysis. Thereafter, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses were conducted using the R packages "clusterProfiler" and "org.Hs.eg.db" to determine the underlying biological mechanisms associated with these DEGs.[16] The mitophagy-related DEGs (mito-DEGs) were identified from the inter-section of the DEGs and mitophagy-related genes using a Venn diagram. Lastly, the receiver operating characteristic (ROC) curve was generated using the "pROC" R program to assess the diagnostic ability of the mito-DEGs for MASH.

Assessment of Immune Cell Infiltration in MASH

The single-sample gene set enrichment analysis (ssGSEA) was used to assess immune cell infiltration in the liver tissues of MASH patients and healthy controls.

Identification of Feature Biomarkers Using Machine Learning (ML) Models

ML stands as a facet of artificial intelligence concentrated on the formulation of algorithms and models facilitating computer systems to assimilate knowledge from data and formulate predictions or decisions devoid of explicit programming. This process entails the creation of mathematical models capable of enhancing their efficacy progressively through experiential learning.

A total of 368 samples (69 healthy controls and 299 MASH) were obtained, which were randomly assigned into training (70%) and test (30%) sets. Thereafter, eight different ML models, namely AdaBoost, Light Gradient Boosting (LightGBM), Decision Tree (DT), eXtreme Gradient Boosting (XGBoost), Random Forest (RF), K-Nearest Neighbors

(KNN), Naïve Bayes, and Support Vector Machines (SVM), were built using genes as variables. Subsequently, a grid search based on 10-fold cross-validation was used to identify the optimal set of hyperparameters for each model.

Several parameters were computed and compared in the test set, including accuracy, F1 score, Matthews correlation coefficient (MCC), true positive, true negative, false positive, and false negative, to assess the performance of the ML models. In addition, the area under the ROC curve (AUC) was computed to evaluate the capacity of the mito-DEGs to discriminate between healthy controls and MASH. The AdaBoost, DT, and KNN were performed using the R package "caret". LightGBM was performed using the R packages "lightgbm" and "tidymodels". RF and XGBoost were performed using the R packages "ranger" and "xgboost", respectively. Lastly, Naïve Bayes and SVM were performed using the R packages "e1071". Models were interpreted using the R package "DALEX".

Construction and Evaluation of Binary Logistic Regression (LR) Model

The binary LR model was used to find the matching coefficients of mito-DEGs and compute the risk score for each sample. The "rms" program was used to produce the nomogram. Decision curve analysis (DCA), calibration curve analysis, and AUC values were to assess the discriminatory performance of the scores. The GSE135251 dataset, from the GEO database, was used as an external validation set to evaluate the generalizability of the model.

Real-Time Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) Assay

The L-O2 cells (Purchased from Punosai Company, Wuhan, China) were cultivated in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and maintained at 37 °C and 5% CO_2. The cells were treated with oleic acid and palmitic acid to generate the MASH cell model.[17] Total RNA was isolated from the cells using a Trizol reagent (Invitrogen) and reverse transcription was performed using the HiScript III RT SuperMix (+ gDNA wiper) kit (Vazyme, R223-01). The qRT-PCR assay was conducted using the AceQ® qPCR SYBR Green Master Mix (Vazyme) to assess the relative mRNA levels. The relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method, with β-actin as the internal reference. The primers used in the qRT-PCR assay are listed in Table S2.

Unsupervised Clustering Analysis and Gene Set Variation Analysis (GSVA)

An unsupervised clustering analysis was performed using the "ConsensusClusterPlus" R package.[18] The k-means approach was run 1000 times with a maximum subtype number of k (k = 6). The optimal subtype number was ascertained using the cumulative distribution function curve, consensus matrix, and consistent cluster score.

The gene set variation analysis (GSVA) was conducted using the "GSVA" R package[19] to assess the variations in the enriched functions and pathways between various mitophagy-related subtypes. The "c2.cp.kegg.v7.4.symbols" and "c5.go.bp.v7.5.1.symbols" gene sets from the MSigDB database were used as the input files. Functions and pathways with a GSVA score $|t| > 2$ were considered to be significantly enriched.

Connectivity Map and Mechanism of Action Analysis

The Connectivity Map (CMap, https://clue.io/) database was used to identify potential small-molecule drugs that target mito-DEGs in the MASH subtypes 1 and 2.[20,21] The top 150 upregulated and downregulated subtype-specific genes were selected as the query in the CMap database. Small-molecule drugs with a compound enrichment score < −98 were chosen for further analysis (Table S3).

Statistical Analysis

The R software (v4.2.3, https://www.r-project.org) was used to conduct all statistical analyses in this study. Two groups were compared using Wilcoxon's rank-sum test, Chi-squared test, and Fisher's exact test, and multiple groups were compared using the Kruskal–Wallis test. The Spearman correlation analysis was performed using the R packages "ggpubr" and "stats". A p-value or adjusted p-value < 0.05 was considered statistically significant.

Results

Identification of DEGs in MASH Patients and Functional Analysis

A total of 767 DEGs (245 downregulated and 522 upregulated) were identified between healthy controls and MASH patients in the GSE213621 dataset at significance thresholds of absolute mean difference < 0.5 and an adjusted p-value < 0.05 (Figure 1A and B). GO enrichment analysis revealed that these DEGs were significantly enriched in biological processes associated with the organization of extracellular matrix, extracellular structure, and external encapsulating
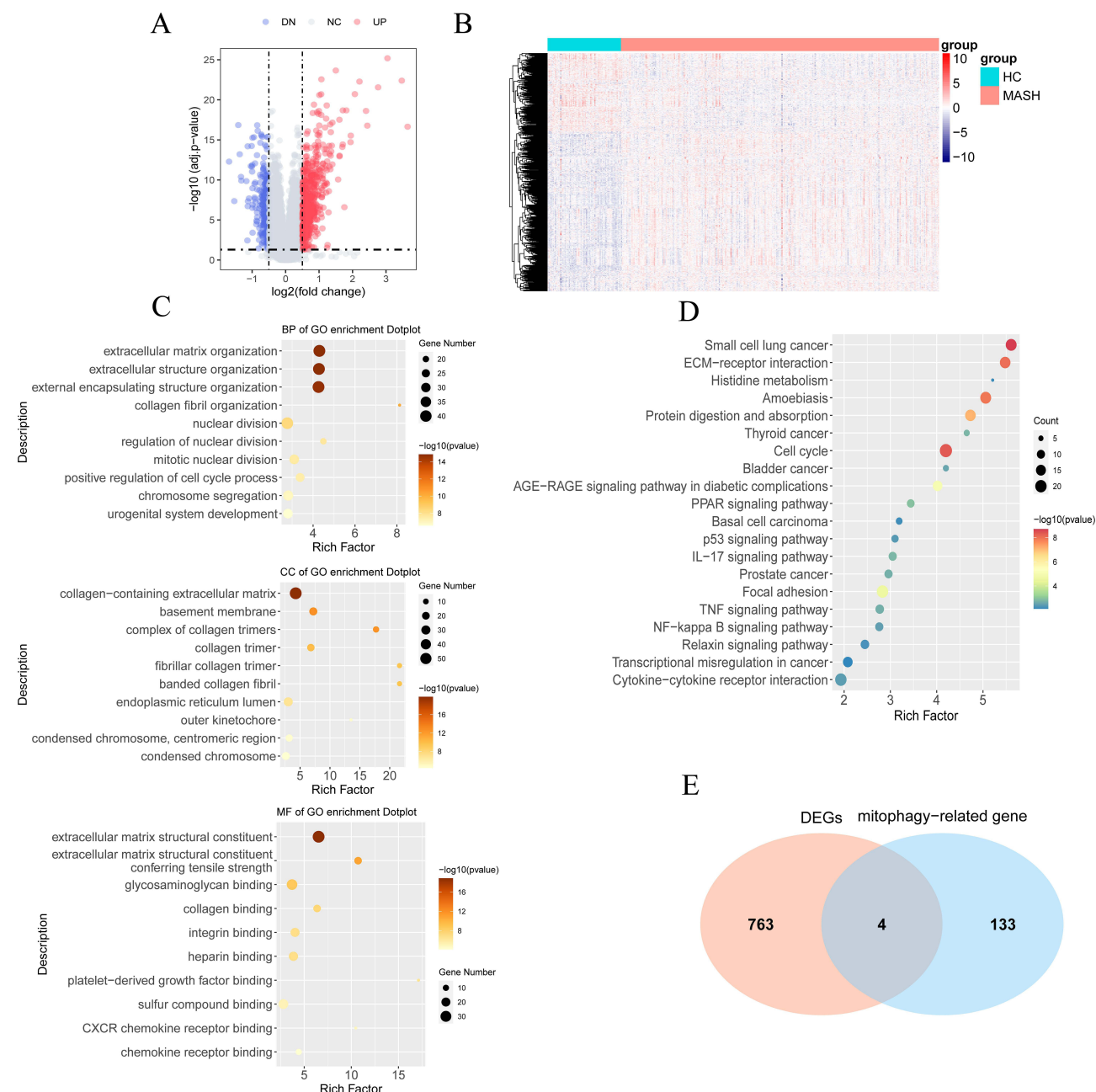
Figure 1 Identification of the DEGs in MASH. (**A**) Volcano map showing the DEGs in the liver tissues of MASH patients and healthy controls, red dots indicate the up-regulated genes, blue dots represent down-regulated genes, and gray dots indicate non-significant genes. (**B**) Heatmap showing the DEG expression profiles of healthy controls and MASH patients. (**C**) GO analyses of the DEGs in MASH. (**D**) KEGG enrichment analyses of the DEGs in MASH, the size of dots indicates the gene number and the shade of color indicate the scale of -log10(pvalue). (**E**) Venn diagram of the DEGs in MASH and mitophagy-related genes.
Abbreviations: DEGs, differentially expressed genes; MASH, metabolic dysfunction-associated steatohepatitis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

structure (Figure 1C). Furthermore, KEGG analysis revealed that the DEGs were associated with extracellular matrix–receptor interaction, histidine metabolism, amoebiasis, cell cycle, AGE-RAGE signaling in diabetic complications, PPAR signaling, p53 signaling, and IL-17 signaling (Figure 1D). Furthermore, we identified four mito-DEGs, namely MRAS, RAB7B, RETREG1, and TIGAR, from the intersection of the DEGs with mitophagy-related genes (Figure 1E).

Immune-Infiltrating Landscape of MASH Patients

To investigate differences in the immune microenvironment of MASH patients and healthy controls, we assessed the distribution and immunological activities of immune cells in the two groups (Figure 2A). Variations in the number of immune cells between the two groups and the correlation between the 28 immune cell types are shown in Figure 2B and C. Our results revealed that MASH patients had higher infiltration of activated CD4 T cells, central memory CD4 T cells,
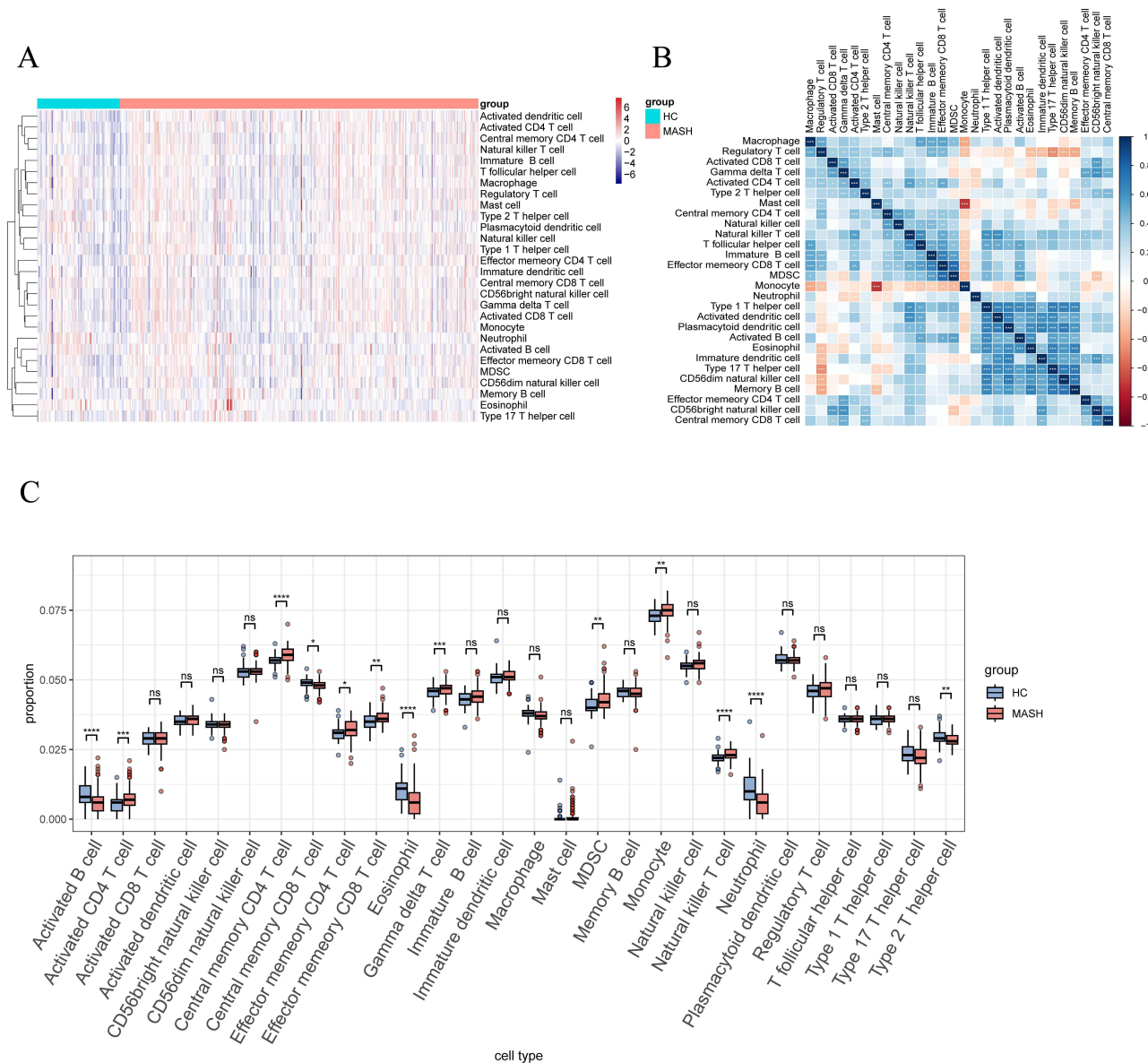

Figure 2 The immune characteristics of the MASH patients and healthy controls in the GSE213621 dataset. (**A**) Heatmap showing the distribution of 28 infiltrating-immune cells in the MASH patients and healthy controls. (**B**) Heatmap showing the relationship between the infiltrating-immune cells, the red indicate negative correlation and blue color indicates positive correlation. The shade of the color indicates the strength of the correlation. (**C**) Boxplot showing the variations in immune cell distribution in the MASH patients compared to the healthy controls. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.
Abbreviations: ns, not statistically significant; mito-DEGs, mitophagy-related DEGs; HC, healthy controls; MASH, metabolic dysfunction-associated steatohepatitis.

effector memory CD4 T cells, gamma/delta (γ/δ) T cells, myeloid-derived suppressor cells (MDSCs), monocytes, and natural killer T cells in the liver tissues, compared to the healthy controls. In contrast, MASH patients showed decreased numbers of activated B cells, central memory CD8 T cells, eosinophils, neutrophils, and type 2 T helper cells.

Subsequently, we examined the association between the infiltrating immune cells and the four mito-DEGs (Figure 3). The results revealed that MRAS (Figure 3A) and RAB7B (Figure 3B) showed the highest correlation with central memory CD4 T cells (r = 0.418 and 0.537, respectively) and eosinophils (r = −0.539 and −0.471, respectively); RETREG1 showed the highest correlation with type 2 T helper cells (r = 0.325) and MDSCs (r = −0.340) (Figure 3C); and TIGAR showed the highest correlation with regulatory T cells (r = 0.356) and eosinophils (r = −0.565) (Figure 3D). These findings contribute to our understanding of the complex interactions between four genes and immune cell populations in the pathogenesis of MASH.

Identification of Feature Biomarkers Using ML Models

We employed AdaBoost, Naïve Bayes, DT, KNN, LightGBM, RF, SVM, and XGBoost ML models to identify the feature genes in MASH. The performance metrics of each model are presented in Table 1. All the ML models exhibited recall rates > 50%. The ROC curves for each model are presented in Figure 4, and the importance plot of the four mito-DEGs across
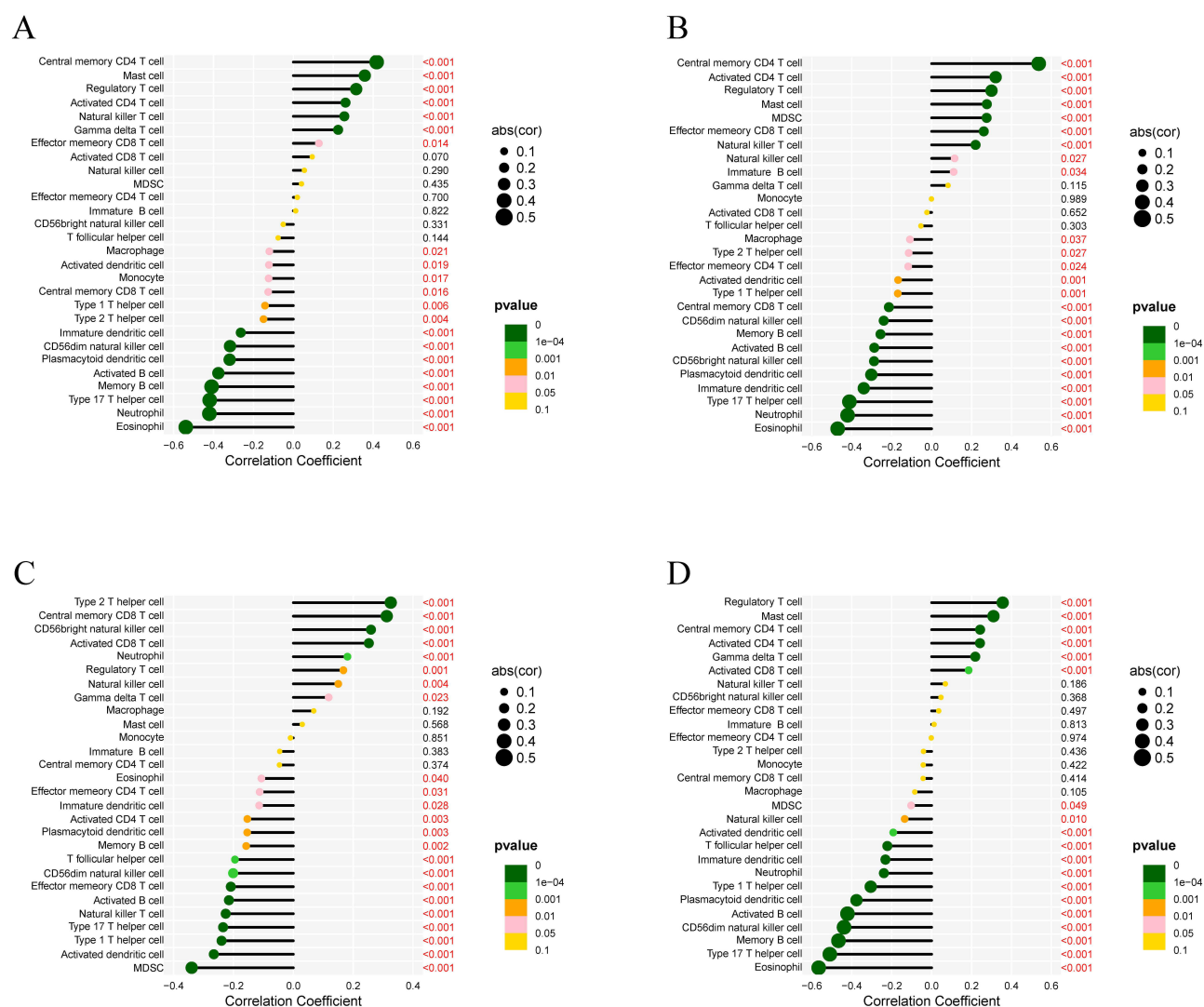

Figure 3 Correlation analysis of the four candidate mito-DEGs and the infiltrating immune cells in MASH. Correlation between the expression of MRAS (**A**), RAB7B (**B**), RETREG1 (**C**), and TIGAR (**D**) and infiltrating immune cells. The size of dots indicates the absolute value of the correlation, and the color indicate the scale of *p* value.

Table 1 Comparison of the Diagnostic Efficacy of the Eight Distinct Machine Learning Models

ML Models	TP	TN	FP	FN	Accuracy	Kappa	Precision	Recall	F1-Score	AUC
AdaBoost	87	13	3	8	0.9009	0.6445	0.9667	0.9158	0.9405	0.913
LightGBM	87	15	3	6	0.9189	0.7204	0.9667	0.9355	0.9508	0.917
DT	88	7	2	14	0.8559	0.3984	0.9778	0.8627	0.9167	0.749
XGBoost	86	11	4	10	0.8739	0.5383	0.9556	0.8958	0.9247	0.888
RF	86	10	4	11	0.8649	0.495	0.9556	0.8866	0.9198	0.910
KNN	88	7	2	14	0.8559	0.3984	0.9778	0.8627	0.9167	0.925
SVM	86	12	4	9	0.8829	0.5799	0.9556	0.9053	0.9297	0.935
Naïve Bayes	84	12	6	9	0.8649	0.534	0.9333	0.9032	0.9180	0.925

Abbreviations: ML, machine learning; TP, true positive; TN, true negative; FP, false positive; FN, false negative; AUC, area under the receiver operating characteristic curve.

the different models is presented in Figure 5. Among these models, SVM exhibited the highest AUC (0.935), whereas the LightGBM demonstrated the highest accuracy (0.9189), kappa (0.7204), and F1-score (0.9508). Partial dependence plots of the LightGBM and SVM were generated to analyze the influence of the four mito-DEGs on the predicted MASH risk (Figure 6). The graph demonstrated that elevated levels of MRAS and RAB7B and decreased levels of RETREG1 had a greater impact on the predicted outcome and increased the likelihood of MASH development.

Construction and Evaluation of the LR Model

Based on the variable importance analysis, the top three mito-DEGs shared by the SVM and LightGBM models were selected as the key mito-DEGs. Therefore, the LR model was constructed using the key mito-DEGs MRAS, RAB7B, and RETREG1, with the final risk score calculated as $(2.9289 \times MRAS) + (2.2171 \times RAB7B) + (-0.9873 \times RETREG1)$. In the training set, the risk scores of MASH patients were substantially greater than those of the healthy controls (Figure 7A). The nomogram based on the LR model is shown in Figure 7B.

The fitting effect of the model was validated using 1000 bootstrap samples in the training set. Calibration curves demonstrated the predictive accuracy of the LR model for the training set (Figure 7C). The DCA showed that the model-based decisions offered benefits to MASH patients, with the red line positioned consistently above the gray line (Figure 7D). The LR model achieved an AUC of 0.960 through 10-fold cross-validation (Figure 7E).

The calibration and DCA curves exhibited strong diagnostic prediction ability (Figure 8A and B), with an AUC value of 0.930 (Figure 8C). Additionally, the risk scores of MASH patients were significantly higher than those of the healthy controls in the test set (Figure 8D). External validation using the GSE135251 dataset revealed that the risk score of MASH patients was significantly higher (Figure 8E). The LR model obtained an AUC value of 0.959 for the validation set (Figure 8H). The calibration curves, and DCA further exhibited the robust diagnostic prediction ability of the LR model (Figure 8F and G).

Verification of Feature Biomarkers and Effectiveness Evaluation

ROC analysis performed on the GSE213621 cohort showed AUC values of 0.885, 0.884, and 0.700 for MRAS, RAB7B, and RETREG1, respectively (Figure 9A–C), suggesting that these three feature biomarkers can be used for the clinical diagnosis of MASH.

Correlation analysis between the expression levels of the three key mito-DEGs and the fibrosis stage (a critical prognostic feature of MASH) revealed that MRAS and RAB7B were significantly up-regulated in the F2 and F3F4 groups of MASH patients compared to the F0F1 group and the healthy controls (Figure 9D–F). Additionally, there was no significant variation in their expression levels between the F2 and F3F4 groups. In contrast, the expression level of RETREG1 was down-regulated in the F3F4 group compared to the F0F1 group.

The expression pattern of the three key mito-DEGs was also examined in the GSE135251 validation dataset. The results revealed that the expression levels of MRAS and RAB7B were higher, while that of RETREG1 was lower in the MASH group in the GSE135251 dataset (Figure 10D). Additionally, ROC analysis revealed AUC values of 0.944, 0.855,
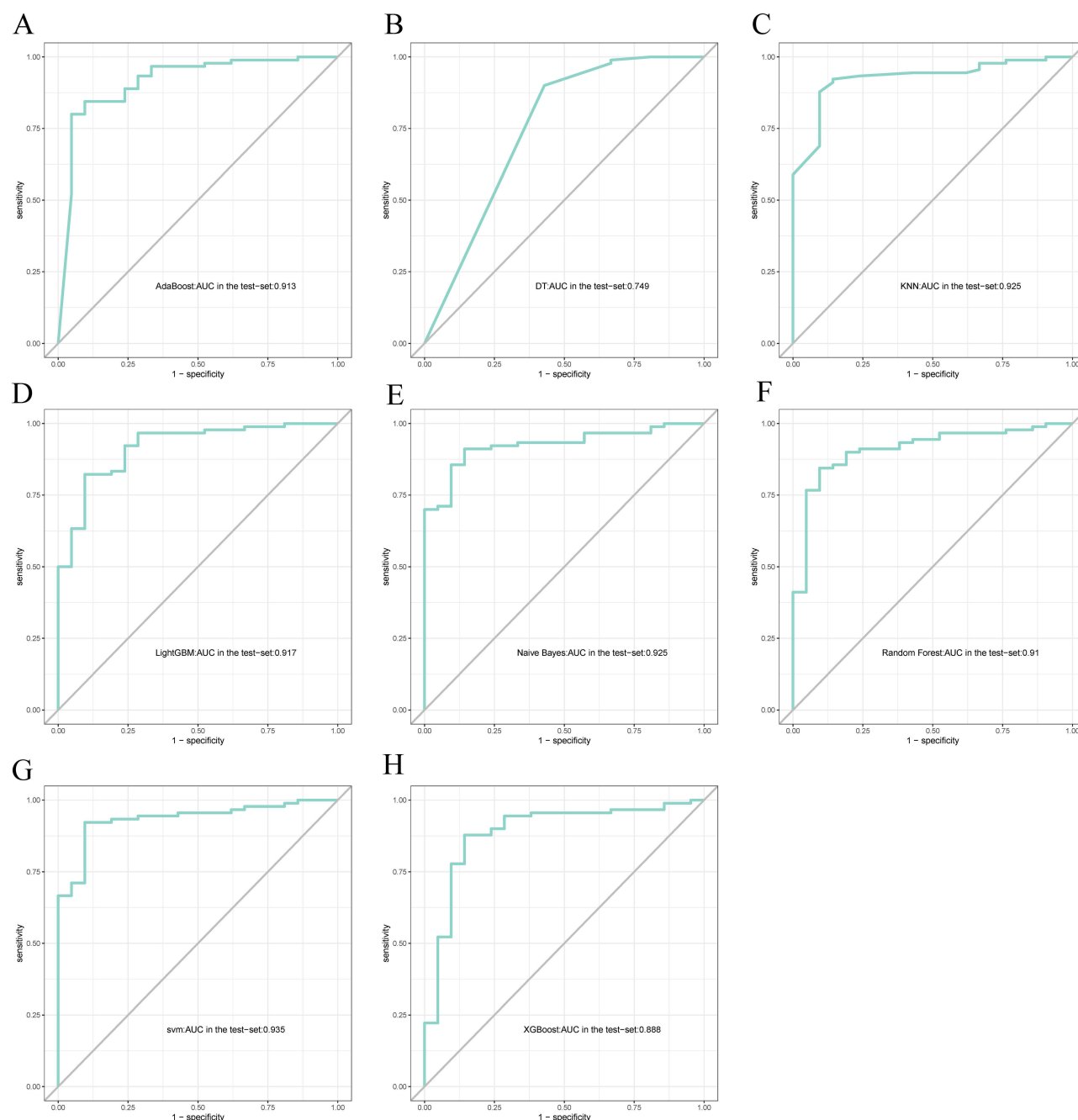
Figure 4 Evaluation of the eight machine learning models based on the area under the ROC curve (AUC). (**A**) AdaBoost; (**B**) Decision Tree; (**C**) KNN; (**D**) LightGBM; (**E**) Naïve Bayes; (**F**) RF; (**G**) SVM; and (**H**) XGBoost.
Abbreviations: KNN, K-nearest neighbors; LightGBM, Light Gradient Boosting; RF, Random Forest; SVM, support vector machines; XGBoost, eXtreme Gradient Boosting; AUC, area under the receiver operating characteristic curve.

and 0.745 for MRAS, RAB7B, and RETREG1 in the validation cohort (Figure 10A–C). Furthermore, our results demonstrated that MRAS and RAB7B were significantly up-regulated in the F0F1, F2, and F3F4 groups of the MASH patients compared to the healthy controls in the GSE135251 dataset (Figure 10E). In contrast, RETREG1 expression was lower in the F0F1, F2, and F3F4 groups of the MASH patients, compared to the healthy controls. However, the F0F1 and F3F4 groups showed significant variations in the expression levels of MRAS and RAB7B.
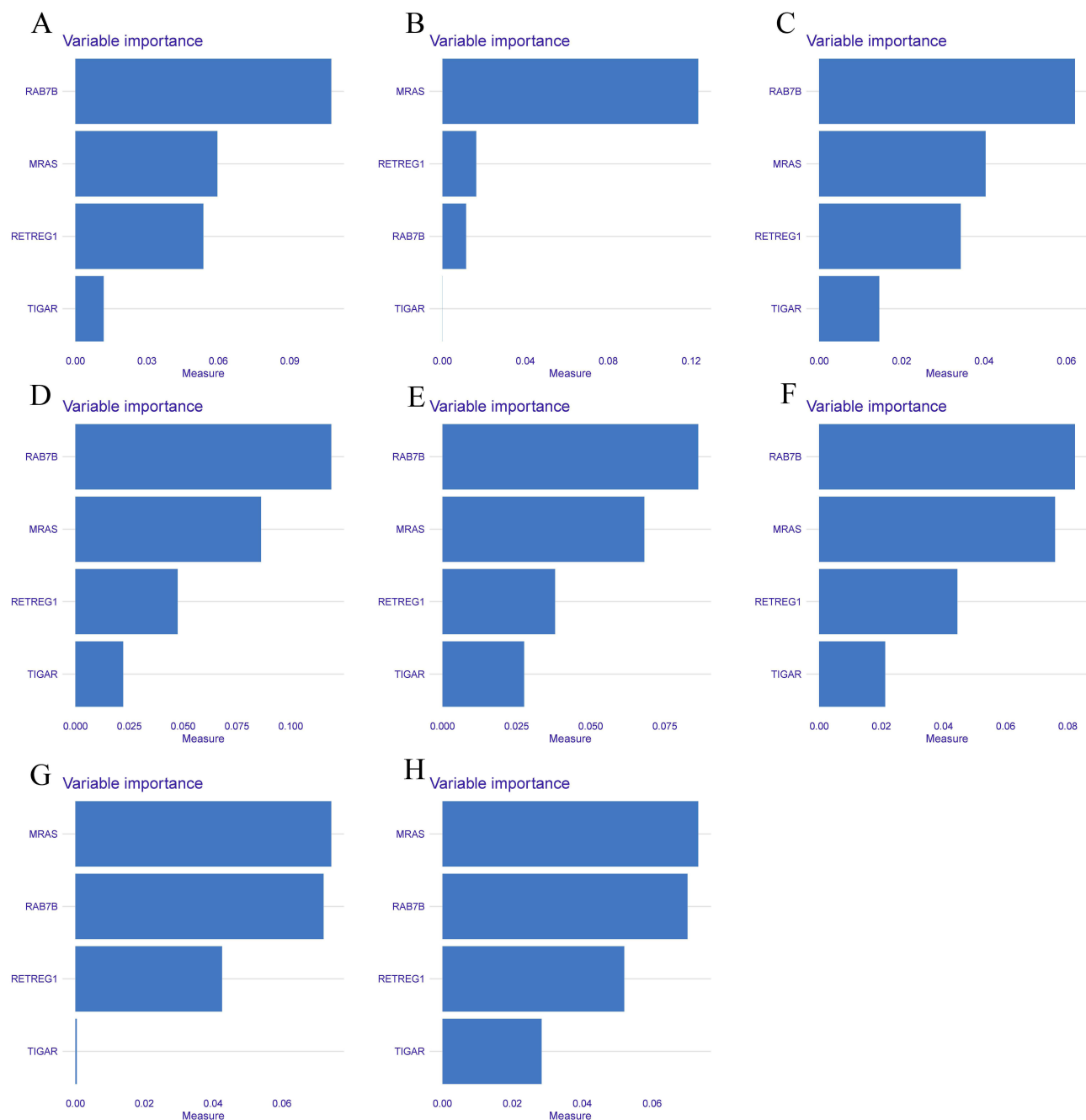
Figure 5 Variable importance plot of the machine learning models. (**A**) AdaBoost; (**B**) Decision Tree; (**C**) KNN; (**D**) LightGBM; (**E**) Naïve Bayes; (**F**) RF; (**G**) SVM; and (**H**) XGBoost.
Abbreviations: KNN, K-nearest neighbors; LightGBM, Light Gradient Boosting; RF, Random Forest; SVM, support vector machines; and XGBoost, eXtreme Gradient Boosting.

These results suggest that MRAS and RAB7B were significantly up-regulated in the early and moderate stages of MASH (Figure 10F) and that their expression increased with disease progression. However, RETREG1 expression was down-regulated in MASH patients compared to the healthy controls, and its expression declined with disease progression.

Effects of the Key Mito-DEGs in MASH

To further validate the expression levels of three potential diagnostic biomarkers in MASH, we conducted qRT-PCR using the L-O2 cells treated with free FAs (1.2 mM) for 0 h and 24 h. The results revealed that the mRNA levels of
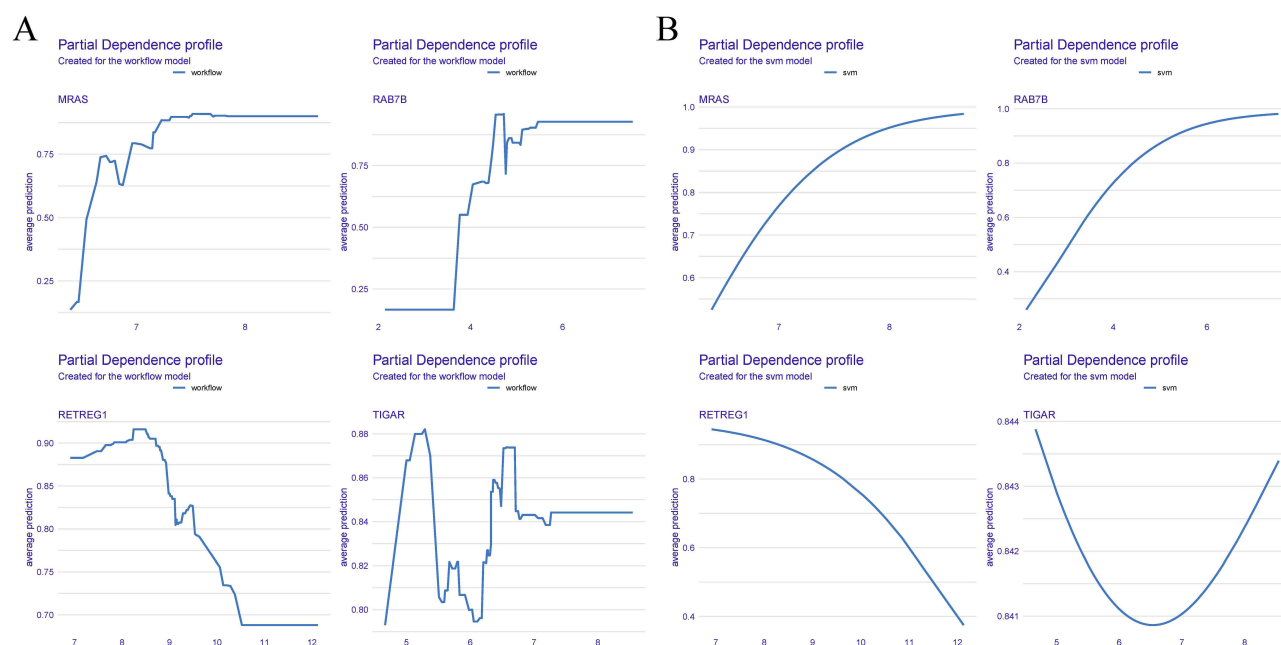
Figure 6 Partial dependence plots of the LightGBM (**A**) and SVM (**B**) models.

MRAS and RAB7B were up-regulated, while that of RETREG1 was down-regulated in the MASH cells, consistent with the bioinformatics predictions (Figure 11). However, the difference in RAB7B expression was insignificant. These results indicate that the three key mito-DEGs MRAS, RAB7B, and RETREG1 can serve as potential diagnostic targets for MASH.

Identifying the Biological Function Features and Immunological Microenvironment in Various Mitophagy-Associated Subtypes

We conducted an unsupervised cluster analysis of 155 MASH samples from the GSE135251 dataset based on the expression of core regulators, and the results revealed two subtypes (Figures 12A–E). Thereafter, we examined variations in the expression of fundamental regulators to connect these subtypes to molecular traits. The results revealed that MRAS and RAB7B were highly up-regulated and RETREG1 was significantly down-regulated in subtype 2 (Figure 12F). Furthermore, the degree of immune cell infiltration varied in subtypes 1 and 2 (Figure 12G). For instance, subtype 1 had more γ/δ T cells and type 2 T helper cells, while subtype 2 had more activated dendritic cells.

Further analysis revealed that the gene expression profiles of the two subtypes differed significantly (Figure 13A and B). The GSVA analysis revealed the up-regulation of several biological processes and pathways in subtype 1, including activin receptor complex; endocardium development; activin-activated receptor activity; endoplasmic reticulum (ER) membrane organization; peroxisome autophagy; negative regulation of ribosome biogenesis (Figure 13C). Additionally, the metabolism of nitrogen, nicotinate and nicotinamide, cysteine and methionine, drug metabolism, cytochrome p450, starch and sucrose metabolism, fatty acid metabolism, and TGF-β signaling pathway were also elevated in this subtype (Figure 13D).

Prediction of Potential Therapeutic Drugs for MASH

CMap analysis was used to identify small-molecule drugs to investigate possible therapeutic targets against MASH subtypes. By applying this method, we were able to identify possible small-molecule drugs from the subtype 1-specific small-molecule drugs. Vinorelbine and ABT-751 shared tubulin inhibitory effects, while LM-1685, isoxicam, and dexketoprofen shared cyclooxygenase inhibitory effects (Figure 14A). The cholinergic receptor antagonist properties of methantheline, procyclidine, and propantheline, as well as the inhibitory effects of dicloxacillin and ethambutol on bacterial cell wall formation, were shared by these subtype 2-specific small-molecule drugs (Figure 14B).
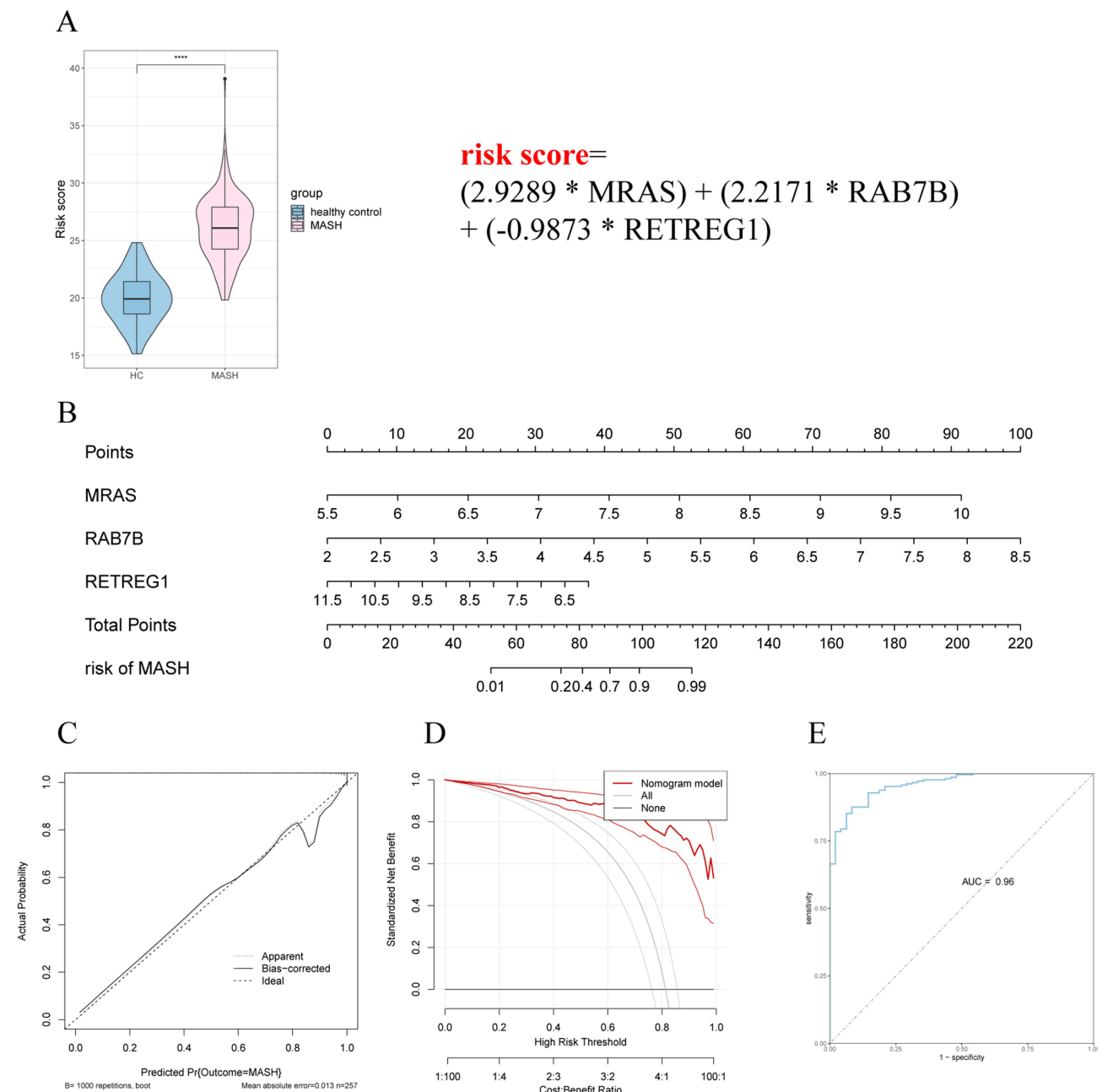
Figure 7 The nomogram assesses the risk of MASH. (**A**) Risk distribution in the training set between MASH patients and healthy controls. (**B**) Nomogram predicting the risk of MASH in the training set based on the feature genes. (**C**) Calibration curve evaluating the prediction efficacy of the nomogram for the training set. (**D**) DCA evaluates the clinical benefits of the nomogram for the training set. (**E**) ROC curves evaluate the diagnostic performance of the feature genes for the training set. ****$p < 0.0001$.
Abbreviations: DCA, Risk decision curve; HC, healthy controls; MASH, metabolic dysfunction-associated steatohepatitis; ROC, the receiver operating characteristic; AUC, area under the receiver operating characteristic curve.

Discussion

The prevalence of MASH poses a significant economic burden and increases the demand for liver transplants in patients suffering from cirrhosis or end-stage liver diseases.[22] Mitophagy eliminates damaged mitochondria and reduces oxidative stress. However, nutritional stress, such as Western diet feeding, can lead to reduced liver mitophagy.[23] The etiology of MAFLD is associated with premature impairment of mitophagy, with the absence of PARKIN further aggravating the MAFLD-associated pathological manifestations.[23] These studies indicate that compromised mitophagy induced by nutritional stress may lead to MAFLD pathogenesis.[23] Therefore, investigating the association between mitophagy and
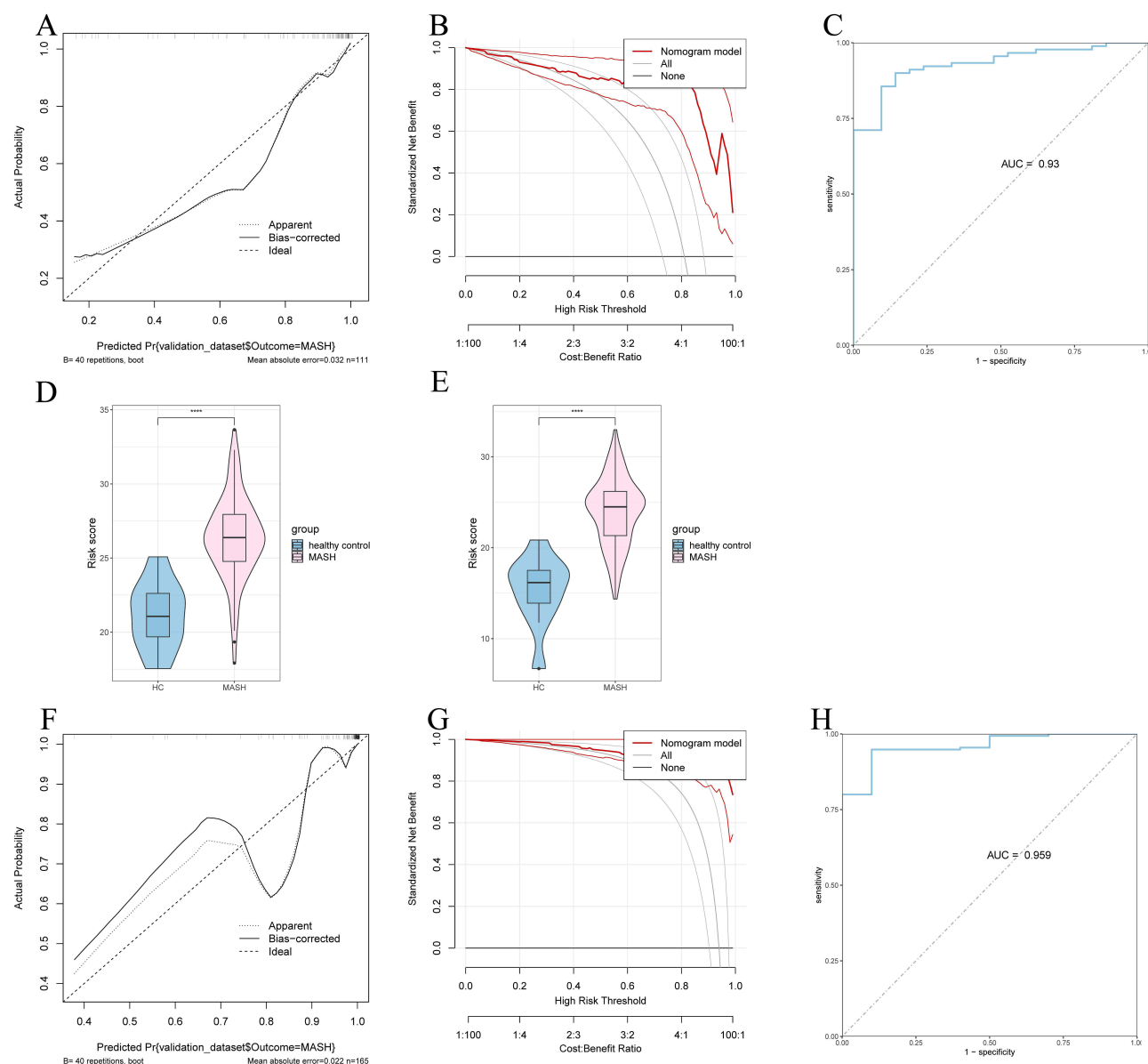
Figure 8 Verification of diagnostic efficacy of model in the test set. (**A**) Use calibration curves to evaluate the predictive ability of the model in the test set. (**B**) DCA evaluates the clinical benefits of the model for the test set. (**C**) ROC curve evaluates the diagnostic efficacy of the model in the test set. (**D**) Risk distribution of MASH and healthy control group in the test set. (**E**) Risk distribution of MASH and healthy control group in the validation set. (**F**) Calibration curves evaluates the the predictive ability of the model in the validation set. (**G**) DCA evaluates the clinical benefits of the model for the validation set. (**H**) ROC curve evaluates the diagnostic efficacy of the model in the validation set. ****p < 0.0001.

Abbreviations: DCA, Risk decision curve; HC, healthy controls; MASH, metabolic dysfunction-associated steatohepatitis; ROC, the receiver operating characteristic; AUC, area under the receiver operating characteristic curve.

MASH may reveal novel prognostic markers, diagnostic indicators, and therapeutic targets and elucidate the underlying mechanisms of the disease.

In this study, we identified four mito-DEGs by overlapping the DEGs from GSE213621 with mitophagy-related genes. Thereafter, we assessed the performance of eight ML models, among which the SVM model had the highest AUC score, while the LightGBM model had the highest F1-score, kappa, and accuracy values. Among the four mito-DEGs, we identified three key mito-DEGs, namely MRAS, RAB7B, and RETREG1, through global and local interpretability analyses, and these DEGs showed high diagnostic prediction ability for MASH. The expression levels of these key mito-DEGs were validated in the MASH cell model and independent dataset, and the experimental results were consistent with the bioinformatics findings.
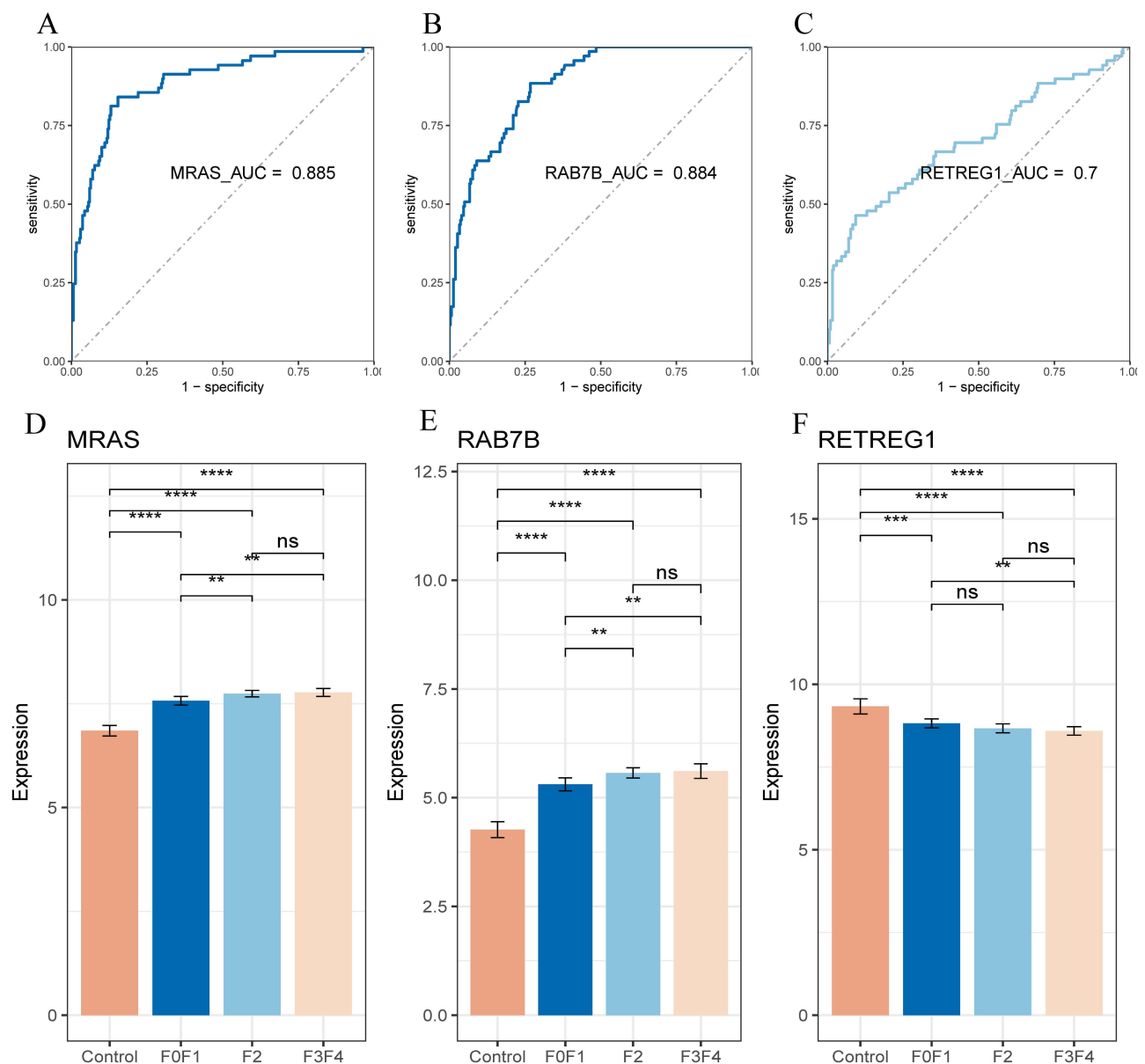
Figure 9 The diagnostic efficacy of the key-mitoDEGs in the GSE213621 dataset. ROC curves of the MRAS (**A**), RAB7B (**B**), and RETREG1 (**C**) for MASH diagnosis in the GSE213621 dataset. Correlation between the expression levels of MRAS (**D**), RAB7B (**E**), and RETREG1 (**F**) and the MASH fibrosis stage. **p < 0.01; ***p < 0.001; ****p < 0.0001. **Abbreviations**: ns, not statistically significant; AUC, area under the receiver operating characteristic curve.

ML has garnered considerable attention in recent years owing to its outstanding performance in clinical diagnosis. In this study, we employed ML models to identify potential mito-DEGs that can predict MASH progression. The findings showed that the SVM had the highest AUC score (0.935), while the LightGBM had the highest accuracy (0.9189), kappa (0.7204), and F1-score (0.9508) values.

Fibrosis in MASH occurs in five stages, ranging from no fibrosis to cirrhosis. In this study, we investigated the association between hub genes and MASH progression, and the results demonstrated a relationship between MRAS, RAB7B, and RETREG1 and MASH progression. Therefore, we developed a diagnostic model based on MRAS, RAB7B, and RETREG1, which exhibited excellent discriminatory ability in MASH. This is the first study to use mitophagy to create a diagnostic model for MASH. In recent years, there has been a gradual increase in research interest in the role of mitophagy in MASH.[24,25] A recent cross-sectional study found that hepatic steatosis patients are characterized by impaired mitochondrial turnover and ER stress. Additionally, the researchers found that mitochondrial quality control
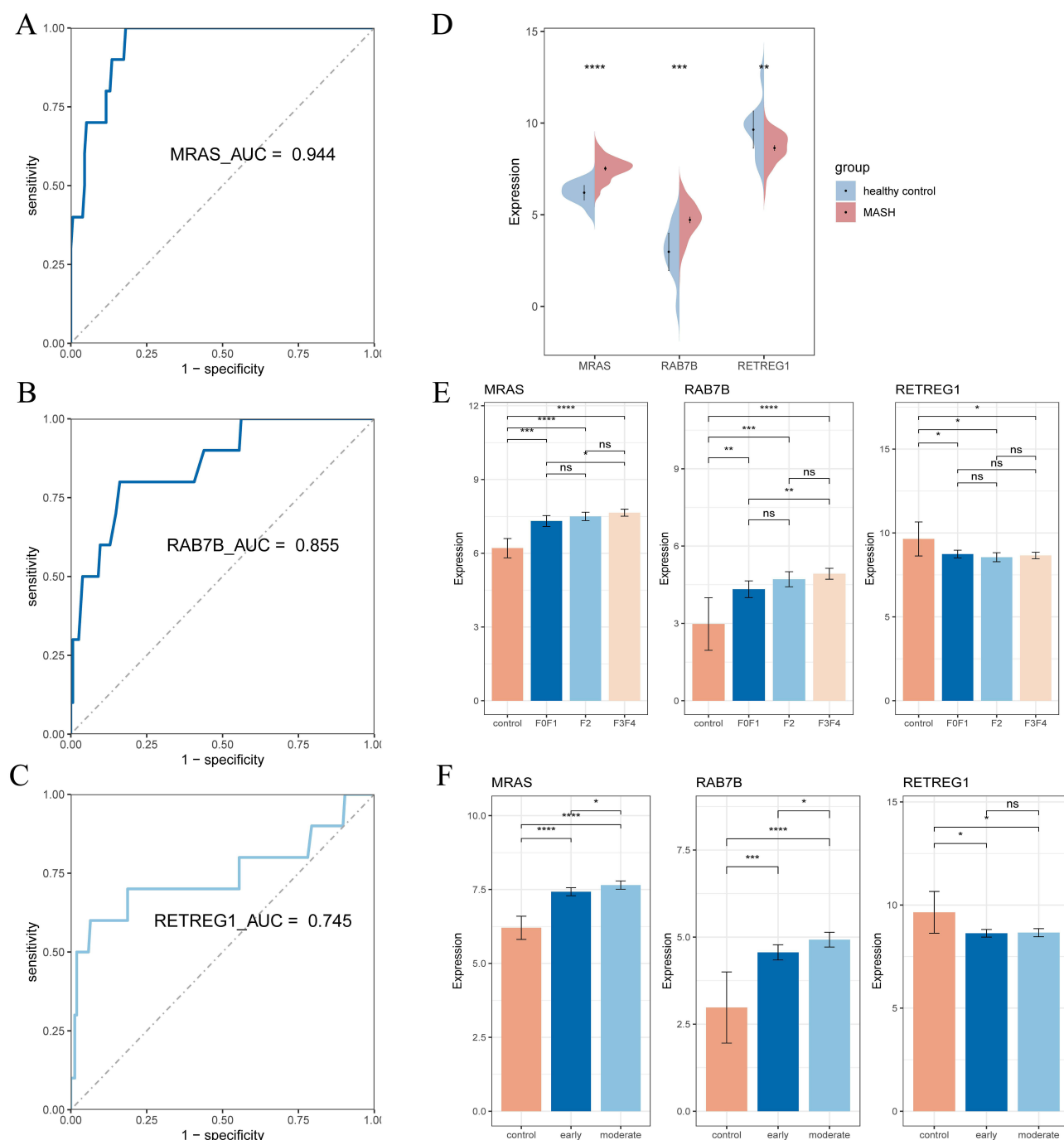
Figure 10 The diagnosis efficacy of the key-mitoDEGs in the GSE135251 dataset. ROC curves of the MRAS (**A**), RAB7B (**B**), and RETREG1 (**C**) for MASH diagnosis in the GSE135251 dataset. (**D**) The split violin plot showed the expression levels of the three mito-DEGs in the MASH patients and healthy controls of the GSE135251 dataset. (**E**) Correlation between the expression levels of the three mito-DEGs and the MASH fibrosis stage. (**F**) The correlation between the stage and the expression level of three mito-DEGs. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

Abbreviations: ns, not statistically significant; MASH, metabolic dysfunction-associated steatohepatitis; AUC, area under the receiver operating characteristic curve.

is critical for energy metabolism and that mitochondrial autophagy-related genes may be potential biological markers for MASH,[26] consistent with the findings of our study. Another study found that epigenetic imbalances in mitochondria-related genes contribute to a metabolically challenging environment that induces mitochondrial dysfunction, leading to MASH progression.[27] These findings may contribute to the auxiliary diagnosis of MASH and provide valuable insights for clinical applications.
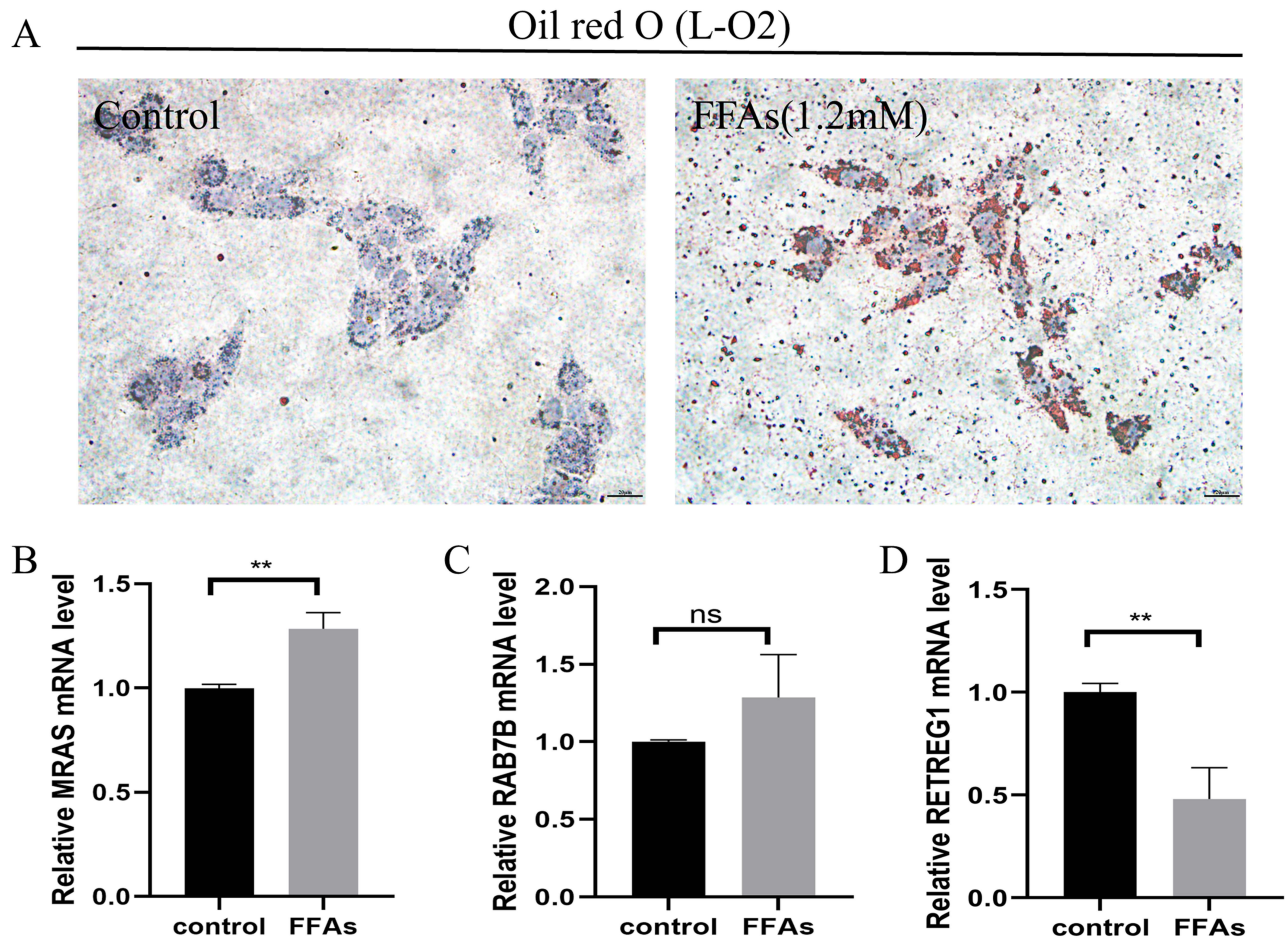
Figure 11 (**A**) Oil red O staining of the L-O2 cells after treatment with 1.2 mM free fatty acids; scale bar = 20 μm; and magnification = 200×. The mRNA levels of MRAS (**B**), RAB7B (**C**), and RETREG1 (**D**) in L-O2 cells treated with FFAs for 0 h and 24 h. **$p < 0.01$.
Abbreviations: ns, not statistically significant; FFAs, free fatty acids.

MRAS, a member of the RRAS subfamily, shares regulatory and effector interactions with classical oncogenic RAS proteins. It functions as a regulatory subunit for phosphatases in complexes with SHOC2 and PP1, contributing to the activation of RAF kinases and providing an essential input for RAS-mediated ERK pathway activation and cellular transformation.[28] Recent studies have reported that MRAS can directly activate RAF on the plasma membrane, potentially substituting for classical RAS.[29] Additionally, a study found that RAS and LZTR1 interact with SQSTM1/ p62 and LC3B, thereby contributing to autophagy, possibly via the Ub-K33-linked chain.[30]

MRAS regulates cell differentiation, proliferation, migration, and polarity.[28] Upon activation, MRAS attracts SHOC2 to cell surface junctions, thereby modifying the connection between E-cadherin and p120-catenin, which is essential for cell migration.[31] Atherosclerotic plaques in high-fat diet-fed mice with dual knockout of Apoe and MRAS exhibited a significant reduction in infiltrating macrophages, demonstrating that the absence of MRAS reduces the migration, adhesion, invasion, and total cell count of macrophages.[32] Further research has shown that MRAS accumulates in the membrane of early-forming phagosomes. Additionally, a study found that macrophages expressing MRAS mutants, whether active or inactive, either show increased or decreased phagosome formation and absorption.[33] Furthermore, studies have reported a significant up-regulation in the MRAS mRNA levels in hepatocellular carcinoma cells compared to normal liver cells.[34]

Rab7b is a small GTPase that regulates transport from the late endosomes/lysosomes to the trans-Golgi network by directly interacting with the actin motor protein myosin II. Rab7b was first identified in DCs and is expressed during monocytic and megakaryocytic differentiation.[35] The polarization of adult DCs is significantly influenced by Rab7b since
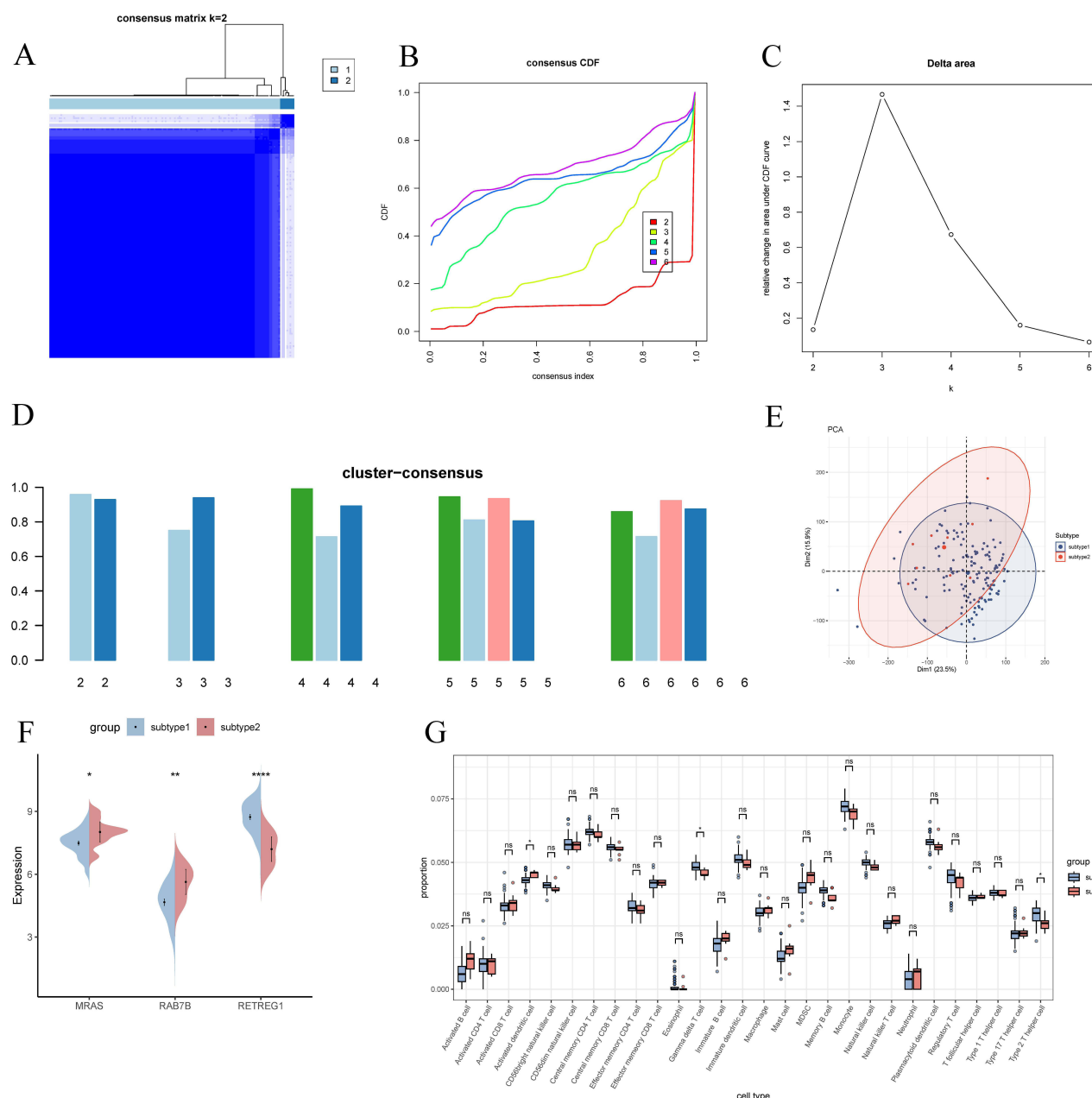
Figure 12 Identification and enrichment analysis of mitophagy-related subtypes of MASH. (**A**) Consensus clustering matrix at k = 2. (**B**) Consensus CDF curves at k = 2–6. (**C**) Relative alterations in the CDF delta area curves. (**D**) Consensus score of each subtype at k = 2–6. (**E**) PCA diagram of subtype 1 (blue) and subtype 2 (red) samples. (**F**) Split violin plot showing the expression of MRAS, RAB7B, and RETREG1 in subtypes 1 and 2. (**G**) Boxplot showing the variations in the infiltrating-immune cells in subtypes 1 and 2. *$p < 0.05$; **$p < 0.01$; ****$p < 0.0001$.
Abbreviations: ns, not statistically significant; CDF, cumulative distribution function; PCA, principal component analysis.

its absence causes improper reorientation of actin and myosin.[36] Additionally, Rab7b interacts with Atg4B, a crucial autophagic enzyme that cleaves LC3 and functions as an autophagy negative regulator.[37] Rab7b depletion increases autophagic flow, which increases autophagic structure size and increases the sequestration and destruction of macroautophages.[37]

The FAM134B family, also known as JK1 and RETREG1, encodes a protein that is anchored to the ER and is found in the Golgi apparatus. RETREG1 mediates the regulation of mitophagy and preadipocyte differentiation.[38] Overexpression of RETREG1 significantly increases mitophagy levels, leading to a decrease in the number of
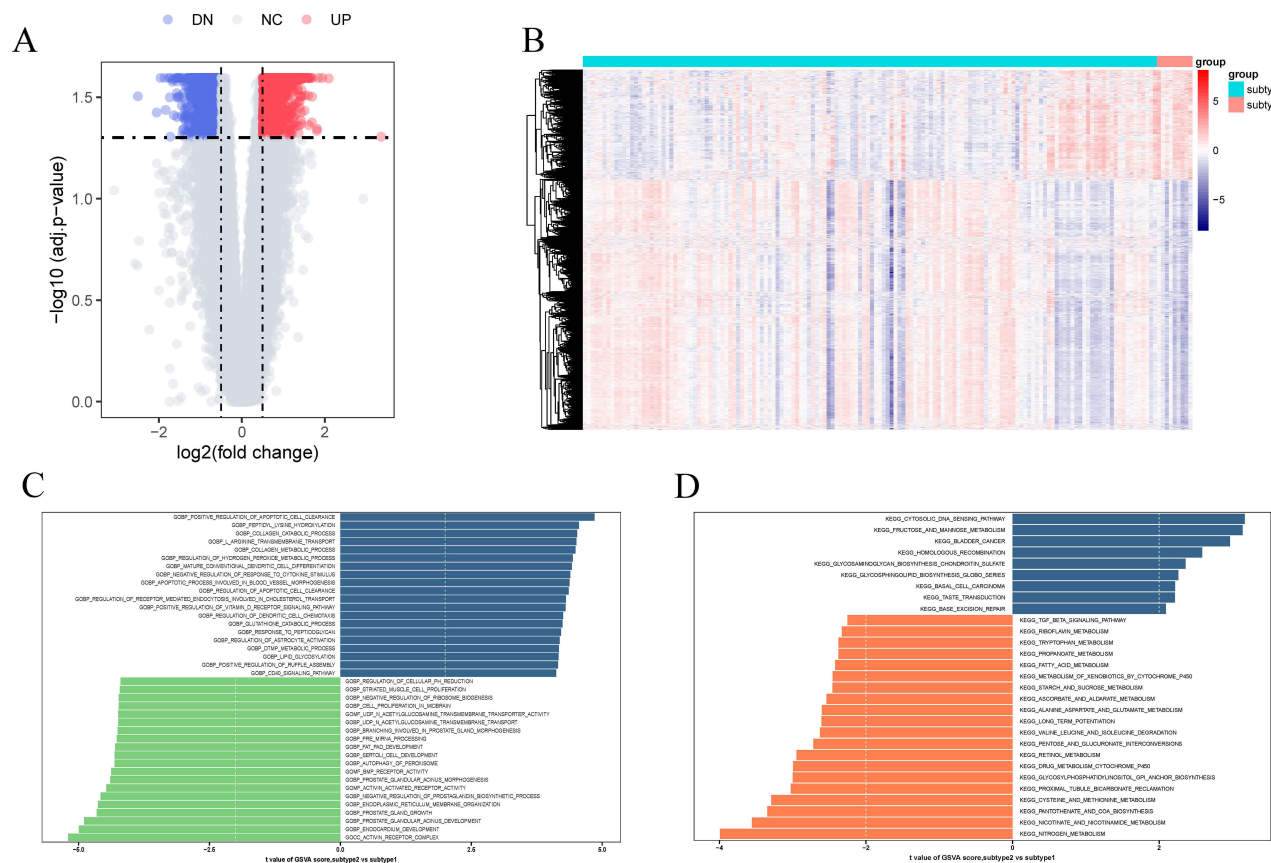
Figure 13 Gene expression profile and functional pathways of the mitophagy-related subtypes. (**A**) Volcano map show the gene expression patterns of mitophagy-related subtypes. (**B**) heatmap show the gene expression patterns of mitophagy-related subtypes. Variations in the biological processes (**C**) and signature pathways (**D**) among the mitophagy-related subtypes, ordered by the t-values of GSVA scores. GSVA, gene set variation analysis. The blue bars indicate upregulated, and the green and Orange bars indicate downregulated biological processes and signature pathways in subtype 2.
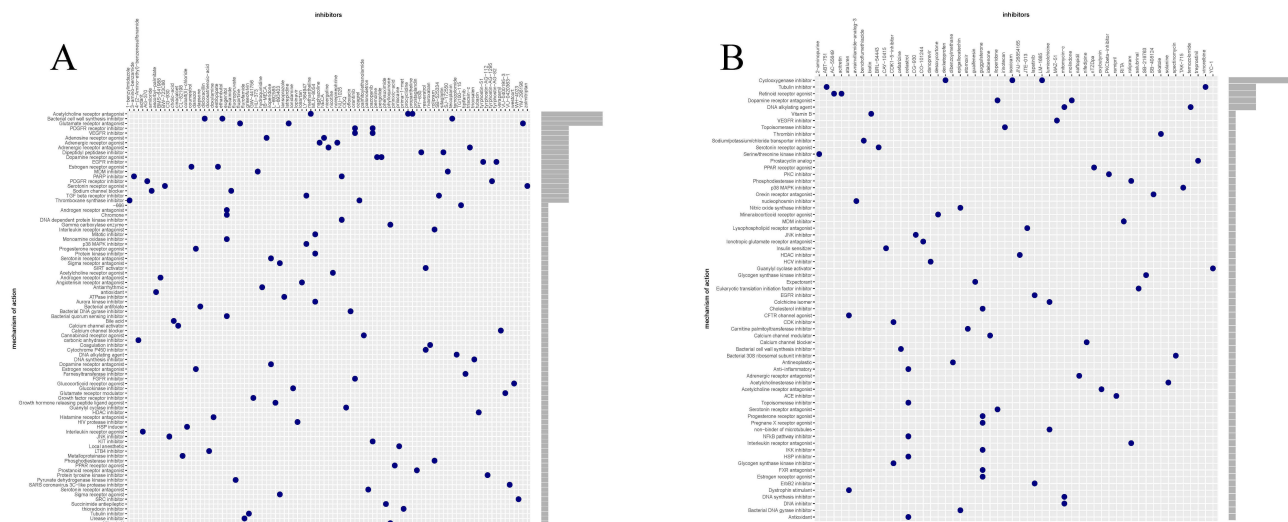

Figure 14 CMap analysis shows the MoA based on the small-molecule compounds against mitophagy-related subtype1 (**A**) and subtype 2 (**B**). CMap, The Connectivity Map; MoA, mechanism of action; The dots indicate the MoA of the small-molecule compounds, and the lengths of the gray bars indicate the number of the small-molecule compounds which share same MoA.

mitochondria.[38] Further research has been conducted to elucidate the mechanism by which RETREG1 mediates selective autophagy of the impaired mitochondria.[39]

Upon mitochondrial damage and mitoplast exposure, OPA1, an inner mitochondrial membrane protein, interacts with RETREG1, directing mitochondria toward the ER. The LC3-interacting region (LIR) motifs within the ER are used to initiate the formation of phagophores around the mitochondria, ultimately resulting in mitoplast degradation.[39] The interaction between RETREG1 and mitochondrial proteins governs the RETREG1-mediated mitophagy. Interaction with the LIR motif enables the recruitment of LC3 by RETREG1 to conduct mitophagy. These findings suggest that the autophagy pathway is the primary mechanism through which the RETREG1-interacting proteins execute their biological functions. Therefore, there is a need for a more comprehensive analysis of the mechanisms underlying protein interactions with the three feature biomarkers to determine their roles in MASH and cellular homeostasis.

In this study, we identified two unique molecular subtypes of MASH by using unsupervised clustering analysis based on the expression profiles of MRAS, RAB7B, and RETREG1 in MASH tissues. Immune infiltration analysis revealed that subtype 2 showed more activated dendritic cells, whereas subtype 1 showed larger amounts of type 2 T helper cells and γ/δ T cells. Functional enrichment analysis showed that subtype 1 was associated with activin receptor complex, endocardium development, activin-activated receptor activity, ER membrane organization, peroxisome autophagy, and negative regulation of ribosome biogenesis.

There are a few limitations to our study. Further research is necessary to confirm the expression levels of MRAS, RAB7B, and RETREG1 in MASH and evaluate the therapeutic efficacy of small-molecule drugs against MASH subtypes. Moreover, the inclusion of additional external validation cohorts is essential to ensure the robustness and reliability of our diagnostic approach. Lastly, our model needs to be regularly updated based on the latest reports on mitophagy-associated genes.

Conclusion

In this study, we employed a sequential biogenic method and ML approach to identify three key mitophagy-related genes in MASH. Subsequently, we developed a LR diagnostic model for predicting MASH based on the expression of these genes. The performance of this model was then evaluated and validated for predicting MASH diagnosis. Furthermore, our findings revealed significant heterogeneity in the gene expression profiles of MASH patients, characterized by distinct mitophagy-related subtypes. Therefore, our results provide novel insights into the management and prevention of MASH and provide the scientific basis for further research in this area.

Data Sharing Statement

These data were derived from the following resources available in the public domain: Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo). The corresponding author can be directed for further requirements.

Ethics and Dissemination

Gene Expression Omnibus (GEO) is a public database supported by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) of United States. In accordance with the Department of Health and Human Services of United States regulations outlined in 45 CFR 46.101, research using certain publicly available data sets does not involve "human subjects". The data contained within these specific data sets are neither identifiable nor private and thus do not meet the federal definition of "human subject" as defined in 45 CFR 46.102. This research adheres to these criteria, as the data from GEO are publicly accessible and de-identified. Therefore, this research project does not need to be reviewed and approved by the Institutional Review Board, (IRB).

Acknowledgment

Funding

Disclosure

The authors have no conflicts of interest to declare in this work.

References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016;64(1):73–84. doi:10.1002/hep.28431
2. Riazi K, Azhari H, Charette JH, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. 2022;7(9):851–861. doi:10.1016/S2468-1253(22)00165-0
3. Powell EE, Wong VWS, Rinella M. Non-alcoholic fatty liver disease. *Lancet*. 2021;397(10290):2212–2224. doi:10.1016/S0140-6736(20)32511-3
4. Ye J, Zhuang X, Li X, et al. Novel metabolic classification for extrahepatic complication of metabolic associated fatty liver disease: a data-driven cluster analysis with international validation. *Metabolism*. 2022;136:155294. doi:10.1016/j.metabol.2022.155294
5. Eslam M, Newsome PN, Sarin SK, et al. A new definition for metabolic dysfunction-associated fatty liver disease: an international expert consensus statement. *J Hepatol*. 2020;73(1):202–209. doi:10.1016/j.jhep.2020.03.039
6. Geier A, Tiniakos D, Denk H, Trauner M. From the origin of NASH to the future of metabolic fatty liver disease. *Gut*. 2021;70(8):1570–1579. doi:10.1136/gutjnl-2020-323202
7. Glancy B. Visualizing mitochondrial form and function within the cell. *Trends Mol Med*. 2020;26(1):58–70. doi:10.1016/j.molmed.2019.09.009
8. Lu Y, Li Z, Zhang S, Zhang T, Liu Y, Zhang L. Cellular mitophagy: mechanism, roles in diseases and small molecule pharmacological regulation. *Theranostics*. 2023;13(2):736–766. doi:10.7150/thno.79876
9. Onishi M, Yamano K, Sato M, Matsuda N, Okamoto K. Molecular mechanisms and physiological functions of mitophagy. *EMBO J*. 2021;40(3): e104705. doi:10.15252/embj.2020104705
10. Pickles S, Vigié P, Youle RJ. Mitophagy and quality control mechanisms in mitochondrial maintenance. *Curr Biol*. 2018;28(4):R170–R185. doi:10.1016/j.cub.2018.01.004
11. Sharma R, Reinstadler B, Engelstad K, et al. Circulating markers of NADH-reductive stress correlate with mitochondrial disease severity. *J Clin Invest*. 2021;131(2):e136055. doi:10.1172/JCI136055
12. Dornas W, Schuppan D. Mitochondrial oxidative injury: a key player in nonalcoholic fatty liver disease. *Am J Physiol Gastrointest Liver Physiol*. 2020;319(3):G400–G411. doi:10.1152/ajpgi.00121.2020
13. Edmunds LR, Xie B, Mills AM, et al. Liver-specific Prkn knockout mice are more susceptible to diet-induced hepatic steatosis and insulin resistance. *Mol Metab*. 2020;41:101051. doi:10.1016/j.molmet.2020.101051
14. Williams JA, Ni HM, Ding Y, Ding WX. Parkin regulates mitophagy and mitochondrial function to protect against alcohol-induced liver injury and steatosis in mice. *Am J Physiol Gastrointest Liver Physiol*. 2015;309(5):G324–340. doi:10.1152/ajpgi.00108.2015
15. Pei Y, Chen S, Zhou F, Xie T, Cao H. Construction and evaluation of Alzheimer's disease diagnostic prediction model based on genes involved in mitophagy. *Front Aging Neurosci*. 2023;15:1146660. doi:10.3389/fnagi.2023.1146660
16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287. doi:10.1089/omi.2011.0118
17. Lan T, Yu Y, Zhang J, et al. Cordycepin ameliorates nonalcoholic steatohepatitis by Activation of the AMP-activated protein kinase signaling pathway. *Hepatology*. 2021;74(2):686–703. doi:10.1002/hep.31749
18. Markovich Gordon M, Moser AM, Rubin E. Unsupervised analysis of classical biomedical markers: robustness and medical relevance of patient clustering using bioinformatics tools. *PLoS One*. 2012;7(3):e29578. doi:10.1371/journal.pone.0029578
19. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf*. 2013;14:7. doi:10.1186/1471-2105-14-7
20. Yang C, Zhang H, Chen M, et al. A survey of optimal strategy for signature-based drug repositioning and an application to liver cancer. *Elife*. 2022;11:e71880. doi:10.7554/eLife.71880
21. Subramanian A, Narayan R, Corsello SM, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 Profiles. *Cell*. 2017;171(6):1437–1452.e17. doi:10.1016/j.cell.2017.10.049
22. Mittal S, El-Serag HB, Sada YH, et al. Hepatocellular carcinoma in the absence of cirrhosis in United States veterans is associated with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol*. 2016;14(1):124–131.e1. doi:10.1016/j.cgh.2015.07.019
23. Undamatla R, Fagunloye OG, Chen J, et al. Reduced mitophagy is an early feature of NAFLD and liver-specific PARKIN knockout hastens the onset of steatosis, inflammation and fibrosis. *Sci Rep*. 2023;13(1):7575. doi:10.1038/s41598-023-34710-x
24. LeFort KR, Rungratanawanich W, Song BJ. Contributing roles of mitochondrial dysfunction and hepatocyte apoptosis in liver diseases through oxidative stress, post-translational modifications, inflammation, and intestinal barrier dysfunction. *Cell Mol Life Sci*. 2024;81(1):34. doi:10.1007/s00018-023-05061-7
25. Ramanathan R, Patwa SA, Ali AH, Ibdah JA. Thyroid hormone and mitochondrial dysfunction: therapeutic implications for Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD). *Cells*. 2023;12(24):2806. doi:10.3390/cells12242806
26. Sarabhai T, Kahl S, Gancheva S, et al. Loss of mitochondrial adaptation associates with deterioration of mitochondrial turnover and structure in metabolic dysfunction-associated steatotic liver disease. *Metabolism*. 2024;151:155762. doi:10.1016/j.metabol.2023.155762
27. Theys C, Ibrahim J, Mateiu L, et al. Mitochondrial GpC and CpG DNA hypermethylation cause metabolic stress-induced mitophagy and cholestophagy. *Int J Mol Sci*. 2023;24(22):16412. doi:10.3390/ijms242216412
28. Young LC, Rodriguez-Viciana P. MRAS: a close but understudied member of the RAS family. *Cold Spring Harb Perspect Med*. 2018;8(12): a033621. doi:10.1101/cshperspect.a033621

29. Lai LP, Fer N, Burgan W, et al. Classical RAS proteins are not essential for paradoxical ERK activation induced by RAF inhibitors. *Proc Natl Acad Sci U S A*. 2022;119(5):e2113491119. doi:10.1073/pnas.2113491119

30. Abe T, Umeki I, Kanno SI, Inoue SI, Niihori T, Aoki Y. LZTR1 facilitates polyubiquitination and degradation of RAS-GTPases. *Cell Death Differ*. 2020;27(3):1023–1035. doi:10.1038/s41418-019-0395-5

31. Kota P, Terrell EM, Ritt DA, Insinna C, Westlake CJ, Morrison DK. M-Ras/Shoc2 signaling modulates E-cadherin turnover and cell-cell adhesion during collective cell migration. *Proc Natl Acad Sci U S A*. 2019;116(9):3536–3545. doi:10.1073/pnas.1805919116

32. Freyer J, Behrensen M, Aherrahrou Z, Berbee JFP, Schunkert H, Erdmann J. MRas-knock-out mouse: b-cell phenotype and reduced macrophage infiltration in atherosclerotic plaques at the aortic root. *Eur Heart J*. 2013;34:442.

33. Egami Y, Araki N. Transient recruitment of M-Ras GTPase to phagocytic cups in RAW264 macrophages during FcγR-mediated phagocytosis. *Microscopy*. 2018;67(2):68–74. doi:10.1093/jmicro/dfx131

34. Zoheir KMA, Abd-Rabou AA, Harisa GI, et al. Gene expression of IQGAPs and Ras families in an experimental mouse model for hepatocellular carcinoma: a mechanistic study of cancer progression. *Int J Clin Exp Pathol*. 2015;8(8):8821–8831.

35. Borg Distefano M, Hofstad Haugen L, Wang Y, et al. TBC1D5 controls the GTPase cycle of Rab7b. *J Cell Sci*. 2018;131(17):jcs216630. doi:10.1242/jcs.216630

36. Vestre K, Persiconi I, Borg Distefano M, et al. Rab7b regulates dendritic cell migration by linking lysosomes to the actomyosin cytoskeleton. *J Cell Sci*. 2021;134(18):jcs259221. doi:10.1242/jcs.259221

37. Kjos I, Borg Distefano M, Sætre F, et al. Rab7b modulates autophagic flux by interacting with Atg4B. *EMBO Rep*. 2017;18(10):1727–1739. doi:10.15252/embr.201744069

38. Cai M, Zhao J, Liu Q, Wang X, Wang Y. FAM134B improves preadipocytes differentiation by enhancing mitophagy. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2019;1864(12):158508. doi:10.1016/j.bbalip.2019.08.004

39. Mookherjee D, Das S, Mukherjee R, et al. RETREG1/FAM134B mediated autophagosomal degradation of AMFR/GP78 and OPA1 -a dual organellar turnover mechanism. *Autophagy*. 2021;17(7):1729–1752. doi:10.1080/15548627.2020.1783118