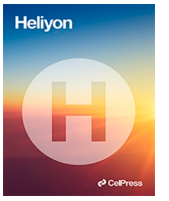
Heliyon 10 (2024) e34432

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Identification of lysosome-related hub genes as potential biomarkers and immune infiltrations of moyamoya disease by multiple bioinformatics methods and machine-learning strategies

Wenyang Li [a,1], Xiang Zhao [b,1], Jinxing Fu [b,1], Lei Cheng [a,*]

[a] Department of Neurosurgery, The Affiliated Hospital of Qingdao University, Qingdao, 266000, China
[b] Department of Neurosurgery, Shengjing Hospital of China Medical University, Shenyang, Liaoning, China

ARTICLE INFO

Keywords:
Moyamoya disease
Lysosome
Bioinformatic analysis
Machine-learning strategies
Drug

ABSTRACT

Background: Moyamoya disease (MMD), characterized by chronic cerebrovascular pathology, poses a rare yet significant clinical challenge, associated with elevated rates of mortality and disability. Despite intensive research endeavors, the exact biomarkers driving its pathogenesis remain enigmatic.

Methods: The expression patterns of GSE189993 and GSE141022 were retrieved from the Gene Expression Omnibus (GEO) repository to procure differentially expressed genes (DEGs) between samples afflicted with MMD and those under control conditions. The Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine with Recursive Feature Elimination (SVM-RFE), and Random Forest (RF) algorithms were employed for identifying candidate diagnostic genes associated with MMD. Subsequently, these candidate genes underwent validation in an independent cohort (GSE157628). The CMAP database was ultimately employed to forecast drugs pertinent to MMD for clinical translation.

Results: A collective of 240 DEGs were discerned. Functional enrichment scrutiny unveiled the enrichment of the cholesterol metabolism pathway, salmonella infection pathway, and allograft rejection pathway within the MMD cohort. EPDR1, DENND3, and NCSTN emerged as discerned diagnostic biomarkers for MMD. The CMAP database was ultimately employed to scrutinize the ten most auspicious pharmaceutical compounds for managing MMD. Finally, after validation through in vitro experiments, EPDR1, DENND3, and NCSTN were identified as the key genes.

Conclusion: EPDR1, DENND3, and NCSTN have emerged as potential novel biomarkers for MMD. The involvement of T lymphocytes, neutrophilic granulocytes, dendritic cells, natural killer cells, and plasma cells could be pivotal in the pathogenesis and advancement of MMD.

1. Introduction

Moyamoya disease (MMD) is a rare cerebrovascular disorder first delineated by Japanese investigators in 1957 [1]. It is characterized by progressive stenosis of the superior internal carotid artery (ICA) and chiefly presents within the Willis circle [2]. The prevalence of MMD in the Asian population exhibits a significant increase, primarily originating from China, Japan, and the Republic

* Corresponding author.
E-mail address: qdfyswcl@126.com (L. Cheng).
[1] Contribute equally.

https://doi.org/10.1016/j.heliyon.2024.e34432
Received 4 July 2024; Received in revised form 9 July 2024; Accepted 9 July 2024
Available online 10 July 2024

Abbreviations list

MMD	Moyamoya disease
GEO	Gene Expression Omnibus
DEGs	differentially expressed genes
LASSO	Least Absolute Shrinkage Selection Operator
SVM-RFE	Support Vector Machine-Recursive Feature Elimination
RF	Random Forest
ICA	superior carotid artery
GSEA	Gene set enrichment analysis
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
ROC:	receiver operating characteristic
AUC	area under the receiver operating characteristic curve
CMAP	The Connectivity Map
BP	biological process
CC	cellular component
MF	molecular function

of Korea, as evidenced by a substantial incidence rate [3]. The primary demographic affected by MMD comprises children (aged 5–9) and young adults (aged 35–45) [4]. Although the etiology of MMD remains elusive, contemporary investigations propose that genetic, immune, and environmental elements may contribute to its pathogenesis [5]. Research findings in the literature indicate that the collective occurrence rate of autoimmune disorders among individuals with Moyamoya disease MMD in the western regions of China could potentially reach 31.0 % [6]. Microarray technology was employed to detect genetic alterations, while bioinformatics analysis was utilized to explore potential causes and pathophysiological mechanisms at the genomic scale [7]. This potent technique enables the comprehensive investigation of gene expression patterns. The burgeoning evidence underscores the involvement of genetic elements in the progression of MMD [8]. Numerous gene variants and monogenic disorders are linked with moyamoya angiopathy [9]. In fact, moyamoya syndrome involves a spectrum of 16 genes and diverse pathways, encompassing Ras and MAPK, Notch signaling, DNA repair mechanisms, inflammasome activation, chromatin modification, actin dynamics, coagulation cascades, and nitric oxide (NO) signaling pathways [10]. Lysosomes, possessing a solitary membrane-bound vesicular architecture, house a variety of hydrolytic enzymes including phosphatases, lipases, proteases, nucleases, glycosidases, sulphate esterases, and other macromolecule-degrading enzymes, thereby exhibiting non-selective intracellular degradation capabilities [11]. Lysosomes play crucial roles in angiogenesis by regulating extracellular matrix degradation and cell signaling pathways [12]. Dysfunctions in lysosomal activities can lead to vascular diseases such as atherosclerosis, attributed to endothelial dysfunction and abnormal vascular cell proliferation [13]. Their involvement in these processes underscores the importance of understanding lysosomal mechanisms in vascular biology [14]. Insights into lysosomal functions provide potential avenues for novel therapeutic strategies targeting vascular-related disorders [15]. Prior investigations into genetic modifications, overall immune infiltration, and the association between central genes and immune cells in microarray datasets have encountered constraints due to the rarity of the ailment and difficulties in procuring samples. The precise involvement of lysosomes in the pathogenesis of MMD remains obscure. Consequently, investigating the interplay of immune cell infiltration amid lysosomal dynamics and MMD pathophysiology bears profound implications for therapeutic strategies, offering a novel vantage point to modulate disease initiation and advancement.

In this investigation, datasets pertaining to MMD were acquired from the Gene Expression Omnibus (GEO) repository and employed for the exploration of candidate diagnostic genes using machine learning techniques including the Least Absolute Shrinkage Selection Operator (LASSO), Support Vector Machine Recursive Feature Elimination (SVM-RFE), and Random Forest (RF) methodologies. The chosen marker genes underwent meticulous scrutiny and validation. Additionally, an exploration into the correlation between these genes and various categories of infiltrating immune cells was conducted. Utilizing bioinformatics methodologies, an investigation was undertaken to forecast small molecule therapeutics, presenting a fresh approach to understanding the molecular mechanisms of MMD and providing insights into effective strategies for intervention in this condition.

2. Materials and methods

2.1. Dataset processing

We acquired three MMD RNA chip datasets (GSE189993, GSE141022, GSE157628) from the Gene Expression Omnibus (GEO) at http://www.ncbi.nlm.nih.gov/geo. GEO serves as a publicly accessible archive, aggregating data from diverse high-throughput experiments. After normalization procedures, samples lacking clinical annotations were excluded from further analysis. Herein, a total of 32 samples were procured from the GSE189993 dataset (comprising 21 MMD samples and 11 controls), while the GSE141022 dataset contributed 8 samples (with 4 MMD and 4 control samples). Additionally, 20 samples were acquired from the GSE157628 dataset (consisting of 11 MMD samples and 9 controls). Each dataset, namely GSE189993, GSE141022, and GSE157628, underwent individual

gene annotation using corresponding platforms. Furthermore, GSE157628 was employed for subsequent validation and subjected to comparative analysis with both MMD and control samples. The flowchart of the research is displayed in Fig. 1.

2.2. Differentially expressed genes (DEGs) analysis

The batch effects were mitigated utilizing the "SVA" package in R programming [16], the DEGs were compared between samples afflicted with MMD and those from the control group utilizing the "limma" package within the R programming environment [17]. DEGs were defined by $|\text{Log 2 FC}| > 1$, adjusted p-value <0.05.

2.3. Functional and pathway enrichment analysis

Gene set enrichment analysis (GSEA) unveiled the markedly modulated Gene Ontology (GO) annotations and functional pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) between the MMD and control cohorts. Appropriate technical terminology was employed, and functional enrichment analysis was conducted on the genes that overlapped between DEGs and lysosomal genes. We clustered pertinent terms according to their shared attributes and then designated the term exhibiting the greatest enrichment level as the representative. To explore the functions and pathways associated with lysosome-related DEGs, we employed the "clusterProfiler" R package (version 4.0) to perform GO and KEGG analyses. Statistical significance was defined as a P-value less than 0.05.

2.4. Machine learning for the diagnostic marker genes

Three machine learning methodologies were utilized to forecast disease status. Initially, a methodology grounded on the least absolute shrinkage and selection operator (LASSO) was utilized to identify marker genes linked with discriminating MMD from control individuals, leveraging the "glmnet" toolbox in R. The most robust discriminatory capacity was harnessed for the identification of gene sets, leveraging support vector machine-recursive feature elimination (SVM-RFE) implemented via the "e1071" R software package. The ultimate selection of marker genes ensued from the shared genes across all three algorithms within the training cohort (GSE189993, GSE141022). The fundamental concept of the random forest (RF) algorithm involves discerning optimal outcomes

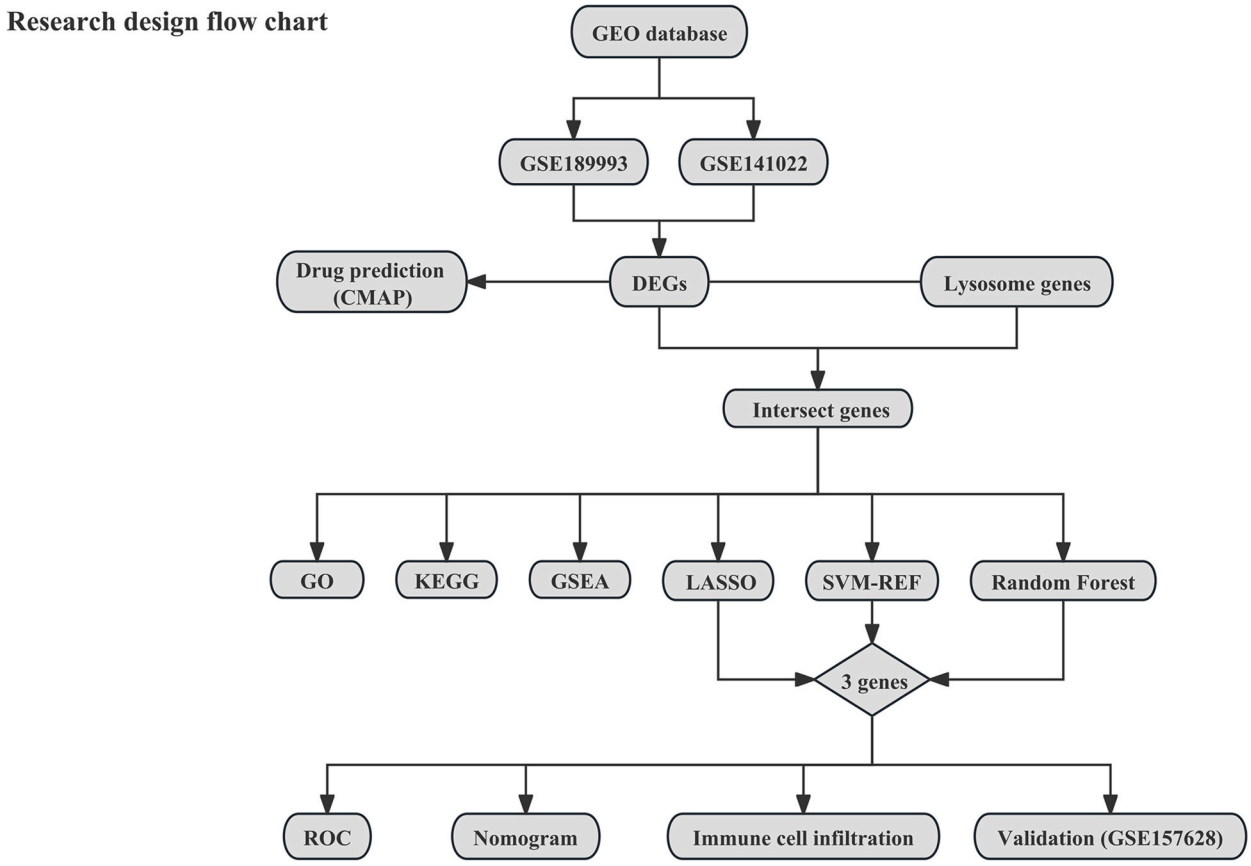


Fig. 1. The research flow chart.

amidst an extensive array of constituent tree models. The assessment of marker gene expression progression was subsequently scrutinized within the validation dataset GSE157628.

2.5. Assessment of the diagnostic value of key biomarkers in MMD

The research utilized receiver operating characteristic (ROC) analysis to assess the predictive performance of potential biomarkers derived from expression data in training datasets through the area under the receiver operating characteristic curve (AUC). An AUC exceeding 0.80 indicated precise classifier models.

2.6. Immune infiltration analysis

The level of immune cell infiltration was evaluated in specimens from both MMD and control cohorts in the training dataset using the CIBERSORT analytical approach, with a "PERM" value established at 1000 and a significance threshold set at $p < 0.05$. Subsequently, a heatmap depicting the distribution of 22 diverse immune cell types was generated employing the "pheatmap" software, in conjunction with the "vioplot" utility. To depict the interrelation among the 22 unique infiltrating immune cell populations, a correlation heatmap was constructed utilizing the "corrplot" package for visualization purposes.

Exploring the correlation between identified gene markers and infiltrating immune cells.

We investigated the association between infiltrating immune cell cohorts and gene biomarkers utilizing Spearman's rank correlation analyses executed within the R programming environment. Following this, correlation evaluations were performed utilizing the "ggplot2" package.

2.7. Drug sensitivity analysis

The Connectivity Map (CMAP) repository encompasses 6100 samples representing 1309 distinct small-molecule therapeutics, each characterized by the gene expression profile of a particular drug and its associated treatment regimen. In this study, we utilized the CMAP repository to predict potential compounds for the treatment of MMD. The scores reflected the correlation between the drug and the DEGs. A negative score signifies a gene expression profile that opposes the disease's expression pattern, suggesting therapeutic potential for the interference.

2.8. Clinical specimen collection and ethics statement

The current investigation adhered to the principles outlined in the Declaration of Helsinki and received approval from the Medical Ethics Committee of The Affiliated Hospital of Qingdao University (Approval Number: QYFY-WZLL-28629). All participants provided written informed consent. Patients diagnosed with MMD were identified based on the diagnostic criteria established by the World Health Organization. Samples of vascular tissue were collected from six patients and six control subjects who underwent surgical procedures at the neurosurgical emergency department of The Affiliated Hospital of Qingdao University (Supplementary table).

2.9. Western blotting

The lysosomal protein expression and transdifferentiation degree were assessed via Western blotting using proteins extracted from the respective experimental groups. Cell lysates were prepared by treating cells with RIPA lysis buffer (composed of RIPA, PMSF, and phosphoprotein inhibitors A and B in a ratio of 100:1:1:1) on ice followed by sonication for 60 s. Clarified supernatants were obtained by centrifugation at 13,000 rpm for 10 min at 4 °C. Protein concentrations were determined using the BCA assay, and subsequent denaturation was achieved by boiling for 10 min. Equal amounts (20 μg) of proteins were separated using SDS-PAGE on gels with varying concentrations (6 %, 10 %, and 12 %) and transferred onto Immobilon®-P PVDF membranes. Membranes were then blocked with NcmBlot blocking buffer and incubated overnight at 4 °C with primary antibodies anti-β-actin (81115-1-RR, 1:10000, Proteintech, Wuhan, China), anti-DENND3 (24414-1-AP, 1:1000, Proteintech, Wuhan, China), anti-NCSTN (14071-1-AP, 1:1000, Proteintech, Wuhan, China), anti-EPRD1 (27252-1-AP, 1:10000, Proteintech, Wuhan, China). Secondary antibodies were applied the next day and incubated at room temperature for 1 h. Protein imaging and development were performed using ECL.

2.10. Statistical analysis

Statistical analyses were performed utilizing R software. Comparison between MMD and control samples was carried out employing a student's t-test. Marker genes were assessed through ROC analysis. Unless specified otherwise, significance level was established at $p < 0.05$.

3. Results

3.1. Identification of DEGs in MMD

The flowchart of this study is shown in Fig. 1. The expression profiles extracted from three GEO datasets (GSE189993, GSE141022)

comprised 40 samples in total, comprising 25 MMD and 15 control specimens. Following preprocessing and the elimination of batch effects, we identified 240 DEGs, consisting of 91 genes exhibiting upregulation and 149 genes displaying downregulation in the MMD samples relative to the control specimens (Supplementary Table). In this study, the significant contrast of DEGs was illustrated using a heatmap (Fig. 2A) and a volcano plot (Fig. 2B).

3.2. Functional enrichment analysis of intersect genes of DEGs and lysosome genes

The initial procedure involves the intersection of DEGs with gene sets associated with lysosomes sourced from the Gene Ontology database (https://geneontology.org/), resulting in the identification of 12 DEGs related to lysosomes in MMD (Fig. 3A). We performed enrichment analyses on GO terms, KEGG pathways, and functional annotations to explore the potential biological roles of the commonly DEGs associated with lysosomes. Our analysis delved into the realms of biological processes (BP), molecular functionalities (MF), and cellular localization (CC), as delineated by the GO annotations. The findings unveiled that BP is predominantly linked with symbiotic symbiosis, lipid conveyance, protein endosomal localization, amyloid-beta metabolism, and phospholipid conveyance. Regarding CC enrichment analysis, it was observed that the overlapping genes notably contributed to the membrane of lysosomes, vacuoles, late endosomes, and granules of azurophilic. In the MF enrichment analysis, the DEGs associated with lysosomes mainly engaged in binding of amides, ceramides, and activities of lipid transporters (Fig. 3B, C, 3D). The KEGG pathway analysis revealed a significant enrichment of intersecting genes in pathways related to cholesterol metabolism, salmonella infection, and allograft rejection (Fig. 3E and F). We additionally performed Gene Set Enrichment Analysis (GSEA), revealing enrichment of pathways associated with colorectal cancer, endometrial cancer, oocyte meiosis, ubiquitin-mediated proteolysis, and vascular smooth muscle contraction in the control cohort (Fig. 4A). Conversely, the MMD cohort exhibited enrichment in pathways related to allograft rejection, antigen processing and presentation, autoimmune thyroid disease, graft-versus-host disease, and type I diabetes mellitus (Fig. 4B). The complete results of GO, KEGG and GSEA analyses can be found in Supplementary Table.

3.3. Identification of immune related diagnostic marker genes in MMD

Three distinct bioinformatics algorithms were employed for the identification of putative biomarkers associated with MMD. Through LASSO regression analysis, the pool of 12 DEGs linked to lysosomes was narrowed down to 10, which were deemed as diagnostic biomarkers for MMD (Fig. 4C and D). Furthermore, a subset consisting of 4 DEGs related to lysosomes was delineated utilizing SVM-RFE (Fig. 4E and F). Ultimately, the top 10 genes in terms of their significance were meticulously selected employing Random Forest (RF) methodology (Fig. 4G and H). Ultimately, we opted for the identification of three overlapping characteristic genes, namely EPDR1, DENND3, and NCSTN, through the implementation of the LASSO algorithm, SVM-RFE algorithm, and RF algorithm, as depicted in Fig. 5A. Subsequently, within the train cohorts GSE189993 and GSE141022, a thorough examination of the mRNA expression levels of these biomarkers was conducted to yield more precise and reproducible findings. Notably, EPDR1 exhibited a significant decrease in expression among MMD patients, whereas DENND3 and NCSTN demonstrated marked upregulation ($p < 0.05$, Fig. 5B, C, and 5D). Importantly, the expression of the three MMD biomarkers produced equivalent outcomes within the validation cohort GSE157628 (Supplementary Fig. 1).

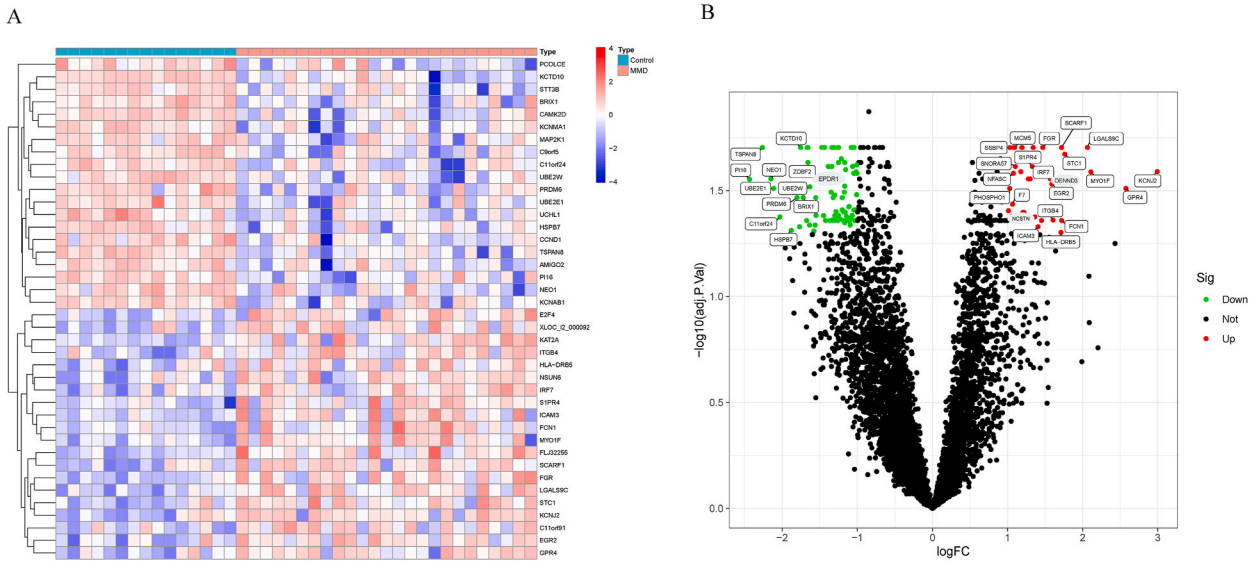


Fig. 2. Visualization of the DEGs. (A) Heatmap clustering of genes with markedly different expression in MMD compared control samples. | log2Foldchange|>1 and adjusted P-value < 0.05 were used to define statistically significant DEGs. MMD: moyamoya disease; DEGs, differentially expressed genes. (B) DEGs volcano map.

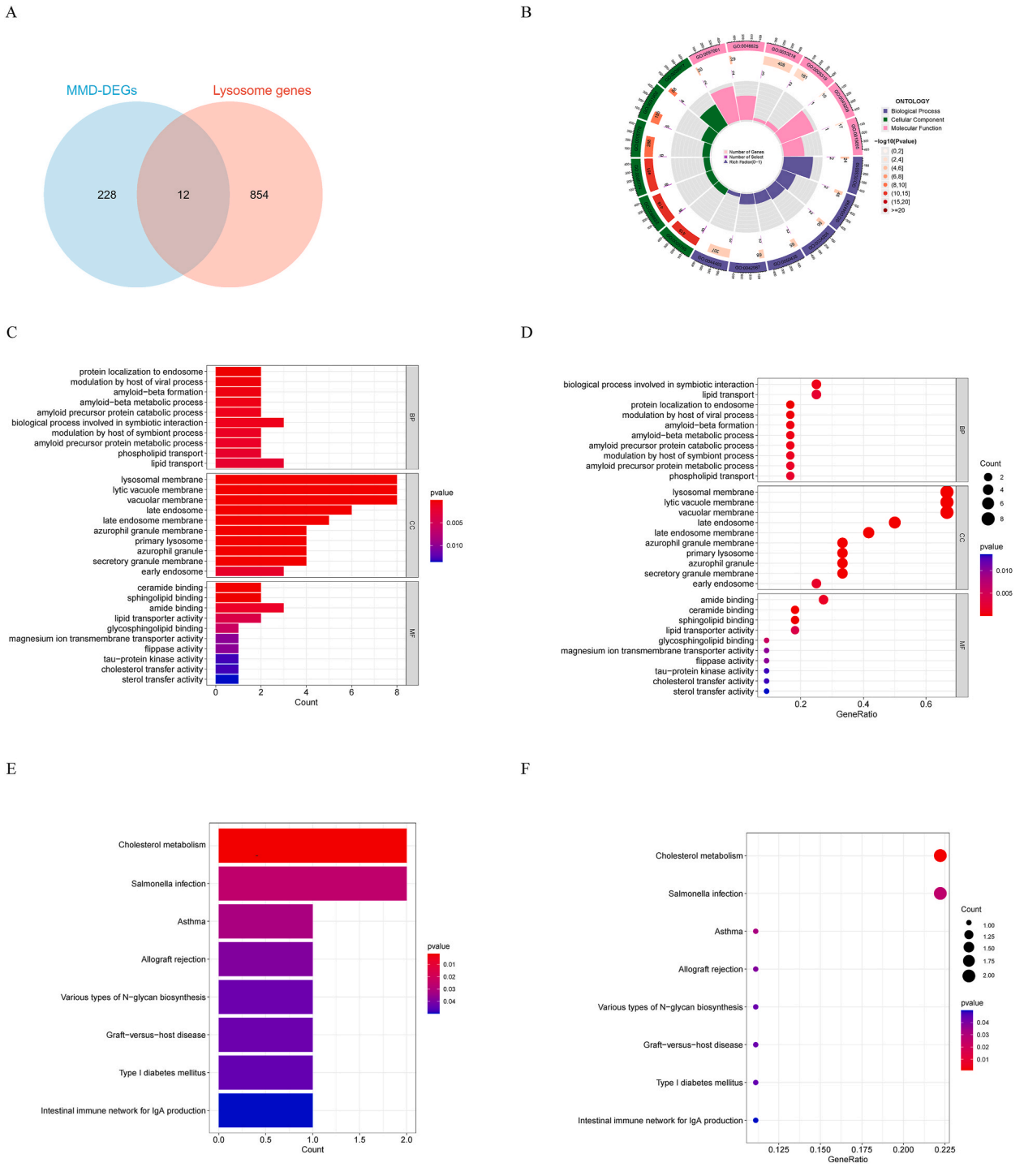


Fig. 3. Functional enrichment of DEGs. (A) Venn diagram of the intersection of DEGs and lysosome gene set. (B) Circle map and cluster of GO analysis. (C, D) Gene Ontology (GO) analyses for intersect genes. (E, F) Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of intersect genes.

3.4. Diagnostic effectiveness of the marker genes

The diagnostic framework was established utilizing the identified genetic markers via a logistic regression methodology. Additionally, the area under the receiver operating characteristic curve (AUC) was employed to precisely assess discriminatory

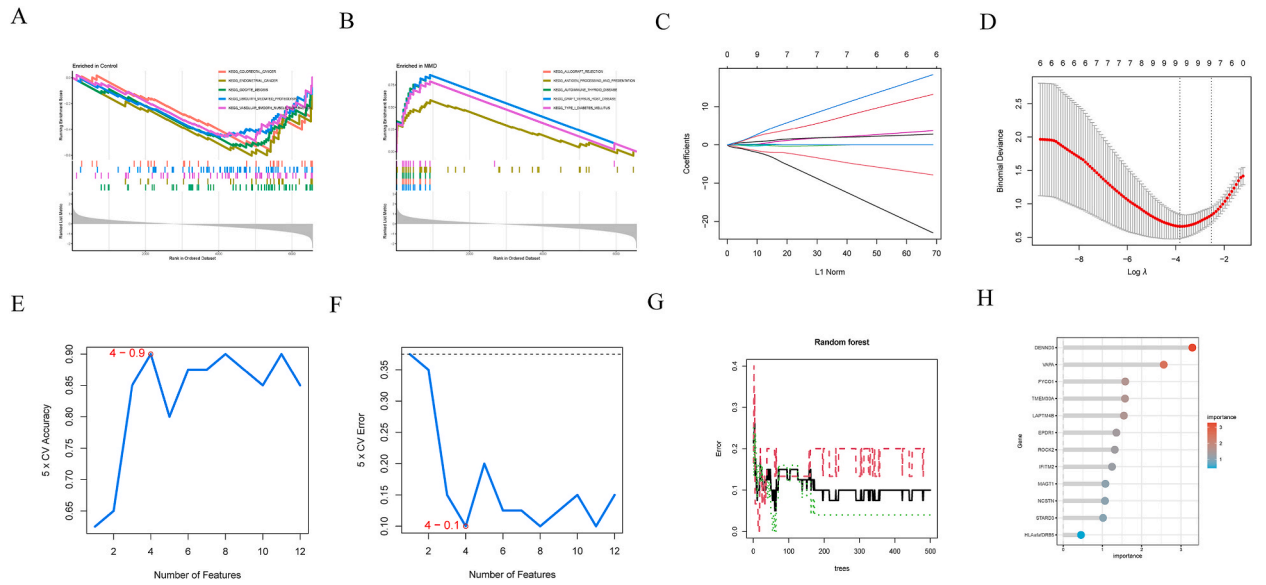


Fig. 4. Machine learning methods to screen for marker genes. (A, B) Gene set enrichment analyses (GSEA) in the MMD and control group. (C, D) Fine-tuning the least absolute shrinkage and selection operator (LASSO) model's feature selection. (E, F) A plot illustrating the process of selecting biomarkers using the support vector machine-recursive feature elimination (SVM-RFE) technique. (G) Random forest (RF) for the relationships between the number of trees and error rate. (H) Top 10 genes with importance > 1 were identified in the Random Forest algorithm.

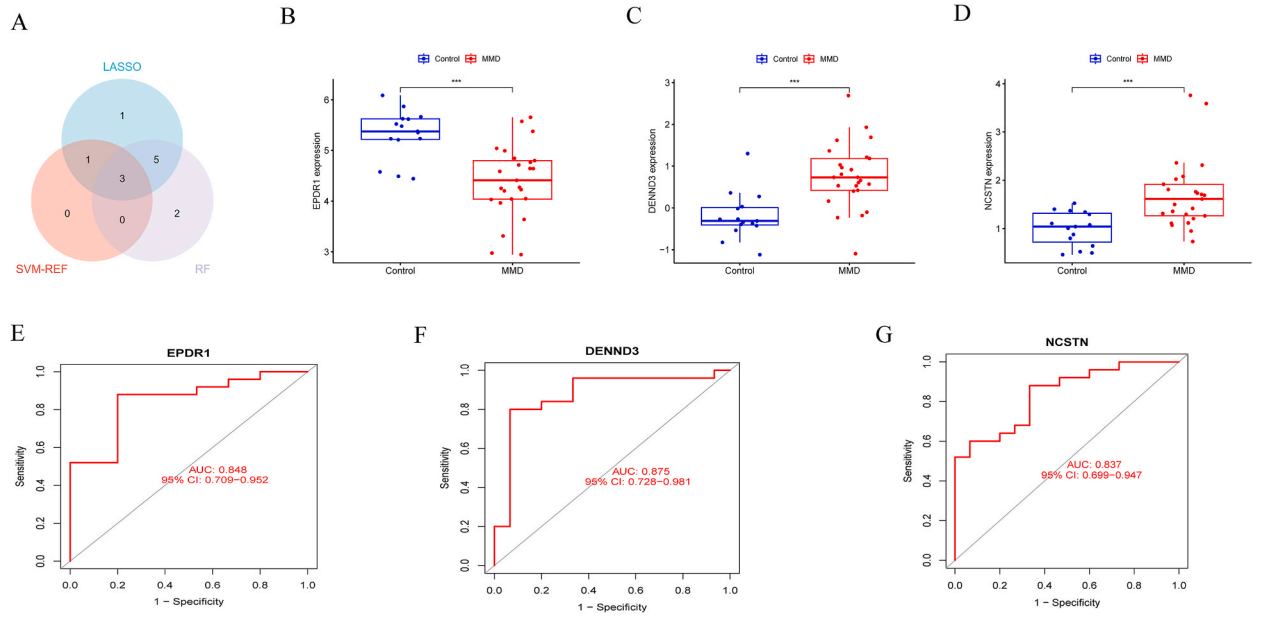


Fig. 5. Obtaining the hub genes of MMD and differential expression analysis. (A) Intersection LASSO, SVM-RFE, and RF was displayed in a Venn diagram. (B, C, D) EPDR1, DENND3 and NCSTN mRNA expression in MMD compared to control samples in the train group GSE189993 and GSE141022. (E) ROC curve construction in public data sets to evaluate the diagnostic accuracy of EPDR1 of MMD. (F) ROC curve construction in public data sets to evaluate the diagnostic accuracy of DENND3 of MMD. (G) ROC curve construction in public data sets to evaluate the diagnostic accuracy of NCSTN of MMD. EPDR1: Ependymin Related 1; DENND3: DENN Domain Containing 3; NCSTN, Nicastrin; ROC: receiver operating characteristic.

performance. As illustrated in Fig. 5E–G, distinctive biomarkers exhibited robust diagnostic efficacy in discriminating MMD from control specimens, yielding AUC values of 0.848 (95 % CI 0.709–0.952) for EPDR1, 0.875 (95 % CI 0.728–0.981) for DENND3, and 0.837 (95 % CI 0.699 to 0.947) for NCSTN. This finding enhances the diagnostic utility of these three unique genes as candidate biomarkers in the diagnosis of MMD. In constructing the nomogram model, we employed the "rms" package within the R environment,

integrating the expression patterns of the three identified biomarkers as pivotal factors (Fig. 6A). We evaluated the precision and dependability of our nomogram through the examination of calibration plots (Fig. 6B). The curves presented compelling evidence regarding the precision of the forecasts generated by our tool, demonstrating a robust correlation between the predictive and observed occurrence of MMD.

3.5. Immune cell infiltration

The CIBERSORT algorithm was employed to estimate the distribution of 22 immune cell types in both MMD and control samples. MMD patients exhibited elevated proportions of activated CD4 memory T cells, follicular helper T cells, and neutrophils ($p < 0.05$) compared to controls. Conversely, the prevalence of plasma cells, resting CD4 memory T cells, activated NK cells, and resting mast cells was significantly diminished in MMD patients ($p < 0.05$) (Fig. 7A–B). The correlation heatmap comprising 22 distinct immune cell types reveals noteworthy associations: T follicular helper cells and resting dendritic cells ($r = 0.61$), resting NK cells and M1 macrophages ($r = 0.53$), M2 macrophages and resting mast cells ($r = 0.55$), memory B cells and T cells gamma delta ($r = 0.65$), T cells gamma delta and resting CD4 memory T cells ($r = 0.51$) demonstrate prominent positive correlations. Conversely, negative correlations are observed between resting NK cells and activated NK cells ($r = -0.75$), resting dendritic cells and mast cells ($r = -0.59$), and follicular helper T cells and resting mast cells ($r = -0.49$), respectively (Fig. 7C).

3.6. Correlation analysis between the marker genes and immune cells

The correlation between feature biomarkers and 22 immune cell phenotypes was assessed via Spearman's rank correlation analysis. According to the correlation findings, EPDR1 exhibited a significant positive association with CD4 memory resting T cells and activated NK cells ($p < 0.05$), whereas it demonstrated a notable inverse correlation with resting dendritic cells ($p < 0.05$) (Fig. 7D). DENND3 exhibited a significant positive association with T follicular helper cells and neutrophil counts ($p < 0.05$), whereas it demonstrated an inverse relationship with plasma cell counts and resting CD4 memory T cell counts ($p < 0.05$) (Fig. 7E). Conversely, NCSTN exhibited a significant negative correlation with activated NK cell counts ($p < 0.05$) (Fig. 7F).

3.7. Validation using western blotting

To validate this outcome, tissue specimens were initially procured from individuals diagnosed with moyamoya disease, followed by a Western blot analysis. Our findings revealed a notable upregulation in the levels of DENND3 and NCSTN proteins among moyamoya patients, accompanied by a considerable downregulation in ERPD1 protein expression in the same cohort (Fig. 8). Then, we compared the expression of the three disease-related genes in MMD patients grouped by different genders and ages. The results revealed that no significant differences in the three MMD-related diagnostic genes between age and gender (Supplementary Fig. 2).

3.8. Drug prediction

We interrogated the CMAP repository employing DEGs derived from the refined MMD gene expression dataset. Compounds eliciting expression alterations contrary to those identified in our MMD profile are delineated in the Supplementary Table. Among the compounds exhibiting the lowest connectivity scores, those acting as inhibitors of MAP kinase and phosphodiesterase were found to be predominant. From this compilation, a selection of ten compounds was made in accordance with the predetermined criteria (Supplementary Fig. 3). Finally, a schematic diagram was used to illustrate the research (Fig. 9).

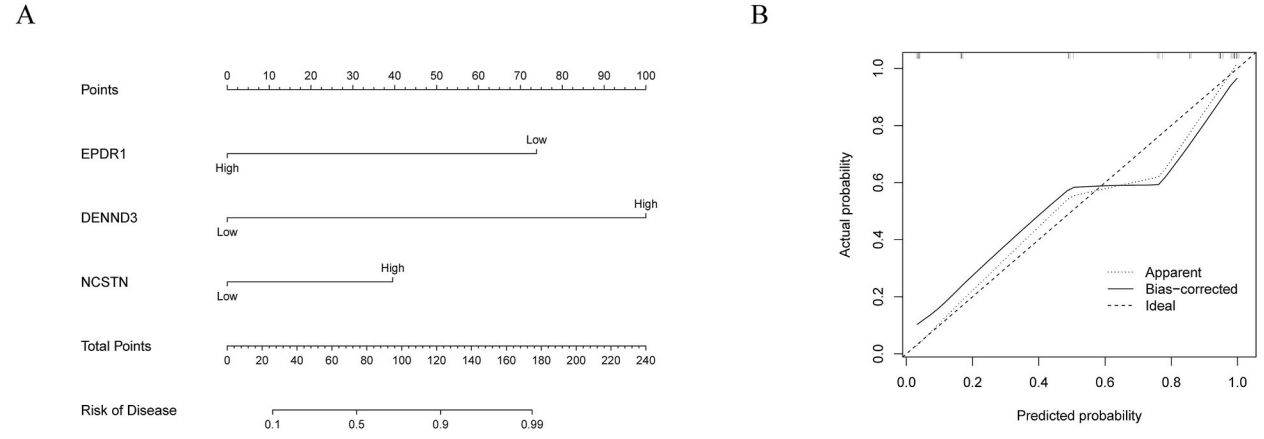


Fig. 6. Validation of the diagnostic efficacy of 3 hub genes. (A) Nomogram showing the predicted risk for MMD based on 3 hub genes. (B) Calibration curve showing predicted performance of the nomogram.

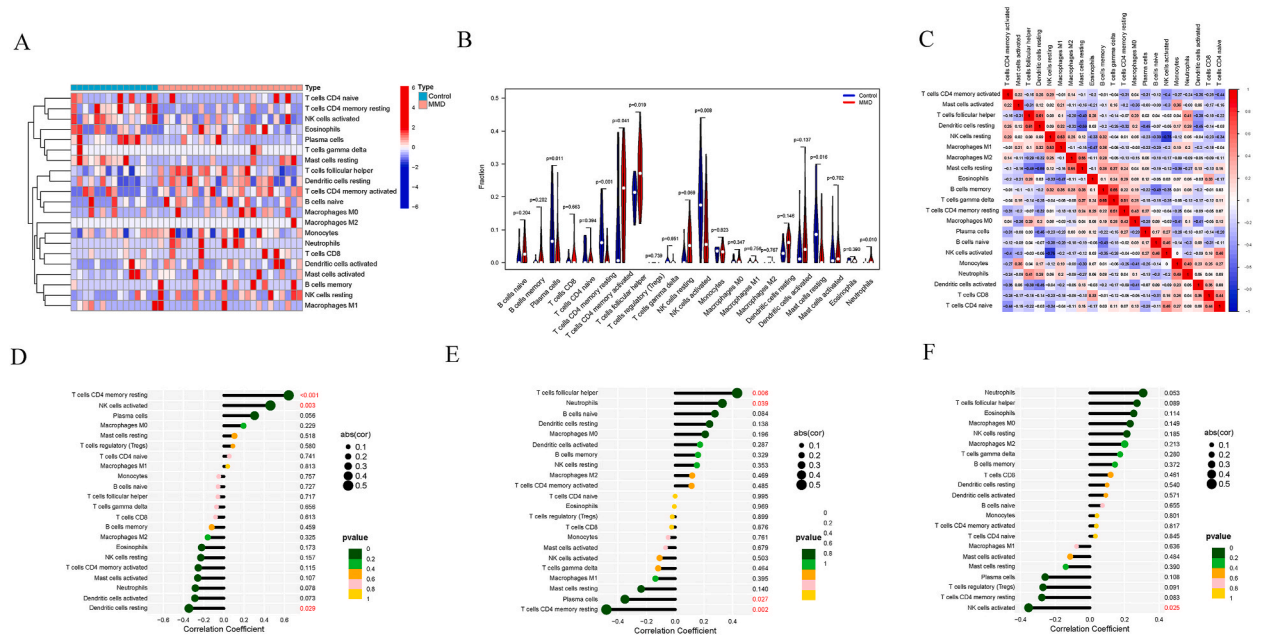


Fig. 7. The composition of immune cells was analyzed and displayed. (A) Heat map of immune cell infiltration in GSE189993 and GSE141022 using cibersort. (B) Violin diagram illustrating the proportion of different kinds of immune cells in MMD and control samples. (C) Heat map showing the correlation between different kinds of immune cells in MMD and control samples. The size of the colored squares indicates the connection' strength; red indicates a positive correlation, while blue indicates a negative correlation. The stronger the connection, the redder the hue. (P-values < 0.05 were considered as statistically significant). (D) Correlation between EPDR1 and infiltrating immune cells. (E) Correlation between DENND3 and infiltrating immune cells. (F) Correlation between NCSTN and infiltrating immune cells. The size of the dots indicates the degree to which genes and immune cells are correlated. Correlation strength is proportional to the size of the dots. The color of the dots indicates the P-value; a yellower hue indicates a lower P-value, while a greener color indicates a higher P-value. P-value < 0.05 was considered statistically significant. EPDR1: Ependymin Related 1; DENND3: DENN Domain Containing 3; NCSTN, Nicastrin.

4. Discussion

Detecting biomarkers associated with MMD is crucial for early diagnosis. Yet, research exploring MMD-related factors is limited. Lysosomes play a pivotal role in numerous vital cellular processes and signaling pathways in physiological contexts [18]. Additionally, research has explored non-lysosomal pathways associated with lysosomes in conditions including atherosclerosis, neurodegenerative disorders, pancreatitis, autoimmune conditions, disorders of lysosomal storage, and malignancies [19]. Thus, this groundbreaking investigation suggests employing an integrated bioinformatics approach to pinpoint lysosome-associated biomarkers for diagnostic purposes and evaluate their correlation with immune cell infiltration in MMD individuals.

A recent study reported that RNF213 plays a role in the pathology of MMD by affecting antigen presentation and presentation pathway [20]. In addition, a study has shown HLA-DRB1*04:10 is associated with thyroid disease in MMD patients [21]. And activation of the autoimmune system is associated with MMD and a variety of diseases such as down syndrome, intracranial hemorrhage and thyroid diseases [1,22,23]. These findings indicate that MMD exhibits significant alterations in immune system dynamics, inflammatory responses, and the infiltration of immune cell populations. The Lasso regression method offers a distinct advantage by mitigating the shortfall in localized optimal estimation observed in both least squares and stepwise regression methodologies. Moreover, it excels in the realm of feature selection and adeptly mitigates the challenge posed by multicollinearity among variables. However, a limitation arises when confronted with highly correlated sets of features, as Lasso regression tends to favor one feature over others, leading to potential instability in outcomes [24,25]. To mitigate experimental bias, we incorporated machine learning methodologies, namely "SVM-RFE" and "RF". Employing LASSO, SVM-RFE, and RF algorithms, we discerned EPDR1, DENND3, and NCSTN as robust diagnostic biomarkers with high sensitivity and specificity for MMD. Conventional diagnostic test assessment techniques, such as logistic regression, entail dichotomizing test outcomes for statistical scrutiny. Conversely, our assessment approach accommodates intermediate states reflective of real-world scenarios, obviating the need for such categorization. Thus, ROC curves offer a broader evaluative spectrum [26,27]. The diagnostic efficiency of the three MMD marker genes was further demonstrated through the development of a nomogram. The protein encoded by EPDR1 is a type II transmembrane protein similar to two families of cell adhesion molecules, procalcitonin and ventricular tubulin. This protein may play a role in calcium-dependent cell adhesion [28]. This protein is glycosylated and the immediate homologous mouse protein is localized to the lysosome. Variable splicing results in multiple transcript variants. A related pseudogene has been identified on chromosome 8 [29]. The EPDR1 gene is a regulator of blood glucose and lipids and has been implicated in the pathogenesis of a number of tumors including bladder cancer, ovarian cancer and acute myeloid leukaemia [30–32]. However, the association between EPDR1 and MMD is unclear, and our study reveals that EPDR1 is

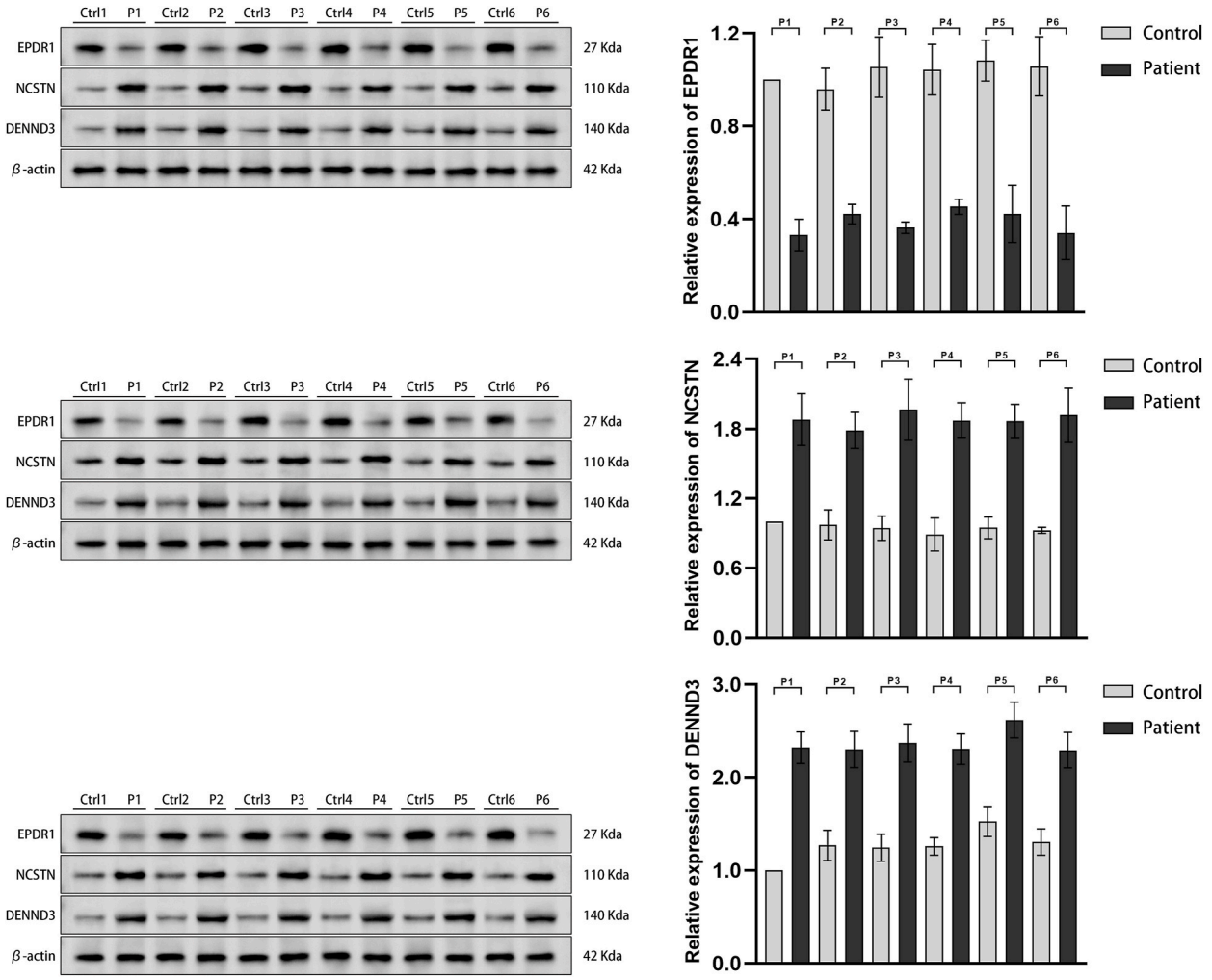


Fig. 8. Validation by WB: The expression of EPDR1, DENND3 and NCSTN proteins by WB. Triple experiments by WB (Western Blot).

lowly expressed in patients with MMD, suggesting that EPDR1 could play a critical protective role in the pathogenesis of MMD. The main functions of the DENND3 gene include enabling guanidino nucleotide exchange factor activity, participation in cellular proteolytic metabolic processes, endosome-to-lysosome transport and regulation of Rab protein signal transduction [33]. DENND3 is associated with the development of lung cancer, prostate cancer and hereditary diseases [33–35]. Whereas it is well known that MMD has a certain tendency of familial aggregation [36]. Our results show that DENND3 is highly expressed in MMD patients, which suggests that DENND3 may be an important biomarker of heritability in MMD. The NCSTN gene encodes a type I transmembrane glycoprotein that is a component of the multimeric γ-secretase complex. The encoded protein cleaves integral membrane proteins, including Notch receptors and β-amyloid precursor protein, and may be a stabilizing cofactor required for assembly of the γ-secretase complex [37]. Cleavage of the β-amyloid precursor protein produces amyloid β peptide, a major component of neuroinflammatory plaques and a hallmark lesion in the brains of Alzheimer's disease patients [38]. NCSTN has been reported to be associated with hidradenitis suppurative and a variety of cancers [37,39]. Interestingly, studies have shown a link between NCSTN and the heritability of several diseases [40,41]. Our findings demonstrate elevated expression of NCSTN in MMD patients, which may potentially be linked to the heritability of MMD. Significantly, our study showed that the AUC values under the ROC curves for three genes, namely EPDR1, DENND3 and NCSTN, were all above 0.8. The findings presented herein underscore the reliability of our results and underscore the prospective utility of EPDR1, DENND3, and NCSTN as robust biomarkers for MMD. Nevertheless, the exact contributions of EPDR1, DENND3, and NCSTN to the etiology of MMD remain elusive and warrant further elucidation.

Moreover, mounting evidence indicates a close association between MMD pathology and immune alterations within the brain, which commence in the early stages and endure throughout the progression of the disease [42]. Our findings indicated elevated proportions of activated memory $CD4^+$ T cells, follicular helper T cells, and neutrophils, alongside diminished levels of resting memory $CD4^+$ T cells, activated NK cells, and resting mast cells, as well as plasma cells, indicating that they might be involved in the development of MMD. During the onset of inflammation within the organism, T lymphocytes and neutrophils undergo metabolic

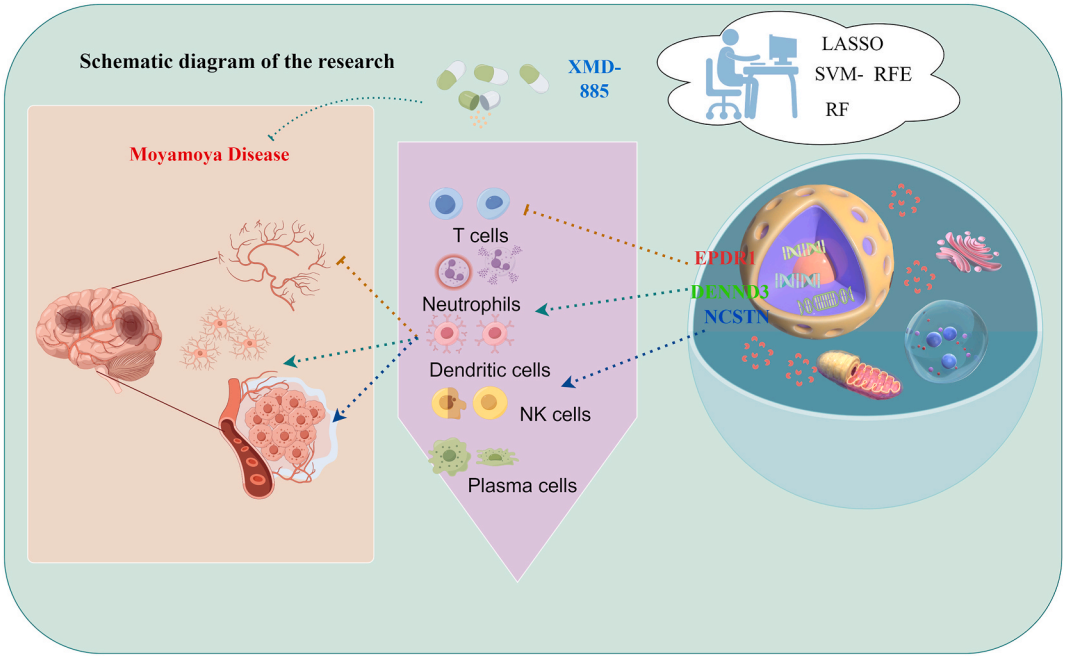


Fig. 9. Schematic diagram of the research.

adjustments in response to alterations in the surrounding microenvironment [43]. Mast cells serve as progenitors to hematopoietic stem cells and play a substantial role in the initiation and expression of allergic responses, thereby exerting a pivotal function in allergic reactions [44]. Natural killer (NK) cells, as cytotoxic lymphocytes, exhibit rapid cytotoxicity, secrete cytokines promptly, undergo clonal expansion, and contribute to immunomodulation across various disease states [45]. Plasma cells represent a class of specialized immune cells that undergo differentiation from B lymphocytes following activation by lymphokines [46]. Our study indicates that specific cell populations may have a role in treating MMD.

Moreover, we explored the correlation between infiltrative immune cell populations and diagnostic biomarkers. EPDR1 displayed a significant and strong correlation with T cells CD4 memory resting, NK cells activated and dendritic cells resting. DENND3 was significantly associated with T cells follicular helper, neutrophils, plasma cells and T cells CD4 memory resting. NCSTN displayed significant correlations to infiltrations of immune cells with NK cells activated. Investigating the intrinsic molecular mechanisms and functional relevance in MMD is imperative, given the dearth of knowledge regarding these intricate genetic and immunological interplays. Ultimately, medicinal drugs that act as MAP kinase inhibitors and phosphodiesterase inhibitors like XMD-885 and enoximone, are the most likely therapy for MMD. Besides, other mechanisms such as inhibitors of retinoid receptor, IGF-1, androgen receptor, estrogen receptor, histamine receptor, EGFR and ACAT also show certain effectiveness for MMD. The schematic illustration delineates the diagnostic biomarkers, potential pathogenic pathways, and therapeutic modalities for MMD, thereby furnishing a valuable resource for managing this condition (Fig. 9). We hope this will help to some extent in the treatment of MMD.

5. Limitations and future directions

The present investigation encountered constraints due to the restricted scope of clinical case samples and the limited depth of inquiry into the specific association mechanisms underlying MMD. Subsequent endeavors aim to amass a more extensive pool of MMD samples and probe into the underlying mechanisms at a heightened granularity.

6. Conclusion

In summary, our investigation has revealed EPDR1, DENND3, and NCSTN as diagnostic indicators for MMD, potentially pivotal in disease onset and advancement. Furthermore, our findings suggest involvement of T lymphocytes, neutrophils, dendritic cells, natural killer cells, and plasma cells in MMD pathogenesis, offering avenues for immunotherapeutic intervention. Ultimately, we envisage our drug prediction insights and associated mechanisms as promising contributions to pharmaceutical advancement and clinical management of MMD.

Data availability statement

The datasets generated in this study can be found in online repository. The names of the repository/repositories and accession

number(s) can be found in the article/supplementary material.

Funding

The authors have no funding to disclose.

Ethics approval and consent to participate

There were no ethical conflicts involved in this study.

CRediT authorship contribution statement

Wenyang Li: Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation. **Xiang Zhao:** Writing – original draft, Visualization, Validation, Formal analysis, Data curation. **Jinxing Fu:** Visualization, Validation, Formal analysis. **Lei Cheng:** Writing – review & editing, Visualization, Validation, Supervision, Conceptualization.

Declaration of competing interest

The authors have declared that no competing interest exists.

Acknowledgments

We acknowledge the Gene Expression Omnibus (GEO) database for providing data of MMD available.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e34432.

References

[1] G. Zhang, E. Liu, X. Tan, C. Liu, S. Yang, Research progress on moyamoya disease combined with thyroid diseases, Front. Endocrinol. 14 (2023) 1233567.
[2] T. Brunet, B. Zott, V. Lieftuchter, D. Lenz, A. Schmidt, P. Peters, R. Kopajtich, M. Zaddach, H. Zimmermann, I. Huning, D. Ballhausen, C. Staufner, A. Bianzano, et al., De novo variants in RNF213 are associated with a clinical spectrum ranging from Leigh syndrome to early-onset stroke, Genet. Med. (2023) 101013.
[3] R. Kleinloog, L. Regli, G.J. Rinkel, C.J. Klijn, Regional differences in incidence and patient characteristics of moyamoya disease: a systematic review, J. Neurol. Neurosurg. Psychiatry 83 (2012) 531–536.
[4] L. Cao, Y. Ai, Y. Dong, D. Li, H. Wang, K. Sun, C. Wang, M. Zhang, D. Yan, H. Li, G. Liang, B. Yang, Bioinformatics analysis reveals the landscape of immune cell infiltration and novel immune-related biomarkers in moyamoya disease, Front. Genet. 14 (2023) 1101612.
[5] H. Ni, Y. Wu, C. Zhou, X. Li, S. Zhou, W. Lan, Z. Zhang, Y. Huang, H. Wang, J. Lin, Application of intraarterial superselective indocyanine green angiography in bypass surgery for adult moyamoya disease, Front. Neurol. 14 (2023) 1241760.
[6] S. He, J. Zhang, Z. Liu, Y. Wang, X. Hao, X. Wang, Z. Zhou, X. Ye, Y. Zhao, Y. Zhao, R. Wang, Upregulated cytoskeletal proteins promote pathological angiogenesis in moyamoya disease, Stroke (2023).
[7] H. Wu, J. Xu, J. Sun, J. Duan, J. Xiao, Q. Ren, P. Zhou, J. Yan, Y. Li, X. Xiong, E. Zeng, APOE as potential biomarkers of moyamoya disease, Front. Neurol. 14 (2023) 1156894.
[8] S. Li, Y. Han, Q. Zhang, D. Tang, J. Li, L. Weng, Comprehensive molecular analyses of an autoimmune-related gene predictive model and immune infiltrations using machine learning methods in moyamoya disease, Front. Mol. Biosci. 9 (2022) 991425.
[9] C. Asselman, D. Hemelsoet, D. Eggermont, B. Dermaut, F. Impens, Moyamoya disease emerging as an immune-related angiopathy, Trends Mol. Med. 28 (2022) 939–950.
[10] M. Ihara, Y. Yamamoto, Y. Hattori, W. Liu, H. Kobayashi, H. Ishiyama, T. Yoshimoto, S. Miyawaki, T. Clausen, O.Y. Bang, G.K. Steinberg, E. Tournier-Lasserve, A. Koizumi, Moyamoya disease: diagnosis and interventions, Lancet Neurol. 21 (2022) 747–758.
[11] J. Knupp, M.L. Pletan, P. Arvan, B. Tsai, Autophagy of the ER: the secretome finds the lysosome, FEBS J. (2023).
[12] M. Zhang, C. Mao, Y. Dai, X. Xu, X. Wang, Qixian granule inhibits ferroptosis in vascular endothelial cells by modulating TRPML1 in the lysosome to prevent postmenopausal atherosclerosis, J. Ethnopharmacol. 328 (2024) 118076.
[13] Y. Gao, L. Zhang, F. Zhang, R. Liu, L. Liu, X. Li, X. Zhu, Y. Liang, Traditional Chinese medicine and its active substances reduce vascular injury in diabetes via regulating autophagic activity, Front. Pharmacol. 15 (2024) 1355246.
[14] N. Lund, H. Wieboldt, L. Fischer, N. Muschol, F. Braun, T. Huber, D. Sorriento, G. Iaccarino, K. Mullerleile, E. Tahir, G. Adam, P. Kirchhof, L. Fabritz, et al., Overexpression of VEGFalpha as a biomarker of endothelial dysfunction in aortic tissue of alpha-GAL-Tg/KO mice and its upregulation in the serum of patients with Fabry's disease, Front Cardiovasc Med 11 (2024) 1355033.
[15] H. Fu, J. Li, C. Zhang, P. Du, G. Gao, Q. Ge, X. Guan, D. Cui, Abeta-aggregation-generated blue autofluorescence illuminates senile plaques as well as complex blood and vascular pathologies in Alzheimer's disease, Neurosci. Bull. (2024).
[16] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, Biostatistics 8 (2007) 118–127.
[17] 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.
[18] C. Yang, X. Wang, Lysosome biogenesis: regulation and functions, J. Cell Biol. (2021) 220.
[19] Z. Zhang, P. Yue, T. Lu, Y. Wang, Y. Wei, X. Wei, Role of lysosomes in physiological activities, diseases, and therapy, J. Hematol. Oncol. 14 (2021) 79.
[20] R. Tashiro, K. Niizuma, J. Kasamatsu, Y. Okuyama, S. Rashad, A. Kikuchi, M. Fujimura, S. Kure, N. Ishii, T. Tominaga, Dysregulation of Rnf 213 gene contributes to T cell response via antigen uptake, processing, and presentation, J. Cell. Physiol. 236 (2021) 7554–7564.

[21] R. Tashiro, K. Niizuma, S.S. Khor, K. Tokunaga, M. Fujimura, H. Sakata, H. Endo, H. Inoko, K. Ogasawara, T. Tominaga, Identification of HLA-DRB1*04:10 allele as risk allele for Japanese moyamoya disease and its association with autoimmune thyroid disease: a case-control study, PLoS One 14 (2019) e0220858.
[22] J.D. Santoro, S. Lee, A.C. Wang, E. Ho, D. Nagesh, M. Khoshnood, R. Tanna, R.A. Durazo-Arvizu, M.A. Manning, B.G. Skotko, G.K. Steinberg, M.S. Rafii, Increased autoimmunity in individuals with down syndrome and moyamoya disease, Front. Neurol. 12 (2021) 724969.
[23] Y.H. Lin, H. Huang, W.Z. Hwang, Moyamoya disease with Sjogren disease and autoimmune thyroiditis presenting with left intracranial hemorrhage after messenger RNA-1273 vaccination: a case report, Medicine (Baltim.) 101 (2022) e28756.
[24] X. Niu, M. Xu, J. Zhu, S. Zhang, Y. Yang, Identification of the immune-associated characteristics and predictive biomarkers of keratoconus based on single-cell RNA-sequencing and bulk RNA-sequencing, Front. Immunol. 14 (2023) 1220646.
[25] H. Wang, Y. Chen, J. Qiu, J. Xie, W. Lu, J. Ma, M. Jia, Machine learning based on SPECT/CT to differentiate bone metastasis and benign bone lesions in lung malignancy patients, Med. Phys. (2023).
[26] D. Sharma, P. Rawat, V. Greiff, V. Janakiraman, M.M. Gromiha, Predicting the immune escape of SARS-CoV-2 neutralizing antibodies upon mutation, Biochim. Biophys. Acta, Mol. Basis Dis. (2023) 166959.
[27] G. Belge, C. Dumlupinar, T. Nestler, M. Klemke, P. Torzsok, E. Trenti, R. Pichler, W. Loidl, Y. Che, A. Hiester, C. Matthies, M. Pichler, P. Paffenholz, et al., Detection of recurrence through microRNA-371a-3p serum levels in a follow-up of stage I testicular germ cell tumors in the DRKS-00019223 study, Clin. Cancer Res. (2023).
[28] R. Agren, S. Patil, X. Zhou, Sahlholm K. FinnGen, S. Paabo, H. Zeberg, Major genetic risk factors for dupuytren's disease are inherited from neandertals, Mol. Biol. Evol. 40 (2023).
[29] K. Takaya, T. Asou, K. Kishi, Aging fibroblasts adversely affect extracellular matrix formation via the senescent humoral factor ependymin-related protein 1, Cells 11 (2022).
[30] L.R. Cataldo, Q. Gao, L. Argemi-Muntadas, O. Hodek, E. Cowan, S. Hladkou, S. Gheibi, P. Spegel, R.B. Prasad, L. Eliasson, C. Scheele, M. Fex, H. Mulder, et al., The human batokine EPDR1 regulates beta-cell metabolism and function, Mol Metab 66 (2022) 101629.
[31] Y. Yang, H. Xu, H. Zhu, D. Yuan, H. Zhang, Z. Liu, F. Zhao, G. Liang, EPDR1 levels and tumor budding predict and affect the prognosis of bladder carcinoma, Front. Oncol. 12 (2022) 986006.
[32] Z. Zhao, Z. Wang, P. Wang, S. Liu, Y. Li, X. Yang, EPDR1, which is negatively regulated by miR-429, suppresses epithelial ovarian cancer progression via PI3K/AKT signaling pathway, Front. Oncol. 11 (2021) 751567.
[33] Y. Li, A. Xu, Q. Ouyang, W. Zhang, C. Zhang, Z. Chen, D. Zhou, B. Zhang, W. Duan, X. Zhao, X. Wang, H. You, X. Ou, et al., DENND3 p.L708V activating variant is involved in the pathogenesis of hereditary hemochromatosis via the RAB12/TFR2 signaling pathway, Hepatol Int 17 (2023) 648–661.
[34] J.Y. Low, M. Ko, B. Hanratty, R.A. Patel, A. Bhamidipati, C.M. Heaphy, E. Sayar, J.K. Lee, S. Li, A.M. De Marzo, W.G. Nelson, A. Gupta, S. Yegnasubramanian, et al., Genomic characterization of prostatic basal cell carcinoma, Am. J. Pathol. 193 (2023) 4–10.
[35] N. Ben-Yosef, M. Frampton, E.R. Schiff, S. Daher, F. Abu Baker, R. Safadi, E. Israeli, A.W. Segal, A.P. Levine, Genetic analysis of four consanguineous multiplex families with inflammatory bowel disease, Gastroenterol Rep (Oxf). 9 (2021) 521–532.
[36] N.R. Gonzalez, S. Amin-Hanjani, O.Y. Bang, C. Coffey, R. Du, J. Fierstra, J.F. Fraser, S. Kuroda, G.E. Tietjen, S. Yaghi, American Heart Association Stroke C, C. Council on, N. Stroke, et al., Adult moyamoya disease and syndrome: current perspectives and future directions: a scientific statement from the American heart association/American stroke association, Stroke 54 (2023) e465–e479.
[37] A. de Oliveira, R.C. de Siqueira, C. Nait-Meddour, P.M. Tricarico, R. Moura, A. Agrelli, A.P. d'Adamo, S. Jamain, S. Crovella, M. de Fatima Medeiros Brito, M. Boniotto, L.A.C. Brandao, A loss-of-function NCSTN mutation associated with familial Dowling Degos disease and hidradenitis suppurativa, Exp. Dermatol. 32 (2023) 1935–1945.
[38] D. Mintoff, N.P. Pace, I. Borg, NCSTN in-frame deletion in Maltese patients with hidradenitis suppurativa, JAMA Dermatol 159 (2023) 939–944.
[39] H. Li, T. Lan, L. Xu, H. Liu, J. Wang, J. Li, X. Chen, J. Huang, X. Li, K. Yuan, Y. Zeng, H. Wu, Correction: NCSTN promotes hepatocellular carcinoma cell growth and metastasis via beta-catenin activation in a Notch1/AKT dependent manner, J. Exp. Clin. Cancer Res. 42 (2023) 47.
[40] Y. He, W. Wang, X. Ma, Z. Duan, B. Wang, M. Li, H. Xu, Discovery and potential functional characterization of long noncoding RNAs associated with familial acne inversa with NCSTN mutation, Dermatology (2023).
[41] C. Liu, X. Liu, R. Wang, L. Chen, H. Zhao, Y. Zhou, A novel NCSTN mutation in a three-generation Chinese family with hidradenitis suppurative, J Healthc Eng 2022 (2022) 1540774.
[42] Y. Zhang, Y. Yuan, L. Jiang, Y. Liu, L. Zhang, The emerging role of E3 ubiquitin ligase RNF213 as an antimicrobial host determinant, Front. Cell. Infect. Microbiol. 13 (2023) 1205355.
[43] M.E. Kim, J.S. Lee, Immune diseases associated with aging: molecular mechanisms and treatment strategies, Int. J. Mol. Sci. 24 (2023).
[44] T.C. Theoharides, A. Twahir, D. Kempuraj, Mast cells in the autonomic nervous system and potential role in disorders with dysautonomia and neuroinflammation, Ann. Allergy Asthma Immunol. (2023).
[45] S. Ma, Y. Tian, J. Peng, C. Chen, X. Peng, F. Zhao, Z. Li, M. Li, F. Zhao, X. Sheng, R. Zong, Y. Li, J. Zhang, et al., Identification of a small-molecule Tim-3 inhibitor to potentiate T cell-mediated antitumor immunotherapy in preclinical mouse models, Sci. Transl. Med. 15 (2023) eadg6752.
[46] Y. Yang, Y. Wang, S. Wei, X. Wang, J. Zhang, Effects and mechanisms of non-thermal plasma-mediated ROS and its applications in animal husbandry and biomedicine, Int. J. Mol. Sci. 24 (2023).