



## OPEN ACCESS

## EDITED BY

Yuan-Ming Zhang,  
Huazhong Agricultural University, China

## REVIEWED BY

Esther Biswas-Fiss,  
University of Delaware, United States  
Jing-Tian Wang,  
Huazhong Agricultural University, China

## \*CORRESPONDENCE

Wei Chen,  
✉ wei.chen@chp.edu  
Yanming Li,  
✉ yli8@kumc.edu

## SPECIALTY SECTION

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

RECEIVED 21 October 2022

ACCEPTED 17 March 2023

PUBLISHED 30 March 2023

## CITATION

Zhou X, Zhang J, Ding Y, Huang H, Li Y  
and Chen W (2023), Predicting late-stage  
age-related macular degeneration by  
integrating marginally weak SNPs in  
GWA studies.  
*Front. Genet.* 14:1075824.  
doi: 10.3389/fgene.2023.1075824

## COPYRIGHT

© 2023 Zhou, Zhang, Ding, Huang, Li and  
Chen. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Predicting late-stage age-related macular degeneration by integrating marginally weak SNPs in GWA studies

Xueping Zhou<sup>1</sup>, Jipeng Zhang<sup>1</sup>, Ying Ding<sup>1</sup>, Heng Huang<sup>2</sup>,  
Yanming Li<sup>3\*</sup> and Wei Chen<sup>4\*</sup>

<sup>1</sup>Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, United States, <sup>2</sup>Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, United States, <sup>3</sup>Department of Biostatistics and Data Science, University of Kansas Medical Center, Kansas, KS, United States, <sup>4</sup>Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA, United States

**Introduction:** Age-related macular degeneration (AMD) is a progressive neurodegenerative disease and the leading cause of blindness in developed countries. Current genome-wide association studies (GWAS) for late-stage age-related macular degeneration are mainly single-marker-based approaches, which investigate one Single-Nucleotide Polymorphism (SNP) at a time and postpone the integration of inter-marker Linkage-disequilibrium (LD) information in the downstream fine mappings. Recent studies showed that directly incorporating inter-marker connection/correlation into variants detection can help discover novel marginally weak single-nucleotide polymorphisms, which are often missed in conventional genome-wide association studies, and can also help improve disease prediction accuracy.

**Methods:** Single-marker analysis is performed first to detect marginally strong single-nucleotide polymorphisms. Then the whole-genome linkage-disequilibrium spectrum is explored and used to search for high-linkage-disequilibrium connected single-nucleotide polymorphism clusters for each strong single-nucleotide polymorphism detected. Marginally weak single-nucleotide polymorphisms are selected *via* a joint linear discriminant model with the detected single-nucleotide polymorphism clusters. Prediction is made based on the selected strong and weak single-nucleotide polymorphisms.

**Results:** Several previously identified late-stage age-related macular degeneration susceptibility genes, for example, *BTBD16*, *C3*, *CFH*, *CFHR3*, *HTARA1*, are confirmed. Novel genes *DENND1B*, *PLK5*, *ARHGAP45*, and *BAG6* are discovered as marginally weak signals. Overall prediction accuracy of 76.8% and 73.2% was achieved with and without the inclusion of the identified marginally weak signals, respectively.

**Conclusion:** Marginally weak single-nucleotide polymorphisms, detected from integrating inter-marker linkage-disequilibrium information, may have strong predictive effects on age-related macular degeneration. Detecting and integrating such marginally weak signals can help with a better understanding of the underlying disease-development mechanisms for age-related macular degeneration and more accurate prognostics.

## KEYWORDS

age-related macular degeneration, biomarker discovery, feature selection, genome-wide association study, linkage disequilibrium, prediction, weak signal

## 1 Introduction

AMD is a progressive neurodegenerative disease characterized by reduced retinal pigment epithelium (RPE) function and loss of photoreceptors in the macula (Fritsche et al., 2016). By 2040, there will likely be 288 million AMD patients worldwide (Wong et al., 2014). Between 6 and 9 percent of blindness worldwide is caused by AMD, which is also the leading cause of blindness in developed countries (Wong et al., 2014; Jonas et al., 2017; Keenan et al., 2021). AMD is divided into early, intermediate, and late stages. In the first two stages, the visual symptoms are typically absent or mild. Late-stage AMD involves two forms: geographic atrophy (GA) and choroidal neovascularization (CNV), though the two manifestations are not mutually exclusive (Keenan et al., 2021; Ung et al., 2021). Either form of late AMD can lead to vision distortion and/or vision loss (Wong et al., 2014; Keenan et al., 2021). Until now, understanding of the disease biology and therapeutic options are limited (Fritsche et al., 2014; Fritsche et al., 2016). Age, genetics, and environmental factors are all expected to cause the disease in a complex interactive way (Keenan et al., 2021). Previous research has identified susceptibility SNPs and genes for the development of late-stage AMD (Fritsche et al., 2013; Kvale et al., 2015; Fritsche et al., 2016). However, the current identified genetic biomarkers only explained a small portion of the disease variability. It would be extremely helpful to identify novel biomarkers pertinent to late-stage AMD for a better understanding of the disease etiology and for developing new tools to aid accurate prognostic and precision treatment.

A recent landmark genome-wide association study (GWAS) has identified 52 independently associated variants across 34 susceptibility loci for AMD by comparing the late-stage AMD patients and unaffected controls of European ancestry (Fritsche et al., 2016). Even though GWAS is a very effective method for identifying SNPs associated with diseases, the majority of GWAS research selects significant biomarkers based on single marker approaches, where each SNP is evaluated separately for its marginal correlation with the outcome. As a result, the correlation spectrum between SNPs across the genome is ignored, and only SNPs with strong marginal effects can be selected. In Fritsche et al. (2016), the 52 identified variants explained only 27.2% of the disease variability. Recent studies have shown that incorporating inter-feature correlations/connections can help identify novel genetic biomarkers and improve outcome prediction accuracy (Li, 2019; Li et al., 2019a; Li et al., 2019b). This suggests that we might be able to discover new AMD variants and explain more disease variability by exploring and integrating the genome-wide correlation/connectivity patterns into the GWAS. Following Li (2019); Li et al. (2019b), we term the variants that cannot be directly detected but would exhibit significant differentiating effects after taking into consideration of their correlations/connections to other genome-wide significant SNPs as the marginally weak SNPs, or weak SNPs. In this paper, we aim to detect weak SNPs for late-stage AMD, investigate their

biological function, and confirm whether they can help to improve disease prediction.

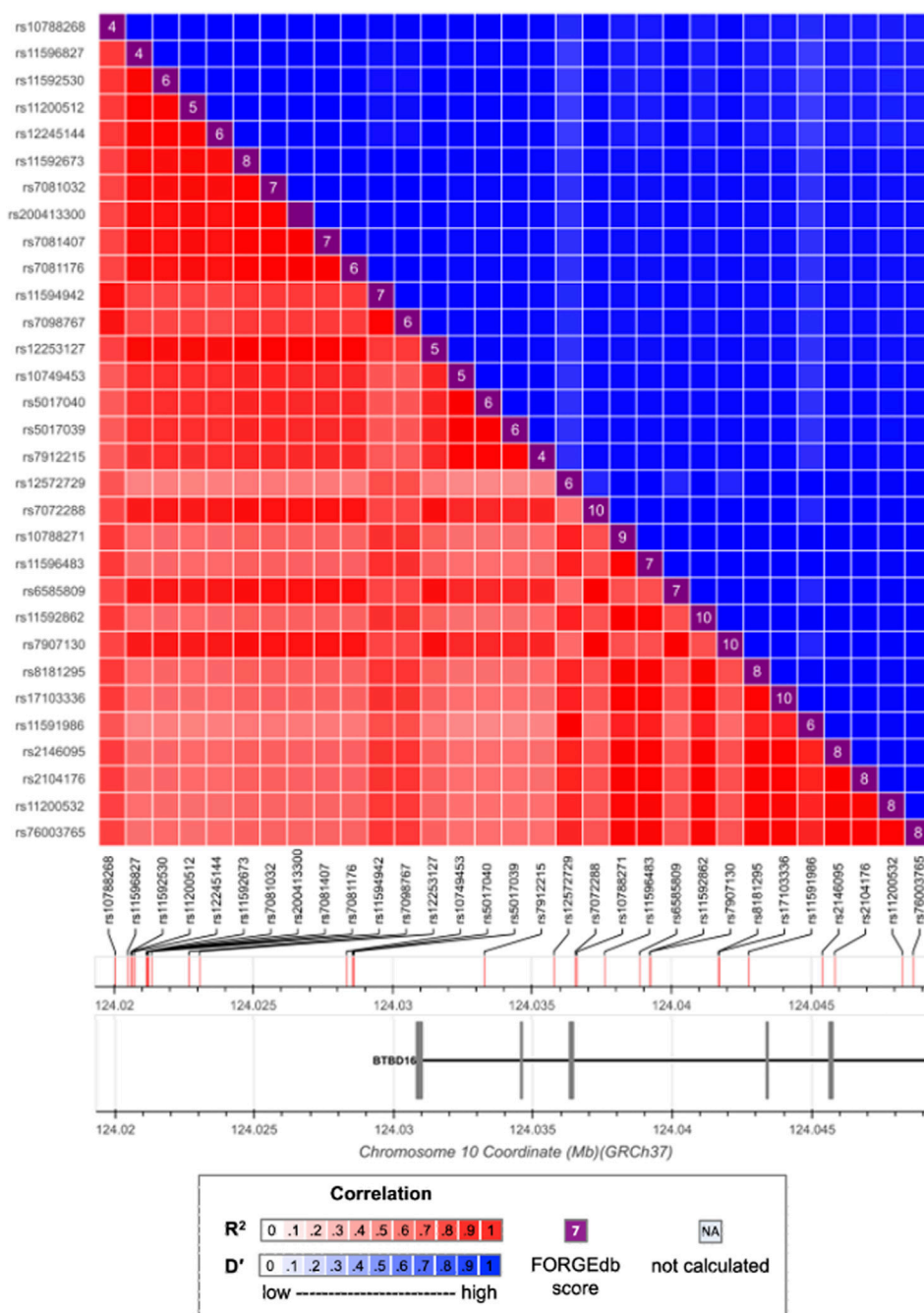
We use linkage disequilibrium (LD) as the metric to measure the inter-SNP correlations. In genetics, “LD” refers to the non-random correlations between alleles at different loci (Slatkin, 2008). It has been used to determine correlated disease-related SNP sets and is a crucial metric for determining inter-SNP interaction (Carlson et al., 2004). The squared correlation, or  $r^2$  statistic, which spans from 0 to 1, is one of the most widely used LD metrics. A higher  $r^2$  value indicates a stronger correlation between two SNPs. In this study, we use  $r^2$  to explore the whole-genome LD spectrum. For each genomewide significant SNP identified from a GWA scan, we detect its high-LD connected SNP cluster using connected component searching algorithms (Li et al., 2019b; Li, 2019), such that any pair of SNPs within the same cluster are connected through a path of high LD links. A marginally weak SNP can be then selected based on not only its own marginally differentiating effect but also how many other SNPs within the cluster it connects to and the disease-differentiating capabilities of those connected SNPs (Li et al., 2019b).

## 2 Materials and methods

### 2.1 Data

International Age-Related Macular Degeneration Genomics Consortium (IAMDGCC) has collected genetic data from AMD patients and control subjects across 26 different studies. IAMDGCC dataset is by far the largest-scaled AMD GWAS dataset (Fritsche et al., 2016). In this study, we used a total of 27,301 unrelated individuals of European ancestry in the IAMDGCC dataset, including 14,590 late-stage AMD patients with GA and/or CNV and 12,711 normal controls. Genotypes and phenotypes for these 27,301 samples were downloaded from NCBI dbGap database.

Our analysis included a total of 5,002,684 common SNPs from chromosomes 1 through 22 with minor allele frequency (MAF) larger than 0.1. Each SNP is coded by its number of minor alleles (0, 1, or 2). It is known from previous studies that this sample population does not contain an obvious heterogeneous structure. We also carried out a principal component analysis (PCA). The PCA results showed no obvious population structures. We calculated the kinship matrix using 100,000 randomly sampled SNPs across the whole genome. The absolute values of the KING-robust estimators on 100,000 randomly chosen sub-principal matrices were very small (mean = 0.0077, standard deviation = 0.0081). All the above evidence suggests that the samples used in our study are independent. Following previous studies using the same data cohort (Fritsche et al., 2016), we did not adjust for PC scores in our study. The missing genotype values were imputed by the means of the non-missing ones. There are 2.1% of the participants older than 90 years. We capped their ages at 90 in the analysis. There are 0.4% of the samples with missing ages and their ages were imputed by the average age of non-missing subjects.



**FIGURE 1**  
 LD pattern between strong and weak SNPs. Marginally strong SNPs: rs57581057, rs10788268, rs11594942, rs7098767, rs12572729, rs10788271, rs11596483, rs8181295, rs17103336, rs11591986, rs2146095, rs2104176, rs11200532, rs76003765. Marginally weak SNP: rs7907130.

### 3 Methods

#### 3.1 Notation

Denote the observed data matrix by  $X_{n \times p}$ , where  $n$  is the number of subjects and  $p$  is the total number of features. We use  $X_j$  to denote the length- $n$  sample vector for feature  $j, j = 1, \dots, p$ . Let  $y$  denotes the length- $n$  vector of outcome class labels, and  $y_i$  denotes the class indicator for

subject  $i, i = 1, \dots, n$ . We use  $n_k$  to denote the size of class  $k, k = 0, 1$ , where  $k = 0$  for normal controls and  $k = 1$  for late-stage AMD patients.

#### 3.2 Detection of marginally strong SNPs

Marginal Student t-tests were used to select marginally strong signals. Specifically, for each SNP  $j, j = 1, \dots, p$ , we calculated

$$\hat{T}_j = (\bar{X}_j^{[1]} - \bar{X}_j^{[0]}) / s_{\Delta}, \tag{1}$$

where  $\bar{X}_j^{[1]} = \frac{\sum_{v_i=1} X_{ij}}{n_1}$ ,  $\bar{X}_j^{[0]} = \frac{\sum_{v_i=0} X_{ij}}{n_0}$  and  $s_{\Delta} = \sqrt{\sum \sum (X_{i,j} - \bar{X}_j^{[k]})^2 / (n_0 + n_1)}$ . SNPs with  $p$ -values less than  $5 \times 10^{-8}$  were selected as the marginally strong SNPs.

### 3.3 Detection of high-LD connected SNP clusters

We used PLINK 1.9 (Purcell et al., 2007), a tool set for whole-genome association and population-based linkage analyses, to calculate  $r^2$  values for all SNPs on each chromosome. We then thresholded  $r^2$  values at 0.7 so that only SNPs in high LD were considered connected and included in the downstream analyses. For each marginally strong SNP, we employed the connected component searching algorithms in Li (2019); Li et al. (2019b) to detect the cluster of SNPs each of which is connected to the strong SNP through a path of high-LD links. Suppose there are  $B$  such clusters detected, we denote those clusters by  $C_b$ , where  $l = 1, \dots, B$ .

### 3.4 Detection of marginally weak SNPs

LD-adjusted effect sizes were then calculated for SNPs in the detected connected components. Specifically, for each  $C_b$ ,  $l = 1, \dots, B$ , we calculated the multivariate  $t$ -statistic by  $\hat{T}_{C_l}^{ld} = \hat{\Omega}_{C_l} (\bar{X}_j^{[1]} - \bar{X}_j^{[0]})$ , where  $\hat{\Omega}_{C_l} = \hat{\Sigma}_{C_l}^{-1}$  is the inverse of the LD matrix  $\hat{\Sigma}_{C_l}$  on component  $C_l$ .  $\hat{\Omega}_{C_l}$  characterizes the LD connections between SNPs in  $C_l$ . For each SNP  $j$  in  $C_b$ , its LD-adjusted differentiating statistic is the corresponding entry in  $\hat{T}_{C_l}^{ld}$

$$\hat{T}_j^{ld} = \sum_{j' \in C_l} [\hat{\Omega}_{C_l}]_{jj'} (\bar{X}_j^{[1]} - \bar{X}_j^{[0]}). \tag{2}$$

The greater the magnitude of  $\hat{T}_j^{ld}$  is, the higher power SNP  $j$  has in differentiating the diseases and controls after integrating the LD information with its connected SNPs. Marginally weak SNPs were selected based on  $|\hat{T}_j^{ld}|$ . It should be noted that the calculation of the connection matrix  $\Omega$  is essential for marginally weak signal detection (Li et al., 2019b). For an ultra-high dimensional association study like a GWAS, it is impossible to directly calculate  $\hat{\Omega}$  for all SNPs. By leveraging the marginal LD information, we were able to break it down to locally connected components  $C_b$ s, which were of much smaller sizes, thus making the calculation feasible. Not every SNP within an LD-connected cluster would necessarily be informative for classification. Only the top 1,000 SNPs with the greatest  $|\hat{T}_j^{ld}|$  values, aside of the strong SNPs, were selected as the marginally weak SNPs. The optimal hyperparameter value of 1,000 was chosen by a cross-validation procedure.

### 3.5 Calculating $p$ -values for selected SNPs

To evaluate the statistical significance of the selected SNPs, a post-selection non-parametric permutation test was performed. We first constructed a permutation dataset by re-sampling (with

replacement) the class labels but keeping the genotype matrix unchanged. We kept the detected strong SNP set and the connected components unchanged to preserve the connection patterns. We then recalculated the statistics  $T_j^{ld}$  using the original  $\hat{\Omega}_{C_l}$ , and  $\bar{X}_j^{[1]}$  and  $\bar{X}_j^{[0]}$  recalculated from the permuted dataset. An empirical null distribution of  $\hat{T}_j^{cl, (m)}$ ,  $m = 1, \dots, M$ , was then generated by repeating the process  $M = 120,000$  times. The percentage of the occurrences that the permuted statistics  $|\hat{T}_j^{cl, (m)}|$  were equal to or more extreme than the observed statistics  $|\hat{T}_j^{ld}|$  (evaluated from the original non-permuted dataset) was defined to be the permutation test  $p$ -value for each selected SNP  $j$ . Last, to further control for the family-wise error rate, a Bonferroni correction was made for all selected SNPs.

### 3.6 Functional enrichment analysis

All SNPs in the analysis were first mapped onto genes in the Human Genome GRCh37/hg19 assembly using the R package “biomaRt” (Durinck et al., 2009). A SNP not in any gene region was mapped to its closest nearby gene containing it within its 2 kb upstream/downstream ranges. Functional enrichment analysis was then performed to validate the association of the selected SNPs with AMD from a system biology perspective. Ingenuity Pathway Analysis (IPA, QIAGEN, Inc.) was used to identify the significantly enriched canonical pathways. Fisher’s exact test  $p$ -values were provided for each canonical pathway. A pathway with a  $p$ -value less than 0.01 was considered a significantly enriched pathway.

For each significantly enriched pathway, we reported its leading-edge genes. Leading edge genes can be viewed as the pivot genes in the pathway because of which the pathway got enriched (Subramanian et al., 2005). We evaluated how the leading-edge genes of the enriched pathways overlapped with selected genes. A greater overlap indicates more compelling evidence of the biological functions of the selected genes/SNPs on AMD.

### 3.7 Late-stage AMD prediction

We fitted a logistic regression with a Lasso penalty and the selected marginally strong and weak SNPs for outcome prediction. We included age and gender as adjusting covariate variables.

The whole analysis pipeline is summarized in Algorithm 1.

- 1: **Marginally strong SNPs:** From marginal  $t$ -tests.
- 2: **LD connected SNP clusters:** For each strong SNP, detecting its high-LD connected SNP clustered by searching for its connected component in the thresholded LD matrix.
- 3: **Marginally weak SNPs:** Detected by ranking the LD-adjusted statistics  $|\hat{T}_{C_l}^{ld}| = |\hat{\Omega}_{C_l} (\bar{X}_j^{[1]} - \bar{X}_j^{[0]})|$ .
- 4: **Functional enrichment analysis.**
- 5: **Late-stage AMD prediction:** Prediction using Lasso-logistic regression with selected SNPs and adjusting covariates.

**Algorithm 1** Procedure for detecting strong and weak signals and late-stage AMD classification.

TABLE 1 Permutation *p*-values of selected marginally strong SNPs and their mapped genes.

SNP	Chr:pos	Gene	<i>p</i> -value	AMD-related literature for gene
rs6695321	1:196675861	<i>CFH</i>	$< 8.805 \times 10^{-6}$	Guenther et al. (2020), Tian et al. (2012) <sup>†</sup>
rs7524776	1:196623337	<i>CFH</i>	$< 8.805 \times 10^{-6}$	Budzinskaia et al. (2011) <sup>†</sup>
rs201084593	1:197316970	<i>CRB1</i>	$< 8.805 \times 10^{-6}$	Fletcher et al. (2014)
rs4915551	1:197508901	<i>DENND1B</i>	$< 8.805 \times 10^{-6}$	
rs10429910	1:197141253	<i>ZBTB41</i>	$< 8.805 \times 10^{-6}$	Zhang et al. (2008)
rs4865172	4:57907587	<i>IGFBP7</i>	$< 8.805 \times 10^{-6}$	Arakawa et al. (2011)
rs3132451	6:31582025	<i>AIF1</i>	$< 8.805 \times 10^{-6}$	Wolf et al. (2022)
rs201943238	6:32314956	<i>TSBP1, TSBP1-AS1</i>	$< 8.805 \times 10^{-6}$	Strunz (2021)
rs1043618	6:31783507	<i>HSPA1A</i>	$< 8.805 \times 10^{-6}$	Liang et al. (2022)
rs444921	6:31932177	<i>SKIV2L</i>	$< 8.805 \times 10^{-6}$	Kopplin et al. (2010)
rs406658	6:31996524	<i>C4B</i>	$< 8.805 \times 10^{-6}$	Grassmann et al. (2016)
rs6906021	6:32626311	<i>HLA-DQB1-AS1</i>	$< 8.805 \times 10^{-6}$	Jorgenson et al. (2016)
rs5788544	10:124091856	<i>BTBD16</i>	$< 8.805 \times 10^{-6}$	Strunz et al. (2020a)
rs2300431	10:124242817	<i>HTRA1</i>	$< 8.805 \times 10^{-6}$	Gibbs et al. (2008) <sup>†</sup>
rs7093894	10:124234880	<i>HTRA1</i>	$< 8.805 \times 10^{-6}$	Goto et al. (2009) <sup>†</sup>
rs2239586	10:124249235	<i>HTRA1</i>	$< 8.805 \times 10^{-6}$	Tam et al. (2008) <sup>†</sup>
rs6857	19:45392254	<i>NECTIN2/PVRL2</i>	$< 8.805 \times 10^{-6}$	Holliday et al. (2013) <sup>†</sup>
rs265285	19:1534676	<i>PLK5</i>	$< 8.805 \times 10^{-6}$	
rs265284	19:1534685	<i>PLK5</i>	$< 8.805 \times 10^{-6}$	
rs2074453	19:1080189	<i>ARHGAP45/HMHA1</i>	$< 8.805 \times 10^{-6}$	

● Chr:pos—chromosome and position.

● †—the SNP is also previously reported in the study.

## 4 Results

### 4.1 Selected candidate SNPs and associate genes

Variants selection was conducted on each of the 22 autosomes separately. There were 4,679 genome-wide significant SNPs detected as the marginally strong SNPs, of which 26 had been reported by Fritsche et al. (2016). We further detected 3,446 high-LD connected SNPs, each of which had an  $r^2 > 0.7$  with at least one of the marginally strong SNPs. Those 3,446 SNPs were clustered within 410 connected components. Figure 1 shows an example connected component of 32 SNPs, out of which 14 were marginally strong SNPs and 1 was eventually selected as marginally weak SNPs. After the final filtering, the top 1,000 SNPs out of those 3,446 with the greatest  $|T_{C_i}^{LD}|$  values were selected as the marginally weak SNPs. Altogether, there were 5,679 SNPs (4,679 marginally strong and 1,000 marginally weak) selected as the informative SNPs. Those 5,679 SNPs were then mapped to 197 gene regions, with 99 genes mapped by marginally strong SNPs only, 35 genes mapped by marginally weak SNPs only, and 63 genes mapped by both marginally strong and weak SNPs (Supplementary Table S1).

In Figure 1, the marginally weak SNP rs7907130 has a marginal  $p$ -value =  $1.124 \times 10^{-4}$ . As a result, it is likely to be missed from

downstream fine mappings in a conventional GWAS. However, Once the LD information is integrated, SNP rs7907130 has the most significant permutation test  $p$ -value among all the 32 SNPs. Therefore, its predictive effect is manifested through integrating its LD connections with the strong SNPs. Moreover, SNP rs7907130 was mapped to a previously identified gene *BTBD16* (Strunz et al., 2020a) together with the other strong SNPs. It could potentially be a new AMD susceptibility locus.

### 4.2 Statistically significance for selected SNPs and genes

Post-selection inference was made for the 5,679 SNPs selected. A Bonferroni correction gave a family-wise significance level of  $8.805 \times 10^{-6}$  (= 0.05/5,679). Based on the permutation test results, 55 out of the 4,679 selected marginally strong SNPs (Table 1 and the full list in Supplementary Table 4) and 13 out of the 1,000 selected marginally weak SNPs (Table 2) were statistically significant with a permutation  $p$ -value  $< 8.805 \times 10^{-6}$ . Table 1 gives the detailed test results for the top marginally strong SNPs and their mapped genes. Table 2 gives the detailed test results for the significant marginally-weak SNPs and their mapped genes.

TABLE 2 Permutation *p*-values of selected marginally weak SNPs and their mapped genes.

SNP	<i>p</i> -value	Chr:pos	Gene	Gege literature	Strong SNP in high LD
rs10489456	$< 8.805 \times 10^{-6}$	1:196687515	<i>CFH</i>	Ennis et al. (2007) <sup>†</sup>	rs388419 (1:196747669) rs449657 (1:196747834) rs201334308 (1:196747111)
rs72468003	$< 8.805 \times 10^{-6}$	1:196883372	<i>CFHR4</i>	Kubista et al. (2011)	rs114092527 (1:196894694)
rs10442656	$< 8.805 \times 10^{-6}$	1:197606708	<i>DENND1B</i>		rs4915551 (1:197508901)
rs13150593	$< 8.805 \times 10^{-6}$	4:57810321	None		rs4865172 (4:57907587)
rs17197637	$7.311 \times 10^{-6}$	6:31210502	None		rs112340183 (6:31228410)
rs2844603	$< 8.805 \times 10^{-6}$	6:31250854	None		rs112340183 (6:31228410)
rs2853931	$< 8.805 \times 10^{-6}$	6:31255007	None		rs112340183 (6:31228410)
rs2242655	$< 8.805 \times 10^{-6}$	6:31627449	<i>C6orf47</i> , <i>BAG6</i>		rs112340183 (6:31228410)
rs805304	$< 8.805 \times 10^{-6}$	6:31698088	<i>DDAH2</i>	Lange et al. (2016)	rs1043618 (6:31783507)
			<i>CLIC1</i>	Chuang et al. (2022)	rs1043618 (6:31783507)
rs72887130	$< 8.805 \times 10^{-6}$	6:50902701	None		rs3857599 (6:50938247)
rs16880854	$< 8.805 \times 10^{-6}$	6:50904881	None		rs3857599 (6:50938247)
rs3857596	$< 8.805 \times 10^{-6}$	6:50905067	None		rs3857599 (6:50938247)
rs7907130	$< 8.805 \times 10^{-6}$	10:124039235	<i>BTBD16</i>	Strunz et al. (2020a)	rs7098767 (10:124022694)
rs9747347	$< 8.805 \times 10^{-6}$	17:79606820	<i>TSPAN10</i>	Strunz et al. (2020b)	rs6565597 (17:79526821)

● Chr:pos—chromosome and position.

● †—the SNP is also previously reported in the study.

The majority of the detected SNPs and their mapped genes had been reported to be associated with AMD previously. However, five new genes have been identified in our study as the potentially novel susceptibility genes for late-stage AMD: *DENND1B*, *PLK5*, *ARHGAP45/HMHA1*, *C6orf47*, and *BAG6*.

*DENND1B* (differentially expressed in normal versus neoplastic domain containing 1B) is mapped by one marginally strong and one marginally weak SNP. Chronic inflammatory illnesses like asthma (Sleiman et al., 2010), inflammatory bowel disease (Lees et al., 2011), and primary biliary cirrhosis (Mells et al., 2011) have been reported associated with this gene. The disease pathogenesis involves the *DENND1B* encoded protein and its interaction with tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ) (Sleiman et al., 2010). Although *DENND1B* has not been directly linked to AMD, it may play a similar role in AMD pathogenesis through inflammation related to TNF- $\alpha$ , which is a pro-inflammatory cytokine contributing to inflammation-associated angiogenesis and CNV (Jasielska et al., 2010; Khan et al., 2021). *ARHGAP45* (rho GTPase activating protein 45) or *HMHA1* (minor histocompatibility antigen 1) is mapped by one marginally strong SNP. It controls endothelial integrity, whereas endothelial barrier dysfunction can lead to uncontrolled leaks or edema (Amado-Azevedo et al., 2018). In CNV, aberrant blood vessels leak fluid or blood into the retina. *ARHGAP45* may contribute to AMD by regulating the endothelial barrier function of blood vessels. *PLK5* (polo-like kinase 5) is associated with two marginally strong SNPs. It can induce genotoxic stress and DNA damage (Andrysik et al., 2010), which is one of the underlying mechanisms of AMD. Additionally, *PLK5* is specifically expressed in the

eye (de Cárcer et al., 2011) and is a candidate gene modulating the proper development and function of mice eyes (Chiang et al., 2020).

*BAG6* (BAG Cochaperone 6) and *C6orf47* (chromosome 6 open reading frame 47) are mapped by one significant marginally weak SNP. *BAG6* gene is found to be a component of a cluster of immune-relevant genes located within the human major histocompatibility complex class (MHC) III region (Banerji et al., 1990). *BAG6* encodes a multifunctional protein. For instance, the protein is required for the acetylation of p53 in response to DNA damage (Binici and Koch, 2014), which is believed to be involved in the RPE cell death in AMD (Bhattacharya et al., 2012; Vuong et al., 2012). *C6orf47* gene has been reported related to inflammatory diseases (Harney et al., 2008; Goudey et al., 2017). However, its biological functions remain unclear (Ren et al., 2015) and require further investigation.

### 4.3 Functional enrichment analyses

We further conducted two functional enrichment analyses to verify the biological functions of the selected genes, especially the 35 genes mapped from marginally weak SNPs. We examined two candidate gene lists. One list consists of all the 197 genes mapped by both marginally strong and marginally weak SNPs. The other list consists of 162 genes mapped from marginally strong SNPs only. IPA software was used for the enrichment analyses. Any differences between the results from the two enrichment analyses should be due

**TABLE 3** Enriched canonical pathways from Ingenuity Pathway Analysis (IPA).

Ingenuity canonical pathway	Ratio	Pathway genes	p-value
Complement System	0.162	<i>C2, C3, C4A/C4B, CFD, CFH, CFI</i>	$2.63 \times 10^{-7}$
LXR/RXR Activation	0.073	<i>ABCA1, APOC1, APOE, APOM, C3, C4A/C4B, CETP</i> <i>MMP9, VTN</i>	$3.09 \times 10^{-7}$
Antigen Presentation Pathway	0.154	<i>HLA-B, HLA-DQA1, HLA-DQB1</i> <i>HLA-A, HLA-C, HLA-DRA</i>	$3.72 \times 10^{-7}$
Neuroprotective Role of <i>THO1</i> in Alzheimer's Disease	0.067	<i>CFD, HLA-A, HTRA1, PRSS57, PRTN3</i> <i>HLA-B, HLA-C, MMP9</i>	$2.88 \times 10^{-6}$
FXR/RXR Activation	0.064	<i>APOC1, APOE, APOM, C3, C4A/C4B, CETP, LIPC, VTN</i>	$4.07 \times 10^{-6}$
PD-1, PD-L1 cancer immunotherapy*	0.066	<i>HLA-B, HLA-DQA1, HLA-DQB1</i> <i>CSNK2B, HLA-A, HLA-C, HLA-DRA</i>	$1.29 \times 10^{-5}$
B Cell Development	0.114	<i>HLA-B, HLA-DQA1, HLA-DQB1</i> <i>HLA-A, HLA-DRA</i>	$1.70 \times 10^{-5}$
Neuroinflammation Signaling Pathway*	0.035	<i>AGER, GRIN3B, HLA-B, HLA-DQA1, HLA-DQB1, PLA2G12A</i> <i>HLA-A, HLA-C, HLA-DRA, MMP9, TGFBR1</i>	$2.24 \times 10^{-5}$
Crosstalk between Dendritic Cells and Natural Killer Cells*	0.066	<i>HLA-B, MICB, NECTIN2</i> <i>HLA-A, HLA-C, HLA-DRA</i>	$5.50 \times 10^{-5}$
Th2 Pathway	0.051	<i>HLA-B, HLA-DQA1, HLA-DQB1, NOTCH4</i> <i>HLA-A, HLA-DRA, TGFBR1</i>	$6.76 \times 10^{-5}$
Th1 and Th2 Activation Pathway*	0.041	<i>HLA-B, HLA-DQA1, HLA-DQB1, NOTCH4</i> <i>HLA-A, HLA-DRA, TGFBR1</i>	$2.75 \times 10^{-4}$
Th1 Pathway	0.049	<i>HLA-B, HLA-DQA1, HLA-DQB1, NOTCH4</i> <i>HLA-A, HLA-DRA</i>	$2.75 \times 10^{-4}$
IL-4 Signaling*	0.054	<i>HLA-B, HLA-DQA1, HLA-DQB1</i> <i>HLA-A, HLA-DRA</i>	$6.03 \times 10^{-4}$
Caveolar-mediated Endocytosis Signaling	0.053	<i>FLOT1, HLA-B, HLA-A, HLA-C</i>	$2.24 \times 10^{-3}$
Atherosclerosis Signaling	0.038	<i>APOC1, APOE, APOM, PLA2G12A, MMP9</i>	$2.75 \times 10^{-3}$
Glucocorticoid Receptor Signaling*	0.019	<i>HLA-B, HLA-DQA1, HLA-DQB1, NDUFS7, PLA2G12A</i> <i>HLA-A, HLA-C, HLA-DRA, MMP9, POLR2E, TGFBR1</i>	$3.72 \times 10^{-3}$
Role of <i>OCT4</i> in Mammalian Embryonic Stem Cell Pluripotency*	0.065	<i>CASP6, POU5F1, REST</i>	$4.57 \times 10^{-3}$
Phagosome Maturation*	0.031	<i>HLA-B, HLA-A, HLA-C, HLA-DRA, TUBB</i>	$6.31 \times 10^{-3}$

- Pathways with \* are enriched only when both marginally strong and weak genes were included.
- Ratio = the proportion of candidate genes in the enriched pathway.
- Genes in bold typeface are marginally weak genes.

to the marginally weak genes and can help reveal the biological significance of these weak genes. Table 3 summarizes all enriched pathways identified by the two gene lists with a *p*-value of < 0.01. Nine canonical pathways were significantly enriched by both candidate gene lists. Another nine canonical pathways were significantly enriched only when marginally weak genes were included.

All nine canonical pathways enriched by both gene lists had been reported in the AMD literature, including the complement system (Anderson et al., 2010; Ishikawa et al., 2016; Potilinski et al., 2021), the antigen presentation (Ambati et al., 2013; Potilinski et al., 2021), the LXR/RXR activation (Mukwaya et al., 2018; Choudhary et al., 2020) and the FXR/RXR activation (Kurji et al., 2010) pathways. These results agree with the current knowledge that AMD is

associated with inflammatory response, which requires the accumulation of various immune cells and cytokines in the eye (Tan et al., 2020). Additionally, leading-edge gene *THO1* in the Alzheimer's disease pathway, which is significantly enriched, is known to have a neuroprotective role. The elevated expression of thimet oligopeptidase (*THO1*) is a neuroprotective response to amyloid- $\beta$  ( $A\beta$ ), a major component of Alzheimer's disease plaques that may cause increased neuronal death (Pollio et al., 2008). This is consistent with the fact that  $A\beta$  also presents in the drusen of AMD patients (Pollio et al., 2008). Multiple studies indicated that atherosclerosis is a risk factor for AMD (Vingerling et al., 1995; van Leeuwen et al., 2003; Logue et al., 2014). The atherosclerosis signaling pathway was significantly enriched in our study and other studies (Fritsche et al., 2013; Zhang and Zhou, 2020). This supported the hypothesis that AMD and atherosclerosis may share common pathogenic pathways (Chau et al., 2008).

For the list with marginally weak genes, nine more canonical pathways were significantly enriched, including those involved in immune responses and inflammatory activities, in crosstalk between dendritic cells, Th1 and Th2 activation, and natural killer cells IL-4 Signaling. The neuroinflammation signaling pathway became significantly enriched by adding the marginally weak genes. Leading-edge genes in this pathway include *HLA-A*, *HLA-C*, *HLA-DRA*, *MMP9*, and *TGFBR1*. This finding is in line with earlier research that suggested neuroinflammation is involved in late-stage AMD (Buschini et al., 2011; Racic et al., 2021). By incorporating the weak genes *CSNK2B*, *HLA-A*, *HLA-C*, and *HLA-DRA*, the PD-1, PD-L1 cancer immunotherapy pathway was significantly enriched. This may be due to the fact that programmed death receptor-1 (PD-1) and its ligand PD-L1 play critical roles in the immune response and autoimmunity beyond cancer (Qin et al., 2019), including preserving the retinal immunosuppressive micro-environment (Chen et al., 2009). Marginally weak genes *HLA-A*, *HLA-C*, *HLA-DRA*, *MMP9*, *POLR2E*, *TGFBR1* were responsible for the enrichment of the glucocorticoid receptor signaling pathway. Glucocorticoid receptor is associated with anti-inflammatory properties and may play a crucial role in late-stage AMD progression (Jin et al., 2017). A synthetic glucocorticoid called triamcinolone is used to treat people with CNV-type AMD, according to Becerra et al. (2011). With marginally weak gene *REST*, the Role of *OCT4* in mammalian embryonic stem cell pluripotency pathway can be validated. *OCT4*, also known as *Pou5f1*, is important for maintaining and regaining stem cell pluripotency. According to a recent study, human embryonic stem cells (hESCs) can encourage the retro-differentiation of Müller cells in the retina into retinal progenitor cells by regulating the expression of *OCT4* in these cells (Ke et al., 2021). This also indicates that stem cells may provide a new option for treating AMD (Parameswaran and Krishnakumar, 2017; Ke et al., 2021). The enriched phagosome maturation pathway is in concordance with the facts that efficient RPE phagosome maturation is crucial for retinal health and homeostasis, and that the delay in maturation is related to the accumulation of debris in the RPE and AMD (Gordiyenko et al., 2010; Wavre-Shapton et al., 2013; Kwon and Freeman, 2020).

We also conducted a gene set enrichment analysis (GSEA) using the curated gene sets in Molecular Signature Database (MSigDB v7.5.1 c2). Multiple gene sets linked to AMD-related phenotype changes or biological processes were enriched. They include extracellular matrix organization, complement system, lipid

metabolism, immune, and inflammation responses (Supplementary Table S6). Marginally weak genes (*REEP6*, *HLA-A*, *TGFBR1*, *POLDIP2*, *MMP9*, and *VTN*) are confirmed as leading edge genes within the AMD-related gene sets, including "GOBP: activation of the immune response," "GOCC: photoreceptor inner segment," "HP: choroidal neovascularization," "HP: macular degeneration" among others. These serve to further illustrate the role that marginally weak genes have significant co-regulatory effects, even no significant differentiating effects, in the onset and course of the disease. Identification of these weak signals could provide new insights into a better understanding of the underlying disease mechanisms.

#### 4.4 Prediction model based on selected SNPs

For prediction, we applied a five-fold cross-validation, where the whole dataset was partitioned into five folds of sub-datasets with similar sample sizes. In each round of cross-validation, we aggregated four folds of data into a training set for model training. We used the remaining fold as the test set for AMD status prediction. Through the five rounds of cross-validations, each sample in the original dataset would have been predicted once. Prediction performance was evaluated by comparing the predicted labels to the observed ones.

Our method, which takes into account both weak and strong signals, has a prediction accuracy of 0.769 with an area under the receiver operating characteristic curve (AUC) of 0.768. With only the 4,679 marginally strong SNPs, it achieves a prediction accuracy of 0.732 and an AUC of 0.730. This demonstrated that marginally weak signals could help to improve the prediction performance.

## 5 Discussion

Our study employed a novel machine-learning method to detect marginally weak variants in a GWAS. By exploring and integrating the genome-wide LD spectrum into GWAS, we were able to detect novel susceptibility genes for AMD and explain more disease variability. The previously identified genes like *BTBD16*, *C3*, *CFH*, *CFHR3*, *HTARA1* were confirmed in our study. Novel AMD-related genes *DENND1B*, *PLK5*, *ARHGAP45*, and *BAG6* were detected as weak signals. More AMD-related biological pathways were enriched when marginally weak signals were considered.

For the effect size estimates of the selected SNPs, compared to the marginal approaches, the estimates from our approach would be closer to their oracle estimates, which are the estimates from fitting a joint model assuming one knew prior to the analysis the subset of true causal SNPs (Li et al., 2019b). It is possible that a marginally strong SNP turns out to have a weak joint effect. Such SNPs would be false positive findings in conventional GWA studies. Our approach diminishes such cases.

In our approach, each weak signal detected must have a high-LD path linking it to some marginally strong SNPs. It is cognizant that other than linking to strong signals, some weak signals work together by themselves as a group to exude their overall predictive effect on the outcomes, such as a group of rare variants (Li and Leal, 2008). Our approach would not detect such groups of weak signals. It is



promising to combine our approach with the rare variant grouping methods in GWA studies. We will investigate this in future studies.

Our weak variants detection method can be directly applied to GWA studies with other types of trait variables: continuous, ordinal, counting, and survival outcomes. We envision its wide range of applications.

## Data availability statement

Publicly available datasets were analyzed in this study. Genotypes and phenotypes for the samples analyzed in this study can be accessed at [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001039.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001039.v1.p1). R codes for the analysis pipeline can be downloaded at [https://github.com/XuepingZhou/mLDA\\_Pipeline](https://github.com/XuepingZhou/mLDA_Pipeline).

## Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

XZ, YD, YL, WC conceived and designed the study. JZ was responsible for data management and data preprocessing. XZ was responsible for data cleaning, data analysis, and interpretation. XZ wrote the paper. XZ, YL, WC, and HH contributed to the additional analysis in the revision of the paper. All authors read and approved the final article.

## Funding

This study was supported by the National Institutes of Health (R01GM141076, R21EY030488, and R01EB034116) and National

Science Foundation (Award number 2225775). This research is supported in part by the University of Pittsburgh Center for Research Computing through the resources provided. Specifically, this work used the HTC cluster, which is supported by NIH award number S10OD028483, and the H2P cluster, which is supported by NSF award number OAC-2117681.

## Acknowledgments

We thank the participants of all the studies included for enabling this research and strengthening our understanding of AMD genetics through their participation. We appreciate the International AMD Genomics Consortium for aggregating genetic data from research groups around the world, performing quality checks, and making the data available on dbGAP.

## 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.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1075824/full#supplementary-material>

## References

- Amado-Azevedo, J., Reinhard, N., van Bezu, J., van Nieuw Amerongen, G., van Hinsbergh, V., and Hordijk, P. (2018). The minor histocompatibility antigen 1 (hmla1)/arhgap45 is a racgap and a novel regulator of endothelial integrity. *Vasc. Pharmacol.* 101, 38–47. doi:10.1016/j.vph.2017.11.007
- Ambati, J., Atkinson, J. P., and Gelfand, B. D. (2013). Immunology of age-related macular degeneration. *Nat. Rev. Immunol.* 13, 438–451. doi:10.1038/nri3459
- Anderson, D. H., Radeke, M. J., Gallo, N. B., Chapin, E. A., Johnson, P. T., Curletti, C. R., et al. (2010). The pivotal role of the complement system in aging and age-related macular degeneration: Hypothesis re-visited. *Prog. Retin. eye Res.* 29, 95–112. doi:10.1016/j.preteyeres.2009.11.003
- Andrysik, Z., Bernstein, W. Z., Deng, L., Myer, D. L., Li, Y.-Q., Tischfield, J. A., et al. (2010). The novel mouse polo-like kinase 5 responds to dna damage and localizes in the nucleolus. *Nucleic acids Res.* 38, 2931–2943. doi:10.1093/nar/gkq011
- Arakawa, S., Takahashi, A., Ashikawa, K., Hosono, N., Aoi, T., Yasuda, M., et al. (2011). Genome-wide association study identifies two susceptibility loci for exudative age-related macular degeneration in the Japanese population. *Nat. Genet.* 43, 1001–1004. doi:10.1038/ng.938
- Banerji, J., Sands, J., Strominger, J. L., and Spies, T. (1990). A gene pair from the human major histocompatibility complex encodes large proline-rich proteins with multiple repeated motifs and a single ubiquitin-like domain. *Proc. Natl. Acad. Sci.* 87, 2374–2378. doi:10.1073/pnas.87.6.2374
- Becerra, E. M., Morescalchi, F., Gandolfo, F., Danzi, P., Nascimbeni, G., Arcidiacono, B., et al. (2011). Clinical evidence of intravitreal triamcinolone acetonide in the management of age-related macular degeneration. *Curr. drug targets* 12, 149–172. doi:10.2174/138945011794182746
- Bhattacharya, S., Chaum, E., Johnson, D. A., and Johnson, L. R. (2012). Age-related susceptibility to apoptosis in human retinal pigment epithelial cells is triggered by disruption of p53–mdm2 association. *Invest. Ophthalmol. Vis. Sci.* 53, 8350–8366. doi:10.1167/iovs.12-10495
- Binici, J., and Koch, J. (2014). Bag-6, a jack of all trades in health and disease. *Cell. Mol. life Sci.* 71, 1829–1837. doi:10.1007/s00018-013-1522-y
- Budzinskaia, M., Pogoda, T., Generozov, É., Chikun, E., Shchegoleva, I., Kazarian, É., et al. (2011). Influence of genetic mutations on clinical presentation of subretinal neovascularization. report 1: The impact of cfh and il-8 genes polymorphism. *Vestn. Oftalmol.* 127, 3–8.
- Buschini, E., Piras, A., Nuzzi, R., and Vercelli, A. (2011). Age related macular degeneration and drusen: Neuroinflammation in the retina. *Prog. Neurobiol.* 95, 14–25. doi:10.1016/j.pneurobio.2011.05.011

- Carlson, C. S., Eberle, M. A., Rieder, M. J., Yi, Q., Kruglyak, L., and Nickerson, D. A. (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* 74, 106–120. doi:10.1086/381000
- Chau, K., Sivaprasad, S., Patel, N., Donaldson, T., Luthert, P., and Chong, N. (2008). Plasma levels of matrix metalloproteinase-2 and-9 (mmp-2 and mmp-9) in age-related macular degeneration. *Eye* 22, 855–859. doi:10.1038/sj.eye.6702722x
- Chen, L., Pai, V., Levinson, R., Sharpe, A. H., Freeman, G. J., Braun, J., et al. (2009). Constitutive neuronal expression of the immune regulator, programmed death 1 (pd-1), identified during experimental autoimmune uveitis. *Ocular Immunol. Inflamm.* 17, 47–55. doi:10.1080/09273940802491884
- Chiang, C.-Y., Ching, Y.-H., Chang, T.-Y., Hu, L.-S., Yong, Y. S., Keak, P. Y., et al. (2020). Novel eye genes systematically discovered through an integrated analysis of mouse transcriptomes and phenome. *Comput. Struct. Biotechnol. J.* 18, 73–82. doi:10.1016/j.csbj.2019.12.009
- Choudhary, M., Ismail, E. N., Yao, P.-L., Tayyari, F., Radu, R. A., Nusinowitz, S., et al. (2020). Lxrs regulate features of age-related macular degeneration and may be a potential therapeutic target. *JCI insight* 5, e131928. doi:10.1172/jci.insight.131928
- Chuang, J.-Z., Yang, N., Nakajima, N., Otsu, W., Fu, C., Yang, H. H., et al. (2022). Retinal pigment epithelium-specific clic4 mutant is a mouse model of dry age-related macular degeneration. *Nat. Commun.* 13, 374–414. doi:10.1038/s41467-021-27935-9
- de Cárcer, G., Escobar, B., Higuero, A. M., Garcia, L., Ansón, A., Pérez, G., et al. (2011). Plk5, a polo box domain-only protein with specific roles in neuron differentiation and glioblastoma suppression. *Mol. Cell. Biol.* 31, 1225–1239. doi:10.1128/MCB.00607-10
- Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the *r/bioconductor* package biomart. *Nat. Protoc.* 4, 1184–1191. doi:10.1038/nprot.2009.97
- Ennis, S., Goverdhan, S., Cree, A., Hoh, J., Collins, A., and Lotery, A. (2007). Fine-scale linkage disequilibrium mapping of age-related macular degeneration in the complement factor h gene region. *Br. J. Ophthalmol.* 91, 966–970. doi:10.1136/bjo.2007.114090
- Fletcher, E. L., Jobling, A. I., Greferath, U., Mills, S. A., Waugh, M., Ho, T., et al. (2014). Studying age-related macular degeneration using animal models. *Optom. Vis. Sci.* 91, 878–886. doi:10.1097/OPX.0000000000000322
- Fritsche, L. G., Chen, W., Schu, M., Yaspan, B. L., Yu, Y., Thorleifsson, G., et al. (2013). Seven new loci associated with age-related macular degeneration. *Nat. Genet.* 45, 433–439. doi:10.1038/ng.2578
- 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
- Gibbs, D., Yang, Z., Constantine, R., Ma, X., Camp, N. J., Yang, X., et al. (2008). Further mapping of 10q26 supports strong association of htra1 polymorphisms with age-related macular degeneration. *Vis. Res.* 48, 685–689. doi:10.1016/j.visres.2007.10.022
- Gordiyenko, N. V., Fariss, R. N., Zhi, C., and MacDonald, I. M. (2010). Silencing of the chm gene alters phagocytic and secretory pathways in the retinal pigment epithelium. *Invest. Ophthalmol. Vis. Sci.* 51, 1143–1150. doi:10.1167/iovs.09-4117
- Goto, A., Akahori, M., Okamoto, H., Minami, M., Terauchi, N., Haruhata, Y., et al. (2009). Genetic analysis of typical wet-type age-related macular degeneration and polypoidal choroidal vasculopathy in Japanese population. *J. Ocular Biol. Dis. Inf.* 2, 164–175. doi:10.1007/s12177-009-9047-1
- Goudey, B., Abraham, G., Kikianty, E., Wang, Q., Rawlinson, D., Shi, F., et al. (2017). Interactions within the mhc contribute to the genetic architecture of celiac disease. *PLoS one* 12, e0172826. doi:10.1371/journal.pone.0172826
- Grassmann, F., Cantsilieris, S., Schulz-Kuhnt, A.-S., White, S. J., Richardson, A. J., Hewitt, A. W., et al. (2016). Multiallelic copy number variation in the complement component 4a (c4a) gene is associated with late-stage age-related macular degeneration (amd). *J. Neuroinflamm.* 13, 81–89. doi:10.1186/s12974-016-0548-0
- Guenther, F., Brandt, C., Winkler, T. W., Wanner, V., Stark, K., Kuechenhoff, H., et al. (2020). Chances and challenges of machine learning-based disease classification in genetic association studies illustrated on age-related macular degeneration. *Genet. Epidemiol.* 44, 759–777. doi:10.1002/gepi.12336
- Harney, S., Vilariño-Güell, C., Adamopoulos, I., Sims, A.-M., Lawrence, R., Cardon, L., et al. (2008). Fine mapping of the mhc class iii region demonstrates association of aif1 and rheumatoid arthritis. *Rheumatology* 47, 1761–1767. doi:10.1093/rheumatology/ken376
- Holliday, E. G., Smith, A. V., Cornes, B. K., Buitendijk, G. H., Jensen, R. A., Sim, X., et al. (2013). Insights into the genetic architecture of early stage age-related macular degeneration: A genome-wide association study meta-analysis. *PLoS one* 8, e53830. doi:10.1371/journal.pone.0053830
- Ishikawa, K., Kannan, R., and Hinton, D. R. (2016). Molecular mechanisms of subretinal fibrosis in age-related macular degeneration. *Exp. eye Res.* 142, 19–25. doi:10.1016/j.exer.2015.03.009
- Jasielska, M., Semkova, I., Shi, X., Schmidt, K., Karagiannis, D., Kokkinou, D., et al. (2010). Differential role of tumor necrosis factor (tnf)- $\alpha$  receptors in the development of choroidal neovascularization. *Invest. Ophthalmol. Vis. Sci.* 51, 3874–3883. doi:10.1167/iovs.09-5003
- Jin, H. L., Choi, Y., and Jeong, K. W. (2017). Crosstalk between aryl hydrocarbon receptor and glucocorticoid receptor in human retinal pigment epithelial cells. *Int. J. Endocrinol.* 2017, 5679517. doi:10.1155/2017/5679517
- Jonas, J. B., Cheung, C. M. G., and Panda-Jonas, S. (2017). Updates on the epidemiology of age-related macular degeneration. *Asia-Pacific J. Ophthalmol.* 6, 493–497. doi:10.22608/APO.2017251
- Jorgenson, E., Melles, R. B., Hoffmann, T. J., Jia, X., Sakoda, L. C., Kvale, M. N., et al. (2016). Common coding variants in the hla-dqb1 region confer susceptibility to age-related macular degeneration. *Eur. J. Hum. Genet.* 24, 1049–1055. doi:10.1038/ejhg.2015.247
- Ke, Y., Fan, X., Hao, R., Dong, L., Xue, M., Tan, L., et al. (2021). Human embryonic stem cell-derived extracellular vesicles alleviate retinal degeneration by upregulating oct4 to promote retinal müller cell retrodifferentiation via hsp90. *Stem Cell Res. Ther.* 12, 21–14. doi:10.1186/s13287-020-02034-6
- Keenan, T. D., Cukras, C. A., and Chew, E. Y. (2021). Age-related macular degeneration: Epidemiology and clinical aspects. *Age-related Macular Degener. Clin. Genes Back Patient Manag.* 1256, 1–31. doi:10.1007/978-3-030-66014-7\_1
- Khan, A. H., Pierce, C. O., De Salvo, G., Griffiths, H., Nelson, M., Cree, A. J., et al. (2021). The effect of systemic levels of tnf-alpha and complement pathway activity on outcomes of vegf inhibition in neovascular amd. *Eye* 1–8, 2192–2199. doi:10.1038/s41433-021-01824-3
- Kopplin, L. J., Igo, R., Wang, Y., Sivakumaran, T. A., Hagstrom, S. A., Peachey, N. S., et al. (2010). Genome-wide association identifies skiv2l and myrip as protective factors for age-related macular degeneration. *Genes Immun.* 11, 609–621. doi:10.1038/gene.2010.39
- Kubista, K. E., Tosakulwong, N., Wu, Y., Ryu, E., Roeder, J. L., Hecker, L. A., et al. (2011). Copy number variation in the complement factor h-related genes and age-related macular degeneration. *Mol. Vis.* 17, 2080–2092.
- Kurji, K. H., Cui, J. Z., Lin, T., Harriman, D., Prasad, S. S., Kojic, L., et al. (2010). Microarray analysis identifies changes in inflammatory gene expression in response to amyloid- $\beta$  stimulation of cultured human retinal pigment epithelial cells. *Invest. Ophthalmol. Vis. Sci.* 51, 1151–1163. doi:10.1167/iovs.09-3622
- Kvale, M. N., Hesselson, S., Hoffmann, T. J., Cao, Y., Chan, D., Connell, S., et al. (2015). Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (gera) cohort. *Genetics* 200, 1051–1060. doi:10.1534/genetics.115.178905
- Kwon, W., and Freeman, S. A. (2020). Phagocytosis by the retinal pigment epithelium: Recognition, resolution, recycling. *Front. Immunol.* 11, 604205. doi:10.3389/fimmu.2020.604205
- Lange, C., Mowat, F., Sayed, H., Mehad, M., Duluc, L., Piper, S., et al. (2016). Dimethylarginine dimethylaminohydrolase-2 deficiency promotes vascular regeneration and attenuates pathological angiogenesis. *Exp. Eye Res.* 147, 148–155. doi:10.1016/j.exer.2016.05.007
- Lees, C., Barrett, J., Parkes, M., and Satsangi, J. (2011). New ibd genetics: Common pathways with other diseases. *Gut* 60, 1739–1753. doi:10.1136/gut.2009.199679
- Li, B., and Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *Am. J. Hum. Genet.* 83, 311–321. doi:10.1016/j.ajhg.2008.06.024
- Li, Y., Hong, H. G., Ahmed, S. E., and Li, Y. (2019a). Weak signals in high-dimension regression: Detection, estimation and prediction. *Appl. Stoch. Models Bus. Ind.* 35, 283–298. doi:10.1002/asmb.2340
- Li, Y., Hong, H. G., and Li, Y. (2019b). Multiclass linear discriminant analysis with ultrahigh-dimensional features. *Biometrics* 75, 1086–1097. doi:10.1111/biom.13065
- Li, Y. (2019). “A local-network guided linear discriminant analysis for classifying lung cancer subtypes using individual genome-wide methylation profiles,” in Proceedings of the Future Technologies Conference (Springer), 676–687.
- Liang, G., Ma, W., Luo, Y., Yin, J., Hao, L., and Zhong, J. (2022). Identification of differentially expressed and methylated genes and construction of a co-expression network in age-related macular degeneration. *Ann. Transl. Med.* 10, 223. doi:10.21037/atm-21-7043
- Logue, M. W., Schu, M., Vardarajan, B. N., Farrell, J., Lunetta, K. L., Jun, G., et al. (2014). A search for age-related macular degeneration risk variants in alzheimer disease genes and pathways. *Neurobiol. aging* 35, 1510.e7–1510.e18. doi:10.1016/j.neurobiolaging.2013.12.007
- Mells, G. F., Floyd, J. A., Morley, K. I., Cordell, H. J., Franklin, C. S., Shin, S.-Y., et al. (2011). Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* 43, 329–332. doi:10.1038/ng.789
- Mukwaya, A., Lennikov, A., Xeroudaki, M., Mirabelli, P., Lachota, M., Jensen, L., et al. (2018). Time-dependent lxr/txr pathway modulation characterizes capillary remodeling

- in inflammatory corneal neovascularization. *Angiogenesis* 21, 395–413. doi:10.1007/s10456-018-9604-y
- Parameswaran, S., and Krishnakumar, S. (2017). Pluripotent stem cells: A therapeutic source for age-related macular degeneration. *Indian J. Ophthalmol.* 65, 177–183. doi:10.4103/ijo.IJO\_1026\_15
- Pollio, G., Hoozemans, J. J., Andersen, C. A., Roncarati, R., Rosi, M. C., van Haastert, E. S., et al. (2008). Increased expression of the oligopeptidase thop1 is a neuroprotective response to  $\alpha\beta$  toxicity. *Neurobiol. Dis.* 31, 145–158. doi:10.1016/j.nbd.2008.04.004
- Potilinski, M. C., Tate, P. S., Lorenc, V. E., and Gallo, J. E. (2021). New insights into oxidative stress and immune mechanisms involved in age-related macular degeneration tackled by novel therapies. *Neuropharmacology* 188, 108513. doi:10.1016/j.neuropharm.2021.108513
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795
- Qin, W., Hu, L., Zhang, X., Jiang, S., Li, J., Zhang, Z., et al. (2019). The diverse function of pd-1/pd-l pathway beyond cancer. *Front. Immunol.* 10, 2298. doi:10.3389/fimmu.2019.02298
- Racic, T., Chang, A., Fernando, N., Brandli, A., Natoli, R., Penfold, P., et al. (2021). Anti-inflammatory and neuroprotective properties of the corticosteroid fludrocortisone in retinal degeneration. *Exp. Eye Res.* 212, 108765. doi:10.1016/j.exer.2021.108765
- Ren, K., Tang, Z.-L., Jiang, Y., Tan, Y.-M., and Yi, G.-H. (2015). Apolipoprotein m. *Clin. Chim. Acta* 446, 21–29. doi:10.1016/j.cca.2015.03.038
- Slatkin, M. (2008). Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* 9, 477–485. doi:10.1038/nrg2361
- Sleiman, P. M., Flory, J., Imielinski, M., Bradfield, J. P., Annaiah, K., Willis-Owen, S. A., et al. (2010). Variants of dennd1b associated with asthma in children. *N. Engl. J. Med.* 362, 36–44. doi:10.1056/NEJMoa0901867
- Strunz, T., Kiel, C., Sauerbeck, B. L., and Weber, B. H. (2020a). Learning from fifteen years of genome-wide association studies in age-related macular degeneration. *Cells* 9, 2267. doi:10.3390/cells9102267
- Strunz, T., Lauwen, S., Kiel, C., Hollander, A. d., and Weber, B. H. (2020b). A transcriptome-wide association study based on 27 tissues identifies 106 genes potentially relevant for disease pathology in age-related macular degeneration. *Sci. Rep.* 10, 1584–1616. doi:10.1038/s41598-020-58510-9
- Strunz, T. (2021). *Genetic variants with significant association to age-related macular degeneration (AMD) and their role in the regulation of gene expression*. Ph.D. thesis. Regensburg: University of Regensburg.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102
- Tam, P. O., Ng, T. K., Liu, D. T., Chan, W. M., Chiang, S. W., Chen, L. J., et al. (2008). Htra1 variants in exudative age-related macular degeneration and interactions with smoking and cfh. *Invest. Ophthalmol. Vis. Sci.* 49, 2357–2365. doi:10.1167/iovs.07-1520
- Tan, W., Zou, J., Yoshida, S., Jiang, B., and Zhou, Y. (2020). The role of inflammation in age-related macular degeneration. *Int. J. Biol. Sci.* 16, 2989–3001. doi:10.7150/ijbs.49890
- Tian, J., Yu, W., Qin, X., Fang, K., Chen, Q., Hou, J., et al. (2012). Association of genetic polymorphisms and age-related macular degeneration in Chinese population. *Invest. Ophthalmol. Vis. Sci.* 53, 4262–4269. doi:10.1167/iovs.11-8542
- Ung, C., Lains, I., Miller, J. W., and Kim, I. K. (2021). Current management of age-related macular degeneration. *Age-Related Macular Degener. Clin. Genes Back Patient Manag.* 1256, 295–314. doi:10.1007/978-3-030-66014-7\_12
- van Leeuwen, R., Ikram, M. K., Vingerling, J. R., Witteman, J. C., Hofman, A., and de Jong, P. T. (2003). Blood pressure, atherosclerosis, and the incidence of age-related maculopathy: The rotterdam study. *Investigative Ophthalmol. Vis. Sci.* 44, 3771–3777. doi:10.1167/iovs.03-0121
- Vingerling, J. R., Dielemans, I., Bots, M. L., Hofman, A., Grobbee, D. E., and de Jong, P. T. (1995). Age-related macular degeneration is associated with atherosclerosis: The rotterdam study. *Am. J. Epidemiol.* 142, 404–409. doi:10.1093/oxfordjournals.aje.a117648
- Vuong, L., Conley, S. M., and Al-Ubaidi, M. R. (2012). Expression and role of p53 in the retina. *Invest. Ophthalmol. Vis. Sci.* 53, 1362–1371. doi:10.1167/iovs.11-8909
- Wavre-Shapton, S. T., Tolmachova, T., da Silva, M. L., Futter, C. E., and Seabra, M. C. (2013). Conditional ablation of the choroideremia gene causes age-related changes in mouse retinal pigment epithelium. *PLoS one* 8, e57769. doi:10.1371/journal.pone.0057769
- Wolf, J., Schlecht, A., Rosmus, D.-D., Boneva, S., Agostini, H., Schlunck, G., et al. (2022). Comparative transcriptome analysis of human and murine choroidal neovascularization identifies fibroblast growth factor inducible-14 as phylogenetically conserved mediator of neovascular age-related macular degeneration. *Biochim. Biophys. Acta (BBA)-Mol. Basis Dis.* 1868, 166340. doi:10.1016/j.bbdis.2022.166340
- Wong, W. L., Su, X., Li, X., Cheung, C. M. G., Klein, R., Cheng, C.-Y., et al. (2014). Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: A systematic review and meta-analysis. *Lancet Glob. Health* 2, e106–e116. doi:10.1016/S2214-109X(13)70145-1
- Zhang, J., and Zhou, Y. (2020). Identification of key genes and pathways associated with age-related macular degeneration. *J. Ophthalmol.* 2020, 2714746. doi:10.1155/2020/2714746
- Zhang, H., Morrison, M. A., DeWan, A., Adams, S., Andreoli, M., Huynh, N., et al. (2008). The nei/ncbi dbgap database: Genotypes and haplotypes that may specifically predispose to risk of neovascular age-related macular degeneration. *BMC Med. Genet.* 9, 51–10. doi:10.1186/1471-2350-9-51