A ten-gene retinal pigment epithelium (RPE)/choroid complex diagnosis signature for age-related macular degeneration

He-Yan Li
Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University
Li Dong
Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University
Wen-Da Zhou
Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University
Hao-Tian Wu
Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University
Yi-Fan Li
Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University
Rui-Heng Zhang
Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University
Wen Bin Wei (weiwenbintr@163.com)
Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University

Research Article

Keywords: Age-related macular degeneration, gene expression, retinal pigment epithelium /choroid complex, machine learning model

Posted Date: June 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1668543/v2

Abstract

Purpose

Age-related macular degeneration (AMD) is a multifactorial disease in the elderly with a prominent genetic basis. This study aimed to apply machine learning method to develop a novel diagnostic model for AMD based on gene biomarkers in RPE/choroid complex, which may be potential therapeutic targets.

Methods

We collected RPE/choroid tissue gene expression profiles of AMD and normal patients from the Gene Expression Omnibus (GEO) database as training and validation cohorts. After differential expression analysis and the selection of gene biomarkers by random forest algorithms, selected genes were applied to the least absolute shrinkage and selection operator (LASSO) logistic regression to construct a diagnostic model in the training cohort. The diagnostic ability of the model was further tested in the validation cohort. Gene set enrichment analysis (GSEA) and immune cell assessment were also conducted for further analyses.

Results

A noval diagnostic model based on ten genes (BMPR2, CNOT3, CRLF1, FXYD6, HRASLS5, KRTDAP, NUDT16L1, PI16, PLAGL1, SART1) was constructed in the training cohort. The AUC in the training cohort reached 0.908 (95% CI: 0.823–0.975), while it remained 0.809 (95% CI: 0.522–0.889) in the validation cohort. According to the GSEA analysis, glutathione metabolism and phototransduction pathway are the two shared enriched pathways in the training and validation cohorts. Functional enrichment analysis and immune cell evaluation demonstrated that AMD was significantly correlated with both adaptive and innate immune cells, and the levels of neutrophil in the high-risk group were significantly higher than that in the low-risk group in both training and validation datasets

Conclusion

We identified and validated a novel ten-gene-based diagnostic model with high accuracy for AMD. The current study provided a promising tool to be used as a precise and cost-effective non-invasive test in clinical practice.

Introduction

Age-related macular degeneration (AMD) is characterized by the loss of photoreceptors and retinal pigment epithelium (RPE) in the macular, which is a cone-rich and primate-specific region in the eye that is responsible for high-acuity of vision. Risk factors for AMD onset and progression include not only

environment and behaviours, but also genes. Genes influence several biological pathways including complement and immune processes, HDL cholesterol, and mechanisms involving collagen, extra-cellular matix, and angiogenesis pathways [1–3].

AMD is traditionally diagnosed on the basis of clinical examination or assessment of colour fundus photographs. It is classified as early-stage (medium-sized drusen and retinal pigmentary changes) to late-stage (neovascular and atrophic). There are no proven therapies for atrophic disease, but anti-vascular endothelial growth factor (VEGF) agents are useful in treating neovascular AMD. However, not all patients respond to these medications [4]. Dry AMD can be difficult to be detected in the early stage due to the absence of clear disease symptoms, and when it develops to geographic atrophy, it often results in irreversible vision damage. Therefor, early diagnosis of AMD is important and biomarker genes may have the potential for early diagnosis.

Machine learning techniques focus on how computers learn from data [5]. In this study, we applied random forest (RF) model to choose robust genes biomarkers for AMD diagnosis on RPE/choroid complex. We also used the least absolute shrinkage and selection operator (LASSO) logistic regression to construct a streamlined diagnostic model that is capable of extensively implementing in clinical practice. Based on the machine learning model, we aimed to create a novel REP/choroid complex diagnostic model of AMD with optimal selected genes. This research may also promote new gene targeted therapies for AMD patients.

Methods

Data mining from the GEO database

We used the string “age-related macular degeneration”, and “*Homo sapiens*” to search the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The GSE29801 (n = 177) microarray dataset was based on the GPL4133 platform (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F). We selected 38 normal samples and 41 AMD samples from macular tissue as the training cohort. The GSE50195 (n = 16) microarray dataset included 9 AMD and 7 normal samples from the macular tissue, and it was based on the GPL17629 (Affymetrix Human Exon 1.0 ST Array). Microarray datasets were normalized and $\log_2$ transformed through “limma” R packages, and then Z-score scaling was performed [6].

Data collection and preprocessing were fully conducted under GEO data access policies. All analyses were performed under relevant guidelines and regulations.

Differentially expressed genes (DEGs) screening

To filter out genes that specific to AMD development, DEGs were identified from the GSE29801 dataset using “limma” R package [6]. The threshold was set as value of $|\log_2$-fold change (FC)$| > 1.0$ and the p-value < 0.05. Then a volcano plot was drawn by using R package to visualize the analysis of the selected DEGs.

Gene selection by applying machine learning method

Random forest (RF) was implemented in the research to perform a binary classification (AMD vs. control). The RF model is a supervised using a random selection of the predictor variables to grow classification trees on bootstrapped samples [7], which can distinguish samples with or without AMD. Feature importance was ranked by the Gini importance measures in the model by “caret” R package [8]. Top 20 genes selected by RF were included in further analyses.

Construction and validation of the diagnostic model

LASSO logistic regression was constructed using “glmnet” R package [9]. 5-fold cross-validation was applied to detect the best penalty parameter lambada. Genes with non-zero coefficients were picked out and formed the diagnostic model to minimize binomial deviance. The risk score of each sample was calculated through the regression coefficient derived from the LASSO logistic regression model multiplied by gene expression level. The median risk score was used as the cut-off point, which divided the patients in the training cohort into high-risk and low-risk groups. GSE50195 was used for external validation with the same formula.

Receiver operating characteristic (ROC) curves were employed to measure the diagnostic performance and the area under the curve (AUC) was compared between the training and validation cohorts.

Functional enrichment and gene set enrichment analyses

Metascape (https://metascape.org/gp/index.html#/main/step1) is used for gene function analysis by integrating several authoritative data resources [10]. Metascape was applied to analyze the potential signaling pathways in AMD in the present study based on the selected DEGs. The threshold was p-value < 0.05. Besides, to further analyze potential biological processes enriched in the high-risk group, gene set enrichment (GSEA) was performed by applying “clusterProfiler” R package annotated by reference gene set file (c5. Bp.v7.0. entrez.gmt) [11]. NOM p-value < 0.05 and FDR q-value < 0.25 was considered statistically significant.

Evaluation of immune cells

The amounts of immune cell subtypes were quantified by single sample gene set enrichment analysis (ssGSEA) implemented in “GSVA” R package [12]. ssGSEA applies gene signatures expressed by immune cell populations to individual high- and low-risk groups. In our study, we enrolled 28 immune cells of both innate and adaptive immunity. The correlation between levels of immune cells and genes involved in the diagnostic model were also assessed.

Statistical analyses

All statistical analyses were performed by the R software (version 4.0.3; R Foundation for Statistical Computing, Vienna, Austria). The Kruskal-Wallis test verified the differentiation of 28 kinds of immune

cells between high- and low-risk groups. Spearman coefficient test was used for the correlation test between the ten genes and immune cells. Statistical tests were all two sided. The reported results were all considered statistically significant when *p*-value was less than 0.05.

Results

Clinical characteristics

We included three phases to identify and validate the gene-based RPE/choroid complex diagnostic model for AMD (Figure 1). In the discovery phase, GSE29801 was applied to screen DEGs, which were used in the RF model to select key genes biomarkers for AMD. In the training phase, LASSO logistic regression was utilized to identify informative genes and construct the diagnostic model. In the validation phase, the performance of the model was verified in GSE50195. Besides, enrichment and immune-cell analysis were also performed in the two datasets to obtain a robust association between the model and overall immune status in AMD patients.

Identification of DEGs in AMD

A total of 83 DEGs were found between AMD and normal samples in GSE29801, including 29 up-regulated ones and 54 down-regulated ones (Figure 2A). Then the DEGs were selected robust and general genes.

Selection of candidate gene biomarkers

We adopted RF model to identify key gene biomarkers for classifying AMD and normal samples. The top 20 genes were used in the traning cohort (Figure 2 B).

Functional enrichment analysis

Metascape analysis showed the top 20 clusters of enriched biological processes (Figure 2C). Results showed that the DEGs between AMD and normal samples were significantly enriched in regulation of burn wound healing, regulation of system process, NABA SECRETED FACTORS and so on.

Construction of the diagnostic model

The 20 genes selected by RF were applied to LASSO logistic regression to build the diagnostic model. Consequently, ten optimal genes were employed to establish a diagnostic model, including BMPR2, CNOT3, CRLF1, FXYD6, HRASLS5, KRTDAP, NUDT16L1, PI16, PLAGL1, SART1 (Figure 3A, B). The risk score formula was calculated as follows: risk score = -2.50339 + (-0.02804 * expression level of BMPR2) + (0.05620 * expression level of CNOT3) + (-0.00099 * expression level of CRLF1) + (0.00811 * expression level of FXYD6) + (-0.00074 * expression level of HRASLS5) + (0.13072 * expression level of KRTDAP) + (0.01815 * expression level of NUDT16L1) + (0.00053 * expression level of PI16) + (0.00086 * expression level of PLAGL1) + (0.00074 * expression level of SART1) (Figure 3C). The ROC curves indicated a high

diagnostic power of the model, because the AUC was 0.908 (95% CI: 0.823-0.975) with a specificity of 0.782 and a sensitivity of 0.965 (Figure 4A). Patients were classified into high- and low-risk groups based on the median expression levels of the optimal genes (Figure 5A).

External validation of the diagnostic model

To verify the robustness of the ten-gene based diagnostic model, we obtained an independent cohort GSE50195. Patients were divided into high- and low-risk groups according to the fixed formula and the median expression levels of the optimal genes calculated from the training cohort (Figure 5B). The ROC curve in the validation cohort indicated a reliable diagnostic result, with the AUC of 0.809 (95% CI: 0.522-0.889) and the specificity and sensitivity of 0.644 and 0.912, respectively (Figure 4B). The model had consistently high sensitivity in both training and validation cohorts so that the diagnostic performance of the model remained robust and precise.

Gene set enrichment analysis

GSEA was conducted to elucidate the potential biological processes occurring in high-risk patients compared with low-risk ones. We found that glutathione metabolism and phototransduction are the shared pathway that had gene enriched in the training and validation cohort. In the training cohort, gene expression in the two signal pathway significantly reduced (Figure 6A). While in the validation cohort, glutathione metabolism had an increase level of gene expression (Figure 6B). The above analyses may set the foundation for further exploring the molecular mechanisms of AMD.

Inference of immune cells in retinal pigment epithelium (RPE)/choroid complex

We quantified 28 types of immune cells including the B cells, T cells, DCs, macrophages, natural killer cells and so on to investigate the composition of the RPE/choroid complex by applying ssGSEA. As a result, the levels of neutrophil in the high-risk group were significantly higher than that in the low-risk group in both training and validation datasets. In the training cohort, the levels of activated B cell, activated CD4 T cell, central memory CD4 T cell, effector memory CD8 T cell, gamma delta T cell, immature B cell, myeloid-derived suppressor cells (MDCS), natural killer T cell, T follicular helper cell, type 1 T helper cell and type2 helper cell in the high-risk group were significantly higher than that in the low-risk group (Figure 7A). While in the validation cohort, the level of monocyte increased in the high-risk group, but the levels of CD56bright natural killer cell, CD56dim natural killer cell and memory B cell decreased in the high-risk group compared with that in the low-risk group (Figure 7B).

We also analyzed the correlation between the ten genes involved in the diagnostic model and immune cells. In the training cohort, NUDT16L1, FXYD6, HRASLS5 and KRTDAP were negatively correlated with almost all the immune cells, while PLAGL1, ACSL1 and CRLF1 exhibited positive correlations (Figure 8A). In the validation cohort, the correlations between the optimal genes and the immune cells were not that strong as that in the training cohort. It showed that PI16 was negatively correlated with CD56dim natural killer cell, natural killer cell and neutrophil, and CNOT3 and HRASLS5 also had the negative correlations

with CD56dim natural killer cell. While BMPR2 and ACSL exhibited positive correlations with CD56bright natural killer cell (Figure 8B).

Discussion

AMD is a highly complex eye disease resulting in visual loss, but its pathogenesis has not been fully elucidated. Gene-environment interaction causes epigenetic modification and may play a significant role in the development of AMD [13]. Discovering genetic loci associated with AMD was one of the major successes coming from genome-wide association studies (GWAS) [14]. By 2017, 52 common and rare variants at 34 genetic loci had been identified to be independently associated with late AMD on the basis of 16144 cases of late AMD and 17832 controls [15]. GWAS loci is a valuable resource for understanding disease mechanisms, but they do not necessarily pinpoint the causal genes responsible for the disease association. This study wants to analyze the RPE/choroid transcriptomic data to determine the biomarker genes that responsible for the diseases association, which are highly relevant to AMD pathology.

In the present study, after systematically screened the datasets, two cohorts focused on AMD were included for analysis. By applying RF model and LASSO logistic regression, we identified and validated a ten-gene model for AMD diagnosis utilizing publicly RPE/choroid complex gene expression data. Our diagnostic model showed a robust accuracy in both training (AUC = 0.908) and validation (AUC = 0.809) cohorts. Notably, the model demonstrated high sensitivity (0.965 in the training cohort and 0.912 in the validation cohort, which is more important for identifying cases as seen in our results. Therefore, once classified into the high-risk group, the patients had a significantly high possibility to have ongoing AMD and need immediate further diagnosis and treatment. Du and his colleagues have applied random walk (RW) and support vector machine (SVM) models to identify AMD related genes with an AUC REACHING 0.927, which revealed the related mechanisms of AMD [16]. Ajana and his colleagues have built a bootstrap lasso model to predict AUC, which includes not only genetic factors, but also diet quality, education levels and pulse pressure [17]. Although this model achieved an AUC of 0.92, the two studies above did not test their models on the external validation sets. Compared to the former studies, our machine learning model has good result with a robust performance on the external validation set.

AMD is a multifactorial disorder, involving the mechanisms of inflammation, oxidative stress and pathological neovascularization. Using ssGSEA method, we found a higher amount neutrophil cells on RPE/choroid complex among high-risk patients. It has been revealed that variations in the complement factor H (CFH) gene have been associated with the risk of developing AMD [18–19], which is a major protein to regulate the innate immune and inflammatory response. It is also proven in our research that the neutrophil cells have negative correlation with HRASLS5, PI16 and FXYD6P2 genes.

The exposure to visible light and numerous photosensitizers results in the generation of reactive oxygen species (ROS) and subsequent oxidative damage. ROS induces RPE cell damage by eliciting changes in multiple genes and subsequently leading to RPE cell death and apoptosis. Phototransduction pathway ultimately leads to a rapid visual response in rod cells, and it links absorbtion of light to a decrease

cytosolic cGMP [20]. This process is accompanied by a decrease in free Ca^{2+} concentration in the photoreceptor cytosol sensed by Ca^{2+}-binding proteins that modulate phototransduction and activate the recovery phase to reestablish the photoreceptor dark potential [20]. Phototransduction pathway of high-risk patients were suppressed on RPE/choroid complex according to the results of gene set enrichment analysis. Glutathione (GSH) conjugation is an important detoxification mechanism and it plays an extremely important role in protecting cells against toxic injury in RPE cells [21]. GSH is the most prominent antioxidant in RPE cells and is present at high concentration in retina and RPE [22–23]. A novel type of programmed cell death, ferroptosis, has been demonstrated in AMD, which is associated with lipid peroxidation contingent on ROS levels and the availability of iron [24]. However, whether ferroptotic cell death is associated with the effects of GSH depletion on RPE is unknown and the mechanism beneath the effect of GSH depletion-induced oxidative stress in RPE cells is not well established [21]. GSH was suppressed in RPE/choroid complex in GSE29801 among high-risk patients.

Among the ten key genes involved in the current diagnostic model, BMPR2, CNOT3, CRLF1, FXYD6, HRASLS5, KRTDAP, NUDT16L1, PI16, PLAGL1, SART1 are all protein coding genes. CRLF1 and FXYD6 are related to protein heterodimerization activity and cytokine activity, which may lead to cell proliferation [25]. Gene Ontology (GO) annotations related to FXYD6 gene include ion channel activity and sodium channel regulator activity. HRASLS5 catalyzes the calcium-independent release of fatty acids from the sn-1 or sn-2 position of glycerophospholipids. The ligands of BMPR2 are members of the TGF-beta superfamily. Mutations in this gene have been associated with RPE degeneration [26]. CNOT3 (CCR4-NOT Transcription Complex Subunit 3) is one of the major cellular mRNA deadenylates and is linked to various cellular processes including bulk mRNA degradation, and the mutation can cause variable neurodevelopment disorders [27]. KRTDAP acts as a soluble regulator of keratinocyte differentiation. NUDT16L1 is associated with RNA binding and hydrolase activity. PI16 regulates peptidase inhibition, which is a fibroblast-derived protein and the inflammation-regulated inhibitor of MMP2 [28]. PLAGL1 encodes a C2H2 zinc finger protein that functions as a suppressor of cell growth. SART1 encodes two proteins, the SART1(800) protein expressed in the nucleus of the majority of proliferating cells, and the SART1(259) protein expressed in the cytosol of epithelial cancers [29]. Further research is urgently needed to verify the complex interactions between genes and the signal pathways among AMD patients.

There were some limitations in our study. First of all, the model can not specify the classification of AMD so that it can not distinguish wet AMD and dry AMD. Further prospective studies with larger cohorts in more centers were required to validate the accuracy and reproducibility of the model. Besides, the precise biological mechanisms underlying these ten genes were still unclear in AMD progression and needed to be studied in functional experiments. In addition, relationships between overall immune status and AMD progression were necessary to be clarified in the future. Despite those limitations, our study proved that RPE/choroid complex biomarkers might have the potential to alert early stage of AMD, and the ten-gene diagnostic model could be used as a non-invasive test for patients in clinical practice.

Conclusions

In summary, a novel diagnostic model for AMD consisting of ten genes was developed and validated. Measuring the expression levels of these ten genes may provide a cost-effective and accurate individualized method for clinical monitoring and diagnosis in AMD. Besides, our model was closely related to the immune status of the patients on RPE/choroid complex, which provides insights for further investigating potential mechanisms and therapeutic targets for AMD.

Abbreviations

AMD: age-related macular degeneration; GEO: Gene Expression Omnibus; LASSO: least absolute shrinkage and selection operator; GSEA: gene set enrichment analysis;

RPE: retinal pigment epithelium; RF: random forest; ROC: receiver operating characteristic; AUC: area under the curve; ssGSEA: single sample gene set enrichment analysis.

Declarations

Ethics approval and consent to participants

Ethics approval and informed consent were not required for this study because of public accessibility to the data.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available in the NCBI Gene Expression Omnibus repository (GEO: GSE29801), https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29801 and (GEO: GSE50195), https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi.

Competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This study was supported by the National Natural Science Foundation of China (82141128); The Capital Health Research and Development of Special (2020-1-2052); Science & Technology Project of Beijing Municipal Science & Technology Commission (Z201100005520045, Z181100001818003).

Authors' contributions

W.B. Wei, H.Y Li and L. Dong designed the study, H.Y Li and L. Dong wrote the manuscript. H.Y Li, C.Y Yu, W.D Zhou, H.T Wu, Y.F Li, and Y.T Licollected the data and conducted the analyses, W.B. Wei edited and revised the manuscript. All authors have approved the submitted version and agreed with the contributions declarations.

Acknowledgement

Not applicable.

References

1. Neale BM, Fagerness J, Reynolds R, et al. Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). Proc Natl Acad Sci USA 2010; 107: 7395–400.
2. Chen W, Stambolian D, Edwards AO, et al. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. Proc Natl Acad Sci USA 2010; 107: 7401–06.
3. McKay GJ, Patterson CC, Chakravarthy U, et al. Evidence of association of APOE with age-related macular degeneration—a pooled analysis of 15 studies. Hum Mutat 2011; 32: 1407–16.
4. Yang, S., Zhao, J., and Sun, X. Resistance to anti-VEGF therapy in neo- vascular age-related macular degeneration: a comprehensive review. Drug Des. Devel. Ther 2016; 10: 1857–1867.
5. Walsh, J. A., Rozycki, M., Yi, E., and Park, Y. Application of Machine Learning in the Diagnosis of Axial Spondyloarthritis. Curr. Opin. Rheumatol 2019; 31 (4): 362–367.
6. Ritchie, M. E., Phipson, B., Wu, D., et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. Nucleic Acids Res 2015; 43 (7): e47.
7. Segev N, Harel M, Mannor S, Crammer K, El-Yaniv R. Learn on Source, Refine on Target: A Model Transfer Learning Framework with Random Forests. IEEE Trans Pattern Anal Mach Intell. 2017;39(9):1811–1824.
8. Kuhn M. The caret Package[J]. Journal of Statistical Software, 2009, 28.
9. Friedman, J., Hastie, T., and Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw 2010; 33 (1): 1–22.
10. Zhou, Y., Zhou, B., Pache, L., et al. Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. Nat. Commun 2019; 10 (1): 1523.
11. Yu, G., Wang, L.-G., Han, Y., et al. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A J. Integr. Biol 2012. 16 (5), 284–287. doi:10.1089/omi.2011.0118
12. Hänzelmann, S., Castelo, R., and Guinney, J. GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. BMC bioinformatics 2013; 14: 7.
13. Gemenetzi M, Lotery AJ. Epigenetics in age-related macular degeneration: new discoveries and future perspectives. Cell Mol Life Sci. 2020;77(5):807–818.

14. Fritsche LG, Fariss RN, Stambolian D, et al. Age-related macular degeneration: genetics and biology coming together. Annu Rev Genomics Hum Genet 2014; 15: 151–71.
15. Fritsche LG, Igl W, Bailey JN, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat Genet 2016; 48: 134–43.
16. Du Y, Kong N, Zhang J. Genetic Mechanism Revealed of Age-Related Macular Degeneration Based on Fusion of Statistics and Machine Learning Method. Frontiers in genetics. 2021;12.
17. Ajana S, Cougnard-Grégoire A, Colijn JM, et al. Predicting progression to advanced age-related macular degeneration from clinical, genetic, and lifestyle factors using machine learning. Ophthalmology. 2021;128(4):587–97.
18. Hughes AE, Meng W, Bridgett S, and Bradley DT. Rare CFH mutations and early-onset age-related macular degeneration. Acta Ophthalmol 2016; 94: e247–e248.
19. Munch IC, Toft U, Linneberg A, and Larsen M. Precursors of age-related macular degen- eration: associations with vitamin A and interaction with CFHY402H in the Inter99 Eye Study. Acta Ophthalmol 2016; 94: 657–662.
20. Ricardo, Stephen, Sławomir, et al. Ca2+-dependent Regulation of Phototransduction[J]. Photochemistry and Photobiology, 2008.
21. Yun S, Zheng Y, Wang C, et al. Glutathione depletion induces ferroptosis, autophagy, and premature cell senescence in retinal pigment epithelial cells[J]. Cell Death & Disease, 2018, 9(7):753-.
22. Saxena, M., Singhal, S. S., and Awasthi, Y. C. A specific, sensitive, and rapid method for the determination of glutathione and its application in ocular tissues. Exp. Eye Res 1992; 55: 461–468.
23. Huster, D. et al. Subcellular compartmentation of glutathione and glutathione precursors. A high resolution immunogold analysis of the outer retina of guinea pig. Anat. Embryol 1998; 198: 277–287.
24. A.Southon,K.Szostak,K.M.Acevedo,K.A.Dent,I.Volitakis,A.A.Belaidi,etal.,Cu (II) (atsm) inhibits ferroptosis: implications for treatment of neurodegenerative disease, Br. J. Pharmacol 2020; 177 (3): 656–667.
25. Suppression of CRLF1 promotes the chondrogenic differentiation of bone marrow-derived mesenchymal stem and protects cartilage tissue from damage in osteoarthritis via activation of miR-320[J]. Molecular Medicine, 2021, 27(1).
26. Miyagawa K, Shi M, Chen P I, et al. Smooth Muscle Contact Drives Endothelial Regeneration by BMPR2-Notch1–Mediated Metabolic and Epigenetic Changes[J]. Circulation Research, 2019, 124(2):211–224.
27. Martin R, Splitt M, Genevieve D, et al. De novo variants in CNOT3 cause a variable neurodevelopmental disorder[J]. European Journal of Human Genetics, 2019, 27(11):1677–1682.
28. Singhmar P, Trinh R, Ma J, et al. The fibroblast-derived protein PI16 controls neuropathic pain[J]. Proceedings of the National Academy of Sciences, 2020, 117(10):201913444.

29. Narita M, Kanda T, Abe T, et al. Immune responses in patients with esophageal cancer treated with SART1 peptide-pulsed dendritic cell vaccine[J]. International Journal of Oncology, 2015, 46(4).

Figures

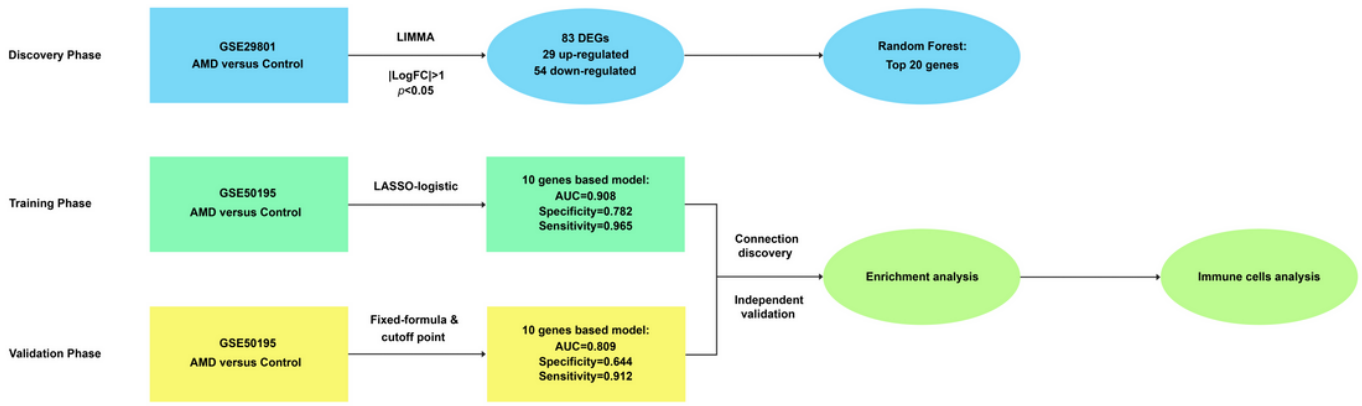

Figure 1

Flow chart of the study. The discovery phase chose candidate genes for AMD, which were used to develop a ten-gene based model in the traning phase. Diagnostic ability of the model was verified in the validation phase. Gene set enrichment analysis and immune cells analysis were conducted to detect potential mechanisms of AMD. AUC: area under the curve.

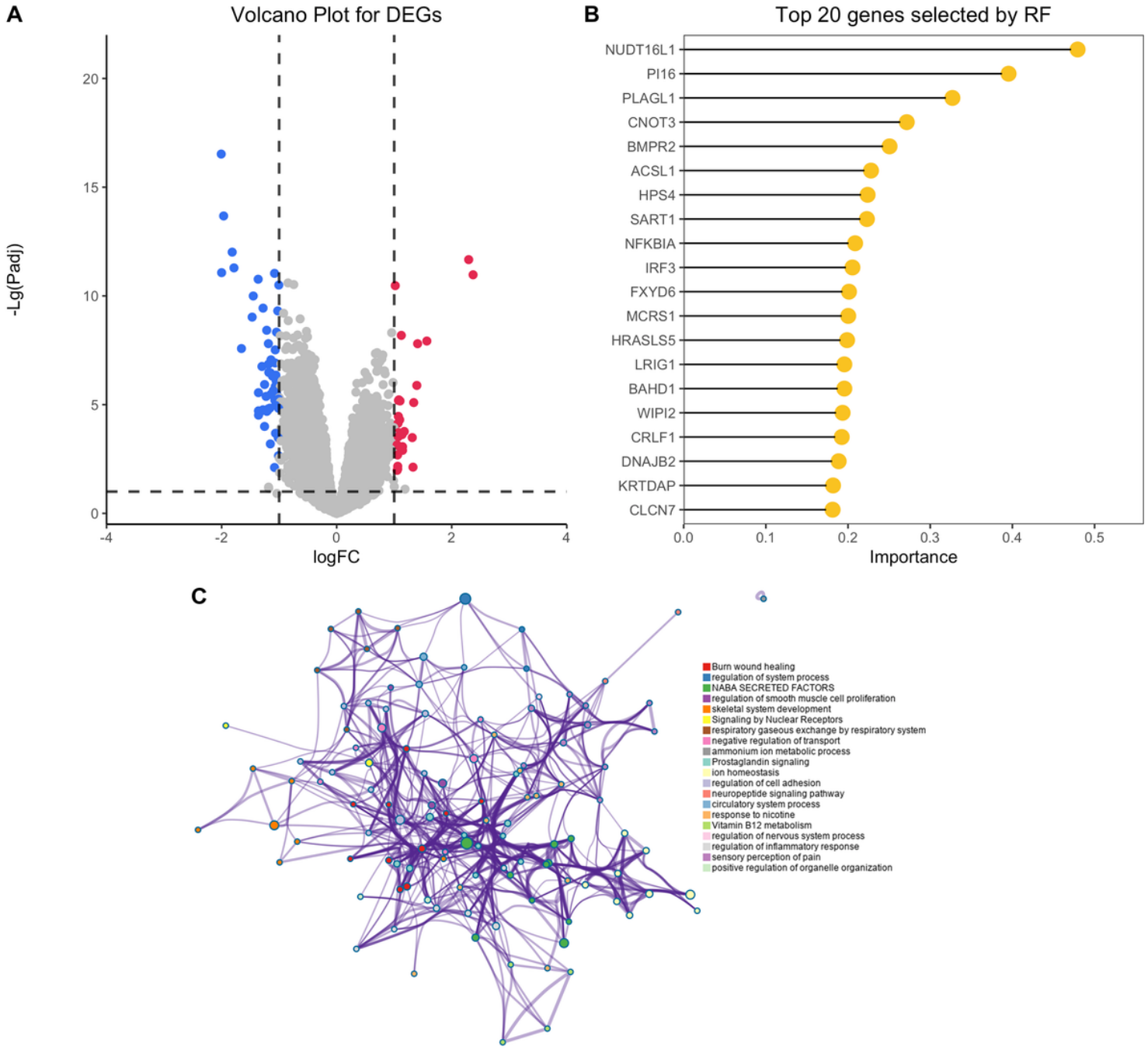

Figure 2

Screening for diagnostic gene biomarkers for AMD. (A) Volcano plot shows DEGs between AMD and control RPE/choroid complex samples. Blue dots denote down-regulated genes and red ones denote up-regulated genes. (B) Lollipop chart shows top ranked 20 genes ordered by Gini-importance through RF. (C) Network of enriched clusters, where each node represents on statistically significant term and terms with similarity of more than 0.3 are connected by edges. RPE: retinal pigment epithelium. DEGs: differentially expressed genes. FC: fold change. RF: random forest.

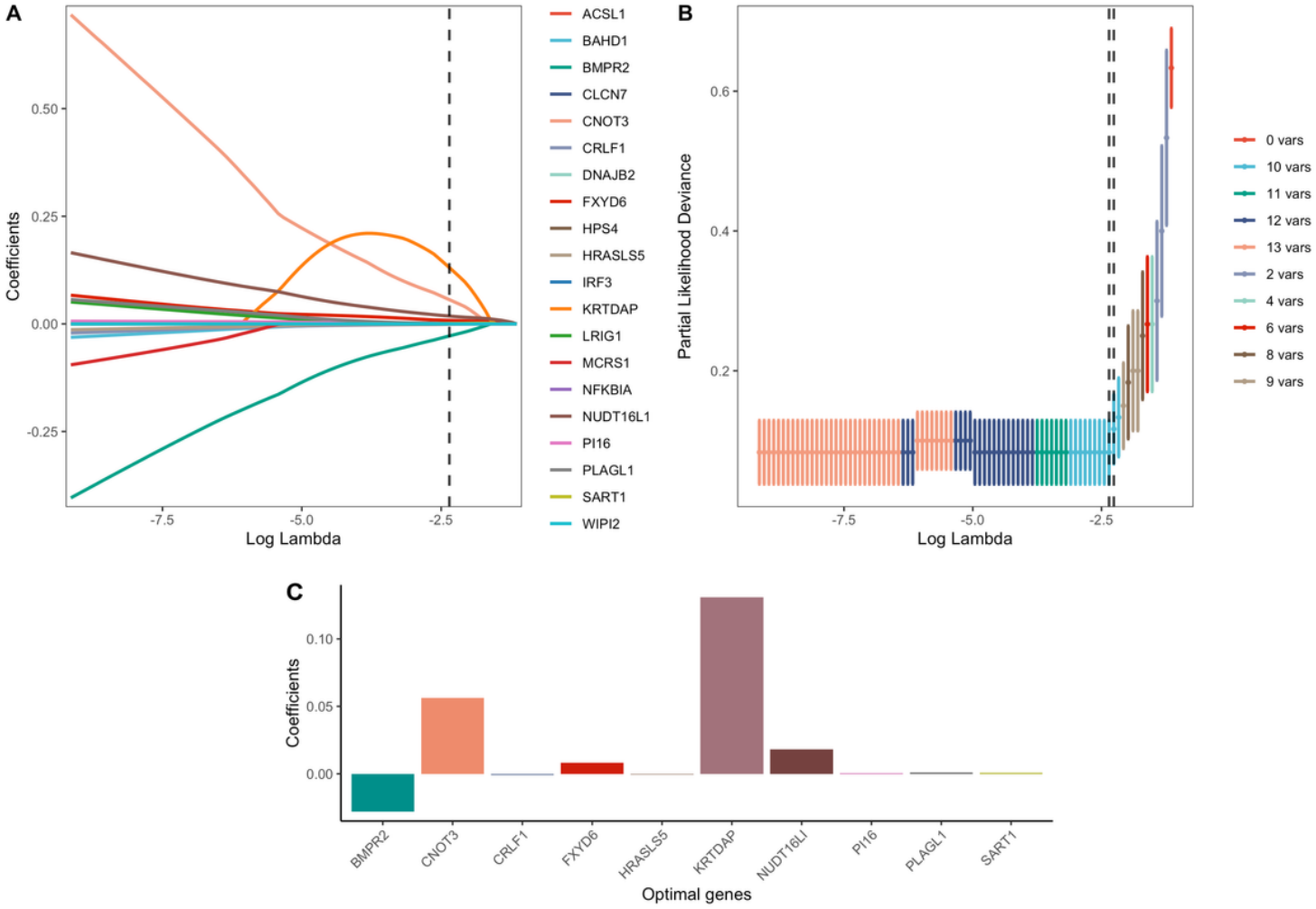

Figure 3

Construction and validation of the ten-gene based diagnostic model. (A) LASSO coefficient progiles, where each curve represenr a DEG. Ten DEGs were finally selected under the optimal lambda to construct a diagnostic model. (B) Selection of the tuning parameter in LASSO logistic regression analysis. Five-fold cross-validation was utilized to calculate the optimal lambda which leads to minimun mean cross-validation error. (C) The bar plot shows coefficients of ten genes in the diagnostic model.

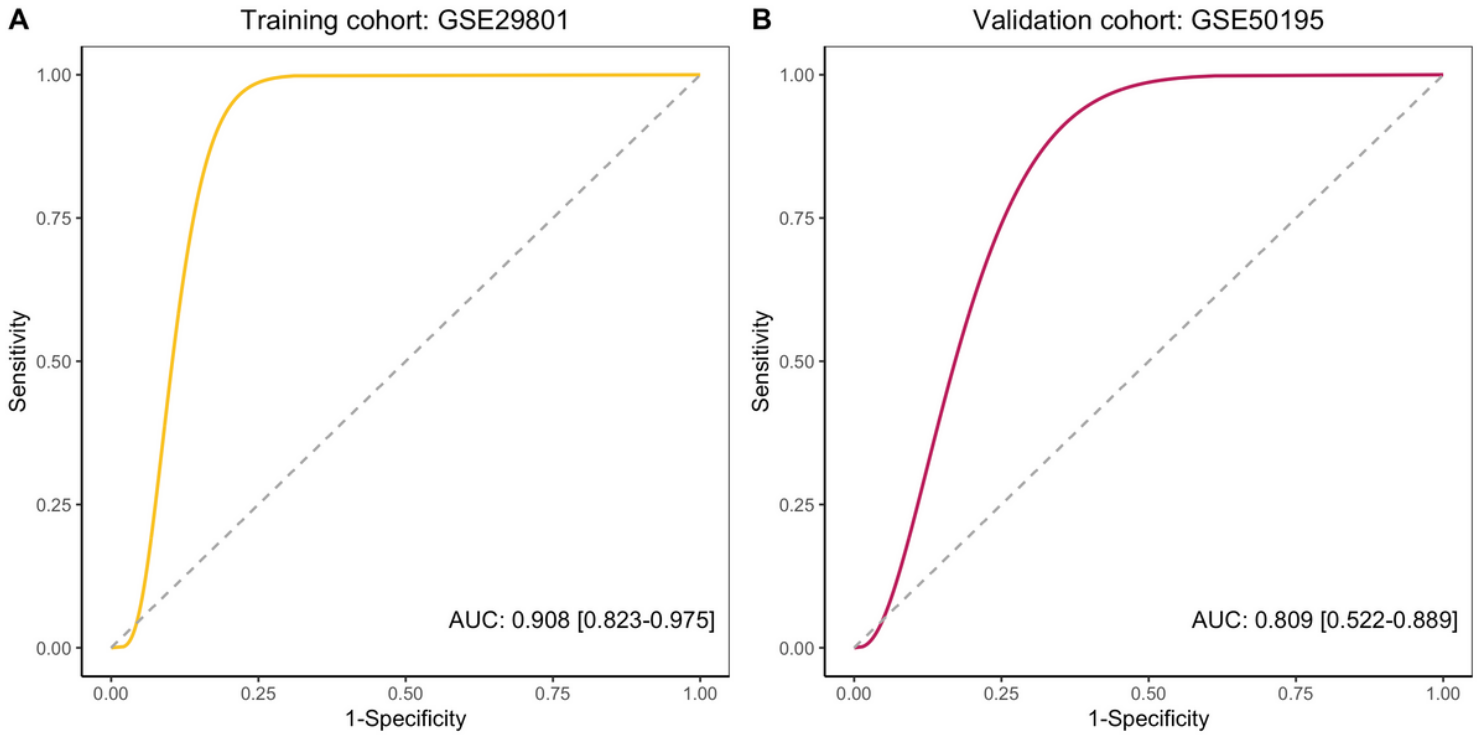

Figure 4

Receiver Operating Characteristic Curve and AUC of LASSO logistic regression for AMD diagnosis prediction. (A) ROC curve in the traning (GSE29801) cohort. (B) ROC curve in the validation (GSEA50195) cohort. ROC curve: receiver operating characteristic curve. AUC: area under the curve.

Figure 5

Heatmaps for expression levels of ten genes. Patients annotations includes disease status and risk score. (A) Heatmap for genes expression in the traning cohort. (B) Heatmap for genes expression in the validation cohort.

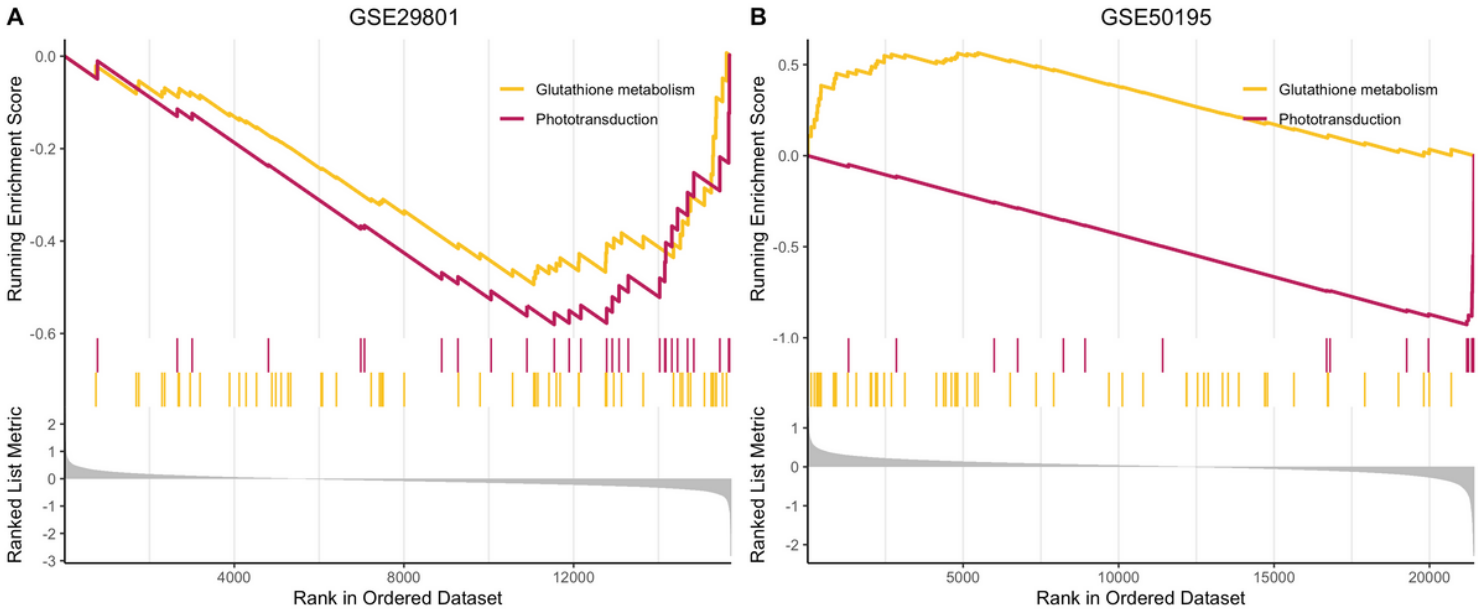

Figure 6

Gene set enrichment analysis (GSEA). GSEA plots show shared biological processes associated with the diagnostic model. (A) GSEA plots in the training (GSE29801) cohort. (B) GSEA plots in the validation (GSEA50195) cohort.

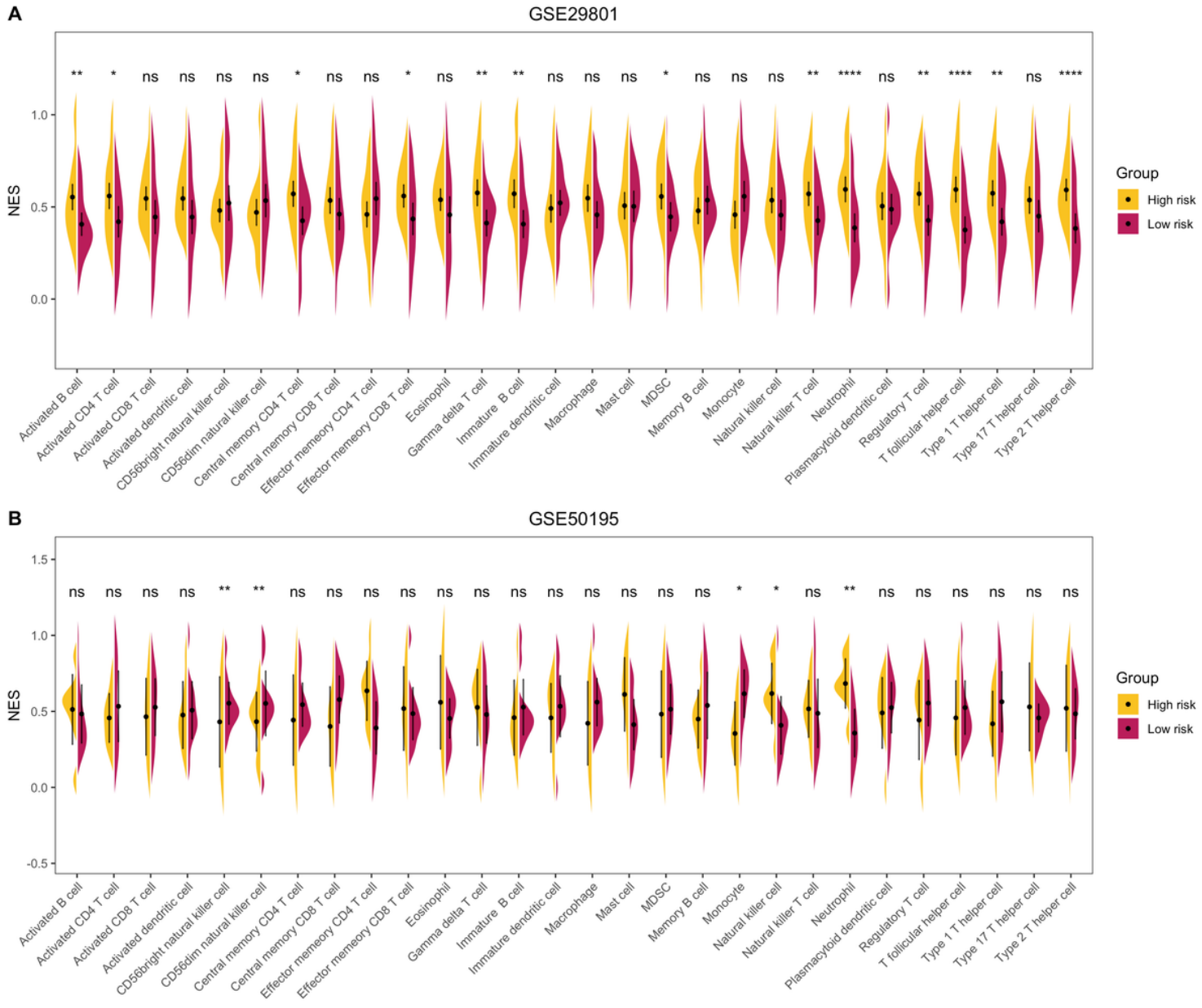

Figure 7

Immune cells analysis. (A) Split-violin plots for comparisons of immune cell levesl between high- and low-risk patients in the traning (GSE29801) cohort. (B) Split-violin plots for comparisons of immune cell levesl between high- and low-risk patients in the validation (GSEA50195) cohort. *$p < 0.05$; **$p < 0.1$; ***$p < 0.001$, ****$p < 0.0001$; ns: not statisticaaly significant.

Figure 8

Immune cells analysis. (A) Correlation heatmaps between the expression levels of ten genes and immune cell levels in the traning (GSE29801) cohort. (B) Correlation heatmaps between the expression levels of ten genes and immune cell levels in the validation (GSEA50195) cohort. *$p < 0.05$; **$p < 0.1$; ***$p < 0.001$;

ns: not statisticaaly significant. NES: normalized enrichment score. MDSC: myeloid-derived suppressor cells.