



Genetic Mechanism Revealed of Age-Related Macular Degeneration Based on Fusion of Statistics and Machine Learning Method

Yongyi Du¹, Ning Kong¹ and Jibin Zhang^{2*}

¹ Department of Ophthalmology, Panyu Central Hospital, Guangzhou, China, ² Department of Stomatology, Panyu Central Hospital, Guangzhou, China

OPEN ACCESS

Edited by:

Lei Deng,
Central South University, China

Reviewed by:

Hong Ju,
Heilongjiang Vocational College
of Biology Science and Technology,
China
Hui Ding,
University of Electronic Science
and Technology of China, China

*Correspondence:

Jibin Zhang
kn13@tom.com

Specialty section:

This article was submitted to
Statistical Genetics and Methodology,
a section of the journal
Frontiers in Genetics

Received: 17 June 2021

Accepted: 13 July 2021

Published: 05 August 2021

Citation:

Du Y, Kong N and Zhang J (2021)
Genetic Mechanism Revealed
of Age-Related Macular Degeneration
Based on Fusion of Statistics
and Machine Learning Method.
Front. Genet. 12:726599.
doi: 10.3389/fgene.2021.726599

Age-related macular degeneration (AMD) is the most common cause of irreversible vision loss in the developed world which affects the quality of life for millions of elderly individuals worldwide. Genome-wide association studies (GWAS) have identified genetic variants at 34 loci contributing to AMD. To better understand the disease pathogenesis and identify causal genes for AMD, we applied random walk (RW) and support vector machine (SVM) to identify AMD-related genes based on gene interaction relationship and significance of genes. Our model achieved 0.927 of area under the curve (AUC), and 65 novel genes have been identified as AMD-related genes. To verify our results, a statistics method called summary data-based Mendelian randomization (SMR) has been implemented to integrate GWAS data and transcriptome data to verify AMD susceptibility-related genes. We found 45 genes are related to AMD by SMR. Among these genes, 37 genes overlap with those found by SVM-RW. Finally, we revealed the biological process of genetic mutations leading to changes in gene expression leading to AMD. Our results reveal the genetic pathogenic factors and related mechanisms of AMD.

Keywords: AMD, GWAS, eQTL, SNPs, disease susceptibility

INTRODUCTION

Age-related macular degeneration (AMD) is the most common cause of irreversible blindness with limited therapeutic options in the elderly in many countries (Lim et al., 2012). AMD causes decreased photoreceptor function in the macular area of the retina (Fritsche et al., 2014). Researchers have found many factors which are related to the development and severity of AMD.

Genetic factors are significantly related to AMD. In 2005, Klein et al. found that CFH gene was related to AMD, which was the first discovered AMD-related gene (Haines et al., 2005). This gene is significantly expressed in retinal pigment epithelial cells. Y402H mutation of CFH impairs the complement pathway regulation function of CFH gene (Landowski et al., 2019). Subsequently, the ARMS2 gene cluster was also found to be related to AMD. Multiple studies have shown that there is a strong correlation between multiple genetic variants in this gene cluster and AMD (Johnson et al., 2001). Recently, it has been discovered that the apolipoprotein E (APOE) gene

has a strong correlation with AMD (Fernández-Vega et al., 2020). The APOE gene plays a role in transporting lipids and cholesterol in the central nervous system, and multiple studies have shown that this gene is associated with neurodegenerative diseases such as Alzheimer's disease and stroke (Feher et al., 2006; Zhao et al., 2019, 2020d). The gene is expressed on photoreceptor cells, retinal ganglion cells, retinal pigment epithelial cells, Bruch's membrane, and the choroid. Most studies have proved APOE can prevent AMD (Pang et al., 2000). The genetic risk of advanced AMD is increased (Heiba et al., 1994). Researchers have found that the heritability estimate for twin studies is 0.45 for early AMD (Hammond et al., 2002) but 0.71 for late AMD (Seddon et al., 2005).

Computational methods have been widely used to discover functions of biological molecules (Zhao et al., 2020a,b, 2021a). AMD-related genome-wide association studies (GWAS) analyses have identified a strong association of 52 independent single-nucleotide polymorphisms (SNPs) at 34 genetic loci accounting for over 50% of the genetic heritability (Fritsche et al., 2016). Machine learning methods can help researchers find disease-related information on a large scale. However, these methods cannot explain the genetic mechanism of the results. GWAS studies are a valuable resource for understanding disease pathologies, but they may not precisely point out the causal genes responsible for the disease of interest. Besides, there have been studies that reported that causal genes are distinct from the nearest genes discovered by GWAS (Smemo et al., 2014; Claussnitzer et al., 2015). However, The gene expression is related to the genetic variant so the gene expression levels are different in different genotypes (Zhao et al., 2020c). Expression quantitative trait locus (eQTL) mapping offers a powerful approach to elucidate the genetic component underlying altered gene expression. Gene expression is vital for complex diseases (Zhao et al., 2021b) and is also differentially regulated across tissues, such as the brain, heart, and pancreas. Ratnapriya et al. (2019) have found potential causal genes in six AMD GWAS loci from human retinal samples. However, that analysis only considered retinal samples and was not comprehensive since it is difficult to obtain multiple living tissues and most eQTL studies so far have been performed with RNA isolated from immortalized lymphoblasts or lymphocytes. In this study, we fused random walk (RW) with support vector machine (SVM) to identify AMD-related genes. Since many GWAS and eQTL studies have been made public, to verify our results, AMD GWAS data and blood eQTL studies are integrated to further find expression of the genes related to AMD. In this method, we referred to the concept of Mendelian randomization (MR) analysis (Davey Smith and Ebrahim, 2003; Katan, 2004), where a genetic variant (such as a SNP) is considered as an instrumental variable (such as gene expression) to validate for the causative effect of an exposure on an outcome (such as a phenotype). Based on this assumption, we can obtain AMD-related genes based on MR. We collected eQTL data from the GTEx database and collected GWAS datasets including 12,711 advanced AMD cases and 14,590 controls of European descent from a study by Han et al. (2020); 707 Caucasian AMD patients and 2,014 controls from a study by Yan et al. (2018); and 14,034 cases, 91,214

controls, and 11 sources of data including the International AMD Genomics Consortium, IAMDGC, and United Kingdom Biobank (UKBB) from a study by Winkler et al. (2020). Based on these GWAS studies and eQTL dataset, we can not only identify genes related to AMD but also speculate on their biological processes.

MATERIALS AND METHODS

Encoding Gene Interaction Network by Random Walk

The RW algorithm is a method that is simple to operate but not easy to fall into a local minimum. We constructed a gene interaction network by known AMD-related genes and a string database. Then, we implemented RW on the gene interaction network.

$f(x)$ is a multivariate function with n variables; $x = (x_1, x_2, \dots, x_n)$ is an n dimension vector.

Step 1: Given the initial iteration point x , λ is the first walking step length, and ϵ is the control accuracy (ϵ is a very small positive number, used to control the end of the algorithm).

Step 2: Given the number of iterations control N , k is the current iteration number; set $k = 1$.

Step 3: When $k < N$, randomly generate an n -dimensional vector between $(-1, 1)$. $u = (u_1, u_2, \dots, u_n)$, $(-1 < u_i < 1, i = 1, 2, \dots, n)$, and standardize it to get u' .

$$u' = \frac{u}{\sqrt{\sum u_i^2}}$$

Let $x_1 = x + \lambda u'$ to complete the first step of walking.

Step 4: Calculate the value of the function, if $f(x_1) < f(x)$, which is a better point than the initial value, then reset k to 1, change x_1 to x , and go back to step 2; otherwise, $k = k + 1$. Go back to step 3.

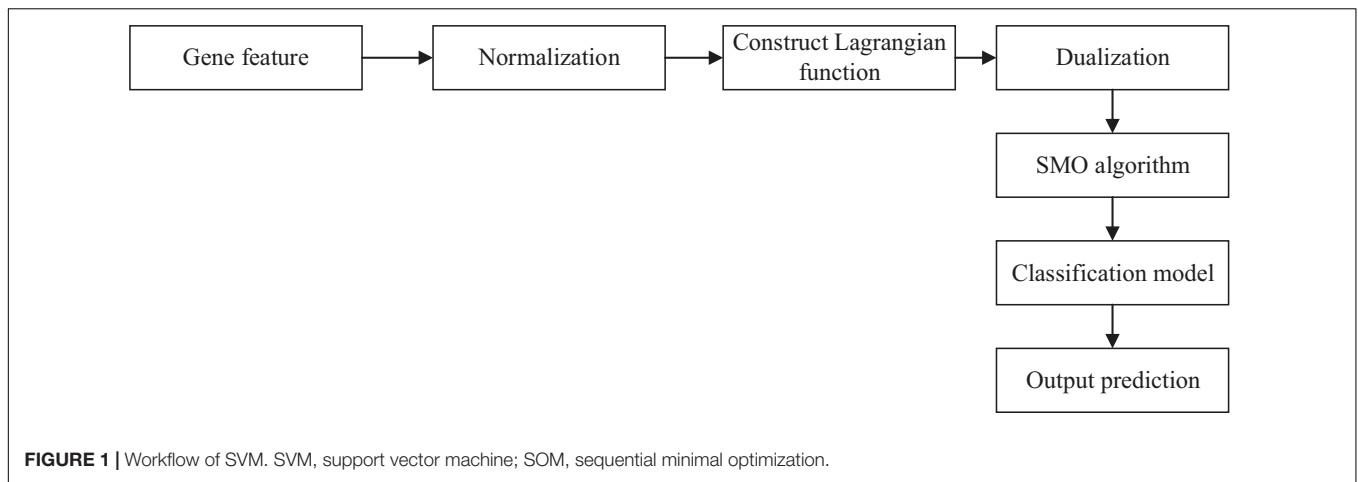
Step 5: If no better value can be found for N consecutive times, it is considered that the optimal solution is within the N -dimensional sphere with the current optimal solution as the center and the current step as the radius (if it is three-dimensional, it just happens to be in the space sphere). At this point, if $\lambda < \epsilon$, the algorithm ends; otherwise, let $\lambda = \lambda/2$, go back to step 1, and start a new round of walking.

Finally, we can get the gene feature after encoding the gene network.

Classification by Support Vector Machine

We obtained the gene feature in the last section. Then, we can input the gene feature and label into SVM to get the relationship between the gene and AMD. The workflow of SVM is shown in **Figure 1**.

First, we used Z-score normalization to process the gene feature. Then, we constructed a Lagrangian function to obtain the values and dualized the original problem. Sequential minimal optimization (SMO) algorithm was used to solve the dualization



problem. Finally, we can obtain the classification model and output the prediction results.

RESULTS

AMD-Related Genes Identification by SVM-RW

We obtained 34 known AMD-related genes from GWAS data. We constructed a gene network which has 239 nodes (genes). We did 10-cross validation by SVM-RW and tested the performance of SVM-RW. The area under the curve (AUC) of SVM-RW is shown in **Figure 2**.

SVM-RW achieved AUC of 0.927 in identifying AMD-related genes. We compared the results of SVM-RW with several other methods. The results are shown in **Table 1**.

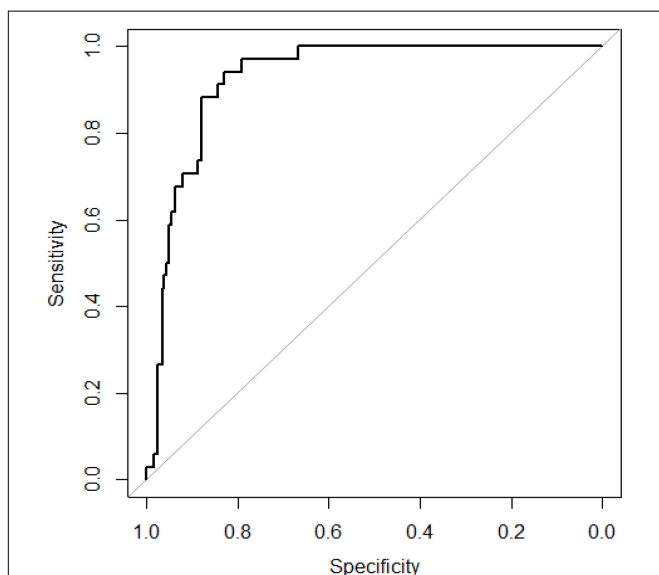


FIGURE 2 | ROC curve of SVM-RW. ROC, receiver-operator characteristic; SVM-RW, support vector machine and random walk.

After verifying the effectiveness of SVM-RW, we randomly selected 34 genes as negative samples and built a final SVM model. SVM-RW predicted 65 novel genes as AMD-related genes.

Verify SVM-RW Results by Summary Data Level-Mendelian Randomization Analysis

If we use g to denote a genetic variant (such as a SNP), x as the expression level of a gene, and y as the trait, then the two-step least-squares (2SLS) estimate of the effect of x on y from an MR analysis can be denoted as:

$$\hat{E}_{xy} = \hat{E}_{zy} / \hat{E}_{zx} \tag{1}$$

where \hat{E}_{zy} and \hat{E}_{zx} indicate the least-squares estimates of y and x on z , respectively, and E_{xy} indicates the effect size of x on y free of confounding from non-genetic factors. Then the sampling variance of the 2SLS estimate of E_{xy} can be denoted as:

$$\text{var}(\hat{E}_{xy}) = \left[\text{var}(y) (1 - P_{xy}^2) \right] / \left[n \text{var}(x) P_{zy}^2 \right] \tag{2}$$

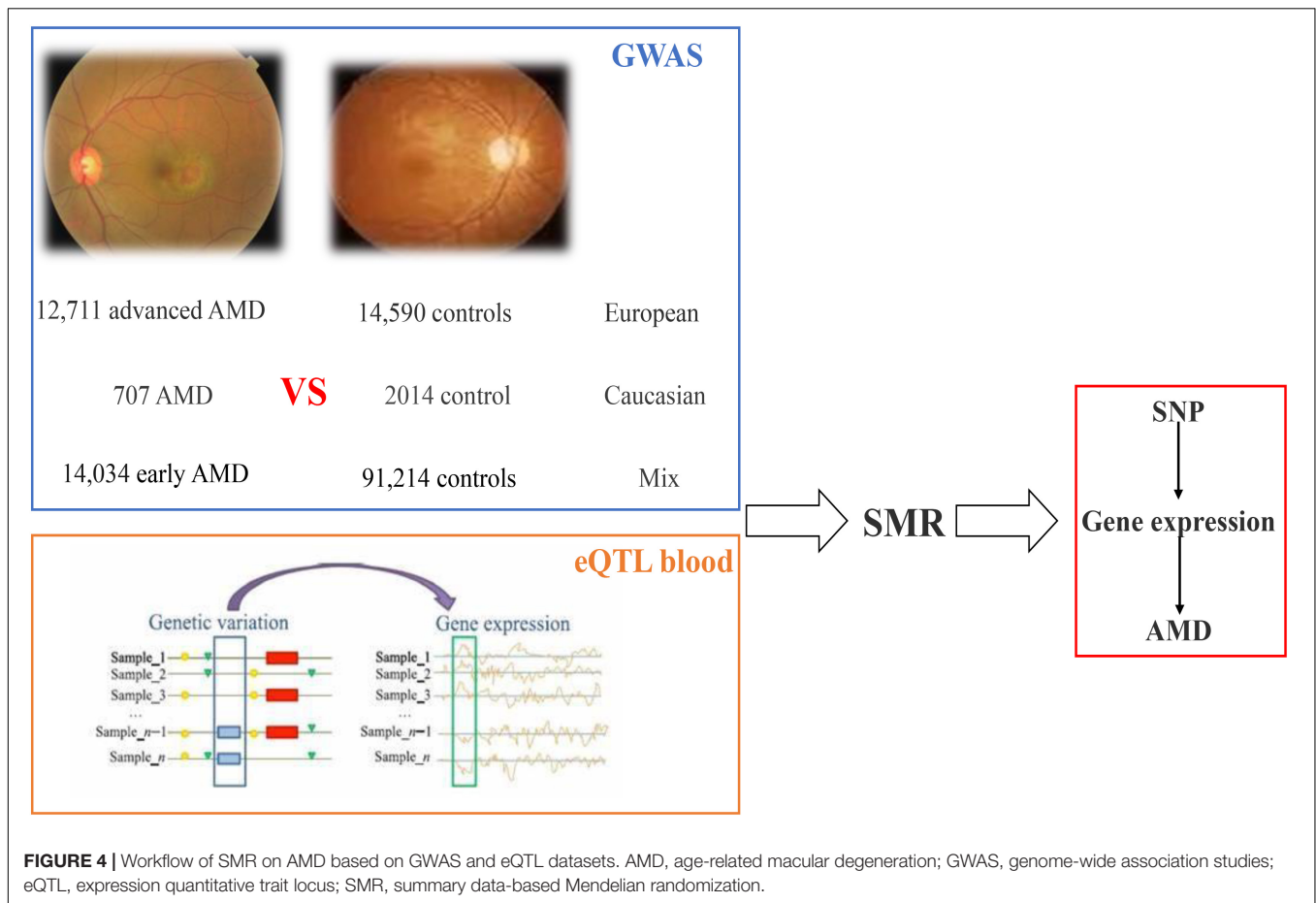
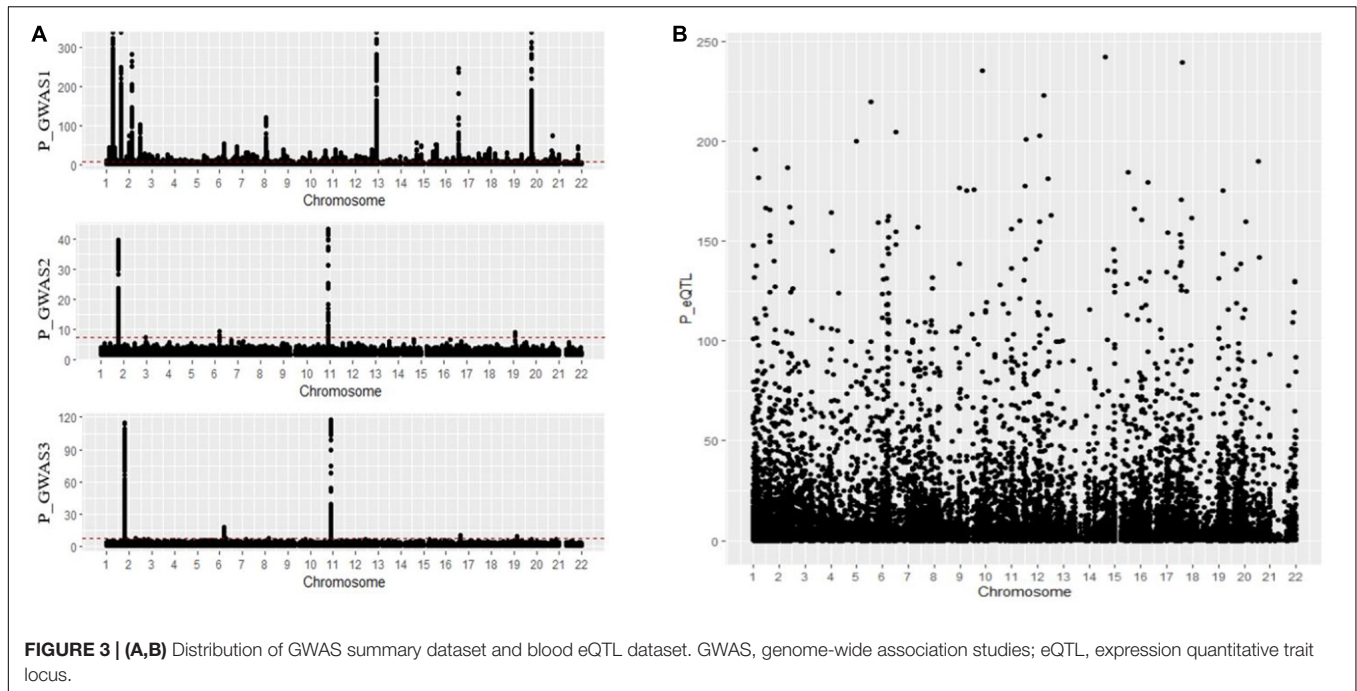
where n denotes the sample size, P_{xy}^2 indicates the proportion of variance in the explanation of y by x , and P_{zy}^2 is the proportion of variance in the explanation of y by z . Therefore, we use the

TABLE 1 | Comparison results.

Algorithm	AUC	AUPR
SVM-RW	0.927	0.781
Random forest-RW	0.852	0.645
Naive Bayes-RW	0.711	0.586
Backpropagation-artificial neural network-RW	0.823	0.692
Logistic regression-RW	0.691	0.531

AUC, area under the curve; AUPR, area under the precision-recall curve; RW, random walk.

Bold values highlight the result of SVM-RW.



statistic T_{MR} to test the significance of E_{xy} ; T_{MR} can be denoted as:

$$T_{MR} = \hat{E}_{xy}^2 / \text{var}(\hat{E}_{xy})$$

where $T_{MR} = \chi_1^2$.

Based on the suggestion that the power of detecting E_{xy} can be significantly increased using a two-sample MR analysis (Inoue and Solon, 2010; Pierce and Burgess, 2013), if GWAS and eQTL datasets share the same population, we can use unbiased estimates $\hat{\epsilon}_{zx}$ to replace E_{zx} . We therefore have

$$\hat{E}_{xy} = \hat{E}_{zy} / \hat{\epsilon}_{zx} \tag{3}$$

where \hat{E}_{zy} is the estimate of a SNP effect from a GWAS for a trait, and $\hat{\epsilon}_{zx}$ is the estimate of a SNP effect on the expression level of

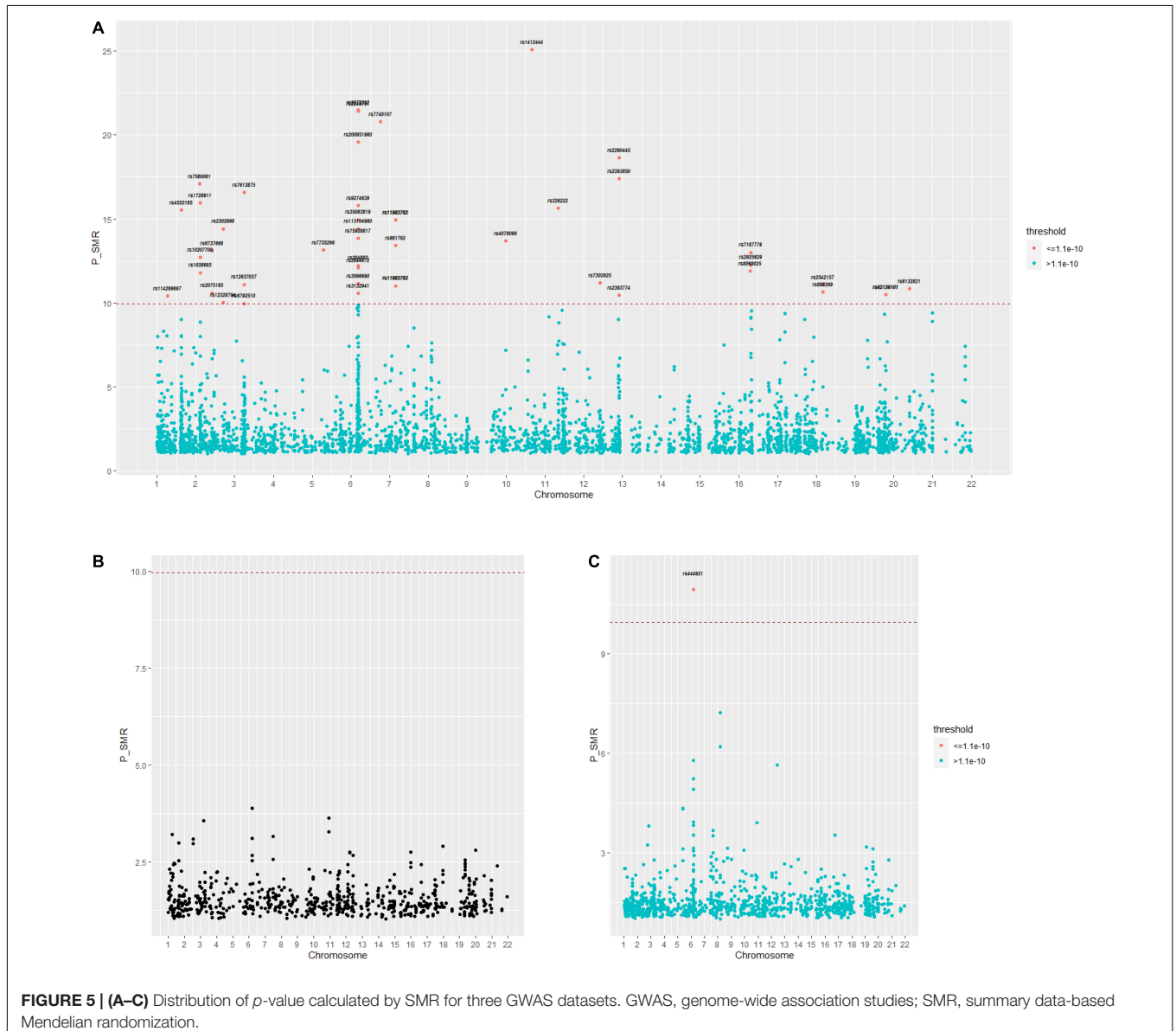
a gene from an eQTL study. The sampling variance of \hat{E}_{xy} can be approximately computed by the Delta method (Lynch and Walsh, 1998) as:

$$\text{var}(\hat{E}_{xy}) \approx \frac{E_{zy}^2}{\epsilon_{zx}^2} \left[\frac{\text{var}(\hat{\epsilon}_{zx})}{\epsilon_{zx}^2} + \frac{\text{var}(\hat{E}_{zy})}{E_{zy}^2} - \frac{2\text{cov}(\hat{\epsilon}_{zx}, \hat{E}_{zy})}{\epsilon_{zx} E_{zy}} \right] \tag{4}$$

where $\text{cov}(\hat{\epsilon}_{zx}, \hat{E}_{zy})$ is 0. Based on experience, we can replace them by their estimates in practice, indicated as an approximate χ^2 test statistic of:

$$T_{SMR} = \frac{\hat{E}_{xy}^2}{\text{var}(\hat{E}_{xy})} \approx \frac{z_{zy}^2 z_{zx}^2}{z_{zy}^2 + z_{zx}^2} \tag{5}$$

where z_{xy} is the z statistic of the GWAS and z_{zx} is the z statistic of the eQTL study.



In an MR analysis, E_{xy} is interpreted as the effect of a phenotype on the gene expression without considering non-genetic confounders. We first collected GWAS summary data and blood eQTL data from available online studies. We first collected a GWAS summary dataset composed of 12,711 advanced AMD cases and 14,590 controls of European descent from the study by Han et al. (2020); 707 Caucasian AMD patients and 2,014 controls from the study by Yan et al. (2018); and 14,034 cases, 91,214 controls, and 11 sourced from datasets including the International AMD Genomics Consortium, IAMDGC, and UKBB from the study by Winkler et al. (2020). The distribution of the above datasets is shown in **Figures 3A,B**.

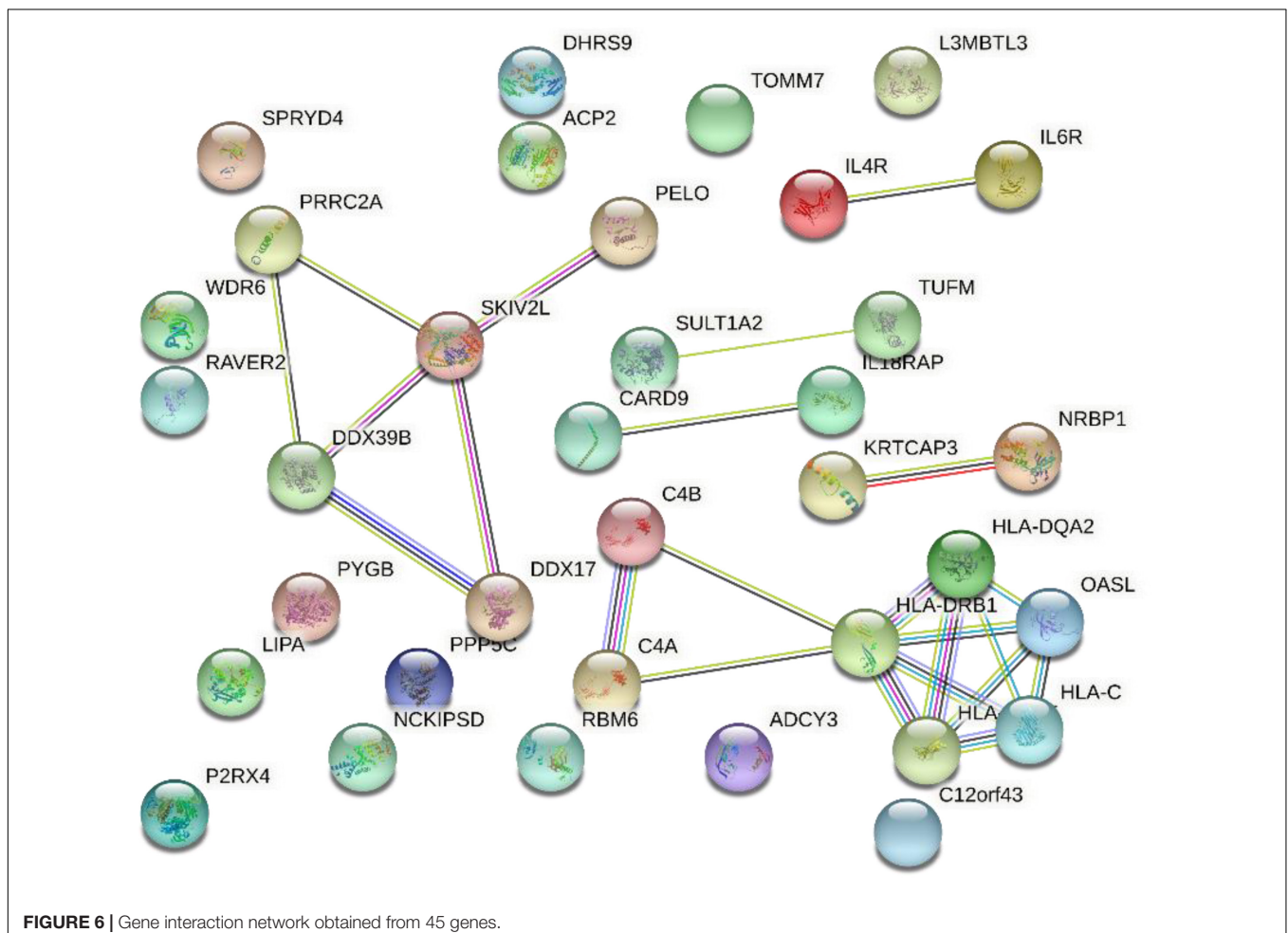
Then summary data-based Mendelian randomization (SMR) analysis is implemented on the blood eQTL data and GWAS data; in this paper, we identified 48 SNPs regulating 45 genes (including 41 coding genes and four non-coding genes) resulting in AMD susceptibility. The workflow is shown in **Figure 4**.

For the first GWAS datasets consisting of 12,711 AMD cases and 14,590 controls from European cohorts, in total we found 3,872 SNPs coexist in both GWAS data and eQTL data; 43 of 3,872 SNPs are significant and regulate 44 genes in gene expression level. In the second GWAS dataset, we found 714 SNPs coexist in both GWAS dataset and eQTL dataset, with

none significant. In the third GWAS dataset, we found 1,149 SNPs coexist both in GWAS dataset and eQTL dataset, with one significant regulating one gene in gene expression level. The distribution of the p -value of SNPs regulating genes tested by SMR is shown in **Figures 5A–C**. A **Supplementary Table 1** indicates the p -values of significant SNPs regulating genes tested by SMR; the last line resulted from GWAS dataset 3, and the rest resulted from GWAS dataset 1.

Case Study

Age-related macular degeneration has been described as a partly genetic disease (Heiba et al., 1994; Stone et al., 2004). Recently, a unifying hypothesis is that immune response gene polymorphisms modulate susceptibility to AMD. Human leukocyte antigen (HLA) polymorphisms, encoded within the major histocompatibility complex (MHC), are the most polymorphic within the human genome. In AMD, researchers detected intense HLA-DR immunoreactivity in not only soft but also hard drusen (Mullins et al., 2000). In the study of Goverdhan et al. (2005), considering the effect of smoking, age, and body mass index (BMI), HLA alleles B*4001, DRB1*1301, and Cw*0701 were found to be related to AMD, which is consistent with our results displayed in **Table 1**.



In a study by Gu et al. (2013), they researched P2RX7 and P2RX4 genes in 744 AMD patients and 557 Caucasian controls and reached a conclusion that a rare functional haplotype of the P2RX4 leads to loss of innate phagocytosis and confers increased risk of AMD. P2RX7 and P2RX4 damage the normal scavenger function of macrophages and microglia through interaction, making individuals susceptible to AMD.

Gene Interaction Network Based on AMD

Figure 6 shows the gene interaction network produced from the results of SMR on AMD. Based on the interaction network, the HLA class intensively interacted and is significantly associated with AMD.

The cluster consisting of DDX39B (aka BAT1), PRRC2A (aka BAT2), and SKIV2L are genes found in the class III region of the MHC (MHC Class III). These genes encode RNA-binding proteins with clear roles in post-transcriptional gene regulation and RNA surveillance. They are likely to have important functions in immunity and are associated with autoimmune diseases (Schott and Garcia-Blanco, 2020). Early work by immunologists have shown that DDX39B promoted gene expression of anti-inflammatory pathways (Allcock et al., 2001). Therefore, understanding the genes interactions may help speculate on the proposed AMD mechanisms and immunotherapy.

CONCLUSION

We applied the SMR method on AMD to test the gene-AMD associations based on GWAS summary data and blood eQTL data. From a total of 27,452 AMD cases and 107,818 controls, we obtained 44 SNPs regulating 45 genes significantly associated with AMD. Among the results, HLA class genes have been proved to be associated with immunologically mediated diseases because of the critical role of HLA in mediating the immune response, and genes from MHC Class III are also associated with autoimmune

REFERENCES

- Allcock, R. J., Williams, J. H., and Price, P. (2001). The central MHC gene, BAT1, may encode a protein that down-regulates cytokine production. *Genes Cells* 6, 487–494. doi: 10.1046/j.1365-2443.2001.00435.x
- Claussnitzer, M., Dankel, S. N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., et al. (2015). FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 373, 895–907. doi: 10.1056/nejmoa1502214
- Davey Smith, G., and Ebrahim, S. (2003). ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1–22. doi: 10.1093/ije/dyg070
- Feher, J., Kovacs, I., Artico, M., Cavallotti, C., Papale, A., and Gabrieli, C. B. (2006). Mitochondrial alterations of retinal pigment epithelium in age-related macular degeneration. *Neurobiol. Aging* 27, 983–993. doi: 10.1016/j.neurobiolaging.2005.05.012
- Fernández-Vega, B., García, M., Olivares, L., Álvarez, L., González-Fernández, A., Artime, E., et al. (2020). The association study of lipid metabolism gene polymorphisms with AMD identifies a protective role for APOE-E2 allele in the wet form in a Northern Spanish population. *Acta Ophthalmol.* 98, e282–e291. doi: 10.1111/aos.14280

diseases. These genes may play important roles in causing AMD susceptibility and need to be further verified with experiments. Since AMD has been considered as a genetic disease, from this perspective, it is helpful in understanding the disease from gene-expression level to speculate about the AMD mechanisms and pathology and propose future treatment options for AMD.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YD, NK, and JZ participated in its design, analyzed the data, and wrote the manuscript. All authors read and approved the published version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.726599/full#supplementary-material>

- Fritsche, L. G., Fariss, R. N., Stambolian, D., Abecasis, G. R., Curcio, C. A., and Swaroop, A. (2014). Age-related macular degeneration: genetics and biology coming together. *Annu. Rev. Genomics Hum. Genet.* 15, 151–171. doi: 10.1146/annurev-genom-090413-025610
- Fritsche, L. G., Igl, W., Bailey, J. N. C., Grassmann, F., Sengupta, S., Bragg-Gresham, J. L., et al. (2016). A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* 48, 134–143. doi: 10.1038/ng.3448
- Goverdhan, S. V., Howell, M. W., Mullins, R. F., Osmond, C., Hodgkins, P. R., Self, J., et al. (2005). Association of HLA class I and class II polymorphisms with age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* 46, 1726–1734. doi: 10.1167/iovs.04-0928
- Gu, B. J., Baird, P. N., Vessey, K. A., Skarratt, K. K., Fletcher, E. L., Fuller, S. J., et al. (2013). A rare functional haplotype of the P2RX4 and P2RX7 genes leads to loss of innate phagocytosis and confers increased risk of age-related macular degeneration. *FASEB J.* 27, 1479–1487. doi: 10.1096/fj.12-215368
- Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., et al. (2005). Complement factor H variant increases the risk of age-related macular degeneration. *Science* 308, 419–421. doi: 10.1126/science.1110359
- Hammond, C. J., Webster, A. R., Snieder, H., Bird, A. C., Gilbert, C. E., and Spector, T. D. (2002). Genetic influence on early age-related maculopathy: a twin study. *Ophthalmology* 109, 730–736. doi: 10.1016/s0161-6420(01)01049-1

- Han, X., Ong, J.-S., An, J., Hewitt, A. W., Gharahkhani, P., and MacGregor, S. (2020). Using mendelian randomization to evaluate the causal relationship between serum C-reactive protein levels and age-related macular degeneration. *Eur. J. Epidemiol.* 35, 139–146. doi: 10.1007/s10654-019-00598-z
- Heiba, I. M., Elston, R. C., Klein, B. E., and Klein, R. (1994). Sibling correlations and segregation analysis of age-related maculopathy: the beaver dam eye study. *Genet. Epidemiol.* 11, 51–67. doi: 10.1002/gepi.1370110106
- Inoue, A., and Solon, G. (2010). Two-sample instrumental variables estimators. *Rev. Econ. Stat.* 92, 557–561. doi: 10.1162/rest_a_00011
- Johnson, L. V., Leitner, W. P., Staples, M. K., and Anderson, D. H. (2001). Complement activation and inflammatory processes in drusen formation and age related macular degeneration. *Exp. Eye Res.* 73, 887–896. doi: 10.1006/exer.2001.1094
- Katan, M. B. (2004). Apolipoprotein E isoforms, serum cholesterol, and cancer. *Int. J. Epidemiol.* 33:9. doi: 10.1093/ije/dyh312
- Landowski, M., Kelly, U., Klingeborn, M., Groelle, M., Ding, J.-D., Grigsby, D., et al. (2019). Human complement factor H Y402H polymorphism causes an age-related macular degeneration phenotype and lipoprotein dysregulation in mice. *Proc. Natl. Acad. Sci. U.S.A.* 116, 3703–3711. doi: 10.1073/pnas.1814014116
- Lim, L. S., Mitchell, P., Seddon, J. M., Holz, F. G., and Wong, T. Y. (2012). Age-related macular degeneration. *Lancet* 379, 1728–1738. doi: 10.1016/S0140-6736(12)60282-7
- Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates Inc.
- Mullins, R. F., Russell, S. R., Anderson, D. H., and Hageman, G. S. (2000). Drusen associated with aging and age-related macular degeneration contain proteins common to extracellular deposits associated with atherosclerosis, elastosis, amyloidosis, and dense deposit disease. *FASEB J.* 14, 835–846. doi: 10.1096/fasebj.14.7.835
- Pang, C., Baum, L., Chan, W., Lau, T., Poon, P., and Lam, D. (2000). The apolipoprotein E $\epsilon 4$ allele is unlikely to be a major risk factor of age-related macular degeneration in Chinese. *Ophthalmologica* 214, 289–291. doi: 10.1159/000027506
- Pierce, B. L., and Burgess, S. (2013). Efficient design for mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am. J. Epidemiol.* 178, 1177–1184. doi: 10.1093/aje/kwt084
- Ratnapriya, R., Sosina, O. A., Starostik, M. R., Kwicklis, M., Kapphahn, R. J., Fritsche, L. G., et al. (2019). Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* 51, 606–610. doi: 10.1038/s41588-019-0351-9
- Schott, G., and Garcia-Blanco, M. A. (2020). MHC class III RNA binding proteins and immunity. *RNA Biol.* 18, 640–646. doi: 10.1080/15476286.2020.1860388
- Seddon, J. M., George, S., Rosner, B., and Rifai, N. (2005). Progression of age-related macular degeneration: prospective assessment of C-reactive protein, interleukin 6, and other cardiovascular biomarkers. *Arch. Ophthalmol.* 123, 774–782. doi: 10.1001/archophth.123.6.774
- Smemo, S., Tena, J. J., Kim, K.-H., Gamazon, E. R., Sakabe, N. J., Gómez-Marín, C., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507, 371–375. doi: 10.1038/nature13138
- Stone, E. M., Braun, T. A., Russell, S. R., Kuehn, M. H., Lotery, A. J., Moore, P. A., et al. (2004). Missense variations in the fibulin 5 gene and age-related macular degeneration. *N. Engl. J. Med.* 351, 346–353. doi: 10.1056/nejmoa040833
- Winkler, T. W., Grassmann, F., Brandl, C., Kiel, C., Günther, F., Strunz, T., et al. (2020). Genome-wide association meta-analysis for early age-related macular degeneration highlights novel loci and insights for advanced disease. *BMC Med. Genomics* 13:120. doi: 10.1186/s12920-020-00760-7
- Yan, Q., Ding, Y., Liu, Y., Sun, T., Fritsche, L. G., Clemons, T., et al. (2018). Genome-wide analysis of disease progression in age-related macular degeneration. *Hum. Mol. Genet.* 27, 929–940. doi: 10.1093/hmg/ddy002
- Zhao, T., Hu, Y., and Cheng, L. (2020a). Deep-DRM: a computational method for identifying disease-related metabolites based on graph deep learning approaches. *Brief. Bioinform.* 9:bbaa212. doi: 10.1093/bib/bbaa212
- Zhao, T., Hu, Y., Peng, J., and Cheng, L. (2020b). DeepLGP: a novel deep learning method for prioritizing lncRNA target genes. *Bioinformatics* 36, 4466–4472. doi: 10.1093/bioinformatics/btaa428
- Zhao, T., Hu, Y., Valsdottir, L. R., Zang, T., and Peng, J. (2021a). Identifying drug-target interactions based on graph convolutional network and deep neural network. *Brief. Bioinform.* 22, 2141–2150. doi: 10.1093/bib/bbaa044
- Zhao, T., Hu, Y., Zang, T., and Cheng, L. (2019). Identifying alzheimer's disease-related proteins by LRRGD. *BMC Bioinformatics* 20:570. doi: 10.1186/s12859-019-3124-7
- Zhao, T., Hu, Y., Zang, T., and Cheng, L. (2020c). MRTFB regulates the expression of NOMO1 in colon. *Proc. Natl. Acad. Sci. U.S.A.* 117, 7568–7569. doi: 10.1073/pnas.2000499117
- Zhao, T., Hu, Y., Zang, T., and Wang, Y. (2020d). Identifying protein biomarkers in blood for alzheimer's disease. *Front. Cell Dev. Biol.* 8:472. doi: 10.3389/fcell.2020.00472
- Zhao, T., Lyu, S., Lu, G., Juan, L., Zeng, X., Wei, Z., et al. (2021b). SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Res.* 49, D1413–D1419. doi: 10.1093/nar/gkaa838

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Du, Kong and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.