

## RESEARCH ARTICLE

# Multi-ancestry GWAS analysis identifies two novel loci associated with diabetic eye disease and highlights *APOL1* as a high risk locus in patients with diabetic macular edema

Amy D. Stockwell<sup>1</sup>, Michael C. Chang<sup>1</sup>, Anubha Mahajan<sup>1</sup>, William Forrest<sup>1</sup>, Neha Anegondi<sup>1</sup>, Rion K. Pendergrass<sup>1</sup>, Suresh Selvaraj<sup>1</sup>, Jens Reeder<sup>1</sup>, Eric Wei<sup>1</sup>, Victor A. Iglesias<sup>2</sup>, Natalie M. Creps<sup>1</sup>, Laura Macri<sup>3</sup>, Andrea N. Neerajan<sup>3</sup>, Marcel P. van der Brug<sup>3</sup>, Suzie J. Scales<sup>1</sup>, Mark I. McCarthy<sup>1</sup>, Brian L. Yaspan<sup>1\*</sup>

**1** Genentech, San Francisco, California, United States of America, **2** F. Hoffmann-La Roche Ltd, Grenzacherstrasse, Basel, Switzerland, **3** Character Biosciences, San Francisco, California, United States of America

\* [yaspan.brian@gene.com](mailto:yaspan.brian@gene.com)



## OPEN ACCESS

**Citation:** Stockwell AD, Chang MC, Mahajan A, Forrest W, Anegondi N, Pendergrass RK, et al. (2023) Multi-ancestry GWAS analysis identifies two novel loci associated with diabetic eye disease and highlights *APOL1* as a high risk locus in patients with diabetic macular edema. *PLoS Genet* 19(8): e1010609. <https://doi.org/10.1371/journal.pgen.1010609>

**Editor:** Gregory S. Barsh, HudsonAlpha Institute for Biotechnology, UNITED STATES

**Received:** January 10, 2023

**Accepted:** June 11, 2023

**Published:** August 16, 2023

**Copyright:** © 2023 Stockwell et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Complete summary statistics from the discovery GWAS multi-ancestry meta analysis are available at [http://research-pub.gene.com/Stockwell\\_et\\_al\\_DME\\_GWAS\\_2023](http://research-pub.gene.com/Stockwell_et_al_DME_GWAS_2023).

**Funding:** Funding for this project was from F. Hoffmann - La Roche. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Diabetic retinopathy (DR) is a common complication of diabetes. Approximately 20% of DR patients have diabetic macular edema (DME) characterized by fluid leakage into the retina. There is a genetic component to DR and DME risk, but few replicable loci. Because not all DR cases have DME, we focused on DME to increase power, and conducted a multi-ancestry GWAS to assess DME risk in a total of 1,502 DME patients and 5,603 non-DME controls in discovery and replication datasets. Two loci reached GWAS significance ( $p < 5 \times 10^{-8}$ ). The strongest association was rs2239785, (K150E) in *APOL1*. The second finding was rs10402468, which co-localized to *PLVAP* and *ANKLE1* in vascular / endothelium tissues. We conducted multiple sensitivity analyses to establish that the associations were specific to DME status and did not reflect diabetes status or other diabetic complications. Here we report two novel loci for risk of DME which replicated in multiple clinical trial and biobank derived datasets. One of these loci, containing the gene *APOL1*, is a risk factor in African American DME and DKD patients, indicating that this locus plays a broader role in diabetic complications for multiple ancestries.

**Trial Registration:** [NCT00473330](https://clinicaltrials.gov/ct2/show/study/NCT00473330), [NCT00473382](https://clinicaltrials.gov/ct2/show/study/NCT00473382), [NCT03622580](https://clinicaltrials.gov/ct2/show/study/NCT03622580), [NCT03622593](https://clinicaltrials.gov/ct2/show/study/NCT03622593), [NCT04108156](https://clinicaltrials.gov/ct2/show/study/NCT04108156).

## Author summary

Although twin and family studies have clearly shown that genes play a role in the development of diabetic macular edema (DME), confirmed risk loci have remained elusive. Here, in the largest genetic study of DME to date with >7,000 individuals, we utilize data from clinical trials and large biobanks to identify and confirm two risk loci associated with risk of DME, one on Chromosome 19 and another on Chromosome 22. The locus on

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: ADS, MCC, WF, NA, AM, RKP, SS, VI, JR, RW, EW, NMC, MIM, and BLY are full time employees of Roche-Genentech with stock or stock options in Roche, or were at the time of study execution. SJS is a full time employee of Roche-Genentech with stock or stock options in Roche, and is a patent holder for Animal model of nephropathy and agents for treating the same. Docket No. P32407-US-PR. US Application no. 15798303.2-1410. Filed November 10th 2015. European patent Published 4.17.19 Application no. 19152929.6-1120. LM, ANN, and MPB are full time employees of Character Biosciences with stock or stock options in Character Biosciences, or were at the time of study execution.

Chromosome 19 contains two genes of interest, *PLVAP* and *ANKLE1*. The locus on Chromosome 22 is tagged by a functional variant in the gene *APOLI1*, a gene previously associated with kidney disease in African descent populations, suggesting *APOLI1* plays a broader role in diabetic complications than previously thought. Further studies of the genes at these loci will shed light on how they affect DME. These findings are important because they represent a significant step forward in the understanding of the genetic component to diabetic eye disease.

## Introduction

Diabetic retinopathy (DR) is a common complication of diabetes and one of the leading causes of blindness worldwide [1]. Approximately 20% of DR patients have diabetic macular edema (DME) characterized by fluid leakage into the retina [2]. DME can occur at any point in the timeline of DR, but is more likely to occur later in the disease. Diabetic patients with African American ancestry have the greatest risk of developing DME of all populations in the United States, estimated at three times greater than those of European-descent [3]. Risk factors include improper glycemic control and duration of diabetes, but these do not explain all of the heterogeneity in DME risk [4–6]. While it is accepted there is a genetic component to DR and DME risk, to date, large-scale genome-wide association studies (GWAS) have not yielded robust insights [7]. Most well powered studies have been in DR, and have been well summarized [8]. DR is an inherently heterogeneous disease marker, and one suggestion cited for identification of replicable loci identified via GWAS was harmonized patient phenotyping [9]. As an example, one recent study containing DR, DME and proliferative diabetic retinopathy cases in a multi-ethnic study population of 43,565 was unable to find loci at genome-wide significance [10]. Because not all DR cases have DME, focusing on that subtype alone could increase power to find novel loci associated with the risk of disease. Furthermore, the increased prevalence of DR and DME in African American individuals indicates there may be genetic risk factors specific to that population. To that end, we performed the largest genetic study of DME to date (a total of 1,502 cases and 5,603 controls) identifying two loci reaching genome-wide significance for risk of DME, one of which contains *APOLI1* and is estimated to account for approximately 5% of the risk of developing DME in African American patients with diabetes.

## Methods

### Study design and populations

We performed a genome wide association study using DNA derived from blood samples obtained from DME patients from clinical trials for ranibizumab (NCT00473330 [RISE] and NCT00473382 [RIDE]), faricimab (NCT03622580 [YOSEMITE] and NCT03622593 [RHINE]), and the port delivery system with ranibizumab (NCT04108156 [PAGODA]) (S1 Table). These study populations were selected for inclusion on the basis of available phenotypic information and DNA availability for whole-genome sequencing (all but PAGODA) and array genotyping (PAGODA). Patients were sequenced and genotyped between 2017–2021.

Samples and data for controls in the discovery study population (cases from the ranibizumab (RISE and RIDE) and faricimab (YOSEMITE and RHINE) studies) were obtained from clinical trial studies of asthma, autoimmune disease, Crohn's disease, chronic obstructive pulmonary disease, inflammatory bowel disease, interstitial lung disease, idiopathic pulmonary fibrosis, rheumatic arthritis, systemic lupus erythematosus and ulcerative colitis. An

independent set of controls for the replication study population (cases from the port delivery study with ranibizumab–PAGODA) were selected from clinical trial studies of cancer and asthma, in addition to healthy controls from a cohort study examining Age-Related Macular Degeneration in subjects with African ancestry (DIVERSITY).

All patients included in this study provided written informed consent for whole-genome sequencing or array genotyping of their DNA. Ethical approval was provided as per the original clinical trials.

### DNA analysis

The WGS data for the discovery study population was generated to a read depth of 30X using the HiSeq platform (Illumina X10, San Diego, CA, USA) processed using the Burrows-Wheeler Aligner (BWA) / Genome Analysis Toolkit (GATK) best practices pipeline. WGS short reads were mapped to hg38 / GRCh38 (GCA\_000001405.15), including alternate assemblies, using BWA version 0.7.9a-r786 to generate BAM files. All WGS data was subject to quality control and checked for concordance with SNP fingerprint data collected before sequencing. After filtering for genotypes with a GATK genotype quality greater than 90, samples with heterozygote concordance with SNP chip data of less than 75% were removed. Sample contamination was determined with VerifyBamID software, and samples with a freemix parameter of more than 0.03 were excluded. Joint variant calling was done using the GATK best practices joint genotyping pipeline to generate a single variant call format (VCF) file. The called variants were then processed using ASDPEX to filter out spurious variant calls in the alternate regions.

Data in the replication study population was generated using the Infinium Global Screening Array-24 Kit v2 and v3 from Illumina. Genotypes with GenCall (GC) score less than 0.15 were set to missing.

### Quality control–Patient DNA samples

Ancestry was assigned using an ancestry threshold of 0.7 obtained from supervised analysis with ADMIXTURE 1.3 using samples from phase 3 of the 1000 Genomes Project as reference samples labeled by their superpopulation (**S1 Fig**) (AFR = African [from the African continent and African ancestry from the US and Caribbean]; AMR = AdMixed American [from North and South America excluding US of African Ancestry and Caribbean of African ancestry]; EAS = East Asian [from China, Japan and Vietnam]; EUR = European [from the United States with Northern and Western European ancestry, Italy, Finland, England and Scotland]).

For the discovery study population, MatchIt [11] was used to do a 4:1 (control:case) ratio using the nearest neighbor matching on the propensity score to select controls based on age, sex and global ancestry.

For both the discovery and replication study populations, samples were excluded if the call rate was less than 90%. Identity by descent analysis was used to detect and filter out relatedness in the dataset; samples were excluded if PI\_HAT was 0.4 or higher. Samples were removed if they showed excess heterozygosity with more than three SDs of the mean. This resulted in 448 in the AFR study population (99 cases and 349 controls), 343 in the AMR study population (87 cases and 347 controls), 439 in the EAS study population (88 cases and 351 controls) and 4,577 in the EUR study population (921 cases and 3,656 controls) for the discovery dataset with a combined total of 1,195 DME cases and 4,703 non-DME controls. Similarly, after quality control there are 976 patient samples in the EUR replication study population (262 cases and 714 controls) and 231 subjects in the AFR replication study population (45 cases and 186 controls) for a combined total of 307 DME cases and 900 non-DME controls.

## Quality control—genotype

For the replication study population, multi-allelic sites were split into biallelic records, variants were left-aligned and normalized, and duplicate variants based on chromosome, position, ref allele, and alt allele were removed prioritizing by missingness rate. Additional quality control was performed by removing variants that had a target majority ancestry alternate allele frequency (AF) difference exceeding 0.2 compared to the reference, as well as removing ambiguous (A/T, G/C) variants where target majority ancestry minor allele frequency (MAF) exceeded 0.4 in either the target or reference. Phasing and imputation were carried out using Beagle 5.1 with hg38 HapMap genetic maps and a reference panel consisting of merged Genentech internal whole-genome sequencing and Haplotype Reference Consortium (HRC) genotypes. This merged panel consists of 134,246,541 bi-allelic sites in 51,455 (chr1), 55,929 (chr2-22), and 54,963 (chrX) samples. The default sliding window length of 40 cM with 4 cM overlap between adjacent sliding windows was used. Imputed VCFs were annotated with rsIDs from dbSNP154. The estimated squared correlation between the estimated allele dose and the true allele dose (DR<sup>2</sup>) was used to filter out poorly imputed variants by setting a DR<sup>2</sup> threshold of 0.3.

Sample genotypes (discovery WGS only) were set to missing if the Genotype Quality score was less than 20 and SNPs were removed if the missingness was higher than 5%. SNPs were filtered if the significance level for the Hardy-Weinberg equilibrium test was less than  $5 \times 10^{-8}$ . The allele depth balance test was performed to test for equal allele depth at heterozygote carriers using a binomial test (discovery WGS only); SNPs were excluded if the p-value was less than  $1 \times 10^{-5}$ .

## Statistical analysis

A common variant (MAF  $\geq 1\%$ ) genome-wide association study (GWAS) was conducted to assess DME risk in the AFR, AMR, EAS and EUR study populations separately. PLINK [12] was used to perform logistic regression two-sided test using an additive model, adjusted for age, sex and genetic ancestry. The first three PCAs were used in as covariates for genetic ancestry and their cumulative proportion of variance was  $>90\%$ . METAL [13] was subsequently used to do a meta-analysis, under a mixed effects model. There were 7,601,242 variants that passed QC in one or more of the cohorts. The ancestry-specific cohorts were QC'd and analyzed separately, rather than filtering on MAF in the combined cohort. For the meta-analysis, variants were included if they were present in two or more of the ancestry groups.

LDSC software was used separately on EUR and AFR GWAS summary stats to calculate heritability [14]. The `parp` function in the CRAN R library `twoxtwo` was used to calculate the population attributable risk proportion in EUR and AFR study populations combined and separately [15].

Samples from other disease areas were used as controls. We used a genotype-on-phenotype reverse regression to test for association with the non-DME controls. We calculated the posterior probability (PP) that a SNP is also associated with each individual disease area included in the control cohort; SNPs passing the 0.6 PP threshold in any of the control cohort disease areas were flagged and excluded from the GWAS results.

## FinnGen ethics statement and methods

A second replication population consisted of patients enrolled in the FinnGen Biobank cohort. Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August

2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.

The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, THL/1524/5.05.00/2020, and THL/2364/14.02/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020, Findata THL/2364/14.02/2020 and Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20).

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 7 include: THL Biobank BB2017\_55, BB2017\_111, BB2018\_19, BB\_2018\_34, BB\_2018\_67, BB2018\_71, BB2019\_7, BB2019\_8, BB2019\_26, BB2020\_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154 and amendment #1 (August 17 2020), Biobank Borealis of Northern Finland\_2017\_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1–2017, and Terveystalo Biobank STB 2018001.

## Colocalization

We utilized colocalization analysis between eQTL data from GTEx, eyeQTL, and Sun et al pqtls to identify putative causal genes in the region [16–18]. The coloc library (CRAN R) uses approximate Bayes factors to estimate posterior probabilities (PP) for common variants in the GWAS study as well as the QTL study. A PP4 estimates the probability that a variant is associated with both traits and a PP4 threshold of 0.8 was used as evidence for colocalization. Variants within +/- 250 bp of the most significant SNP in the locus were included. GTEx tissues relevant to complications of diabetes were included: kidney and vascular / endothelium (artery aorta, artery coronary and artery tibial). EyeQTL included the following tissues: RPE, retina, RPE macula, RPE non-macula, retina macula, and retina non-macula.

## Results

We conducted a multi-ancestry GWAS to identify risk alleles associated with DME. Discovery population cases (N = 1,195) comprised patients from clinical trials for DME involving ranibizumab (NCT00473330 [RISE N = 126], NCT00473382 [RIDE N = 143]) and faricimab (NCT03622580 [YOSEMITE N = 511] and NCT03622593 [RHINE N = 415]) who consented for genetic analysis. Inclusion criteria for all studies allowed for both T1D and T2D patients. Based on self-reported data, from YOSEMITE and RHINE, we estimate 6% of participants have T1D and 94% have T2D. We obtained controls (N = 4,703) from participants in non-ophthalmic clinical trials that were sequenced and processed with the same informatics pipeline; this ensures we minimize batch effects that may be introduced via differing sequencing technologies and bioinformatics processing pipelines (see [methods](#) for details). We grouped DME cases and non-DME controls by genetic ancestry using defined populations in the 1000 Genomes Project. Therefore, population subsets in this study will be referred to by the 1000 Genomes super-population codes (see [methods](#) for details). We performed case and control

Table 1. Cohort demographics.

	Discovery Cases	Discovery Controls	P-value	Replication Cases	Replication Controls	P-value
AFR						
N	99	349		45	186	
Age, mean (sd)	61.2 (9.6)	56.8 (8.6)	0.000041	61.1 (8.6)	71.7 (8.1)	5.50E-07
Female, N (%)	62 (62.6%)	247 (70.7%)	0.15	22 (48.8%)	120 (64.5%)	0.07
AMR						
N	87	347		-	-	-
Age, mean (sd)	59.3 (10.5)	57.8 (11.5)	0.29	-	-	-
Female, N (%)	29 (33.3%)	175 (50.4%)	0.006	-	-	-
EAS						
N	88	351		-	-	-
Age, mean (sd)	62.0 (10.4)	51.2 (14.1)	1.1E-09	-	-	-
Female, N (%)	33 (37.5%)	215 (61.2%)	0.000096	-	-	-
EUR						
N	921	3656		262	714	
Age, mean (sd)	62.8 (9.7)	62.8 (9.7)	0.96	61.3 (9.5)	54.8 (11.6)	8.10E-09
Female, N (%)	375 (40.7%)	1498 (40.9%)	0.88	111 (42.3%)	472 (66.1%)	3.40E-11
Combined						
N	1195	4703		307	900	
Age, mean (sd)	62.3 (9.8)	61.1 (10.7)	0.001	61.3 (9.3)	59.1 (13.1)	0.04
Female, N (%)	499 (41.7%)	2135 (45.4%)	0.02	133 (43.3)	592 (65.7%)	6.40E-12

\* AFR = African descent; AMR = American descent; EAS = East Asian descent; EUR = European descent

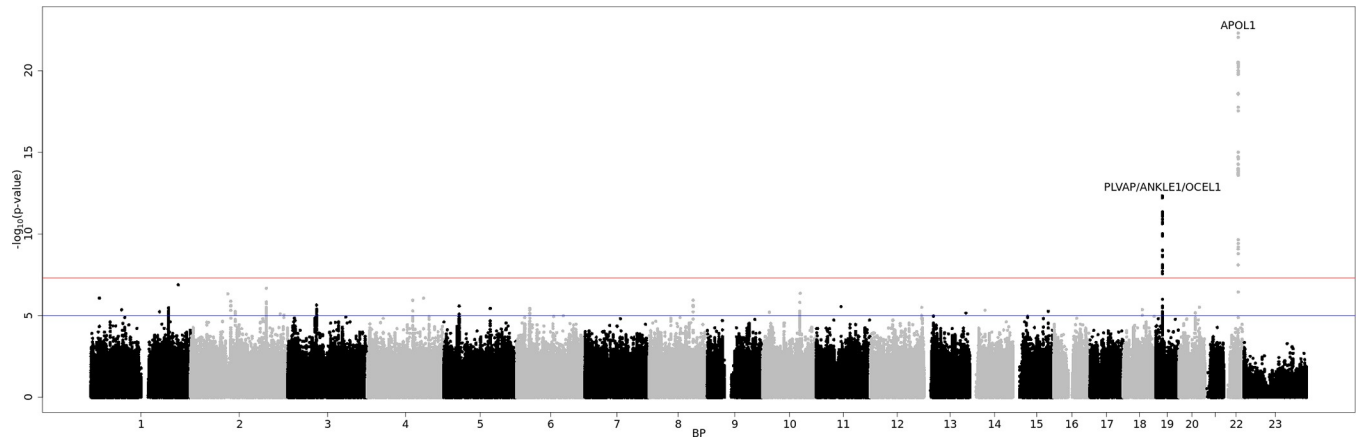
<https://doi.org/10.1371/journal.pgen.1010609.t001>

analyses for each of the four super-populations (AFR—African, EUR—European, EAS—East Asian and AMR—AdMixed American), controlling for age, sex and genetically defined ancestry, combining results via meta-analysis. Due to the use of individuals with other, non-ophthalmic diseases as controls, we used a genotype-on-phenotype reverse regression to remove non-DME specific findings, which have been removed from the presented analysis [19]. We sought to replicate any genome-wide significant signals in a sample of 307 multi-ancestry DME cases from a clinical trial evaluating ranibizumab in the Port Delivery System (NCT04108156 [PAGODA]) and 900 non-DME controls obtained in a similar fashion to the discovery dataset. Case and control demographics for the discovery and replication populations are available in Table 1.

## GWAS identifies loci on chr 22q12 and chr19p13 as associated with risk of DME

There was no evidence of genomic inflation in this meta-analysis ( $\lambda_{gc} = 0.98$  S2 Fig). We estimated the SNP heritability ( $h^2$ ) in the largest two super-populations, EUR and AFR, to be 0.41 and 0.48 respectively.

After genome-wide removal of non-DME specific loci via reverse regression, we found two novel loci at chr22q12.3 and chr19p13.11 reaching the pre-defined genome-wide significance level of  $P < 5 \times 10^{-8}$  (Fig 1 and Table 2; super-population specific results—S3A–S3D Fig), (Locus specific plots at Figs 2 and 3 (chr19p13.11, chr22q12.3 respectively) and super-population specific results—S4A–S4H Fig).



**Fig 1. Manhattan plot of the discovery study population containing 1,195 DME cases and 4,703 non-DME controls.** Gene names in peaks containing SNPs with a  $P < 5 \times 10^{-8}$  are labeled. The red line signifies the threshold for genome-wide significance ( $P < 5 \times 10^{-8}$ ) and the blue line indicates the threshold for suggestive significance ( $P < 1 \times 10^{-5}$ ).

<https://doi.org/10.1371/journal.pgen.1010609.g001>

### Chr22q12 –*APOL1*

The strongest signal of association in the discovery study population was at chr22q12.3 (S2 Table). The top SNP at this locus, rs2239785, results in a missense variant (K150E) in *APOL1*, encoding apolipoprotein L1 (G allele— $OR_{meta} = 1.71$ ,  $P_{meta} = 5.05 \times 10^{-23}$ ). The risk (G) allele encodes the E150 allele of the peptide. Previous investigation has shown the E150 allele is important, since in the context of I228 and K255 (the AFR "EIK" haplotype) it increases APO-L1-induced cytotoxicity in HEK293 cells compared to K150 (KIK) expressed at similar levels, whereas in the context of the refseq haplotype (EMR; E150/M228/R255), there is almost no cytotoxicity [20]. Furthermore, in Africans the minor K150 allele was shown to be protective for diabetic, non-diabetic and all-cause end-stage kidney disease in a dominant model (ORs 0.76–0.8), implying the major (E150) allele exacerbates kidney disease (presumably in the context of I228 and K255 since this is the major African haplotype) [21]. However, to our knowledge, this is from a single study and more investigation is warranted.

The direction of effect was consistent across all four super-populations ( $OR_{eur} = 1.73$ ,  $P_{eur} = 4.07 \times 10^{-20}$ ;  $OR_{afr} = 1.84$ ,  $P_{afr} = 0.002$ ;  $OR_{amr} = 1.43$ ,  $P_{amr} = 0.14$ ;  $OR_{eas} = 1.43$ ,  $P_{eas} = 0.12$ ) (Fig 4). The frequency of this allele was highest in AFR DME patients ( $EAF_{afr} = 0.76$ ), greater than twice that of EUR DME patients ( $EAF_{eur} = 0.31$ ). We replicated the association at rs2239785 in an independent multi-ancestry DME case (PAGODA) and control study sample ( $OR_{meta} = 1.41$ ,  $P_{meta} = 0.003$ ;  $OR_{eur} = 1.31$ ,  $P_{eur} = 0.03$ ;  $OR_{afr} = 2.06$ ,  $P_{afr} = 0.02$ ), with a combined sample  $OR = 1.65$  and  $P = 2.0 \times 10^{-24}$  (Table 3).

To further investigate this association, we performed a PheWAS for rs2239785 in the FinnGen biobank study population (Freeze 7, <https://r7.finnngen.fi>). FinnGen is a large biobank derived study where case assignment is based on physician coding, and cases were compared to the remainder of FinnGen study participants. In FinnGen, diabetic maculopathy was the only coded indication specific to diabetic complications of the macula, the area of the retina impacted in DME. For the index SNP at the *APOL1* locus found in this GWAS (rs2239785), we saw an association with diabetic maculopathy with the same direction of effect as the primary DME GWAS ( $OR = 1.17$ ,  $P = 5.5 \times 10^{-5}$ ,  $MAF_{cases} = 0.18$ ,  $MAF_{controls} = 0.16$ ;  $N = 2,790$  cases, 284,826 controls). This was the second top association in all coded phenotypes for this variant, with the first being intrahepatic cholestasis of pregnancy (S3 Table). Finally we queried the summary statistics for a recent DME GWAS in Australian patients of EUR ancestry

Table 2. Independent list of SNPs that reach the suggestive p-value (1E-5) threshold in the meta-analysis.

CHR	BP	SNP	A1	A2	OR	P	LOCUS
22	36265284	rs2239785	G	A	1.707	5.05E-23	APOL1 (missense)
19	17273194	rs10402468	A	C	1.551	4.62E-13	BABAM1 (intron)
22	36268440	rs136179	T	C	1.367	2.2E-10	LOC112268296 (intron)
22	36246324	rs713791	G	A	0.738	6.32E-10	* APOL2-APOL1
19	17242176	rs10419742	C	G	1.529	1.96E-08	NR2F6 (intron)
2	188778589	rs11901929	A	G	1.294	2.07E-07	DIRC1 (intron)
22	36256843	rs136147	T	G	0.781	3.57E-07	APOL1 (intron)
1	21640340	rs12142081	T	G	1.274	0.00000084	RAP1GAP (intron)
19	17357281	rs73016427	A	G	1.481	0.000001	PLVAP (intron)
8	110170017	rs11984657	G	A	1.712	0.00000114	* RPSAP48-RP11-211F2.1
4	113000224	rs7655898	T	A	1.276	0.00000115	ANK2 (intron)
2	99866558	rs13005475	G	T	0.791	0.0000013	AFF3 (intron)
10	92670740	rs2185756	C	A	0.79	0.00000156	* EIF2S2P3-HHEX
5	38633105	rs327275	C	G	0.792	0.00000257	LIFR-AS1 (intron)
2	173576901	rs10930604	G	T	1.733	0.00000294	LOC107985961 (intron)
20	50686424	rs17726420	A	G	1.6	0.00000302	RIPOR3 (intron)
12	127471508	rs11059148	T	C	1.255	0.00000311	* RP11-483I24.1-RP11-351O2.1
1	193892611	rs555119	T	C	1.309	0.00000321	* RPL23AP22-RP11-563D10.1
5	115856933	rs7713314	A	G	1.312	0.00000361	AP3S1 (intron)
6	33097934	rs3117211	A	C	1.353	0.00000364	* HLA-DPA2-COL11A2P1
18	47409305	rs80057143	A	G	1.378	0.00000425	MIR4257HG (intron)
15	33638551	rs74005915	G	A	1.516	0.00000448	RYR3 (intron)
14	37754326	rs68016111	C	G	1.293	0.00000468	TTC6 (intron)
11	1221460	rs868903	T	C	1.246	0.00000499	* MUC5AC-MUC5B
2	111207578	rs56693625	A	G	0.774	0.00000528	MIR4435-2HG (intron)
15	88311946	rs9920776	A	G	1.579	0.00000537	LOC105370958 (intron)
2	188494506	rs13005128	A	G	1.26	0.00000561	GULP1 (intron)
20	40492611	rs35425049	T	C	1.643	0.0000068	* RP1-191L6.2-MAFB
14	22329352	rs1555350954	T	C	1.268	0.00000936	* TRAV41-LOC105370401

\* Denotes intergenic SNPs

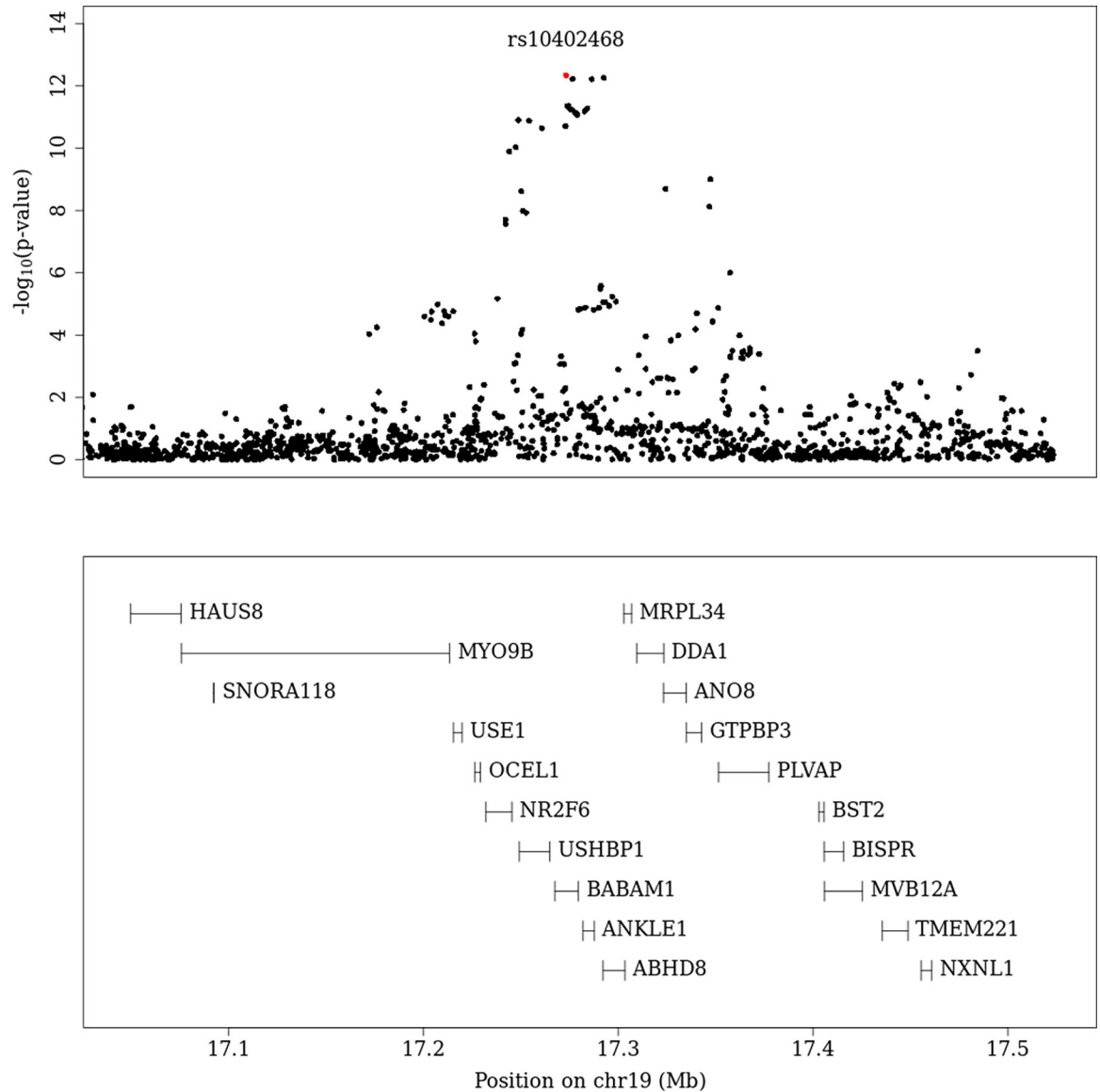
<https://doi.org/10.1371/journal.pgen.1010609.t002>

[22]. While rs2239785 was not in the available data, we find a perfect proxy in rs136175 ( $r^2 = 1.0$ ,  $D' = 1.0$  with rs2239785 in EUR) has a similar direction of effect (OR = 1.32,  $P = 0.09$ ;  $N = 270$  cases, 435 controls). When combining the aforementioned studies in a meta-analysis for this locus, we see  $OR_{meta} = 1.30$ ,  $P_{meta} = 1.30 \times 10^{-20}$ . Therefore, we are able to replicate the observed association between DME (or closely related phenotypes) and rs2239785, an amino acid changing SNP in *APOL1*, in three further studies.

Conditional analysis did not identify secondary signals at the locus in EUR; however SNP rs6518994 retained significance in the AFR sample after conditioning on rs2239785 in the primary GWAS dataset (original OR = 1.92,  $P = 2 \times 10^{-4}$ ; conditioned OR = 1.74,  $P = 0.002$ ;  $r^2$  with rs2239785 = 0.04). Further conditional analysis incorporating rs6518994 did not reveal any novel signals with  $P < 0.01$ .

Due to the use of non-DME controls in this study for whom diabetes was not ascertained, it was important to establish whether the association we observed was driven diabetes status, rather than DME itself. We employed multiple approaches to exclude this. First, in the FinnGen PheWAS, we did not observe association with rs2239785 and T1D or T2D status (both



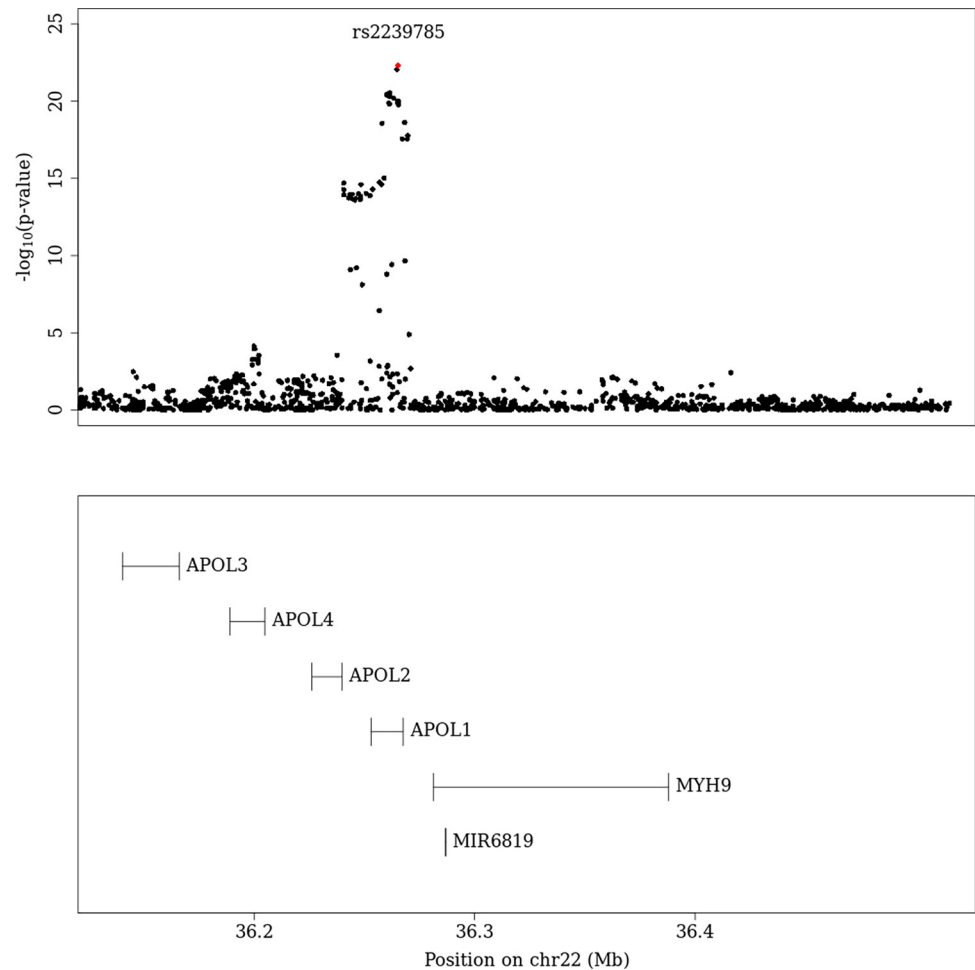


**Fig 2. Regional plot (meta-analysis) for chr19.**

<https://doi.org/10.1371/journal.pgen.1010609.g002>

$P > 0.53$ ). Furthermore, we queried recent highly powered GWAS for type-1 (T1D) [23] and a multi-ancestry type-2 (T2D) diabetes finding no association for rs2239785 in either (both  $P > 0.4$ ). Further investigation in the EUR and AFR patient subsets of the T2D multi-ancestry study did not alter this (both  $P > 0.5$ ). Thus, we find this association is not driven by diabetes predisposition itself.

Multiple variants in *APOLI*, including rs2239785, have been associated in a recessive model with diabetic and non-diabetic nephropathy in African American individuals [21,24–28]. These findings identified an AFR specific haplotype tagged by two SNPs (rs73885319, encoding S342G, and rs60910145, encoding I284M (G1)) and a 6-bp deletion (rs75781853 (G2)) associated with disease risk [24]. G1 and G2, found in AFR only, are not strongly correlated ( $r^2 < 0.18$ ) but reside on the same haplotype block ( $D' = 1.0$ ) with rs2239785. Due to this haplotypic structure, it is difficult to disentangle the effect of rs2239785, G1 and G2 in AFR, and

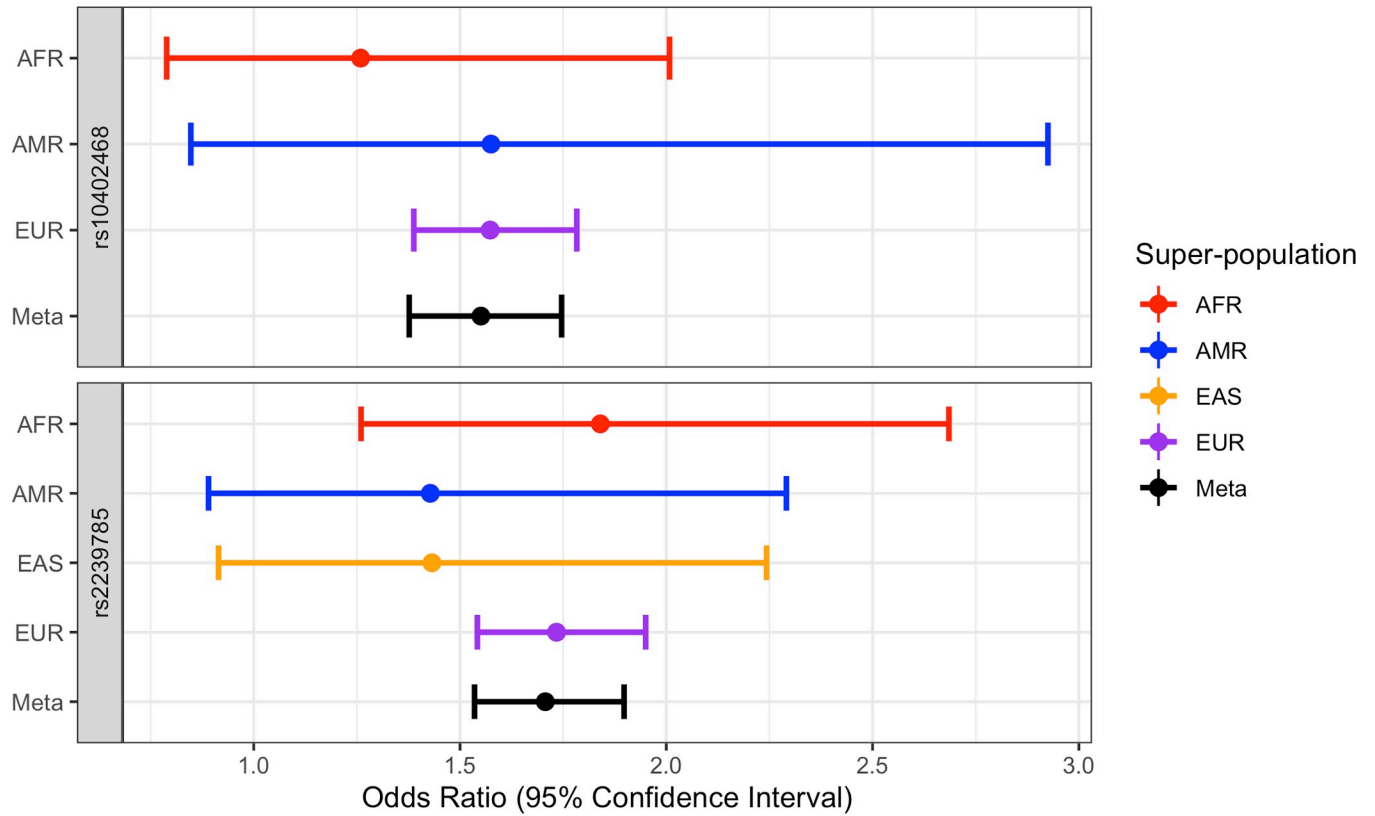


**Fig 3. Regional plot (meta-analysis) for chr22.** As the meta-analysis contains patient DNA samples from multiple ancestries, linkage disequilibrium information is not detailed here.

<https://doi.org/10.1371/journal.pgen.1010609.g003>

raises the possibility that the DME association we observed could be confounded to some degree by nephropathy. The G1 and G2 *APOL1* variants are not found in individuals with European ancestry in our cohort (S4 Table). As such, the European population enables investigation of rs2239785 (K150E) on its own, and whether this SNP affects kidney disease.

First, we see repeated evidence of association between rs2239785 and risk of DME in individuals of European descent: the RISE, RIDE, YOSEMITE and RHINE trials in the discovery dataset, the PAGODA trial in the replication dataset and the FinnGen BioBank dataset. Second, in FinnGen, despite the association between rs2239785 and diabetic maculopathy, we found no evidence of association between rs2239785 and renal dysfunction in T1D or T2D patients (both  $P > 0.37$ ). Furthermore, we were unable to find an association between this SNP and DKD in recent GWAS conducted in European ancestry patients with either T1D or T2D (both  $P > 0.71$ ) [29,30]. Additionally, we saw no association in a large meta-analysis for CKD and kidney function (BUN, eGFR) in European patients (all  $P > 0.22$ ) [31]. While renal failure or anticipated need for dialysis were exclusion criteria in the DME clinical trials from which these cases were obtained, a DKD diagnosis was not. There were kidney measurements for 912 of the DME patients in the YOSEMITE and RHINE trials. As a final test, we found that removing 290 patients from the 912 of whom had albumin to creatinine ratio (S5 Table), blood urea



SNP	OR	P	DME EAF	Control EAF
rs10402468 (EA=A)				
AFR	1.26	0.33	0.15	0.12
AMR	1.57	0.16	0.09	0.05
EUR	1.57	1.41x10 <sup>-12</sup>	0.24	0.17
Meta	1.55	4.62x10 <sup>-13</sup>		
rs2239785 (EA=G)				
AFR	1.84	1.58x10 <sup>-3</sup>	0.76	0.62
AMR	1.43	0.14	0.19	0.13
EAS	1.43	0.12	0.24	0.17
EUR	1.73	4.06x10 <sup>-20</sup>	0.31	0.20
Meta	1.71	5.05x10 <sup>-23</sup>		

Fig 4. Forest plot and summary statistics for the top SNP at chr 22 and chr 19 loci.

<https://doi.org/10.1371/journal.pgen.1010609.g004>

nitrogen (BUN) or protein levels outside of the ideal range had no effect on the association with DME in the remaining 622 patients with normal kidney function ( $OR_{meta} = 1.79$ ,  $P_{meta} = 7.91 \times 10^{-17}$ ;  $OR_{eur} = 1.89$ ,  $P_{eur} = 1.25 \times 10^{-16}$ ;  $OR_{afr} = 1.76$ ,  $P_{afr} = 0.03$ ).

These data are consistent with a model whereby the G1 and G2 *APOLI* variants drive DKD susceptibility in individuals of African descent, but that rs2239785 (or another allele with which it is in linkage disequilibrium) represents an independent, and distinct, association that contributes directly to DME predisposition across ancestries.

### Chr19p13 –*PLVAP* and *ANKLE1*

The second genome-wide significant peak in this meta-analysis was located at chr19p13.11 (**S6 Table**) with the top SNP, rs10402468, located within intron 3 of *BABAM1* (A allele— $OR_{meta} = 1.55$ ,  $P_{meta} = 4.62 \times 10^{-13}$ ). The effect size and direction was similar across the super-populations tested ( $OR_{eur} = 1.57$ ,  $P_{eur} = 1.42 \times 10^{-12}$ ;  $OR_{afr} = 1.26$ ,  $P_{afr} = 0.33$ ;  $OR_{amr} = 1.58$ ,  $P_{amr} = 0.15$ ; monomorphic in EAS (**Fig 4**). The risk allele was most common in EUR ( $EAF_{eur} = 0.25$ ). We were able to replicate rs10402468 in an independent set of DME cases from the PAGODA trial and non-DME controls ( $OR_{meta} = 1.31$ ,  $P_{meta} = 0.03$ ;  $OR_{eur} = 1.34$ ,  $P_{eur} = 0.03$ ;  $OR_{afr} = 1.10$ ,  $P_{afr} = 0.79$ ) (**Table 1**), with a combined discovery and replication  $OR = 1.50$  and  $P = 1.0 \times 10^{-13}$  (**Table 3**). Conditional analysis did not identify a secondary signal at this locus. We derived the 95% credible set of variants for fine-mapping of the signal, identifying 4 potentially causal variants at the locus with 95% probability (**S7 Table**). Finally, we performed PheWAS in the aforementioned FinnGen study. While there was no association for diabetic maculopathy ( $P = 0.11$ ), there were several associations with ocular phenotypes when examining the locus at chr19p13.11 (rs10402468). The top phenotype in the PheWAS is macular cyst ( $OR = 1.42$ ,  $P = 4.6 \times 10^{-5}$ ,  $MAF_{cases} = 0.31$ ,  $MAF_{controls} = 0.23$ ;  $N = 297$  cases, 284,826 controls), and there are also strong associations with wet age-related macular degeneration (AMD) and “degeneration of macula and posterior pole” ( $P = 1.3 \times 10^{-4}$  and  $P = 2.1 \times 10^{-4}$  respectively) (**S8 Table**). The AMD association did not replicate in a recent AMD GWAS ( $P = 0.25$ ; wet AMD vs controls) [32].

To investigate the possibility that this locus could be associated with diabetes status rather than DME we performed similar analyses as described for the locus at chr22. First, in the FinnGen PheWAS, we did not observe an association with T1D or T2D (both  $P > 0.42$ ). Second, there was no association for rs10402468 in large T1D [23] or T2D [33] GWAS studies (both  $P > 0.3$ ). Further investigation in T2D in the EUR patient subset of the multi-ancestry study did not alter this ( $P > 0.4$ ). Thus, we find this association is specific to DME, and not due to diabetes status.

This SNP, rs10402468, is in strong LD ( $r^2 = 1.0$ ) with a missense SNP in *ANKLE1* in all super-populations (rs10425939, A505V). We found that the peak co-localized with cis-eQTL peaks from bulk RNA-seq data [16–18] to *PLVAP* and *ANKLE1* in vascular/endothelium tissues (*ANKLE1*  $Prob_{coloc}$  artery aorta = 0.96; *PLVAP* artery coronary = 0.88, artery tibial = 0.60) (**S5A–S5C Fig**). Thus, we identify both *PLVAP*, encoding plasmalemma vesicle associated protein, and *ANKLE1*, encoding ankyrin repeat and LEM domain-containing protein 1 as candidate effectors at the locus.

**Table 3. Replication of independent genome-wide significant hits.**

					Discovery	Replication	Meta-Analysis
SNP	CHR	BP	A1	A2	OR (P)	OR (P)	OR (P)
rs10402468	19	17273194	A	C	1.55 (4.6E-13)	1.31 (0.03)	1.50 (1E-13)
rs2239785	22	36265284	G	A	1.71 (5E-23)	1.41 (0.003)	1.65 (2E-24)

<https://doi.org/10.1371/journal.pgen.1010609.t003>

**Table 4. Population attributable fraction for chr22 and chr19 loci by ancestry.**

Locus / Haplotype	Cohort	PARP
rs2239785	EUR	1.85
rs2239785	AFR	5.16
rs10402468	EUR	1.03
rs10402468	AFR	0.74
GG (EIK) Haplotype	AFR	4.35

\* PARP = population attributable risk percent; AFR = African descent; EUR = European descent

<https://doi.org/10.1371/journal.pgen.1010609.t004>

### Population attributable fraction at the *APOL1* locus in AFR

The increased effect allele frequency of rs2239785 in AFR indicates a greater overall burden in this patient population. As such, we calculated the fraction of DME risk in patients with diabetes that is attributable to the *APOL1* locus (Table 4). DME prevalence in United States patients with diabetes for non-Hispanic white and non-Hispanic black individuals is estimated at 2.6% and 8.4% respectively [3]. We estimate that rs2239785 accounts for 5.2% of the risk of developing DME in non-Hispanic black patients with diabetes compared to 1.9% in non-Hispanic white patients. As previously mentioned, the G1 and G2 variants are only seen in AFR individuals and are on the same haplotype as rs2239785. We hypothesized that the population attributable fraction for rs2239785 in AFR who do not possess either the G1 or G2 haplotypes would be similar to that of EUR. Looking at this subsample in AFR, we estimate that rs2239785 accounts for 1.60% of the risk of developing DME in non-Hispanic black patients with diabetes. Thus, it appears the haplotype containing the rs2239785 risk allele (E150) itself, without the presence of G1 and G2 in AFR (most likely a result of recombination with the European rs2239785-only haplotype in the admixed African American population used in this study), accounts for a similar proportion of overall population risk in individuals of EUR and AFR ancestry. This provides further evidence that rs2239785 is a separate, DME specific locus which increases risk of DME regardless of presence of G1 and G2 in patients of African ancestry.

### Association with baseline DME clinical and imaging measures

We sought to determine if these two loci impact DME clinical and imaging measures at baseline. We did not see an association for the *APOL1* and *ANKLE1/PLVAP* SNPs with baseline Diabetic Retinopathy Severity Score (DRSS) ( $N = 1,168$ ;  $P > 0.1$ ). Nor was there any association of the *APOL1* SNP with three retinal measures derived from Optical Coherence Tomography (OCT) data: (1) center subfield thickness (CST) ( $N = 1,339$ ), (2) intraretinal fluid in central subfield ( $N = 1,020$ ), and (3) subretinal fluid ( $N = 387$ ). However, we did observe an association between CST and the *ANKLE1/PLVAP* SNP ( $\beta = 19.3$ ;  $P = 0.006$ ). Thus, it appears the association of the risk allele with greater CST is consistent with previous reports of CST being higher in severe DME patients as well as when comparing DME patients to controls [34–36].

### Discussion

Despite multiple GWAS for DR and subtypes performed to date, few replicable associations have been reported. In this analysis we focused on the DME subtype to increase power by reducing disease heterogeneity.

In this multi-ethnic common variant GWAS, we report two novel loci for risk of DME. Both loci achieve genome-wide significance in the discovery dataset and replication in both an

additional multi-ancestry DME clinical trial patient sample, and the FinnGen biobank study. We observed consistent effect sizes and directions across ancestries in both the discovery and replication datasets at both loci.

We also observed super-population differences in disease heritability, where the SNP-heritability was 0.41 in EUR and 0.48 in AFR. This was partially due to variation at the *APOL1* locus in AFR, where the effect allele frequency for the top SNP was over two times that of EUR. This increased allele frequency at the *APOL1* locus resulted in an estimated 5% population attributable fraction of DME in non-Hispanic black patients with diabetes, over two times that of non-Hispanic white patients, despite the similar effect size for the SNP in each super-population. This indicates that non-Hispanic black patients with DME could preferentially benefit from therapies designed around APOL1, however, follow-up studies with increased sample size are warranted.

To date there are no published reports of APOL1 staining in the eye, but scRNA seq data indicates expression in immune and vascular cells of the retina and choroid endothelia [37]. The development of macular edema requires a greater rate of fluid entry from leaky vessels into the retina than that of fluid reabsorption [38]. Given that *APOL1* renal risk variants expressed under an endothelial promoter in mice caused vascular leak in the skin [39], one possible mechanism is that APOL1 could similarly exacerbate vascular leak in the retinal or choroidal endothelia, directly resulting in edema. Alternatively, the leaky vasculature could allow infiltration of cytokines into the eye, which would upregulate APOL1 beyond a critical threshold in the cells where it is expressed, causing cytotoxicity [40]. Another possibility is that the cation channel activity of APOL1 (e.g. K<sup>+</sup> efflux) could interfere with K<sup>+</sup> co-transporter mediated water reabsorption by Müller or retinal pigment epithelial cells, contributing to edema [38,41]. Determining the localization of APOL1 protein in DME eyes would help narrow down the possibilities.

The second locus identified includes the genes *PLVAP* and *ANKLE1*. PLVAP, plasma-lemma vesicle-associated protein, is also called PV-1 and recognized by the PAL-E antibody. PLVAP is an endothelial protein that regulates basal permeability. In the eye, PAL-E staining is specific to the fenestrated endothelium of the choriocapillaris and ciliary processes [42]. *Plvap* deficient mice develop edema in some vascular beds [43] and loss of Plvap has been shown to disrupt the choriocapillaris structure and lead to retinal degeneration [44]. As such, these murine studies are consistent with our results suggesting lower *PLVAP* mRNA levels are a risk factor for DME. However, *PLVAP* expression is also seen at the blood retinal barrier (BRB), but only after insult by factors that induce leakage, such as diabetic retinopathy [45]. In a properly functioning BRB, *PLVAP* expression is low or non-existent [46]. This is contrary to our results and further studies will be needed to associate the risk allele with differential PLVAP expression and function.

*ANKLE1* (which encodes the protein LEM3) was also identified at this locus by our co-localization analysis. LEM domain proteins, such as LEM3 bind BAF, a chromatin protein. *ANKLE1* has been shown in *C. elegans* to be important in the formation and stabilization of chromatin bridges in the end stages of cellular division [47]. Expression appears somewhat ubiquitous with enrichment in brain and eye tissue [48], however mechanistic information on *ANKLE1* and retina or eye is absent from the literature. While further perturbation studies will be required to assess causality, based on what is currently known, *PLVAP* is the most likely causal gene at this locus.

Notably, we observed no prior DME, DR or PDR loci at the GWAS significant level in this DME analysis (S9 Table) [8,10,22]. This could be for multiple reasons. First, this study contains at least three times more cases than previous DME studies. Additionally, other studies have focused on DR patients as a whole. As DR is a heterogeneous condition, homogenizing based on phenotypes within DR can increase power to find an association.

A weakness of our study is the use of controls from non-ophthalmologic clinical trials. This has the potential to lead to associations that are confounded by diabetes status or other diabetic complications such as kidney disease. This is especially important at the *APOL1* locus, for which multiple alleles in *APOL1* have strong associations to risk of kidney disease in individuals of African descent. We have addressed this in several ways, including leveraging genetic architecture differences at the *APOL1* locus in individuals of European and African descent. We queried highly powered GWAS for T1D and T2D, finding no hint of an association for either locus in patients of European ancestry. For the *APOL1* locus we queried DKD, CKD and kidney function GWAS, again finding no association in European ancestry patients. Through PheWAS in the FinnGen population, we found strong associations for ocular complications due to diabetes, but not T1D, T2D or kidney disease. Finally, we conducted a sensitivity analysis, where we removed patients with signs of kidney malfunction at baseline finding no effect on the association at the *APOL1* locus. Based on these efforts, these findings appear to be specific for DME risk in multiple populations and not related to the concomitant diagnosis of diabetes or nephropathy, however, as for any genetic association, there remains the possibility that the findings in this study are the consequence of confounding by other unmeasured factors. Historically, the standard design when assessing determinants of complications from diabetes has been to compare patients with complications from diabetes to patients with diabetes and no complications. Our findings question the need for such a constraint on study design, especially when considering there are multiple highly powered GWAS one can query to detect possible confounding due to diabetes risk itself or other co-morbid complications.

In summary, here we present results from the largest GWAS on DME to date which highlights *APOL1*, a gene canonically associated with kidney disease, and broadens its impact to a second diabetic complication. We find this gene to be associated with risk of DME in multiple populations, and note a substantial increased genetic risk of DME in non-Hispanic black patients with diabetes due to the increased allele frequency of rs2239785 in patients of African ancestry. Additionally, we uncovered an additional locus containing the genes *ANKLE1* and *PLVAP*. Historically, replicable loci for diabetic retinopathy and sub-phenotypes have been difficult to uncover. The present study indicates that power to detect genetic effects influencing diabetic ocular phenotypes can be improved by categorizing DR cases by sub-phenotype prior to genetic analysis.

## Supporting information

**S1 Table. Plot of admixture proportions by superpopulation.** \* AFR = African descent; AMR = Admixed American descent; EAS = East Asian descent; EUR = European descent. \*\* Red = proportion of AFR ancestry; Blue = proportion of AMR ancestry; Orange = proportion of EAS ancestry; Grey = proportion of EUR ancestry; Yellow = proportion of SAS ancestry. (XLSX)

**S2 Table. QQ plot for discovery analysis results.** (XLSX)

**S3 Table. Ancestry specific manhattan plots for the discovery cohort.** \* AFR = African descent; AMR = Admixed American descent; EAS = East Asian descent; EUR = European descent. (XLSX)

**S4 Table. Regional association plots for ancestry groups in the discovery cohort.** \* AFR = African descent; AMR = Admixed American descent; EAS = East Asian descent;

EUR = European descent.  
(XLSX)

**S5 Table. Colocalization of discovery analysis with eQTL data.**  
(XLSX)

**S6 Table. Annotation of SNPs in Chr 19 locus.**  
(XLSX)

**S7 Table. Credible Sets for genome-wide significant chr19.**  
(XLSX)

**S8 Table. FinnGen PheWAS results for rs10402468.**  
(XLSX)

**S9 Table. Summary statistics for GWAS Catalog hits for diabetic retinopathy and diabetic macular edema.**  
(XLSX)

**S1 Fig. Plot of admixture proportions by superpopulation.**  
(XLSX)

**S2 Fig. QQ plot from the discovery analysis.**  
(XLSX)

**S3 Fig. Ancestry specific manhattan plots for the discovery cohort.**  
(XLSX)

**S4 Fig. Regional association plots for ancestry groups in the discovery cohort.**  
(XLSX)

**S5 Fig. Colocalization of discovery analysis with eQTL data.**  
(XLSX)

## Acknowledgments

We would like to acknowledge Hsu-Hsin Chen and Luz Orozco for their efforts to identify APOL1 expression in retina. We would like to acknowledge the efforts of Marie Waldvogel, Aurelie Felez and Nicole Bienz in the Pharma Biosample Services (PBS) Department of F. Hoffmann-La Roche for their efforts on handling and genotyping the patient samples used in this study. We also would like to acknowledge the participants and investigators of FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme Corp, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Jansen Biotech Inc, Novartis AG, and Boehringer Ingelheim. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank ([www.auria.fi/biopankki](http://www.auria.fi/biopankki)), THL Biobank ([www.thl.fi/biobank](http://www.thl.fi/biobank)), Helsinki Biobank ([www.helsinginbiopankki.fi](http://www.helsinginbiopankki.fi)), Biobank Borealis of Northern Finland (<https://www.ppsbp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx>), Finnish Clinical Biobank Tampere ([www.tays.fi/en-US/Research\\_and\\_development/Finnish\\_Clinical\\_Biobank\\_Tampere](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Biobank of Eastern Finland ([www.ita-suomenbiopankki.fi/en](http://www.ita-suomenbiopankki.fi/en)), Central Finland Biobank ([www.ksshp.fi/fi-FI/Potilaalle/Biopankki](http://www.ksshp.fi/fi-FI/Potilaalle/Biopankki)), Finnish Red Cross Blood Service Biobank ([www.veripalvelu.fi/](http://www.veripalvelu.fi/)



[verenluovutus/biopankkitoiminta](https://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/)) and Terveystalo Biobank ([www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/](http://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/)). All Finnish Biobanks are members of BBMRI.fi infrastructure ([www.bbmri.fi](http://www.bbmri.fi)). Finnish Biobank Cooperative -FINBB (<https://finbb.fi/>) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious services (<https://site.fingenious.fi/en/>) managed by FINBB.

## Author Contributions

**Conceptualization:** Amy D. Stockwell, Brian L. Yaspan.

**Data curation:** Amy D. Stockwell, Jens Reeder, Eric Wei.

**Formal analysis:** Amy D. Stockwell.

**Funding acquisition:** Michael C. Chang.

**Investigation:** Amy D. Stockwell, Anubha Mahajan, Neha Anegondi, Suzie J. Scales, Mark I. McCarthy, Brian L. Yaspan.

**Methodology:** Amy D. Stockwell, William Forrest, Jens Reeder, Brian L. Yaspan.

**Project administration:** Natalie M. Creps.

**Resources:** Michael C. Chang, Rion K. Pendergrass, Suresh Selvaraj, Jens Reeder, Eric Wei, Victor A. Iglesias, Natalie M. Creps, Laura Macri, Andrea N. Neeranjana, Marcel P. van der Brug.

**Supervision:** Mark I. McCarthy.

**Validation:** Amy D. Stockwell, Brian L. Yaspan.

**Visualization:** Amy D. Stockwell, Brian L. Yaspan.

**Writing – original draft:** Amy D. Stockwell, Brian L. Yaspan.

**Writing – review & editing:** Suzie J. Scales, Mark I. McCarthy.

## References

1. Ruta LM, Magliano DJ, Lemesurier R, Taylor HR, Zimmet PZ, Shaw JE. Prevalence of diabetic retinopathy in Type 2 diabetes in developing and developed countries. *Diabet Med*. 2013; 30(4):387–98. Epub 2013/01/22. <https://doi.org/10.1111/dme.12119> PMID: 23331210.
2. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*. 2012; 35(3):556–64. Epub 2012/02/04. <https://doi.org/10.2337/dc11-1909> PMID: 22301125; PubMed Central PMCID: PMC3322721.
3. Varma R, Bressler NM, Doan QV, Gleeson M, Danese M, Bower JK, et al. Prevalence of and risk factors for diabetic macular edema in the United States. *JAMA Ophthalmol*. 2014; 132(11):1334–40. Epub 2014/08/16. <https://doi.org/10.1001/jamaophthalmol.2014.2854> PMID: 25125075; PubMed Central PMCID: PMC4576994.
4. Klein R, Klein BE, Moss SE, Cruickshanks KJ. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: XVII. The 14-year incidence and progression of diabetic retinopathy and associated risk factors in type 1 diabetes. *Ophthalmology*. 1998; 105(10):1801–15. Epub 1998/10/27. [https://doi.org/10.1016/S0161-6420\(98\)91020-X](https://doi.org/10.1016/S0161-6420(98)91020-X) PMID: 9787347.
5. Klein R, Klein BE, Moss SE, Cruickshanks KJ. Relationship of hyperglycemia to the long-term incidence and progression of diabetic retinopathy. *Arch Intern Med*. 1994; 154(19):2169–78. Epub 1994/10/10. PMID: 7944837.
6. Klein R, Klein BE, Moss SE, Linton KL. The Beaver Dam Eye Study. Retinopathy in adults with newly discovered and previously diagnosed diabetes mellitus. *Ophthalmology*. 1992; 99(1):58–62. Epub 1992/01/01. [https://doi.org/10.1016/s0161-6420\(92\)32011-1](https://doi.org/10.1016/s0161-6420(92)32011-1) PMID: 1741141.
7. Cho H, Sobrin L. Genetics of diabetic retinopathy. *Curr Diab Rep*. 2014; 14(8):515. Epub 2014/06/22. <https://doi.org/10.1007/s11892-014-0515-z> PMID: 24952107; PubMed Central PMCID: PMC4125976.

8. Liu C, Chen G, Bentley AR, Doumatey A, Zhou J, Adeyemo A, et al. Genome-wide association study for proliferative diabetic retinopathy in Africans. *NPJ Genom Med*. 2019; 4:20. Epub 2019/09/05. <https://doi.org/10.1038/s41525-019-0094-7> PMID: 31482010; PubMed Central PMCID: PMC6715701.
9. Cooke Bailey JN, Sobrin L, Pericak-Vance MA, Haines JL, Hammond CJ, Wiggs JL. Advances in the genomics of common eye diseases. *Hum Mol Genet*. 2013; 22(R1):R59–65. Epub 2013/08/22. <https://doi.org/10.1093/hmg/ddt396> PMID: 23962718; PubMed Central PMCID: PMC3782072.
10. Pollack S, Igo RP Jr., Jensen RA, Christiansen M, Li X, Cheng CY, et al. Multiethnic Genome-Wide Association Study of Diabetic Retinopathy Using Liability Threshold Modeling of Duration of Diabetes and Glycemic Control. *Diabetes*. 2019; 68(2):441–56. Epub 2018/11/30. <https://doi.org/10.2337/db18-0567> PMID: 30487263; PubMed Central PMCID: PMC6341299.
11. Ho D, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*. 2011; 42(8):1–28. <https://doi.org/10.18637/jss.v042.i08>
12. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. Epub 2007/08/19. <https://doi.org/10.1086/519795> PMID: 17701901; PubMed Central PMCID: PMC1950838.
13. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26(17):2190–1. Epub 2010/07/10. <https://doi.org/10.1093/bioinformatics/btq340> PMID: 20616382; PubMed Central PMCID: PMC2922887.
14. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015; 47(3):291–5. Epub 2015/02/03. <https://doi.org/10.1038/ng.3211> PMID: 25642630; PubMed Central PMCID: PMC4495769.
15. Nagraj V. twotwo: Work with Two-by-Two Tables. R package 2021 [cited 2022 03/01/2022]. 0.1.0. [Available from: <https://CRAN.R-project.org/package=twotwo>.
16. Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020; 369(6509):1318–30. Epub 2020/09/12. <https://doi.org/10.1126/science.aaz1776> PMID: 32913098; PubMed Central PMCID: PMC7737656.
17. Orozco LD, Chen HH, Cox C, Katschke KJ Jr., Arceo R, Espiritu C, et al. Integration of eQTL and a Single-Cell Atlas in the Human Eye Identifies Causal Genes for Age-Related Macular Degeneration. *Cell Rep*. 2020; 30(4):1246–59 e6. Epub 2020/01/30. <https://doi.org/10.1016/j.celrep.2019.12.082> PMID: 31995762.
18. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018; 558(7708):73–9. Epub 2018/06/08. <https://doi.org/10.1038/s41586-018-0175-2> PMID: 29875488; PubMed Central PMCID: PMC6697541.
19. Tom J, Chang D, Wuster A, Mukhyala K, Cuenco K, Cowgill A, et al. Enabling genome-wide association testing with multiple diseases and no healthy controls. *Gene*. 2019; 684:118–23. Epub 2018/10/27. <https://doi.org/10.1016/j.gene.2018.10.047> PMID: 30366082.
20. Lannon H, Shah SS, Dias L, Blackler D, Alper SL, Pollak MR, et al. Apolipoprotein L1 (APOL1) risk variant toxicity depends on the haplotype background. *Kidney Int*. 2019; 96(6):1303–7. Epub 2019/10/16. <https://doi.org/10.1016/j.kint.2019.07.010> PMID: 31611067; PubMed Central PMCID: PMC6907738.
21. Bailey JNC, Palmer ND, Ng MCY, Bonomo JA, Hicks PJ, Hester JM, et al. Analysis of coding variants identified from exome sequencing resources for association with diabetic and non-diabetic nephropathy in African Americans. *Hum Genet*. 2014; 133(6):769–79. Epub 2014/01/05. <https://doi.org/10.1007/s00439-013-1415-z> PMID: 24385048; PubMed Central PMCID: PMC4024071.
22. Graham PS, Kaidonis G, Abhary S, Gillies MC, Daniell M, Essex RW, et al. Genome-wide association studies for diabetic macular edema and proliferative diabetic retinopathy. *BMC Med Genet*. 2018; 19(1):71. Epub 2018/05/10. <https://doi.org/10.1186/s12881-018-0587-8> PMID: 29739359; PubMed Central PMCID: PMC5941644.
23. Chiou J, Geusz RJ, Okino ML, Han JY, Miller M, Melton R, et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature*. 2021; 594(7863):398–402. Epub 2021/05/21. <https://doi.org/10.1038/s41586-021-03552-w> PMID: 34012112.
24. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*. 2010; 329(5993):841–5. Epub 2010/07/22. <https://doi.org/10.1126/science.1193032> PMID: 20647424; PubMed Central PMCID: PMC2980843.
25. Genovese G, Tonna SJ, Knob AU, Appel GB, Katz A, Bernhardt AJ, et al. A risk allele for focal segmental glomerulosclerosis in African Americans is located within a region containing APOL1 and MYH9. *Kidney Int*. 2010; 78(7):698–704. Epub 2010/07/30. <https://doi.org/10.1038/ki.2010.251> PMID: 20668430; PubMed Central PMCID: PMC3001190.

26. Tzur S, Rosset S, Shemer R, Yudkovsky G, Selig S, Tarekegn A, et al. Missense mutations in the APOL1 gene are highly associated with end stage kidney disease risk previously attributed to the MYH9 gene. *Hum Genet.* 2010; 128(3):345–50. Epub 2010/07/17. <https://doi.org/10.1007/s00439-010-0861-0> PMID: 20635188; PubMed Central PMCID: PMC2921485.
27. Friedman DJ, Kozlitina J, Genovese G, Jog P, Pollak MR. Population-based risk assessment of APOL1 on renal disease. *J Am Soc Nephrol.* 2011; 22(11):2098–105. Epub 2011/10/15. <https://doi.org/10.1681/ASN.2011050519> PMID: 21997396; PubMed Central PMCID: PMC3231785.
28. Friedman DJ, Pollak MR. APOL1 Nephropathy: From Genetics to Clinical Applications. *Clin J Am Soc Nephrol.* 2021; 16(2):294–303. Epub 2020/07/04. <https://doi.org/10.2215/CJN.15161219> PMID: 32616495; PubMed Central PMCID: PMC7863644.
29. Salem RM, Todd JN, Sandholm N, Cole JB, Chen WM, Andrews D, et al. Genome-Wide Association Study of Diabetic Kidney Disease Highlights Biology Involved in Glomerular Basement Membrane Collagen. *J Am Soc Nephrol.* 2019; 30(10):2000–16. Epub 2019/09/21. <https://doi.org/10.1681/ASN.2019030218> PMID: 31537649; PubMed Central PMCID: PMC6779358.
30. van Zuydam NR, Ahlqvist E, Sandholm N, Deshmukh H, Rayner NW, Abdalla M, et al. A Genome-Wide Association Study of Diabetic Kidney Disease in Subjects With Type 2 Diabetes. *Diabetes.* 2018; 67(7):1414–27. Epub 2018/04/29. <https://doi.org/10.2337/db17-0914> PMID: 29703844; PubMed Central PMCID: PMC6014557.
31. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019; 51(6):957–72. Epub 2019/06/04. <https://doi.org/10.1038/s41588-019-0407-x> PMID: 31152163; PubMed Central PMCID: PMC6698888.
32. Fritsche LG, Igl W, Bailey JN, Grassmann F, Sengupta S, Bragg-Gresham JL, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet.* 2016; 48(2):134–43. Epub 2015/12/23. <https://doi.org/10.1038/ng.3448> PMID: 26691988; PubMed Central PMCID: PMC4745342.
33. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet.* 2022; 54(5):560–72. Epub 2022/05/14. <https://doi.org/10.1038/s41588-022-01058-3> PMID: 35551307; PubMed Central PMCID: PMC9179018.
34. Hannouche RZ, Avila MP, Isaac DL, Silva RS, Rassi AR. Correlation between central subfield thickness, visual acuity and structural changes in diabetic macular edema. *Arq Bras Oftalmol.* 2012; 75(3):183–7. Epub 2012/08/09. <https://doi.org/10.1590/s0004-27492012000300007> PMID: 22872201.
35. Kwon JW, Kim B, Jee D, Cho YK. Aqueous humor analyses of diabetic macular edema patients with subretinal fluid. *Sci Rep.* 2021; 11(1):20985. Epub 2021/10/27. <https://doi.org/10.1038/s41598-021-00442-z> PMID: 34697354; PubMed Central PMCID: PMC8546094.
36. Halim MS, Afridi R, Hasanreisoglu M, Hassan M, Ibrahim-Ahmed M, Do DV, et al. Differences in the characteristics of subjects achieving complete, partial, or no resolution of macular edema in the READ-3 study. *Graefes Arch Clin Exp Ophthalmol.* 2021; 259(10):2941–8. Epub 2021/04/02. <https://doi.org/10.1007/s00417-021-05148-6> PMID: 33792788.
37. Cowan CS, Renner M, De Gennaro M, Gross-Scherf B, Goldblum D, Hou Y, et al. Cell Types of the Human Retina and Its Organoids at Single-Cell Resolution. *Cell.* 2020; 182(6):1623–40 e34. Epub 2020/09/19. <https://doi.org/10.1016/j.cell.2020.08.013> PMID: 32946783; PubMed Central PMCID: PMC7505495.
38. Reichenbach A, Wurm A, Pannicke T, Iandiev I, Wiedemann P, Bringmann A. Muller cells as players in retinal degeneration and edema. *Graefes Arch Clin Exp Ophthalmol.* 2007; 245(5):627–36. Epub 2007/01/16. <https://doi.org/10.1007/s00417-006-0516-y> PMID: 17219109.
39. Wu J, Ma Z, Raman A, Beckerman P, Dhillon P, Mukhi D, et al. APOL1 risk variants in individuals of African genetic ancestry drive endothelial cell defects that exacerbate sepsis. *Immunity.* 2021; 54(11):2632–49 e6. Epub 2021/10/30. <https://doi.org/10.1016/j.immuni.2021.10.004> PMID: 34715018; PubMed Central PMCID: PMC9338439.
40. Datta S, Kataria R, Zhang JY, Moore S, Petitpas K, Mohamed A, et al. Kidney Disease-Associated APOL1 Variants Have Dose-Dependent, Dominant Toxic Gain-of-Function. *J Am Soc Nephrol.* 2020; 31(9):2083–96. Epub 2020/07/18. <https://doi.org/10.1681/ASN.2020010079> PMID: 32675303; PubMed Central PMCID: PMC7461666.
41. Olabisi OA, Heneghan JF. APOL1 Nephrotoxicity: What Does Ion Transport Have to Do With It? *Semin Nephrol.* 2017; 37(6):546–51. Epub 2017/11/08. <https://doi.org/10.1016/j.semnephrol.2017.07.008> PMID: 29110762; PubMed Central PMCID: PMC5726439.
42. Bosma EK, van Noorden CJF, Schlingemann RO, Klaassen I. The role of plasmalemma vesicle-associated protein in pathological breakdown of blood-brain and blood-retinal barriers: potential novel

- therapeutic target for cerebral edema and diabetic macular edema. *Fluids Barriers CNS*. 2018; 15(1):24. Epub 2018/09/21. <https://doi.org/10.1186/s12987-018-0109-2> PMID: 30231925; PubMed Central PMCID: PMC6146740.
43. Herrnberger L, Seitz R, Kuespert S, Bosl MR, Fuchshofer R, Tamm ER. Lack of endothelial diaphragms in fenestrae and caveolae of mutant Plvap-deficient mice. *Histochem Cell Biol*. 2012; 138(5):709–24. Epub 2012/07/12. <https://doi.org/10.1007/s00418-012-0987-3> PMID: 22782339.
  44. Kim SA, Kim SJ, Choi YA, Yoon HJ, Kim A, Lee J. Retinal VEGFA maintains the ultrastructure and function of choriocapillaris by preserving the endothelial PLVAP. *Biochem Biophys Res Commun*. 2020; 522(1):240–6. Epub 2019/11/25. <https://doi.org/10.1016/j.bbrc.2019.11.085> PMID: 31759628.
  45. Schlingemann RO, Hofman P, Vrensen GF, Blaauwgeers HG. Increased expression of endothelial antigen PAL-E in human diabetic retinopathy correlates with microvascular leakage. *Diabetologia*. 1999; 42(5):596–602. Epub 1999/05/20. <https://doi.org/10.1007/s001250051200> PMID: 10333053.
  46. Wisniewska-Kruk J, van der Wijk AE, van Veen HA, Gorgels TG, Vogels IM, Versteeg D, et al. Plasma-lemma Vesicle-Associated Protein Has a Key Role in Blood-Retinal Barrier Loss. *Am J Pathol*. 2016; 186(4):1044–54. Epub 2016/02/16. <https://doi.org/10.1016/j.ajpath.2015.11.019> PMID: 26878208.
  47. Hong Y, Zhang H, Gartner A. The Last Chance Saloon. *Front Cell Dev Biol*. 2021; 9:671297. Epub 2021/06/01. <https://doi.org/10.3389/fcell.2021.671297> PMID: 34055803; PubMed Central PMCID: PMC8160109.
  48. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015; 347(6220):1260419. Epub 2015/01/24. <https://doi.org/10.1126/science.1260419> PMID: 25613900.