

## **Genome-wide association studies found *CCDC7* and *ITGB1* associated with diabetic retinopathy**

Tengda Cai<sup>1</sup>, Qi Pan<sup>1</sup>, Yiwen Tao<sup>1</sup>, Luning Yang<sup>1</sup>, Charvi Nangia<sup>2</sup>, Aravind L Rajendrakumar<sup>3</sup>, Yu Huang<sup>4</sup>, Yongqing Shao<sup>5</sup>, Yunyan Ye<sup>6</sup>, Tania Dottorini<sup>7</sup>, Mainul Haque<sup>8</sup>, Colin NA Palmer<sup>2</sup>, Weihua Meng<sup>1,2,9\*</sup>

1 Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute, University of Nottingham Ningbo China, Ningbo, China, 315100

2 Division of Population Health and Genomics, School of Medicine, University of Dundee, Dundee, UK, DD2 4BF

3 Biodemography of Aging Research Unit, Duke University, Durham, NC, USA, 27708-0408

4 NIHR Global Health Research Unit on Global Diabetes Outcomes Research, Division of Population Health and Genomics, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK, DD1 9SY

5 Department of Ophthalmology, The Affiliated Ningbo Eye Hospital of Wenzhou Medical University, Ningbo, China, 315040

6 Department of Ophthalmology, Lihuili Hospital affiliated with Ningbo University, Ningbo, China, 315040.

7 School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK, LE12 5RD

8 School of Mathematical Sciences, University of Nottingham Ningbo China, Ningbo, China, 315100

9 Center for Public Health, Faculty of Medicine, Health and Life Sciences, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK BT12 6BA

Corresponding Author: Dr Weihua Meng

Address: Division of Population Health and Genomics, Ninewells Hospital and Medical

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**  
School, University of Dundee, Dundee, UK, DD2 4BF; Nottingham Ningbo China Beacons

of Excellence Research and Innovation Institute, University of Nottingham Ningbo China,  
Ningbo, China, 315100. Tel: +44 1382 383419; +86 574 88188632. Email:  
[w.meng@dundee.ac.uk](mailto:w.meng@dundee.ac.uk); [weihua.meng@nottingham.edu.cn](mailto:weihua.meng@nottingham.edu.cn). ORCID: 0000-0001-5388-8494

ORCID codes

Tengda Cai: 0000-0003-1617-506X

Qi Pan: 0000-0003-0135-9122

Yiwen Tao: 0009-0000-0102-4205

Luning Yang: 0000-0001-9056-5950

Charvi Nangia: 0000-0003-3984-2603

Aravind L Rajendrakumar: 0000-0002-4286-7265

Yu Huang: 0000-0001-5355-0047

Yongqing Shao: 0009-0006-1319-8521

Yunyan Ye: 0009-0005-2325-1405

Tania Dottorini: 0000-0001-7733-906X

Mainul Haque: 0000-0003-1836-8076

Colin NA Palmer: 0000-0002-6415-6560

Weihua Meng: 0000-0001-5388-8494

## Abstract

*Purpose:* Diabetic retinopathy (DR), a complication affecting the eyes, is associated with diabetes. This study aims to identify genetic variants associated with DR in patients with type 1 diabetes in the UK Biobank cohort (n = 1,004).

*Methods:* A genome-wide association study (GWAS) was conducted to identify significant genetic variants of DR in type 1 diabetes. The findings are set to undergo validation during the replication and meta-analysis stages by using six cohorts: African American, European, FinnGen, GoSHARE, GoDARTS and Caucasian Australians.

*Results:* In a locus, top single nucleotide polymorphism (SNP) rs184619214 in *CCDC7* reached a GWAS significance level ( $p = 6.38 \times 10^{-9}$ ) and rs79853754 in *ITGB1* ( $p = 3.24 \times 10^{-8}$ ), with both genes being adjacent to each other. The SNP-based heritability was estimated to be 31.09%. Rs184619214 was replicated and reached statistical significance ( $p < 5.0 \times 10^{-8}$ ) in the meta-analysis stage. Pathway analysis revealed that *ITGB1* is involved in the generation of biomolecules that impact the progression of DR. PheWAS analysis revealed that osteoarthritis (OA) of the hip was significantly associated with most of the SNPs of the locus. Mendelian Randomization further confirmed an association between OA and DR.

*Conclusions:* Our study has identified a novel genomic risk locus associated with DR in type 1 diabetes, located in the intergenic region between the *CCDC7* and *ITGB1* genes, providing insights for DR researchers.

**Keywords:** Diabetic retinopathy; genome-wide association study; meta-analysis; Phenome-Wide Association Study; type 1 diabetes

## Introduction

Diabetic retinopathy (DR), a major microvascular complication of diabetes mellitus (DM), is a leading cause of vision loss due to diabetes-induced retinal damage. Research suggests that among individuals with diabetes, the global occurrence of DR is estimated to be 34.6%, 6.96% for proliferative diabetic retinopathy (PDR) <sup>1</sup>. The incidence of DR increases over time, affecting up to 50% of patients within 10 years of a diabetes diagnosis, and 90% by 25 years <sup>2</sup>. There are well-described risk factors for the development of DR. The most clearly recognized are the duration of diabetes (DoD) and the degree of glycaemic control measured as glycosylated hemoglobin (HbA1c) <sup>3</sup>. The burden of DR is anticipated to grow due to factors such as an aging population, increasing obesity rates, and the expanding DM epidemic.

In patients with glycated hemoglobin (HbA1c)  $\leq$  7.0%, the prevalence of any type of DR was 18.0%, compared with 51.2% in patients with HbA1c  $>$  9.0% <sup>4</sup>. The Wisconsin Epidemiological Study of Diabetic Retinopathy (WESDR) shows that DM duration is a key predictor of DR progression, with DR prevalence increasing from 17% in type 1 DM and 29% in type 2 DM for those with less than 5 years of DM, to nearly 100% in type 1 DM and 78% in type 2 DM for those with the disease for over 15 years <sup>5</sup>. This trend is further supported by a study showing that the prevalence of DR escalates with the DoD, with 21.1% of patients affected at less than 10 years, soaring to 76.3% at 20 years or more <sup>6</sup>. Studies suggest that there is a clear familial nature to DR, with twin and family studies suggesting a genetic basis, yet the specific genes contributing to DR risk remain elusive, with only a handful of genetic associations identified, none of which demonstrate strong links <sup>5,7</sup>.

In a study of 4,172 patients diagnosed with T1D from 12 years of age, 26.7% had background diabetic retinopathy, 10.7% had initial diabetic retinopathy, and 4.1% had PDR <sup>8</sup>. Notably, adults with T1D indicated that the prevalence of DR increases with the disease duration, affecting more than 80% of patients who have lived with the disease for over 40 years <sup>9</sup>. In the United States, T1D affects 244,000 youths and 1.6 million adults over 20 years old, with around 64,000 new cases diagnosed annually <sup>10</sup>. This means that cases of DR caused solely by T1D are not uncommon. T1D itself results from the immune-mediated destruction of insulin-secreting beta cells in the pancreas, leading to chronic hyperglycemia and subsequent organ impairment, including the eyes <sup>8,11</sup>. Alarming, the incidence of T1D and related vision loss is on the rise, growing annually at a rate of 3% to 5% <sup>12</sup>. Effective

interventions such as strict glucose and blood pressure management, lipid-lowering treatments, and laser therapies can slow disease progression and help preserve vision <sup>13</sup>.

Various genome-wide association studies (GWAS) have identified numerous genetic loci potentially linked to DR, highlighting a significant genetic component in its pathology <sup>13,14,15,16</sup>. Specifically, for T1D-related DR, GWAS have pinpointed candidate genes such as *MIR449b* <sup>11</sup>, *CACNB2* <sup>17</sup>, and others including *LOC339529*, *RBFOX1* <sup>18</sup>, *KLRD1*, *TNF* <sup>19</sup>, as well as *KIF16B*, *FUT4* <sup>20</sup>. However, these identified GWAS loci for DR often vary and lack consistent replicability across studies, which complicates deeper genetic analyses and hinders the identification of novel biological pathways for targeted interventions. This gap in understanding exacerbates the healthcare challenges and costs associated with managing DR.

To elucidate the genetic mechanisms linked to DR in T1D patients, we conducted a GWAS based on the UK Biobank (UKB) cohort, with efforts using additional cohorts for replication. This research aims to uncover new genetic loci related to DR, providing valuable information for researchers in this area.

## **Research Design and Methods**

### **Information on Cohorts and Patients**

The UKB functions as a large-scale biomedical database and research resource, housing detailed genetic and phenotypic information for approximately 500,000 participants throughout England, Scotland, and Wales. For further information on the UKB cohort, refer to [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk). Conducted in accordance with the Declaration of Helsinki's tenets, this study secured informed consent from all participants after they were fully informed about the UKB study's purpose. The research received the necessary ethical approval from the National Health Service National Research Ethics Service (reference 11/NW/0382).

In this research, the genetic data came from the UK Biobank cohort. They used a standardized process for DNA extraction and quality control (QC), in which detailed methods can be found at [https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/genotyping\\_sample\\_workflow.pdf](https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/genotyping_sample_workflow.pdf). We defined samples clearly for cases and controls by phenotype data filtering using the following UKB field ID codes (Fig.1). Our study used the participants of white British (UKB field ID: 21000). T1D with prolonged

DoD was identified using four criteria: a diagnosis of diabetes by doctors (UKB field ID: 2443), a response of 'Yes' to the question 'Started insulin within one year of diabetes diagnosis' (UKB field ID: 2986), "age diabetes diagnosed" <= 40 years old (UKB field ID: 2976), and diabetes samples that have a DoD at least 20 years. Among T1D patients, DR cases were identified using the ICD-10 code 'H360' from the World Health Organization International Classification of Diseases (UKB field ID: 41202). Conversely, control subjects were those without the ICD-10 code 'H360', indicating no diagnosis of DR. Furthermore, we selected the HbA1c level that ranks in the top 95% of the above-selected samples to focus on individuals with relatively poor glycemic control, which is a critical risk factor for the development of DR. The clinical characteristics of the case and control groups are shown in Table 1. Six cohorts were used for replication and meta-analysis, including African American & European cohorts<sup>16</sup>, FinnGen cohorts<sup>21</sup>, GoSHARE & GoDARTS (Rajendrakumar, 2022), Caucasian Australians<sup>22</sup>. More information is shown in Supplementary Table S1.

Fig.1. Definition of diabetic retinopathy in type 1 diabetes based on UKB cohort.

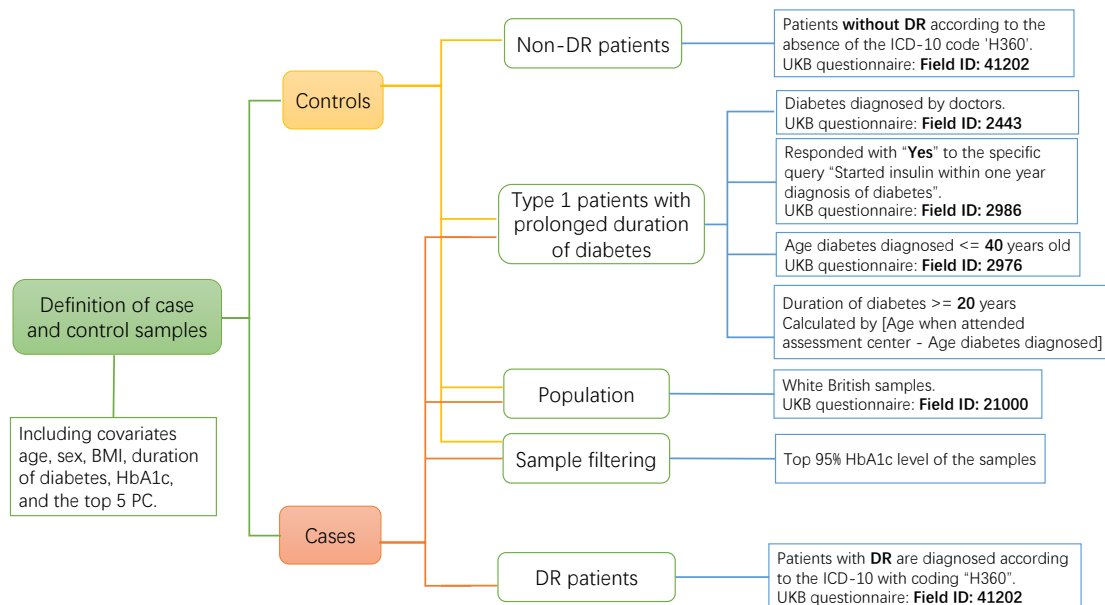


Table 1. Clinical characteristics using a dataset based on UK Biobank

Covariates	Clinical Characteristics		
	Cases	Controls	<i>P</i>
Sex(male:female)	256:192(57.14%:42.86%)	315:241(56.65%:43.35%)	0.87 > 0.05
Age(years)	57.1(8.02)	55.1(7.66)	$5.81 \times 10^{-5} < 0.05$
BMI(kg/m <sup>2</sup> )	27.7(4.58)	27.0(4.59)	0.03 < 0.05
DoD(duration)	36.4(9.74)	33.7(9.29)	$1.05 \times 10^{-5} < 0.05$
HbA1c(mmol/mol)	64.9(13.25)	60.7(10.98)	$1.15 \times 10^{-7} < 0.05$

DoD: duration of diabetes. A chi-square test assessed the differences in sex distribution between cases and controls, while an independent t-test evaluated other covariates. Continuous covariates were shown as 'mean(standard deviation)'.

### Statistical analysis

Genome-Wide Complex Trait Analysis (GCTA v1.93.3) was used for GWAS association analysis (refer to <https://yanglab.westlake.edu.cn/software/gcta/#Overview>). Our study performed the genome-wide association methods utilizing the fastGWA-GLMM tool<sup>23</sup>. Covariates incorporated into the model were age, sex, BMI, DoD, HbA1c, and the top 5 genetic principal components. In the QC step, SNPs with imputation INFO scores < 0.3 and minor allele frequency < 0.5% were removed by using GCTA, as well as SNPs that failed Hardy-Weinberg < 0.00001 and samples with missing call rates < 0.1 were filtered out by using Plink v1.90. SNPs situated on the X and Y chromosomes and mitochondrial SNPs were eliminated from consideration. SNPs were considered to reach genome-wide association significance if they had a *p* value less than  $5 \times 10^{-8}$ . SNP-based heritability was calculated using the “BLD-LDAK Model” in SumHer<sup>24,25</sup>.

### Replication studies and Meta-analysis

Our replication approach is to replicate SNPs or their LD proxies in other studies that matched the significant SNPs ( $p < 5 \times 10^{-8}$ ) found in our GWAS and record their *p* values. Fixed-effects meta-analyses with inverse weighting were conducted using the R package 'metafor'. Notably, to standardize and align the strand orientation for alleles across different cohorts, we have inverted the strand designation (switching the effect allele, EA, with the non-effect allele, NEA) in summary statistics results produced

from our UKB study. Correspondingly, the beta values and their 95% confidence intervals have also been adjusted to ensure consistency in the results.

### **GWAS analysis by FUMA and LDSC**

The GRCh37 genome assembly was followed for SNP annotation. SNP functional annotations and Manhattan plots were applied and produced using FUMA v1.5.2, a web platform designed to annotate, prioritize, visualize, and interpret GWAS results <sup>26</sup>. Gene mapping was conducted for cis-expression quantitative trait loci (cis-eQTL), chromatin interaction, and positional mapping studies. Gene expression heatmaps were generated, and tissue specificity analyses were performed for genes identified through positional mapping, utilizing data from 54 specific tissue types and 30 general tissue types in GTEx v8. The gene analysis, gene-set analysis, and tissue expression analysis were performed by MAGMA v1.08, which was integrated into FUMA. In gene analysis, SNP summary statistics were consolidated at the gene level, assessing the combined association of all SNPs within a gene with the phenotype. Gene-set analysis involved grouping genes based on common biological, functional, or other traits, which helped to unravel the role of specific biological pathways or cellular functions in the genetic basis of a phenotype. To test the relationship between highly expressed genes in a specific tissue and genetic associations with DR, tissue expression analysis, based on GTEx (<https://www.gtexportal.org/home/>), was performed separately for 30 general tissue types and 53 specific tissue types. The parameters in the FUMA platform are set in Supplementary Table S2. Q-Q plots and regional visualization were provided by LocusZoom (<http://locuszoom.org/>).

### **Enrichment of genes and Pathway analysis**

Enrichment analysis of positional gene sets was performed using the 'GENE2FUNC' feature in FUMA. Utilizing the Molecular Signatures Database (MSigDB) <sup>27,28</sup> and the Reactome <sup>29</sup>, we identified relevant pathways and gene interactions for genes of interest, particularly *ITGB1* and *CCDC7*. This analysis involved mapping our significantly associated SNPs to corresponding gene sets in MSigDB and exploring gene interactions through Reactome pathways to assess the potential biological mechanisms influencing diabetic retinopathy.



## **Transcriptome-wide association studies (TWAS)**

We employed Transcriptome-Wide Association Studies (TWAS) (<http://gusevlab.org/projects/fusion/>)<sup>30</sup> to assess the impact of gene expression modulated by genetic variants on DR risk. Focusing on specific tissues from panel GTEx v7, TWAS integrated SNP-gene expression associations from eQTL studies with SNP-disease associations from GWAS summary statistics.

## **Phenome-Wide Association Study**

To investigate potential associations between the top SNPs identified in our current analysis and other related traits, we conducted a Phenome-Wide Association Study (PheWAS). In this analysis, a PheWAS was performed using a comprehensive dataset of 4,756 GWAS summary statistics available on the GWASATLAS platform (<https://atlas.ctglab.nl/PheWAS>). Our inclusion criteria included all GWAS. To construct the PheWAS SNP heatmap, we only included phenotypes that had p values less than 0.05 in their association with the SNPs. Construct a data frame with 12 SNPs associated with different traits. Adjustment for multiple comparisons was performed using the Bonferroni correction method.

## **Mendelian Randomization**

To investigate whether significant phenotypic results from PheWAS are causally related to DR, we used a bidirectional Mendelian randomization (MR) approach using the “TwoSampleMR” package. Except for DR summary statistics, data sources were obtained from the IEU OpenGWAS database (<https://gwas.mrcieu.ac.uk/>). Prior to MR analysis, the data were clumped to address the issue of linkage disequilibrium. Our MR methods included inverse variance weighted (IVW), MR Egger, weighted median, simple median, and weighted mode.

## **Results**

### **Significant GWAS results**

We identified 448 cases and 556 controls after genetic and phenotype quality control. Altogether, 11,185,683 imputed SNPs passed from routine QC checking and imputation INFO quality scores that were larger than 0.3. The most significant SNP in this region was rs184619214 in the *CCDC7* gene with a  $p$  value of  $6.38 \times 10^{-9}$  in Chromosome 10 (see Table 2). The significant SNP rs79853754 in *ITGB1*, with a  $p$  value of  $3.24 \times 10^{-8}$ , is in high linkage disequilibrium with rs184619214 ( $r^2 = 0.95$ ). A cluster of top SNPs reached a significant level in the Manhattan plot, with regional plots at the top (Fig.2). The corresponding Q-Q plot is shown in Fig.3. The SNP-based heritability of DR was estimated to be 31.09% in this T1D population.

Table 2. Significant locus of GWAS based on the UKB.

Marker	Chr	Position	Nearest gene	EA	NEA	AF1	Beta	SE	$p$ value	$r^2$
rs184619214	10	32794461	<i>CCDC7</i>	A	G	0.98	-0.47	0.08	6.38E-09	1.00
rs76802012	10	32913261	<i>CCDC7</i>	A	T	0.98	-0.45	0.08	1.53E-08	0.96
rs77948029	10	33162616	<i>CCDC7</i>	A	T	0.98	-0.45	0.08	1.55E-08	0.96
rs79853754	10	33192606	<i>ITGB1</i>	G	C	0.98	-0.44	0.08	3.24E-08	0.95
rs187213927	10	32900213	<i>CCDC7</i>	A	G	0.98	-0.45	0.08	4.04E-08	0.95
rs41299216	10	33227260	<i>ITGB1</i>	G	A	0.98	-0.45	0.08	4.29E-08	0.93
rs77605282	10	32733099	<i>CCDC7</i>	C	T	0.98	-0.45	0.08	5.93E-08	0.97
rs41276070	10	33134064	<i>CCDC7</i>	A	G	0.98	-0.45	0.08	6.48E-08	0.95
rs117300946	10	32814555	<i>CCDC7</i>	A	G	0.98	-0.43	0.08	9.73E-08	0.92
rs117055050	10	32712210	<i>CCDC7</i>	G	A	0.98	-0.48	0.09	1.37E-07	0.91
rs117250131	10	33252757	<i>ITGB1-DT</i>	A	C	0.98	-0.36	0.07	1.33E-06	0.89

\* indicates SNPs reach genome-wide significance with  $p$  value  $< 5 \times 10^{-8}$ , while others are suggestive.

Chr: chromosome, EA: effect allele, NEA: non effect allele, AF1: frequency of effect allele

Fig.2 The Manhattan and regional plots of the GWAS analysis for diabetic retinopathy in type 1 diabetes.

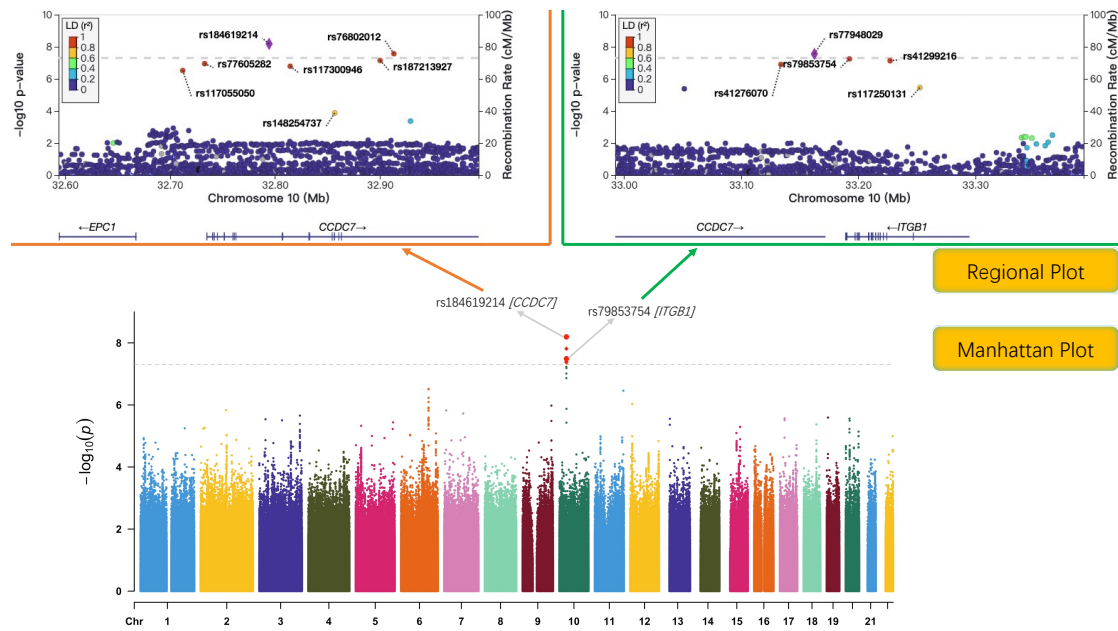
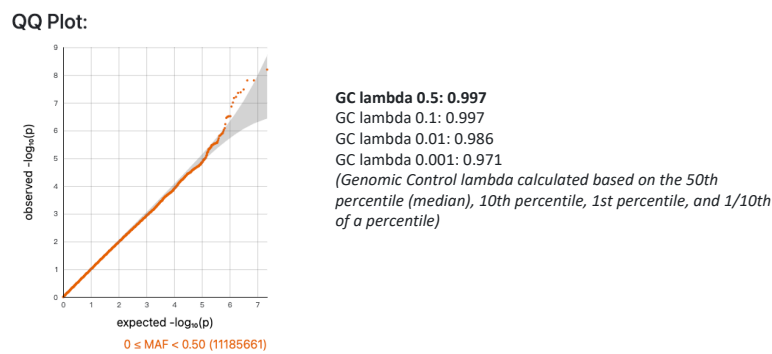


Fig.3 The QQ plot of the GWAS



## Replication & Meta-analyses

Rs184619214 was replicated in the African American and European cohorts with a  $p$  value of 0.05 and 0.04. Meta-analyses conducted with six cohorts revealed that rs184619214 reached genome-wide significance with  $p$  value of  $4.47 \times 10^{-11}$  (Table 3). Figure S1 shows a forest plot, funnel plot and radial plot of meta-analysis results using UKB, European, FinnGen, GoSHARE, GoDARTS, and Caucasian Australian cohorts.

## Gene, gene-set and tissue expression analysis by MAGMA

The entire dataset of GWAS summary statistics was utilized to conduct a comprehensive assessment of gene-disease associations.

In the gene analysis, all the SNPs located within genes were mapped to 19,295 protein-coding genes. Genome-wide significance was defined (red dashed line in the plot) at  $p = 0.05/19295 = 2.59 \times 10^{-6}$ . No gene demonstrated the strongest association with DR, see Figure S2.

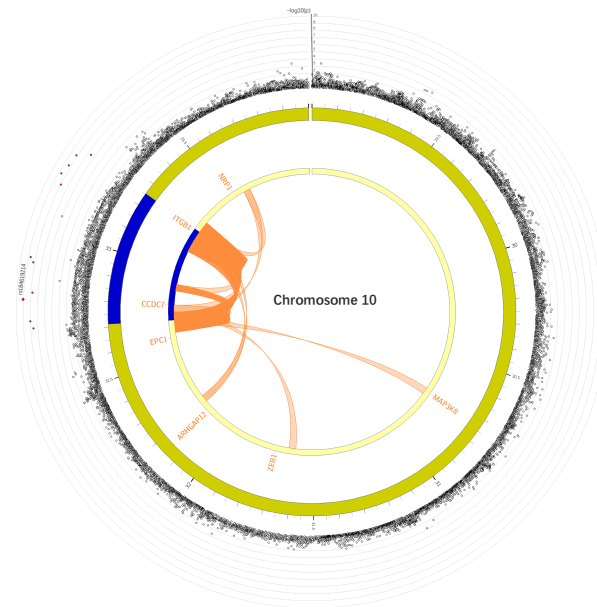
In the gene-set analysis, 10,678 gene sets from MsigDB v6.2 were tested. The top first gene set GOMF\_LEUKOTRIENE\_C4\_SYNTHASE\_ACTIVITY has not achieved a statistically significant association after Bonferroni correction, demonstrating a  $p$ -value of  $1.61 \times 10^{-5}$  ( $> 0.05/10,678 = 4.68 \times 10^{-6}$ ) and  $p_{bon} = 0.27 < 0.05$ . This analysis's top ten gene sets are shown in Supplementary Table S3.

Tissue expression analysis demonstrated no tissue achieved statistical significance in the expression analysis of 30 general tissue types from multiple organs and the 53 specific tissue types within some organs. Results are shown in Figure S3.

### **Cis-expression quantitative trait loci (cis-eQTL), chromatin interaction analysis and positional mapping**

Cis-eQTL analysis did not identify any significant SNPs associated with GTEx tissue types. Chromatin interaction analysis revealed that significant occurring tissues or cells in *CCDC7* are Left\_Ventricle, IMR90, Mesenchymal\_Stem\_Cell, Mesendoderm, Trophoblast-like\_Cell, hESC, and in *ITGB1* are Left\_Ventricle, Liver, GM12878, IMR90, Mesenchymal\_Stem\_Cell, Mesendoderm, Trophoblast-like\_Cell, hESC respectively. The results of chromatin interactions are presented in the Fig.4, genes interacting with *CCDC7* and *ITGB1* were *NRP1*, *CI0ORF68*, *EPC1*, *ARHGAP12*, *ZEB1*, *ZNF438* and *MAP3K8*. Positional mapping identified 10 genes as potentially related to the DR, including *CCDC7*, *ITGB1*, *NRP1*, *EPC1*, *ARHGAP12*, etc (details are listed in Supplementary Table S4).

Fig.4 The circus plot of gene mapping of chromatin interactions and cis-eQTL

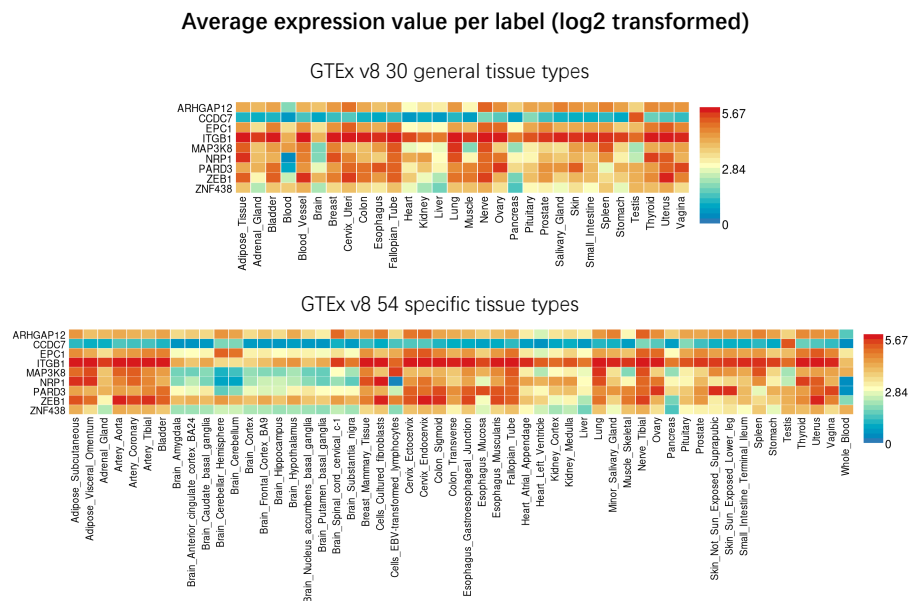


### Gene expression heatmap and tissue specificity

Gene expression heatmaps using average expression value per label are shown in the Fig.5 Cells filled in red represent higher expression compared to cells filled in blue across genes and labels. Cells filled in red represents higher expression of the genes in a corresponding label compared to other labels, but it does not represent higher expression compared to other genes. The expression of gene *CCDC7* in the testis is obviously higher than in other tissue types. Nerve tibial shows a relatively higher expression among all these genes.

Tissue specificity is tested using the differentially expressed genes (DEG) defined for each label of each expression data set. An enrichment test of DEG for both GTEx v8 54 tissue types and GTEx v8 30 general tissue types was performed, only testing the enrichment of positional mapping genes instead of using the full distribution of SNP  $p$  values in MAGMA tissue expression analysis. Tissue specificity testing found that brain cortex showed significant enrichment in the down-regulated DEG in GTEx v8 54 tissue types. See Figure S4.

Fig.5. Gene expression heatmaps of positional mapping genes.



## Enrichment of genes and Pathway analysis

The results of the gene enrichment are shown in Figure S5. *CCDC7*, *C10ORF68*, *ITGB1* and *NRP1* are overlapping genes. The pathway analysis revealed enrichment and interactions involving *ITGB1* and *CCDC7*. Utilizing the Human Tissue Compendium from MSigDB, we observed associations among these genes and also identified the varying strengths of connections between these genes and various tissues (Figure S6). In the Reactome pathway diagram (Figure S7), *LICAM* is involved in the assembly of integrin complexes and suggests that *CHL1* mediates interactions between *ITGB1* and *NRP1*.

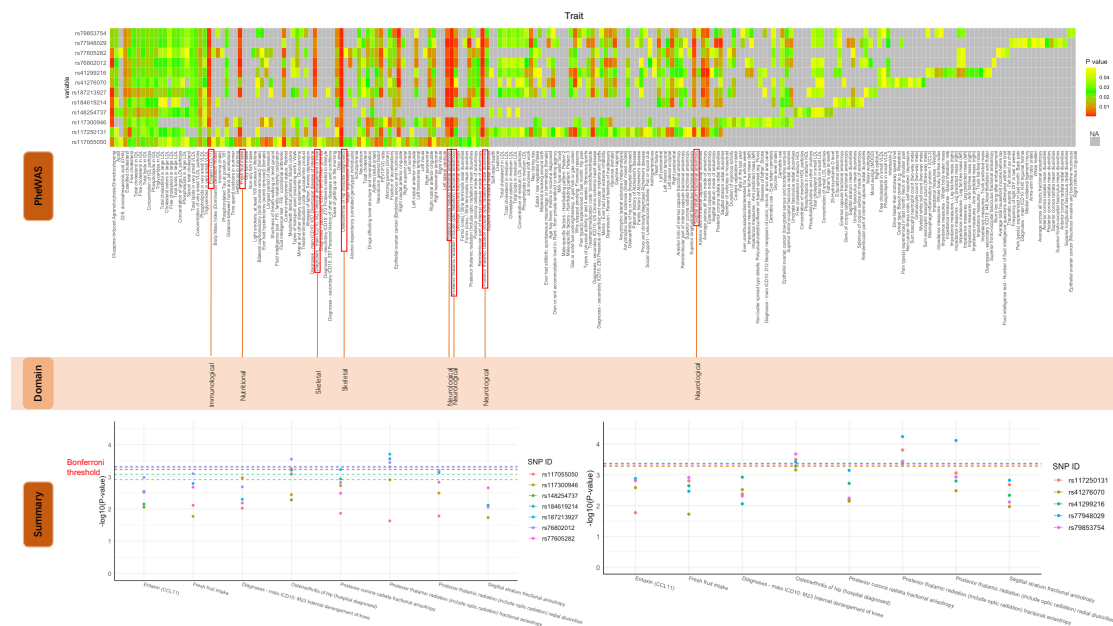
## TWAS results

TWAS results across multiple tissue types using GTEx v7 data did not reveal statistically significant associations for the SNPs associated with our GWAS results (Supplementary Table S5). Despite extensive testing, both the *CCDC7* and *ITGB1* genes did not show significant expression-trait associations in the tissues analysed. Specifically, *CCDC7* was occasionally identified but not consistently across different tissues, while *ITGB1* was rarely detected, which may indicate a lack of statistical power or the broader biological roles of the genes not limited to tissue-specific expression.

## Phenome-Wide Association Study (PheWAS)

Totally 8 potentially significant phenotypes were selected according to their high association with the 12 SNPs (upper part of Fig.6), including eotaxin\_(CCL11), fresh fruit intake, M23 internal derangement of knee, osteoarthritis of hip, posterior corona radiata fractional anisotropy, posterior thalamic radiation (include optic radiation) fractional anisotropy, posterior thalamic radiation (include optic radiation) radial diusivities, sagittal stratum fractional anisotropy, as indicated by red coloring predominantly in the heatmap ( $p < 0.01$ ). After Bonferroni correction, where each SNP has its own threshold (lower part of Fig.6), we found that three phenotypes Posterior thalamic radiation (include optic radiation) fractional anisotropy, Posterior thalamic radiation (include optic radiation) radial diusivities, Osteoarthritis of hip (hospital diagnosed) were significant with most of the SNPs.

Fig.6 Results of PheWAS for the top 12 SNPs in significant GWAS.

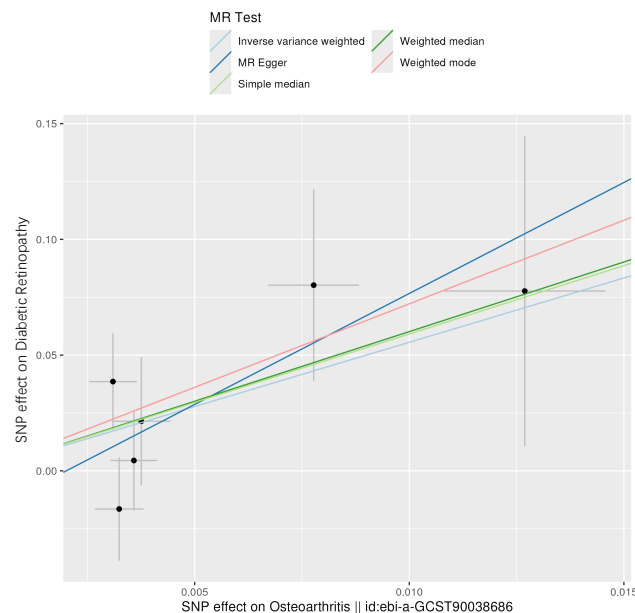


### Mendelian Randomization

The results revealed that osteoarthritis (OA) (data from IEU ID: "ebi-a-GCST90038686"), when considered as an exposure, significantly led to DR, with an IVW  $p$  value of 0.02, which is below the significant threshold of 0.05. In the heterogeneity analysis, both the IVW and MR Egger  $p$ -values are greater than 0.05, indicating no significant heterogeneity between the genetic tools used, suggesting

consistency in the magnitude and direction of their effects. Conversely, DR as an exposure did not significantly lead to OA, as indicated by an IVW  $p$  value of  $0.10 > 0.05$ . More detailed results can be found in Supplementary Table S6. The accompanying figure illustrates the MR analysis, including methods: IVW, MR Egger, simple median, weighted median and weighted mode. It plots the SNP effects on osteoarthritis against their effects on diabetic retinopathy in Fig. 7 and Figure S8.

Fig.7 Plot of SNP effects on osteoarthritis (exposure) and diabetic retinopathy (outcome) using multiple MR methods.



## Discussion

In this GWAS study of DR on type 1 patients using the UKB dataset, we identified novel significant genetic variants at a genomic risk locus in the region of gene *CCDC7* and *ITGB1*, highlighting the complex genetic basis associated with DR. The top variants rs184619214 in *CCDC7* and rs79853754 in *ITGB1* have reached genome-wide significance ( $p < 5 \times 10^{-8}$ ), and they exhibit high linkage disequilibrium with  $r^2 = 0.95$ . The estimated genomic inflation factor ( $\lambda$ ) calculated from GWAS was 0.997 (Fig.4), suggesting a homogeneous population. Notably the heritability of DR was calculated to be 31.09% in the GWAS, emphasizing that DR exhibits a relatively high heritability in the population with T1D.



The rs184619214 variant is located in the *CCDC7* gene (Coiled-Coil Domain Containing 7). Its gene synonyms include *BIOT2*, *C10ORF68*, *FLJ13031*, and *FLJ32762*. Current research on *CCDC7* primarily focuses on its roles in vascular functions and male-specific characteristics. *CCDC7* is involved in VEGF activation to promote sustained angiogenesis, along with many *CCDC* proteins that play a role in vascular stability and inflammation<sup>31</sup>. According to Priyanka<sup>32</sup>, *CCDC* proteins are primarily expressed in the human male reproductive system, similar to our gene expression heatmaps showing that *CCDC7* has relatively high expression levels in the testes, suggesting specificity in its function and activity within this system. Research indicates that male sex is a risk factor for DR, with men more prone to advanced retinopathy in T1D and seeming to exhibit more severe conditions at diagnosis in T2D<sup>33,34</sup>. Conversely, experimental models in mice show no gender differences in the early stages of DR<sup>35</sup>. In our study according to Table 1 clinical characteristics, there is no significant difference in sex distribution between cases and controls, with a ratio of approximately 57% male to 43% female. This finding indicates that our current study does not provide sufficient evidence to suggest a higher prevalence of DR in males compared to females within T1D. Furthermore, the total sample size of 1,004 may also limit the statistical power of these results.

The rs79853754 variant is located in *ITGB1* (Integrin Subunit Beta 1), which has been found to be associated with DR in previous studies. The *ITGB1* plays a pivotal role in PDR and proliferative vitreo-retinopathy due to its contribution to the terminal event of tractional retinal detachment<sup>36</sup>. A result suggested a unique mechanism through which *SPARC-ITGB1* interaction and hyperglycemia modulate VEGF signaling to inhibit neovascularization in DR<sup>37</sup>. Additionally, lysine acetyltransferase 1 (KAT1) induces YTHDF2-mediated instability of *ITGB1* mRNA, which alleviates the progression of DR in a mouse model<sup>37</sup>. Reactome pathway analysis reveals that *LICAM* is involved in the assembly of integrin complexes, with *CHL1* mediating interactions between *ITGB1* and *NRP1*. Elevated levels of *LICAM* have been associated with retinopathy<sup>38</sup>. Additionally, there is a partial overlap between the *ITGB1* and *NRP1* genes, with *NRP1* identified as being linked to DR<sup>39,40</sup>. Inhibition of *NRP1* has been shown to reduce the development of early DR<sup>41</sup>. Furthermore, under hypoxic conditions, *CHL1* upregulation disrupts the homeostatic balance between BMP-4 and VEGF, enhancing VEGF's role in promoting retinal angiogenesis<sup>42</sup>. Other comprehensive analyses using tissue expression analysis, TWAS, and eQTL studies across multiple tissue types from the GTEx database did not yield significant findings for

our significant SNPs. This suggests that while *ITGB1* is known to play a crucial role at the cellular level, there is no evidence from our data to indicate that these SNPs have tissue-specific expression patterns.

Gene expression heatmaps indicate higher expression in both the artery and nerve tibial regions (Fig.5). Research links tibial nerve characteristics to DR. patients with mild NPDR show reduced tibial motor conduction velocity compared to those without DR <sup>43</sup>. Moreover, diabetic peripheral neuropathy (DPN) patients exhibit increased cross-sectional areas of the tibial nerve, while DPN and DR have a positive correlation <sup>44</sup>, with similar abnormalities noted in type 1 diabetes (T1D) patients with neuropathy <sup>45</sup>, suggesting the tibial nerve's involvement in both T1D and DR. It is noteworthy that *CCDC7* has been linked to tibial involvement, as it was identified in patients with symptoms similar to centronuclear myopathy, featuring lower limb pain and reduced muscle strength in the tibia <sup>46</sup>. Chromatin interaction analysis revealed significant interactions in *CCDC7* and *ITGB1*, where common cell and tissue are Mesenchymal Stem Cell and Left Ventricle. Mesenchymal stem cells have been discovered as potential therapy in diabetic retinopathy <sup>47,48</sup>, and left ventricle structure and dysfunction were associated with DR <sup>49</sup>.

The PheWAS results have provided us with many important insights. Eotaxin (CCL11) is a chemokine closely associated with inflammatory processes and has been recognized as a representative phenotype indicating the presence of inflammation in the body. Furthermore, Eotaxin (CCL11) is proposed as a potential candidate for developing targeted therapeutic strategies, which could be utilized either alone or in conjunction with anti-VEGF drugs to treat PDR <sup>50</sup>. Radiation retinopathy is more likely to develop in eyes with preexisting diabetic retinopathy and, due to their clinical and histopathological similarities, is treated using established diabetic retinopathy therapies like laser photocoagulation <sup>51,52</sup>. M23 Internal derangement of the knee and osteoarthritis of the hip from the skeletal domain showed significant associations with most of the SNPs. Research indicates that PDR is significantly associated with the absence of clinical OA in long-term diabetes mellitus patients <sup>53</sup>. Other studies found that retinal arteriolar narrowing, associated with an increased risk of knee replacement for OA <sup>54</sup>, also correlated with diabetic retinopathy in T1D patients <sup>55</sup>. Our MR analysis demonstrates that OA acts as a significant exposure factor associated with the development of DR. This suggests that OA may play a mechanistic role in the pathogenesis of DR, potentially mediated by systemic inflammatory processes or microvascular dysfunction that are common to both conditions. Identifying OA as a contributory factor

to DR provides a valuable early warning indicator, which could be critical for preemptive medical intervention to mitigate the progression of DR.

Our study's limitation stems from the UKB's unclear definition of T1D. We established our T1D definition based on clinical insights, considering both the sample size (a minimum of 1000 participants) and the precision (age at diabetes diagnosis) to finalize our definition of T1D (Fig.1). Compared to another study's definition of T1D—diagnosed under 30 years and requiring insulin within a year of diagnosis<sup>19</sup>—our clinical characteristics showed an average age at diabetes diagnosis of approximately 22 years. Furthermore, unlike Forrest<sup>56</sup>, which included diverse ethnic backgrounds from the UKB cohort, our study targeted White British individuals with T1D-related DR. We also incorporated critical covariates such as DoD and HbA1c levels in our GWAS, enhancing both the accuracy and relevance of our findings. Our significant SNPs have proven challenging to replicate strongly. This may be attributed to the limited number of studies focusing on this specific association, as detailed in Supplementary Table S7, resulting in insufficient robust data to support our findings. Although the meta-analysis indicates significant results, the use of high linkage disequilibrium proxies for these SNPs introduces a potential bias and limits the statistical power of our analyses.

## **Conclusion**

Overall, our GWAS study identified a new locus at the genetic position of the *CCDC7* and *ITGB1* genes, which will help to understand the mechanisms of DR.

## **Statements and Declarations**

### **Data availability**

This research complies with all ethical standards and data privacy measures the UKB sets. Detailed summary statistics data about DR within the UKB are retrievable from <https://figshare.com/s/68c2e17003c694c6f453>. Should any additional data pertinent to this study need to be presented within this paper or its other files, the authors can provide such data upon reasonable request.

### **Author information**

#### **Authors and Affiliations**

**Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute,  
University of Nottingham Ningbo China, Ningbo, China, 315100**

Tengda Cai, Qi Pan, Yiwen Tao, Luning Yang, Weihua Meng

**Division of Population Health and Genomics, School of Medicine, University of Dundee,  
Dundee, UK, DD2 4BF**

Charvi Nangia, Colin NA Palmer

**Biodemography of Aging Research Unit, Duke University, Durham, NC, USA, 27708-0408**

Aravind L Rajendrakumar

**NIHR Global Health Research Unit on Global Diabetes Outcomes Research, Division of  
Population Health and Genomics, University of Dundee, Ninewells Hospital and Medical  
School, Dundee, UK, DD1 9SY**

Yu Huang

**Department of Ophthalmology, The Affiliated Ningbo Eye Hospital of Wenzhou Medical  
University, Ningbo, China. 315040**

Yongqing Shao

**Department of Ophthalmology, Lihuili Hospital affiliated with Ningbo University, Yinzhou  
District, Ningbo, China**

Yunyan Ye

**School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK, LE12  
5RD**

Tania Dottorini

**School of Mathematical Sciences, University of Nottingham Ningbo China, Ningbo, China,  
315100**

Mainul Haque

### **Authors' Contributions**

TC drafted the paper and performed GWAS analysis. QP, YT, LY contributed to data formatting and correction. CN, AR, YH, YS, YY, TD, MH, CP provided comments on the paper. WM organized the project and provided comments.

### **Acknowledgment**

This study adheres to all ethical guidelines and data protection protocols of the UK Biobank. The current study was conducted under approved UK Biobank data application number 50604. This study was mainly funded by the Pioneer and Leading Goose R&D Program of Zhejiang Province 2023 with reference number 2023C04049 and Ningbo International Collaboration Program 2023 with reference number 2023H025. The authors gratefully thank all the participants and professionals contributing to the UKB and FinnGen studies.

### **Corresponding authors**

Correspondence to Weihua Meng

### **Consent to Publish**

All authors have consent for publication.

### **Conflict of Interest**

The authors have no conflicts of interest to disclose.

### **Ethics approval**

This study was approved by the Ethics Committee of the University of Nottingham, Ningbo, China.

## References

1. Yuan, W. et al. (2024) 'Exploration of potential novel drug targets for diabetic retinopathy by plasma proteome screening', *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-62069-0>.
2. Wang, Y. et al. (2021) 'Proteomic analysis of aqueous humor proteins associated with neovascular glaucoma secondary to proliferative diabetic retinopathy', *Current Proteomics*, 18(5), pp. 717–729. <https://doi.org/10.2174/1570164618999201210224640>.
3. Burdon, K.P., McComish, B.J. and Charlesworth, J.C. (2018) Progress and challenges in genome-wide studies to understand the genetics of diabetic retinopathy, *Annals of Eye Science*. Available at: <https://doi.org/10.21037/aes.2018.08.04>.
4. Yau, J.W.Y. et al. (2012) 'Global prevalence and major risk factors of diabetic retinopathy', *Diabetes Care*, 35(3), pp. 556–564. <https://doi.org/10.2337/dc11-1909>.
5. Kuo, J.Z., Wong, T.Y. and Rotter, J.I. (2014) 'Challenges in elucidating the genetics of diabetic retinopathy', *JAMA Ophthalmology*, 132(1), p. 96. <https://doi.org/10.1001/jamaophthalmol.2013.5024>.
6. Hainsworth, D.P. et al. (2019) 'Risk factors for retinopathy in type 1 diabetes: The DCCT/Edic Study', *Diabetes Care*, 42(5), pp. 875–882. <https://doi.org/10.2337/dc18-2308>.
7. Hosseini, S.M. et al. (2014) 'The association of previously reported polymorphisms for microvascular complications in a meta-analysis of diabetic retinopathy', *Human Genetics*, 134(2), pp. 247–257. <https://doi.org/10.1007/s00439-014-1517-2>.
8. Wysocka-Mincewicz, M. et al. (2021) 'Diabetic retinopathy in children with type 1 diabetes—occurrence and screening using optical coherence tomography', *Life*, 11(6), p. 590. <https://doi.org/10.3390/life11060590>.
9. Hammes, H.P. et al. (2011) 'Diabetic retinopathy in type 1 diabetes—a contemporary analysis of 8,784 patients', *Diabetologia*, 54(8), pp. 1977–1984. <https://doi.org/10.1007/s00125-011-2198-1>.
10. Hesterman, J. et al. (2022) Discovering autoantibodies to nitrated proteins in type 1 diabetes, *Handle Proxy*. Available at: <https://hdl.handle.net/2286/R.2.N.171311> (Accessed: 10 November 2023).
11. Liu, E. et al. (2019) 'MicroRNA-related genetic variants are associated with diabetic retinopathy in type 1 diabetes mellitus', *Investigative Ophthalmology & Visual Science*, 60(12), p. 3937. <https://doi.org/10.1167/iovs.18-25570>
12. Maffi, P. and Secchi, A. (2017) 'The burden of diabetes: Emerging data', *Management of Diabetic Retinopathy*, pp. 1–5. <https://doi.org/10.1159/000459641>.
13. Yu X and Rong S (2023) Genome-wide associations and confirmatory meta-analyses in diabetic retinopathy. *Genes* 14(3): 653. <https://doi.org/10.3390/genes14030653>
14. Liu, C. et al. (2019) 'Genome-wide association study for proliferative diabetic retinopathy in Africans', *npj Genomic Medicine*, 4(1). <https://doi.org/10.1038/s41525-019-0094-7>.
15. Meng W, Shah KP, Pollack S et al (2018) A genome-wide association study suggests new evidence for an association of the NADPH oxidase 4 (nox4) gene with severe diabetic retinopathy in type 2 diabetes. *Acta Ophthalmologica* 96(7). <https://doi.org/10.1111/aos.13769>
16. Pollack S, Jr RP, Jensen RA et al (2018) Multiethnic genome-wide association study of diabetic retinopathy using liability threshold modeling of duration of diabetes and glycemic control. *Diabetes* 68: 441–456. <https://doi.org/10.2337/db18-0567>
17. Vuori, N. et al. (2019) 'cacnb2 is a novel susceptibility gene for diabetic retinopathy in type 1 diabetes', *Diabetes*, 68(11), pp. 2165–2174. <https://doi.org/10.2337/db19-0130>.
18. Grassi, M.A. et al. (2011) 'Genome-wide meta-analysis for severe diabetic retinopathy', *Human Molecular Genetics*, 20(12), pp. 2472–2481. <https://doi.org/10.1093/hmg/ddr121>.
19. Agardh, E. et al. (2015a) 'Genome-wide analysis of DNA methylation in subjects with type 1 diabetes identifies epigenetic modifications associated with proliferative diabetic retinopathy', *BMC Medicine*, 13(1). <https://doi.org/10.1186/s12916-015-0421-5>.
20. Miller, R.G. et al. (2024) 'An epigenome-wide association study of DNA methylation and proliferative retinopathy over 28 years in type 1 diabetes', *Ophthalmology Science*, 4(4), p. 100497. <https://doi.org/10.1016/j.xops.2024.100497>.
21. Kurki MI, Karjalainen J, Palta P et al (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 613: 508–518. <https://doi.org/10.1038/s41586-022-05473-8>
22. Graham PS, Kaidonis G, Abhary S et al (2018) Genome-wide association studies for diabetic macular edema and proliferative diabetic retinopathy. *BMC Medical Genetics* 19. <https://doi.org/10.1186/s12881-018-0587-8>
23. Jiang L, Zheng Z, Fang H, Yang J (2021) A generalized linear mixed model association tool for biobank-scale data. *Nature Genetics* 53(11): 1616–1621. <https://doi.org/10.1038/s41588-021-00954-4>.
24. Speed D, Balding DJ (2018) Sumher better estimates the SNP heritability of complex traits from summary statistics. *Nature Genetics* 51(2): 277–284. <https://doi.org/10.1038/s41588-018-0279-5>.

25. Speed D, Holmes J, Balding DJ (2020) Evaluating and improving heritability models using summary statistics. *Nature Genetics* 52(4): 458–462. <https://doi.org/10.1038/s41588-020-0600-y>.
26. Watanabe K, Taskesen E, Arjen van B, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8(1): 1826. <https://doi.org/10.1038/s41467-017-01261-5>.
27. Subramanian, A. et al. (2005) ‘Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles’, *Proceedings of the National Academy of Sciences*, 102(43), pp. 15545–15550. <https://doi.org/10.1073/pnas.0506580102>
28. Liberzon, A. et al. (2015) ‘The Molecular Signatures Database Hallmark Gene Set Collection’, *Cell Systems*, 1(6), pp. 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>.
29. Milacic, M. et al. (2023) ‘The reactome pathway knowledgebase 2024’, *Nucleic Acids Research*, 52(D1). <https://doi.org/10.1093/nar/gkad1025>.
30. Gusev A, Ko A, Shi H et al (2016) Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* 48(3): 245–252. <https://doi.org/10.1038/ng.3506>
31. Tan, H. et al. (2024) Rare mutations in *CCDC7* contribute to early-onset preeclampsia by inhibiting trophoblast migration and invasion, *MDPI*. Available at: <https://doi.org/10.3390/jpm14030253>.
32. Priyanka, P.P. and Yenugu, S. (2021) ‘Coiled-coil domain-containing (CCDC) proteins: Functional roles in general and male reproductive physiology’, *Reproductive Sciences*, 28(10), pp. 2725–2734. doi:10.1007/s43032-021-00595-2. <https://doi.org/10.1007/s43032-021-00595-2>
33. Qian, J. et al. (2022) ‘Male sex increases the risk of diabetic retinopathy in an urban safety-net hospital population without impacting the relationship between axial length and retinopathy’, *Scientific Reports*, 12(1). doi:10.1038/s41598-022-13593-4. <https://doi.org/10.1038/s41598-022-13593-4>
34. Ozawa, G.Y., Bearnse, M.A. and Adams, A.J. (2014) ‘Male–female differences in diabetic retinopathy?’, *Current Eye Research*, 40(2), pp. 234–246. <https://doi.org/10.3109/02713683.2014.958500>.
35. Saadane, A. et al. (2020) ‘Successful induction of diabetes in mice demonstrates no gender difference in development of early diabetic retinopathy’, *PLOS ONE*, 15(9). <https://doi.org/10.1371/journal.pone.0238727>
36. Sharma, A. et al. (2019) ‘Unraveling the role of genetics in the pathogenesis of diabetic retinopathy’, *Eye*, 33(4), pp. 534–541. doi:10.1038/s41433-019-0337-y. <https://doi.org/10.1038/s41433-019-0337-y>.
37. Luo, L. et al. (2022) ‘Secreted protein acidic and rich in cysteine mediates the development and progression of diabetic retinopathy’, *Frontiers in Endocrinology*, 13. <https://doi.org/10.3389/fendo.2022.869519>.
38. Yeh, S.-H. et al. (2015) ‘Differentiation of type 2 diabetes mellitus with different complications by proteomic analysis of plasma low abundance proteins’, *Journal of Diabetes & Metabolic Disorders*, 15(1). <https://doi.org/10.1186/s40200-016-0246-6>.
39. Raimondi, C. et al. (2016) ‘NRP1 function and targeting in neurovascular development and eye disease’, *Progress in Retinal and Eye Research*, 52, pp. 64–83. <https://doi.org/10.1016/j.preteyeres.2016.02.003>.
40. Thomas, L. et al. (2023) ‘BI-Y, an neuropilin-1 antagonist, enhances revascularization and prevents vascular endothelial growth factor- $\alpha$  induced retinal hyperpermeability in rodent models of retinopathies’, *Journal of Pharmacology and Experimental Therapeutics*, 385(3), pp. 214–221. <https://doi.org/10.1124/jpet.122.001473>.
41. Wang, J. et al. (2015) ‘The neuropilin-1 inhibitor, ATWLPPR peptide, prevents experimental diabetes-induced retinal injury by preserving vascular integrity and decreasing oxidative stress’, *PLOS ONE*, 10(11). <https://doi.org/10.1371/journal.pone.0142571>.
42. Kane, R., Godson, C., & O'Brien, C. (2008). Chordin-like 1, a bone morphogenetic protein-4 antagonist, is upregulated by hypoxia in human retinal pericytes and plays a role in regulating angiogenesis. *Molecular vision*, 14, 1138–1148. [PMC2435163](https://pubmed.ncbi.nlm.nih.gov/182435163/)
43. Ito, A. et al. (2018) ‘The relationship between peripheral nerve conduction velocity and ophthalmological findings in type 2 diabetes patients with early diabetic retinopathy’, *Journal of Ophthalmology*, 2018, pp. 1–7. <https://doi.org/10.1155/2018/2439691>.
44. Chen, W. et al. (2022) ‘Optical coherence tomography of the retina combined with color doppler ultrasound of the tibial nerve in the diagnosis of diabetic peripheral neuropathy’, *Frontiers in Endocrinology*, 13. doi:10.3389/fendo.2022.938659. <https://doi.org/10.3389/fendo.2022.938659>
45. Vaeggemose, M. et al. (2017) ‘Magnetic resonance neurography visualizes abnormalities in sciatic and tibial nerves in patients with type 1 diabetes and neuropathy’, *Diabetes*, 66(7), pp. 1779–1788. <https://doi.org/10.2337/db16-1049>.
46. Cassandrini, D. et al. (2017) ‘Congenital myopathies: Clinical phenotypes and New Diagnostic Tools’, *Italian Journal of Pediatrics*, 43(1). <https://doi.org/10.1186/s13052-017-0419-z>.
47. Fiori, A. et al. (2018) ‘Mesenchymal stromal/stem cells as potential therapy in diabetic retinopathy’, *Immunobiology*, 223(12), pp. 729–743. <https://doi.org/10.1016/j.imbio.2018.01.001>.

48. Gu, X. et al. (2018) 'Efficacy and safety of autologous bone marrow mesenchymal stem cell transplantation in patients with diabetic retinopathy', *Cellular Physiology and Biochemistry*, 49(1), pp. 40–52. <https://doi.org/10.1159/000492838>.
49. Aguilar, D. et al. (2009) 'Adverse association between diabetic retinopathy and cardiac structure and function', *American Heart Journal*, 157(3), pp. 563–568. <https://doi.org/10.1016/j.ahj.2008.10.019>.
50. Dai, Y. et al. (2014) 'Identification of chemokines and growth factors in proliferative diabetic retinopathy vitreous', *BioMed Research International*, 2014, pp. 1–9. <https://doi.org/10.1155/2014/486386>.
51. Viebahn, M., Barricks, M.E. and Osterloh, M.D. (1991) 'Synergism between diabetic and radiation retinopathy: Case report and review.', *British Journal of Ophthalmology*, 75(10), pp. 629–632. <https://doi.org/10.1136/bjo.75.10.629>.
52. Jeganathan, V.S., Wang, J.J. and Wong, T.Y. (2008) 'Ocular associations of diabetes other than diabetic retinopathy', *Diabetes Care*, 31(9), pp. 1905–1912. <https://doi.org/10.2337/dc08-0342>.
53. Skolik, S.A. and Caspers, L.E. (2012) 'Proliferative diabetic retinopathy in long-term diabetic patients with and without clinical osteoarthritis', *Eye*, 27(1), pp. 84–91. <https://doi.org/10.1038/eye.2012.187>.
54. Hussain, S.M. et al. (2015) 'Retinal arteriolar narrowing and incidence of knee replacement for osteoarthritis: A prospective cohort study', *Osteoarthritis and Cartilage*, 23(4), pp. 589–593. <https://doi.org/10.1016/j.joca.2015.01.007>.
55. Klein, R. et al. (2003) 'Retinal vascular abnormalities in persons with type 1 diabetes', *Ophthalmology*, 110(11), pp. 2118–2125. [https://doi.org/10.1016/S0161-6420\(03\)00863-7](https://doi.org/10.1016/S0161-6420(03)00863-7).
56. Forrest IS, Chaudhary K, Paranjpe I et al (2021) Genome-wide polygenic risk score for retinopathy of type 2 diabetes. *Human Molecular Genetics* 30(10): 952–960. <https://doi.org/10.1093/hmg/ddab067>.



## Figure legends

**Fig.1** Definition of diabetic retinopathy in type 1 diabetes based on UKB cohort. The control samples consist of non-DR patients, T1D with prolonged duration of diabetes, white British ethnicity, and those within the top 95% of HbA1c levels. The case samples include DR patients, T1D with prolonged duration of diabetes, white British ethnicity, and those within the top 95% of HbA1c levels. The GWAS model incorporates covariates age, sex, BMI, duration of diabetes, HbA1c levels, and the top 5 principal components.

**Fig.2** The Manhattan and regional plots of the GWAS analysis for diabetic retinopathy in type 1 diabetes. Figure below displays a cluster of top SNPs that reached a significant level in the Manhattan plot, with regional plots positioned above it. The regional plot on the left primarily includes SNPs within the *CCDC7* gene, while the regional plot on the right predominantly features SNPs within the *ITGB1* gene.

**Fig.3** The Q-Q plot of GWAS for DR in type 1 diabetes. This QQ plot visually assesses the distribution of observed p-values (plotted on the y-axis) against the expected distribution under the null hypothesis (plotted on the x-axis) for the GWAS on DR. The linearity of the points along the diagonal suggests that the observed p-values largely conform to the expected distribution, indicating well-controlled population stratification with genomic control lambda value being very close to 1.

**Fig.4** The circus plot of gene mapping of chromatin interactions and cis-eQTL. The top SNPs in each risk locus are displayed in the most outer layer (Manhattan plot). The second and third layers are chromosome rings, where genomic risk loci are highlighted in blue to indicate their locations. Genes mapped exclusively by chromatin interactions are colored orange.

**Fig.5** Gene expression heatmaps of positional mapping genes. This figure presents two heatmaps illustrating the average expression values (log<sub>2</sub> transformed) of selected positional mapping genes across diverse tissue types, as sourced from the GTEx v8 dataset. The upper heatmap displays gene expression in 30 general tissue types, while the lower heatmap shows expression across 54 specific tissue types. Expression values are color-coded from low (blue) to high (red), offering a detailed view of the tissue-specific expression patterns of these genes.

**Fig.6** Results of PheWAS for the top 12 SNPs in significant GWAS. This figure presents a multi-part analysis of 12 top SNPs in high LD and their phenome-wide association study results. The top figure (heatmap) displays the pheWAS results, where each cell represents the association of SNPs with various traits, color-coded by p-value significance. The middle panel, labeled 'Domain,' highlights the phenotypes concentrated across multiple

SNPs (marked in red) and categorizes them into their respective domains. The bottom figure is divided into two sections based on the regional SNP locations from the regional plot. Each section shows the significance of individual SNPs, with dashed lines representing the Bonferroni correction thresholds for each SNP.

**Fig.7** Scatter plot of SNP effects on osteoarthritis (exposure) and diabetic retinopathy (outcome) using multiple MR methods. This scatter plot displays the MR estimates of SNP effects on the relationship between osteoarthritis and diabetic retinopathy using multiple MR methods. The x-axis represents the SNP effects on osteoarthritis, and the y-axis represents the SNP effects on diabetic retinopathy. Different MR methods are indicated by lines of varying colors and styles: inverse variance weighted (green line), weighted median (blue line), MR Egger (red line), simple median (black line), and weighted mode (light blue line). Each point represents a specific SNP, with error bars showing confidence intervals.

**Figure S1** (A) The forest plots of the meta-analyses results from GWAS conducted by UKB study with European, FinnGen, GoSHARE, GoDARTS and Caucasian Australian study. Lines with green colors represent the effect sizes of SNPs from different cohorts. Each green line corresponds to a specific cohort, indicating the estimated effect size along with its 95% confidence interval. The overall effect beta (with CIs) is summarized at the bottom of the plot, shown as a red diamond. (B) Funnel plot of meta-analysis results. The horizontal axis represents the effect estimates beta, and the vertical axis denotes the standard error of each study included in the meta-analysis. The funnel plot is primarily symmetric around the vertical dashed line, which represents the overall effect estimate. This symmetry suggests a minimal risk of publication bias within the included studies. (C) Radial plot of meta-analysis results. Displaying the influence of individual studies on the overall effect estimate. The horizontal axis represents the square root of the inverse of the variance, which corresponds to the precision of each study. The vertical axis shows the standardized effect sizes. The grey shaded area delineates the 95% confidence region around the combined effect estimate, illustrating the consistency of individual study results with the overall meta-analytic conclusion.

**Figure S2** The Manhattan plot of gene analysis in the primary GWAS. Genome-wide significance was defined and established after Bonferroni correction ( $p = 0.05/19295 = 2.59 \times 10^{-6}$ ), as illustrated by the red dashed line in the plot.

**Figure S3** Tissue expression results on 30 general tissue types and 53 specific tissue types by GTEx in the FUMA. The dashed line shows the cut-off  $p$  value for significance with Bonferroni adjustment for multiple hypothesis testing. No tissues achieved statistical significance

**Figure S4** Tissue Specificity of Differentially Expressed Genes in GTEx v8 Data. The figure presents bar plots illustrating the results of an enrichment test for differentially expressed genes (DEGs) across two datasets: GTEx v8 with 54 specific tissue types (left panel) and GTEx v8 with 30 general tissue types (right panel). Each bar represents the  $-\log_{10}$  transformed  $p$  values of DEGs tested for enrichment in positional mapping genes within each tissue type. Significant enrichment is highlighted in the brain cortex in the 54 tissue type dataset, shown by the red bar, indicating a notable down-regulation of DEGs. This analysis focuses on the enrichment of positional mapping genes and avoids the broader SNP  $p$ -value distribution typically used in MAGMA tissue expression analyses.

**Figure S5** Enrichment Analysis of Genetic Associations. This figure presents the results of enrichment analyses for positional gene sets (MSigDB c1), Reactome pathways (MSigDB c2), and GWAS catalog reported genes related to specific traits. The analysis reveals varying degrees of gene overlap and enrichment significance across different datasets. The proportion of overlapping genes and their enrichment  $p$ -values are displayed for each category, highlighting notable interactions in biological pathways.

**Figure S6** Gene Association Patterns Across Human Tissues from MSigDB Chr10p11. This heatmap depicts the expression and association patterns of multiple genes across a variety of human tissues, as sourced from the Novartis Human Tissue Compendium. Notably, the genes *ITGB1* and *NRP1* are located under the same detailed node in the tree diagram.

**Figure S7** Pathway Diagram of *ITGB1* from Reactome. (A) showed interaction pathways involving *LICAM*, *ITGB1*, and *NRP1*. *LICAM* is shown to participate in the assembly of integrin complexes. (B) depicts *CHL1*'s role in facilitating interactions between *ITGB1* and *NRP1*.

**Figure S8** Mendelian Randomization Analysis between Osteoarthritis and Diabetic Retinopathy. Utilizing genetic variants as instrumental variables to estimate the causal effects between OA and DR. The first plot (top panel) shows the forest plot, funnel plot, and leave-one-out plot of MR results when osteoarthritis is considered the exposure and diabetic retinopathy the outcome. The second plot (bottom panel) reverses the roles, treating DR as the exposure and OA as the outcome. The plot of SNP effects using multiple MR methods is shown, as well as the forest plot, funnel plot, and leave-one-out plot.

Table 3. *p* values of top significant SNP rs184619214 from UKB GWAS in replication cohorts and meta-analysis results <sup>16,21,23,(Rajendrakumar,2022)</sup>

GWAS	Replication results for significant SNP rs184619214 from UKB										Meta-analysis results				
	SNPs(LDproxies)	CHR	POS	EA	NEA	AF1	BETA	SE	<i>P</i>	DISTANCE	ESTIMATE	SE	<i>P</i>	Q_heterogeneity	Q_Pval
African American	rs139204448	10	31864160	C	T	0.017	/	/	0.050	-930,301	0.329	0.05	4.47E-11	8.91	0.11
European	rs78963187	10	32073079	G	T	0.035	0.453	0.13	0.040	-721,382					
FinnGen	rs184619214	10	32505533	G	A	0.054	0.120	0.31	0.699	0					
GoSHARE	rs11008927	10	32703688	A	C	0.250	0.126	0.16	0.436	-90,773					
GoDARTS	rs17230291	10	32646193	G	A	0.036	0.168	0.10	0.095	-148,268					
Caucasian Australian	rs2488320	10	33238917	T	C	0.354	0.245	0.17	0.150	404,450					

"/" indicates missing value. Chr: chromosome, EA: effect allele, NEA: non-effect allele, AF1: frequency of effect allele.

The effect alleles and beta values of the SNPs in the UKB were adjusted to ensure that the strand direction was consistent across all cohorts during meta-analysis.