中山大學 中山眼科中心 ZHONGSHAN OPHTHALMIC CENTER, SUN YAT-SEN UNIVERSITY

Contents lists available at Lighthouse Scholar

EYE SCIENCE

journal homepage: journal.gzzoc. com

doi: 10.12419/es24012703
View this article at: https://dx.doi.org/10.12419/es24012703

· Original Article ·

Identification of novel drug targets for diabetic retinopathy: proteome-wide mendelian randomization and colocalization analyses

Shaopeng Yang(杨少鹏)[1,2§], Ziyu Zhu(朱梓瑜)[1,2§], Riqian Liu(刘日乾)[1§], Wenyong Huang(黄文勇)[1], Wei Wang(王伟)[1,2]

1. State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, Guangzhou 510060, China
2. Hainan Eye Hospital and Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Haikou 570100, China

HIGHLIGHTS

· This Proteome-wide mendelian randomization (PW-MR) study identified 24 plasma proteins linked to diabetic retinopathy (DR), including 18 previously unrecognized ones, offering new avenues for drug targeting. Four proteins highly supported by evidence from co-location analyses were identified, along with an overview of their status in drug development.

· The study focused on cis-pQTLs that directly influence protein functionality and meticulously filtered instrumental variables to reduce weak instrumental biases. By combining bidirectional MR, Bayesian co-localization, and phenotype scanning, confounding factors and reverse causation have been effectively minimized.

· Highlighting the synergy between proteomics and genomics, the study underscored the potential for novel drug targets. The overlap of causal proteins in different diseases suggested broader implications for drug target identification.

Abstract: **Background:** Diabetic retinopathy (DR) urgently needs novel and effective therapeutic targets. Integrated analyses of plasma proteomic and genetic markers can clarify the causal relevance of proteins and discover novel targets for diseases, but no systematic screening for DR has been performed. **Methods:** Summary statistics of plasma protein quantitative trait loci (pQTL) were derived from two extensive genome-wide analysis study (GWAS) datasets and one systematic review, with over 100 thousand participants covering thousands of plasma proteins. DR data were sourced from the largest FinnGen study, comprising 10,413 DR cases and 308,633 European

Received date: 2024-01-27; **Accepted date:** 2024-03-20; **Published online:** 2024-03-28

§Contributing equally

Corresponding author: Wenyong Huang, E-mail:huangwenyong@gzzoc.com; Wei Wang, E-mail:wangwei@gzzoc.com.

Full Text

controls. Genetic instrumental variables were identified using multiple filters. In the two-sample MR analysis, Wald ratio and inverse variance-weighted (IVW) MR were utilized to investigate the causality of plasma proteins with DR. Bidirectional MR, Bayesian Co-localization, and phenotype scanning were employed to test for potential reverse causality and confounding factors in the main MR analyses. By systemically searching druggable gene lists, the ChEMBL database, DrugBank, and Gene Ontology database, the druggability and relevant functional pathways of the identified proteins were systematically evaluated. **Results:** Genetically predicted levels of 24 proteins were significantly associated with DR risk at a false discovery rate <0.05 including 11 with positive associations and 13 with negative associations. For each standard deviation increase in plasm protein levels, the odds ratios (ORs) for DR varied from 0.51 (95% CI: 0.36-0.73; P=2.22×10^{-5}) for tubulin polymerization-promoting protein family member 3 (TPPP3) to 2.02 (95% CI: 1.44-2.83; P=5.01×10^{-5}) for olfactomedin like 3 (OLFML3). Bidirectional MR indicated there was no reverse causality that interfered with the results of the main MR analyses. Four proteins exhibited strong co-localization evidence (PH4 ≥0.8): cytoplasmic tRNA synthetase (WARS), acrosin binding protein(ACRBP), and intercellular adhesion molecule 1 (ICAM1) were negatively associated with DR risk, while neurogenic locus notch homolog protein 2 (NOTCH2) showed a positive association. No confounding factors were detected between pQTLs and DR according to the phenotypic scan. Drugability assessments highlighted 6 proteins already in drug development endeavor and 18 novel drug targets, with metalloproteinase inhibitor 3 (TIMP) currently in phase I clinical trials for DR. GO analysis identified 18 of 24 plasma proteins enriching 22 pathways related to cell differentiation and proliferation regulation. **Conclusions:** Twenty-four promising drug targets for DR were identified, including four plasma proteins with particular co-localization evidence. These findings offer new insights into DR's etiology and therapeutic targeting, exemplifying the value of genomic and proteomic data in drug target discovery.

Keywords: plasma proteome; mendelian randomization; therapeutic targets; genome-wide analysis study; colocalization analysis; diabetic retinopathy

Cite this article as: Yang SP, Zhu ZY, Liu RQ, Huang WY, Wang W. Identification of novel drug targets for diabetic retinopathy: proteome-wide mendelian randomization and colocalization analyses. Eye Science, 2024, 1(1): 26-44. doi: 10.12419/es24012703

INTRODUCTION

Diabetic retinopathy (DR) is the leading cause of blindness in the working-age population, affecting 103.12 million individuals globally.[1] Although the VEGF pathway plays a pivotal role in DR pathogenesis, responses to anti-VEGF treatments vary significantly among patients.[2] Even with close monitoring and frequent injections, achieving desired outcomes remains challenging, with up to 40% cases of diabetic macular edema (DME) experience unresolved edema.[3] This underscores the urgent need for new DR therapeutic targets.

The plasma proteome, reflecting human health and disease, offers crucial insights into an individual's molecular profile.[4] High-throughput proteomics

techniques have revolutionized the discovery of disease-related proteins (e.g., dementia, heart failure, cancer) at an unprecedented pace.[5-6] Moreover, the extensive coverage and ready accessibility of plasma proteins make them promising targets for pharmaceutical interventions, with 75% of 2017 FDA-approved drugs targeting human proteins.[7]

Mendelian randomization (MR) minimizes confounding and enables causal inference for exposure (plasma proteins)-outcome (DR) relationships. Genome-wide association studies (GWAS) have identified over 18,000 protein quantitative trait loci (pQTLs), forming a foundation for MR analysis.[8-14] Based on these, proteome-wide MR (PW-MR) has recently yielded significant insights into the etiology of various conditions, including stroke, psychiatric disorders, and breast cancer.[15-17] Moreover, PW-MR offers exceptional prospects for identifying and prioritizing drug targets compared to other approaches.[18] However, no large-scale PW-MR study for DR is currently available. Therefore, this study aimed to identify plasma proteins that could serve as potential therapeutic targets for DR by integrating human plasma proteomic and genomic data, and systematically assessing the druggability of potential DR therapeutic target proteins.

MATERIALS AND METHODS

Figure 1 presents the study design. In brief, pQTLs data from seven large-scale proteomic studies were utilized to examine their correlation with DR using a two-stage PW-MR methodology. Bayesian co-localization was subsequently applied to validate the causal linkage between proteins and DR. A thorough assessment of the drugability of the identified proteins followed, with a specific emphasis on their potential as therapeutic targets.

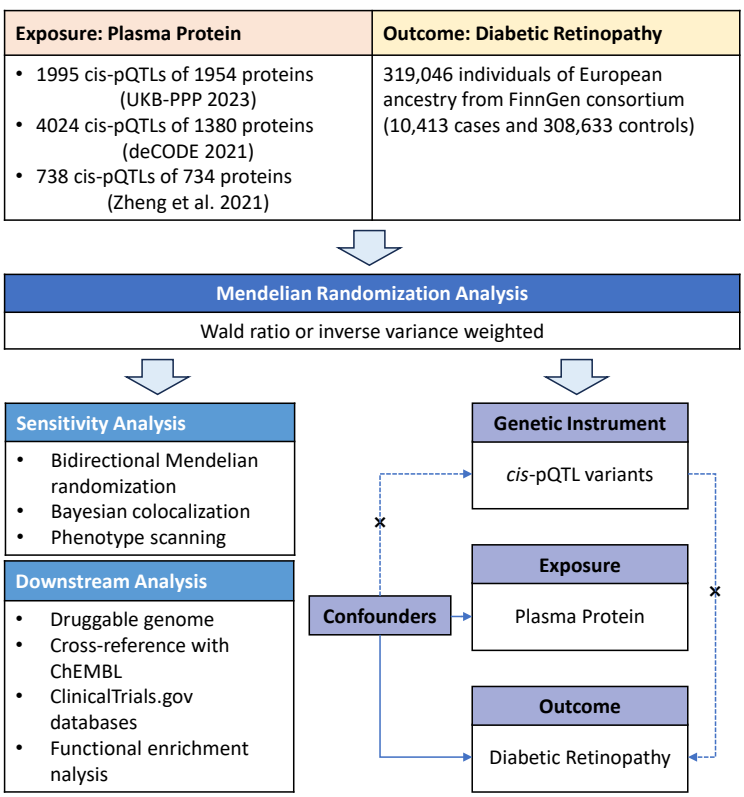

Figure 1 A flow chart of the study design and a schematic illustration of cis-MR

Study Exposures: pQTL

Instrument variable (IV) for plasma proteins, specifically *cis*-pQTLs, were sourced from two proteomic GWASs and a PW-MR study. These *cis*-pQTLs were gathered from two large-scale GWAS datasets: the UK Biobank Pharma Protein Project (UKB-PPP)[19] and the deCODE Health study.[9] We also obtained data from Zheng et al.,[20] which encompassed 734 plasma proteins with cis-pQTLs. Our selection criteria for inclusion included: (I) genome-wide significance ($P < 5 \times 10^{-8}$); (II) independence (linkage disequilibrium [LD] clump $R^2 < 0.001$); and (III) proximity as *cis*-pQTLs within 1 Mb of the respective protein-encoding gene.

We identified 1,995 *cis*-pQTLs for 1,954 plasma proteins from the UKB-PPP dataset (Olink platform, n=54,306), and 4,024 *cis*-pQTLs for 1,380 circulating proteins from the deCODE Health study (Somascan platform, n=35,559). Furthermore, we integrated data on 738 *cis*-pQTLs linked to 734 plasma proteins, as reported by Zheng et al.[20] These data were derived from GWAS studies utilizing the Somascan platform by Sun et al.,[13] Emilsson et al.,[12] Suhre et al.,[21] and Yao et al.,[11] as well as the GWAS study conducted by Folkersen et al. using the Olink platform (Table 1).

Study Outcome: DR

Data on the plasma protein-related SNPs and DR association were extracted from the FinnGen R9 study, involving 10,413 individuals of European descent as cases and 308,633 control subjects. To ensure the consistency and precision of the clinical endpoint, the diagnostic criteria and levels for DR were constructed from the register codes using the Finnish version of the International Classification of Diseases, 10th revision (ICD-10) diagnosis codes harmonized with definitions from ICD-8 and ICD-9.[22] The whole process has been subjected to multiple evaluations by specialized committees. Stringent protocols were applied to guarantee ethnic homogeneity across the exposure and outcome datasets. Ethical approval and informed consent were acquired from the relevant institutional review boards, acknowledging the study's reliance on established publications and publicly accessible databases.

STATISTICAL ANALYSIS

Two-sample MR Analysis

Genetic tool robustness was assessed through the integration of plasma protein *cis*-pQTLs from

Table 1 Data sources for studied phenotypes

Study	Phenotype	Case	Control	Platform	Country
UKB-PPP	Plasma protein	54,306	NA	Olink	United Kindom
deCODE	Plasma protein	35,559	NA	SomaScan	Iceland
Sun et al.	Plasma protein	3,301	NA	SomaScan	United Kindom
Emilsson et al.	Plasma protein	5,457	NA	SomaScan	Iceland
Suhre et al.	Plasma protein	997	NA	SomaScan	Germany
Folkersen et al.	Plasma protein	3,394	NA	Olink	Finland, France, Italy, Netherlands, Sweden
Yao et al.	Plasma protein	7,333	NA	SomaScan	America
FinnGen	DR	10,413	308,633	NA	Finland

UKB-PPP=The UK Biobank Pharma Proteomics Project; DR=diabetic retinopathy; NA=not available.

three distinct sources. This involved determining the proportion of variance (R^2) explained by the genetic IV for each risk factor and assessing the strength of association (F-statistics) between these genetic tools and the risk factors.[23] To ensure dataset homogeneity, plasma proteins lacking viable genetic tools for the outcome were excluded. An R^2 of IV (R^2> 0.1%) identified 6,133 *cis*-pQTLs representing 2,358 proteins as IV for the analysis.

In the two-sample MR analysis, plasma proteins served as the exposure, with DR as the endpoint. The analysis utilized the 'TwoSampleMR' software package. Association estimates between plasma proteins and DR were derived using the Wald ratio (for proteins linked to a single *cis*-pQTL) and inverse variance-weighted MR (for proteins associated with two or more *cis*-pQTLs). Confidence intervals (CIs) were computed using the delta method.[24] Results were presented as MR odds ratios (OR) and their corresponding 95% CIs, quantifying the effect of a 1-SD change in gene-predicted plasma protein levels on DR risk. To mitigate the risk of false positives arising from multiple comparisons, the Benjamini and Hochberg correction was applied, setting the False Discovery Rate (FDR) at 0.05. Furthermore, a more stringent correction using the Bonferroni method was implemented, with significance established at $P< 2.12\times10^{-5}$ (0.05/2,358). For *cis*-pQTLs significantly associated with DR in the MR study post-Bonferroni correction, the proportion of variance explained (R^2) for these specific proteins was calculated.

Reverse Causality Detection

Bidirectional MR analysis explored the potential reverse causality in the association between DR risk and identified plasma protein levels. Complete GWAS data for proteins associated with DR were sourced from three prior studies. Effect estimates were derived via the MR-IVW method and validated using MR-Egger, weighted median, simple mode, and weighted mode approaches. Statistical significance was set at P<0.05. A Bonferroni correction was applied to account for multiple comparisons, setting significance at $P < 2.08 \times 10^{-3}$.

Bayesian Co-localization Analysis

Co-localization analysis was used to investigate the potential influence of linkage disequilibrium (LD) on the association between identified proteins and DR. This analysis employed a Bayesian model to assign posterior probabilities to five hypotheses related to the presence of a shared variant between the two traits.[25] The hypotheses considered were as follows: (I) no causal variant exists for either protein or DR in the genomic locus (H0); (II) a causal variant is specific to one protein (H1); (III) a causal variant is specific to one DR (H2); (IV) two distinct causal variants independently affect protein and DR (H3); and (V) a shared causal-variant influences both protein and DR (H4). We evaluated the posterior probabilities for H3 (PH3) and H4 (PH4). High-strength co-localization support was indicated when the PH4 was ≥0.8, while moderate-strength co-localization support was defined as 0.5 < PH4 < 0.8. These analyses were conducted using the 'coloc' package in R software (version 4.4.1).

MR Assumptions and Phenotype Scanning

MR relies on three foundational assumptions. Firstly, it requires robust associations between SNPs and exposures. Given that all *cis*-pQTLs exhibited significant associations with their respective protein levels at a genome-wide scale (with P-values < 5×10^{-8} and were identified as the SNPs with the lowest P-values within 1 Mb of the transcriptional start sites of the relevant

genes, this assumption is unlikely to pose issues. To gauge the strength of the relationship between genetic instruments and exposure, we computed the F-statistic for the *cis*-pQTLs, typically exceeding 10 for an SNP to be considered a robust MR tool.

The second MR hypothesis posits that *cis*-pQTLs should not be linked to a confounding factor in the relationship between exposure and outcome. Ancestry represents a potential source of such confounding, which can be mitigated by ensuring that all participants in the GWAS share European ancestry.

The third MR hypothesis requires that genetic variants do not influence the outcome except through their impact on exposure (horizontal pleiotropy). In this study, potential bias stemming from horizontal pleiotropy is significantly reduced, as the primary MR tool is a *cis*-acting SNP located within 300 kb of the gene encoding the protein. To explore associations beyond our primary analysis, we employed the 'phenoscanner' software package for a comprehensive scan of previously published large-scale genetic association studies.[26] This scan aimed to unveil potential connections between the identified pQTLs and other traits. SNPs were considered pleiotropic if they met specific criteria: (I) their associations achieved genome-wide significance ($P < 5 \times 10^{-8}$); (II) the GWAS involved populations of European ancestry; and (III) the SNPs exhibited associations with established risk factors for DR, including metabolic traits, proteins, or clinical features. In addition, we calculated the LD r^2 between pQTLs for prioritized proteins to unveil potential associations.

Downstream Analysis of Drug Target Proteins

Plasma proteins represent a significant source of potential drug targets. To determine if MR-preferred proteins overlap with genes within the druggable genome, we extracted druggability profiles of target proteins from the updated list of druggable genes.[27] For assessing the clinical development status of drug candidates targeting these proteins, we interrogated the ChEMBL database (version 33) for data on drug molecule classifications, approved indications, and clinical trial outcomes.[28] Furthermore, data regarding drug-target proteins was obtained by searching DrugBank (https://wwws.go.drugbank.com) and manually screening for each candidate target.[29] To assess druggability, proteins were categorized into four groups: approved (for specific proteins with one or more approved drugs); clinical trials (for drugs currently in clinical investigation); preclinical (representing drugs in preclinical development); druggable (relating to proteins designated as potential druggable targets but not found in drug databases). To comprehensively reveal the potential mechanism of MR priority proteins involved in the pathological process of DR, gene function enrichment analysis of all proteins was further completed based on the information from the Gene Ontology database (http://geneontology.org).[30]

Data Availability

Genome-wide summary statistics for Cis-pQTL can be found in the original study,[9, 19-20] while GWAS summary statistics for DR are accessible via the FinnGen consortium website (R9 release) at https://www.finngen.fi/.

RESULTS

Figure 1 presents the framework of this study. The three separate datasets consolidated *cis*-pQTL data pertaining to plasma proteins, resulting in the discovery of 6,362 *cis*-pQTLs associated with 2,602 unique proteins (Table 1). The subsequent MR analysis incorporated

2,358 proteins from the FinnGen R9 dataset, following a rigorous screening process for genetic instrument tools with an R^2 exceeding 0.01%, and the exclusion of plasma proteins lacking suitable genetic instruments for further analysis.

Genetic Association Between Plasma Proteins and DR

Using 6,362 *cis*-pQTLs as genetic instruments for the corresponding proteins, the MR analysis unveiled associations between 24 plasma proteins and DR risk, following FDR correction (Table 2). With each SD increase in genetically predicted protein levels, the OR for DR spanned from 0.51(95% CI: 0.36–0.73; P=2.22×10^{-4}) for tubulin polymerization-promoting protein family member 3 (TPPP3), to 2.02 (95% CI: 1.44–2.83; P=5.01×10^{-5}) for olfactomedin like 3 (OLFML3). There are 11 proteins positively associated with DR risk and 13 proteins inversely associated (Table 2, Figure 2). After Bonferroni correction (P < 2.16×10^{-5}), the MR analysis sustained significant associations for 8 proteins with DR risk, with all causal proteins displaying F-statistic values exceeding 10, excluding bias from weak instrumental variables.

Sensitivity Analyses

To evaluate the robustness of estimates derived from primary MR analyses, multiple sensitivity analyses were conducted on the 24 causal proteins. In the bidirectional MR analysis, after applying a stringent Bonferroni correction (P < 2.08 × 10^{-3}), none of the causal plasma proteins showed signs of reverse causation (Table 3).

To examine potential confounding due to LD, colocalization analysis was conducted to assess whether genetic factors influencing plasma protein levels intersected causally with those affecting DR. Among the 24 DR-associated proteins, acrosin-binding protein (ACRBP), intercellular adhesion molecule 1 (ICAM1), neurogenic locus notch homolog protein 2 (NOTCH2), and cytoplasmic tRNA synthetase (WARS) exhibited strong evidence of co-localization (PH4≥0.8). The endoribonuclease LACTB2 (LACTB2), integrin beta-7 (ITGB7), signal regulatory protein gamma (SIRPG), tubulin polymerization-promoting protein family member

Table 2 Plasma proteins significantly associated with diabetic retinopathy after false discovery rate (FDR) correction

Protein	Chr	Position	rsnumber *cis*-pQTL	EAF	EA	MR analysis						Source
						β (SE)	OR	95%CI	P value	R^2	F statistics	
ACRBP	12	6754191	rs7959658	0.36	G	0.37(0.10)	1.45	1.19-1.76	1.84×10^{-4}	0.012	404.99	UKB-PPP
AIF1	6	31603591	rs2261033	0.43	G	-0.47(0.13)	0.62	0.48-0.80	2.90×10^{-4}	0.006	192.50	UKB-PPP
BCL2L15	1	114428086	rs1230715	0.67	C	0.32(0.09)	1.38	1.16-1.64	3.41×10^{-4}	0.014	489.49	UKB-PPP
BTN3A2	6	26367833	rs9393710	0.13	G	-0.10(0.03)	0.91	0.86-0.95	9.65×10^{-5}	0.183	7,552.74	UKB-PPP
GALNT3	2	166723533	rs2116546	0.30	C	-0.17(0.04)	0.85	0.78-0.92	4.32×10^{-5}	0.065	2,364.25	UKB-PPP
GCKR	2	27730940	rs1260326	0.34	C	0.67(0.17)	1.96	1.39-2.76	1.17×10^{-4}	0.003	106.48	deCODE
ITGB7	12	53593632	rs12232003	0.07	C	0.40(0.12)	1.49	1.19-1.87	4.96×10^{-4}	0.006	193.84	UKB-PPP
LACTB2	8	71529866	rs191588099	0.01	A	-0.39(0.11)	0.67	0.54-0.84	3.74×10^{-4}	0.016	557.73	UKB-PPP

Table 2 (continued)

Protein	Chr	Position	rsnumber *cis*-pQTL	EAF	EA	MR analysis						Source
						β (SE)	OR	95%CI	*P* value	R^2	F statistics	
NOTCH2†	1	120437884	rs2641348	0.11	G	-0.51(0.12)	0.60	0.48-0.75	8.95×10^{-6}	0.006	205.43	UKB-PPP
OLFML3	1	114489769	rs4381184	0.30	C	0.70(0.17)	2.02	1.44-2.83	5.01×10^{-5}	0.003	104.78	deCODE
SIRPG	20	1616206	rs6043409	0.32	G	0.37(0.10)	1.45	1.19-1.78	3.11×10^{-4}	0.009	311.86	deCODE
TNXB†	6	32030700	rs9267797	0.01	T	0.27(0.05)	1.31	1.18-1.45	4.46×10^{-7}	0.078	2,870.72	UKB-PPP
TPPP3	16	67332365	rs13334364	0.08	C	-0.66(0.18)	0.51	0.36-0.73	2.22×10^{-4}	0.002	60.84	UKB-PPP
BPIFA2*	20	31742114	rs141715080	0.97	C	-0.26(0.07)	0.77	0.67-0.88	1.96×10^{-4}	0.006	31.85	Emilsson et al.
	20	31812923	rs3818222	0.49	A					0.002	87.86	deCODE
	20	31754236	rs6059134	0.04	A					0.010	339.32	UKB-PPP
CLPS*	6	35753227	rs2766588	0.61	C	0.12(0.03)	1.13	1.06-1.20	2.01×10^{-4}	0.123	4,726.28	UKB-PPP
	6	35741963	rs2817064	0.39	A					0.063	2,369.49	deCODE
COL18A1*	21	46865359	rs11909509	0.21	A	0.21(0.05)	1.23	1.11-1.36	9.83×10^{-5}	0.008	274.54	deCODE
	21	45922658	rs150169381	0.03	T					0.002	56.59	deCODE
	21	46932618	rs17004785	0.08	C					0.003	110.63	deCODE
	21	46906711	rs2274809	0.65	G					0.008	56.92	Yao et al.
	21	46897214	rs9976834	0.18	T					0.014	479.10	UKB-PPP
GBP1*	1	90150895	rs12410795	0.02	C	-0.10(0.03)	0.91	0.86-0.95	1.32×10^{-4}	0.009	328.03	deCODE
	1	89523756	rs148526074	0.03	T					0.088	3,263.69	UKB-PPP
	1	88852286	rs9428012	0.09	A					0.002	76.81	deCODE
	1	88890193	rs144525174	0.01	C					0.017	610.87	deCODE
GSTA1*†	6	52477669	rs111461400	0.04	A	0.13(0.02)	1.13	1.08-1.19	9.35×10^{-8}	0.002	88.01	deCODE
	6	52227833	rs12525609	0.15	G					0.001	47.98	deCODE
	6	52662153	rs2290758	0.57	A					0.083	300.26	Sun et al.
	6	52813089	rs35928674	0.03	T					0.002	57.49	deCODE
	6	52613697	rs4712015	0.43	A					0.057	2,128.28	deCODE
	6	52665977	rs6458871	0.57	T					0.057	2,024.88	UKB-PPP
ICAM1*†	19	10404519	rs17852402	0.12	A	0.06(0.01)	1.06	1.04-1.08	2.54×10^{-8}	0.159	6,711.17	deCODE
	19	9998774	rs35193916	0.03	T					0.006	202.49	deCODE
	19	11342599	rs4804153	0.06	T					0.003	90.16	deCODE
	19	10395683	rs5498	0.43	G					0.690	7,330.26	Sun et al.
	19	10386599	rs62130660	0.38	G					0.287	14,264.93	deCODE
	19	10579474	rs7256672	0.36	G					0.008	282.17	deCODE

Table 2 (continued)

Protein	Chr	Position	rsnumber *cis*-pQTL	EAF	EA	MR analysis						Source
						β (SE)	OR	95%CI	*P* value	R^2	*F* statistics	
MANSC4*	12	27925037	rs34311349	0.23	A	-0.06(0.01)	0.94	0.92-0.97	9.83×10^{-5}	0.243	10,861.63	UKB-PPP
	12	27927881	rs36138811	0.23	C					0.136	517.25	Sun et al.
	12	27965469	rs61915452	0.11	C					0.012	415.65	deCODE
	12	27858007	rs61917498	0.13	T					0.063	2,393.09	deCODE
	12	27799008	rs73294067	0.02	G					0.003	95.82	deCODE
PAM*†	5	101578769	rs1116046	0.08	T	-0.11(0.02)	0.90	0.87-0.93	5.47×10^{-10}	0.074	2,806.42	deCODE
	5	101195428	rs115362035	0.02	T					0.010	352.89	deCODE
	5	102034963	rs13188927	0.38	T					0.106	4,187.68	deCODE
	5	102346866	rs149802978	0.05	G					0.110	4,178.99	UKB-PPP
	5	101565016	rs191278394	0.03	C					0.018	651.25	deCODE
	5	102418604	rs257309	0.35	G					0.095	347.24	Sun et al.
	5	102290553	rs4703249	0.04	A					0.002	58.64	deCODE
	5	101853615	rs76028248	0.01	A					0.002	87.70	deCODE
TIGIT*†	3	114013194	rs6792290	0.29	G	-0.65(0.14)	0.52	0.40-0.69	2.33×10^{-6}	0.004	141.01	UKB-PPP
	3	114014080	rs73238029	0.24	C					0.002	83.41	deCODE
TIMP3*†	22	33113151	rs111284156	0.05	G	0.06(0.01)	1.06	1.03-1.09	2.80×10^{-6}	0.023	825.98	deCODE
	22	33677720	rs117555418	0.04	A					0.007	243.93	deCODE
	22	33595991	rs138361770	0.01	T					0.001	39.47	deCODE
	22	34127425	rs149539274	0.02	A					0.001	35.88	deCODE
	22	33165020	rs2097326	0.28	G					0.351	540.64	Suhre et al.
	22	33181787	rs35201299	0.12	A					0.033	1,213.51	deCODE
	22	33162831	rs5749504	0.73	T					0.164	6,641.86	UKB-PPP
	22	33140654	rs9606990	0.19	T					0.079	3,050.71	deCODE
	22	32822217	rs9609547	0.04	T					0.003	95.78	deCODE
WARS*†	14	100842637	rs2273804	0.26	G	0.32(0.05)	1.38	1.26-1.52	3.17×10^{-11}	0.024	821.73	UKB-PPP
	14	100813077	rs941923	0.77	C					0.014	77.29	Emilsson et al.

Source indicates the protein GWAS providing the estimate of the effect of the *cis*-pQTL on protein level. Results express changes in type 1 diabetes risk per 1-SD increase in protein level. The human genome reference sequence version GRCh37 is employed as the reference genome panel. EA=effect allele; EAF=effect allele frequency; *β*=effect on diabetic retinopathy; OR=odds ratio.

* Mendelian randomization using the inverse variance weighted methods.

† Plasma protein passed Bonferroni correction ($P < 2.16\times10^{-5}$).

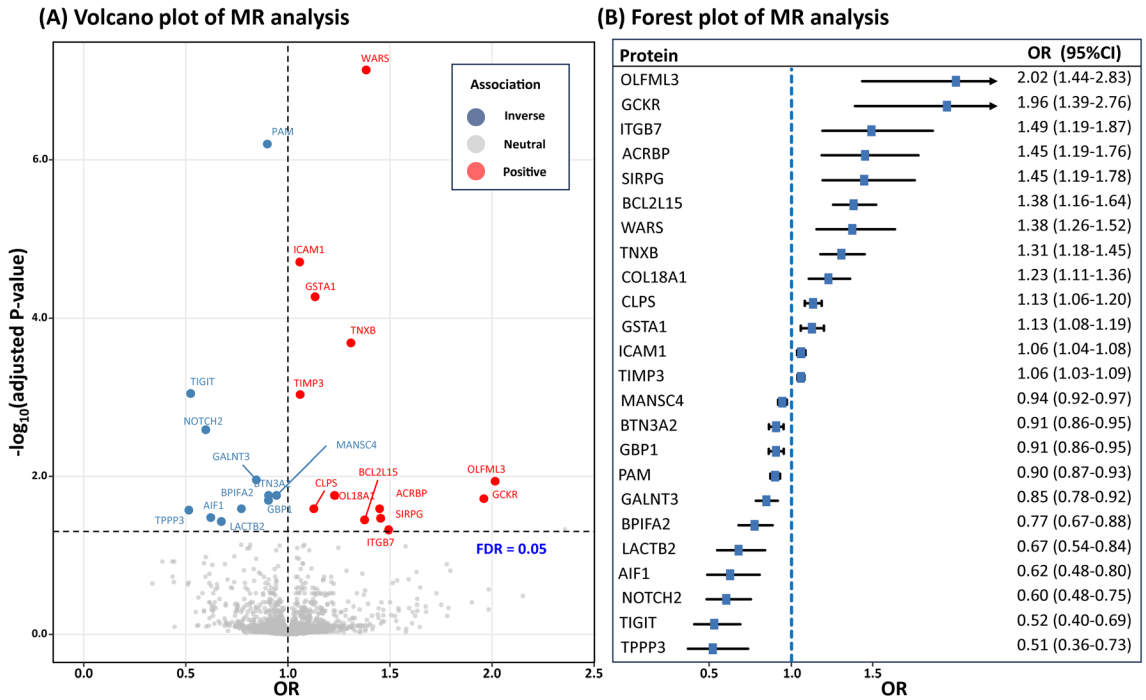

Figure 2 Result summary of MR and colocalization analysis on the associations between plasma proteins and the risk of diabetic retinopathy. OR=odds ratio; FDR=false discovery rate

3 (TPPP3), and polypeptide N-acetylgalactosaminyltransferase 3 (GALNT3) demonstrated moderate-strength evidence of co-localization analysis (0.8 > PH4≥0.5) (Table 3; Figure 3). These findings suggest a potential shared pool of genetic variants between these proteins and DR.

The phenotypic scan of *cis*-pQTLs linked to all DR-associated proteins revealed connections of CGKR (*rs1260326*), NOTCH2 (*rs2641348*), PAM (*rs1116046*), and TNXB (*rs9267797*) with diabetes mellitus. However, no such associations were detected between these *cis*-pQTLs and DR (Table 3). In addition, the phenotypic scan unveiled correlations of specific DR-associated proteins, including AIF1 (*rs2261033*), BTN3A2 (*rs9393710*), GCKR (*rs1260326*), GSTA1 (*rs2290758*), ICAM1 (*rs5498*), and TIMP3 (*2097326*), with various blood constituents. Particular plasma protein *cis*-pQTLs displayed links to various diseases, such as IgA deficiency, idiopathic membranous nephropathy, rheumatoid arthritis, primary sclerosing cholangitis, ankylosing spondylitis, multiple sclerosis, psoriasis, schizophrenia, sarcoidosis, benign prostatic hyperplasia, and schizophrenia, gout, nonalcoholic fatty liver disease. LACTB2 (*rs191588099*) showed a connection with cerebrovascular disease, NOTCH2 (*rs2641348*) with Crohn's disease, and TNXB (*rs9267797*) with hypothyroidism and rheumatoid arthritis (Table 3). Relaxing the association significance threshold to ($P<1\times10^{-5}$) revealed no *cis*-pQTLs for any plasma proteins maintaining associations with DR, thus affirming the robustness of the MR analysis results in the presence of potential confounding factors.

Downstream Analysis of MR-derived Protein Targets

To evaluate the druggability and developmental

Table 3 Summary of reverse causality detection, Bayesian co-localization analysis and phenotype scanning on 24 potential causal proteins

Protein	Bidirectional MR		Co-localization		Previously reported associations
	OR (95%CI)	*P*	P_{H3}	P_{H4}	
WARS	1.00(0.97-1.03)	0.88	0.039	0.960	NA
NOTCH2	0.98(0.96-1.01)	0.21	0.101	0.898	Crohn's disease, DM
ACRBP	0.99(0.96-1.01)	0.24	0.049	0.849	NA
ICAM1	0.99(0.93-1.06)	0.73	0.062	0.821	Blood cell, body measurement, lipid metabolism
LACTB2	1.00(0.98-1.02)	0.99	0.090	0.757	CVD
ITGB7	1.03(1.00-1.06)	0.02	0.095	0.731	NA
SIRPG	0.97(0.93-1.03)	0.41	0.079	0.716	NA
TPPP3	0.99(0.97-1.02)	0.52	0.244	0.710	Body measurement
GALNT3	1.01(0.98-1.03)	0.71	0.229	0.548	NA
GBP1	1.01(0.99-1.03)	0.52	0.259	0.235	NA
GSTA1	1.03(1.00-1.05)	0.04	0.272	0.339	Blood cell
PAM	0.99(0.97-1.02)	0.62	0.665	0.240	DM, body measurement
TIGIT	1.02(1.00-1.05)	0.08	0.799	0.166	NA
MANSC4	1.02(1.00-1.04)	0.10	0.351	0.134	NA
COL18A1	1.00(0.97-1.03)	0.98	0.407	0.090	NA
TIMP3	1.02(1.00-1.05)	0.08	0.475	0.056	Blood protein
BPIFA2	1.00(0.97-1.02)	0.76	0.660	0.035	NA
OLFML3	0.96(0.91-1.03)	0.27	0.089	0.012	NA
GCKR	1.01(0.99-1.03)	0.33	0.026	0.009	Blood cell, lipid metabolism, CRP, IFT172, eGFR, DM, HSA, Gout, NAFLD
AIF1	0.97(0.94-1.01)	0.12	1.000	<0.001	Blood cell, body measurement, IgAD, IMN, RA, PSC, AS, MS, PSO, SP
BCL2L15	1.02(0.99-1.05)	0.16	1.000	<0.001	RA
BTN3A2	0.91(0.82-1.01)	0.07	1.000	<0.001	Blood cell, IgAD, PSC, sarcoidosis, BPH, SP
CLPS	0.99(0.97-1.01)	0.21	1.000	<0.001	NA
TNXB	1.06(1.00-1.12)	0.04	1.000	<0.001	HT, RA, DM

AS=ankylosing spondylitis; BPH=benign prostatic hyperplasia; CRP=C reactive protein; CVD=cardiovascular disease; DM=diabetes mellitus; eGFR=estimated glomerular filtration rate; HSA=human serum albumin; PSO=psoriasis; HT=hypothyroidism; IgAD=IgA deficiency; IMN=idiopathic membranous nephropathy; MS=multiple sclerosis; NAFLD=nonalcoholic fatty liver disease; PSC=primary sclerosing cholangitis; RA=rheumatoid arthritis; SP=schizophrenia; NA=not available.

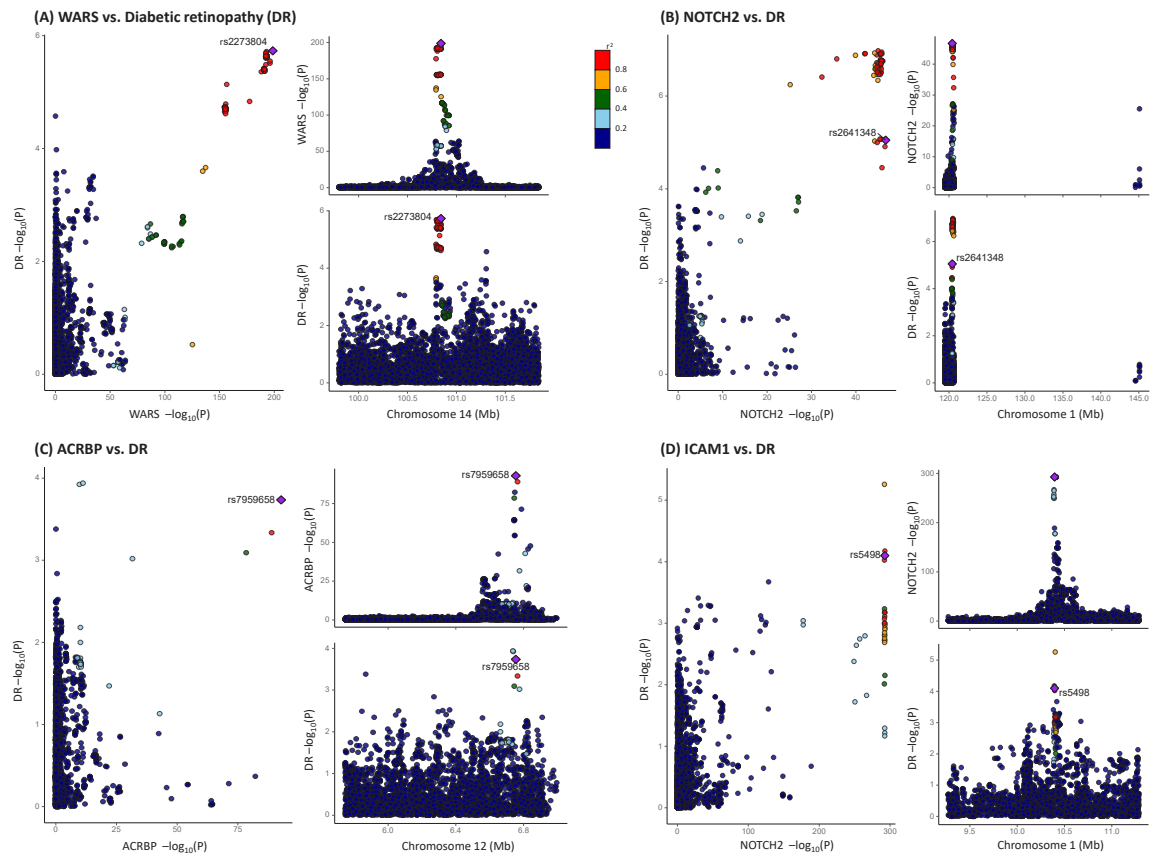

Figure 3 Co-localization analysis of the association between plasma protein level and diabetic retinopathy (DR)

(A) Co-localization analysis of the association between WARS level and DR: the *cis*-pQTL *rs2273804* is shown as a purple diamond. The P values of the SNPs in the WARS locus were extracted from the GWAS by UKB-PPP and from the diabetic retinopathy GWAS by the FinnGen study. (B) Co-localization analysis of the association between NOTCH2 level and DR: the *cis*-pQTL *rs2641348* is shown as a purple diamond. The P values of the SNPs in the NOTCH2 locus were extracted from the GWAS by UKB-PPP and from the GWAS by FinnGen study. (C) Co-localization analysis of the association between ACRBP level and DR: the *cis*-pQTL *rs7959658* is shown as a purple diamond. The P values of the SNPs in the ACRBP locus were extracted from the GWAS by UKB-PPP and from the GWAS by FinnGen study. (D) Co-localization analysis of the association between ICAM1 level and DR: the *cis*-pQTL *rs5498* is shown as a purple diamond. The P values of the SNPs in the ICAM1 locus were extracted from the GWAS by deCODE HEALTH and from the GWAS by FinnGen study. The dots in the scatter plots are colored according to their LD to the identified casual *cis*-pQTL. LD is calculated based on the 1000 Genomes phase 3 European references. Mb=Megabase pairs.

status of plasma protein targets associated with DR, an extensive investigation utilizing druggable gene lists, ChEMBL (release 33) drug discovery database, and DrugBank (release 5.1.10) was conducted. These targets were categorized into three groups based on their progress in drug development: approved (targets of drugs already approved for one or more conditions), in development (currently in clinical trials), and druggable (potential targets; Table 4). The analysis of drug development activities identified six plasma proteins (COL18A1, ICAM1, ITGB7, NOTCH2, TIGHT, and TIMP3) as the focus of drug development endeavors.

COL18A1 was targeted by drugs approved for various conditions, including penile induration, dupuytren contracture, ulcer, skin ulcer (e.g., collagenase clostridium histolyticum), and eye diseases (e.g., ocriplasmin). Similarly, ICAM1 was a subject of drug development for dry eye disease (e.g., lifitegrast) and other conditions like joint pain, wound healing, ophthalmologicand cosmetic treatments (e.g., hyaluronic

acid). ITGB7 attracted attention from drugs such as vedolizumab and natalizumab, addressing immune system diseases, ulcerative colitis, Crohn's disease, inflammation, and multiple sclerosis. Among the six DR-related plasma proteins, three have associations with six approved drugs, and all six of them are linked to one or more drugs undergoing Phase I-III clinical trials. Notably, pimagedine, targeting TIMP3, was in Phase I trials for diabetic complications, including DR. However, there was no evidence of drug development activities for the remaining 18 plasma proteins. Based on the list of druggable genes, ACRBP, BPIFA2, BTN3A2, CLPS, GSTA1, OLFML3, PAM, SIRPG, and TNXB were identified as druggable genes, offering potential targets for future drug development initiatives.

GO enrichment analysis revealed functional pathways enriched by genes for novel drug target proteins. The Supplementary Figure shows that 18 out of 24 plasma proteins are involved in 22 gene functional pathways. These pathways include epithelial cell differentiation, positive regulation of cytokine production, supramolecular fiber organization, negative regulation of cell population proliferation, and blood vessel development, among others.

DISCUSSION

Summary of the Findings

This first PW-MR study successfully identified 24 DR causal plasma proteins with potential for drug targeting, comprising 18 novel DR susceptibility proteins and six previously known (COL18A1, ICAM1, ITGB7, NOTCH2, TIGHT, and TIMP3) proteins. Co-localization analysis revealed that genetically higher levels of WARS, ACRBP, and ICAM1 were inversely related to DR risk, while elevated levels of NOTCH2 increased DR risk. Three of these 24 proteins are being developed as therapeutic agents for dry eye disease and an immune disorder, with two additional proteins undergoing phase I clinical trials as targeted treatments for DR. This study reinforced the strategies for DR drug development, and underscored the potential of PW-MR for unraveling the causal biology of complex diseases.

Table 4 Summary of druggability and clinical development activity for diabetic retinopathy associated with causal associations on MR analysis

Target	Compound name	Clinical development activities
		Approved
COL18A1	COLLAGENASE CLOSTRIDIUM HISTOLYTICUM† (CHEMBL2108709)	Approved for penile induration, dupuytren contracture, ulcer, skin ulcer; Phase III: pressure ulcer; Phase II: urethral stricture, lipoma, bursitis, contracture; Phase I: tendinopathy, diabetic foot
ICAM1	LIFITEGRAST† (CHEMBL2048028)	Approved for dry eye disease; Phase II: allergic conjunctivitis
ICAM1	Hyaluronic acid‡	Approved for joint pain, wound healing, ophthalmologic treatment, cosmetic treatment
ITGB7	VEDOLIZUMAB† (CHEMBL1743087)	Approved for immune system disease, ulcerative colitis, Crohn's disease; Phase II: HIV infection/ HIV-1 infection, infection, colitis, celiac disease; Phase I: inflammatory bowel disease, melanoma, type 1 diabetes mellitus
ITGB7	NATALIZUMAB† (CHEMBL1201607)	Approved for immune system disease, Crohn's disease, inflammation, multiple sclerosis; Phase III: multiple Sclerosis; Phase II: arthritis, graft vs host disease, ischemic stroke; Phase I: myositis, schizophrenia, multiple myeloma

Table 4 (continued)

Target	Compound name	Clinical development activities
		In development
COL18A1	OCRIPLASMIN† (CHEMBL2095222)	Approved: eye diseases; Phase III: retinal perforations; Phase II: stroke, macular degeneration, venous thrombosis; Phase I: uveitis, retinal vein occlusion
ICAM1	ENLIMOMAB† (CHEMBL2109090)	Phase II: multiple myeloma
ITGB7	ETROLIZUMAB† (CHEMBL1743015)	Phase III: ulcerative colitis, Crohn's disease
ITGB7	ABRILUMAB† (CHEMBL3137351)	Phase II: ulcerative colitis, Crohn's disease;
ITGB7	FIRATEGRAST† (CHEMBL2104967)	Phase II: multiple sclerosis; Phase I: relapsing-remitting multiple sclerosis
NOTCH2	TAREXTUMAB† (CHEMBL3301588)	Phase I: pancreatic carcinoma, neoplasm, small cell lung carcinoma
TIGIT	TIRAGOLUMAB† (CHEMBL4298050)	Phase III: small cell lung carcinoma, carcinoma, esophageal neoplasms; Phase II: carcinoma, neoplasms, uterine cervical neoplasms, rectal neoplasms, melanoma; Phase I: breast neoplasms, liver neoplasms, urinary bladder neoplasms, lymphoma, stomach neoplasms, endometrial neoplasms
TIGIT	ETIGILIMAB† (CHEMBL4298205)	Phase I: neoplasm
TIGIT	ASP-8374† (CHEMBL4297894)	Phase I: neoplasm, glioblastoma multiforme
TIGIT	VIBOSTOLIMAB† (CHEMBL4297785)	Phase III: carcinoma, small cell lung carcinoma; Phase II: hematologic neoplasms, colorectal neoplasms; Phase I: prostatic neoplasms, neoplasms, melanoma, carcinoma
TIGIT	OCIPERLIMAB† (CHEMBL4650411)	Phase III: non-small cell lung carcinoma; Phase II: cervical cancer, esophageal squamous cell carcinoma, hepatocellular carcinoma, small cell lung carcinoma; Phase I: neoplasm, diffuse large B-cell lymphoma
TIGIT	DOMVANALIMAB† (CHEMBL4650392)	Phase III: non-small cell lung carcinoma; Phase II: neoplasm|cancer, melanoma;
TIGIT	IBI-939† (CHEMBL4650473)	Phase I: lung cancer, cancer
TIMP3	PIMAGEDINE† (CHEMBL225304)	Phase I: diabetic retinopathy, diabetic kidney, diabetic nephropathy‡
		Druggable *
		ACRBP, BPIFA2, BTN3A2, CLPS, GSTA1, ICAM1, OLFML3, PAM, SIRPG, TNXB, TPPP3
		Not currently listed as druggable
		WARS, AIF1, BCL2L15, LACTB2, MANSC4, GALNT3, GBP1, GCKR

* Data from druggable gene list.
† Data from ChEMBL release 33 (compound ID in brackets).
‡ Data from Drugbank online.

MR-derived Novel Biomarkers and Potential Targets for DR

This study identified 18 novel drug target proteins for DR, notably focusing on OLFML3, ACRBP, WARS, and TPPP3. Among them, ACRBP and WARS demonstrate significant potential in their involvement with the pathogenesis of DR, supported by strong evidence of colocalization. Acrosin-binding protein (ACRBP) is recognized as a glycoprotein that plays a pivotal role in regulating protease activity during sperm penetration through the zona pellucida.[31] Recent studies have shown that ACRBP is expressed abnormally in various types of tumors,[32] which suggests that ACRBP may counteract pathological interferences with microtubule dynamics and centrosome functions, ensuring normal mitosis during angiogenesis.[33] This study suggests that ACRBP may be a causative protein in the development of DR at the genetic level, suggesting the need to further explore specific drug-targeting mechanisms. Additionally, tryptophanyl-tRNA synthetase (WARS) is an essential housekeeping enzyme that attaches tryptophan to its corresponding tRNA for protein synthesis.[34] Studies have shown a positive correlation between WARS and the expression of T-cell markers,[35] and the anterior segment of the retina in patients with DR has been found to contain numerous immune-inflammatory cells, including T-cells.[36] Based on previous findings, this study further supports the notion that WARS is a plasma protein that increases the risk of DR. However, these preliminary insights necessitate further validation through comprehensive experimental and clinical studies.

Compared to Previous Experimental and Observational Studies

Among the proteins identified as causal through MR, 6 have been indicated by previous studies. Intercellular adhesion molecule 1 (ICAM1), critical for leukocyte adhesion and migration, consistently exhibits higher levels in DR patients than in controls, a finding substantiated by a meta-analysis of 11 studies, indicating a potential correlation with DR severity.[37] Vascular endothelial growth factor, a key element in retinopathy pathogenesis, promotes the release of serum sICAM1 from ICAM1's outer segment, intensifying inflammation and DR progression.[37-38] This study underscored the causal relationship between increased ICAM1 levels and DR risk, with drugs targeting ICAM1, such as LIFITEGRAST, having gained FDA approval for clinical treatment of dry eye disease. NOTCH signaling, is crucial for retinal vascular development and eye diseases, but previous studies show conflicting outcomes, particularly regarding NOTCH2.[39-40] This study revealed that genetic NOTCH2 levels confer protection against DR, underscoring the need for further studies to elucidate the precise nature of this association.

Aligns with MR Assumptions and Co-localization Evidence

To enhance the robustness and reliability of the findings, a comprehensive method was implemented, ensuring adherence to the MR assumptions. Firstly, we focused on selecting *cis*-pQTLs due to their inherently higher biological likelihood of influencing protein expression in nearby genes, thus significantly affecting protein function and activity, compared to trans-pQTLs. The selection criteria for instrumental variables for plasma protein levels included LD clustering with an r^2 below 0.001, *cis*-pQTLs with a P-value of 5.0×10^{-8} or less, and a location within 1 Mb of the relevant protein-coding gene. We also assessed the strength of these genetic IVs, requiring F-statistics greater than 10, to reduce bias from weaker instrumental variables. Secondly, utilizing an integrated analysis that combined

MR and co-localization, we identified novel drug targets for DR. Thirdly, we conducted a bidirectional MR analysis to address potential reverse causality and genetic confounding linked to LD. Furthermore, Bayesian co-localization analysis effectively ruled out confounding by LD. This led to the identification of four proteins (WARS, NOTCH2, ACRBP, ICAM1) as likely sharing genetic variants with DR. However, non-co-localization does not undermine these findings, given the high false-negative rate (approximately 60%) associated with Bayesian co-localization methods. Phenotypic scanning also linked two proteins (ACRBP, LACTB2) with other traits, such as blood cells and body measurements, but these links do not fully clarify their relationship with DR. Therefore, WARS, NOTCH2, ACRBP, and ICAM1 are promising drug targets for DR.

Scientific and Clinical Implications

This study had crucial insights for scientific research and clinical practice. Firstly, it underscores the potential of expanding omic biomarker testing in discovering novel drug targets. By aggregating genetic data from extensive plasma proteomics studies, we analyzed 2 602 circulating proteins related to DR, representing a fraction (10%) of all known proteins. Secondly, our PW-MR findings suggest that proteins causally associated with DR may also modulate the risk of various non-DR diseases. Several proteins identified are either approved for other conditions or in clinical trials, suggesting a potential overlap in causal proteins across different diseases. Thirdly, PW-MR stands out as an effective preclinical approach for prioritizing drug targets and anticipating target-related side effects. The ongoing exploration of these biomarkers is crucial, promising to unveil immediate mediators and enhance the effectiveness and safety of emerging drug targets.

Strength and Limitations

The strength of this study lies in using the most recent, comprehensive plasma protein dataset and the largest GWAS dataset for DR, thereby improving our ability to detect associations. By focusing on individuals of European ancestry, we minimized population stratification bias. Our findings are validated through various sensitivity analyses, including bidirectional MR and co-localization analyses, which are crucial for identifying shared genetic variants between exposure outcomes. This study also has several limitations. Firstly, genetic variants and qualitative protein changes, such as amino acid substitutions and post-translational modifications, may compromise the specificity of protein quantification. These changes can affect binding affinity in assays, potentially leading to misinterpretations.[41] To address this, we adopted multiple protein quantification methods including Olink, Somascan, and various genetic IVs, thereby minimizing the impact of technology-specific biases on our results. Secondly, the exclusive focus on *cis*-acting SNPs limited our choice of analytical methods. However, our meticulous selection of SNPs from multiple studies allowed for the use of robust IVs. The study's scope was confined to proteins associated with identifiable *cis*-pQTL signaling, possibly overlooking other relevant therapeutic targets. Despite this, we conducted a comprehensive co-localization of circulating proteins to identify potential causal genes for DR. Thirdly, our standardized units of protein expression do not directly convey the absolute impact on DR, nor facilitate direct comparisons across proteins.[42] However, the inferred causal directions provide valuable insights for predicting efficacy and side effects, which can be substantiated in future clinical trials. Finally, the inclusion of only the European population limits the applicability of our findings to other ethnic groups. This highlights the need for additional studies to validate these results in more diverse populations.

CONCLUSION

In summary, we leveraged two-sample MR analysis using extensive proteomic data to identify 24 potential drug targets for diabetic retinopathy (DR). Four plasma proteins (WARS, ACRBP, ICAM1, and NOTCH2) emerged as particularly promising candidates, supported by robust co-localization evidence. This study offers unprecedented insights into DR's etiology and potential therapeutic strategies. It also highlights how burgeoning genomic and proteomic datasets can be pivotal in identifying novel drug targets. Further experimental and clinical studies are warranted to evaluate the utility and efficacy of these candidate proteins.

Correction notice

None

Acknowledgement

None

Author Contributions

(I) Conception and design: W.W., W.H

(II) Administrative support: S.Y., Z.Z., W.H

(III) Provision of study materials or patients: All authors

(IV) Collection and assembly of data: W.W., S.Y., Z.Z

(V) Data analysis and interpretation: S.Y., Z.Z., X.Z

(VI) Manuscript writing: All authors

(VII) Final approval of manuscript: All authors

Fundings

This study was funded by the Hainan Province Clinical Medical Center(82171084), the National Natural Science Foundation of China (82371086).

Conflict of Interests

None of the authors has any conflicts of interest to disclose. All authors have declared in the completed the ICMJE uniform disclosure form.

Patient consent for publication

None

Ethical Statement

The study is based on publicly available data from large-scale genome-wide association studies. Included studies had been approved by corresponding ethical review committees.

Provenance and Peer Review

This article was a standard submission to our journal. The article has undergone peer review with our anonymous review system.

Data Sharing Statement

All data analyzed in this study can be obtained by a reasonable request to the corresponding authors.

References

1. Teo ZL, Tham YC, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. Ophthalmology. 2021, 128(11): 1580-91. DOI: 10.1016/j.ophtha.2021.04.027.
2. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. Lancet. 2010, 376(9735): 124-136. DOI: 10.1016/s0140-6736(09)62124-3.
3. Boyer DS, Yoon YH, Belfort RJ, et al. Three-year, randomized, sham-controlled trial of dexamethasone intravitreal implant in patients with diabetic macular

edema. Ophthalmology, 2014, 121(10): 1904-1914. DOI: 10.1016/j.ophtha.2014.04.024.

4. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002, 1(11): 845-867. DOI: 10.1074/mcp.r200007-mcp200.
5. Chouliaras L, Thomas A, Malpetti M, et al. Differential levels of plasma biomarkers of neurodegeneration in Lewy body dementia, Alzheimer's disease, frontotemporal dementia and progressive supranuclear palsy. J Neurol Neurosurg Psychiatry. 2022, 93(6): 651-658. DOI: 10.1136/jnnp-2021-327788.
6. Chan MY, Efthymios M, Tan SH, et al. Prioritizing candidates of post-myocardial infarction heart failure using plasma proteomics and single-cell transcriptomics. Circulation. 2020, 142(15): 1408-1421. DOI: 10.1161/CIRCULATIONAHA.119.045158.
7. Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets. Nat Rev Drug Discov. 2017, 16(1): 19-34. DOI: 10.1038/nrd.2016.230.
8. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Mapping the proteo-genomic convergence of human diseases. Science. 2021, 374(6569): eabj1541. DOI: 10.1126/science.abj1541.
9. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021, 53(12): 1712-1721. DOI: 10.1038/s41588-021-00978-w.
10. Ursu O, Glick M, Oprea T. Novel drug targets in 2018. Nat Rev Drug Discov. 2019. DOI: 10.1038/d41573-019-00052-5.
11. Yao C, Chen G, Song C, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat Commun. 2018, 9(1): 3268. DOI: 10.1038/s41467-018-05512-x.
12. Emilsson V, Ilkov M, Lamb JR, et al. Co-regulatory networks of human serum proteins link genetics to disease. Science. 2018, 361(6404): 769-773. DOI: 10.1126/science.aaq1327.
13. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature. 2018, 558(7708): 73-79. DOI: 10.1038/s41586-018-0175-2.
14. Hulur I, Gamazon ER, Skol AD, et al. Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. BMC Genomics. 2015, 16(1): 138. DOI: 10.1186/s12864-015-1292-z.
15. Lu T, Forgetta V, Greenwood C, et al. Circulating proteins influencing psychiatric disease: a mendelian randomization study. Biol Psychiatry. 2023, 93(1): 82-91. DOI: 10.1016/j.biopsych.2022.08.015.
16. Mälarstig A, Grassmann F, Dahl L, et al. Evaluation of circulating plasma proteins in breast cancer using Mendelian randomisation. Nat Commun. 2023, 14(1): 7680. DOI: 10.1038/s41467-023-43485-8.
17. Chen L, Peters JE, Prins B, et al. Systematic Mendelian randomization using the human plasma proteome to discover potential therapeutic targets for stroke. Nat Commun. 2022, 13(1): 6143. DOI: 10.1038/s41467-022-33675-1.
18. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015, 47(8): 856-860. DOI: 10.1038/ng.3314.
19. Sun BB, Chiou J, Traylor M, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023, 622(7982): 329-338. DOI: 10.1038/s41586-023-06592-6.
20. Zheng J, Haberland V, Baird D, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet. 2020, 52(10): 1122-1131. DOI: 10.1038/s41588-020-0682-6.
21. Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017, 8: 15345. DOI: 10.1038/ncomms15345.
22. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023, 613(7944): 508-518. DOI: 10.1038/s41586-022-05473-8.
23. Burgess S, Thompson SG. Avoiding bias from weak

instruments in Mendelian randomization studies. Int J Epidemiol. 2011, 40(3): 755-764. DOI: 10.1093/ije/dyr036.

24. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. Am J Epidemiol. 2013, 178(7): 1177-1184. DOI: 10.1093/aje/kwt084.
25. Foley CN, Staley JR, Breen PG, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. Nat Commun. 2021, 12(1): 764. DOI: 10.1038/s41467-020-20885-8.
26. Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics. 2019, 35(22): 4851-4853. DOI: 10.1093/bioinformatics/btz469.
27. Finan C, Gaulton A, Kruger FA, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med. 2017, 9(383): eaag1166. DOI: 10.1126/scitranslmed.aag1166.
28. Mendez D, Gaulton A, Bento AP, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 2019, 47(D1): D930-940. DOI: 10.1093/nar/gky1075.
29. Knox C, Wilson M, Klinger CM, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. Nucleic Acids Res. 2024, 52(D1): D1265-D1275. DOI: 10.1093/nar/gkad976.
30. Mi H, Muruganujan A, Huang X, et al. Protocol update for large-scale genome and gene function analysis with the PANTHER classificationsystem(v.14.0). Nat Protoc. 2019, 14(3): 703-721. DOI: 10.1038/s41596-019-0128-8.
31. Mohaqiq M, Movahedin M, Mazaheri Z, et al. In vitro transplantation of spermatogonial stem cells isolated from human frozen-thawed testis tissue can induce spermatogenesis under 3-dimensional tissue culture conditions. Biol Res. 2019, 52(1): 16. DOI: 10.1186/s40659-019-0223-x.
32. Tammela J, Uenaka A, Ono T, et al. OY-TES-1 expression and serum immunoreactivity in epithelial ovarian cancer. Int J Oncol. 2006, 29(4): 903-910.
33. Whitehurst AW, Xie Y, Purinton SC, et al. Tumor antigen acrosin binding protein normalizes mitotic spindle function to promote cancer cell proliferation. Cancer Res. 2010, 70(19): 7652-7661. DOI: 10.1158/0008-5472.CAN-10-0840.
34. Jin M. Unique roles of tryptophanyl-tRNA synthetase in immune control and its therapeutic implications. Exp Mol Med. 2019, 51(1): 1-10. DOI: 10.1038/s12276-018-0196-9.
35. Ahn YH, Oh SC, Zhou S, et al. Tryptophanyl-tRNA synthetase as a potential therapeutic target. Int J Mol Sci. 2021, 22(9): 4523. DOI: 10.3390/ijms22094523.
36. Kastelan S, Zjacic-Rotkvic V, Kastelan Z. Could diabetic retinopathy be an autoimmune disease?. Med Hypotheses, 2007, 68(5): 1016-1018. DOI: 10.1016/j.mehy.2006.05.073.
37. Yao Y, Du J, Li R, et al. Association between ICAM-1 level and diabetic retinopathy: a review and meta-analysis. Postgrad Med J. 2019, 95(1121): 162-168. DOI: 10.1136/postgradmedj-2018-136102.
38. Funatsu H, Yamashita H, Sakata K, et al. Vitreous levels of vascular endothelial growth factor and intercellular adhesion molecule 1 are related to diabetic macular edema. Ophthalmology. 2005, 112(5): 806-816. DOI: 10.1016/j.ophtha.2004.11.045.
39. Dou GR, Wang L, Wang YS, et al. Notch signaling in ocular vasculature development and diseases. Mol Med. 2012, 18(1): 47-55. DOI: 10.2119/molmed.2011.00256.
40. Zhang J, Zeng Y, Chen J, et al. miR-29a/b cluster suppresses high glucose-induced endothelial-mesenchymal transition in human retinal microvascular endothelial cells by targeting Notch2. Exp Ther Med. 2019, 17(4): 3108-3116. DOI: 10.3892/etm.2019.7323.
41. Joshi A, Mayr M. In aptamers they trust: the caveats of the somascan biomarker discovery platform from somalogic. Circulation. 2018, 138(22): 2482-2485. DOI: 10.1161/CIRCULATIONAHA.118.036823.
42. Henry A, Gordillo-Maranon M, Finan C, et al. Therapeutic targets for heart failure identified using proteomics and mendelian randomization. Circulation. 2022, 145(16): 1205-1217. DOI: 10.1161/CIRCULATIONAHA.121.056663.