HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2025 March 01.

Published in final edited form as:

Nat Med. 2024 September ; 30(9): 2480–2488. doi:10.1038/s41591-024-03089-1.

Adaptive Selection at *G6PD* Exacerbates Disparities in Diabetes Complications

Joseph H. Breeyear[1,2,3], **Jacklyn N. Hellwege**[3,4,5], **Philip H. Schroeder**[6,7,8,9], **John S. House**[1], **Hannah M. Poisner**[5], **Sabrina L. Mitchell**[3,10], **Brian Charest**[11], **Anjali Khakharia**[12,13], **Til B. Basnet**[2,14], **Christopher W. Halladay**[15], **Peter D. Reaven**[16,17], **James B Meigs**[7,18,19], **Mary K Rhee**[12,20], **Yang Sun**[21], **Mary G. Lynch**[12], **Alexander G. Bick**[4,5], **Otis D. Wilson**[3], **Adriana M. Hung**[3], **Cari L. Nealon**[22,23], **Sudha K. Iyengar**[24,25], **Daniel M. Rotroff**[26,27,28], **John B. Buse**[29], **Aaron Leong**[6,7,8,9,18,19,30], **Josep M. Mercader**[6,7,8,9,18,30], **Lucia Sobrin**[31], **Milam A. Brantley Jr.**[3,10], **Neal S. Peachey**[23,32,33], **Alison A. Motsinger-Reif**[1], **Peter W. Wilson**[12,34], **Yan V. Sun**[12,35,36], **Ayush Giri**[3,5,14], **Lawrence S. Phillips**[12,20], **Todd L. Edwards**[2,3], **VA Million Veteran Program**

[1]Biostatistics and Computational Biology Branch, Division of Intramural Research, National Institute of Environmental Health Sciences, Durham, NC

[2]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

[3]VA Tennessee Valley Healthcare System (626), Nashville, TN

[4]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

[5]Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN

[6]Program in Metabolism, Broad Institute of MIT and Harvard, Cambridge, MA

[7]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA

[8]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA

[9]Diabetes Unit, Endocrine Division, Department of Medicine, Massachusetts General Hospital, Boston, MA

Addresses for Co-Correspondence: Todd L. Edwards, 2525 West End Avenue, 6th Floor, Suite 600, Office 615, Nashville, TN, 37203, Ayush Giri, 2525 West End Avenue, 6th Floor, Suite 600, Office 631, Nashville, TN, 37203, todd.l.edwards@vumc.org, ayush.giri@vumc.org.

CONTRIBUTIONS

Conceptualization: J.H.B., P.W.W., Y.V.S., L.S.P., and T.L.E.; methodology: J.H.B., S.L.M., T.B.B., A.M.H., C.L.N., D.M.R., J.B.B., A.S.L., J.M.M., Y.V.S., A.G., L.S.P., and T.L.E; investigation: J.H.B., J.N.H., P.H.S., J.S.H., and H.M.P.; writing (original draft): J.H.B., L.S.P., and T.L.E.; writing (review and editing): J.H.B., J.N.H., J.S.H., S.L.M., T.B.B., P.D.R., J.B.M., M.K.R, Y.S., M.G.L., A.G.B., A.M.H., S.K.I, D.M.R, J.B.B., J.M.M., L.S., M.A.B., N.S.P., A.A.M.R., P.W.W., Y.V.S., A.G., L.S.P., and T.L.E.; visualization: J.H.B., J.S.H., and A.G.; resources: A.M.H., J.M.M., L.S., Y.V.S., A.G., L.S.P., and T.L.E.; data curation: B.C., A.K., C.W.H., O.D.W.; supervision: P.W.W., Y.V.S., A.G., L.S.P., and T.L.E.; funding acquisition: P.W.W., Y.V.S., A.G., L.S.P., and T.L.E.

INCLUSION AND ETHICS STATEMENT

The central Veterans Affairs Institutional Review Board (IRB) and site-specific IRBs approved the Million Veteran Program study. The Vanderbilt University Medical Center IRB approved the use of BioVU data for this study. The Massachusetts General Brigham Biobank IRB approved the use of data for this study. All relevant ethical regulations were followed.

[10]Department of Ophthalmology and Visual Sciences, Vanderbilt University Medical Center, Nashville, TN

[11]Massachusetts Veterans Epidemiology Research and Information Center, Boston, MA

[12]Atlanta VA Medical Center, Decatur, GA

[13]Department of Medicine and Geriatrics, Emory University School of Medicine, Atlanta, GA

[14]Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN

[15]Providence VA Medical Center, Providence, RI

[16]Phoenix VA Health Care System, Phoenix, AZ

[17]College of Medicine, University of Arizona, Phoenix, AZ

[18]Department of Medicine, Harvard Medical School, Boston, MA

[19]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA

[20]Division of Endocrinology, Metabolism, and Lipids, Department of Medicine, Emory University School of Medicine, Atlanta, GA

[21]Department of Ophthalmology, Stanford University School of Medicine, Palo Alto, CA

[22]Eye Clinic, VA Northeast Ohio Healthcare System, Cleveland, OH

[23]Department of Ophthalmology & Visual Sciences, Case Western Reserve University School of Medicine, Cleveland, OH

[24]Research Service, VA Northeast Ohio Healthcare System, Cleveland, OH

[25]Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH

[26]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH

[27]Endocrinology and Metabolism Institute, Cleveland Clinic, Cleveland, OH

[28]Center for Quantitative Metabolic Research, Cleveland Clinic, Cleveland, OH

[29]Division of Endocrinology & Metabolism, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina

[30]Department of Medicine, Massachusetts General Hospital, Boston, MA

[31]Department of Ophthalmology, Mass Eye and Ear Infirmary, Harvard Medical School, Boston, MA

[32]Cole Eye Institute, Cleveland Clinic, Cleveland, OH

[33]Department of Ophthalmology, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH

[34]Division of Cardiology, Department of Medicine, Emory University School of Medicine, Atlanta, GA

[35]Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA

[36]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA

Abstract

Diabetes complications occur at higher rates in individuals of African ancestry. Glucose 6-phosphate dehydrogenase deficiency (G6PDdef) is common in some African populations, confers malaria resistance, and reduces HbA1c levels by shortening erythrocyte lifespan. In a combined-ancestry GWAS of diabetic retinopathy, we identified nine novel loci including a G6PDdef causal variant, rs1050828-T (Val98Met), which was also associated with increased risk of other diabetes complications. The effect of rs1050828-T on retinopathy was fully mediated by glucose levels. In the years preceding diabetes diagnosis and insulin prescription, glucose levels were significantly higher and HbA1c significantly lower in those with vs. without G6PDdef. In the ACCORD trial, participants with G6PDdef had significantly higher hazards of incident retinopathy and neuropathy. At the same HbA1c levels, G6PDdef participants in both ACCORD and the MVP had significantly increased risk of retinopathy. We estimate that 12% and 23% of diabetic retinopathy and neuropathy cases, respectively, in African ancestry participants reflect this exposure. In patients with G6PDdef, management guided only by HbA1c levels could lead to delayed diagnosis and intensification of therapy, increasing the risk of diabetes complications. Across continentally defined ancestral populations, the differences in frequency of rs1050828-T and other G6PDdef alleles contribute to disparities in diabetes complications.

Keywords

Diabetic Retinopathy; Diabetic Complications; Racial Disparities

INTRODUCTION

Diabetic retinopathy, a complication of diabetes mellitus, is a leading cause of visual impairment and preventable blindness among working-age adults.[1] Diabetic retinopathy was previously thought to be solely due to damage to the retinal microvasculature but there is increasing evidence that there is parallel damage to the neural retina.[2] The pathogenesis of the microvasculopathy has been thought to be due in part to the high metabolic rate, induced by derangement of glucose metabolism, causing dysregulation in the network of intraretinal vasculature that maintains retinal homeostasis.[3] Progression of diabetic retinopathy can be characterized by the presence of vascular lesions and dependent on the degree of neovascularization.

Diabetic retinopathy is thought to be due mainly to cumulative glycemic exposure, as shown in analyses of the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT-EDIC) cohort, but the severity varies substantially among individuals, and is not fully explained by known risk factors.[4] There are racial disparities in both the prevalence of diabetes and its complications,[1,5–8] and a recent review of datasets in the United States estimated the age- and sex-standardized prevalence rates

of diabetic retinopathy to be 34.4% and 24.4% for non-Hispanic Black and non-Hispanic White individuals, respectively.[8]

At present, only a small number of genetic loci have been associated with diabetic retinopathy, even though twin-based heritability estimates are between 18% and 52% for diabetic retinopathy and proliferative diabetic retinopathy (PDR).[9–17] These loci have been identified primarily in studies of individuals with European (EUR) and Asian (ASN) ancestry, with some studies utilizing controls with and without diabetes. To address this gap in identified genetic loci compared to estimates of heritability, we conducted a combined-ancestry genome-wide association study (GWAS) of diabetic retinopathy and PDR. We used an electronic health records (EHR) algorithm to identify individuals with diabetes who did or did not have diabetic retinopathy (cases and controls, respectively). We combined evidence from the Million Veteran Program (MVP), the United Kingdom Biobank, Vanderbilt University Medical Center's BioVU, the Mass General Brigham Biobank (MGBB), and summary statistics from Pollack et al., for a total of 68,169 cases with diabetic retinopathy and 129,188 controls with diabetes.[15]

We conducted meta-analyses to identify single nucleotide polymorphisms (SNPs) associated with diabetic retinopathy. The most significant variant, rs1050828-T (Val98Met), only exists at common frequencies in African (AFR) ancestry populations. This allele, previously been shown to be under positive selection, is causal for glucose 6-phosphate dehydrogenase deficiency (G6PDdef), an enzymopathy protective against severe malaria.[18–20] This allele has also been previously associated with decreases in HbA1c levels, which is believed to be due to shorter erythrocyte lifespan in G6PDdef individuals that results in less hemoglobin glycosylation.[21–23] Therefore, this raises concern regarding the usage of HbA1c as an appropriate indicator of glycemic control in diabetics with G6PDdef. We hypothesized that G6PDdef individuals with diabetes may experience more severe hyperglycemia due to delayed diagnosis and inadequate treatment intensification due to clinical dependency on HbA1c measures. In this paper, we present evidence that the G6PDdef phenotype may increase risk for diabetes complications and exacerbate disparities of diabetes complications.

RESULTS

Genetic Architecture of Diabetic Retinopathy

We combined evidence of SNP – diabetic retinopathy associations through inverse-variance weighted fixed-effects meta-analyses, both within and across genetic ancestry groups (Table 1). We identified SNPs corresponding to nine unique genome-wide significant loci across the combined- (N_{loci} = 9), Non-Hispanic African (NH-AFR-) (N_{loci} = 1), and Non-Hispanic European (NH-EUR) -ancestry (N_{loci} = 1) meta-analyses (Figure 1, Table 2, Supplemental Tables 1 – 3). These loci corresponded to nine genes based on proximity (<1Mb): *MFSD4A*, *TENM2*, *GRHL2*, *TCF7L2*, *FBXW8, COL4A1*, *SLC16A3*, *TMPRSS6*, and *G6PD.* The most significant SNP in the combined- and NH-AFR-ancestry meta-analyses was rs1050828-T (Val98Met) (odds ratio [OR] = 1.48, 95% confidence interval [CI] = (1.45 – 1.51), $p = 1.99 \times 10^{90}$) in *G6PD*. The most significant SNP in the NH-EUR ancestry analysis was rs7903146-T (OR = 1.08 (1.07 – 1.09), $p = 3.14 \times 10^{-13}$), intronic in *TCF7L2*, corresponding to the same lead variant of the strongest type 2 diabetes risk association.[24]

Additionally, rs1050828-T was significant in the combined- and NH-AFR-ancestry meta-analyses of PDR cases against diabetic retinopathy controls (OR = 1.36 (1.31 – 1.42), $p = 3.28 \times 10^{-15}$) (Supplemental Table 4 – 6). We did not identify any conditionally independent SNPs at this or any locus.

We examined the genomic inflation factor (λ_{GC}), linkage disequilibrium score regression intercept, and heritability on the liability scale (h^2) of the meta-analyses as well as the genetic correlation of diabetic retinopathy between NH-AFR- and NH-EUR-ancestries in the MVP with linkage disequilibrium score regression (LDSC) (Supplemental Table 7). There was little evidence for confounding due to population stratification in ancestry specific analyses as suggested by λ_{GC} and LDSC intercept values for NH-AFR (λ_{GC}: 1.0225; LDSC intercept: 0.9817 (0.0059)) and NH-EUR ancestry (λ_{GC}: 1.0405; LDSC intercept: 0.9841 (0.0077)) analyses. Heritability (h^2) estimates were larger for NH-AFR ancestry (8.0% (5.8% – 10.2%)) than for NH-EUR ancestry (3.6% (3.0% – 4.2%)), though were similar to previously published SNP-based heritability estimates (7.0%).[13] Additionally, there was a high correlation between the NH-EUR- and NH-AFR-ancestry analyses (0.76 ± 0.11, $p = 1.72 \times 10^{-12}$).

We estimated the association between genetically predicted gene expression (GPGE) and diabetic retinopathy in 49 tissues from GTEx v8 with S-PrediXcan.[25] We identified 12 statistically significant gene-tissue pairs associated with diabetic retinopathy in the combined-ancestry analyses and 13 pairs in the NH-AFR ancestry analyses (Supplemental Tables 8 – 9). These gene-tissue combinations corresponded to nine unique genes (including *G6PD*), none of which were previously associated with diabetic retinopathy as reported in the NHGRI-EBI catalog of human GWAS at the time of publication. We identified seven statistically significant gene-tissue pairs associated with PDR in the combined-ancestry analyses and 14 pairs in the NH-AFR ancestry analyses (Supplemental Tables 10 – 11). Four genes were statistically significant in both diabetic retinopathy and PDR: *FLNA*, *GAB3*, *G6PD*, and *RENBP*, though we did not detect any gene-tissue pairs where there were statistically significant GPGE results and high posterior probability (PP.H4>0.9) of colocalization. We did not identify any statistically significant GPGE associations in the NH-EUR ancestry analyses.

Because rs1050828 is located on the X chromosome we investigated the effect of the variant on diabetic retinopathy risk separately for men and women of NH-AFR ancestry. We observed similar effect sizes for the variant on diabetic retinopathy risk for men (OR = 1.46 (1.40 – 1.52), $p = 7.52 \times 10^{-67}$) and women (OR = 1.54 (1.27 – 1.88), $p = 1.38 \times 10^{-05}$) (Supplemental Table 12) although the variant did not reach genome-wide significance for women due to reduced sample size. We did not detect additional associations on the X chromosome in a conditional analysis adjusting for rs1050828 genotypes.

G6PDdef Risk Allele and Diabetes Complications

We utilized EHR data from the MVP to compare mean hemoglobin A1c (HbA1c %) and outpatient random plasma glucose (mg/dL) values post diabetes diagnosis for men of NH-AFR ancestry with and without the G6PDdef risk allele (rs1050828-T) stratified by diabetic retinopathy status. Individuals with the G6PDdef risk allele with diabetic

retinopathy had similar levels of HbA1c (7.27% vs. 7.29%; $p = 0.46$), yet substantially higher levels of plasma glucose (185 mg/dL vs. 145 mg/dL; $p = 4.93 \times 10^{-186}$) compared to individuals without the G6PDdef risk allele and without diabetic retinopathy (Figure 2a – 2b, Supplemental Table 13).

We compared HbA1c and glucose levels at different times during the natural history of diabetes in men of NH-AFR ancestry in the MVP with and without the G6PDdef risk allele. In each participant, we determined the mean outpatient random plasma/serum glucose (mg/dL) and mean HbA1c (%) a year prior to diabetes diagnosis as well as a year prior to insulin prescription. During the period prior to diabetes diagnosis, individuals with the G6PDdef risk allele had a significantly lower HbA1c level than those without (6.70% vs. 7.06%; $p = 6.05 \times 10^{-12}$), but a significantly higher plasma glucose level (165 mg/dL vs. 135 mg/dL; $p = 1.12 \times 10^{-46}$) (Supplemental Table 14). HbA1c and glucose levels were higher, but patterns were similar during the year prior to the initial outpatient prescription of insulin (Supplemental Table 14).

We compared the difference between mean HbA1c and predicted HbA1c (generated by regressing HbA1c on outpatient random plasma glucose) for all individuals with diabetes in the MVP (Figure 2c). Of all the individuals of NH-AFR ancestry, 76% were concentrated in the top and middle tertile, suggesting an HbA1c either higher than or similar to the predicted HbA1c (Figure 2d). However, 80% of individuals of NH-AFR ancestry with the G6PDdef risk allele were in the bottom tertile, suggesting an HbA1c lower than the predicted HbA1c (Figure 2e).

We demonstrate that individuals with the G6PDdef risk allele have a higher predicted odds of diabetic retinopathy compared to individuals without the G6PDdef risk allele at lower ranges of HbA1c, with risk of diabetic retinopathy converging near an HbA1c of 10% (Figure 3, Supplemental Figure 1). To investigate whether the effect of the G6PDdef risk allele on diabetic retinopathy risk could be attributed to differences in plasma glucose levels, we performed a mediation analysis in individuals of NH-AFR ancestry in the MVP. For men, we estimated a 2.39% (total effect, $p < 2.0 \times 10^{-16}$) increase in diabetic retinopathy risk due to the G6PDdef risk allele. However, the indirect effect of plasma glucose completely mediated the effect of the G6Pddef risk allele on diabetic retinopathy (average indirect effect = 2.37%, $p < 2.0 \times 10^{-16}$, average direct effect = 0.03%, $p = 0.95$) (Supplemental Table 15). We identified similar results in women (Supplemental Table 16). As the effect of the G6Pddef risk allele was mediated by increased plasma glucose, the increase in risk of diabetic retinopathy due to the G6Pddef risk allele is most likely due to suboptimal control of glycemia.

We further investigate if the G6Pddef risk allele increases risk of diabetic macular edema, a severe subtype of diabetic retinopathy, as well as other diabetes complications identified with an EHR based algorithm in men of NH-AFR ancestry. We identified associations between the G6PDdef risk allele and diabetes complications including diabetic macular edema (OR = 1.55 (1.45 – 1.65) $p = 1.31 \times 10^{-41}$) and diabetic kidney disease (OR = 1.24 (1.18 – 1.31) $p = 2.59 \times 10^{-15}$). To evaluate whether the effect of the G6PDdef risk allele on diabetic retinopathy is driven by the association with diabetic macular edema, we

performed a separate analysis for diabetic retinopathy excluding individuals with diabetic macular edema. The G6PDdef risk allele was still significantly associated with diabetic retinopathy without the presence of diabetic macular edema (OR = 1.33 (1.27 – 1.39) $p = 5.7 \times 10^{-35}$).

We further investigated the effect of the G6PDdef risk allele on the risk of diabetes complications in longitudinal data from the Action to Control Cardiovascular Risk in Diabetes (ACCORD) clinical trial in individuals of NH-AFR ancestry (Supplemental Table 17). Two diabetes complications or a composite outcome of diabetic retinopathy or diabetic nephropathy were significantly associated with the G6PDdef risk allele in the ACCORD clinical trial data, regardless of medication regimen. Individuals of NH-AFR ancestry with the G6PDdef risk allele had a higher likelihood of proliferative diabetic retinopathy (HR = 1.78 (1.56 – 2.04) $p = 1.95 \times 10^{-05}$) (characterized as photocoagulation or vitrectomy), diabetic neuropathy (HR = 1.37 (1.22 – 1.53) $p = 5.45 \times 10^{-03}$) (characterized as vibratory sensation loss), and a composite outcome of diabetic retinopathy and diabetic nephropathy (HR = 1.61 (1.41 – 1.82) $p = 2.07 \times 10^{-04}$) (specified as renal failure, ESRD, serum creatinine>3.3, photocoagulation, or vitrectomy) (Supplemental Table 18 – 19). With the available incidence data, we estimated the excess risk of diabetic complications due to the G6PDef by computing a crude and an adjusted risk difference (RD) for diabetic retinopathy (adjusted RD = 133 (95% CI: 3 – 264) per 1,000 individuals) and diabetic neuropathy (adjusted RD = 245 (95% CI: 84 – 507) per 1,000 individuals) in men of NH-AFR ancestry (Supplemental Table 20). Assuming the RD is an estimate of the attributable risk and that all of the excess risk is due to inadequate management of hyperglycemia due to reliance on HbA1c, we estimated 53 (95% CI: 30 – 254) of individuals with NH-AFR ancestry with diabetes would need to be screened for the G6PDdef variant and adequately managed for hyperglycemia to prevent one additional case of diabetic retinopathy. Finally, we estimated screening 407 (95% CI: 228 –1,948) individuals of NH-AFR ancestry would prevent one case of diabetic retinopathy (Supplementary Tables 21 – 22).

PheWAS of G6PDdef Risk Allele

To further investigate the effects of the G6PDdef risk allele on risk of diabetes complications, we conducted a PheWAS to investigate phenotypes associated with the G6PDdef risk allele separately in individuals with and without diabetes of NH-AFR ancestry in MVP and BioVU. We identified 30 PheCodes in individuals with diabetes and 19 PheCodes (higher order definitions using ICD codes) significantly associated with the G6PDdef risk allele for individuals without diabetes (Figure 4, Supplemental Table 23 – 24). Of these PheCodes, ten were significant in both analyses with the same direction of effect between groups, including anemia (other hereditary hemolytic anemia), metabolic (intestinal disaccharidase deficiencies and disaccharide malabsorption), and digestive (jaundice [not of newborn], cholelithiasis and cholecystitis) PheCodes.

The 20 significant PheCodes in individuals with diabetes included those in the infectious disease (septicemia (1.12 (1.09 – 1.15), $p = 1.82 \times 10^{-5}$)), endocrine/metabolic (diabetic retinopathy (1.12 (1.10 – 1.14), $p = 1.36 \times 10^{-8}$)), circulatory system (hypertensive heart and/or renal disease (1.10 (1.08 – 1.12) $p = 2.75 \times 10^{-7}$)), and genitourinary (acute renal

failure (1.12 (1.10 – 1.14) $p = 2.14 \times 10^{-9}$)) PheCode groups. We did not detect any significant PheCodes in the infectious disease, circulatory system, or genitourinary PheCode groups associated with the G6PDdef risk allele in individuals without diabetes.

We identified PheCodes associated with the G6PDdef risk allele in the analysis of individuals without diabetes that were not found in the analysis of individuals with diabetes. Specifically, we identified 'obesity' (1.06 (1.05 – 1.07) $p = 4.21 \times 10^{-06}$), 'poisoning/allergy of sulfonamides' (1.48 (1.35 – 1.62) $p = 1.85 \times 10^{-5}$), 'other diseases of blood and blood-forming organs' (1.24 (1.19 – 1.28) $p = 3.31 \times 10^{10}$), and 'other deficiencies of circulating enzymes' (4.29 (3.68 – 4.54) $p = 1.22 \times 10^{-40}$).

DISCUSSION

With an EHR-based algorithm to identify diabetic retinopathy cases and controls with diabetes, we conducted a combined-ancestry GWAS of diabetic retinopathy, totaling 68,169 cases and 129,188 controls. Individuals of NH-AFR ancestry are overrepresented in the MVP compared to the U.S. population, allowing us to investigate the genetic architecture of diabetic retinopathy in individuals of NH-AFR ancestry on an unprecedented scale. Additionally, unlike previous studies, we interrogated the effects of variants on the X chromosome on diabetic retinopathy risk. We also present the largest ancestry-stratified SNP-based estimation of heritability of diabetic retinopathy ranging from 3.6% to 8.0% between ancestries, though these estimations do not include other contributing factors to heritability, such as rare variants, gene-gene interactions, and gene-environment interactions. We detected nine previously unreported loci associated with diabetic retinopathy, including an evolutionarily adaptive genetic variant with potential impacts on racial disparities in diabetes complications in individuals of NH-AFR ancestry, and we propose a functional mechanism for this association.

The previously unreported loci associated with diabetic retinopathy in the combined-ancestry analysis included nine genes based on proximity to lead SNPs (<250 Kb): *MFSD4A*, *TENM2*, *GRHL2*, *TCF7L2*, *FBXW8, COL4A1*, *SLC16A3*, *TMPRSS6*, and *G6PD*. While these genes have varied functions, previous studies have reported associations with retinal phenotypes (*TENM2* and *COL4A1*), diabetes (*TCF7L2*, *MFSD4A*, and *GRHL2*), and changes in reactive oxygen species (*G6PD* and *GRHL2*).[26–29] We also observed differing genetic architectures of diabetic retinopathy for individuals of EUR- and AFR-ancestry.

The *TCF7L2* locus, significant in the NH-EUR ancestry analysis, was previously reported to have the strongest genetic association with risk of type 2 diabetes. *TCF7L2* expression impacts insulin secretion, particularly through the expression of splicing variants.[30–32] Our results suggest that these non-coding variants may increase risk of diabetes severity in individuals of NH-EUR ancestry, as hyperglycemia is a major risk factor for diabetes complications. An intronic locus in *COL4A1,* most significant in the NH-EUR ancestry analysis, was associated with increased risk of diabetic retinopathy. *COL4A1* has been associated with retinal tortuosity (increased twisting, bending, or turning of blood vessels), with tortuous vessels at increased risk of atherosclerosis and rupture.[28,33–36] Therefore,

in individuals who are at high risk of diabetic retinopathy, as retinal ischemia worsens, neovascularization occurs and retinal permeability increases, higher tortuosity may further add to the risk of retinal hemorrhages. These findings shed light into non-glycemic mechanisms that increase risk of diabetic retinopathy.

We identified a single significant locus that is exonic in the glucose 6-phosphate dehydrogenase (*G6PD*) gene in the NH-AFR ancestry analysis. The *G6PD* locus was also identified in the NH-AFR ancestry PDR GWAS and GPGE analyses, as well as in DME. The lead variant in the analyses was rs1050828, where the risk allele (T) is a missense mutation resulting in a valine to methionine substitution and has only been observed in AFR ancestry populations.[37] This variant 1) is causal for G6PDdef, 2) is associated with a 20% to 60% reduction in G6PD levels, 3) causes acute intermittent hemolysis, 4) is associated with reduced HbA1c, and 5) has been shown to be protective from severe malaria disease.[18,23,38] The G6PDdef risk allele has also been shown to reduce an individual's HbA1c due to the decreased lifespan of red blood cells which makes HbA1c an unreliable marker for detection of hyperglycemia in individuals with this deficiency.[21–23]

G6PD, encoded by the X-linked gene *G6PD*, is involved in the pentose phosphate pathway and the generation of NADPH, and enzymatic deficiencies are known to increase levels of reactive oxygen species. G6PDdef is the most common enzymopathy, affecting nearly 500 million individuals worldwide.[9,10] Studies have identified approximately 220 genetic variants associated with G6PDdef. In parts of the world where malaria is endemic, these mutations have been shown to be under recent positive selection due to protection from severe *Plasmodium falciparum* malaria disease.[18] As *P. falciparum* malaria is endemic in Africa and Asia, G6PDdef disproportionately affects individuals of AFR (7.5%) and ASN (4.7%) ancestry, compared to individuals of EUR ancestry (3.4%).[10] G6PDdef is often undiagnosed, as individuals carrying the mutation are often asymptomatic unless they are exposed to an exogenous agent, such as fava beans, alcohol, or medications such as sulfa drugs that can increase oxidative stress and trigger an acute hemolytic anemia.[14]

Individuals with G6PDdef have reduced hemoglobin A1c (HbA1c) levels relative to their underlying glucose levels, as the hemolytic anemia associated with G6PDdef reduces the erythrocyte lifespan, shortening the length of time available for the glycation of hemoglobin.[14,15] As a result, these individuals' plasma glucose levels are a more accurate measure of their glycemic control than their HbA1c levels which tend to be misleadingly low in individuals with G6PDdef.[14]

With an allele frequency between 8% and 12%, approximately 1.6 to 2.4 million men, and 3.1 to 4.5 million women of NH-AFR ancestry carry this allele. Considering the prevalence of diabetes in NH-AFR ancestry individuals is as much as 20% in the United States, we estimate that there are approximately 270,000 men and 510,000 women of NH-AFR ancestry with diabetes and some level of G6PDdef derived from rs1050828 genotype.[39] This is broadly consistent with estimated rates of diabetes misdiagnosis due to rs1050828-induced G6PDdef and HbA1c depression of approximately 650,000 missed diabetes cases in the U.S.[21] With a crude excess risk of diabetic retinopathy for the G6PDdef risk allele of 119 per 1000 men of NH-AFR ancestry with diabetes, we estimate that

nearly 12% of all diabetic retinopathy cases in men of NH-AFR ancestry in the U.S. could be attributed to the G6PDdef risk allele.[40] We estimate that screening as few as 364 individuals of NH-AFR ancestry and altering diabetes diagnosis and treatment to include glucose as well as genotype-adjusted HbA1c would prevent one case of diabetic retinopathy. Alternatively, if only NH-AFR people diagnosed with diabetes were screened, with the optimistic assumption that the effects of G6PDdef risk allele were completely mitigated by adjusted treatment after diagnosis, screening as few as 48 people could prevent one case of diabetic retinopathy.

We demonstrate that there is no distinguishable difference in mean HbA1c between men of NH-AFR ancestry with diabetes without the G6PDdef risk allele and men of NH-AFR ancestry with diabetic retinopathy and the G6PDdef risk allele; however, there is a significant difference in the mean plasma glucose between the two groups (Figure 2a – 2b). This suggests that in a system that predominantly relies on HbA1c to manage glycemic control, individuals with the G6PDdef risk allele are more likely to evade detection of hyperglycemia. Furthermore, this trend was evident prior to diabetes diagnosis as well as prior to insulin prescription, suggesting individuals with the G6PDdef risk allele may be exposed to a higher glycemic load over the course of their pre-diabetes and post-diagnosis diabetes management. This evidence is additionally supported with the mediation analysis, which revealed complete mediation of the effect of the G6PDdef risk allele on diabetic retinopathy through plasma glucose levels. This suggests that the G6PDdef risk alleles contribute to diabetic retinopathy risk through a reduced tendency to intensify treatment since diagnostic classification and management may be based on HbA1c levels rather than plasma glucose levels.

We hypothesized, if the mechanism of action for the association between G6PDdef risk allele and diabetic retinopathy is indeed due to inadequate control of hyperglycemia in diabetes, then we should also detect similar associations with other diabetes complications. We identified associations with both diabetic kidney disease and diabetic macular edema in large-scale EHR data. Primary findings for diabetic retinopathy were not driven by the association observed for diabetic macular edema. Furthermore, when investigating the risk of incident diabetes complications in longitudinal data from the ACCORD clinical trial, we identified that individuals of NH-AFR ancestry with the G6PDdef risk allele had a significantly higher likelihood of diabetic retinopathy and diabetic neuropathy than individuals those without. These findings demonstrate that individuals with the G6PDdef risk allele receiving standard-of-care treatment have an increased incidence of diabetes complications compared to individuals without the G6PDdef risk allele.

We demonstrate the inverse of this scenario where an individual with a higher HbA1c compared to glucose levels are at a reduced risk of diabetic retinopathy. We identified the *TMPRSS6* locus, in which a non-synonymous variant has been previously shown to increase HbA1c (rs855791-A, $\beta = 0.027 \pm 0.004$, $p = 2.74 \times 10^{-14}$).[41,42] While the exact variant was nominally associated with diabetic retinopathy in the MVP NH-EUR ancestry analysis (rs855791-A, OR = 0.95 (0.93 – 0.97), $p = 3.35 \times 10^{-06}$) a variant in strong LD ($r^2 > 0.75$) was significantly associated with a reduced risk of diabetic retinopathy (rs2076085-C, OR = 0.956 (0.949 – 0.964), $p = 2.52 \times 10^{-08}$) in the combined-ancestry meta-analysis.

While HbA1c in most individuals may be an accurate indicator of average glycemia, this evidence further supports our hypothesis that factors other than glycemia may affect HbA1c levels and with greater identification of markers, may provide an opportunity for precision medicine based management of glycemia.

To further investigate the clinical consequences of G6PDdef, we conducted a PheWAS of the G6PDdef risk allele separately in individuals with and without diabetes. In individuals with diabetes, this revealed associations with multiple diabetes sequelae, including diabetic retinopathy, chronic kidney disease, end-stage renal disease, and congestive heart failure. In individuals without diabetes, we did not find associations with these outcomes, nor other microvascular complications, despite a substantially larger sample size. These results suggest G6PD deficiency increases risk of cardiovascular disease and diabetes complications in individuals with diabetes likely due to lack of proper management of diabetes. We cannot definitively rule out the role of increased in reactive oxygen species in individuals with G6PDdef, although the evidence does suggest the relative magnitude of effect for disease processes related to differential hyperglycemia is much larger.

In addition to the G6PDdef risk allele, other genetic variants under selective pressure due to protection from infectious diseases have been identified for malaria, leprosy, tuberculosis, cholera, influenza, and other conditions.[43] These include not only the G6PDdef risk allele, which disproportionately affects individuals of AFR- and ASN-ancestry compared to individuals of EUR ancestry, but also the *APOL1* G alleles which confers protection from sleeping sickness, sickle cell disease which confers protection from malaria, and the hypoxia-inducible factor risk alleles which confers protection from metazoans.[44–46] These results collectively suggest that adaptive processes influence disease risk and disparities in populations underrepresented in biomedical research.

Our findings raise several practical and clinical implications for management of diabetes in individuals with the G6PDdef risk allele. We provide evidence for a mechanism of action for this genetic variant, which likely increases risk of diabetic retinopathy and other diabetes complications through delayed diabetes diagnosis coupled with inadequate intervention based on HbA1c levels relative to underlying plasma glucose levels. Given the common occurrence of this deficiency and its impact on diabetic complications, screening individuals of NH-AFR ancestry with diabetes for this deficiency and adjusting HbA1c measures may be an effective approach. Alternatively, concurrently checking plasma glucose levels along with HbA1c in all individuals of NH-AFR ancestry could be another strategy; however, the cost-benefit assessment of screening for G6PDdef once vs. continuous dual screening with HbA1c and plasma glucose worthy of consideration and economic analysis. With comprehensive screening of individuals of NH-AFR ancestry and subsequent standard-of-care treatment, possibly aimed at glucose rather than HbA1c targets, nearly 12% of diabetic retinopathy cases and 23% of diabetic neuropathy in individuals of NH-AFR ancestry could be avoided in the U.S. alone. Our results suggest reductions in rates for other complications of diabetes in NH-AFR ancestry people as well, but we were unable to estimate risk differences with the resources we have in this study. Based on the evidence from this study, we suggest that incorporation of this screening into clinical processes would mitigate

disparities of prevalence and severity for other diabetes sequelae among individuals of NH-AFR ancestry.

LIMITATIONS

The main limitation of this study was the small number of women of AFR ancestry homozygous for the G6PDdef risk allele. As such, we were unable extrapolate the results observed in men of AFR ancestry confidently to women of AFR ancestry. Furthermore, the G6PDdef risk allele is present in populations with ASN ancestry, we were unable to investigate the risk of diabetic complications due to sample size limitations.

ONLINE METHODS

Study Populations

Vanderbilt University Medical Center's DNA Repository (BioVU) is a de-identified database of electronic health records (EHR) that are linked to patient DNA samples.[47] The Veterans Affairs Million Veteran Program (MVP) is a national research program that incorporates genomic data and health record data, collected from veterans, to investigate how genes, lifestyle, military experiences, and exposures affect health and wellness.[48,49] The United Kingdom Biobank (UKB) is a prospective cohort study with genetic and phenotypic data collected on approximately 500,000 individuals across the United Kingdom between 40 and 69 years of age at recruitment.[50,51] Mass General Brigham Biobank (MGBB; formerly Partners HealthCare Biobank) is a large repository of biospecimens and data linked to extensive electronic health record and survey data.[52] We utilized the discovery stage summary statistics from Pollack et al.[15] The Action to Control Cardiovascular Risk in Diabetes (ACCORD) study was designed to examine treatment intensity of glycemic control and blood pressure control and the corresponding effects on the prevention of major cardiovascular event in individuals with type 2 diabetes mellitus. In total we identified 66,179 cases and 126,227 controls across the five contributing sources (Table 1).

Race and Ethnicity

Race and ethnicity information for Vanderbilt University Medical Center's BioVU, the MVP, the UKB, and the MGBB have been previously described.[48,53–55] Briefly, contributing sites collected self-reported race and ethnicity information. The MVP utilized the harmonized ancestry and race/ethnicity (HARE) method to merge self-reported and ancestral data.[56] This resulted in four categories: individuals of predominately non-Hispanic African (NH-AFR)ancestry, individuals of predominately non-Hispanic European (NH-EUR) ancestry, individuals of predominately Asian (ASN) ancestry, and Hispanic individuals.

Identification of Diabetic Retinopathy

In BioVU, the MGBB, and the MVP, diabetic retinopathy cases and controls with diabetes were identified with a previously published algorithm.[57] Briefly, the algorithm utilized EHR data to identify individuals with International Classification of Diseases (ICD) and Current Procedural Terminology (CPT) code-based evidence of diabetic retinopathy, while controls

were individuals with ICD evidence of diabetes, no ICD evidence of diabetic retinopathy, as well as a comprehensive ophthalmology exam. Cases and controls were identified in the UK Biobank through the diabetic retinopathy PheCode. Proliferative diabetic retinopathy was identified in a similar manner as diabetic retinopathy in BioVU and the MVP, where cases had ICD and CPT code-based evidence of PDR and controls were individuals with ICD and CPT code-based evidence of diabetic retinopathy but not PDR.

In VUMC and the MVP, age was designated as the age of diabetic retinopathy diagnosis or age at last ophthalmology exam. Sex was classified based on self-reported data and individuals with discordant self-reported sex and genetically inferred sex were excluded from analyses. Duration of diabetes was classified in a similar manner, taking the difference between the date of diabetes mellitus diagnosis and the date of diabetic retinopathy diagnosis or date of the last ophthalmology exam, for cases and controls, respectively. Individuals with a negative or zero duration of diabetes (suggesting the individual received normal care outside of either VUMC or the VA) were removed from the analyses. Three mean plasma glucoses were calculated with an individual's outpatient plasma glucose tests (excluding point-of-care and inpatient tests), one beginning at the first criteria of diabetes, the second 365 days prior to first criteria of diabetes, and 365 days prior to the first prescription of insulin in MVP. Three mean HbA1c were calculated with all of an individual's HbA1c tests (excluding point-of-care), one beginning at the first criteria of diabetes, the second 365 days prior to first criteria of diabetes, and 365 days prior to the first prescription of insulin in MVP. We excluded individuals with a mean HbA1c (%) lower than 4% and greater than 20%. We generated predicted HbA1c by regressing mean HbA1c on mean plasma glucose of individuals with diabetes, resulting in the equation: $pHbA1c = 0.4384 + 0.01807(Plasma\ \ Glucose)$. We then calculated the difference in mean HbA1c and predicted HbA1c for each individual.

We defined type 2 diabetes using an algorithm designed specifically for UKB[58]. Type 2 diabetes, diabetic retinopathy, coronary artery disease, and chronic kidney disease were defined using the ICD 9 and 10 codes (Supplementary Table 25). MGBB included a total of 6,623 T2D cases. T2D status was defined based on "curated phenotypes" developed by the MGBB Portal team using structured and unstructured electronic medical record data and clinical, computational, and statistical methods. Natural Language Processing was used to extract data from narrative text. Chart reviews by disease experts helped identify features and variables associated with particular phenotypes and were also used to validate the results of the algorithms. The process produced robust phenotype algorithms that were evaluated using metrics such as positive predictive value (PPV) and negative predictive value (NPV).

Single Variant Analyses

We identified diabetic retinopathy – SNP and PDR – SNP associations in the BioVU and MVP with logistic regression using PLINK2, adjusting for age, sex, diabetes duration, mean HbA1c values post diabetes diagnosis, and the top 10 principal components of ancestry, henceforth defined as 'standard covariates', separately in individuals of EUR-, AFR-, and ASN-ancestry as well as Hispanic ethnicity. Data for the X chromosome was available in BioVU, MGBB, MVP, and the UKB. In BioVU and MVP, logistic regression analyses

were conducted in PLINK2 with the "xchr-model 2" option. We performed whole-genome regression analysis using REGENIE version 2 for UKB and MGBB. We used diabetic retinopathy as a binary outcome and included age, sex, body mass index (BMI) and 10 PCs and imputation batch for each cohort. We used a block size of 1,000 for step 1 and 400 for step 2. All variants with a MAC < 3 for UKB, and MGBB. To provide better-calibrated test statistics, REGENIE supports the option (--firth --approx --firth-se) to use the Firth correction for variants where the p-value from the standard logistic regression test is below a threshold, as defined by pthresh. For UKB and MGBB, we set pthresh to 0.999, causing the Firth correction to be applied. We conducted an inverse-variance weighted fixed effects meta-analysis, within and across ancestral groups with METAL.[59] We utilized LDSC to calculate the genomic inflation factor (λ_{GC}), linkage disequilibrium score regression intercept, heritability on the liability scale (h2), and the genetic correlation of diabetic retinopathy between EUR- and AFR-ancestries in the MVP.[60]

Secondary Analyses

We identified G6PDdef risk allele associations with a subtype of diabetic retinopathy, diabetic macular edema (n_{cases} = 3,812, $n_{controls}$ = 32,921), and diabetic kidney disease (n_{cases} = 3,919, $n_{controls}$ = 49,417), in individuals of NH-AFR ancestry in the MVP. Diabetic macular edema was identified through an ICD code-based method, requiring ≥ 2 ICD code days for diabetic macular edema, while controls were individuals with diabetic retinopathy and no ICD evidence of diabetic macular edema. Diabetic kidney disease was characterized as a >50% decline in eGFR from baseline over, confirmed by a second eGFR value more than 90 days apart with no recovery in between from the baseline kidney function at MVP enrollment or reaching incident end-stage renal disease. Both analyses were conducted with logistic regression using PLINK2, adjusting for age, sex, diabetes duration, mean HbA1c values, and the top 10 principal components of ancestry.

Sensitivity Analyses

We conducted sensitivity analyses for the diabetic retinopathy – G6PDdef risk allele association in individuals of NH-AFR ancestry. The first was a logistic regression with diabetic retinopathy cases and diabetes mellitus controls, separated by sex (n_{men} = 33,408, n_{women} = 3,487) and adjusted for age, sex, diabetes duration, mean HbA1c values, and the top 10 principal components of ancestry in the MVP. The second was a logistic regression in the MVP with diabetic retinopathy cases and diabetes mellitus controls, adjusted for age, sex, diabetes duration, mean HbA1c values, and the top 10 principal components of ancestry as well as the G6PDdef risk allele. The final analysis was conducted for the diabetic macular edema – G6PDdef risk allele association with logistic regression, adjusting for age, sex, diabetes duration, mean HbA1c values, and the top 10 principal components of ancestry. The analysis was restricted to cases with diabetic retinopathy and without diabetic macular edema (n_{max} = 10,556) and controls with diabetes (n_{max} = 22,365). The third sensitivity analysis was to investigate the fitted log-odds of diabetic retinopathy with a logistic regression of diabetic retinopathy in men of NH-AFR ancestry in the MVP (n_{max} = 45,159), separately for individuals with and without the G6PDdef risk allele, adjusting for age, duration of diabetes, and the top two principal components. The fitted odds were generated for men of NH-AFR ancestry of median age (64 years), median duration

of diabetes (7.5 years), median principal component 1 (−0.144), and median principal component 2 (−0.024). The final sensitivity analysis utilized the ACCORD clinical trial data to investigate if different medication regimes impacted the association between the G6PDdef risk allele and diabetic retinopathy in individuals of NH-AFR ancestry (n_{max} = 1,337). We conducted Cox proportional hazard models investigating the association between the G6PDdef risk allele and diabetic retinopathy including standard covariates as well as clinical trial treatment arm, metformin use, sulfonylurea use, thiazolidinedione use, and insulin use (individually and jointly).

Conditional Analyses

To identify independently associated SNPs, we used the Genome-wide Complex Traits Analysis (GCTA) software. The GCTA-COJO software performs iterative conditional stepwise selection of independently significant SNPs and joint analysis simultaneously with stepwise model selection.[61] The summary statistics from meta-analyses were used as the input summary data, and we utilized 5,000 unrelated individuals of NH-EUR ancestry from the UK Biobank or 2,217 unrelated individuals of NH-AFR ancestry from BioVU for the linkage disequilibrium structure. A p value threshold of 5×10^{-8} was used as the selection threshold within GCTA, and the collinearity threshold was set at the default value of 0.9, so that SNPs are not selected if the multiple regression with the current SNPs in the model has $r^2 \geq 0.9$. For any sets of SNPs in LD ($r^2 \geq 0.1$), we selected the most significant SNP with the minimum p value from the GCTA model.

Genetically Predicted Gene Expression and Colocalization

We evaluated genetically predicted gene expression (GPGE) with S-PrediXcan, a gene-level approach that estimates the genetically determined component of gene expression in a given tissue and tests it for association with an outcome using SNP-level summary statistics.[62] We used the common variants (minor allele frequency > 0.01) from the meta-analyses and the 49 tissues from GTEx V8, incorporating elastic net gene models and covariance matrices developed for either EUR- or AFR-ancestry from 1000 Genomes.[63] The Bonferroni-corrected significance threshold was 1.55×10^{-6} to account for the total number of gene models assessed across the tissues. Coloc, a Bayesian gene-level test, was used to test the hypothesis that a single variant underlies GWAS and eQTL associations at a given locus (i.e. colocalization).[62] Input included the common variants (MAF >= 1%) from the meta-analyses and eQTL summary statistics corresponding to the gene-expression references used in GPGE analysis, restricting to only variants included in the GTEx V8 models. A statistically significant GPGE result along with a corresponding posterior probability greater than 80% that SNPs were associated with both expression and the outcome was considered as strong evidence of colocalization.

Mediation Analyses

We conducted a mediation analysis to examine if the effect of variant rs1050828-T on diabetic retinopathy was fully or partially mediated through plasma glucose with the ‘mediate’ function, part of the R package ‘Mediation’.[64] We investigated the effect of the G6PDdef risk allele and plasma glucose on diabetic retinopathy with logistic regression,

utilizing the glm() function in R. We utilized the lm() function to investigate the effect of the G6PDdef risk allele on plasma glucose.

Phenome-Wide Association Study

We performed a phenome-wide association study (PheWAS) of the G6PDdef risk allele, rs1050828-T, in individuals of NH-AFR ancestry in the MVP ($n_{diabetes}$ = 33,282, $n_{no\text{-}diabetes}$ = 42,853) and BioVU ($n_{diabetes}$ = 1,117, $n_{no\text{-}diabetes}$ = 4,427).[65] We used logistic regression to separately model the ~1,800 PheCode traits as a function of the variant, adjusted for age, sex, and the top 10 principal components of ancestry separately for individuals with and without diabetes, as well as separately for men and women. Interpretation of results were limited to phenotypes with 20 or more cases. We meta-analyzed the PheWAS between sites. Bonferroni-corrected significance threshold was $p \leq 2.75 \times 10^{-5}$ to account for the total number of PheCode models assessed in these analyses. All PheWAS analyses were conducted using the PheWAS R package.[66]

Longitudinal Diabetic Complications

We performed a time-to-event analysis with the Cox proportional hazard model in R (‘coxph’ in the ‘survival’ package) using diabetes duration until event with data from the ACCORD trial, including individuals of NH-AFR ancestry (n_{max} = 1,337) (regardless of trial arm) with 9 diabetic complication outcomes with the G6PDdef risk allele, adjusting for age, sex, diabetes duration, and HbA1c.[67,68] Diabetes complication outcomes were excluded if there were less than or equal to five G6PDdef carriers with an event. The multiple testing correction threshold was 5.60×10^{-3} to account for the nine tests conducted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This research is based on data from the Million Veteran Program, Office of Research and Development, and Veterans Health Administration. This research was funded in part by the Extramural Research Programs of the NIH. This research was funded in part by the Intramural Research Program of the National Institute of the Environmental Health Sciences. This publication does not represent the views of the Department of Veterans Affairs or the United States Government. Mass General Brigham Biobank provided samples, genomic data, and health information data.

FUNDING

Efforts were supported by the NEI (F31 EY033663 (J.H.B.), T32 EY021453-10 (J.H.B.), R01 EY025295 (Y.S.), R01 EY032159 (Y.S.), P30 EY026877 (Y.S.), P30 EY025885 (N.S.P.), P30 EY011373 (S.K.I.)), NICHD (K12 HD043483 (J.N.H.)), NIAMS (R01 AR074989 (A.G.), K12 AR084232-24 (T.L.E.), NIDDK ((R01 DK127083 (M.K.R.), R01 DK127083 (L.S.P.), K01 DK120631 (A.G.)), NHLBI (R01 HL110380 (J.B.B.), R01 HL161516 (A.G.)), NIAID (R21 AI156161 (L.S.P.)), NHGRI (U01 HG011723 (J.M.M.), U01 HG011723 (A.L.)), NCATS (TL1 TR002244 (J.H.B), UM1 TR00406 (J.B.B.), UL1 TR002378 (L.S.P.)), NCCDPHP (U18 DP006711 L.S.P.), ‘VA Office of Research & Development (CSP 2002 (M.K.R.), CSP 2008 (L.S.P.), I01 CX001899 (L.S.P.), I01 CX001737 (L.S.P.), I01 BX005831 (L.S.P.), IK6 BX005233 (N.S.P.), I01 BX004557 (N.S.P.),, I01 CX001481 (Y.S.)), the Doris Duke Foundation (2020096 (A.L.)), the American Diabetes Association (1-19-ICTS-068 (J.M.M), 11-22-ICTSPM-16 (J.M.M.), 7-22-ICTSPM-23 (A.L.)), a Cystic Fibrosis Foundation Award (PHILLI12A0 (L.S.P.)), Unrestricted Grant from Research to Prevent Blindness (Stanford University, Y.S.), the Intramural Research Program of the National Institute of Environmental Health Sciences (J.H.B., J.S.H., A.M.R.).

DISCLOSURE STATEMENT:

L.S.P. declares that there is no duality of interest associated with this manuscript. With regard to potential conflicts of interest, L.S.P. has or had research support from Merck, Pfizer, Eli Lilly, Novo Nordisk, Sanofi, PhaseBio, Roche, Abbvie, Vascular Pharmaceuticals, Janssen, Glaxo SmithKline, and the Cystic Fibrosis Foundation. L.S.P. is also a cofounder and Officer and Board member and stockholder for a company, Diasyst, Inc., which markets software aimed to help improve diabetes management.

DATA AVAILABILITY

The individual-level data from BioVU, MVP, and MGBB are not able to be shared. Individual level data from the UK Biobank are freely available to approved researchers. Individual-level phenotype data are also available to UK Biobank approved researchers for the health-record datasets from which our trait of interest was derived. Instructions for access to UK Biobank data are available at https://www.ukbiobank.ac.uk/enable-your-research. The published article includes all significant results generated during this study. Summary statistics for genome-wide significant variants are available in supplementary tables. Statistically significant reports for S-PrediXcan results for all tissues and PheWAS analyses are also available in the Supplementary Tables.

REFERENCES

1. Teo ZL, et al. Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis. Ophthalmology 128, 1580–1591 (2021). [PubMed: 33940045]
2. Sachdeva MM Retinal Neurodegeneration in Diabetes: an Emerging Concept in Diabetic Retinopathy. Curr Diab Rep 21, 65 (2021). [PubMed: 34902066]
3. Yumnamcha T, Guerra M, Singh LP & Ibrahim AS Metabolic Dysregulation and Neurovascular Dysfunction in Diabetic Retinopathy. Antioxidants (Basel) 9(2020).
4. Miller RG & Orchard TJ Understanding Metabolic Memory: A Tale of Two Studies. Diabetes 69, 291–299 (2020). [PubMed: 32079705]
5. Wong TY, et al. Diabetic retinopathy in a multi-ethnic cohort in the United States. Am J Ophthalmol 141, 446–455 (2006). [PubMed: 16490489]
6. Varma R, Torres M, Peña F, Klein R. & Azen SP Prevalence of diabetic retinopathy in adult Latinos: the Los Angeles Latino eye study. Ophthalmology 111, 1298–1306 (2004). [PubMed: 15234129]
7. Zhang X, et al. Prevalence of Diabetic Retinopathy in the United States, 2005–2008. JAMA 304, 649–656 (2010). [PubMed: 20699456]
8. Lundeen EA, et al. Prevalence of Diabetic Retinopathy in the US in 2021. JAMA Ophthalmol 141, 747–754 (2023). [PubMed: 37318810]
9. Burdon KP, et al. Genome-wide association study for sight-threatening diabetic retinopathy reveals association with genetic variation near the GRB2 gene. Diabetologia 58, 2288–2297 (2015). [PubMed: 26188370]
10. Graham PS, et al. Genome-wide association studies for diabetic macular edema and proliferative diabetic retinopathy. BMC Med Genet 19, 71 (2018). [PubMed: 29739359]
11. Imamura M, et al. Genome-Wide Association Studies Identify Two Novel Loci Conferring Susceptibility to Diabetic Retinopathy in Japanese Patients with Type 2 Diabetes. Hum Mol Genet (2021).
12. Liu C, et al. Genome-wide association study for proliferative diabetic retinopathy in Africans. NPJ Genom Med 4, 20 (2019). [PubMed: 31482010]
13. Meng W, et al. 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 Ophthalmol 96, e811–e819 (2018). [PubMed: 30178632]

14. Peng D, et al. Common variants in or near ZNRF1, COLEC12, SCYL1BP1 and API5 are associated with diabetic retinopathy in Chinese patients with type 2 diabetes. Diabetologia 58, 1231–1238 (2015). [PubMed: 25819896]
15. Pollack S, et al. Multiethnic Genome-Wide Association Study of Diabetic Retinopathy Using Liability Threshold Modeling of Duration of Diabetes and Glycemic Control. Diabetes 68, 441–456 (2019). [PubMed: 30487263]
16. Shtir C, et al. Exome-based case-control association study using extreme phenotype design reveals novel candidates with protective effect in diabetic retinopathy. Hum Genet 135, 193–200 (2016). [PubMed: 26693933]
17. Looker HC, et al. Genome-wide linkage analyses to identify Loci for diabetic retinopathy. Diabetes 56, 1160–1166 (2007). [PubMed: 17395753]
18. Liang XY, et al. Evidence of positively selected G6PD A- allele reduces risk of Plasmodium falciparum infection in African population on Bioko Island. Mol Genet Genomic Med 8, e1061 (2020). [PubMed: 31872983]
19. Tishkoff SA, et al. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293, 455–462 (2001). [PubMed: 11423617]
20. Sabeti PC, et al. Positive natural selection in the human lineage. Science 312, 1614–1620 (2006). [PubMed: 16778047]
21. Leong A. & Wheeler E. Genetics of HbA1c: a case study in clinical translation. Curr Opin Genet Dev 50, 79–85 (2018). [PubMed: 29522974]
22. Chen Z, et al. Genome-wide association analysis of red blood cell traits in African Americans: the COGENT Network. Hum Mol Genet 22, 2529–2538 (2013). [PubMed: 23446634]
23. da Rocha JEB, et al. G6PD distribution in sub-Saharan Africa and potential risks of using chloroquine/hydroxychloroquine based treatments for COVID-19. Pharmacogenomics J 21, 649–656 (2021). [PubMed: 34302047]
24. Mahajan A, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 54, 560–572 (2022). [PubMed: 35551307]
25. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL & Im HK Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet 15, e1007889 (2019).
26. Farris JC, Pifer PM, Zheng L, Gottlieb E, Denvir J. & Frisch SM Grainyhead-like 2 Reverses the Metabolic Changes Induced by the Oncogenic Epithelial-Mesenchymal Transition: Effects on Anoikis. Mol Cancer Res 14, 528–538 (2016). [PubMed: 27084311]
27. Tomasoni M, et al. Genome-wide Association Studies of Retinal Vessel Tortuosity Identify Numerous Novel Loci Revealing Genes and Pathways Associated With Ocular and Cardiometabolic Diseases. Ophthalmol Sci 3, 100288 (2023).
28. Veluchamy A, et al. Novel Genetic Locus Influencing Retinal Venular Tortuosity Is Also Associated With Risk of Coronary Artery Disease. Arterioscler Thromb Vasc Biol 39, 2542–2552 (2019). [PubMed: 31597446]
29. Bansal A, Balasubramanian S, Dhawan S, Leung A, Chen Z. & Natarajan R. Integrative Omics Analyses Reveal Epigenetic Memory in Diabetic Renal Cells Regulating Genes Associated With Kidney Dysfunction. Diabetes 69, 2490–2502 (2020). [PubMed: 32747424]
30. Jin T. & Liu L. The Wnt signaling pathway effector TCF7L2 and type 2 diabetes mellitus. Mol Endocrinol 22, 2383–2392 (2008). [PubMed: 18599616]
31. Gloyn AL, Braun M. & Rorsman P. Type 2 diabetes susceptibility gene TCF7L2 and its role in beta-cell function. Diabetes 58, 800–802 (2009). [PubMed: 19336690]
32. Del Bosque-Plata L, Martínez-Martínez E, Espinoza-Camacho M. & Gragnoli C. The Role of TCF7L2 in Type 2 Diabetes. Diabetes 70, 1220–1228 (2021). [PubMed: 34016596]
33. Alavi MV, et al. Col4a1 mutations cause progressive retinal neovascular defects and retinopathy. Sci Rep 6, 18602 (2016). [PubMed: 26813606]
34. Han HC Twisted blood vessels: symptoms, etiology and biomechanical mechanisms. J Vasc Res 49, 185–197 (2012). [PubMed: 22433458]
35. Sears J, Gilman J. & Sternberg P Jr. Inherited Retinal Arteriolar Tortuosity With Retinal Hemorrhages. Archives of Ophthalmology 116, 1185–1188 (1998). [PubMed: 9747676]

36. Han HC, Chesnutt JK, Garcia JR, Liu Q. & Wen Q. Artery buckling: new phenotypes, models, and applications. Ann Biomed Eng 41, 1399–1410 (2013). [PubMed: 23192265]
37. Leong A, et al. Association of G6PD variants with hemoglobin A1c and impact on diabetes diagnosis in East Asian individuals. BMJ Open Diabetes Res Care 8(2020).
38. Mbanefo EC, et al. Association of glucose-6-phosphate dehydrogenase deficiency and malaria: a systematic review and meta-analysis. Sci Rep 7, 45963 (2017).
39. Cheng YJ, et al. Prevalence of Diabetes by Race and Ethnicity in the United States, 2011–2016. Jama 322, 23892398 (2019).
40. McKean-Cowdin R, et al. Prevalence and Risk Factors for DR in the African American Eye Disease Study. Investigative Ophthalmology & Visual Science 60, 1089–1089 (2019).
41. Soranzo N, et al. Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic pathways. Diabetes 59, 3229–3239 (2010). [PubMed: 20858683]
42. Wheeler E, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med 14, e1002383 (2017).
43. Karlsson EK, Kwiatkowski DP & Sabeti PC Natural selection and infectious disease in human populations. Nat Rev Genet 15, 379–393 (2014). [PubMed: 24776769]
44. Genovese G, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science 329, 841–845 (2010). [PubMed: 20647424]
45. Pauling L, Itano HA & et al. Sickle cell anemia a molecular disease. Science 110, 543–548 (1949). [PubMed: 15395398]
46. Bigham AW & Lee FS Human high-altitude adaptation: forward genetics meets the HIF pathway. Genes Dev 28, 2189–2204 (2014). [PubMed: 25319824]
47. Roden DM, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther 84, 362–369 (2008). [PubMed: 18500243]
48. Gaziano JM, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70, 214–223 (2016). [PubMed: 26441289]
49. Hunter-Zinck H, et al. Genotyping Array Design and Data Quality Control in the Million Veteran Program. Am J Hum Genet 106, 535–548 (2020). [PubMed: 32243820]
50. Allen NE, Sudlow C, Peakman T. & Collins R. UK biobank data: come and get it. Sci Transl Med 6, 224ed224 (2014).
51. Sudlow C, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12, e1001779 (2015).
52. Karlson EW, Boutin NT, Hoffnagle AG & Allen NL Building the Partners HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of Research Results, Recruitment Lessons and Operational Considerations. J Pers Med 6(2016).
53. Bycroft C, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018). [PubMed: 30305743]
54. Dumitrescu L, et al. Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. Genet Med 12, 648–650 (2010). [PubMed: 20733501]
55. Boutin NT, et al. The Evolution of a Large Biobank at Mass General Brigham. J Pers Med 12(2022).
56. Fang H, et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. Am J Hum Genet 105, 763–772 (2019). [PubMed: 31564439]
57. Breeyear JH, et al. Development of Portable Electronic Health Record Based Algorithms to Identify Individuals with Diabetic Retinopathy. medRxiv, 2023.2011.2010.23298311 (2023).
58. Eastwood SV, et al. Algorithms for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank. PLoS One 11, e0162388 (2016).
59. Willer CJ, Li Y. & Abecasis GR METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010). [PubMed: 20616382]
60. Bulik-Sullivan BK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics 47, 291–295 (2015). [PubMed: 25642630]

61. Yang J, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nature Genetics 44, 369–375 (2012). [PubMed: 22426310]

62. Barbeira AN, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nature Communications 9, 1825 (2018).

63. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015). [PubMed: 25954001]

64. Tingley D, Yamamoto T, Hirose K, Keele L. & Imai K. mediation: R Package for Causal Mediation Analysis. Journal of Statistical Software 59, 1 – 38 (2014). [PubMed: 26917999]

65. Denny JC, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 26, 1205–1210 (2010). [PubMed: 20335276]

66. Carroll RJ, Bastarache L. & Denny JC R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics 30, 2375–2376 (2014). [PubMed: 24733291]

67. Gerstein HC, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 358, 2545–2559 (2008). [PubMed: 18539917]

68. Therneau TM A Package for Survival Analysis in {R}. (2020).

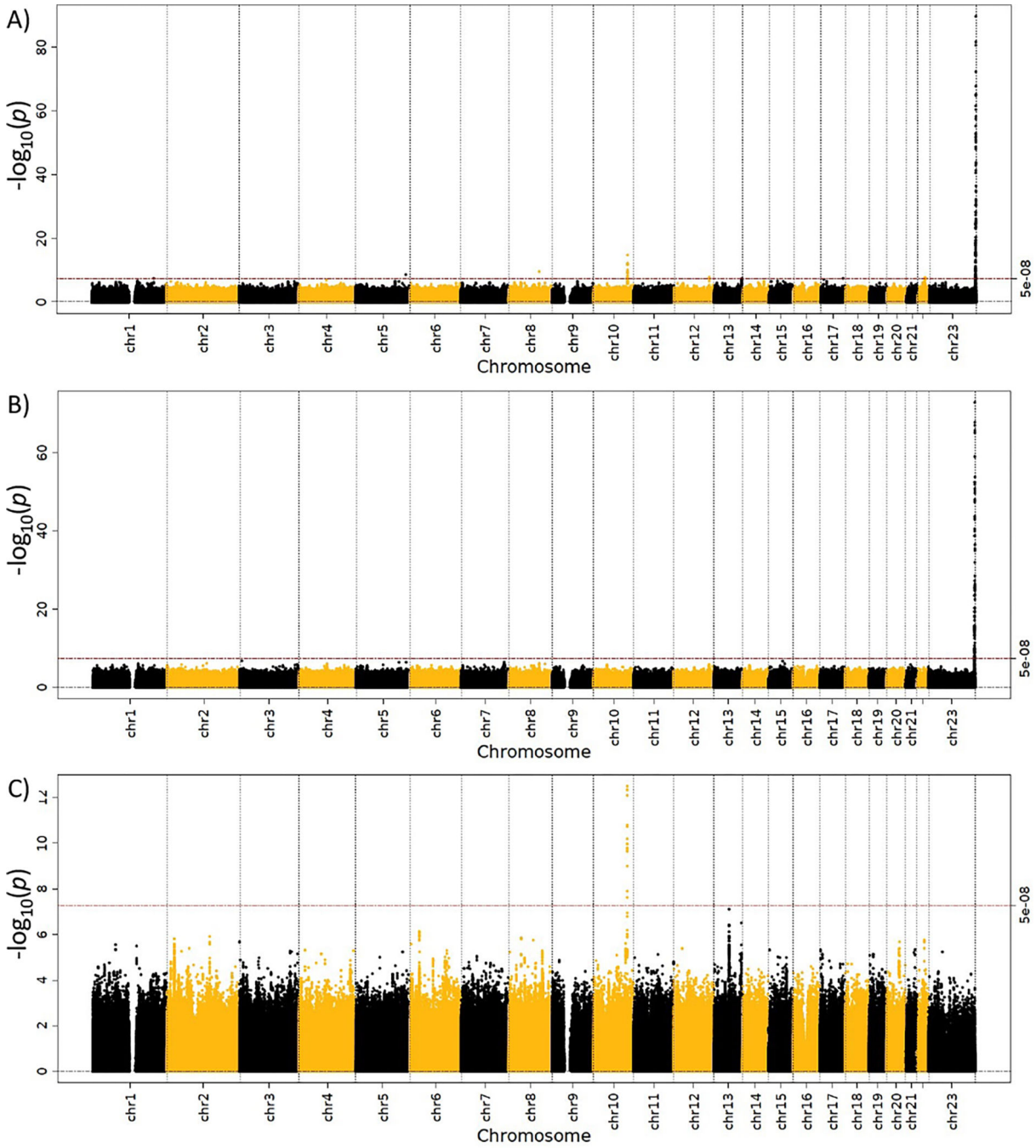

Figure 1.

Manhattan plots depicting the diabetic retinopathy meta-analysis. **A–C**, Manhattan plot of the diabetic retinopathy meta-analysis for (**A**) combined-, (**B**) NH-AFR-, and (**C**) NH-EUR-Ancestry. The y-axis shows the -log10[*p*], and the x-axis shows the chromosomal positions. The horizontal red line represents the threshold of $p = 5 \times 10^{-8}$ for genome-wide significance. All *p* values are computed for associations between genotyped or imputed SNPs and diabetic retinopathy as dependent variables in multivariable adjusted logistic regression models.

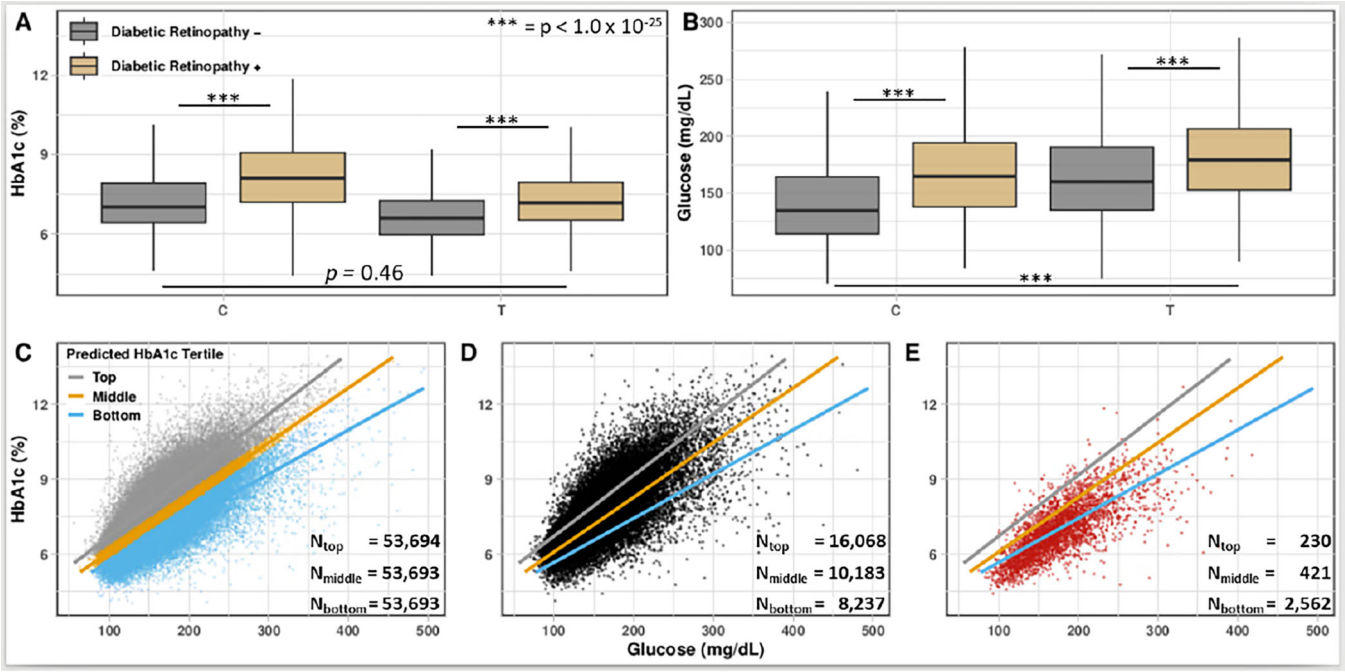

Figure 2.

Box plots summarizing differences in HbA1c, mean plasma glucose, and predicted HbA1c. **A–E**, Box plot of (**A**) mean HbA1c by diabetic retinopathy and G6PDdef allele status, (**B**) mean plasma glucose by diabetic retinopathy and G6PDdef allele status. Scatter plot of (**C**) mean HbA1c and mean plasma glucose in men with diabetes, (**D**) men of NH-AFR ancestry with diabetes, and (**E**) men with the G6PDdef risk allele with an overlay of the difference in HbA1c and predicted HbA1c from (**A**). The y-axis shows either mean HbA1c (%) or plasma glucose (mg/dL), the x-axis shows either G6PDdef allele status or mean plasma glucose (mg/dL), the colored lines of (**C** – **E**) represent difference in HbA1c and predicted HbA1c (blue = top tertile, orange = middle tertile, grey = bottom tertile).

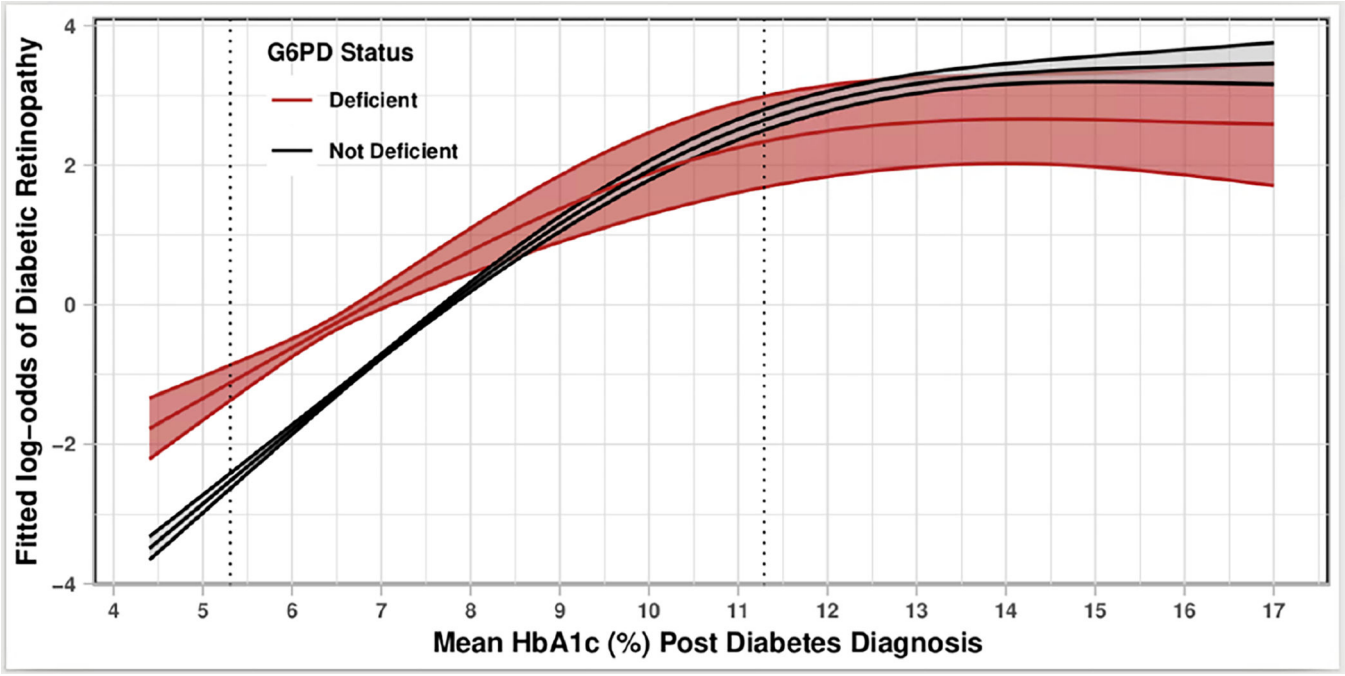

Figure 3.

Line graph summarizing the fitted log-odds of diabetic retinopathy and corresponding 95% confidence band by mean HbA1c post diabetes diagnosis, stratified by G6PDdef risk allele status in men of NH-AFR ancestry with diabetes. The y-axis shows mean fitted log-odds of diabetic retinopathy, the x-axis shows HbA1c (%), the colored lines represent G6PDdef risk allele status, between the dotted lines represents 99% of the data.

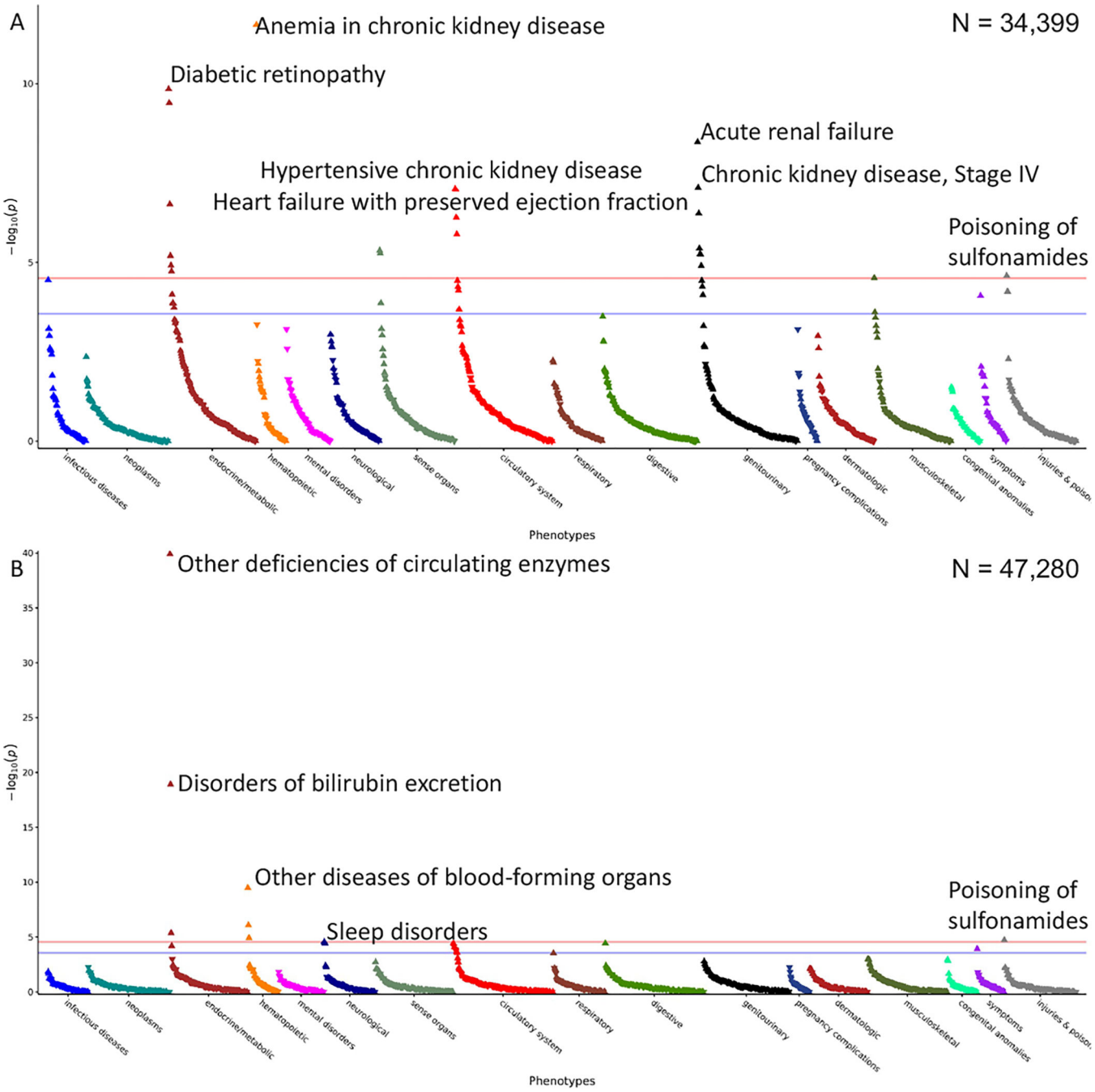

Figure 4.

PheWAS plots summarizing the association of the G6PDdef risk allele with PheCodes in men with and without diabetes with NH-AFR ancestry. **A–B**, Plot of PheWAS results for individuals with diabetes (**A**) and plot of PheWAS results for individuals without diabetes (**B**). The y-axis shows the -log10[*p*], and the x-axis shows the PheCode Groups. The horizontal red line represents the threshold of $p = 1.55 \times 10^{-5}$ for genome-wide significance.

Table 1:

Patient Characteristics from BioVU, Million Veteran Program, UK Biobank, and Mass General Brigham Biobank

Characteristic	BioVU		Million Veteran Program		UK Biobank		Mass General Brigham Biobank	
	DR Cases (N=1,124)	Controls (N=1,375)	DR Cases (N=63,394)	Controls (N=107,163)	DR Cases (N=1,318)	Controls (N=16,807)	DR Cases (N=343)	Controls (N=932)
Mean Age ± SD, years	61.95 ± 13.65	64.04 ± 14.34	63.08 ± 9.77	68.48 ± 10.36	62.28 ± 6.30	60.59 ± 7.07	67.8 ± 12.4	69.0 ± 13.1
Diabetes Duration ± SD, years	4.90 ± 5.02	7.88 ± 5.21	6.16 ± 5.05	9.34 ± 5.43	9.36	8.17	–	–
Mean HbA1c ± SD, %	6.9 ± 1.5	5.9 ± 0.8	7.84 ± 1.19	7.03 ± 0.96	7.71 ± 1.56	6.97 ± 1.24	8.0 ± 1.3	6.9 ± 1.0
Mean Glucose ± SD, % mg/dL	79.0 ± 61.5	62.73 ± 28.61	170 ± 38.4	145 ± 34	163.3 ± 80.8	133.7 ± 58.0	–	–
Women, N (%)	583 (52%)	762 (55%)	2,648 (4.3%)	7,229 (6.5%)	469 (31.9%)	9,175 (37.6%)	219 (47%)	512 (45)
Proliferative DR, N (%)	308 (27%)	–	7,673 (12%)	–	–	–	–	–
Ancestry/ethnicity								
NH-EUR Ancestry, N (%)	827 (74%)	1,059 (77%)	38,944 (61%)	69,416 (65%)	1,168 (89%)	15,286 (91%)	268 (57%)	809 (71%)
NH-AFR Ancestry, N (%)	297 (26%)	316 (23%)	17,587 (28%)	27,572 (26%)	45 (3%)	419 (2%)	75 (16%)	123 (11%)
Hispanic, N (%)	–	–	6,193 (10%)	8,910 (8%)	–	–	–	–
Asian Ancestry, N (%)	–	–	670 (1%)	1,265 (1%)	105 (8%)	1,102 (7%)	–	–

Table 2.

Significant Loci and Corresponding Sentinel SNPs of the Diabetic Retinopathy Meta-Analyses

Chr	SNP	Effect Allele	Freq.	OR (95% CI)	*p*	Location	Nearest Gene
1	rs115170237	A	0.93	1.19 (1.15 – 1.23)	4.04×10^{-08}	Intronic	*MFSD4A*
5	rs11950268	T	0.05	0.81 (0.78 – 0.84)	2.56×10^{-09}	Intronic	*TENM2*
8	rs76624501	T	0.06	1.22 (1.18 – 1.26)	2.57×10^{-10}	Intergenic	*GRHL2*
10	rs7903146	T	0.33	1.07 (1.06 – 1.08)	2.22×10^{-15}	Intronic	*TCF7L2*
12	rs114926776	T	0.07	0.84 (0.81 – 0.86)	1.56×10^{-08}	Intronic	*FBXW8*
13	rs7333159	A	0.87	1.13 (1.11 – 1.16)	4.86×10^{-08}	Intronic	*COL4A1*
17	rs73995757	C	0.89	1.14 (1.12 – 1.17)	4.02×10^{-08}	Intronic	*SLC16A5*
22	rs2076085	C	0.43	0.96 (0.95 – 0.96)	2.48×10^{-08}	Intronic	*TMPRSS6*
X	rs1050828	T	0.09	1.48 (1.45 – 1.51)	1.99×10^{-90}	Exonic	*G6PD*