

Article

Not peer-reviewed version

Exploring Prognostic Gene Expression Signature for Thyroid Eye Disease

[Bi Lv](#) , [Xianghui Wei](#) , Wenlian Zheng , Diyao Wei , [Minghui Hou](#) , Haiyan Dai , Quan Xiao , Hong-Gang Liang , Hansheng Huang , Lulu Qin , Jingrong Li , [Huanyan Wang](#) *

Posted Date: 10 January 2025

doi: 10.20944/preprints202501.0754.v1

Keywords: thyroid eye disease; mitochondrial energy metabolism-related gene; immune; gene expression



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Exploring Prognostic Gene Expression Signature for Thyroid Eye Disease

Bi Lv ¹, Xianghui Wei ¹, Wenlian Zheng ², Diyao Wei ¹, Minghui Hou ³, Haiyan Dai ¹,
Quan Xiao ¹, Hong-Gang Liang ¹, Hansheng Huang ¹, Lulu Qin ¹, Jingrong Li ¹
and Huanyan Wang ^{1,*}

¹ Liuzhou Red Cross Hospital/Eye Hospital of Liuzhou City, China; lvbi681234@163.com; lzs778888@163.com; weidiyao1996@163.com; 18707727896@163.com; lzhhmzk@163.com; lianghg6563@163.com; 18877282159@163.com; 19101256751@163.com; lijngrng@163.com

² Shaoguan Hospital Southern Medical University, China; zwlxwk623@163.com

³ Third Affiliated Hospital of Sun Yat-sen University, China; houmh3@mail2.sysu.edu.cn

* Correspondence: 13397725199@163.com

Abstract: Background: Thyroid-related eye disease (TED) is the most prevalent orbital disorder of autoimmune origin, yet research on its mitochondrial energy metabolism-related genes remains limited with no definitive biomarkers or therapeutic targets identified. **Methods:** We downloaded datasets from the GEO database, forming TED and Control groups. Mitochondrial energy metabolism-related genes (MEMRGs) were sourced from GeneCards. We conducted differential gene expression analysis, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment, gene set enrichment analysis (GSEA), and identified Key Genes. A diagnostic logistic regression model was then built, expression differences validated, immune infiltration analyzed via ssGSEA, and network interaction assessed. **Results:** After correcting batch effects in datasets GSE58331 and GSE105149, we identified 26 mitochondrial energy metabolism-related differentially expressed genes (MEMRDEGs). Through one-way logistic analysis, Random Forest, and a LASSO risk model, we pinpointed 10 Key Genes crucial for diagnosis. The diagnostic model demonstrated superior performance and accuracy, with Key Genes showing significantly higher expression in the TED group. Additionally, activated CD8 T cells, immature B cells, and T follicular helper cells exhibited notable differences. **Conclusions:** The expression of mitochondrial energy metabolism-related genes in TED, stratified by risk score levels, exhibits a positive correlation with immune activity, unveiling a promising new pathway for identifying therapeutic targets. This finding calls for deeper biological exploration.

Keywords: thyroid eye disease; mitochondrial energy metabolism-related gene; immune; gene expression

Background

Thyroid eye disease (TED) occurs in about 25% of individuals with Graves' disease, making it a prevalent orbital condition. It is an autoimmune disorder that mainly impacts orbital tissues, particularly the extraocular muscles and connective tissue[1]. TED is commonly linked to hyperthyroidism, particularly Graves' disease, but can also affect individuals with normal or reduced thyroid function. The development of TED involves immune system dysregulation, where the immune system targets not only the thyroid gland but also the surrounding eye tissues, leading to inflammation and fibrosis. Clinical symptoms include eye bulging, eyelid swelling, restricted eye movement, double vision, and in severe cases, vision impairment. The severity of TED can vary from mild discomfort to significant visual issues, sometimes requiring surgery. Treatment typically includes managing thyroid function, using immunosuppressants to reduce inflammation, and surgical intervention for severe cases. However, both medical and surgical treatments may not

always yield satisfactory results. Numerous studies are currently investigating the pathogenesis of TED[2-4], focusing on potential therapeutic targets, including receptor proteins, immune cells, adipose inhibition, anti-fibrosis, transcription factors and metabolism-regulating enzymes[1]. However, the pathogenesis has not been fully elucidated.

Past TED treatments relied on immunotherapy and immunosuppression. Research on immune responses in TED has primarily centered on inflammation, oxidative stress, adipose tissue changes, and fibrosis. Connective tissue emerged as a significant autoimmune target during active TED phases, as shown by the dense presence of T lymphocytes, mast cells, and sporadic B cells among extraocular muscle fibers and orbital fat[2-4]. Fibroblasts in the orbital tissues of patients with TED overexpressed the insulin-like growth factor receptor (IGF-1R)[5], which stimulated autoimmune responses[6]. Under the combined action of IL-1 β , CD154, and interleukins, excessive production of prostaglandin E2 (PGE2) may lead to inflammation, fat swelling and deposition of excessive extracellular matrix components (e.g., hyaluronic acid) in TED[7]. Mycophenolate mofetil (MMF) selectively inhibits the differentiation proliferation and activation of T and B lymphocytes and exerts immunosuppressive effects[8]. Additionally, it reduced mRNA and protein levels of tumor necrosis factor- α , interleukin β , and vascular endothelial growth factor, lowered inflammatory exudate levels, and demonstrated anti-inflammatory properties[9]. Teprotumumab (TAB), a human monoclonal antibody that inhibited IGF-1R and well controlled the inflammatory response of TED in active phase[10], has become an emerging targeted therapeutic option for TED. However, the specific pathogenesis of TED remain incompletely understood.

Recent research indicates that mitochondrial energy metabolism significantly influences the development and advancement of thyroid eye disease (TED). Orbital fibroblasts in TED are crucial pathological cells that can transform into adipocytes or myofibroblasts, causing orbital tissue expansion and fibrosis. Mitochondrial dysfunction in these cells leads to overproduction of reactive oxygen species (ROS), activating pro-inflammatory and pro-fibrotic signaling pathways like NF- κ B and TGF- β . These pathways worsen inflammation and tissue changes, propelling disease progression. Moreover, increased ROS from mitochondria triggers oxidative stress, harming cellular components and activating immune cells to release pro-inflammatory cytokines such as IL-6 and TNF- α , escalating inflammation. Oxidative stress also triggers hypoxia-inducible factor (HIF-1 α), promoting angiogenesis and fibrosis, common in advanced TED stages. Additionally, mitochondrial dysfunction can prevent diseased cells from undergoing apoptosis, sustaining inflammation and tissue alterations. Consequently, mitochondrial function intricately influences TED's pathological processes, suggesting that targeting mitochondrial metabolism could offer new treatment avenues. Current research in various fields explores the mitochondrial energy metabolism pathway, linking abnormalities to oxidative stress caused by excessive free radicals, ROS release, and an imbalance in antioxidant capabilities[11,12]. In cancer research, disruptions in mitochondrial pathways and metabolic imbalances can alter gene expression, contributing to cancer initiation, progression, and immune evasion[13]. In degenerative diseases, inhibiting the mitochondrial pyruvate carrier (MPC) has shown significant neuroprotective effects in Parkinson's disease models. Furthermore, the complement system and mitochondrial energy metabolism are linked to myopia prevention and management, offering potential therapeutic targets[14]. There is currently a lack of literature on the investigation of mitochondrial energy metabolism-related genes (MEMRGs) in TED. Exploring these genes could be crucial for diagnosing and predicting TED outcomes, making it a valuable area for further research.

In this study, we identified MEMRGs linked to TED from databases. As there were no previous reports on their connection to TED immunity, we compared gene enrichment in biological processes between TED and normal individuals. We investigated the relationship between MEMRG expression and TED immunity, sought bioinformatic approaches for potential therapeutic targets in TED, and assessed their therapeutic and prognostic relevance in this condition.

1. Materials and Methods

1.1. Dataset and Data Processing

The TED datasets GSE58331[15-20] (35 TED samples and 29 Control samples), GSE105149[21] (4 TED samples and 7 Control samples) were obtained by downloading them from the GEO database[22] (<https://www.ncbi.nlm.nih.gov/geo/>), separately. The GSE105149 dataset analyzes gene expression in lacrimal gland biospecimens and surgical waste tissues, while GSE58331 focuses on gene expression in orbital biopsies. Both datasets utilized the GPL570 microarray platform, ensuring technical consistency. MEMRGs were sourced from the GeneCards database[23] (<https://www.genecards.org/>). The set of MEMRGs from the published literature[24] was obtained from the PubMed website (<https://pubmed.ncbi.nlm.nih.gov/>), and after merging and de-duplication, a total of 386 MEMRGs. Datasets GSE58331 and GSE105149 were de-batched using R package sva[25], which were normalised by R package limma[26], annotated probes were standardized and normalized. Distribution boxplots indicate successful removal of batch effects, reducing the impact of sample type variations on results. All chosen samples from GSE58331 were part of the test set, while those from GSE105149 were the validation set for further analyses.

1.2. Analysis of Differentially Expressed Genes

Genes in the TED and Control groups were differentially analyzed using the R package limma. A threshold of $|\log FC| > 0.30$ and $p\text{-value} < 0.05$ was applied for identifying Differentially Expressed Genes (DEGs). Among these, genes with $\log FC > 0.30$ and $p\text{-value} < 0.05$ were classified as Up-regulated Genes, while genes with $\log FC < -0.30$ and $p\text{-value} < 0.05$ were categorized as Down-regulated Genes. Mitochondrial Energy Metabolism-Related Differentially Expressed Genes (MEMRDEGs) were identified by intersecting genes from all TED samples in the DEGs dataset with MEMRGs. The cutoff ($|\log_2 FC| > 0.3$) chosen for the DEG analysis was carefully considered. We aimed to capture genes potentially linked to TED's pathological processes, ensuring inclusion of candidates with biological relevance. Additionally, similar or more permissive $\log_2 FC$ thresholds have been used in previous studies (e.g., PMID: 39021997), supporting our decision.

1.3. Functional Enrichment Analysis of Differentially Expressed MMRGs

For Gene Ontology (GO)[27] and KEGG[28] enrichment analysis, the "clusterProfiler"[29] software package was used to explore the enrichment degree of biological process (BP), cellular component (CC), molecular function (MF), and biological pathway (KEGG) of the differentially expressed MMRGs. P value < 0.05 and FDR value (q value) < 0.25 was considered statistically significant. Data visualization was performed.

1.4. GSEA

Gene Set Enrichment Analysis (GSEA)[30] was conducted on all genes in dataset GSE58331 using the "clusterProfiler" package in R, sorted based on $\log FC$ values. TED samples from GSE58331 were then divided into high-risk and low-risk groups using the median RiskScore value. Genes meeting the criteria of $|\log FC| > 0.30$ and $p\text{-value} < 0.05$ in these groups were selected for variance analysis with the limma package. Subsequently, GSEA was repeated on all genes within the different risk groups using the "clusterProfiler" package.

The analysis utilized the seed 2020, 1000 permutations, a minimum of 10 genes per set, and a maximum of 500 genes per set. The gene set c2.cp.all.v2022.1.Hs.symbols.gmt [All Canonical Pathways] (3050 sets) was sourced from the Molecular Signatures Database (MSigDB)[31] for GSEA, and the gene set GSEA screening criteria were $\text{adj. } p < 0.05$ and FDR value (q value) < 0.25 , and the p value correction method was Benjamini-Hochberg (BH).

1.5. Screening of Key Genes

To identify Key Genes and build diagnostic models, we initially conducted one-way logistic regression on MEMRDEGs, considering a significance threshold of $p < 0.05$ for screening. The selected MEMRDEGs were then included in the subsequent random forest analysis. Model construction was performed using the Random Forest package[32] with the expression of the MEMRDEGs obtained from the screening of the one-way logistic regression in the expression matrix of the dataset GSE58331 with the parameters $\text{set.seed}(234)$, $\text{ntree} = 300$. MeanDecreaseGini is the average decrease in the Gini

coefficient. the Gini coefficient represents the impurity of a node, the larger the Gini coefficient the less pure it is and the more impurities it contains. We then trade-off the number of variables by performing five ten-fold cross-validations in conjunction with the cross-validation curves. We used the training set itself for cross-validation, retaining a relatively small number of variables with relatively small errors, and then selected our significant variables for subsequent analyses based on MeanDecreaseGini. Least Absolute Shrinkage and Selection Operator(LASSO)[33] Regression analyses were conducted following random forest screening using the "glmnet" R package, with set.seed(300) to set the seed and a run period of 200 to prevent overfitting. The outcomes of the LASSO regression analyses were visualized using diagnostic model plots and variable trajectory plots. MEMRDEGs incorporated in the ultimate LASSO regression model were identified as the Key Genes for further analyses.

We selected LASSO regression for its robust feature selection, effectively managing high-dimensional data while minimizing overfitting. Its penalty term inherently eliminates less relevant variables, making it ideal for identifying key genes. Compared to elastic net, LASSO offers greater simplicity and interpretability, which is crucial when prioritizing model clarity.

1.6. Key Genes Constructs Diagnostic Logistic Regression Models

We developed a multifactorial logistic model to derive coefficients for each Key Gene. These coefficients were multiplied by their respective expressions to calculate the Risk Score for each sample. Based on the median Risk Score, the dataset's disease groups were categorized into high and low-risk groups (High, Low). The Risk Score was computed using the following formula:

$$\text{RiskScore} = \sum_i \text{Coefficient}(\text{gene}_i) * \text{mRNA Expression}(\text{gene}_i)$$

Nomogram[35] is a graphical representation of the functional relationships among multiple independent variables in a two-dimensional rectangular coordinate system, typically showing disjointed line segments. The results of the multifactorial logistic model were visualized using the "rms" R package. Plotting the columns illustrated the interconnections among the Key Genes featured in the multifactorial logistic model.

Calibration Curve to evaluate the accuracy and discriminatory power of the model based on the results of the multifactorial logistic regression analysis through Calibration analysis. Decision Curve Analysis (DCA) was plotted through Decision Curve Analysis (DCA)[36] using the R package "ggDCA" based on Key Genes from the dataset GSE58331.

Finally, the R package pRO was used to plot the ROC curves of the Logistic Risk Score (Risk Score) in datasets GSE58331 and GSE105149 and to calculate the Area Under the Curve (AUC) to assess the expression of Logistic Risk Score (Risk Score) volume on the diagnostic effect of TED occurrence. The functional correlation of Key Genes was calculated by R package "GOSemSim"[37], and the functional correlation between Key Genes was analysed for Friends. Finally, the chromosomal localisation map was drawn to show the position of Key Genes on chromosomes by R package "RCircos"[38].

1.7. Validation of Differential Expression of Key Genes

To further explore the expression differences of Key Genes in the datasets GSE58331 and GSE105149 in the TED group and on the Control group, group comparison plots were drawn based on the expression of Key Genes. Finally, the R package "pROC" was used to plot the ROC curve of Key Genes and calculate the Area Under the Curve (AUC) to assess the diagnostic effect of Key Genes expression on the occurrence of TED. The AUC of the ROC curve is generally between 0.5 and 1. The closer the AUC is to 1, the better the diagnostic effect. The AUC of 0.5 to 0.7 has low accuracy, the AUC of 0.7 to 0.9 has some accuracy, and the AUC of 0.9 or more has high accuracy.

1.8. Key Genes' ssGSEA Algorithm for Immune Infiltration Analysis

Single-Sample Gene-Set Enrichment Analysis (ssGSEA)[39] quantifies the relative abundance of each immune cell infiltration. First, each infiltrating immune cell type is labeled, such as Activated CD8 T cell, Activated dendritic cell, Gamma delta T cell, Natural killer cell, Regulatory T cell, and

many other human immune cell subtypes. Next, the enrichment scores calculated using ssGSEA analysis were used to represent the relative abundance of each immune cell infiltration in each sample to derive the immune cell infiltration matrix. Subsequently, immune cells with significant differences in the two groups were screened for subsequent analysis, the correlation between immune cells was calculated based on the Spearman algorithm, and the correlation heatmap was drawn using the R package pheatmap to demonstrate the results of the correlation analysis of the immune cells themselves. The correlation between Key Genes and immune cells was calculated based on the Spearman algorithm, and the result of p value < 0.05 was retained, and the correlation bubble map was plotted using the R package “ggplot2” to show the results of the correlation analysis between Key Genes and immune cells.

1.9. mRNA-miRNA, mRNA-TF, mRNA-RBP Interaction Network Analysis of Key Genes

ENCORI Database[40] is the version 3.0 of starBase database. miRNA-ncRNA, miRNA-mRNA, ncRNA-RNA, RNA-RNA, RBP-ncRNA and RBP-mRNA interactions in ENCORI database are based on the data mining of CLIP-seq and degradome sequencing (for plants), and we utilized ENCORI database to predict miRNAs interacting with Key Genes and screen mRNA-miRNA interaction pairs using pancancerNum > 7 as a screening criterion, and visualize the mRNA-miRNA interaction network using Cytoscape software.

CHIPBase database (version 3.0)[41] (<https://rna.sysu.edu.cn/chipbase/>) identifies thousands of binding motif matrices and their binding sites from ChIP-seq data of DNA-binding proteins and predicts transcriptional regulatory relationships between millions of transcription factors (TFs) and genes. We searched for TFs binding to Key Genes through the CHIPBase database (version 3.0) and used the sum of Number of samples found (upstream) and Number of samples found (downstream) greater than 10 as a filtering criterion for mRNA-TF interaction pairs. The mRNA-TF interaction pairs were screened and the mRNA-TF interaction network was visualized using Cytoscape software.

RNA-Binding Protein (RBP)[42] plays a key role in gene regulation, and is critical for the regulation of life activities such as RNA synthesis, selective splicing, modification, translocation and translation. Based on the StarBase v3.0 database (<https://starbase.sysu.edu.cn/>), the target RBPs of Key Genes were predicted and the mRNA-RBP interaction pairs were screened using clusterNum > 12 as the screening criterion. mRNA-RBP interaction networks were visualized by Cytoscape software (mRNA-RBP Interaction Network).

1.10. Statistical Analysis

All data processing and analyses in this study were conducted using R software (Version 4.2.2). For comparing two groups of continuous variables, statistical significance of normally distributed variables was assessed using the independent Student's T-Test, while non-normally distributed variables were analyzed with the Mann-Whitney U-Test (Wilcoxon Rank Sum Test). Comparisons involving three or more groups utilized the Kruskal-Wallis test. Spearman correlation analysis was employed to calculate correlation coefficients between different molecules. Unless specified otherwise, all p values were two-sided, with $p < 0.05$ considered statistically significant.

2. Results

2.1. MEMRGs Landscape in TED

After using R package “sva” to remove the batch effect on TED dataset GSE58331, GSE105149, the difference in the expression value of the dataset before and after the removal of the batch effect was compared by distribution box line plots (Fig.1A-B), (Fig.1C-D), The results showed that the batch effect of the samples in TED dataset is basically eliminated after removal of the batch treatment.

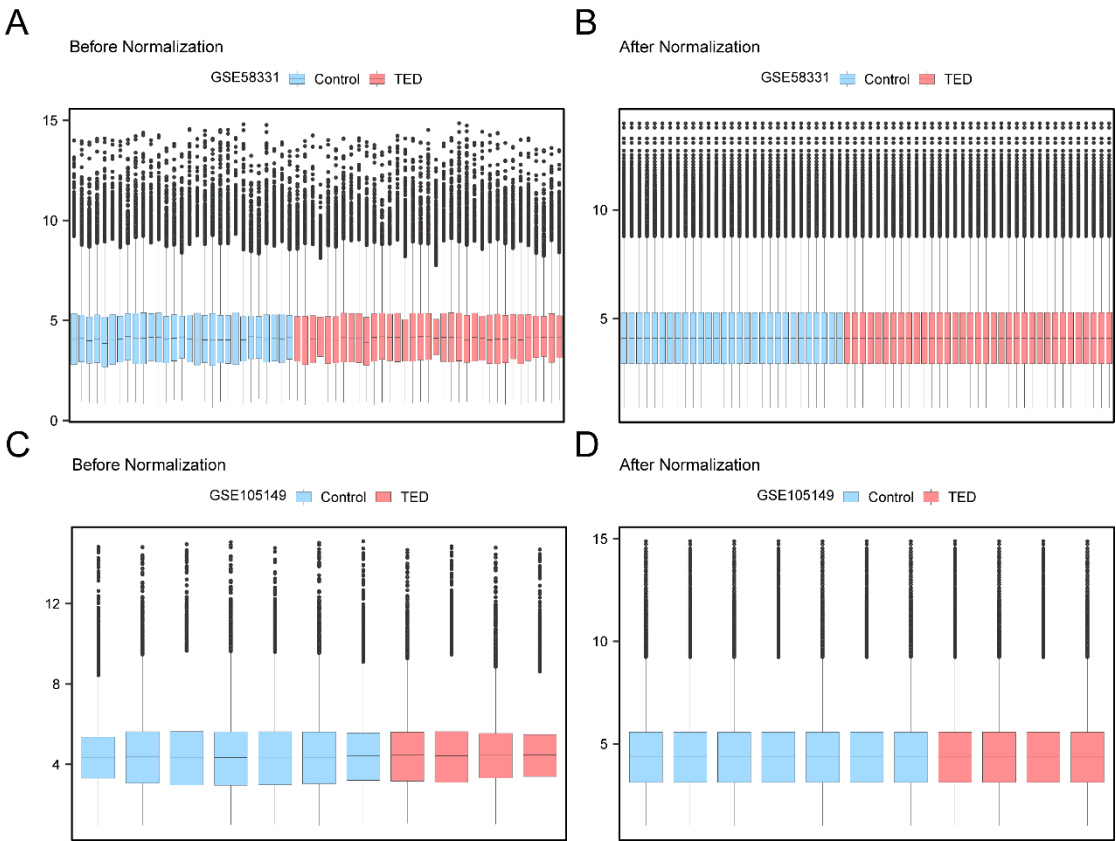


Figure 1. data de-batching process. A. Boxplot of the distribution of dataset GSE58331 before de-batching. B. Boxplot of the distribution of dataset GSE58331 after de-batching. C. Boxplot of the distribution of dataset GSE105149 before de-batching. D. Boxplot of the distribution of dataset GSE105149 after de-batching. The light red color is for the TED group and the light blue color is for the Control group.

2.2. TED-related Differentially Expressed MMRG-Mediated Signaling Pathways

The GSE58331 dataset had a total of 3994 DEGs; up-expressed genes had a total of 1972, and down-expressed genes had a total of 2022, and the GSE105149 dataset had a total of 3952 DEGs; the up-expressed genes had a total of 2027, and the down-expressed genes had a total of 1925, and the results of the differential analysis of the dataset (Fig.2A, Fig.2B).

The intersection (Fig.2C) was taken for all TED samples of genes and MEMRGs in all GSE58331-DEGs, GSE105149-DEGs obtained. A total of 26 MEMRDEGs were obtained, namely *UQCRC1*, *ALDH1A1*, *GPI*, *IMMT*, *GCK*, *NDUFS3*, *IDH1*, *ETFA*, *UQCRC2*, *PGK1*, *MFF*, *COX6C*, *BOLA3*, *SOAT2*, *MDH1*, *C1QBP*, *ALDH3A1*, *NDUFV2*, *NDUFS1*, *AKR1A1*, *ALDH3A2*, *DHTKD1*, *FXN*, *ACADS*, *SLC25A47*, *EPO*. Based on the intersection results, the expression differences of MEMRDEGs between different sample subgroups in GSE58331 and GSE105149 datasets were analyzed separately and heatmaps were plotted to show the results of the analysis using the R package pheatmap (Fig.2D-E).

BOLA3 displays contrasting expression patterns in the GSE58331 and GSE105149 datasets, possibly due to various factors. Variations in sample origins and processing methods, with GSE58331 utilizing anterior orbital tissues and GSE105149 using lacrimal glands, may impact gene expression levels. The smaller sample size in GSE105149 could result in greater result instability and variability. Furthermore, clinical sample heterogeneity, including diverse biological backgrounds and disease stages among patients, may contribute to expression disparities of the same gene across datasets.

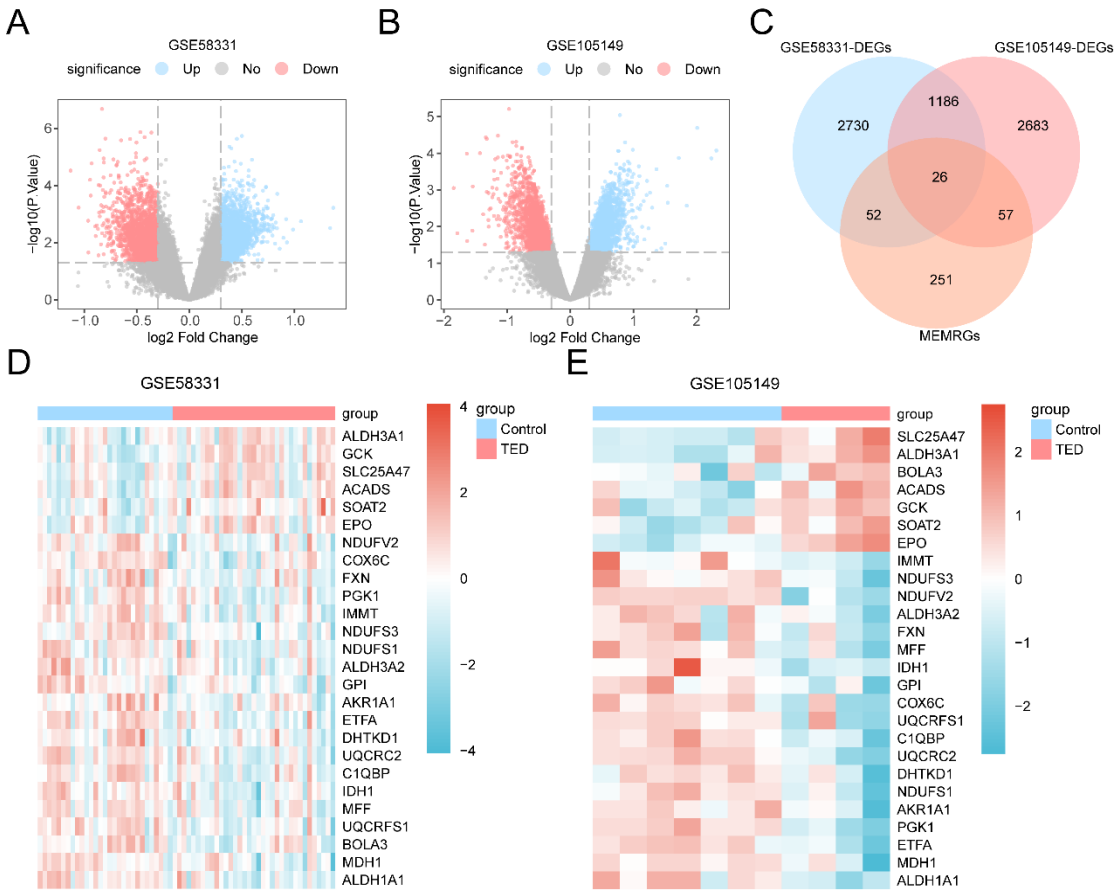


Figure 2. differential gene expression analysis. A-B. Volcano plots of DEGs analyzed in the TED group and Control group in the GSE58331 (A) and GSE105149 datasets (B). C. Wayne plots of all the genes in TED and MEMRGs in the DEGs in the GSE58331 and GSE105149 datasets. D-E. Heatmaps of all the genes in DEGs and MEMRGs in the GSE58331 (D) and Heatmap of MEMRDEGs in GSE105149 dataset (E). DEGs: Differentially Expressed Genes; MEMRGs: Mitochondrial Energy Metabolism-Related Genes; MEMRDEGs: Mitochondrial Energy Metabolism-Related Differentially Expressed Genes. The light red color is for the TED group and the light blue color is for the Control group. Red color in the heatmap represents high expression and blue color represents low expression.

2.3. Analysis of GO and KEGG

A total of 26 MEMRDEGs were subjected to GO and KEGG enrichment analyses as shown in Table 1. The results showed that 26 MEMRDEGs were predominantly enriched in TED. The results of the analyses were visualized by bubble plots (Fig.3A). Meanwhile, the network diagrams of biological process (BP), cellular component (CC), molecular function (MF) and biological pathway (KEGG) were plotted based on the GO and KEGG enrichment analysis (Fig.3B-E). The connecting lines illustrated the corresponding molecules and the annotations of the corresponding entries; the larger the node, the more molecules the entry contained.

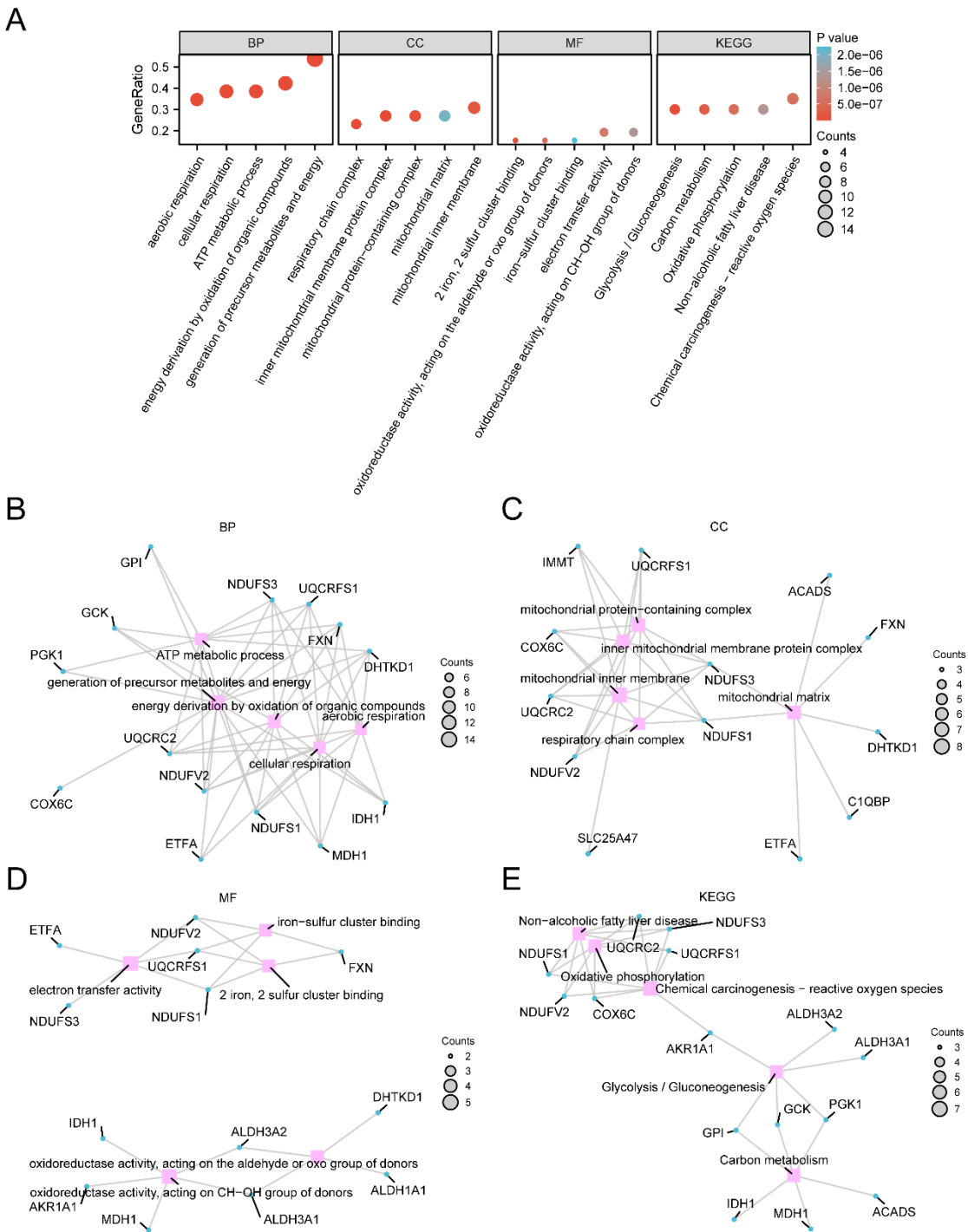


Figure 3. GO and KEGG enrichment analysis of MEMRDEGs. A. Bubble diagram presentation of GO and KEGG enrichment analysis resulted for MEMRDEGs : BP, CC, MF and KEGG. Horizontal coordinates are GO and KEGG terms. B-E. MEMRDEGs of GO and KEGG Enrichment analysis results network graphs showing: BP (B), CC (C), MF (D) and KEGG (E). The pink nodes represent the entries, the blue nodes represent the molecules, and the connecting lines represent the relationship between the entries and the molecules. MEMRDEGs: Mitochondrial Energy Metabolism-Related Differentially Expressed Genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; BP: Biological Process; CC: Cellular Component; MF: Molecular Function. The size of the bubbles in the bubble diagram represents the number of genes, and the color of the bubbles represents the magnitude of the p value, and the redder the color, the smaller the p value, the bluer the color the larger the p value.

2.4. GSEA in TED

The association between the expression of all genes in the dataset GSE58331 and the biological processes involved, the cellular components affected, and the molecular functions performed was

investigated by GSEA (Fig.4 A), and the specific results were shown in Table 2. The results showed that all genes in the dataset GSE58331 were significantly enriched in Fatty Acids (Fig.4 B), the Overview of Proinflammatory and Profibrotic Mediators (Fig.4 C), Dilated Cardiomyopathy (Fig.4 D), ADORA2B Mediated Anti Inflammatory Cytokines Production (Fig.4 E) and other biologically relevant functions and signaling pathways.

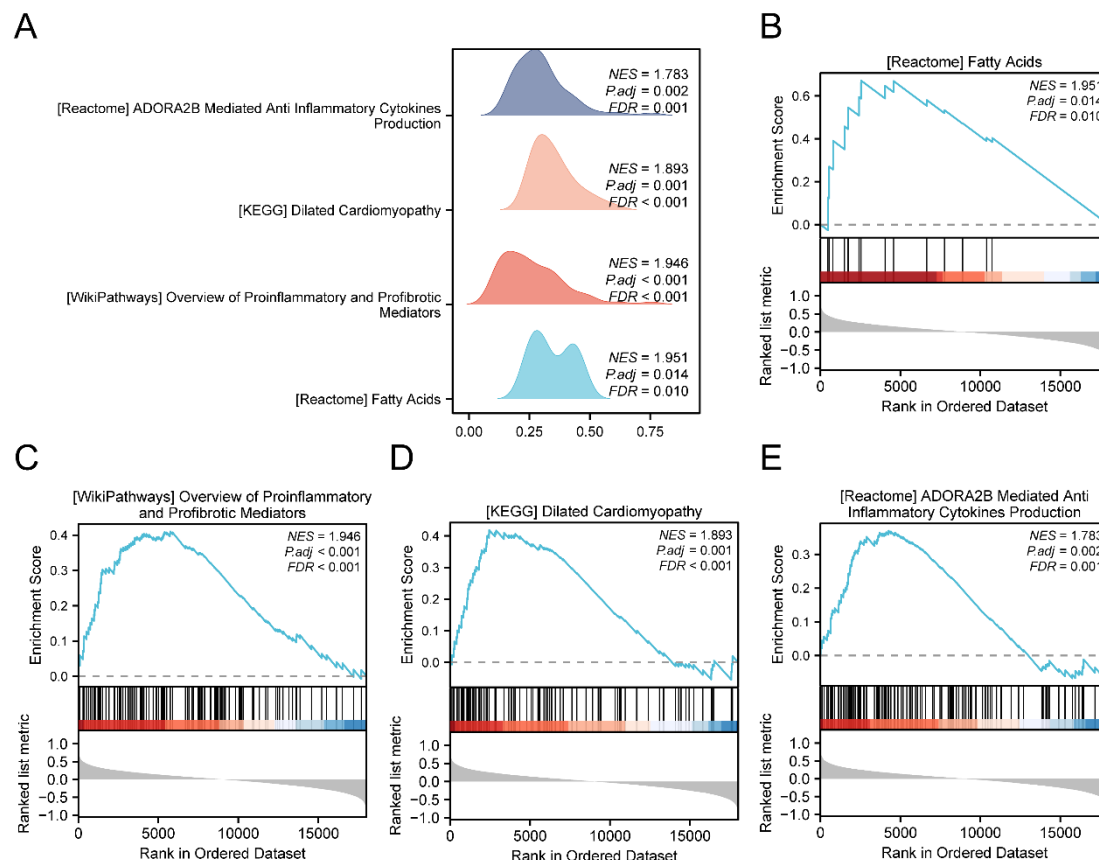


Figure 4. GSEA analysis of thyroid-related eye disease. The GSEA 4 biological functions of dataset GSE58331 were shown in mountain range plots.B-E. GSEA showed significant enrichment of all genes in Fatty Acids (B), Overview of Proinflammatory and Profibrotic Mediators (C), Dilated Cardiomyopathy (D), ADORA2B Mediated Anti Inflammatory Cytokines Production (E). GSEA: Gene Set Enrichment Analysis.

2.5. Screening of Key Genes

To determine the diagnostic value of MEMRDEGs in TED, 26 MEMRDEGs were screened for significance using one-way logistic regression (p value <0.05). The RF algorithm was used to analyze the expression of these 26 MEMRDEGs in the TED/Control disease subgroup of the dataset GSE58331. Setting the seed to 234 and the number of decision trees to 300, we plotted the decision tree error curve (Fig.5A), which showed that the error reached its minimum and leveled off when the number of decision trees was around 50. We then plotted the MeanDecreaseGini scatterplot (Fig.5B) for important gene screening. The number of genes was rounded off by performing five ten-fold cross-validation and plotting the cross-validation error plot (Fig.5C). The graph showed that the model error was minimized when the number of genes was 12, which was then combined with MeanDecreaseGini to select specific genes for subsequent analysis. The results showed (Fig.5B-C) that the 12 MEMRDEGs screened by this algorithm that had a significant impact on the diagnosis of TED were, in descending order of importance, *GPI*, *COX6C*, *NDUFS3*, *MDH1*, *ALDH1A1*, *IDH1*, *FXN*, *IMMT*, *MFF*, *GCK*, *UQCRRS1*, *NDUFV2*.

LASSO risk models were constructed by LASSO regression analysis for these 12 MEMRDEGs. The LASSO regression model was plotted (Fig.5D) and the trajectory of LASSO variables (Fig.5E). The results showed that the LASSO risk model contained a total of 10 MEMRDEGs, which were *IDH1*,

IMMT, *MDH1*, *NDUFS3*, *NDUFV2*, *GPI*, *UQCRCF1*, *GCK*, *ALDH1A1*, and *MFF*. These genes were taken as Key Genes for our subsequent study, and the forest plot of Key Genes was drawn (Fig.5F).

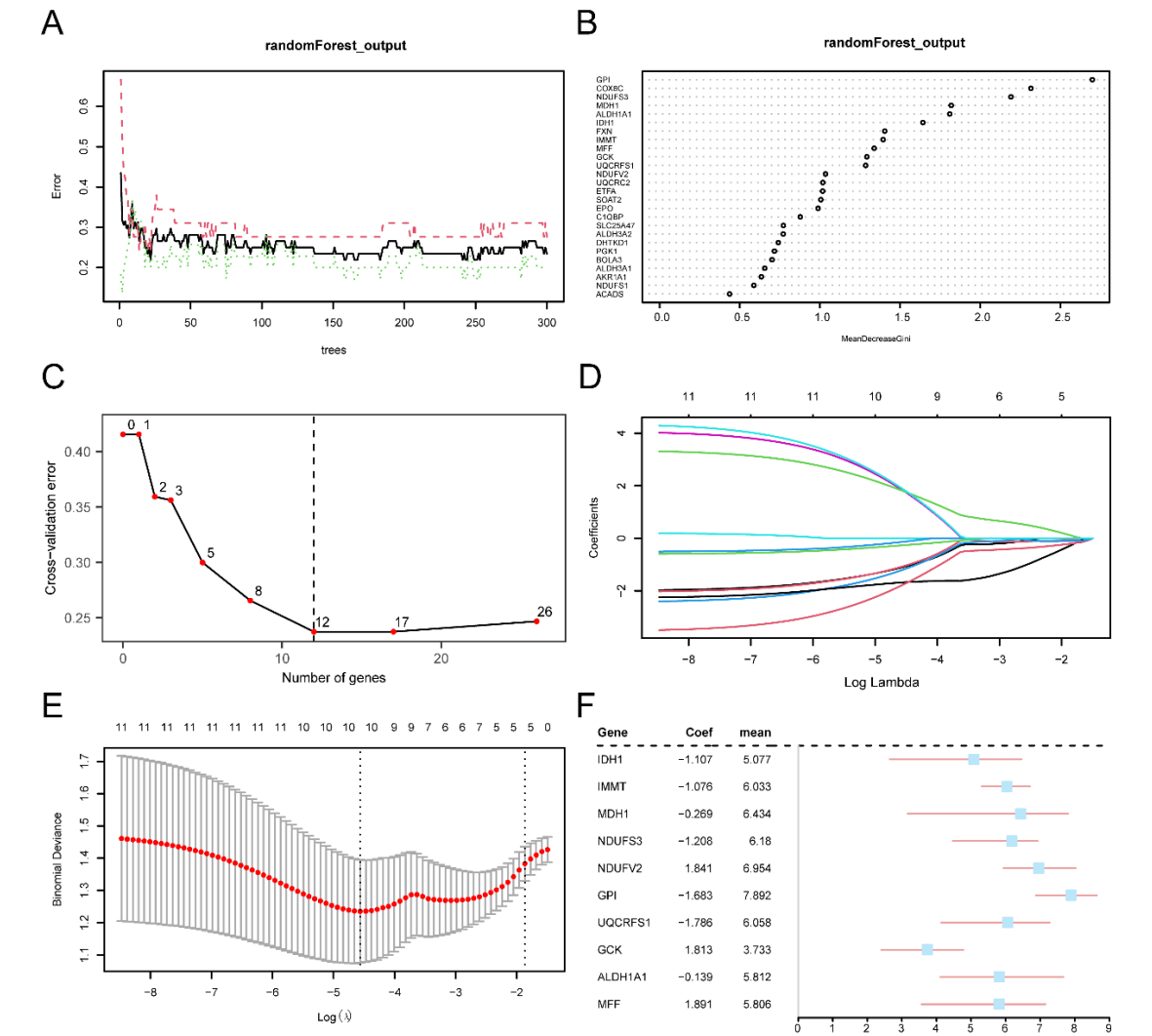


Figure 5. Screening of Key Genes. A. Model training error plot for the Random Forest algorithm. B. MeanDecreaseGini scatterplot for MEMRDEGs (in descending order of MeanDecreaseGini). C. Cross validation error plot. D. Diagnostic model plot for the LASSO regression model. E. Variable trajectories for the LASSO regression model. F. Forest plot for Key Genes in the LASSO regression model. MEMRDEGs: Mitochondrial Energy Metabolism-Related Differentially Expressed Genes; LASSO: Least Absolute Shrinkage and Selection Operator.

2.6. Key Genes Constructing Diagnostic Logistic Regression Models

The Risk Score formula for the dataset GSE58331 was as follows:
$$\text{RiskScore} = -1.9687 * IDH1 - 2.0103 * IMMT - 0.6120 * MDH1 - 2.3690 * NDUFS3 + 4.1636 * NDUFV2 - 2.2142 * GPI - 3.5576 * UQCRCF1 + 3.3602 * GCK - 0.5756 * ALDH1A1 + 4.3804 * MFF$$

The TED patients were classified into high-risk and low-risk groups using the risk score. Then we plotted the column line graph (Fig. 6A) to show the connection between the 10 Key Genes, and we could see that the expression of *MFF* contributes the most to the multifactorial logistic model.

In order to judge the accuracy and discriminative power of the multifactor logistic model, the Calibration Curve was plotted by Calibration analysis, and the predictive effect of the model on the actual results was assessed based on the fit between the actual probabilities and the probabilities predicted by the model in different cases plotted in the graph (Fig.6B). The calibration curve plot of the multifactor logistic model showed that the calibration line shown by the dashed line basically

coincided with the diagonal line of the ideal model, which indicated that the model had high accuracy and discriminatory power.

The diagnostic effectiveness of the diagnostic multifactorial logistic model for TED diagnosis was evaluated by decision curve analysis (DCA) based on the dataset GSE58331(Fig.6C). The results showed that the line of the model was stably higher than All and None in a certain range, and the net gain of the model was higher, indicating that the model was more effective in diagnosis.

Finally, ROC curves were plotted using the R package “pROC” based on the Risk Score in the datasets GSE58331 and GSE105149 to validate the diagnostic utility of TED with the diagnostic multifactorial logistic model. showing that the multifactorial logistic model presented a high accuracy for the diagnosis of TED (Fig.6D-E. AUC = 0.910, AUC = 1.000).

The high or low score of Friends analysis was used to determine the genes that TED played an important role in the biological process (Fig.6F). The results showed that *NDUFS3* played an important role in TED as the gene closest to the cut-off value (cut-off value = 0.60).

The positions of 10 Key Genes on human chromosomes were analyzed by R package “RCircos” (Fig.6G). The chromosome localization map showed that more Key Genes were located on chromosome 2, namely, *MDH1*, *IMMT*, *IDH1*, and *MFF* were located on chromosome 2(Fig.6G).

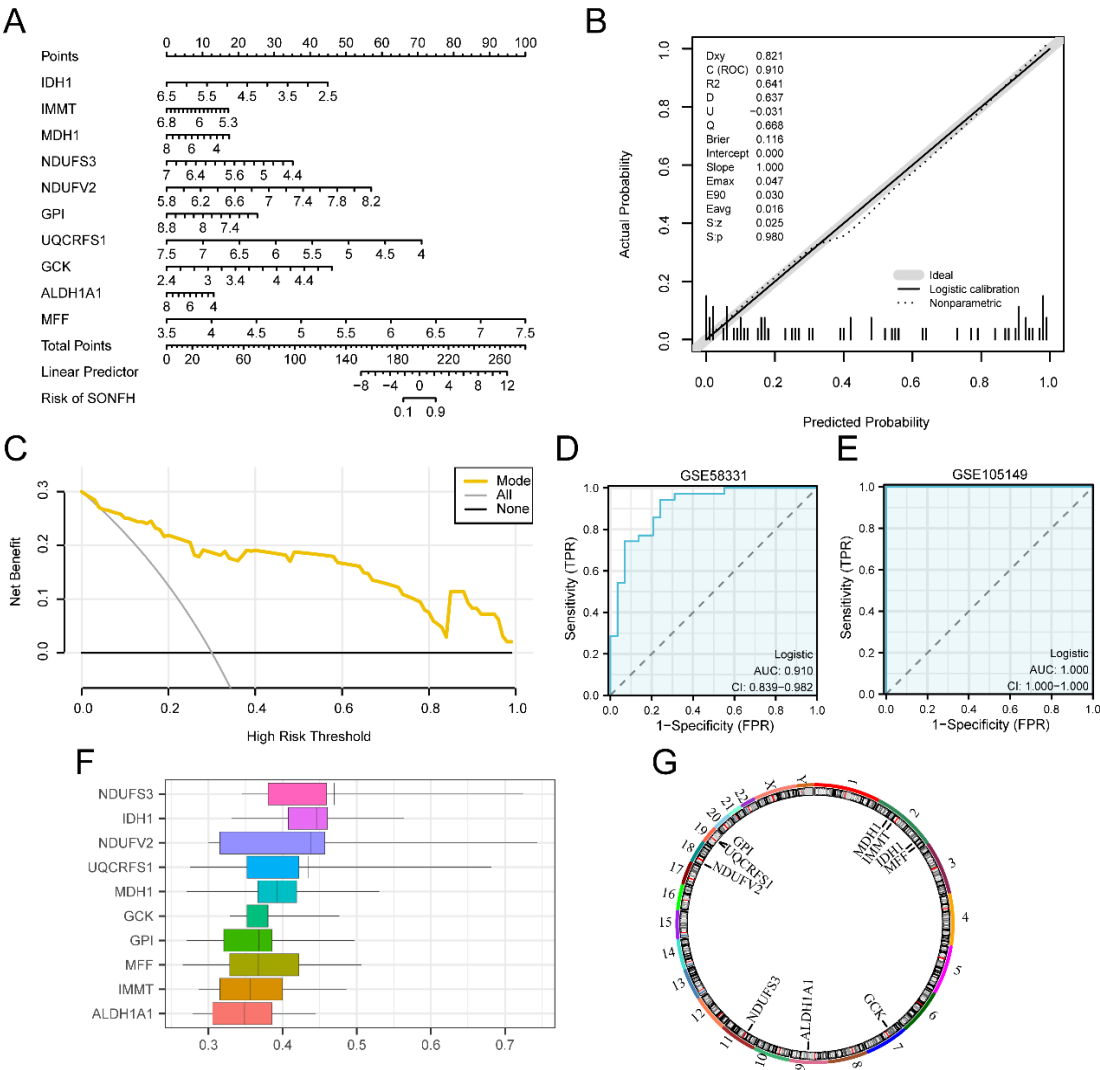


Figure 6. Key Genes constructs diagnostic logistic regression models. A. Column line plot (Nomogram) of Key Genes in diagnostic multifactor logistic model based on dataset GSE58331. B. Calibration curve plot of Key Genes in diagnostic multifactor logistic model based on dataset GSE58331. C. DCA plot of Key Genes based on the diagnostic multifactor logistic model for dataset GSE58331. D-E. Diagnostic ROC curves of diagnostic multifactorial logistic model Risk Score in datasets GSE58331 (D) and GSE105149 (E). F. Friends plots of Key Genes. G. Chromosomal localization map of Key Genes. DCA plots have net gain in the vertical coordinate and

probability threshold or Threshold Probability (TPR) in the horizontal coordinate. DCA: Decision Curve Analysis; ROC: Receiver Operating Characteristic; TPR: True Positive Rate; FPR: False Positive Rate.

2.7. Expression Differential Validation Analysis and Functional Similarity Analysis of Key Genes

In order to explore the expression differences of 10 Key Genes, the results of the difference analysis of in dataset GSE58331 in TED group and Control group with high and low expression were demonstrated by group comparison plot (Fig.7A). The difference results showed (Fig.7A) that six Key Genes had highly statistically significant (p value < 0.001) expression in the TED group and Control group of dataset GSE58331, namely, *IMMT*, *NDUFS3*, *GPI*, *UQCRFS1*, *GCK*, and *ALDH1A1*. Finally, using the R package "pROC" to draw ROC curves based on the expression of Key Genes in the dataset GSE58331. The ROC curves (Fig.7B-F) showed that the expression levels of all Key Genes were very similar to those in the classification of TED group and Control group presented some accuracy ($0.7 < \text{AUC} < 0.9$).

Using the same methodology, the analysis of differences in expression levels of 10 Key Genes between the TED and Control groups in dataset GSE105149 revealed significant findings. A group comparison plot (Fig.7G) displayed that *IDH1*, *IMMT*, and *NDUFS3* exhibited highly significant expression variances ($p < 0.01$) between the TED and Control groups. Subsequently, the R package "pROC" was utilized to generate ROC curves based on dataset GSE105149. The ROC curves (Fig.7H-L) indicated that *IDH1*, *IMMT*, *MDH1*, *NDUFS3*, *NDUFV2*, *ALDH1A1*, and *MFF* expression levels among the Key Genes were highly accurate in distinguishing the TED and Control groups ($\text{AUC} > 0.9$). Additionally, *GPI*, *UQCRFS1*, and *GCK* expression levels showed moderate accuracy in classifying the TED and Control groups ($0.7 < \text{AUC} < 0.9$).

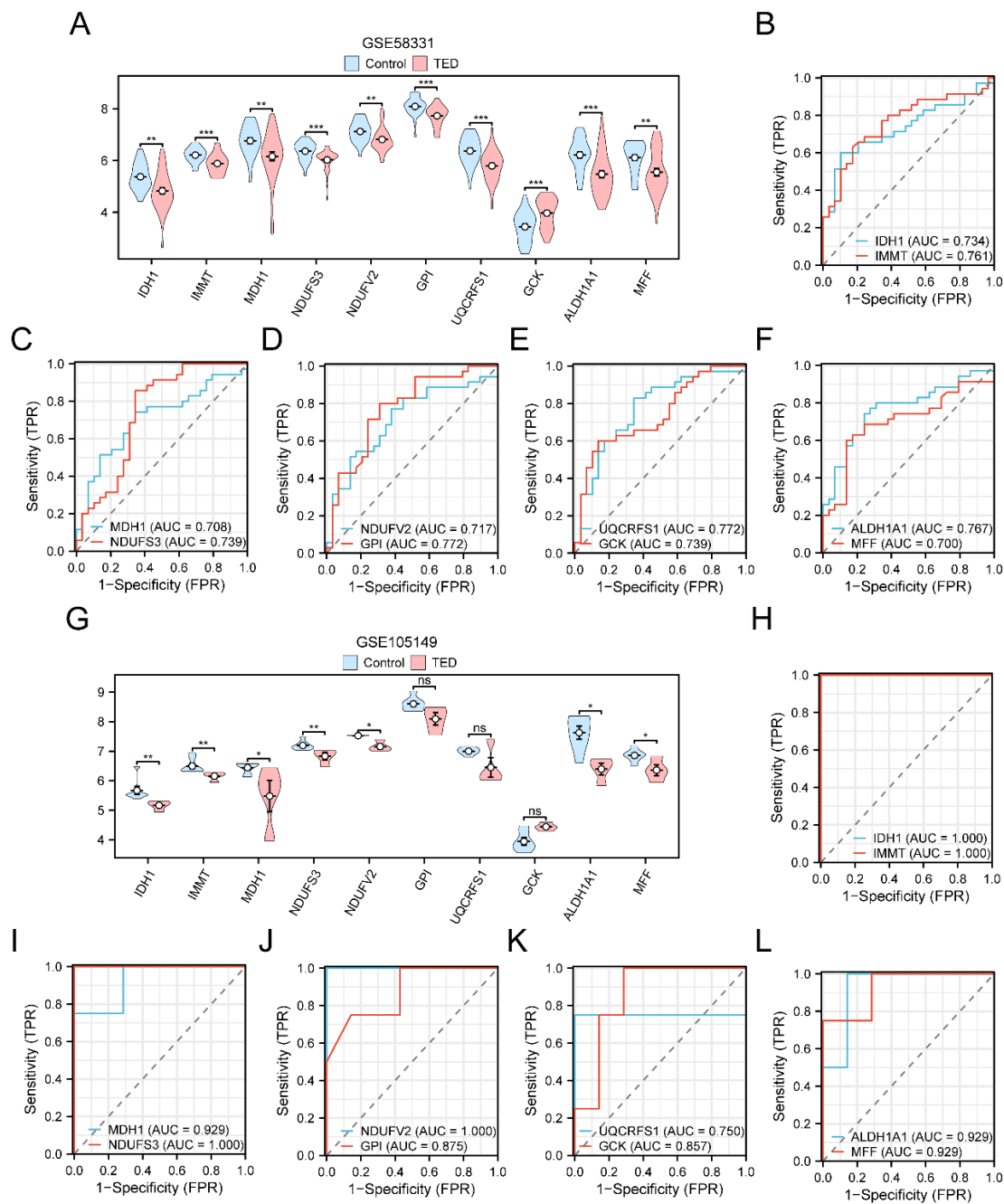


Figure 7. Expression difference validation analysis of Key Genes. A. Group comparison plots of Key Genes in the TED group and Control group of dataset GSE58331. B-F. Key Genes in *IDH1* and *IMMT* (B), *MDH1* and *NDUFS3* (C), *NDUFV2h* and *GPI* (D), *UQCRCF1* and *GCK* (E), *ALDH1A1* and *MFF* (F) in the ROC curves in dataset GSE58331. G. Group comparison plots of Key Genes in dataset GSE105149 in the TED group and Control group. H-L. Key Genes in *IDH1* and *IMMT* (H), *MDH1* and *NDUFS3* (I), *NDUFV2* and *GPI* (J), *UQCRCF1* and *GCK* (K), *ALDH1A1* and *MFF* (L) in the ROC curves in dataset GSE105149. ns represents p value ≥ 0.05 , not statistically significant; * represents p value < 0.05 , statistically significant; ** represents p value < 0.01 , highly statistically significant; *** represents p value ≤ 0.001 , extremely high statistically significant. When AUC > 0.5 , it indicates that the expression of the molecule is a trend to promote the event. The closer the AUC is to 1, the better the diagnosis. ROC, Receiver Operating Characteristic; AUC: Area Under the Curve; TPR: True Positive Rate; FPR: False Positive Rate. Light blue color represents the Control group and light red color represents the TED group.

2.8. GSEA Based on Logistic Risk Score High/Low Grouping

In order to determine the effect of the expression levels of all genes in dataset GSE58331 on TED between high and low risk score subgroups, the association between the expression of all 21655 genes

in dataset GSE58331 of TED between different high and low risk score subgroups (High/Low) and the biological processes involved, cellular components affected, and molecular functions performed were investigated by GSEA . (Fig.8A), the specific results were shown in Table 3. The results showed that all genes in the dataset GSE58331 were significantly enriched in Activation of the Tfp2 Ap 2 Family of Transcription Factors (Fig.8B), Negative Regulation of Activity of the Tfp2 Ap 2 Family of Transcription Factors (Fig.8C), Signaling By Interleukins (Fig.8D), IL18 Signaling Pathway (Fig.8E), and other biologically relevant functions and signaling pathways.

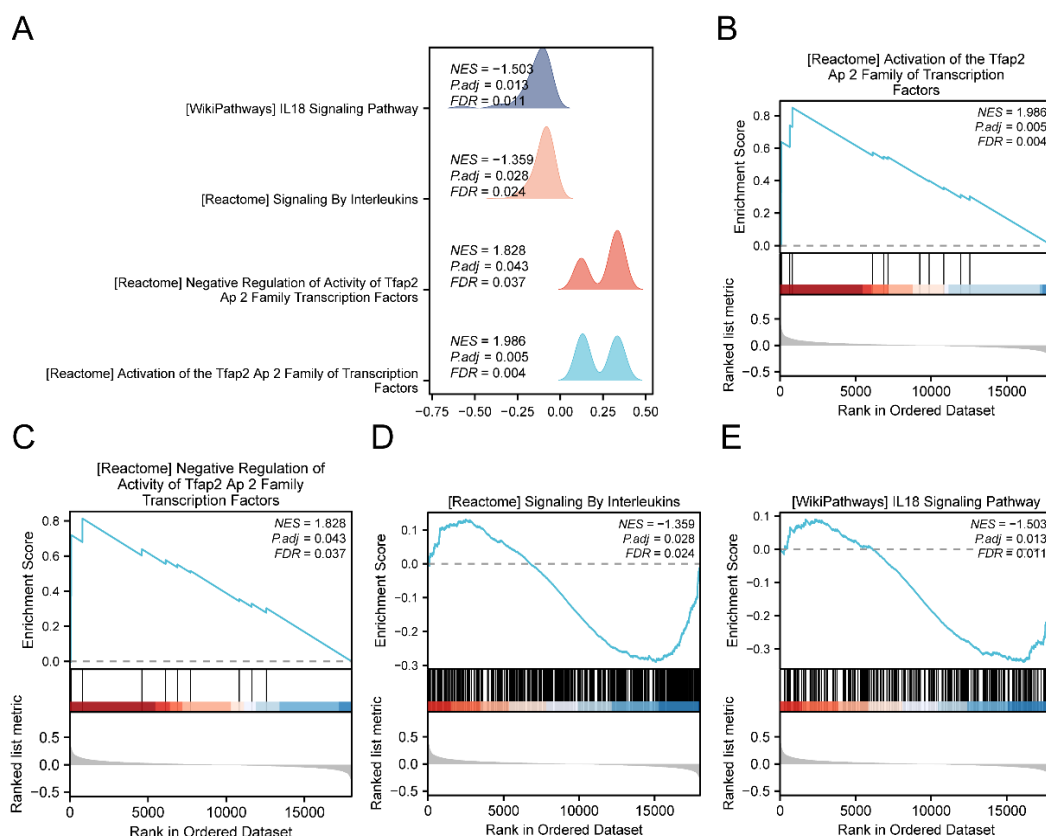


Figure 8. GSEA of high and low risk groups in TED. A. GSEA of the dataset GSE58331 for 4 biological functions shown in mountain range plots. B-E. GSEA showed significant enrichment of all genes in Activation of the Tfp2 Ap 2 Family of Transcription Factors (B), Negative Regulation of Activity of the Tfp2 Ap 2 Family Transcription Factors (C), Signaling By Interleukins (D), IL18 Signaling Pathway (E). GSEA: Gene Set Enrichment Analysis.

2.9. ssGSEA Algorithm Based on Logistic Risk Score High/Low Grouping for Immune Infiltration Analysis

The expression matrix of TED samples from dataset GSE58331 was applied to calculate the immune infiltration abundance of 28 immune cells in TED the high and low risk groups by the ssGSEA algorithm. First, the differences in the expression of immune cell infiltration abundance in different subgroups were demonstrated by group comparison plots (Fig.9A) , showing that activated CD8 T cell, Immature B cell and T follicular helper cell were statistically significant (p value < 0.05). Finally, the correlation of Key Genes with the abundance of immune cell infiltration was demonstrated by correlation bubble plots (Fig.9B-C). The results of correlation bubble plots showed that TED Key Genes showed a strong positive correlation with immune cells in the Low Risk (LR) and High Risk (HR) groups. Among them, *UQCERS1* showed the strongest positive correlation with T follicular helper cells in the Low Risk group (r value = 0.529, p value < 0.05); *MDH1* showed the strongest positive correlation with T follicular helper cells in the High Risk group (r value = 0.733, p value < 0.05).

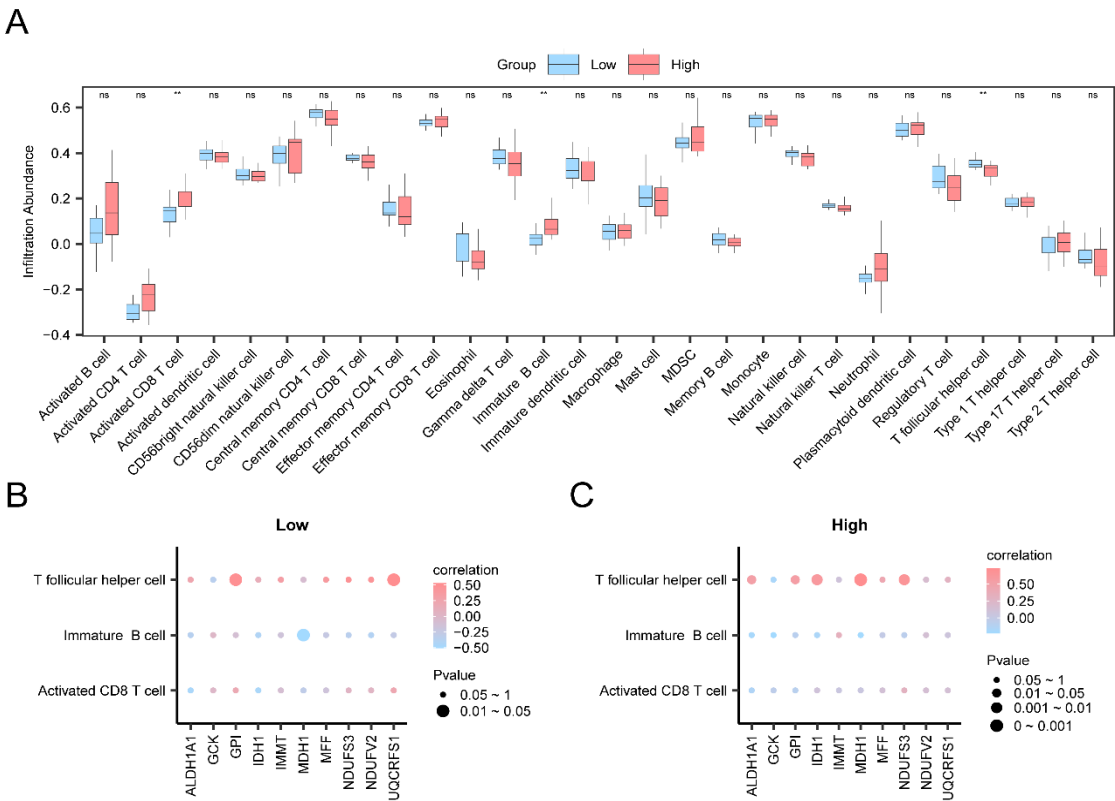


Figure 9. Immune infiltration analysis based on ssGSEA algorithm. A. Comparative plots of immune cells grouped in the TED group and Control group of dataset GSE58331. B-C. Bubble plots of correlation between immune cell infiltration abundance and Key Genes in the low risk (Low Risk) group (B) and the high risk (High Risk) group (C) of the TED . ssGSEA: single-sample Gene-Set Enrichment Analysis; TED: Thyroid Eye Disease. ns represents a p value ≥ 0.05 with no statistical significance; ** represents p value < 0.01 , highly statistically significant. Correlation coefficients (r value) with absolute values between 0.5 and 0.8 were considered moderately correlated. Light purple color was for the Control group and light red color was for the TED group. Red color was positive correlation and blue color was negative correlation. Shades of color represent the strength of the correlation.

2.10. Interaction Networks Analysis of mRNA-miRNA, mRNA-TF, and mRNA-RBP at Key Genes

Firstly, miRNAs associated with 10 Key Genes were obtained from ENCORI database, and the mRNA-miRNA Interaction Network was constructed and visualized using Cytoscape software (Fig.10A). Among them, 6 Key Genes (*ALDH1A1*, *GPI*, *IDH1*, *MFF*, *NDUFS3*, *UQCRCF1*) and 44 miRNAs were included.

Next, the TFs combined with Key Genes were obtained from ChIPBase database, and the mRNA-TF Interaction Network was constructed and visualized using Cytoscape software (Fig.10B). Among them, 8 Key Genes (*IDH1*, *ALDH1A1*, *GSK*, *IMMT*, *MDH1*, *MFF*, *NDUFV2*, *UQCRCF1*) and 33 TFs were included.

Finally, RBPs associated with Key Genes were predicted by the StarBase database, and the mRNA-RBP Interaction Network was constructed and visualized using Cytoscape software (Fig.10C). Among them, 8 Key Genes (*ALDH1A1*, *GPI*, *IDH1*, *IMMT*, *MDH1*, *MFF*, *NDUFS3*, *NDUFV2*) and 36 RBPs were included.

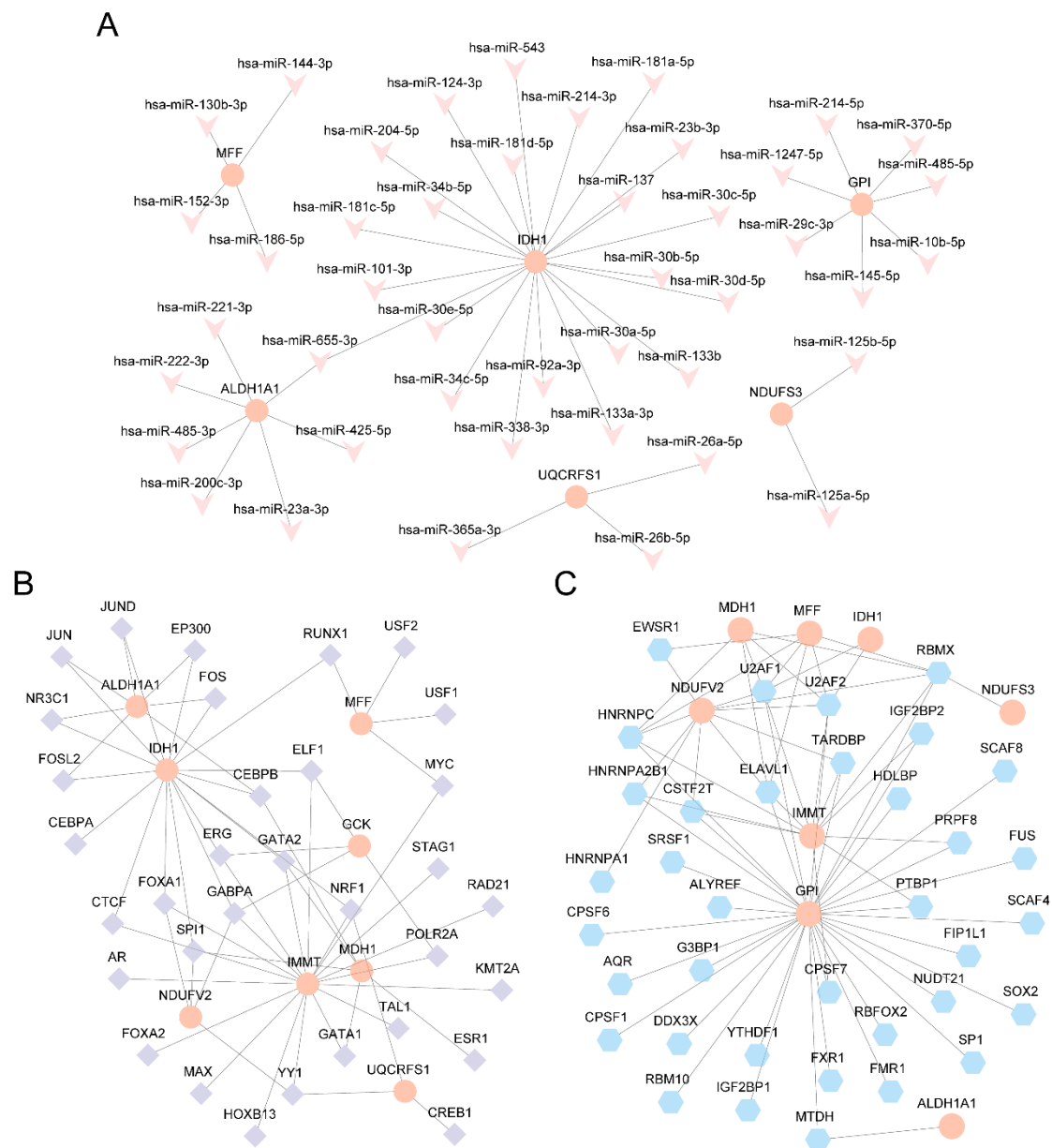


Figure 10. Interaction network analysis of Key Genes. A. Key Genes' mRNA-miRNA Interaction Network. B. Key Genes' mRNA-TF Interaction Network. C. Key Genes' mRNA-RBP Interaction Network. TF: Transcription Factor; RBP: RNA-Binding Protein. Orange is mRNA, purple is TF, Pink is miRNA and blue is RBP.

Table 1. Results of GO and KEGG Enrichment Analysis for MEMRDEGs.

ONTO			Gene	BgRa	pval	p.adj	qvalu
LOGY	ID	Description	Ratio	tio	ue	ust	e
BP	GO:00	generation of precursor metabolites		494/1	4.512	3.091	2.247
	06091	and energy	14/26	8800	9E-16	3E-13	E-13
BP	GO:00	energy derivation by oxidation of		321/1	1.864	6.386	4.641
	15980	organic compounds	11/26	8800	6E-13	2E-11	8E-11
BP	GO:00			231/1	2.886	6.590	4.790
	45333	cellular respiration	10/26	8800	4E-13	6E-11	4E-11
BP	GO:00			273/1	1.530	2.621	1.905
	46034	ATP metabolic process	10/26	8800	9E-12	7E-10	6E-10
BP	GO:00			187/1	2.121	2.906	2.112
	09060	aerobic respiration	9/26	8800	7E-12	7E-10	7E-10
CC	GO:00			491/1	1.542	1.388	6.495
	05743	mitochondrial inner membrane	8/26	9594	6E-07	3E-06	E-07

	GO:00	inner mitochondrial membrane		155/1	9.809	4.640	2.170
CC	98800	protein complex	7/26	9594	2E-10	3E-08	9E-08
	GO:00	mitochondrial protein-containing		281/1	6.031	6.333	2.963
CC	98798	complex	7/26	9594	8E-08	4E-07	E-07
	GO:00			473/1	2.019	1.590	7.440
CC	05759	mitochondrial matrix	7/26	9594	5E-06	3E-05	2E-06
	GO:00			91/19	1.813	4.640	2.170
CC	98803	respiratory chain complex	6/26	594	7E-09	3E-08	9E-08
	GO:00			125/1	7.812	2.167	1.237
MF	09055	electron transfer activity	5/26	8410	9E-07	2E-05	7E-05
	GO:00	oxidoreductase activity, acting on CH-		140/1	1.369	2.167	1.237
MF	16614	OH group of donors	5/26	8410	3E-06	2E-05	7E-05
	GO:00			24/18	3.257	3.061	1.748
MF	51537	2 iron, 2 sulfur cluster binding	4/26	410	2E-08	7E-06	6E-06
	GO:00	oxidoreductase activity, acting on the		46/18	4.897	2.167	1.237
MF	16903	aldehyde or oxo group of donors	4/26	410	9E-07	2E-05	7E-05
	GO:00			67/18	2.254	2.649	1.513
MF	51536	iron-sulfur cluster binding	4/26	410	8E-06	3E-05	1E-05
	hsa05	Chemical carcinogenesis - reactive		223/8	5.914	9.610	4.824
KEGG	208	oxygen species	7/20	164	4E-07	9E-06	9E-06
	hsa00			67/81	8.612	5.598	2.810
KEGG	010	Glycolysis / Gluconeogenesis	6/20	64	3E-09	E-07	3E-07
	hsa01			115/8	2.261	7.350	3.689
KEGG	200	Carbon metabolism	6/20	164	6E-07	1E-06	9E-06
	hsa00			134/8	5.608	9.610	4.824
KEGG	190	Oxidative phosphorylation	6/20	164	8E-07	9E-06	9E-06
	hsa04			155/8	1.322	1.719	8.631
KEGG	932	Non-alcoholic fatty liver disease	6/20	164	6E-06	4E-05	6E-06

GO, Gene Ontology; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; KEGG, Kyoto Encyclopedia of Genes and Genomes; MEMRDEGs. Mitochondrial Energy Metabolism-Related Differentially Expressed Genes.

Table 2. Results of GSEA for GSE58331.

ID	set Siz e	enrichm entScor e	NES	pval ue	p.adj ust	qval ue
			1.950	0.001	0.013	0.010
		0.671421	9898	4869	8059	1944
REACTOME_FATTY_ACIDS	14	52	5	6	3	8
			1.945	5.686	0.000	
WP_OVERVIEW_OF_PROINFLAMMATORY_ AND_PROFIBROTIC_MEDIATORS	117	7	0.410293	9672	9E-06	1931
			7	6.988	2	0.000
		0.418312	1.893	3E-05	0.001	9650
KEGG_DILATED_CARDIOMYOPATHY	88	41	3955	05	3069	3
REACTOME_ADORA2B_MEDIATED_ANTI_I NFLAMMATORY_CYTOKINES_PRODUCTIO N	127	97	1.782	0.000	0.002	0.001
		0.370305	5569	1225	0250	4953
			2	3	8	4

GSEA, Gene Set Enrichment Analysis.

Table 3. Results of GSEA (Low/High) for GSE58331.

ID	set Size	enrichmentScore	NES	pvalue	p.adjust	qvalue
			1.986	9.73	0.00	0.00
REACTOME_ACTIVATION_OF_THE_TFAP2_AP_2_FAMILY_OF_TRANSCRIPTION_FACTORS	12	0.84970469	25648	75E-05	524769	449219
REACTOME_NEGATIVE_REGULATION_OF_ACTIVITY_OF_TFAP2_AP_2_FAMILY_TRANSCRIPTION_FACTORS	10	0.81338831	1.8275564	0.00231583	0.04332249	0.03708547
		-	-	-	-	-
	44	-	1.3589434	0.001078	0.027515	0.023554
REACTOME_SIGNALING_BY_INTERLEUKINS	1	0.29000623	94345	107874	751538	355407
		-	-	-	-	-
	25	-	1.5025658	0.000333	0.013329	0.011410
WP_IL18_SIGNALING_PATHWAY	9	0.33916019	56587	033336	33291	141015

3. Discussion

The prevalence of Thyroid Eye Disease (TED) among Graves' disease patients is approximately 25%, making it the most prevalent orbital condition. Existing treatments such as medications, surgery, and radiation have not consistently yielded satisfactory outcomes and can potentially result in complications and adverse effects. While numerous studies have explored the pathogenesis of TED and potential treatment targets, the precise mechanisms remain elusive. Contrasting gene expression profiles between TED and other orbital inflammatory conditions have shown weaker associations with inflammatory markers in TED, indicating a relatively subdued inflammatory response in TED[19]. This implies that targeting metabolic or alternative pathways could be more effective in treating Thyroid Eye Disease (TED). Dysregulation in mitochondrial energy metabolism pathways is linked to oxidative stress. Inflammation is implicated in a range of degenerative conditions, acute illnesses, and the aging process. Mitochondria are key players in amplifying inflammatory signals, leading to heightened mitochondrial oxidative stress and perpetuating a detrimental cycle of inflammation[43]. There is currently no definitive target or treatment mechanism in the study of mitochondrial energy metabolism-related genes in TED, which holds potential and guiding significance for the diagnosis and treatment of TED.

Following GO and KEGG analyses, 26 differentially expressed mitochondria-related genes (MEMRDEGs) were identified as enriched in various biological processes, cellular components, molecular functions, and pathways in Thyroid Eye Disease (TED). The primary aim of this study is to evaluate the diagnostic potential of these 26 MEMRDEGs in TED. While the specific regulation status (upregulation or downregulation) of these genes in TED was not detailed, their enrichment in the disease context suggests diagnostic relevance. Our analysis focuses on the collective impact of these genes in TED pathogenesis. Gene Set Enrichment Analysis (GSEA) revealed significant enrichment of gene expression levels related to TED in dataset GSE58331, particularly in pathways involving fatty acids, proinflammatory and profibrotic mediators, and ADORA2B-mediated anti-inflammatory cytokine production. Mitochondrial energy metabolism pathways encompass glycolysis, the tricarboxylic acid cycle, oxidative phosphorylation, ketone body metabolism, lipid metabolism, and gluconeogenesis. Oxidative phosphorylation, crucial for ATP synthesis, stands out as a central pathway in energy metabolism. The activation of the ADORA2B receptor triggers the cyclic adenosine monophosphate (cAMP) pathway and associated downstream pathways, linked to immunosuppression and potential applications in asthma and tumor immunotherapy[44]. GSEA indicated significant enrichment related to oxidative phosphorylation and ATP energy production within mitochondria. By categorizing high and low-risk groups using Logistic risk scoring, GSEA demonstrated significant enrichment of TED genes across pathways such as Signaling By Interleukins, IL18 Signaling Pathway, and other relevant biological functions. In the context of

intracytoplasmic mitochondrial DNA and ROS, the activation of inflammatory vesicles can lead to IL-1 β and IL-18 secretion upon release from dysfunctional mitochondria[45]. We suggest that the mitochondrial energy metabolism pathway in TED involves inflammasome signaling pathways. These findings suggested that MEMRGs were involved in the immune aspects of TED and were associated with the pathogenesis of TED. Therefore, we further compared the immune infiltration abundance in TED grouped by high and low risk score.

Using the ssGSEA algorithm to assess immune cell infiltration abundance in Thyroid Eye Disease (TED) high and low-risk groups, significant findings ($p < 0.05$) were observed for Activated CD8 T cells, Immature B cells, and T follicular helper cells. Key Genes displayed strong positive correlations with immune cells in both Low Risk and High Risk TED groups. Notably, *UQCERS1* exhibited the strongest positive correlation with T follicular helper cells in the Low Risk group ($r = 0.529$, $p < 0.05$), while *MDH1* showed the strongest positive correlation with T follicular helper cells in the High Risk group ($r = 0.733$, $p < 0.05$). Previous immunohistochemical studies have linked cytokines such as interferon- γ (IFN- γ), tumor necrosis factor- α (TNF- α), and interleukin-1 α (IL-1 α) with T-cell infiltration in TED connective tissues. During the Active phase of TED, CD4+ and CD45RO+ T cells are abundant in early orbital infiltration, secreting cytokines and chemokines to enhance the immune response. Follicular helper T cells (Tfh) play a crucial role as regulators of germinal center (GC) B cells and in sustaining long-term protective humoral responses[46,47]. Tfh cells aided in immunoglobulin class switch reorganization and supported germinal center responses, thereby promoting immunoglobulin affinity maturation and the generation of humoral immune memory[48]. While their primary function was to promote B-cell responses, Tfh cells also exhibited phenotypic and functional diversity depending on the immunologic and spatial environment they encounter[49]. *UQCERS1* was highly expressed in gastric and breast cancers. Knocking out *UQCERS1* cells can reduce cell proliferation, induce cell cycle arrest in G1 phase, increase the proportion of apoptosis, ROS production and DNA damage gene expression, and inhibit the AKT/mTOR pathway[50]. Dysregulation of Tfh and follicular regulation T cells had been reported to be positively associated with the pathogenic process of autoimmune diseases[51].

The LASSO regression model identified 10 key genes in Thyroid Eye Disease (TED): *IDH1*, *IMMT*, *MDH1*, *NDUFS3*, *NDUFV2*, *GPI*, *UQCERS1*, *GCK*, *ALDH1A1*, and *MFF*. ROC curves revealed that *IDH1*, *IMMT*, *MDH1*, *NDUFS3*, *NDUFV2*, *ALDH1A1*, and *MFF* expression levels are highly accurate in distinguishing between the TED and Control groups ($AUC > 0.9$). *NDUFS3*, highlighted by Friends analysis, was deemed crucial in TED's biological processes (cut-off value = 0.60). In invasive ductal breast carcinoma samples, the mitochondrial complex I subunit *NDUFS3* exhibited a distinct signature as a marker for hypoxia/necrosis. *NDUFS3* expression levels were significantly and positively associated with nuclear grading, enabling improved diagnosis and survival rate predictions for breast cancer patients[52]. Furthermore, Complex I (CI) is the largest enzyme in the mitochondrial respiratory chain, and its defects were a major cause of mitochondrial disease. CI disassembly processes were identified as occurring in a hierarchical and modular manner when the amount of *NDUFS3* gradually decreases[53]. Inner membrane mitochondrial protein (*IMMT*) is crucial for preserving mitochondrial structural integrity, regulating mitochondrial dynamics, and contributing to various cellular functions. It plays a role in cell cycle advancement and mitochondrial antioxidant protection. Elevated *IMMT* levels are linked to lung adenocarcinoma and independently predict a poor prognosis in patients with this condition. *IMMT* has the potential to serve as a novel prognostic marker and therapeutic target in lung adenocarcinoma[54]. *IMMT* played a crucial role in shaping the inner mitochondrial membrane, regulating metabolism and innate immunity. Low expression of *IMMT* in tumors was associated with mitochondrial inhibition and activation of angiogenesis, indicating a poor prognosis for patients with kidney renal clear cell carcinoma (KIRC) and being linked to the progression of KIRC[55]. *NDUFS3* and *IMMT* exhibited high expression levels in Thyroid Eye Disease (TED) datasets and were linked to mitochondrial energy metabolism pathways, potentially influencing TED's immune responses. These genes are considered promising research directions and potential therapeutic targets for TED. However, our study has limitations. While we identified associations between 26 differentially expressed genes and TED, lack of biological validation hinders a comprehensive understanding of their roles in the disease's

pathophysiology. Resource constraints prevent further wet lab experiments, and the absence of clinical relevance studies limits analysis with clinical data. Future research aims to address these gaps through experimental validation. Despite combining data from GSE105149 and GSE58331 to enhance statistical power, the relatively small sample size may impact result reliability. Relying on publicly available data introduces potential bias, necessitating future studies with independent datasets and experimental confirmation to bolster our findings.

4. Conclusions

The study examined differential expression of mitochondria-related genes (MEMRGs) in Thyroid Eye Disease (TED) and Control groups, identifying 10 Key Genes for a TED diagnostic model. Immunoinfiltration analysis demonstrated a notable positive correlation between these key genes and immune cells in TED, offering valuable insights for TED diagnosis and prognosis. Nevertheless, additional validation is necessary to uncover precise pathogenic mechanisms and target effects, necessitating thorough investigation.

Author Contributions : Conceptualization, Jingrong Li; Data curation, Wenlian Zheng and Diyao Wei; Methodology, Hong-Gang Liang, Hansheng Huang and Lulu Qin; Resources, Haiyan Dai and Quan Xiao; Supervision, Minghui Hou; Writing – original draft, Bi Lv and Xianghui Wei; Writing – review & editing, Huanyan Wang. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), GeneCards database (<https://www.genecards.org/>), PubMed website (<https://pubmed.ncbi.nlm.nih.gov/>).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. J.H. Yu, Y.H. Wang, H.F. Liao, [Research status of potential therapeutic targets for thyroid-associated ophthalmopathy]., *Zhonghua Yan Ke Za Zhi* 60 (2024) 282-288. 10.3760/cma.j.cn112142-20231116-00235.
2. V.M. Naik, M.N. Naik, R.A. Goldberg, T.J. Smith, R.S. Douglas, Immunopathogenesis of thyroid eye disease: emerging paradigms, *Surv Ophthalmol* 55 (2010) 215-226. 10.1016/j.survophthal.2009.06.009.
3. J. Pritchard, R. Han, N. Horst, W.W. Cruikshank, T.J. Smith, Immunoglobulin activation of T cell chemoattractant expression in fibroblasts from patients with Graves' disease is mediated through the insulin-like growth factor I receptor pathway, *J Immunol* 170 (2003) 6348-6354. 10.4049/jimmunol.170.12.6348.
4. A.P. Weetman, M.E. Yateman, P.A. Ealey, C.M. Black, C.B. Reimer, R.C. Williams, Jr., B. Shine, N.J. Marshall, Thyroid-stimulating antibody activity between different immunoglobulin G subclasses, *J Clin Invest* 86 (1990) 723-727. 10.1172/jci114768.
5. T.J. Smith, J. Janssen, Insulin-like Growth Factor-I Receptor and Thyroid-Associated Ophthalmopathy, *Endocr Rev* 40 (2019) 236-267. 10.1210/er.2018-00066.
6. L. Wang, J.M. Ma, [Progression of the pathogenesis of thyroid associated ophthalmopathy]., *Zhonghua Yan Ke Za Zhi* 53 (2017) 474-480. 10.3760/cma.j.issn.0412-4081.2017.06.018.
7. R. Han, S. Tsui, T.J. Smith, Up-regulation of prostaglandin E2 synthesis by interleukin-1beta in human orbital fibroblasts involves coordinate induction of prostaglandin-endoperoxide H synthase-2 and glutathione-dependent prostaglandin E2 synthase expression, *J Biol Chem* 277 (2002) 16355-16364. 10.1074/jbc.M111246200.
8. G.J. Kahaly, Immunotherapies for thyroid eye disease, *Curr Opin Endocrinol Diabetes Obes* 26 (2019) 250-255. 10.1097/med.0000000000000493.
9. S. Yasuda, T. Atsumi, S. Shimamura, K. Ono, K. Hiromura, K. Sada, M. Mori, S. Takei, Y. Kawaguchi, N. Tamura, Y. Takasaki, Surveillance for the use of mycophenolate mofetil for adult patients with lupus nephritis in Japan, *Mod Rheumatol* 25 (2015) 854-857. 10.3109/14397595.2015.1031444.

10. T. Nie, Y.N. Lamb, Teprotumumab: A Review in Thyroid Eye Disease, *Drugs* 82 (2022) 1663-1670. 10.1007/s40265-022-01804-1.
11. A.C. Boese, S. Kang, Mitochondrial metabolism-mediated redox regulation in cancer progression, *Redox Biol* 42 (2021) 101870. 10.1016/j.redox.2021.101870.
12. A.A. Starkov, The role of mitochondria in reactive oxygen species metabolism and signaling, *Ann N Y Acad Sci* 1147 (2008) 37-52. 10.1196/annals.1427.015.
13. G. Egan, D.H. Khan, J.B. Lee, S. Mirali, L. Zhang, A.D. Schimmer, Mitochondrial and Metabolic Pathways Regulate Nuclear Gene Expression to Control Differentiation, Stem Cell Function, and Immune Response in Leukemia, *Cancer Discov* 11 (2021) 1052-1066. 10.1158/2159-8290.Cd-20-1227.
14. T. Huang, Y. Wang, Z. Wang, Q. Long, Y. Li, D. Chen, Complement-mediated inflammation and mitochondrial energy metabolism in the proteomic profile of myopic human corneas, *J Proteomics* 285 (2023) 104949. 10.1016/j.jprot.2023.104949.
15. J.T. Rosenbaum, D. Choi, D.J. Wilson, H.E. Grossniklaus, C.A. Harrington, R.A. Dailey, J.D. Ng, E.A. Steele, C.N. Czyz, J.A. Foster, D. Tse, C. Alabiad, S. Dubovy, P. Parekh, G.J. Harris, M. Kazim, P. Patel, V. White, P. Dolman, D.P. Edward, H. Alkatan, H. Al Hussain, D. Selva, P. Yeatts, B. Korn, D. Kikkawa, P. Stauffer, S.R. Planck, Fibrosis, gene expression and orbital inflammatory disease, *Br J Ophthalmol* 99 (2015) 1424-1429. 10.1136/bjophthalmol-2015-306614.
16. J.T. Rosenbaum, D. Choi, D.J. Wilson, H.E. Grossniklaus, C.A. Harrington, C.H. Sibley, R.A. Dailey, J.D. Ng, E.A. Steele, C.N. Czyz, J.A. Foster, D. Tse, C. Alabiad, S. Dubovy, P. Parekh, G.J. Harris, M. Kazim, P. Patel, V. White, P. Dolman, B.S. Korn, D. Kikkawa, D.P. Edward, H. Alkatan, H. Al-Hussain, R.P. Yeatts, D. Selva, P. Stauffer, S.R. Planck, Parallel Gene Expression Changes in Sarcoidosis Involving the Lacrimal Gland, Orbital Tissue, or Blood, *JAMA Ophthalmol* 133 (2015) 770-777. 10.1001/jamaophthalmol.2015.0726.
17. J.T. Rosenbaum, D. Choi, D.J. Wilson, H.E. Grossniklaus, C.A. Harrington, C.H. Sibley, R.A. Dailey, J.D. Ng, E.A. Steele, C.N. Czyz, J.A. Foster, D. Tse, C. Alabiad, S. Dubovy, P.K. Parekh, G.J. Harris, M. Kazim, P.J. Patel, V.A. White, P.J. Dolman, B.S. Korn, D.O. Kikkawa, D.P. Edward, H.M. Alkatan, H. al-Hussain, R.P. Yeatts, D. Selva, P. Stauffer, S.R. Planck, Orbital pseudotumor can be a localized form of granulomatosis with polyangiitis as revealed by gene expression profiling, *Exp Mol Pathol* 99 (2015) 271-278. 10.1016/j.yexmp.2015.07.002.
18. J.T. Rosenbaum, D. Choi, D.J. Wilson, H.E. Grossniklaus, C.H. Sibley, C.A. Harrington, S.R. Planck, Molecular diagnosis of orbital inflammatory disease, *Exp Mol Pathol* 98 (2015) 225-229. 10.1016/j.yexmp.2015.01.009.
19. J.T. Rosenbaum, D. Choi, A. Wong, D.J. Wilson, H.E. Grossniklaus, C.A. Harrington, R.A. Dailey, J.D. Ng, E.A. Steele, C.N. Czyz, J.A. Foster, D. Tse, C. Alabiad, S. Dubovy, P.K. Parekh, G.J. Harris, M. Kazim, P.J. Patel, V.A. White, P.J. Dolman, D.P. Edward, H.M. Alkatan, H. Al Hussain, D. Selva, R.P. Yeatts, B.S. Korn, D.O. Kikkawa, P. Stauffer, S.R. Planck, The Role of the Immune Response in the Pathogenesis of Thyroid Eye Disease: A Reassessment, *PLoS One* 10 (2015) e0137654. 10.1371/journal.pone.0137654.
20. A.J. Wong, S.R. Planck, D. Choi, C.A. Harrington, M.L. Troxell, D.C. Houghton, P. Stauffer, D.J. Wilson, H.E. Grossniklaus, R.A. Dailey, J.D. Ng, E.A. Steele, G.J. Harris, C. Czyz, J.A. Foster, V.A. White, P.J. Dolman, M. Kazim, P.J. Patel, D.P. Edward, H. al Katan, H. al Hussain, D. Selva, R.P. Yeatts, B.S. Korn, D.O. Kikkawa, J.T. Rosenbaum, IgG4 immunostaining and its implications in orbital inflammatory disease, *PLoS One* 9 (2014) e109847. 10.1371/journal.pone.0109847.
21. J.T. Rosenbaum, D. Choi, C.A. Harrington, D.J. Wilson, H.E. Grossniklaus, C.H. Sibley, S.S. Salek, J.D. Ng, R.A. Dailey, E.A. Steele, B. Hayek, C.M. Craven, D.P. Edward, A.M.Y. Maktabi, H. Al Hussain, V.A. White, P.J. Dolman, C.N. Czyz, J.A. Foster, G.J. Harris, Y.S. Bee, D.T. Tse, C.R. Alabiad, S.R. Dubovy, M. Kazim, D. Selva, R.P. Yeatts, B.S. Korn, D.O. Kikkawa, R.Z. Silkiss, J.A. Sivak-Callcott, P. Stauffer, S.R. Planck, Gene Expression Profiling and Heterogeneity of Nonspecific Orbital Inflammation Affecting the Lacrimal Gland, *JAMA Ophthalmol* 135 (2017) 1156-1162. 10.1001/jamaophthalmol.2017.3458.
22. K. Taguchi, S. Hamamoto, A. Okada, R. Unno, H. Kamisawa, T. Naiki, R. Ando, K. Mizuno, N. Kawai, K. Tozawa, K. Kohri, T. Yasui, Genome-Wide Gene Expression Profiling of Randall's Plaques in Calcium Oxalate Stone Formers, *J Am Soc Nephrol* 28 (2017) 333-347. 10.1681/asn.2015111271.

23. G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T.I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran, D. Lancet, The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses, *Curr Protoc Bioinformatics* 54 (2016) 1.30.31-31.30.33. 10.1002/cpbi.5.
24. Z. Ye, H. Zhang, F. Kong, J. Lan, S. Yi, W. Jia, S. Zheng, Y. Guo, X. Zhan, Comprehensive Analysis of Alteration Landscape and Its Clinical Significance of Mitochondrial Energy Metabolism Pathway-Related Genes in Lung Cancers, *Oxid Med Cell Longev* 2021 (2021) 9259297. 10.1155/2021/9259297.
25. J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, *Bioinformatics* 28 (2012) 882-883. 10.1093/bioinformatics/bts034.
26. M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res* 43 (2015) e47. 10.1093/nar/gkv007.
27. H. Mi, A. Muruganujan, D. Ebert, X. Huang, P.D. Thomas, PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools, *Nucleic Acids Res* 47 (2019) D419-d426. 10.1093/nar/gky1038.
28. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res* 28 (2000) 27-30. 10.1093/nar/28.1.27.
29. G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, *Omics* 16 (2012) 284-287. 10.1089/omi.2011.0118.
30. A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc Natl Acad Sci U S A* 102 (2005) 15545-15550. 10.1073/pnas.0506580102.
31. A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J.P. Mesirov, Molecular signatures database (MSigDB) 3.0, *Bioinformatics* 27 (2011) 1739-1740. 10.1093/bioinformatics/btr260.
32. Y. Liu, H. Zhao, Variable importance-weighted Random Forests, *Quant Biol* 5 (2017) 338-351.
33. [W. Cai, M. van der Laan, Nonparametric bootstrap inference for the targeted highly adaptive least absolute shrinkage and selection operator (LASSO) estimator, *Int J Biostat* (2020). 10.1515/ijb-2017-0070.
34. S. Engebretsen, J. Bohlin, Statistical predictions with glmnet, *Clin Epigenetics* 11 (2019) 123. 10.1186/s13148-019-0730-1.
35. J. Wu, H. Zhang, L. Li, M. Hu, L. Chen, B. Xu, Q. Song, A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis, *Cancer Commun (Lond)* 40 (2020) 301-312. 10.1002/cac2.12067.
36. B. Van Calster, L. Wynants, J.F.M. Verbeek, J.Y. Verbakel, E. Christodoulou, A.J. Vickers, M.J. Roobol, E.W. Steyerberg, Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators, *Eur Urol* 74 (2018) 796-804. 10.1016/j.eururo.2018.08.038.
37. G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, S. Wang, GOSemSim: an R package for measuring semantic similarity among GO terms and gene products, *Bioinformatics* 26 (2010) 976-978. 10.1093/bioinformatics/btq064.
38. H. Zhang, P. Meltzer, S. Davis, RCircos: an R package for Circos 2D track plots, *BMC Bioinformatics* 14 (2013) 244. 10.1186/1471-2105-14-244.
39. B. Xiao, L. Liu, A. Li, C. Xiang, P. Wang, H. Li, T. Xiao, Identification and Verification of Immune-Related Gene Prognostic Signature Based on ssGSEA for Osteosarcoma, *Front Oncol* 10 (2020) 607622. 10.3389/fonc.2020.607622.
40. K.R. Zhou, S. Liu, W.J. Sun, L.L. Zheng, H. Zhou, J.H. Yang, L.H. Qu, ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data, *Nucleic Acids Res* 45 (2017) D43-d50. 10.1093/nar/gkw965.
41. J. Huang, W. Zheng, P. Zhang, Q. Lin, Z. Chen, J. Xuan, C. Liu, D. Wu, Q. Huang, L. Zheng, S. Liu, K. Zhou, L. Qu, B. Li, J. Yang, ChIPBase v3.0: the encyclopedia of transcriptional regulations of non-coding RNAs and protein-coding genes, *Nucleic Acids Res* 51 (2023) D46-d56. 10.1093/nar/gkac1067.
42. A. Singh, RNA-binding protein kinetics, *Nat Methods* 18 (2021) 335. 10.1038/s41592-021-01122-6.

43. M.J. López-Armada, R.R. Riveiro-Naveira, C. Vaamonde-García, M.N. Valcárcel-Ares, Mitochondrial dysfunction and the inflammatory response, *Mitochondrion* 13 (2013) 106-118. 10.1016/j.mito.2013.01.003.
44. L.N. Strickland, E.Y. Faraoni, W. Ruan, X. Yuan, H.K. Eltzschig, J.M. Bailey-Lundberg, The resurgence of the Adora2b receptor as an immunotherapeutic target in pancreatic cancer, *Front Immunol* 14 (2023) 1163585. 10.3389/fimmu.2023.1163585.
45. S. Marchi, E. Guilbaud, S.W.G. Tait, T. Yamazaki, L. Galluzzi, Mitochondrial control of inflammation, *Nat Rev Immunol* 23 (2023) 159-173. 10.1038/s41577-022-00760-x.
46. J.L. Cannons, K.T. Lu, P.L. Schwartzberg, T follicular helper cell diversity and plasticity, *Trends Immunol* 34 (2013) 200-207. 10.1016/j.it.2013.01.001.
47. M.G. Scherm, V.B. Ott, C. Daniel, Follicular Helper T Cells in Autoimmunity, *Curr Diab Rep* 16 (2016) 75. 10.1007/s11892-016-0770-2.
48. C.G. Vinuesa, M.A. Linterman, D. Yu, I.C. MacLennan, Follicular Helper T Cells, *Annu Rev Immunol* 34 (2016) 335-368. 10.1146/annurev-immunol-041015-055605.
49. W. Song, J. Craft, T Follicular Helper Cell Heterogeneity, *Annu Rev Immunol* (2023). 10.1146/annurev-immunol-090222-102834.
50. Q. Sun, J. Li, H. Dong, J. Zhan, X. Xiong, J. Ding, Y. Li, L. He, J. Wang, UQCERS1 serves as a prognostic biomarker and promotes the progression of ovarian cancer, *Sci Rep* 13 (2023) 8335. 10.1038/s41598-023-35572-z.
51. J. Qi, C. Liu, Z. Bai, X. Li, G. Yao, T follicular helper cells and T follicular regulatory cells in autoimmune diseases, *Front Immunol* 14 (2023) 1178792. 10.3389/fimmu.2023.1178792.
52. S. Suhane, D. Berel, V.K. Ramanujan, Biomarker signatures of mitochondrial NDUF53 in invasive breast carcinoma, *Biochem Biophys Res Commun* 412 (2011) 590-595. 10.1016/j.bbrc.2011.08.003.
53. L. D'Angelo, E. Astro, M. De Luise, I. Kurelac, N. Umesh-Ganesh, S. Ding, I.M. Fearnley, G. Gasparre, M. Zeviani, A.M. Porcelli, E. Fernandez-Vizarra, L. Iommarini, NDUF53 depletion permits complex I maturation and reveals TMEM126A/OPA7 as an assembly factor binding the ND4-module intermediate, *Cell Rep* 35 (2021) 109002. 10.1016/j.celrep.2021.109002.
54. Y. Hiyoshi, Y. Sato, M. Ichinoe, R. Nagashio, D. Hagiuda, M. Kobayashi, S. Kusuhaara, S. Igawa, K. Shiomi, N. Goshima, Y. Murakumo, M. Saegusa, Y. Satoh, N. Masuda, K. Naoki, Prognostic significance of IMMT expression in surgically-resected lung adenocarcinoma, *Thorac Cancer* 10 (2019) 2142-2151. 10.1111/1759-7714.13200.
55. C.C. Chen, P.Y. Chu, H.Y. Lin, Supervised Learning and Multi-Omics Integration Reveals Clinical Significance of Inner Membrane Mitochondrial Protein (IMMT) in Prognostic Prediction, Tumor Immune Microenvironment and Precision Medicine for Kidney Renal Clear Cell Carcinoma, *Int J Mol Sci* 24 (2023). 10.3390/ijms24108807.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.