Multiethnic Genome-wide Association Study of Diabetic Retinopathy using Liability Threshold Modeling of Duration of Diabetes and Glycemic Control

Samuela Pollack[1], Robert P. Igo, Jr.[2], Richard A. Jensen[3], Mark Christiansen[3], Xiaohui Li[4], Ching-Yu Cheng[5,6], Maggie C.Y. Ng[7,8], Albert V. Smith[9], Elizabeth J. Rossin[10], Ayellet V. Segrè[10], Samaneh Davoudi[10], Gavin S. Tan[5,6], Yii-Der Ida Chen[4], Jane Z. Kuo[4,11], Latchezar M. Dimitrov[7,8], Lynn K. Stanwyck[10], Weihua Meng[12], S. Mohsen Hosseini[13ˈ], Minako Imamura[14, 15, 16], Darryl Nousome[17], Jihye Kim[18], Yang Hai[4], Yucheng Jia[4], Jeeyun Ahn[19], Aaron Leong[20], Kaanan Shah[21], Kyu Hyung Park[22], Xiuqing Guo[4], Eli Ipp[23], Kent D. Taylor[4], Sharon G. Adler[24], John R. Sedor[25, 26, 27], Barry I. Freedman[28], Family Investigation of Nephropathy and Diabetes-Eye Research Group, DCCT/EDIC Research Group, I-Te Lee[29,30,31], Wayne H-H Sheu[29,30,31,32], Michiaki Kubo[33], Atsushi Takahashi[34,35], Samy Hadjadj [36,37,38,39], Michel Marre[40,41,42], David-Alexandre Tregouet[43,44], Roberta Mckean-Cowdin[17,45], Rohit Varma[17,45], Mark I. McCarthy[46,47,48], Leif Groop[49], Emma Ahlqvist[49], Valeriya Lyssenko[49,50], Elisabet Agardh[49], Andrew Morris[51], Alex S.F. Doney[52], Helen M. Colhoun[53], Iiro Toppila[54,55,56], Niina Sandholm[54,55,56], Per-Henrik Groop[54,55,56,57], Shiro Maeda[14, 15, 16], Craig L. Hanis[18], Alan Penman[58], Ching J. Chen[59], Heather Hancock[59], Paul Mitchell[60], Jamie E. Craig[61], Emily Y. Chew[62], Andrew D. Paterson[63,64,65], Michael A. Grassi[66,67], Colin Palmer[68], Donald W. Bowden[7,8], Brian L. Yaspan[69], David Siscovick[70], Mary Frances Cotch[62], Jie Jin Wang[5, 60], Kathryn P. Burdon[71], Tien Y. Wong[5,72] , Barbara E. K. Klein[73], Ronald Klein[73], Jerome I. Rotter[4], Sudha K. Iyengar[2], Alkes Price[1]*, Lucia Sobrin[10]*

1. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA
2. Department of Population and Quantitative Health Sciences, Case Western University, Cleveland, OH
3. Cardiovascular Health Research Unit, Department of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA
4. Institute for Translational Genomics and Population Sciences, LABioMed and Department of Pediatrics at Harbor-UCLA Medical Center, Torrance, CA
5. Duke-NUS Medical School, Singapore
6. Singapore Eye Research Institute, Singapore National Eye Centre, Singapore
7. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC
8. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC
9. Department of Medicine, University of Iceland, Reykjavík, Iceland
10. Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA
11. Medical Affairs, Ophthalmology, Sun Pharmaceutical Industries, Inc, Princeton, NJ
12. Division of Population Health Sciences, Ninewells Hospital and Medical School, University of Dundee School of Medicine, Scotland, UK
13. Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada ˈCurrently at: Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA
14. Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan

Diabetes Publish Ahead of Print, published online November 28, 2018

15. Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara, Japan
16. Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Nishihara, Japan
17. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA
18. Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX
19. Department of Ophthalmology, Seoul National University College of Medicine, SMG-SNU Boramae Medical Center, Seoul, Korea
20. Endocrine Unit, Diabetes Unit, Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA
21. Section of Genetic Medicine, The University of Chicago, Chicago, IL
22. Department of Ophthalmology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Korea
23. Section of Diabetes and Metabolism, Harbor-UCLA Medical Center, University of California, Los Angeles, Los Angeles County, CA
24. Department of Nephrology and Hypertension, Los Angeles Biomedical Research Institute at Harbor-University of California, Torrance, CA
25. Department of Medicine, Case Western Reserve University, Cleveland, OH
26. Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, OH
27. Division of Nephrology, MetroHealth System, Cleveland, OH
28. Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC
29. Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan
30. School of Medicine, Chung Shan Medical University, Taichung, Taiwan
31. School of Medicine, National Yang-Ming University, Taipei, Taiwan
32. School of Medicine, National Defense Medical Center, Taipei, Taiwan
33. RIKEN Center for Integrative Medical Sciences, Yokohama, 230-0045 Japan
34. Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Kanagawa 230-0045, Japan.
35. National Cerebral and Cardiovascular Center, Research Institute, Department of Genomic Medicine, Osaka 565-8565, Japan
36. CHU de Poitiers, Centre d'Investigation Clinique, Poitiers, France
37. Université de Poitiers, UFR Médecine Pharmacie, CIC1402, Poitiers, France
38. INSERM, CIC1402, Poitiers, France
39. L'Institut du Thorax, INSERM, CNRS, CHU Nantes, Nantes, France
40. Université Paris Diderot, Sorbonne Paris Cité, Paris, France
41. Diabetology, Endocrinology and Nutrition Department, DHU FIRE, Bichat Hospital, AP-HP, Paris, France
42. INSERM U1138, Centre de Recherche des Cordeliers, Paris, France

43. Sorbonne Université, UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Paris, France
44. ICAN Institute for Cardiometabolism and Nutrition, Paris, France
45. USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, CA
46. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ UK
47. Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK
48. Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ UK
49. Department of Clinical Sciences, Faculty of Medicine, Lund University, Malmo, Sweden
50. Department of Clinical Science, KG Jebsen Center for Diabetes Research, University of Bergen, Norway
51. The Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK, EH8 9AG
52. Molecular and Clinical Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, DD1 9SY, UK
53. Institute of Genetics and Molecular Medicine, Western General Hospital, Crewe Road, University of Edinburgh, Edinburgh, UK, EH4 2XUT
54. Folkhälsan Institute of Genetics, Folkhälsan Research Center, 00290, Helsinki, Finland
55. Abdominal Center, Nephrology, University of Helsinki and Helsinki University Hospital, 00290, Helsinki, Finland
56. Research Programs Unit, Diabetes and Obesity, University of Helsinki, 00290, Helsinki, Finland
57. Department of Diabetes, Central Clinical School, Monash University, Melbourne, Victoria, Australia
58. Department of Preventive Medicine, John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, MS
59. Department of Ophthalmology, University of Mississippi Medical Center, Jackson, MS
¦Currently at: Retina Center, North Mississippi Medical Center, Tupelo, MS
60. Centre for Vision Research, Westmead Institute for Medical Research, University of Sydney, Sydney, Australia
61. Department of Ophthalmology, Flinders University, Bedford Park SA, Australia
62. Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD
63. Institute of Medical Sciences, University of Toronto, Toronto, Canada
64. Program in Genetics and Genome Biology Hospital for Sick Children's, Toronto, Ontario, Canada
65. Epidemiology & Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
66. Grassi Retina, Naperville, IL
67. Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, 1012 95th St., Suite 9, Chicago, IL

68. Pat MacPherson Centre for Pharmacogenetics and Pharmacogenomics, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK
69. Genentech Inc., South San Francisco, CA
70. Institute for Urban Health, New York Academy of Medicine, New York City, New York
71. Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia
72. Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
73. Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison, Madison, WI

*These authors contributed equally to this work

Corresponding Author:
Lucia Sobrin, MD, MPH
Massachusetts Eye and Ear Infirmary
Department of Ophthalmology, Harvard Medical School
243 Charles Street
Boston, Massachusetts 02114
Tel: 617-573-4279 Fax: 617-573-3011
Email: Lucia_Sobrin@meei.harvard.edu

Running title: Genetic Study of Diabetic Retinopathy
Word Count: 3997
Number of tables: 6 (main); 34 (online Supplemental)
Number of figures: 2 (main); 7 (online Supplemental)

Abstract

To identify genetic variants associated with diabetic retinopathy (DR), we performed a large, multiethnic genome-wide association study (GWAS). Discovery included eight European cohorts (n = 3,246) and seven African American cohorts (n = 2,611). We meta-analyzed across cohorts using inverse-variance weighting, with and without liability threshold modeling of glycemic control and duration of diabetes. Variants with a P value < 1 X 10^{-5} were investigated in replication cohorts that included 18,545 Europeans, 16,453 Asians and 2,710 Hispanics. After correction for multiple testing, the C allele of rs142293996 in an intron of nuclear VCP-like (*NVL*) was associated with DR in European discovery cohorts (P = 2.1 x 10^{-9}), but did not reach genome-wide significance after meta-analysis with replication cohorts. We applied the Disease Association Protein-Protein Link Evaluator (DAPPLE) to our discovery results to test for evidence of risk being spread across underlying molecular pathways. One protein-protein interaction network built from genes in regions associated with proliferative DR (PDR) was found to have significant connectivity (P=0.0009) and corroborated with gene set enrichment analyses. These findings suggest that genetic variation in *NVL,* as well as variation within a protein-protein interaction network that includes genes implicated in inflammation, may influence risk for DR.

Diabetic retinopathy (DR) is a leading cause of blindness.(1) Established risk factors include longer duration of diabetes (DoD) and poor glycemic control.(2) Genetic factors are also implicated, with heritability of 52% for proliferative diabetic retinopathy (PDR).(3, 4) Several candidate gene and genome-wide association studies (GWAS) have been conducted.(5-11) While several polymorphisms have been suggested to be associated with DR, few have been convincingly replicated.(10, 12-15)

There are several reasons why studies have not yielded consistent findings. The genetic effects are likely modest and identification requires large sample sizes. Previous studies have not consistently accounted for the strongest two covariates, DoD and glycemic control. Liability threshold (LT) modeling is one way to incorporate these covariates while also increasing statistical power.(16) Finally, previous genetic studies have largely examined individual variants. Techniques that examine top GWAS findings collectively for variants that cluster in biological networks based on known protein-protein interactions have the potential to identify variants where there is insufficient power to detect their individual effects.

The purpose of this study was to identify genetic variants associated with DR by (1) assembling a large sample size through inclusion of multiple ethnicities, (2) incorporating DoD and glycemic control via LT modeling and (3) collectively examining variants that cluster in biological networks.

Research Design and Methods

All studies conformed to the Declaration of Helsinki tenets and were Health Insurance Portability and Accountability Act (HIPAA) compliant. Written informed consent was obtained from all participants. Institutional Review Board /Ethics Committee approval was obtained by each individual study.

Discovery Sample Description

The discovery sample, encompassing seven African American and eight European cohorts, arose from a consortium of 11 DR studies for a total of 3246 Europeans and 2611 African Americans. (6-8, 12, 13, 17, 18) Inclusion criteria for the discovery stage were (1) type 2 diabetes and (2) European or African American ethnicity. Type 2 diabetes was defined as a fasting plasma glucose (FPG) $\geq$ 126 mg/dL (7.0 mmol/L) or a hemoglobin A_{1C} (HbA_{1C}) $\geq$ 6.5% (48 mmol/mol) (19) with onset of the diabetes after age 30 years. Table 1 summarizes the DR phenotyping protocols and covariates by discovery cohort. Phenotyping protocols have been previously described (4, 20-29) and additional details are in the Supplemental Materials.

DR Case-Control Definitions

The analysis plan pre-specified four DR case-control definitions with varying Early Treatment Diabetic Retinopathy Study (ETDRS) score thresholds for cases and controls (Table 2).(30) The primary case-control definition compared any DR to no DR (ETDRS $\geq$ 14 vs. ETDRS < 14, henceforth referred to as the any DR analysis). There were three secondary case-control definitions. The first compared patients with PDR to those without PDR (ETDRS $\geq$ 60 vs.

ETDRS < 60, henceforth the PDR analysis). The second compared those with non-proliferative DR (NPDR) or worse to those without DR (ETDRS ≥ 30 vs. ETDRS < 14, henceforth the NPDR analysis). The third compared those with PDR to those without DR (ETDRS ≥ 60 vs. ETDRS < 14, henceforth the extremes of DR analysis). The rationale for four definitions is in the Supplemental Materials. Table 1 shows the available samples by cohort and ETDRS score thresholds. Supplemental Table 1 summarizes the mean values for glycemic control and DoD.

Statistical Analyses

The genotyping platforms and the numbers of single nucleotide polymorphisms (SNPs) genotyped are summarized in Supplemental Table 2. Details about quality control, imputation, and data filtering are in the Supplemental Materials. Supplemental Figure 1 provides a flow chart of the discovery and replication analyses. For the four main case-control definition analyses, we performed each of the analyses (1) without incorporating DoD and glycemic control using EIGENSOFT (16, 31) and (2) with LT modeling of DoD and glycemic control using LTSCORE.(16) LT modeling details are in the Supplemental Materials. Both the EIGENSOFT and LTSCORE tests were implemented in LTSOFT version 2.0 (see Web Resources in Supplemental Material). For the discovery analyses, we ran principal components (PC) analysis with EIGENSTRAT using only typed SNPs and five PCs, separately by ethnicity and case-control definition.(32) We computed association analyses for each of the seven African American and eight European cohorts separately and then meta-analyzed by ethnicity. Meta-analysis was performed using inverse-variance weighting, accounting for both effective sample size (defined as $4/[1/N_{case} + 1/N_{control}]$) and allele frequency.(33) We also performed multiethnic (Europeans and African Americans together) meta-analyses for the any DR and PDR analyses

using inverse-variance weighting and a sensitivity analysis of the any DR meta-analyses in African Americans and Europeans (see Supplemental Materials). Because we included rare variants in this GWAS, we also tested the robustness of the top associations ($P < 5 \times 10^{-8}$) by performing two additional tests: (1) a Fisher's exact test on cases or controls aggregated across all cohorts tested per variant and on each cohort separately and (2) an inverse variance-weighted meta-analysis across cohorts using the natural logarithm of the odds ratio as the effect size (34) without adjusting for covariates.

P-value thresholds for genome-wide significance

The P-value thresholds for genome-wide significance were based on empirically-determined thresholds for different ancestral populations that account for the GWAS multiple testing burden, as well as population-specific linkage disequilibrium (LD) patterns (35):

(1) $P < 3.24 \times 10^{-8}$ for SNPs ascertained in African ancestry populations

(2) $P < 5.0 \times 10^{-8}$ for SNPs ascertained in European ancestry populations

(3) $P < 3.24 \times 10^{-8}$ for SNPs ascertained in multiethnic meta-analyses

We further corrected these thresholds for additional multiple testing from examination of four case-control definitions, each with and without covariate incorporation, for eight tests total. This yielded the following P-value thresholds for our study:

(4) $P < 3.75 \times 10^{-9}$ for SNPs ascertained in African ancestry populations

(5) $P < 6.25 \times 10^{-9}$ for SNPs ascertained in European ancestry populations

(6) $P < 3.75 \times 10^{-9}$ for SNPs ascertained in multiethnic meta-analyses

We note that correction for eight tests is conservative, because the case-control definitions are not completely independent. We did not apply further multiple testing correction for the different ancestries analyzed.

Replication Meta-Analysis

Eight European, eight Asian and four Hispanic replications cohorts provided summary statistics on SNPs with $P < 1 \times 10^{-5}$ in the discovery analyses (Table 3). Their phenotyping/genotyping protocols have been previously described and details are in the Supplemental Material. (6-8, 12, 13, 17, 18) The rationale for including additional ethnicities in the replication phase is that high trans-ethnic genetic correlations have been documented for type 2 diabetes and other traits/diseases and support the use of multiethnic studies to increase sample size.(36) Supplemental Table 3 summarizes the replication cohorts' mean values for HbA_{1C}, FPG, and DoD. Replication was *in silico* with existing genotyping. LT modeling was not applied to the replication cohort analyses. The replication cohorts used standard covariate adjustment in their regression models. Replication meta-analysis was also performed using inverse-variance weighting – first individually by each ethnicity (Europeans, Hispanics, Asians) followed by all cohorts combined. Replicated genome-wide significance had to meet the aforementioned thresholds after meta-analysis of the discovery and replication results.

Protein-Protein Interaction Analysis of Top GWAS Loci

To identify significantly-enriched protein networks among the loci with the highest statistical evidence for association to DR, we applied the Disease Association Protein-Protein Link Evaluator (DAPPLE) to our discovery GWAS.(37) It has been shown that top associated loci, despite not being genome-wide significant, tend to cluster in biological networks.(37, 38) For this reason, we examined the top 1000 loci from the discovery GWAS in the two mono-ethnic analyses (European and African American) and for each of the four case-control definitions analyses which incorporated DoD and glycemic control (eight network analyses in total). Our threshold for significance was therefore $P < 0.00625$ (0.05 corrected for eight tests). We used the publically available version of DAPPLE, and the protocol is outlined in the Supplemental Materials. This methodology has been used successfully with previous GWAS to identify protein networks with biological relevance.(37-39)

Gene Set Enrichment Analysis of DAPPLE significant genes

To further support the protein-protein interaction results from the DAPPLE analysis, we applied gene set enrichment analysis (GSEA) using MAGENTA (Meta-Analysis Gene-Set Enrichment of variaNT Associations) (40) to the set of genes significantly enriched for protein-protein interactions in the DAPPLE analysis (details in Supplemental Materials).

Type 2 diabetes and Associated Glycemic Traits Loci

To understand to what extent genetic determination of DR might reflect enrichment for type 2 diabetes or glycemic control genes, we computed a correlation between case status in the any DR analysis and the sum of the beta*risk allele (for quantitative glycemic traits) or logOR*risk allele

(for type 2 diabetes) of the trait-associated SNPs for each cohort and each trait (see Supplemental Materials for details).

Results

Discovery Meta-Analysis

Supplemental Figure 2 shows the PC analysis. We observed little inflation in the association statistic distribution (Supplemental Figure 3), indicating no significant population stratification as a confounder. Supplemental Figure 4 shows the Manhattan plots for the any DR analyses. Supplemental Tables 4 - 25 show the top 10 SNPs for independent loci with the lowest P values for each discovery analysis, including the sensitivity analyses (full results available at https://www.ncbi.nlm.nih.gov/gap).

Table 4 shows SNPs that met the traditional nominal threshold for genome wide significance of $P < 5 \times 10^{-8}$ from the discovery analyses. All of the SNPs in Table 4 were either from the PDR or extremes of DR analyses; Figure 1 shows the QQ and Manhattan plots for the PDR and extremes of DR analyses. The results for the associations in Table 4 are shown for each cohort separately in Supplemental Table 26. Results for these SNPs after meta-analysis with replication samples both combined and separated by ethnicity are shown in Table 5 and Supplemental Table 27, respectively.

Genome-Wide Significant Finding from the Discovery Analyses in NVL Gene

Using the corrected significance thresholds, only one SNP in the discovery meta-analyses met genome-wide significance: rs142293996 for the extremes of DR analysis incorporating DoD and glycemic control in Europeans ($P = 2.1 \times 10^{-9}$). The association was not significant without adjusting for covariates based on a Fisher's exact test (Supplemental Table 28). This is an intronic variant in the nuclear VCP-like (*NVL*) gene which encodes a member of the AAA (ATPases associated with diverse cellular activities) superfamily.(41) The *NVL* gene is widely expressed *in vivo* with highest expression in retina (https://www.proteinatlas.org/ENSG00000143748- NVL/tissue#top).

We tested whether this association was a significant *cis*-expression quantitative trait locus (eQTL) in the Genotype-Tissue Expression (GTEx) Project release v7 (see Supplemental Materials for eQTL analysis details). This variant, rs142293996, lies in the 22nd intron of *NVL* and is in LD (r^2=0.62) with variant rs41271487 in the 24th intron of *NVL*. Rs41271487 is a significant eQTL ($P = 6.4 \times 10^{-6}$, effect size=1.27) in the GTEx spinal cord cervical c-1 tissue, targeting calpain 2 (*CAPN2*), a calcium-activated neutral protease (Supplemental Figure 5). Common variants in the intron or regulatory region of *CAPN2*, 527-576 kb upstream of the DR association, are associated with variation in serum alpha-carotene levels (42), a vitamin A precursor required for sight, supporting a functional role for this gene. Based on the eQTL analysis, increased expression of *CAPN2* is associated with decreased risk of DR (Supplemental Figure 6). *CAPN2* is expressed in the retina (https://www.proteinatlas.org/ENSG00000162909-CAPN2/tissue).

When examined in the replication analyses (which included a more diverse population), the direction of effect in the replication cohorts for rs142293996 was the same but the meta-analysis P-value was not genome-wide significant ($P = 4.10 \times 10^{-6}$).

Top Finding from the African American Discovery Analyses

In African Americans, the SNP with the lowest P value was rs115523882 from the PDR analysis ($P = 5.37 \times 10^{-9}$). This was short of the 3.75×10^{-9} threshold for significance in African Americans. We could not reproduce this finding in the replication cohorts. This variant is located near the *GOLIM4* gene, which helps process proteins and mediates protein transport. The SNP rs115523882 specifically changes a motif which is a binding site for Nlx3, a transcription factor in blood, suggesting it plays a regulatory role. This variant is mainly present in people of African ancestry [minor allele frequency (MAF) = 0.0393] and not common in other ethnic groups, suggesting we may have had insufficient power to replicate it.

Of note, there was one SNP, rs184340784, suggestively associated with DR ($P = 3.52 \times 10^{-8}$) in the extremes of DR analysis without covariates in African Americans that was not present in our replication cohorts (due to low MAF) and thus could not be replicated. Neither rs115523882 nor rs184340784 were analyzed for eQTL activity in GTEx due to their low MAF (MAF<0.01 in GTEx tissues).

Table 6 and Supplemental Table 29 show the discovery variants with $P < 1 \times 10^{-5}$ that achieved a nominal $P < 0.05$ in the complete replication sample or in one of the replication ethnicities,

respectively, and had the same direction as the discovery samples. None of these variants achieved genome-wide significance after discovery and replication meta-analysis, as defined above.

DAPPLE Results Protein-Protein Interactions

One protein network from the African American PDR analysis was significant (P=0.0009) for average binding degree within the network (Figure 2). The aforementioned top ranked SNP (rs115523882) could not be included in the DAPPLE analysis since its nearby gene (*GOLIM4*) is not in the protein database. The significant protein network includes genes with primary roles in inflammation including *IFNG*, *IL22RA1*, *CFH* and *SELL*. *INFG* encodes INF-γ which is highly expressed in ocular tissues from PDR patients.(43) *IL22RA1* encodes the IL-22 receptor and *CFH* encodes complement factor H; both proteins are suspected to play a role in PDR.(44, 45) *SELL* encodes L-selectin, which is expressed at higher levels in lymphocytes from DR patients and associated with increased endothelial adhesion.(46) We did not identify any statistically significant protein networks for any of the other case-control definitions in African Americans or in Europeans.

MAGENTA Confirmation of DAPPLE Results

We examined the 41 genes in the significant network identified by the DAPPLE analysis via GSEA using MAGENTA. The genes showed a significant (16.5-fold) enrichment of low association P-values in the African American PDR analysis ($P < 1 \times 10^{-6}$; Supplemental Figure 7 and Supplemental Table 30) and to a lesser extent in African American extremes of DR analysis ($P = 2 \times 10^{-4}$; Supplemental Table 30), suggesting new DR associations of modest effects in

African Americans (Supplemental Table 31). No significant gene set enrichment was found for the PDR and extremes of DR analyses in Europeans.

Loci Associated with Type 2 Diabetes and Glycemic Traits

The results of the correlation analysis between type 2 diabetes/glycemic trait-associated SNPs and DR case status are shown in Supplemental Table 32. The Z-score for type 2 diabetes was +2.256 (P=0.024). The correlation coefficient R was positive, indicating that a greater burden of SNPs that increase type 2 diabetes risk is correlated with having DR. However, this Z-score was not significant after correcting for the six hypotheses (six traits) tested.

Previously associated SNPs from Prior Studies

We extracted results from our discovery meta-analysis for the variants with the lowest association P-values from previously published DR GWAS or large candidate gene studies (Supplemental Table 33). There were three variants that were nominally significant ($P < 0.05$) in our sample and had the same direction of effect as the previously published studies. Two of the variants—rs9896052 and rs6128—were from previous studies whose samples overlapped with some samples in our discovery meta-analysis, and therefore do not represent independent replication.(10, 20) Variant rs1399634, originally found in Chinese patients ($P = 2 \times 10^{-6}$), was nominally significant in our European discovery cohort ($P = 0.0124$). Meta-analysis of the original study and our cohorts was performed using the same method as our discovery and replication meta-analyses and was short of genome-wide significance ($OR = 1.47$, $P = 9.63 \times 10^{-8}$).

Discussion

To our knowledge, this study represents the largest GWAS performed for DR. The discovery analysis included 3,246 Europeans and 2,611 African Americans. The replication analysis included 18,545 Europeans, 16,453 Asians, and 2,710 Hispanics. Despite the relatively large sample size, we did not identify any individual variants that were associated at a genome-wide significant level after meta-analysis with multiethnic replication cohorts. However, among the most significant results in the African American PDR analysis, we did identify a statistically significant enrichment for a network of genes using DAPPLE which was corroborated by GSEA using MAGENTA.

In the discovery meta-analyses, several variants from the PDR and extremes of DR analyses achieved nominal genome-wide significance of $P < 5 \times 10^{-8}$, but the only variant to achieve genome-wide significance after conservative multiple testing correction was rs142293996 in the European analysis for extremes of DR ($P = 2.1 \times 10^{-9}$). It is notable that the variants with the most significant findings came from the two case-control definitions that have PDR as their case definition. This is consistent with the fact that PDR has a higher heritability than overall DR.(4) While the most strongly associated variants in the discovery analyses (rs142293996 in *NVL* in Europeans and rs115523882 in *GOLIM4* in African Americans) did not reach genome-wide significance with replication, it is still possible that they do play a role in DR pathogenesis. *NVL* is highly expressed in the retina and the implicated variant is in LD with an eQTL acting on *CAPN2* with functional implications in neural tissue. The eQTL variant falls in a binding site of a transcription factor.(47) The *GOLIM4* variant also has a known regulatory role.

We could not replicate the association with rs142293996 when we used the Fisher's exact test, although the Fisher's exact test did not allow for covariate incorporation. There is potential for inflated false positive rate when standard association methods are applied to rare (e.g. MAF < 1%) variants in imbalanced (e.g. case fraction < 10%) case-control cohorts at modest sample sizes.(48) However, most cohorts in this study did not have case fraction <10%. Larger sample sizes will help determine the confidence in these top associations.

There was one variant suggestively associated in the extremes of DR discovery analysis in African Americans, rs184340784, which was not present in any replication datasets. The T allele of this variant has a frequency of 0.0023 in African populations and 0 in European, East Asian, South Asian and Hispanic populations in the 1000 Genomes Phase 3 panel. In the discovery analysis, the $P = 3.52 \times 10^{-8}$ was shy of the genome-wide significance threshold of 3.75×10^{-9} for variants discovered from the African ancestry analyses. This variant is within an intronic region upstream of adherens junctions associated protein 1 (*AJAP1*) which has its highest expression in brain frontal cortex but is also expressed in the retina (https://www.proteinatlas.org/ENSG00000196581-AJAP1/tissue).

In the DAPPLE analysis, we did find that the top signals for the PDR analyses in African Americans analysis were enriched for a biologic network. The advantage of DAPPLE is that it can identify a protein pathway which may not be evident solely from the primary individual variant GWAS. The presence of an underlying network amongst the top loci suggests there are likely true associations within top findings that have yet to reach genome-wide significance due to limited power. Multiple pathways including inflammatory pathways are implicated by this

network. To confirm biological significance, these results will need to be followed up with functional *in vitro* studies.

The DAPPLE results were corroborated by the MAGENTA GSEA in the African American PDR and extremes of DR analyses. This network of genes, however, was not enriched for in Europeans. This could either be due to technical differences, e.g., the number of African American cases is ~3-fold larger than the number of European cases, or to biological reasons. For example, we found that the allele frequencies of the most significant variant per gene for 40% of these protein interacting genes are rare in Europeans (MAF < 0.2%), while common in African Americans (MAF > 1%), according to the Genome Aggregation Database (gnomAD, see Web Resources).

In the analysis between type 2 diabetes/glycemic trait SNPs and DR case status, only type 2 diabetes variants were significantly associated with DR prior to, but not after, multiple testing correction. One previous study examined aggregate effects of 76 type 2 diabetes-associated variants in Asian patients.(49) Participants in the top tertile of type 2 diabetes-risk score were 2.56-fold more likely to have DR compared with lowest tertile participants. Our study's result showed the same direction of effect as the prior study, with type 2 diabetes risk raising alleles increasing DR risk. The prior study did not examine glycemic traits. Our inability to detect a correlation for glycemic traits may be due to the small amount of glycemic variance captured by these variants. In European patients, HbA_{1C} SNPs explain approximately 5% of HbA_{1C} variance.(50)

We were unable to replicate findings from previous studies.(6-8, 12, 13, 17, 18) We did have the same direction of effect in our European discovery sample for rs1399634 *(LRP2)* which was initially reported in an Asian population. However, the meta-analysis was shy of genome-wide significance. The overall lack of replication of previous reports' findings is not surprising, given the heterogeneity in phenotyping, case-control definitions, ethnicities and analytic approaches, although we did try to match our case-control definitions to the original studies' definitions.

There are many potential reasons why we were unable to identify replicable, significant associations from our discovery GWAS. First, the genetic risk in DR development may be quite small in proportion to the non-genetic risk factors. Therefore, even though we assembled the largest sample, it may not be sufficient to detect very modest effects. There was heterogeneity between the discovery and replication cohorts that could contribute to inability to replicate. The discovery cohort included individuals with type 2 diabetes while the replication cohorts included individuals with either type 1 or type 2 diabetes. It is not known definitively whether genetic variants for DR differ between type 1 and type 2 diabetes. Clinically, DR phenotypes are similar in patients with type 1 and type 2 diabetes, so we hypothesize that at least some of the genetic risk is shared. However, we cannot be certain of this and heterogeneity of diabetes type might have contributed to lack of replication. The discovery cohort included individuals who were of either European or African American descent while the replication cohorts included individuals of European, Hispanic, or Asian descent. This heterogeneity could also have led to lack of replication. Europeans were represented in both the discovery and replication phases, but even our European discovery analysis has limited power. Power calculations show that our discovery GWAS for the any DR analysis in Europeans had 100% power to detect a variant with a MAF of

0.40 with a heterozygous genotypic relative risk (GRR) of 1.5 with a P-value < 5 X 10^{-8}, whereas the power decreases to 5% for the same variant with GRR of 1.2.

We attempted to harmonize the phenotypes as much as possible, but there were some limits to complete harmonization, particularly for cohorts with limited-field or no photography. Misclassification of participants because of limited DR ascertainment could have biased the results to the null. Although we did use LTSCORE modeling to account for DoD, we may have had some misclassification bias because we did not have a minimum DoD for controls – i.e. some controls could have developed DR with longer DoD - which would also bias our result towards the null. We only had one HbA_{1c} measure. Repeated HbA_{1c} measures would reflect long-term glycemia more accurately.

In summary, we have executed the largest GWAS of DR to date. There were no genome-wide significant findings but analysis of protein-protein interaction networks point to possible candidate pathways for PDR in African Americans. Future studies examining DR genetics would benefit from a greater international collaboration encompassing larger samples that would allow strict case-control definitions that define a minimal DoD without sacrificing power. Furthermore, these studies should focus case definitions on the advanced forms of DR—PDR and diabetic macular edema (DME)—and incorporate more refined phenotyping, particularly optical coherence tomography for DME. Finally, whole genome sequencing might reveal a role for very rare variants, particularly for the DR phenotypic extremes.

Acknowledgments

Author Contributions

SP, EJR, AVS, SD, LKS, AP, LS contributed to the writing of the manuscript. RPI, RAJ, MC, XL, C-YC, MCYN, AVS, GST, Y-DIC, JZK, LMD, WM, SMH, MI, DN, JK, YH, YJ, JA, AL, KS, KHP, XG, EI, KDT, SGA, JRS, BIF, I-TL, WH-HS, MK, AT, SH, MM, D-AT, RM-C, RV, MIM, LG, EA, VL, EA, AM, ASFD, HMC, IT, NS, P-HG, SM, CLH, AP, CJC, HH, PM, JEC, EYC, ADP, MAG, CP, DWB, BLY, DS, MFC, JJW, KPB, TYW, BEKK, RK, JIR, SKI reviewed and edited the manuscript. SP, RPI, RAJ, C-YC, MCYN, AVS, GT, Y-DIC, JZK, WM, MH, MI, JK, JA, AL, KS, KHP, XG, BIF, IT, NS, P-HG, SH, CLH, MM, D-AT, RM-C, SM, AP, CJC, HH, PM, JEC, EC, AP, MAG, CP, DWB, BLY, DS, MFC, JJW, KPB, TYW, BEKK, RK, JIR, SI, AP, LS collected and researched data. SP, RPI, RAJ, MC, XL, C-YC, MCYN, AVS, EJR, AS, SD, GT, Y-DIC, JZK, LMD, LKS, WM, MH, MI, DN, JK, YH, YJ, JA, AL, KS, IT, NS, SH, MM, KPB, BLY, AP, LS performed the analysis.

Guarantor Statement

LS is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of Interest

P-HG has received investigator-initiated research grants from Eli Lilly and Roche, is an advisory board member for AbbVie, AstraZeneca, Boehringer Ingelheim, Cebix, Eli Lilly, Janssen, Medscape, Merck Sharp & Dohme, Novartis, Novo Nordisk and Sanofi; and has received lecture fees from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Elo Water, Genzyme, Merck Sharp &

Dohme, Medscape, Novo Nordisk and Sanofi. BLY is a full-time employee of Genentech Inc. and holds stock and stock options in Roche. JZK is employed by Sun Pharmaceutical Industries, Inc.; however, the current employer is not in any way involved in this study. All other authors declare no conflicts of interest.

Financial Support

We gratefully acknowledge support from the following organizations for this research: Research to Prevent Blindness, Inc., New York; National Eye Institute (EY16335; EY22302; EY11753; R01 EY023644; Core Grant EY001792); Massachusetts Lions Eye Research Fund; Alcon Research Institute; American Diabetes Association (1-11-CT-51); Harvard Catalyst.

The Age, Gene, Environment, Susceptibility - Reykjavik Study (AGES) was supported by the U.S. National Institutes of Health (NIH) through the Intramural Research Program of the National Institute of Aging (ZIAAG007380) and the National Eye Institute (ZIAEY00401), NIH contract number N01-AG-1-2100, Hjartavernd (the Icelandic Heart Association), the Althingi (Icelandic Parliament), and the University of Iceland Research Fund. We are indebted to the staff at the Icelandic Heart Association and to the AGES participants who volunteered their time and allowed us to contribute their data to this international project. The funders had no role in collection, management, analysis or interpretation of data nor were funders involved in the preparation, writing, or approval of the article, or the decision to submit the article for publication.

The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The ARIC study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419).

The Australian Genetics of Diabetic Retinopathy Study was supported by the National Health and Medical Research Council (NHMRC) of Australia [no. 595918] and the Ophthalmic Research Institute of Australia. K.P.B is supported by a Senior Research Fellowship from the NHMRC and J.E.C by a Practitioner Fellowship from the NHMRC.

The Blue Mountains Eye Study (BMES) was supported by the Australian National Health & Medical Research Council (NHMRC), Canberra Australia (NHMRC project grant IDs 974159,

211069, 302068, and Centre for Clinical Research Excellence in Translational Clinical Research in Eye Diseases, CCRE in TCR-Eye, grant ID 529923). The BMES GWAS and genotyping costs was supported by Australian NHMRC, Canberra Australia (NHMRC project grant IDs 512423, 475604 and 529912), and the Wellcome Trust, UK as part of Wellcome Trust Case Control Consortium 2 (IDs 085475/B/08/Z and 085475/08/Z).

The Cardiovascular Health Study was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, N01HC75150, and grants U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Subjects included in the present analysis consented to the use of their genetic information.

The authors of this manuscript would like to thank all the patients recruited in the GoDARTS and in other European and African American cohorts. The authors are especially grateful to the Health Informatics Centre in the School of Medicine, University of Dundee, for their help with data access.

This work was supported by Tenovus Scotland (2015 T15/40 to W.M.). The GoDARTS project was supported by Chief Scientist Office Scotland and Diabetes UK. The genotyping costs were granted by Wellcome Trust for WTCCC2 samples and by the Innovative Medicines Initiative for SUMMIT samples.

The SUMMIT consortium was supported by the European Union's Seventh Framework Program (FP7/2007-2013) for the Innovative Medicine Initiative under grant agreement IMI/115006 (the SUMMIT consortium). FinnDiane was supported by grants from the Folkhälsan Research Foundation, the Wilhelm and Else Stockmann Foundation, the Liv och Hälsa Foundation, Helsinki University Central Hospital Research Funds (EVO), the Novo Nordisk Foundation (NNF14SA0003), and the Academy of Finland (134379, and 275614, and 299200).

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201300049C and HHSN268201300050C), Tougaloo College (HHSN268201300048C), and the University of Mississippi Medical Center (HHSN268201300046C and HHSN268201300047C) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

The Multiethnic Study of Atherosclerosis (MESA) and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001881, and DK063491. Additional funding provided by the Intramural Research Program of the National Eye Institute (ZIAEY000403). Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

The Singapore Epidemiology of Eye Diseases study was supported by the National Medical Research Council, Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council, Singapore (08/1/35/19/550 and 09/1/35/19/616). C.Y.C is supported by an award from NMRC (CSA-SI/0012/2017). The funding organization had no role in the design or conduct of this research.

The Starr County Health Studies were supported, in part, by the State of Texas and EY012386, DK047487, and DK073541 from the National Institutes of Health.

The Korean Study of Diabetic Retinopathy was supported by by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (No NRF-2017R1A2B2011436 and NRF-2012R1A1A2008943).

For Wake Forest School of Medicine study (WFU), genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSC268200782096C. This work was supported by National Institutes of Health grants R01 DK087914, R01 DK066358, R01 DK053591, DK070941, DK084149 and by the Wake Forest School of Medicine grant M01 RR07122 and Venture Fund.

This paper was supported in part by the Genetics of Latinos Diabetic Retinopathy (GOLDR) Study grant EY14684.

This study was supported by the National Eye Institute of the National Institutes of Health (EY014684 to J.I.R. and Y.-D.I.C.) and ARRA Supplement (EY014684-03S1, -04S1), the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center, the Eye Birth Defects Foundation Inc., the National Science Council, Taiwan (NSC 98-2314-B-075A-002-MY3 to W.H.S.) and the Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1003001C to W.H.S .).

The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant, UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Research Center.

The FIND study was supported by grants U01DK57292, U01DK57329, U01DK057300, U01DK057298, U01DK057249, U01DK57295, U01DK070657, U01DK057303, and U01DK57304 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and, in part, by the Intramural Research Program of the NIDDK. Support was also received from the National Heart, Lung and Blood Institute grants U01HL065520, U01HL041654, and U01HL041652. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health (NIH), under contract N01-CO-12400 and the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. This work was also supported by the National Center for Research Resources for the General Clinical Research Center grants: Case Western Reserve University, M01-RR-000080; Wake Forest University, M01-RR-07122; Harbor-University of California, Los Angeles Medical Center, M01-RR-00425; College of Medicine, University of California, Irvine, M01-RR-00827–29; University of New Mexico, HSC M01-RR-00997; and Frederic C. Bartter, M01-RR-01346. Computing resources were provided, in part, by the Wake Forest School of Medicine Center for Public Health Genomics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

MMcC is a Wellcome Senior Investigator supported by Wellcome awards 090532, 106130, 098381 and 203141. MMcC is also a National Institute for Health Research (NIHR) Senior Investigator. The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

References

1. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, Chen SJ, Dekker JM, Fletcher A, Grauslund J, Haffner S, Hamman RF, Ikram MK, Kayama T, Klein BE, Klein R, Krishnaiah S, Mayurasakorn K, O'Hare JP, Orchard TJ, Porta M, Rema M, Roy MS, Sharma T, Shaw J, Taylor H, Tielsch JM, Varma R, Wang JJ, Wang N, West S, Xu L, Yasuda M, Zhang X, Mitchell P, Wong TY. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35:556-64

2. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin epidemiologic study of diabetic retinopathy. III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. Arch Ophthalmol. 1984;102:527-32

3. Looker HC, Nelson RG, Chew E, Klein R, Klein BE, Knowler WC, Hanson RL. Genome-wide linkage analyses to identify Loci for diabetic retinopathy. Diabetes. 2007;56:1160-6

4. Arar NH, Freedman BI, Adler SG, Iyengar SK, Chew EY, Davis MD, Satko SG, Bowden DW, Duggirala R, Elston RC, Guo X, Hanson RL, Igo RP, Jr., Ipp E, Kimmel PL, Knowler WC, Molineros J, Nelson RG, Pahl MV, Quade SR, Rasooly RS, Rotter JI, Saad MF, Scavini M, Schelling JR, Sedor JR, Shah VO, Zager PG, Abboud HE, Family Investigation of N, Diabetes Research G. Heritability of the severity of diabetic retinopathy: the FIND-Eye study. Invest Ophthalmol Vis Sci. 2008;49:3839-45

5. Sheu WH, Kuo JZ, Lee IT, Hung YJ, Lee WJ, Tsai HY, Wang JS, Goodarzi MO, Klein R, Klein BE, Ipp E, Lin SY, Guo X, Hsieh CH, Taylor KD, Fu CP, Rotter JI, Chen YD.

Genome-wide association study in a Chinese population with diabetic retinopathy. Hum Mol Genet. 2013;22:3165-73

6. Awata T, Yamashita H, Kurihara S, Morita-Ohkubo T, Miyashita Y, Katayama S, Mori K, Yoneya S, Kohda M, Okazaki Y, Maruyama T, Shimada A, Yasuda K, Nishida N, Tokunaga K, Koike A. A genome-wide association study for diabetic retinopathy in a Japanese population: potential association with a long intergenic non-coding RNA. PLoS One. 2014;9:e111715

7. Fu YP, Hallman DM, Gonzalez VH, Klein BE, Klein R, Hayes MG, Cox NJ, Bell GI, Hanis CL. Identification of Diabetic Retinopathy Genes through a Genome-Wide Association Study among Mexican-Americans from Starr County, Texas. J Ophthalmol. 2010; 2010.

8. Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL. Genome-wide meta-analysis for severe diabetic retinopathy. Hum Mol Genet. 2011;20:2472-81

9. Huang YC, Lin JM, Lin HJ, Chen CC, Chen SY, Tsai CH, Tsai FJ. Genome-wide association study of diabetic retinopathy in a Taiwanese population. Ophthalmology. 2011;118:642-8

10. Burdon KP, Fogarty RD, Shen W, Abhary S, Kaidonis G, Appukuttan B, Hewitt AW, Sharma S, Daniell M, Essex RW, Chang JH, Klebe S, Lake SR, Pal B, Jenkins A, Govindarjan G, Sundaresan P, Lamoureux EL, Ramasamy K, Pefkianaki M, Hykin PG, Petrovsky N, Brown MA, Gillies MC, Craig JE. Genome-wide association study for sight-threatening diabetic retinopathy reveals association with genetic variation near the GRB2 gene. Diabetologia. 2015;58:2288-97

11. Shtir C, Aldahmesh MA, Al-Dahmash S, Abboud E, Alkuraya H, Abouammoh MA, Nowailaty SR, Al-Thubaiti G, Naim EA, B AL, Binhumaid FS, AB AL, Altamimi AS, Alamer FH, Hashem M, Abouelhoda M, Monies D, Alkuraya FS. Exome-based case-control association

study using extreme phenotype design reveals novel candidates with protective effect in diabetic retinopathy. Hum Genet. 2016;135:193-200

12. Hosseini SM, Boright AP, Sun L, Canty AJ, Bull SB, Klein BE, Klein R, Paterson AD. The association of previously reported polymorphisms for microvascular complications in a meta-analysis of diabetic retinopathy. Hum Genet. 2015;134:247-57

13. Grassi MA, Tikhomirov A, Ramalingam S, Lee KE, Hosseini SM, Klein BE, Klein R, Lussier YA, Cox NJ, Nicolae DL. Replication analysis for severe diabetic retinopathy. Invest Ophthalmol Vis Sci. 2012;53:2377-81

14. Peng D, Wang J, Zhang R, Jiang F, Tang S, Chen M, Yan J, Sun X, Wang S, Wang T, Yan D, Bao Y, Hu C, Jia W. Common variants in or near ZNRF1, COLEC12, SCYL1BP1 and API5 are associated with diabetic retinopathy in Chinese patients with type 2 diabetes. Diabetologia. 2015;58:1231-8

15. Cheung CY, Hui EY, Lee CH, Kwok KH, Gangwani RA, Li KK, Chan JC, Woo YC, Chow WS, Yuen MM, Wong RL, Fong CH, Xu A, Wong DS, Sham PC, Lam KS. Impact of Genetic Loci Identified in Genome-Wide Association Studies on Diabetic Retinopathy in Chinese Patients With Type 2 Diabetes. Invest Ophthalmol Vis Sci. 2016;57:5518-24

16. Zaitlen N, Lindstrom S, Pasaniuc B, Cornelis M, Genovese G, Pollack S, Barton A, Bickeboller H, Bowden DW, Eyre S, Freedman BI, Friedman DJ, Field JK, Groop L, Haugen A, Heinrich J, Henderson BE, Hicks PJ, Hocking LJ, Kolonel LN, Landi MT, Langefeld CD, Le Marchand L, Meister M, Morgan AW, Raji OY, Risch A, Rosenberger A, Scherf D, Steer S, Walshaw M, Waters KM, Wilson AG, Wordsworth P, Zienolddiny S, Tchetgen ET, Haiman C, Hunter DJ, Plenge RM, Worthington J, Christiani DC, Schaumberg DA, Chasman DI, Altshuler

D, Voight B, Kraft P, Patterson N, Price AL. Informed conditioning on clinical covariates increases power in case-control association studies. PLoS genetics. 2012;8:e1003032

17. Grassi MA, Mazzulla DA, Knudtson MD, Huang WW, Lee KE, Klein BE, Nicolae DL, Klein R. Patient self-report of prior laser treatment reliably indicates presence of severe diabetic retinopathy. Am J Ophthalmol. 2009;147:501-4

18. Meng W, Shah KP, Pollack S, Toppila I, Hebert HL, McCarthy MI, Groop L, Ahlqvist E, Lyssenko V, Agardh E, Daniell M, Kaidonis G, Craig JE, Mitchell P, Liew G, Kifley A, Wang JJ, Christiansen MW, Jensen RA, Penman A, Hancock HA, Chen CJ, Correa A, Kuo JZ, Li X, Chen YI, Rotter JI, Klein R, Klein B, Wong TY, Morris AD, Doney ASF, Colhoun HM, Price AL, Burdon KP, Groop PH, Sandholm N, Grassi MA, Sobrin L, Palmer CNA, Wellcome Trust Case Control Consortium SmfM, Macro-vascular hard endpoints for Innovative diabetes Tools study g. 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. 2018

19. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care. 2003;26 Suppl 1:S5-20

20. Sobrin L, Green T, Sim X, Jensen RA, Tai ES, Tay WT, Wang JJ, Mitchell P, Sandholm N, Liu Y, Hietala K, Iyengar SK, Brooks M, Buraczynska M, Van Zuydam N, Smith AV, Gudnason V, Doney AS, Morris AD, Leese GP, Palmer CN, Swaroop A, Taylor HA, Jr., Wilson JG, Penman A, Chen CJ, Groop PH, Saw SM, Aung T, Klein BE, Rotter JI, Siscovick DS, Cotch MF, Klein R, Daly MJ, Wong TY. Candidate gene association study for diabetic retinopathy in persons with type 2 diabetes: the Candidate gene Association Resource (CARe). Invest Ophthalmol Vis Sci. 2011;52:7593-602

21. Wong TY, Klein R, Islam FM, Cotch MF, Folsom AR, Klein BE, Sharrett AR, Shea S. Diabetic retinopathy in a multi-ethnic cohort in the United States. Am J Ophthalmol. 2006;141:446-55

22. Klein R, Marino EK, Kuller LH, Polak JF, Tracy RP, Gottdiener JS, Burke GL, Hubbard LD, Boineau R. The relation of atherosclerotic cardiovascular disease to retinopathy in people with diabetes in the Cardiovascular Health Study. Br J Ophthalmol. 2002;86:84-90

23. Klein R, Sharrett AR, Klein BE, Moss SE, Folsom AR, Wong TY, Brancati FL, Hubbard LD, Couper D. The association of atherosclerosis, vascular risk factors, and retinopathy in adults with diabetes: the Atherosclerosis Risk in Communities Study. Ophthalmology. 2002;109:1225-34

24. Kaidonis G, Abhary S, Daniell M, Gillies M, Fogarty R, Petrovsky N, Jenkins A, Essex R, Chang JH, Pal B, Hewitt AW, Burdon KP, Craig JE. Genetic study of diabetic retinopathy: recruitment methodology and analysis of baseline characteristics. Clin Experiment Ophthalmol. 2014;42:486-93

25. Mitchell P, Smith W, Wang JJ, Attebo K. Prevalence of diabetic retinopathy in an older community. The Blue Mountains Eye Study. Ophthalmology. 1998;105:406-11

26. Gunnlaugsdottir E, Halldorsdottir S, Klein R, Eiriksdottir G, Klein BE, Benediktsson R, Harris TB, Launer LJ, Aspelund T, Gudnason V, Cotch MF, Jonasson F. Retinopathy in old persons with and without diabetes mellitus: the Age, Gene/Environment Susceptibility--Reykjavik Study (AGES-R). Diabetologia. 2012;55:671-80

27. Nguyen QD, Brown DM, Marcus DM, Boyer DS, Patel S, Feiner L, Gibson A, Sy J, Rundle AC, Hopkins JJ, Rubio RG, Ehrlich JS. Ranibizumab for diabetic macular edema: results from 2 phase III randomized trials: RISE and RIDE. Ophthalmology. 2012;119:789-801

28. Penman A, Hoadley S, Wilson JG, Taylor HA, Chen CJ, Sobrin L. P-selectin Plasma Levels and Genetic Variant Associated With Diabetic Retinopathy in African Americans. Am J Ophthalmol 2015;159:1152-60 e2

29. Kuo JZ, Guo X, Klein R, Klein BE, Cui J, Rotter JI, Ipp E, Chen YD. Systemic soluble tumor necrosis factor receptors 1 and 2 are associated with severity of diabetic retinopathy in Hispanics. Ophthalmology. 2012;119:1041-6

30. Grading diabetic retinopathy from stereoscopic color fundus photographs--an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. Ophthalmology. 1991;98:786-806

31. Armitage P. Tests for linear trends in proportions and frequencies. Biometrics. 1955;11:375-86

32. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38:904-9

33. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M,

Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008;40:638-45

34. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013;14:379-89

35. Kanai M, Tanaka T, Okada Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. J Hum Genet. 2016;61:861-6

36. Brown BC, Asian Genetic Epidemiology Network Type 2 Diabetes C, Ye CJ, Price AL, Zaitlen N. Transethnic Genetic-Correlation Estimates from Summary Statistics. Am J Hum Genet. 2016;99:76-88

37. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, International Inflammatory Bowel Disease Genetics C, Cotsapas C, Daly MJ. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011;7:e1001273

38. Lundby A, Rossin EJ, Steffensen AB, Acha MR, Newton-Cheh C, Pfeufer A, Lynch SN, Consortium QTIIG, Olesen SP, Brunak S, Ellinor PT, Jukema JW, Trompet S, Ford I, Macfarlane PW, Krijthe BP, Hofman A, Uitterlinden AG, Stricker BH, Nathoe HM, Spiering W, Daly MJ, Asselbergs FW, van der Harst P, Milan DJ, de Bakker PI, Lage K, Olsen JV. Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. Nat Methods. 2014;11:868-74

39. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfelser R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, Lewis L, Han Y, Voight BF, Lim E, Rossin E, Kirby A, Flannick J, Fromer M, Shakir K, Fennell T, Garimella K, Banks E, Poplin R, Gabriel S, DePristo M, Wimbish JR, Boone BE, Levy SE, Betancur C, Sunyaev S, Boerwinkle E, Buxbaum JD, Cook EH, Jr., Devlin B, Gibbs RA, Roeder K, Schellenberg GD, Sutcliffe JS, Daly MJ. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012;485:242-5

40. Segre AV, Wei N, Consortium D, Investigators M, Altshuler D, Florez JC. Pathways targeted by antidiabetes drugs are enriched for multiple genes associated with type 2 diabetes risk. Diabetes. 2015;64:1470-83

41. Her J, Chung IK. The AAA-ATPase NVL2 is a telomerase component essential for holoenzyme assembly. Biochem Biophys Res Commun. 2012;417:1086-92

42. D'Adamo CR, Dawson VJ, Ryan KA, Yerges-Armstrong LM, Semba RD, Steinle NI, Mitchell BD, Shuldiner AR, McArdle PF. The CAPN2/CAPN8 Locus on Chromosome 1q Is Associated with Variation in Serum Alpha-Carotene Concentrations. J Nutrigenet Nutrigenomics. 2016;9:254-64

43. Paine SK, Basu A, Mondal LK, Sen A, Choudhuri S, Chowdhury IH, Saha A, Bhadhuri G, Mukherjee A, Bhattacharya B. Association of vascular endothelial growth factor, transforming growth factor beta, and interferon gamma gene polymorphisms with proliferative diabetic retinopathy in patients with type 2 diabetes. Mol Vis. 2012;18:2749-57

44. Takeuchi M, Sato T, Tanaka A, Muraoka T, Taguchi M, Sakurai Y, Karasawa Y, Ito M. Elevated Levels of Cytokines Associated with Th2 and Th17 Cells in Vitreous Fluid of Proliferative Diabetic Retinopathy Patients. PLoS One. 2015;10:e0137358

45. Wang J, Yang MM, Li YB, Liu GD, Teng Y, Liu XM. Association of CFH and CFB gene polymorphisms with retinopathy in type 2 diabetic patients. Mediators Inflamm. 2013;2013:748435

46. MacKinnon JR, Knott RM, Forrester JV. Altered L-selectin expression in lymphocytes and increased adhesion to endothelium in patients with diabetic retinopathy. Br J Ophthalmol. 2004;88:1137-41

47. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40:D930-4

48. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, LeFaive J, VandeHaar P, Gagliano SA, Gifford A, Bastarache LA, Wei WQ, Denny JC, Lin M, Hveem K, Kang HM, Abecasis GR, Willer CJ, Lee S. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018;50:1335-41

49. Chong YH, Fan Q, Tham YC, Gan A, Tan SP, Tan G, Wang JJ, Mitchell P, Wong TY, Cheng CY. Type 2 Diabetes Genetic Variants and Risk of Diabetic Retinopathy. Ophthalmology. 2017;124:336-42

50. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, Chu AY, Zhang W, Wang X, Chen P, Maruthur NM, Porneala BC, Sharp SJ, Jia Y, Kabagambe EK, Chang LC, Chen WM, Elks CE, Evans DS, Fan Q, Giulianini F, Go MJ, Hottenga JJ, Hu Y, Jackson AU, Kanoni S, Kim YJ, Kleber ME, Ladenvall C, Lecoeur C, Lim

SH, Lu Y, Mahajan A, Marzi C, Nalls MA, Navarro P, Nolte IM, Rose LM, Rybin DV, Sanna S, Shi Y, Stram DO, Takeuchi F, Tan SP, van der Most PJ, Van Vliet-Ostaptchouk JV, Wong A, Yengo L, Zhao W, Goel A, Martinez Larrad MT, Radke D, Salo P, Tanaka T, van Iperen EPA, Abecasis G, Afaq S, Alizadeh BZ, Bertoni AG, Bonnefond A, Bottcher Y, Bottinger EP, Campbell H, Carlson OD, Chen CH, Cho YS, Garvey WT, Gieger C, Goodarzi MO, Grallert H, Hamsten A, Hartman CA, Herder C, Hsiung CA, Huang J, Igase M, Isono M, Katsuya T, Khor CC, Kiess W, Kohara K, Kovacs P, Lee J, Lee WJ, Lehne B, Li H, Liu J, Lobbens S, Luan J, Lyssenko V, Meitinger T, Miki T, Miljkovic I, Moon S, Mulas A, Muller G, Muller-Nurasyid M, Nagaraja R, Nauck M, Pankow JS, Polasek O, Prokopenko I, Ramos PS, Rasmussen-Torvik L, Rathmann W, Rich SS, Robertson NR, Roden M, Roussel R, Rudan I, Scott RA, Scott WR, Sennblad B, Siscovick DS, Strauch K, Sun L, Swertz M, Tajuddin SM, Taylor KD, Teo YY, Tham YC, Tonjes A, Wareham NJ, Willemsen G, Wilsgaard T, Hingorani AD, Consortium E-C, Consortium EP-I, Lifelines Cohort S, Egan J, Ferrucci L, Hovingh GK, Jula A, Kivimaki M, Kumari M, Njolstad I, Palmer CNA, Serrano Rios M, Stumvoll M, Watkins H, Aung T, Bluher M, Boehnke M, Boomsma DI, Bornstein SR, Chambers JC, Chasman DI, Chen YI, Chen YT, Cheng CY, Cucca F, de Geus EJC, Deloukas P, Evans MK, Fornage M, Friedlander Y, Froguel P, Groop L, Gross MD, Harris TB, Hayward C, Heng CK, Ingelsson E, Kato N, Kim BJ, Koh WP, Kooner JS, Korner A, Kuh D, Kuusisto J, Laakso M, Lin X, Liu Y, Loos RJF, Magnusson PKE, Marz W, McCarthy MI, Oldehinkel AJ, Ong KK, Pedersen NL, Pereira MA, Peters A, Ridker PM, Sabanayagam C, Sale M, Saleheen D, Saltevo J, Schwarz PE, Sheu WHH, Snieder H, Spector TD, Tabara Y, Tuomilehto J, van Dam RM, Wilson JG, Wilson JF, Wolffenbuttel BHR, Wong TY, Wu JY, Yuan JM, Zonderman AB, Soranzo N, Guo X, Roberts DJ, Florez JC, Sladek R, Dupuis J, Morris AP, Tai ES, Selvin E, Rotter JI, Langenberg C, Barroso I, Meigs JB.

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. 2017;14:e1002383

Table 1. Studies included in the discovery sample

Study	Population	Diabetes Type	# of Eyes/ # of Fields/ Size of Fields Photographed	Diabetes Duration	Glycemic Control Measure	Cases (ETDRS ≥ 14)	Ctrls (ETDRS < 14)	Cases (ETDRS ≥ 60)	Ctrls (ETDRS < 60)	Cases (ETDRS ≥ 30)
AAPDR	AA	2	2/7/30 deg.	Y	HbA_{1C}	274	56	255	75	261
AGES*	EUR	2	2/2/45 deg.	Y	HbA_{1C}	85	222	3	304	8
ARIC	AA	2	1/1/45 deg.	Y	HbA_{1C}	96	265	3	358	73
	EUR	2	1/1/45 deg.	Y	HbA_{1C}	126	632	6	752	80
AUST	EUR	2	NA‡	Y	HbA_{1C}	522	435	187	770	346
BMES	EUR	2	2/5/30 deg.	Y	FPG	124	208	1	331	37
CHS	AA	2	1/1/45 deg.	Y	FPG	19	35	4	50	14
	EUR	2	1/1/45 deg.	Y	FPG	26	119	4	141	16
FIND-Eye*	AA	2	2/2/45 deg.†	Y	HbA_{1C}	330	167	264	233	303
	EUR	2	2/2/45 deg.†	Y	HbA_{1C}	158	154	115	197	145
JHS	AA	2	2/7/30 deg.	Y	HbA_{1C}	91	160	12	239	57
MESA	AA	2	2/2/45 deg.	Y	HbA_{1C}	101	258	11	348	60
	EUR	2	2/2/45 deg.	Y	HbA_{1C}	38	200	2	236	12
RISE/RIDE	EUR	2	2/7/30 deg.	Y	HbA_{1C}	--	--	80	117	--
WFU	AA	2	NA‡	Y	HbA_{1C}	--	--	548	211	--
TOTAL	AA	2	--	Y	Varies	911	941	1097	1514	768
TOTAL	EUR	2	--	Y	Varies	1079	1970	398	2848	644

Ctrls= Controls, AAPDR = African American Proliferative Diabetic Retinopathy Study, AGES = Age, Gene/Environment Susceptibility Study, ARIC = Atherosclerosis Risk In Communities Study, AUST= Australian Genetics of Diabetic Retinopathy Study, BMES = Blue Mountains Eye Study, CHS=Cardiovascular Health Study, FIND-Eye = Family Study of Nephropathy and Diabetes-Eye, JHS = Jackson Heart Study, MESA = Multiethnic Study of Atherosclerosis, RIDE/RISE= Ranibizumab Injection in Subjects with Clinically Significant Macular Edema with Center Involvement Secondary to Diabetes, WFU=Wake Forest University, AA=African American, EUR = European, Illum=Illumina, Affy=Affymetrix, NA=not available, Y=information on diabetes duration is available, HbA_{1C}=hemoglobin A_{1C}, FPG=fasting plasma glucose, deg.= degrees, SNPs= single nucleotide polymorphisms, QC=quality control

* Cohorts without access to raw genotype information

† Not all FIND-Eye subjects had photographs but all participants had harmonization of exam and clinical data to an ETDRS score.
‡ The AUST study used examination by an ophthalmologist to ascertain diabetic retinopathy. The WFU study used a questionnaire to ascertain diabetic retinopathy.

Table 2. Four case-control definitions and the number of samples available for discovery for each definition.

	Controls			Cases		
Analysis Name	Score	*n* AA	*n* EUR	Score	*n* AA	*n* EUR
Any DR (Primary Analysis)	< 14	941	1970	≥ 14	911	1079
PDR	< 60	1514	2848	≥ 60	1097	398
NPDR	< 14	941	1970	≥ 30	768	644
Extremes of DR	< 14	941	1970	≥ 60	1097	398

DR= diabetic retinopathy, PDR = proliferative diabetic retinopathy, NPDR = non-proliferative diabetic retinopathy, Score = ETDRS score range, AA = African American, EUR= European

Table 3. Studies included in the replication meta-analyses

Cohort by Ancestry	Ethnicity/ Nationality	DM Type	Any DR Analysis		PDR Analysis		NPDR Analysis		Extremes of DR Analysis	
			Cases	Controls	Cases	Controls	Cases	Controls	Case	Controls
Asian										
KSDR	Korean	2	1516	571	918	1167	1300	571	918	571
MESA	Chinese	2	28	83	--	--	17	83	--	--
RIKEN	Japanese	2	5532	5565	--	--	2371	5565	--	--
SCES I	Chinese	2	75	228	--	--	--	--	--	--
SCES II	Chinese	2	27	78	--	--	--	--	--	--
SiMES	Malay	2	214	557	--	--	--	--	--	--
SINDI	Indian	2	315	669	--	--	--	--	--	--
TUDR	Chinese	2	--	--	--	--	--	--	436	559
European										
DCCT/EDIC Primary cohort	North American	1	--	--	53	598	--	--	--	--
DCCT/EDIC Secondary cohort, conventional treatment	North American	1	--	--	114	209	--	--	--	--
DCCT/EDIC Secondary cohort, intensive treatment	North American	1	--	--	42	288	--	--	--	--
GENESIS/GENEDIAB	French	1	277	999	808	468	277	607	277	468
GoDARTS	Scottish		2506	2412	574	4345	1381	2412	574	2412
GoKinD	North American	1	--	--	138	581	--	--	--	--
SUMMIT	European	1 and 2	5422	4302	--	--	--	--	--	--
WESDR	North American	1	--	--	309	294	--	--	--	--
Hispanic										
GOLDR	Hispanics	2	298	301	76	523	215	301	76	301

LALES	Hispanics	2	552	500	53	999	341	500	53	500
MESA	Hispanics	2	92	192	--	--	52	192	--	--
SCHS	Mexican Americans	2	528	247	103	672	406	247	103	247
Total			17382	16704	3188	10144	6360	10478	2437	5058

DM = diabetes mellitus, KSDR = Korean Study of Diabetic Retinopathy, MESA = Multiethnic Study of Atherosclerosis, RIKEN = Rikagaku Kenkyusho - Institute of Physical and Chemical Research, SCES= Singapore Chinese Eye Study, SiMES = Singapore Malay Eye Study, SINDI = Singapore Indian Eye Study, DCCT/EDIC = Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications, GENESIS/GENEDIAB=Genetics nephropathy and sib pair study/Genetics, nephropathy, diabetes, GoDARTS =Genetics of Diabetes and Audit Research Tayside Study, GoKinD = Genetics of Kidneys in Diabetes, SUMMIT = Surrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools, WESDR = Wisconsin Epidemiologic Study of Diabetic Retinopathy, GOLDR = Genetics of Latino Diabetic Retinopathy, LALES = Los Angeles Latino Eye Study, SCHS = Starr County Health Studies, TUDR = Taiwan–US Diabetic Retinopathy Study

The SUMMIT cohort is a meta-analysis of three European studies: The Finnish Diabetic Nephropathy (FinnDiane) Study; Scania Diabetes Registry; and the Eurodiab study.

Table 4. Variants with P < 5 x 10^{-8} (traditional, nominal threshold for genome wide significance) in the discovery analyses

							CASES		CONTROLS					
Case Control Definition	Population/ LT Modeling	RSID	CHR	Position	Nearest Gene	REF	N	RAF	N	RAF	NEFF	P	OR	95% CI
PDR	AA/no	rs115523882	3	167876205	*GOLIM4*	A	1105	0.9823	1119	0.9611	1452	9.42 X 10^{-9}	3.10	2.12, 4.53
PDR	AA/yes	rs115523882	3	167876205	*GOLIM4*	A	1105	0.9823	1119	0.9611	1452	5.37 X 10^{-9}	3.10	2.14, 4.50
PDR	EUR/no	rs139205645	2	201949806	*NDUFB3*	T	309	0.9725	975	0.9959	907	3.93 X 10^{-8}	0.13	0.06, 0.27
PDR	EUR/yes	rs17791488	17	26232732	*NOS2/LYRM9*	T	309	0.9871	975	0.9661	907	7.26 X 10^{-9}	3.70	2.40, 5.71
Extremes of DR	AA/no	rs184340784	1	4589883	*AJAP1*	C	520	0.999	230	0.9784	603	3.52 X 10^{-8}	NA	NA
Extremes of DR	EUR/yes	rs142293996	1	224448059	*NVL*	C	187	0.9947	435	0.9874	523	2.10 X 10^{-9}	2.38	1.80, 3.14
Extremes of DR	EUR/yes	rs17706958	3	73837141	*PDZRN3*	T	308	0.8139	594	0.7332	797	3.04 X 10^{-8}	1.58	1.35, 1.85
Extremes of DR	EUR/yes	rs80117617	2	40855125	*SLC8A1*	T	308	0.9838	594	0.9445	797	4.04 X 10^{-8}	3.78	2.37, 6.02

LT= liability threshold, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, RAF= reference allele frequency, OR= odds ratio for reference allele, CI = confidence interval, AA= African Americans, EUR = European

Table 5. Replication results for variants with P < 5 x 10^{-8} (traditional, nominal threshold for genome wide significance) in the discovery analysis

Discovery Population/ LT modeling	RSID	Nearest Gene	REF	Disc NEFF	Disc RAF	Disc P	Disc OR	All Rep NEFF	All Rep RAF	All Rep OR	All Rep P	Disc + REP OR (95% CI)	Disc + Rep P
Variants identified in the PDR Discovery Analysis													
AA/no	rs115523882	*GOLIM4*	A	1452	0.9721	9.42 X 10^{-9}	3.10	571	0.9975	0.20	0.13	2.89 (1.97, 4.23)	8.51 X 10^{-8}
AA/yes	rs115523882	*GOLIM4*	A	1452	0.9721	5.37 X 10^{-9}	3.10	571	0.9975	0.20	0.18	2.89 (1.99, 4.20)	4.25 X 10^{-8}
European/no	rs139205645	*NDUFB3*	T	907	0.9907	3.93 X 10^{-8}	0.13	3431	0.9900	0.74	0.77	0.48 (0.29, 0.79)	0.004
European/yes	rs17791488	*NOS2/LYRM9*	T	907	0.9705	7.26 X 10^{-9}	3.70	5883	0.9772	0.82	0.33	1.08 (0.98, 1.19)	0.12
Variants identified in the Extremes of DR Analysis													
AA/no	rs184340784	*AJAP1*	C	603	0.0063	3.52 X 10^{-8}	NA	*	*	*	*	--	--
European/yes	rs142293996	*NVL*	C	523	0.9895	2.10 X 10^{-9}	2.38	1229	0.9910	3.23	0.16	2.91 (1.85, 4.57)	4.10 x 10^{-6}
European/yes	rs17706958	*PDZRN3*	T	797	0.7615	3.04 X 10^{-8}	1.58	4194	0.9828	1.28	0.02	1.39 (1.24, 1.56)	7.41 x 10^{-8}
European/yes	rs80117617	*SLC8A1*	T	797	0.9598	4.04 X 10^{-8}	3.78	3345	0.9726	1.29	0.24	1.71 (1.30, 2.25)	1.35 x 10^{-4}

LT= liability threshold, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, RAF= reference allele frequency in sample, ALL= all replication cohorts, OR= odds ratio for reference allele, CI = confidence interval, PDR = proliferative diabetic retinopathy, DR = diabetic retinopathy, AA= African Americans

* None of the replication cohorts were able to provide data for this SNP.

Table 6. Replication results for variants with nominal significance (P <0.05) in the combined (Hispanic, African American, and European cohorts) replication meta-analyses

Discovery Population/ LT modeling	RSID	Nearest Gene	REF*	DISC EAF	DISC OR	DISC P	ALL REP OR	ALL REP P	DISC + REP OR	DISC + REP P
Variants identified in the Any DR Discovery Analysis										
European (Sens)/no	rs1394919	*PPEF2/NAAA*	C	0.72	0.73	8.51×10^{-6}	0.91	0.003	0.88	6.35×10^{-6}
AA (Sens)/no	rs75360147	*SLC28A3*	T	0.93	2.08	7.07×10^{-6}	2.65	0.009	2.17	2.29×10^{-7}
European/no	rs1508244	*HTR1E*	A	0.98	0.33	3.74×10^{-6}	0.92	0.01	0.90	0.002
ME/no	rs10432638	*UBXN2A*	C	0.73	0.78	2.60×10^{-6}	0.93	0.01	0.89	7.74×10^{-6}
EU/no	rs150775408	*BC031225*	C	0.95	1.97	7.24×10^{-6}	1.27	0.04	1.46	2.54×10^{-5}
AA/yes	rs143894698	*GCM1*	G	0.98	3.14	4.62×10^{-6}	1.45	0.004	1.58	2.53×10^{-5}
European/yes	rs13006587	*ATAD2B*	G	0.58	0.79	7.52×10^{-6}	0.93	0.006	0.92	4.74×10^{-5}
European/yes	rs73642012	*PTPRD*	C	0.91	0.67	9.58×10^{-6}	0.90	0.02	0.87	8.67×10^{-5}
Variants identified in the PDR Discovery Analysis										
Europeans/no	rs139921826	*PRSS35*	G	0.98	0.33	7.92×10^{-6}	0.66	0.03	0.62	0.0008
AA/yes	rs1414474	*C1orf94*	C	0.14	1.62	1.46×10^{-7}	1.12	0.01	1.19	1.90×10^{-5}
AA/yes	rs9998354	*BTF3P13*	T	0.44	0.73	8.74×10^{-6}	0.92	0.04	0.87	0.0001
European/yes	rs142293996	*NVL*	C	0.99	1.83	1.14×10^{-6}	2.40	0.04	2.29	0.0001
Variants identified in the NPDR Discovery Analysis										
European/no	rs1508244	*RN7SL643P*	A	0.98	0.32	8.13×10^{-6}	0.89	0.005	0.87	0.0005
European/no	rs7944308	*KCNA4*	G	0.42	0.71	7.76×10^{-7}	0.94	0.02	0.90	5.80×10^{-5}
Variants identified in the Extremes of DR Discovery Analysis										
AA/no	rs74161190	*TCERG1L*	A	0.94	0.32	4.57 X 10-6	0.40	0.03	0.32	7.16×10^{-7}
European/yes	rs17706958	*PDZRN3*	T	0.76	1.58	3.04×10^{-8}	1.28	0.02	1.39	7.41×10^{-8}
European/yes	rs10932347	*CPS1*	A	0.04	0.33	4.22×10^{-7}	0.64	0.02	0.55	1.30×10^{-5}
AA/yes	rs2690028	*KAZN*	C	0.32	0.62	4.52×10^{-6}	0.80	0.03	0.74	1.72×10^{-5}
European/yes	rs116972715	*DSC3*	C	0.99	2.60	2.48×10^{-6}	3.62	0.03	3.29	1.59×10^{-5}
European/yes	rs75167957	*CTNNA2*	C	0.99	3.26	3.36×10^{-6}	9.77	0.04	6.34	5.83×10^{-6}
AA/yes	rs6577631	*LOC339862*	G	0.86	0.53	3.45×10^{-6}	0.89	0.04	0.84	0.0006

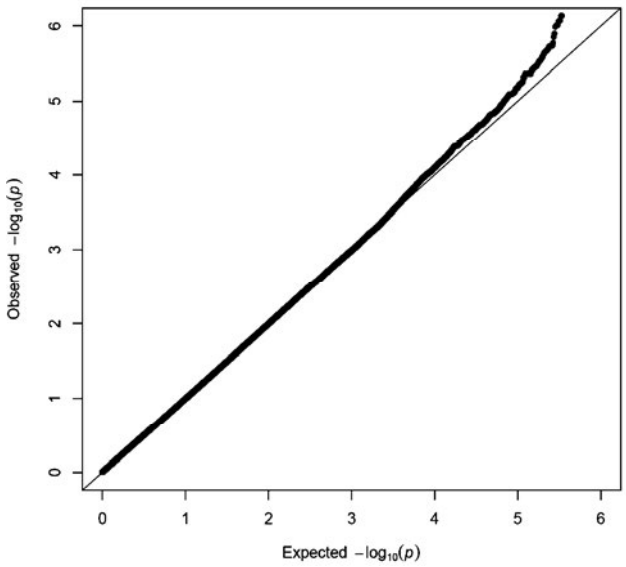
Sens= Sensitivity Analysis, ME = Multiethnic, AA = African American, DR = diabetic retinopathy, PDR = proliferative diabetic retinopathy, NPDR = non-proliferative diabetic retinopathy. * For insertions-deletion, the reference allele is shown first followed by the alternate allele

Figure Legends

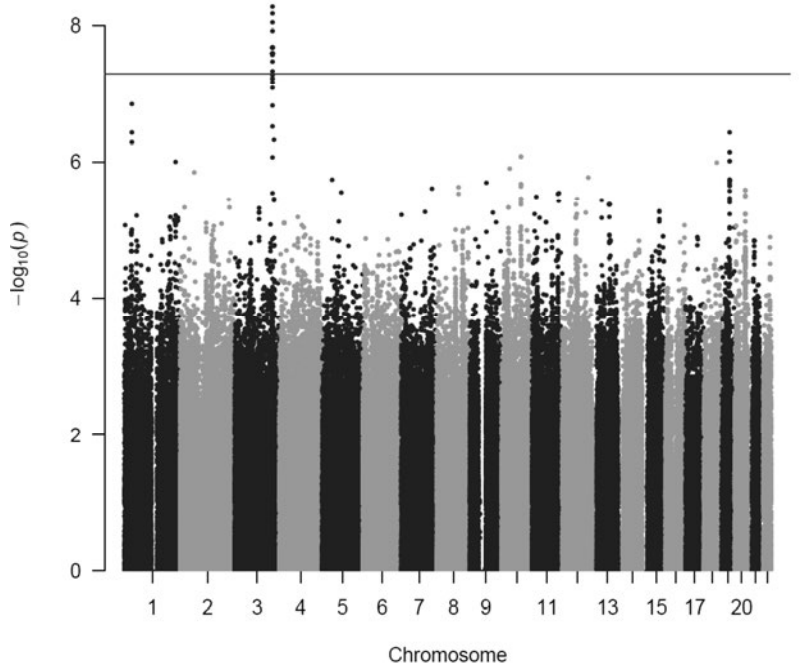
Figure 1. Quantile-quantile and Manhattan plots for the PDR and extremes of DR discovery meta-analyses for: (A and B) PDR analysis in African American participants with liability threshold modeling of duration of diabetes and glycemic control; (C and D) PDR analysis in European participants with liability threshold modeling of duration of diabetes and glycemic control; (E and F) Extremes of DR analysis in African American participants with liability threshold modeling of duration of diabetes and glycemic control; and (G and H) Extremes of DR analysis in European participants with liability threshold modeling of duration of diabetes and glycemic control. The horizontal line in each of the Manhattan plots indicates the nominal threshold for genome-wide significance ($P = 5 \times 10^{-8}$).

Figure 2. Protein network from the African American proliferative diabetic retinopathy discovery analysis that was significant in the DAPPLE analysis. This significant protein network includes genes with primary roles in inflammation (*IFNG*, *IL22RA1*, *CFH*, *SELL)*, protein function/endoplasmic reticulum function (*ADAMT30, ERP44, HSP90B1, SPON1, CNAX, WFS1*), catabolic processing/metabolism (*PPT1, ALDH1B1*), gene expression/transcription factor activity (*HNRNPH1, TAF4, POLR2E, TCEB1, COMMD1, PLAGL1, THRB, SIN3A*), macromolecule transport (*NUP153, NUP50*), protein localization (*SEC61B, SEC61A2*), and DNA repair/cell cycle (*RBBP8, ATM, EEF1E1*).

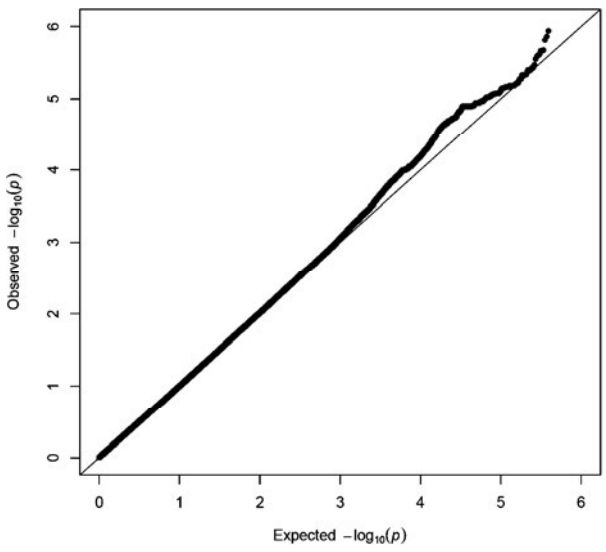
A.


Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
1
2
3
4
5
6

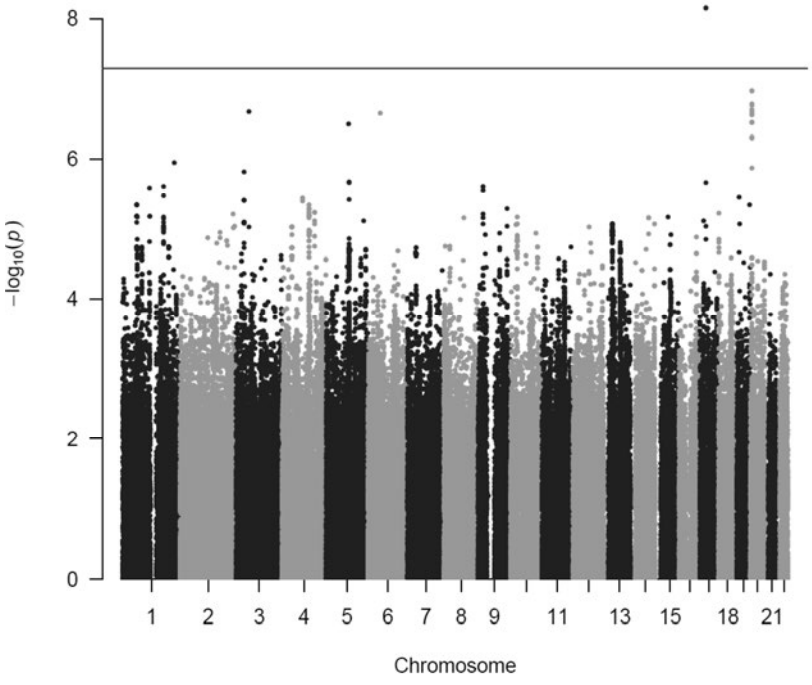
B.


$-\log_{10}(p)$
Chromosome
0
2
4
6
8
1
2
3
4
5
6
7
8
9
11
13
15
17
20

C.

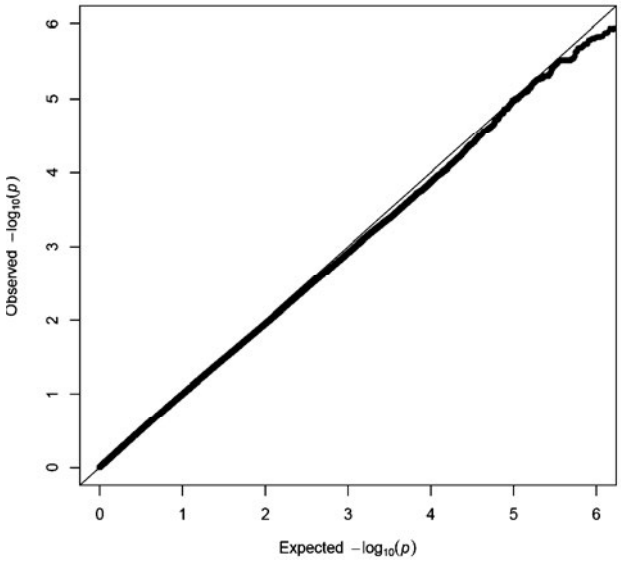

Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
1
2
3
4
5
6

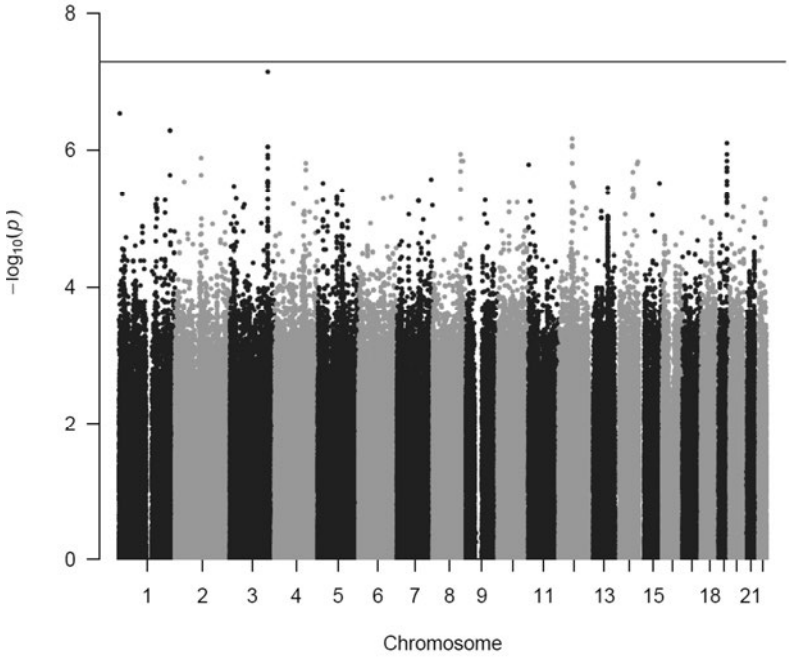
D.


$-\log_{10}(p)$
Chromosome
0
2
4
6
8
1
2
3
4
5
6
7
8
9
11
13
15
18
21

E.

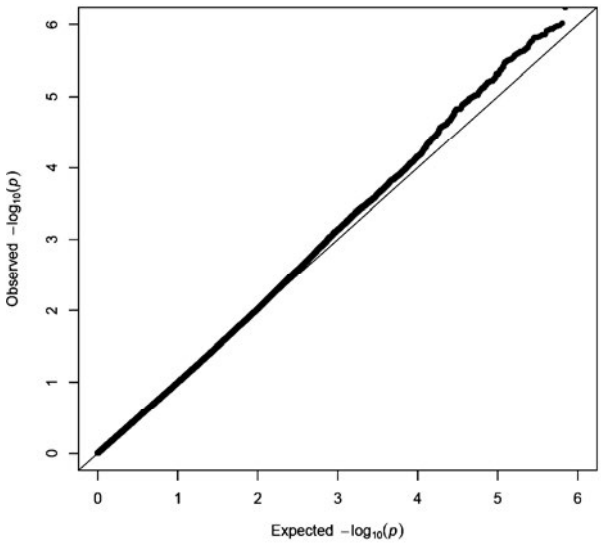
F.

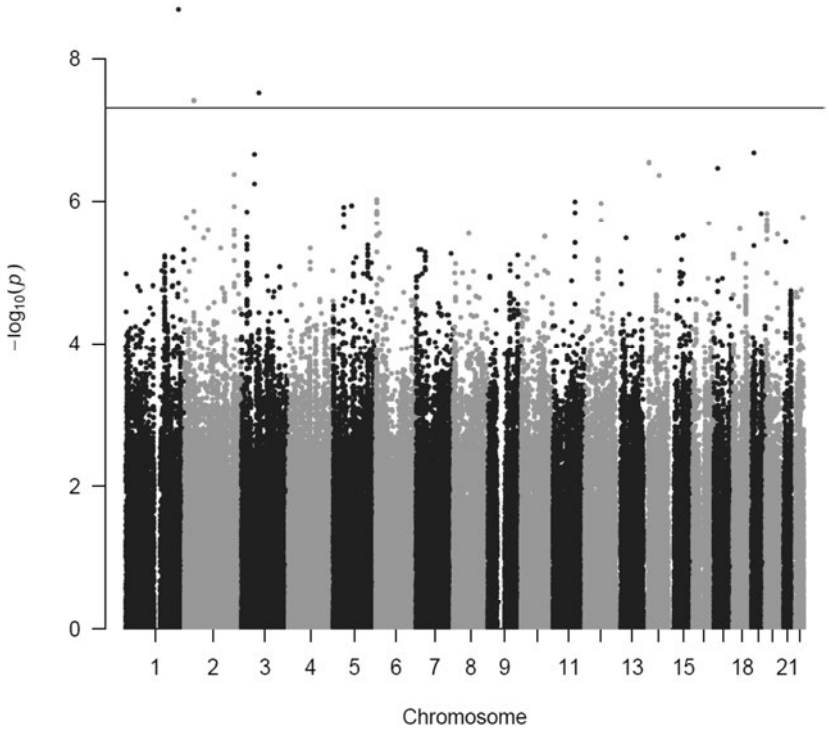

Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0 1 2 3 4 5 6

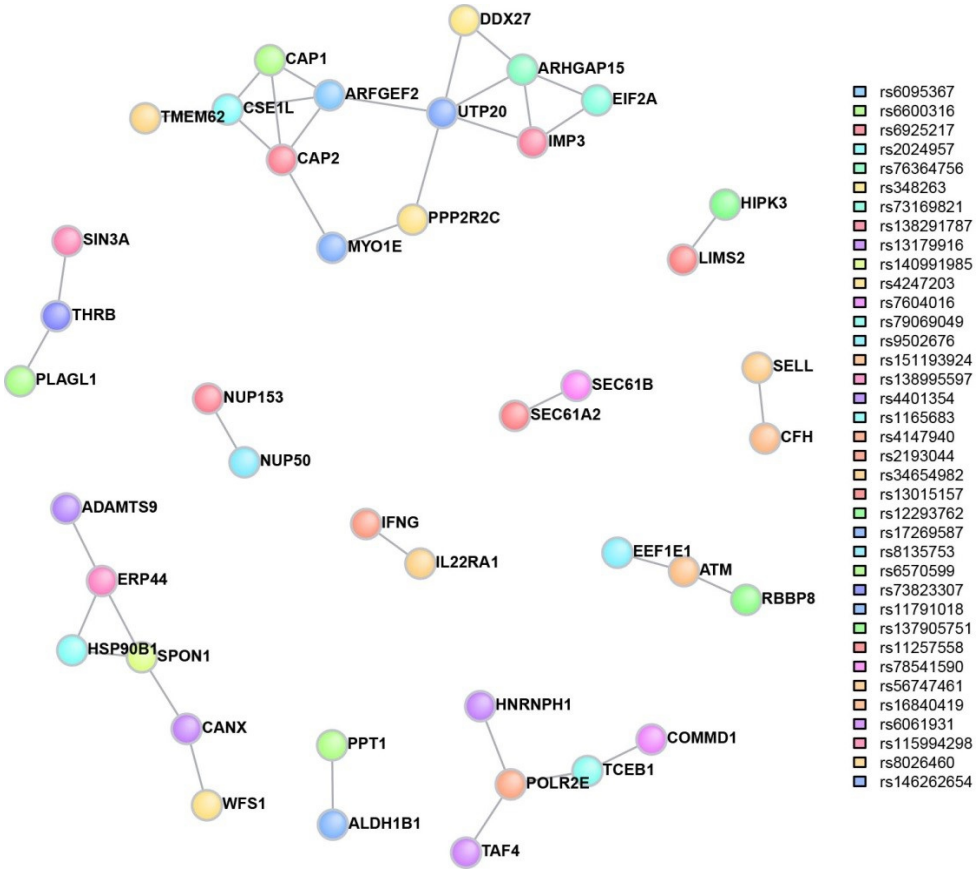

$-\log_{10}(p)$
Chromosome
0 2 4 6 8
1 2 3 4 5 6 7 8 9 11 13 15 18 21

G.

H.


Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0 1 2 3 4 5 6


$-\log_{10}(p)$
Chromosome
0 2 4 6 8
1 2 3 4 5 6 7 8 9 11 13 15 18 21



132x114mm (220 x 220 DPI)

Supplemental Materials

Web Resources:

EIGENSOFT: https://www.hsph.harvard.edu/alkes-price/software/

LTSOFT: https://www.hsph.harvard.edu/alkes-price/software/

MAGENTA (Meta-Analysis Gene-Set Enrichment of variaNT Associations**)**: http://www.broadinstitute.org/mpg/magenta

Genome Aggregation Database (gnomAD): http://gnomad.broadinstitute.org

PLINK 1.9: https://www.cog-genomics.org/plink/1.9/ld

GTEx Portal: https://www.gtexportal.org

Supplemental Materials

Additional Details Regarding Diabetic Retinopathy Phenotyping in the Discovery Cohorts
With the exception of the Wake Forest University (WFU) cohort, all of the discovery cohort samples used photography with Early Treatment Diabetic Retinopathy Study (ETDRS) grading for diabetic retinopathy (DR) phenotyping. The WFU cohort based DR diagnosis on questionnaire items of previous laser treatment for DR. Therefore, WFU could only be included in the proliferative DR (PDR) analysis as it was not possible to determine the degree of retinopathy in patients without laser treatment to include them in the other analyses. The Ranibizumab Injection in Subjects with Clinically Significant Macular Edema with Center Involvement Secondary to Diabetes (RISE/RIDE) cohort did not include any patients without retinopathy and therefore could only be incorporated in the PDR analyses as well.

Rationale for the Four Different Case-Control Definitions
The rationale for using four DR case-control definitions is multiple-fold. First, it allows us to replicate findings from previous studies across a range of case-control definitions. Similarly, it also allows us the greatest opportunity for well-powered replication of our discovery analysis findings since the available cohorts for replication use different case-control definitions. The any DR and PDR analyses have the advantages of incorporating the largest sample sizes, thus maximizing power. They also help distinguish if there are genetic variants that increase risk at the earliest stage of retinopathy development (any DR) vs. advanced retinopathy (PDR). To specifically address the genetic basis for the response to retinal ischemia, which is responsible for the transition from NPDR to PDR, we performed a *post hoc* sensitivity analysis of the PDR analysis where we excluded participants with no DR and compared participants with NPDR to those with PDR. The non-proliferative DR (NPDR) analysis addresses whether there are genetic variants that influence DR at its intermediate stage. Because many of the discovery cohorts were population-based studies, they did not have large numbers of patients with advanced DR. Thus, the NPDR analysis preserves sample size to some degree while trying to examine a more rigorous case-control definition than the any DR analysis. The extremes of DR analysis was performed with the expectation that the decrease in power from smaller sample size would be compensated by enrichment for genetic variants of strong effects in participants at the extremes of DR development.

Initial Quality Control and Data Filtering

The following data filtering and quality-control measures were employed. First, for each study (e.g. WFU) separately, we removed any SNP with minor allele frequency (MAF) <0.1% and with >1% missing data, and that did not have a human genome build 37 physical position and rs id. Next, for each cohort separately, we inferred the continental (EUR, ASI, AFR) ancestry proportions of each individual using SNPweights.(Chen et al., 2013) European individuals were defined based on >90% European ancestry. African American individuals were defined based on >50% African ancestry and <10% East Asian ancestry. We observed that there existed some individuals with African ancestry between 10% and 50% and less than 10% East Asian ancestry, and these were removed for the primary analyses.

From this point forward, all data filtering and quality control steps continued to be performed for each study separately and, in addition, were performed separately for African Americans and Europeans if that study had both African Americans and Europeans. A Hardy-Weinberg filter was applied for each SNP. We compared the actual number of heterozygotes to the expected number of heterozygotes based on MAF. We removed the SNP if the actual number of heterozygotes was greater than 5 standard deviations from expected.

We removed one of each pair of duplicate samples. Two of the studies recruited patients from overlapping geographic areas and there was a higher chance for duplicate samples between them. The Jackson Heart Study (JHS) overlapped in recruitment area with the Atherosclerosis Risk in Communities (ARIC) Study. For these studies, we merged the cohorts prior to the duplicate check, performed the duplicate check on the reduced set of merged SNPs, removed the duplicate sample, and then restored the full set of cohort-specific SNPs to each cohort before proceeding with the analyses below.

1000 Genomes Imputation

Next, imputation was done with the 2.3.0 version of impute2. We used the cosmopolitan reference sample based on 1,092 1000 Genomes Phase 1, version 3 samples. After imputation, any SNP with low imputation accuracy (INFO <0.6), MAF<1%, or < 5 copies in imputed data was removed from that cohort. SNPs that were retained in only a subset of cohorts were included in the meta-analysis. Association analysis was run on rounded dosages, retaining highly uncertain genotypes (e.g. P(most likely genotype)<0.9) so long as the SNP has imputation quality score ≥ 0.6.

Related individuals

After imputation, we removed one of each pair of related individuals separately for each cohort, except for JHS and ARIC where overlap was expected as described above. We removed one of each pair of individuals with genome-wide relatedness >0.10 with the smartrel program in EIGENSOFT (see Web Resources). This genome-wide relatedness coefficient is analogous to the kinship coefficient.

The quality control steps and analyses were performed for all studies at the Harvard School of Public Health except for the Family Investigation of Nephropathy and Diabetes (FIND) and Age/Gene Susceptibility Study (AGES). For these two studies, the consents and permissions did not allow for sharing of individual-level data so these cohorts performed the above analyses at the primary study site and shared their data for meta-analysis. Because there was also overlap between WFU and FIND with regards to recruitment area, the FIND study identified cross-study duplicate individuals and those were removed from the WFU study for the analyses.

Liability Threshold Modeling

To address the issue of incorporating covariates while also optimizing statistical power, a method based on the classical technique of liability threshold (LT) modeling was developed. Individuals who progress to DR shortly after diabetes diagnosis or with tighter glycemic control may be more likely to harbor putative genetic risk variants whereas individuals who do not develop DR after having diabetes for several decades and with poor glycemic control might harbor protective genetic variants.(Aschard et al., 2017)This modeling method can adjust for variates like more standard approaches (Armitage, 1955; Thomas, 2010) but it also increases the statistical power of the analysis when compared with standard approaches by a wide margin. This has been shown in simulation and in empirical case-control studies of type 2 diabetes, prostate cancer, breast cancer, rheumatoid arthritis, macular degeneration, and end stage renal disease.(Zaitlen et al., 2012a)

The idea behind LT modeling is that there exists an underlying unobserved quantitative trait, which is a function of covariates, a genetic effect being evaluated, and statistical noise from a standard normal distribution.(Falconer, 1967) An individual is a case if the quantitative trait is above some threshold, and a control otherwise. Only case-control status is observed, and from this information the posterior mean value of the quantitative trait after adjusting for covariates can be computed for each individual. Standard association tests for quantitative traits can then be applied to these posterior mean values. Application of these methods to DR could substantially increase the ability to discover genes. We expect that subjects with DR and short duration of diabetes (DoD) and/or good glycemic control harbor greater genetic risk than those with DR and long DoD and poor glycemic control; LT modeling leverages this understanding of covariates' effects to maximize this study's power.

The LTSCORE test has been previously described.(Falconer, 1967) Let x be a phenotypic covariate with mean $\bar{x}$, and let $\varphi(x) = \alpha(x - \bar{x}) + \beta + \varepsilon$ be an underlying quantitative phenotype such that individuals with $\varphi(x) \geq 0$ are disease cases. We assume $\varepsilon \sim N(0,1)$ and model disease prevalence using the affine term β, precluding the need for a nonzero threshold. For example, for DR with x=DoD, we fit α=0.03 and β=–0.5, based on higher DR prevalence with higher values of DoD (e.g. 22%, 47% and 57% for DoD 0-9 yrs, 10-19 yrs and 20-29 yrs; see below).
Instead of conducting an association test using case-control phenotype z, we use the posterior mean estimate

$$E(\varepsilon \mid z,x) = \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \varepsilon \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\varepsilon^2}{2}\right) d\varepsilon \Big/ \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\varepsilon^2}{2}\right) d\varepsilon$$

of the underlying quantitative phenotype ε, where $(\varepsilon_{\min},\varepsilon_{\max}) = (-\alpha(x - \bar{x}) - \beta, \infty)$ for z=1 (cases) and $(\varepsilon_{\min},\varepsilon_{\max}) = (-\infty, -\alpha(x - \bar{x}) - \beta)$ for z=0 (controls). We define our LT statistic as the number of samples times the squared correlation between genotype g and posterior mean $E(\varepsilon \mid z,x)$ across samples, generalizing the standard Armitage trend test.(Armitage, 1955) This statistic is a χ^2(1 d.f.) statistic and is equivalent to a score test, which is commonly used in genetic association studies.(Marchini, Howie, Myers, McVean, & Donnelly, 2007) The appropriate χ^2(1 d.f.) distribution is observed in simulations and in permuted real data.(Zaitlen et al., 2012b)

For this study, the LTSCORE test made use of information on two clinical covariates: DoD and glycemic control. Glycemic control could be captured by either hemoglobin A_{1C} (HbA_{1C}) or fasting plasma glucose (FPG). We used phenotype and covariate information to compute the posterior mean $E(\varepsilon \mid z,x)$ of each sample, based on the above LT model parameters. Covariate files specifying liability-scale effect sizes of DoD and HbA1c or DoD and FPG covariates in the population were created separately for African Americans and Europeans. If HbA_{1C} data was available, that was used preferentially over FPG as it is more precise measure of glycemic control. We estimated liability-scale effect sizes of DoD, HbA_{1C} and FPG

using cohorts that ascertained/enrolled diabetic participants without regard to DR status. For African Americans, these cohorts were JHS, ARIC, Cardiovascular Health Study (CHS) and Multi-Ethnic Study of Atherosclerosis (MESA). For Europeans, these cohorts were ARIC, CHS, MESA, and Blue Mountains Eye Study (BMES). The numbers of African American participants who contributed information on DoD, FPG and HbA_{1c} to calculate liability-scale effect sizes were 702, 988 and 866, respectively. The numbers of European participants who contributed information on DoD, FPG and HbA_{1c} to calculate liability-scale effect sizes were 917, 1389 and 1013, respectively.

Duration of actual type 2 diabetes is difficult to precisely ascertain in the absence of longitudinal measurement of HbA_{1C} up to disease onset. For this reason, all of our analyses instead use duration of *diagnosed* type 2 diabetes as a covariate. Because all covariate effect size estimates are based on duration of diagnosed type 2 diabetes and not on duration of actual type 2 diabetes, there is no mismatch between estimated and true effects of duration of diagnosed type 2 diabetes. We acknowledge that the gain in power attained by including the duration of diagnosed type 2 diabetes covariate may be lower than the gain in power that can be attained by including the duration of actual type 2 diabetes covariate (with correspondingly higher covariate effect size), however the latter analysis is not possible based on available data. We further note that using a less than maximally informative covariate would not introduce false positives, since the purpose of including this covariate is to increase statistical power (DR cases with lower duration of diagnosed type 2 diabetes harbor stronger genetic signals) and not to avoid confounding.

Sensitivity Analyses and Meta-analysis Minimum Sample Sizes

We performed sensitivity analyses of the meta-analysis for the primary any DR analysis by removing cohorts that did not have bilateral grading for DR (ARIC, CHS) or did not have fundus photograph grading for retinopathy (FIND). The rationale for this sensitivity analysis was to limit the discovery sample to cohorts with high quality documentation of DR status and thus decrease misclassification that might bias results to the null.

Meta-analysis statistics were computed only for SNPs with at least half of the total available effective sample size. For any given analysis, only studies with at least 10 cases and at least 10 controls (for a given ethnicity) were included. Such low numbers of cases or controls may lead to unstable statistical findings.

Replication Cohorts

The replication cohorts chose the index SNP if it was available in their dataset. The replication cohorts chose the case-control definition from our discovery analyses that most closely matched their existing case-control definitions. If possible, the replication cohorts re-coded their case-control definitions to match all of the case-control definitions in the discovery analyses. However, this was not always feasible and therefore, every cohort did not provide information on every case-control definition. The replication cohorts incorporated DoD and glycemic control using traditional adjustment in logistic regression.

Protein-Protein Interaction Analysis of Top GWAS Loci (DAPPLE) Protocol

We applied the Disease Association Protein-Protein Link Evaluator (DAPPLE) to our discovery GWAS to identify significantly-enriched protein networks among the loci with the highest statistical evidence for association to DR. Each locus was defined based on linkage disequilibrium ($r^2 > 0.5$ from the associated SNP, then expand to the nearest recombination hotspot). Genes were identified by whether their transcription start or

transcription stop (with 50kB upstream and downstream to account for regulatory DNA) overlapped with the locus. Using these genes, we identified direct and indirect protein-binding networks based on the InWeb [published by Lage et al.(Lage et al., 2007) and used by DAPPLE] protein-protein interaction database (a curated database of published protein-protein interactions) and tested whether cross-locus connections were present. Permutation methods were used to score the resultant network. The P-values generated represent the probability of drawing such a network from a random distribution (controlled for the binding properties of each individual protein).

Gene Set Enrichment Analysis with MAGENTA of DAPPLE significant genes

We applied gene set enrichment analysis (GSEA) to our GWAS using MAGENTA (Meta-Analysis Gene-Set Enrichment of variaNT Associations) to further support the protein-protein interaction results from the DAPPLE analysis. Adjusted gene association P-values were computed for all genes genome-wide based on the association P-value of the most significant variant within 110kb upstream to the transcript start site and 40kb downstream to the transcript end site, correcting for gene size, density of variants per gene, and population-specific LD properties. Gene set enrichment was assessed using a hypergeometric approach and permutation analysis (MAGENTA version 2.6; see Web Resources) and the 95th percentile of all gene association scores as the enrichment cutoff.

Type 2 Diabetes and Associated Glycemic Traits Loci – Meta-analysis Computation of Correlation

For the correlation computation, beta and logOR were obtained from the original discovery cohorts. These traits were tested using the any DR case-control definition for Europeans only without covariates. This was performed in the European cohorts only because majority of the trait SNPs and their effect sizes were originally assessed in Europeans. The SNPs were obtained from the largest powered meta-analyses for the following traits: type 2 diabetes (Scott et al., 2017), fasting glucose (Manning et al., 2012), fasting insulin (Manning et al., 2012), 2-hour glucose (Saxena et al., 2010), pro-insulin (Saxena et al., 2010), and HbA1c (Wheeler et al., 2017). The full list of SNPs is in Supplemental Table 34. For each trait, we computed a meta-analyzed correlation (R) across cohorts as the (effective) sample size-weighted average of per-cohort correlations (N_{indiv} = number of individuals in each cohort including cases and controls; ρ = correlation found in each cohort), and computed a Z score (Z) for the meta-analyzed correlation by multiplying it by the square root of the sum of (effective) sample sizes.:

$$(1)\quad R = \frac{\sum_{cohorts} N_{indiv}\rho}{\sum_{cohorts} N_{indiv}}$$

$$(2)\quad Z = R\sqrt{\sum_{cohorts} N_{indiv}}$$

Expression quantitative trait loci (eQTL) analysis of DR associations

We tested whether the two genome-wide significant DR associations: rs142293996 (found in the European population) and rs184340784 (found in the African American population), or variants in LD with these variants were significant *cis*-eQTLs [false discovery rate (FDR) ≤ 0.05] in any of the 48 tissues tested in Genotype-Tissue Expression (GTEx) release v7 (datasets and methods description available on GTEx Portal: see Web Resources). We used 1000 Genomes Project Phase 3 genotypes (dated 20130502) from the European or African American super populations (503

and 661 individuals, respectively), to identify LD-proxy variants at $r^2>0.5$ [a cutoff shown to capture a substantial proportion of colocalizing eQTL-GWAS signals(Ongen et al., 2017)], using Plink 1.9 (see Web Resources). If no significant eQTL was found, we checked whether the variants and genes in the GWAS locus were tested for eQTL analysis in any of the 48 tissues in GTEx.

Colocalization analysis of eQTL and GWAS associations

For the significant rs41271487-*CAPN2* brain spinal cord eQTL found to be in LD ($r^2=0.62$) with the rs142293996 DR association, we assessed the probability that the co-occurring GWAS and eQTL signals are tagging the same causal variant, using two Bayesian-based colocalization analysis methods: eCAVIAR (Hormozdiari et al., 2016), which assumes allelic heterogeneity in a locus (using the default of 2 causal variants per locus), and the *coloc.abf* function (Giambartolomei et al., 2014), which assumes only one causal variant per locus. For both methods, we used the extremes of DR GWAS summary statistics. We followed recommended guidelines for each method. For eCAVIAR, we considered 100 variants upstream and downstream of the lead GWAS variant that were tested in both the GWAS and eQTL studies (179 variants total), and used as input z-scores from both studies. For the *coloc* analysis, we analyzed all variant-gene pairs tested for eQTL analysis in GTEx, within ±1Mb of the transcript start site of *CAPN2*, which were also tested in the DR GWAS (4,611 variants total). The variant p-values, study-specific MAF, sample sizes, and the GWAS case/control ratio were inputted into the analysis. For prior probabilities that a variant is associated with gene expression or disease status we used: $p_1=p_2=10^{-4}$, and for the probability that the variant is associated with both traits we used: $p_{12}=10^{-6}$. We used both methods' recommended significance cutoffs: colocalization posterior probability (CLPP)>0.01 for eCAVIAR, and posterior probability for the same variant affecting both traits (PP4) > 0.9 for coloc.abf. We found no statistical support for the co-occurring DR association, rs142293996, and the rs41271487-CAPN2 spinal cord eQTL tagging the same causal variant [eCAVIAR colocalization posterior probability (CLPP) < 0.0056; coloc posterior probability (PP4)=0.002]. However, given that the CLPP of the lead GWAS variant, rs142293996, was substantially larger than the CLPP of the surrounding variants (average CLPP of other variants=$1.5\text{x}10^{-5}$), the lack of colocalization may be due to the GWAS and eQTL studies being under-powered (N_{eQTL}= 83 and $N_{effective}$ of GWAS=797).

References

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics, 11*, 375-386.

Aschard, H., Guillemot, V., Vilhjalmsson, B., Patel, C. J., Skurnik, D., Ye, C. J., . . . Zaitlen, N. (2017). Covariate selection for association screening in multiphenotype genetic studies. *Nat Genet, 49*(12), 1789-1795. doi:10.1038/ng.3975

Chen, C. Y., Pollack, S., Hunter, D. J., Hirschhorn, J. N., Kraft, P., & Price, A. L. (2013). Improved ancestry inference using weights from external reference panels. *Bioinformatics, 29*(11), 1399-1406. doi:10.1093/bioinformatics/btt144

Falconer, D. S. (1967). The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann Hum Genet, 31*(1), 1-20.

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., & Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet, 10*(5), e1004383. doi:10.1371/journal.pgen.1004383

Hormozdiari, F., van de Bunt, M., Segre, A. V., Li, X., Joo, J. W. J., Bilow, M., . . . Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet, 99*(6), 1245-1260. doi:10.1016/j.ajhg.2016.10.003

Lage, K., Karlberg, E. O., Storling, Z. M., Olason, P. I., Pedersen, A. G., Rigina, O., . . . Brunak, S. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol, 25*(3), 309-316. doi:10.1038/nbt1295

Manning, A. K., Hivert, M. F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H., . . . Langenberg, C. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet, 44*(6), 659-669. doi:10.1038/ng.2274

Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet, 39*(7), 906-913. doi:ng2088 [pii]
10.1038/ng2088

Ongen, H., Brown, A. A., Delaneau, O., Panousis, N. I., Nica, A. C., Consortium, G. T., & Dermitzakis, E. T. (2017). Estimating the causal tissues for complex traits and diseases. *Nat Genet, 49*(12), 1676-1683. doi:10.1038/ng.3981

Saxena, R., Hivert, M. F., Langenberg, C., Tanaka, T., Pankow, J. S., Vollenweider, P., . . . Watanabe, R. M. (2010). Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet, 42*(2), 142-148. doi:10.1038/ng.521

Scott, R. A., Scott, L. J., Magi, R., Marullo, L., Gaulton, K. J., Kaakinen, M., . . . Meta-analysis, C. (2017). An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes, 66*(11), 2888-2902. doi:10.2337/db16-1253

Thomas, D. (2010). Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet, 11*(4), 259-272. doi:nrg2764 [pii]
10.1038/nrg2764

Wheeler, E., Leong, A., Liu, C. T., Hivert, M. F., Strawbridge, R. J., Podmore, C., . . . Meigs, J. B. (2017). 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 medicine, 14*(9), e1002383. doi:10.1371/journal.pmed.1002383

Zaitlen, N., Lindstrom, S., Pasaniuc, B., Cornelis, M., Genovese, G., Pollack, S., . . . Price, A. L. (2012a). Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS genetics, 8*(11), e1003032. doi:10.1371/journal.pgen.1003032

Zaitlen, N., Lindstrom, S., Pasaniuc, B., Cornelis, M., Genovese, G., Pollack, S., . . . Price, A. L. (2012b). Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet, 8*(11), e1003032. doi:10.1371/journal.pgen.1003032

Supplemental Table 1. Clinical Covariate Information by Study and Ancestry for Discovery Cohorts

Cohort	Ancestry	Cases (ETDRS ≥ 14)			Controls (ETDRS < 14)		
		DoD, months	HbA_{1C} %(mmol/mol)	FPG mg/dL(mmol/L)	DoD, months	HbA_{1C} %(mmol/mol)	FPG mg/dL(mmol/L)
AAPDR	AA	249.3	8.47 (69)	NA	188.8	7.55 (59)	NA
AGES	European	176.9	6.78 (51)	150.5 (8.35)	119.7	6.32 (46)	135.9 (7.54)
ARIC	AA	147.1	9.65 (82)	244.2 (13.55)	80.1	7.36 (57)	183.1 (10.16)
	European	164.5	8.84 (73)	202.8 (11.26)	103.5	6.71 (50)	164.9 (9.15)
AUST	European	217.9	8.41 (68)	NA	151.5	7.45 (58)	NA
BMES	European	109.7	NA	173.0 (9.60)	90.27	NA	139.6 (7.75)
CHS	AA	168.7	NA	145.2 (8.06)	108.7	NA	184.2 (10.22)
	European	191.5	NA	164.4 (9.12)	114.5	NA	140.6 (7.80)
FIND-Eye	AA	279.4	7.44 (58)	NA	230.5	6.84 (51)	NA
	European	269.9	7.20 (55)	NA	189.4	6.85 (51)	NA
JHS	AA	180.6	8.45 (69)	162.4 (9.01)	87.9	6.66 (49)	123.5 (6.85)
MESA	AA	81.8	7.64 (60)	146.3 (8.12)	59.7	6.99 (53)	134.4 (7.46)
	European	14.9	6.91 (52)	140.1 (7.78)	30.1	6.61 (49)	130.7 (7.25)
RISE/RIDE	AA	228.7	8.45 (69)	150.0 (8.32)	--	--	--
	European	194.1	7.60 (60)	152.1 (8.44)	--	--	--
WFU	AA	294.5	7.63 (60)	150.0 (8.32)	--	--	--

AAPDR = African American Proliferative Diabetic Retinopathy Study, AGES = Age, Gene/Environment Susceptibility Study, ARIC = Atherosclerosis Risk In Communities Study, AUST= Australian Genetics of Diabetic Retinopathy Study, BMES = Blue Mountains Eye Study, CHS=Cardiovascular Health Study, FIND-Eye = Family Study of Nephropathy and Diabetes-Eye, JHS = Jackson Heart Study, MESA = Multi-Ethnic Study of Atherosclerosis, RIDE/RISE= Ranibizumab Injection in Subjects with Clinically Significant Macular Edema with Center Involvement Secondary to Diabetes, WFU=Wake Forest University, AA=African American, NA=not available, DoD=Duration of Diabetes, HbA_{1C}=hemoglobin A_{1C}, FPG=fasting plasma glucose

Supplemental Table 2. Discovery Study Cohort Details

Study	Population	Genotyping Platform	SNPs genotyped	SNPs after QC	SNPs after imputation
AAPDR	AA	Affy 6.0	883851	579834	12194970
AGES*	EUR	Illum 370	353202	304677	7811063
ARIC	AA	Affy 6.0	603146	505092	12308412
	EUR	Affy 6.0	603146	505092	6665985
AUST	EUR	Illum OmniExpress	605258	603947	6933006
BMES	EUR	Illum 670	544799	525303	6886208
CHS	AA	Illum Omni Quad	886457	844504	8501510
	EUR	Illum 370	886457	844504	2168325
FIND-Eye*	AA	Affy 6.0	837,210	798920	19463075
	EUR	Affy 6.0	727,731	690150	8813873
JHS	AA	Affy 6.0	633267	580336	12441123
MESA	AA	Affy 6.0	735518	671745	12733230
	EUR	Affy 6.0	735518	671745	6727876
RISE/RIDE	EUR	Illum 2.5M	2379855	§	§
WFU	AA	Affy 6.0	868155	817755	7044061

AAPDR = African American Proliferative Diabetic Retinopathy Study, AGES = Age, Gene/Environment Susceptibility Study, ARIC = Atherosclerosis Risk In Communities Study, AUST= Australian Genetics of Diabetic Retinopathy Study, BMES = Blue Mountains Eye Study, CHS=Cardiovascular Health Study, FIND-Eye = Family Study of Nephropathy and Diabetes-Eye, JHS = Jackson Heart Study, MESA = Multi-Ethnic Study of Atherosclerosis, RIDE/RISE= Ranibizumab Injection in Subjects with Clinically Significant Macular Edema with Center Involvement Secondary to Diabetes, WFU=Wake Forest University, AA = African American, EUR = European, HbA_{1C}=hemoglobin A_{1C}, FPG=fasting plasma glucose

§ The RISE and RIDE studies were imputed separately, RISE has 1174886 SNPs after QC and 7417483 SNPs after imputation. RIDE had 1607163 SNPs after QC and 7888767 SNPs after imputation.

* Cohorts without access to raw genotype information

Supplemental Table 3. Clinical Covariate Information by Study and Ancestry for Replication Cohorts

Cohort	Cases				Controls			
	Case Definition	DoD, months	HbA_{1C} %(mmol/mol)	FPG mg/dL(mmol/L)	Control Definition	DoD, months	HbA_{1C} %(mmol/mol)	FPG mg/dL(mmol/L)
Asian								
KSDR	Any DR	202	7.9 (63)	NA	No DR	NA	6.8 (51)	NA
MESA-Chinese	Any DR	NA	7.5 (58)	154 (8.6)	No DR	NA	7.1 (54)	136 (7.6)
RIKEN – Stage 1	Any DR	143	NA	NA	No DR	159	NA	NA
RIKEN- Stage 2	Any DR	165	NA	NA	No DR	142	NA	NA
SCES I/ SCES II/ SiMES/ SINDI	Any DR	148	8.3 (67)	202 (11.2)	No DR	65	7.5 (58)	168 (9.0)
TUDR	PDR	156	8.9 (74)	NA	No DR	168	9 (75)	NA
European								
DCCT/EDIC – primary	Severe DR	233	9.9 (85)	NA	No severe DR	216	8 (6.4)	NA
DCCT/EDIC – secondary, conventional	Severe DR	313	9.1 (76)	NA	No severe DR	294	8.2 (66)	NA
DCCT/EDIC – secondary, intensive	Severe DR	316	8.7 (72)	NA	No severe DR	306	7.5 (58)	NA
GENESIS/GENEDIAB	DR	190	7.8 (62)	NA	No DR	315	8.5 (69)	NA
GoDARTS	no or mild DR	286	8.1 (65)	NA	severe NPDR/PDR	169	7.3 (56)	NA
GoKIND	laser	385	7.5 (58)	NA	no laser	290	7.5 (58)	NA
SUMMIT – FinnDiane	severe DR	379	8.7 (72)	NA	no/mild DR	340	8.2 (66)	NA
SUMMIT- SDR*	DKD	388	7.9 (63)	NA	no DKD	378	7.1 (54)	NA
SUMMIT – EuroDiab*	DKD	284	8.3 (67)	NA	no DKD	300	7.7 (61)	NA
WESDR	No PDR or DME	425	9.3 (78)	NA	PDR +/- DME	397	8.5 (69)	NA
Hispanic								
GOLDR	Any DR	86	8.9 (74)	NA	no DR	162	7.8 (62)	NA
LALES	Any DR	150	8.8 (73)	NA	no DR	82	8.1 (65)	NA
MESA - Hispanics	NA	NA	8.3 (67)	169 (9.4)	NA	NA	7.2 (55)	148 (8.2)
SCHS	severe NPDR/PDR	217	12.3 (111)	202 (11.2)	no or early NPDR	144	10.9 (96)	186 (10.3)

DoD=Duration of Diabetes, HbA_{1C}=hemoglobin A_{1C}, FPG=fasting plasma glucose, NA=not available, DR= diabetic retinopathy, NPDR = non-proliferative diabetic retinopathy, PDR = proliferative diabetic retinopathy, DKD = diabetic kidney disease, KSDR = Korean Study of Diabetic Retinopathy, MESA = Multiethnic Study of Atherosclerosis, RIKEN = Rikagaku Kenkyusho - Institute of Physical and Chemical Research, SCES= Singapore Chinese Eye Study, SiMES = Singapore Malay Eye Study, SINDI = Singapore Indian Eye Study, TUDR = Taiwan–US Diabetic Retinopathy Study, DCCT/EDIC = Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications, GENESIS/GENEDIAB=Genetics nephropathy and sib pair study/Genetics, nephropathy, diabetes, GoDARTS =Genetics of Diabetes and Audit Research Tayside Study, GoKinD = Genetics of Kidneys in Diabetes, SUMMIT = Surrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools, FinnDiane = Finnish Diabetic Nephropathy Study, SDR = Scania Diabetes Registry, WESDR = Wisconsin Epidemiologic Study of Diabetic Retinopathy, GOLDR = Genetics of Latino Diabetic Retinopathy, LALES = Los Angeles Latino Eye Study, SCHS = Starr County Health Studies

* For these two studies we were unable to obtain a breakdown of the duration of diabetes and glycemic control indicators by diabetic retinopathy status so we present the breakdown by diabetic nephropathy status.

Supplemental Table 4. Ten independent variants with the lowest P values any DR analysis, African American participants, without accounting for duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs2104455	6	9368518	A	1495	911	0.8105	948	0.7464	1.57	1.18 X 10-7
rs11771617	7	125373426	A	1445	892	0.9715	913	0.9421	3.12	1.35 X 10-6
rs115489684	5	142116464	C	1495	911	0.8845	948	0.9295	0.57	1.51 X 10-6
rs184033309	5	103622556	T	1445	892	0.9583	913	0.9841	0.33	1.55 X 10-6
rs4901258	14	52696282	G	1495	911	0.5716	948	0.49	1.44	1.91 X 10-6
rs115634195	10	25735812	G	969	517	0.9734	599	0.9934	0.16	1.92 X 10-6
rs7139352	12	61345749	G	1495	911	0.7007	948	0.6342	1.46	2.01 X 10-6
rs202189921	2	236562108	CAAAAATAAATAAAA/C	1445	892	0.9875	913	0.9717	3.73	2.40 X 10-6
rs12095420	1	226788952	C	932	705	0.9965	488	0.9806	7.04	2.58 X 10-6
rs1559674	15	62106284	T	1495	911	0.2852	948	0.342	0.71	3.03 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 5. Ten independent variants with the lowest P values for any DR analysis, African American participants, with liability threshold modeling of duration of diabetes and glycemic control

						CASES		CONTROLS		
RSID	CHR	POSITION	REF	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs12095420	1	226788952	C	932	705	0.9965	488	0.9806	7.04	9.37 X 10-8
rs74152685	1	247811291	G	978	522	0.9564	592	0.9814	0.39	2.66 X 10-7
rs2300993	5	109038629	G	1495	911	0.9519	948	0.9232	1.83	4.51 X 10-7
rs11771617	7	125373426	A	1445	892	0.9715	913	0.9421	3.12	1.19 X 10-6
rs114790220	8	5163956	T	1445	892	0.9658	913	0.9863	0.34	1.26 X 10-6
rs11662496	18	41275397	A	1495	911	0.495	948	0.5679	0.72	1.38 X 10-6
rs2104455	6	9368518	A	1495	911	0.8105	948	0.7464	1.57	1.54 X 10-6
rs9882204	3	167566283	G	1495	911	0.5373	948	0.4807	1.32	1.83 X 10-6
rs190634129	4	143341715	T	969	517	0.9857	599	0.9933	0.29	1.99 X 10-6
rs73194819	2	13252874	T	1495	911	0.8586	948	0.8884	0.65	2.21 X 10-6

DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 6. Ten independent variants with the lowest P values for any DR, European participants, without accounting for duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs66598691	2	51060137	G	2464	1088	0.9431	1975	0.9053	1.87	3.49 X 10-6
rs1508244	6	87133646	A	1520	774	0.9626	816	0.9877	0.33	3.74 X 10-6
rs55921628	15	47814176	C	2464	1088	0.8802	1975	0.9124	0.66	3.83 X 10-6
rs76557325	5	76299200	A	2464	1088	0.8893	1975	0.9271	0.65	4.02 X 10-6
rs62489286	7	1365867507	C	2464	1088	0.7943	1975	0.8413	0.71	5.78 X 10-6
rs200366339	3	12298312	TTG/T	2218	1003	0.6269	1753	0.6831	0.75	7.09 X 10-6
rs150775408	3	82302346	C	2044	962	0.9652	1343	0.9326	1.97	7.24 X 10-6
rs3113743	4	181471147	C	2464	1088	0.5859	1975	0.5314	1.28	7.27 X 10-6
rs201856528	15	82286425	AAAG/A	1798	877	0.9556	1121	0.9777	0.43	7.43 X 10-6
rs12653353	5	145780135	A	2464	1088	0.2015	1975	0.2486	0.74	9.57 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 7. Ten independent variants with the lowest P values for any DR analysis, European participants, with liability threshold modeling of duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs142293996	1	224448059	C	1259	646	0.9938	643	0.986	2.2	2.30 X 10-7
rs9905843	17	74869651	C	2464	1088	0.5831	1975	0.6306	0.8	8.74 X 10-7
rs2699116	7	9216486	A	2464	1088	0.9174	1975	0.8876	1.5	9.76 X 10-7
rs8130730	21	45243204	C	2464	1088	0.8040	1975	0.8395	0.8	1.09 X 10-6
rs140675300	8	137699220	C	2044	962	0.9621	1343	0.9348	1.7	1.48 X 10-6
rs7241056	18	3381626	A	2464	1088	0.2553	1975	0.2266	1.2	1.91 X 10-6
rs151166811	9	110284659	C/CAAAAT	1798	877	0.9709	1121	0.9501	1.8	3.52 X 10-6
rs200071549	16	70563931	G/GGAGGCTGA	1670	839	0.1212	921	0.0852	1.5	4.29 X 10-6
rs117958555	19	41628014	G	1387	684	0.8004	843	0.7562	1.3	4.53 X 10-6
rs2727270	11	61603237	C	2044	962	0.9029	1343	0.8690	1.4	4.56 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 8. Ten independent variants with the lowest P values for PDR Analysis, African American participants, without accounting for duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs115523882	3	167876205	A	1452	1105	0.9823	1119	0.9611	3.1	9.42 X 10-9
rs137949823	5	43743109	G	1452	1105	0.9072	1119	0.941	0.43	6.15 X 10-8
rs7533141	1	217716877	T	1452	1105	0.6213	1119	0.5509	1.47	2.85 X 10-7
rs78340493	4	43638196	C	1452	1105	0.8579	1119	0.8186	1.66	3.01 X 10-7
rs1144964	12	69338637	G	1452	1105	0.3674	1119	0.4254	0.67	4.08 X 10-7
rs149201869	2	216352519	T	1364	1082	0.9945	532	0.9765	5.57	8.62 X 10-7
rs73228199	3	111316364	G	1452	1105	0.9254	1119	0.9522	0.36	1.13 X 10-6
rs184340784	1	4589883	C	754	534	0.999	321	0.9829	Undefined	1.15 X 10-6
rs73347124	10	54680020	A	1452	1105	0.9688	1119	0.9857	0.2	1.54 X 10-6
rs2037601	2	196282058	T	1452	1105	0.3294	1119	0.2806	1.44	1.57 X 10-6

PDR= Proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 9. Ten independent variants with the lowest P values for PDR Analysis, African American participants, with liability threshold modeling of duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs115523882	3	167876205	A	1452	1105	0.9823	1119	0.9611	3.1	5.37 X 10-9
rs1414474	1	34663411	C	1452	1105	0.8792	1119	0.8324	1.6	1.46 X 10-7
rs71354195	19	36876318	G	1452	1105	0.9814	1119	0.962	2.3	3.72 X 10-7
rs73050171	3	174519478	G	1452	1105	0.8561	1119	0.8061	1.5	4.79 X 10-7
rs11201335	10	86778199	C	1452	1105	0.586	1119	0.55	1.3	8.49 X 10-7
rs11488711	1	229241233	C	1452	1105	0.7751	1119	0.8074	0.7	1.00 X 10-6
rs138683663	18	56862782	C	1174	838	0.9952	805	0.9832	4.9	1.03 X 10-6
rs75348186	10	37071613	T	1452	1105	0.9665	1119	0.9423	1.9	1.26 X 10-6
rs7604016	2	62192135	A	1406	1093	0.0238	880	0.0438	0.4	1.41 X 10-6
rs7139429	12	116118551	A	1452	1105	0.8606	1119	0.8311	1.5	1.69 X 10-6

PDR= Proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 10. Ten independent variants with the lowest P values for the PDR Analysis, European participants, without accounting for duration of diabetes and glycemic control

						CASES		CONTROLS		
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs139205645	2	201949806	T	907	309	0.9725	975	0.9959	0.13	3.93 X 10-8
rs202105116	17	15899617	C/CA	907	309	0.9482	975	0.9851	0.26	1.71 X 10-7
rs202159532	11	96159584	A/AT	601	187	0.9626	770	0.9942	0.15	4.24 X 10-7
rs1087427	5	99705657	G	1097	389	0.8278	1092	0.7399	1.69	6.86 X 10-7
rs141785043	2	105038245	C	1097	389	0.7481	1092	0.8242	0.61	8.88 X 10-7
rs141145674	1	187073283	T	907	309	0.966	975	0.9923	0.22	9.82 X 10-7
rs115487191	11	125879341	T	907	309	0.9693	975	0.9934	0.23	1.74 X 10-6
rs35174795	10	87399545	G	1018	356	0.9241	1043	0.9669	0.41	1.80 X 10-6
rs7833313	8	121236971	A	1097	389	0.4229	1092	0.3324	1.51	1.87 X 10-6
rs118027686	21	29806610	T	601	187	0.9545	770	0.989	0.23	2.76 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. PDR= Proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 11. Ten independent variants with the lowest P values for the PDR Analysis, European participants, with liability threshold modeling of duration of diabetes and glycemic control

						CASES		CONTROLS		
RSID	CHR	POSITION	REF	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs17791488	17	26232732	T	907	309	0.9871	975	0.9661	3.7	7.26 X 10-9
rs6038419	20	6294054	A	1097	389	0.9177	1092	0.8686	1.8	1.09 X 10-7
rs17256891	3	55104767	A	907	309	0.9984	975	0.9852	3.6	2.17 X 10-7
rs62413933	6	53330900	C	1097	389	0.982	1092	0.9579	2.9	2.28 X 10-7
rs77017556	5	97479293	T	1018	356	0.9888	1043	0.9727	2.8	3.26 X 10-7
rs142293996	1	224448059	C	601	187	0.9947	770	0.9903	1.8	1.14 X 10-6
rs115630068	3	34231058	T	907	309	0.9871	975	0.9671	2.8	1.54 X 10-6
rs1087427	5	99705657	G	1097	389	0.8278	1092	0.7399	1.7	2.11 X 10-6
rs140790006	1	177641541	T	1097	389	0.5154	1092	0.4350	1.4	2.45 X 10-6
rs828579	9	22166971	A	1097	389	0.6157	1092	0.5659	1.3	2.48 X 10-6

PDR= Proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 12. Ten independent variants with the lowest P values for the NPDR Analysis, African American participants without accounting for duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs148995025	5	7085207	A	1218	754	0.9884	913	0.9689	3.41	6.39 X 10-7
rs61741249	7	157309591	C	1029	695	0.9642	683	0.9854	0.16	1.08 X 10-6
rs1360935	13	55143426	T	865	436	0.9788	697	0.9971	0.27	1.64 X 10-6
rs115885880	10	105720104	A	865	436	0.9853	697	0.9971	0.10	1.95 X 10-6
rs10878791	12	68609046	G	1258	768	0.5444	948	0.6104	0.67	1.96 X 10-6
rs150557761	7	63671293	C	1258	768	0.9000	948	0.9442	0.48	3.00 X 10-6
rs7332766	13	106007555	A	1258	768	0.6218	948	0.6974	0.67	3.42 X 10-6
rs6763376	3	167604119	G	1258	768	0.1737	948	0.1332	1.57	3.51 X 10-6
rs145279814	1	72796513	A	1218	754	0.9648	913	0.9416	2.78	3.84 X 10-6
rs78141810	18	35931012	C	1218	754	0.9442	913	0.9710	0.44	4.12 X 10-6

NPDR= non-proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 13. Ten independent variants with the lowest P values for the NPDR Analysis, African American participants, with liability threshold modeling of duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs202069793	9	107475294	TGA/T	1339	768	0.3651	948	0.3010	1.39	6.29 X 10-8
rs184340784	1	4589883	C	707	NA	NA	NA	NA	Undefined	5.62 X 10-7
rs12656571	5	109078040	G	1339	768	0.948	948	0.9227	1.89	8.01 X 10-7
rs142610219	8	116918264	A	919	433	0.9967	599	0.9850	Undefined	9.62 X 10-7
rs200295620	3	167890185	AT/A	1299	754	0.9889	913	0.9732	2.42	1.36 X 10-6
rs74816310	5	73266734	G	1299	754	0.9863	913	0.9688	Undefined	1.95 X 10-6
rs74966374	5	55905540	C	1299	754	0.9792	913	0.9705	2.13	2.19 X 10-6
rs145764941	5	67857981	C	1299	754	0.9573	913	0.9399	1.63	2.34 X 10-6
rs201685870	5	98692371	TTA/T	1339	768	0.1665	948	0.2068	0.69	2.36 X 10-6
rs201581084	1	158733766	G/GAAATA	1339	768	0.7511	948	0.6878	1.44	2.48 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. NPDR= non-proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio, NA = not available

Supplemental Table 14. Ten independent variants with the lowest P values for the NPDR analysis, European participants without accounting for duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs200063790	20	2229666	C/CT	1593	643	0.9597	1753	0.9886	Undefined	1.62 X 10-7
rs426500	5	122467316	T	1711	677	0.326	1975	0.2552	1.45	5.08 X 10-7
rs17702306	20	896753	T	1081	498	0.9830	594	0.9478	3.18	5.30 X 10-7
rs7944308	11	30054459	G	1711	677	0.3644	1975	0.4347	0.70	7.76 X 10-7
rs143467814	15	65765671	A	1309	563	0.9610	1121	0.9844	Undefined	9.99 X 10-7
rs1316491	1	245999838	T	1711	677	0.6372	1975	0.7040	0.71	1.81 X 10-6
rs1829255	15	24315588	G	1711	677	0.6069	1975	0.5466	1.33	2.34 X 10-6
rs12199358	6	158655375	A	1711	677	0.5936	1975	0.6585	0.71	4.20 X 10-6
rs76512108	11	5613131	G	1381	585	0.7713	1143	0.7021	1.49	4.36 X 10-6
rs4634045	3	74108823	T	1711	677	0.7728	1975	0.8225	0.70	6.70 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. NPDR= non-proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, FREQ = frequency, OR = odds ratio

Supplemental Table 15. Ten independent variants with the lowest P values for the NPDR Analysis, European participants, with liability threshold modeling of duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs80117617	2	40855125	T	1122	510	0.9794	794	0.9490	Undefined	7.80 X 10-8
rs142293996	1	224448059	C	896	383	0.9935	643	0.9860	Undefined	1.03 X 10-7
rs181278228	19	10414126	C	888	380	0.9961	657	0.9856	Undefined	8.87 X 10-7
rs117958555	19	41628014	G	941	395	0.8089	843	0.7562	1.35	1.15 X 10-6
rs117241684	16	78093913	C	1247	547	0.9954	1002	0.9821	Undefined	1.66 X 10-6
rs17017690	3	316189	T	1706	677	0.9122	1975	0.8774	1.57	2.43 X 10-6
rs200183045	17	16247098	G/GC	941	395	0.9848	843	0.9751	Undefined	3.12 X 10-6
rs79120131	17	73886532	G	1014	417	0.9940	865	0.9809	Undefined	3.13 X 10-6
rs80187193	10	20038077	A	1133	514	0.9903	713	0.9740	2.35	3.64 X 10-6
rs148819275	2	171987611	ATATATC/A	1588	643	0.9038	1753	0.9304	0.67	4.08 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. NPDR= non-proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = sample size, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 16. Ten independent variants with the lowest P values for the Extremes of DR Analysis, African American participants, without accounting for duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs184340784	1	4589883	C	603	520	0.9990	230	0.9784	Undefined	3.52 X 10-8
rs4726066	7	151322272	G	690	543	0.9864	648	0.9616	Undefined	5.02 X 10-8
rs78464534	2	115991851	C	690	543	0.9836	648	0.9524	Undefined	2.64 X 10-7
rs61811867	1	154775244	C	690	543	0.9731	648	0.9385	Undefined	1.19 X 10-6
rs2064196	6	144587183	C	690	543	0.2493	648	0.3276	0.55	1.22 X 10-6
rs12447665	16	5537598	C	690	543	0.786	648	0.7271	1.83	1.65 X 10-6
rs11575234	12	56744276	C	690	543	0.4168	648	0.5038	0.57	1.97 X 10-6
rs1566115	6	99632323	G	690	543	0.3349	648	0.2632	1.87	2.12 X 10-6
rs9446832	6	73783508	T	690	543	0.9829	648	0.967	Undefined	2.28 X 10-6
rs4129798	3	18280998	A	690	543	0.7332	648	0.7991	0.55	2.47 X 10-6

DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = number of subjects, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 17. Ten independent variants with the lowest P values for the Extremes of DR Analysis, African American participants, with liability threshold modeling of duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs200295620	3	167890185	AT/A	690	543	0.9927	648	0.9715	1.39	7.06 X 10-8
rs184340784	1	4589883	C	603	520	0.999	230	0.9784	Undefined	2.97 X 10-7
rs114921230	1	226782862	G	646	531	0.9963	488	0.9806	1.90	5.37 X 10-7
rs1000708	12	60083488	C	690	543	0.2048	648	0.2706	Undefined	6.91 X 10-7
rs71354195	19	36876318	G	690	543	0.982	648	0.96	2.42	8.00 X 10-7
rs116396065	8	122129270	T	690	543	0.7778	648	0.7246	Undefined	1.16 X 10-6
rs78464534	2	115991851	C	690	543	0.9836	648	0.9524	2.14	1.33 X 10-6
rs201584991	8	125991724	CATT/C	690	543	0.8901	648	0.8485	1.63	1.47 X 10-6
rs74705672	8	133054627	A	648	532	0.987	390	0.9643	0.69	1.47 X 10-6
rs200197449	14	101156145	CTTTCTTTCTT/C	506	288	0.7446	592	0.7969	1.44	1.51 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = number of subjects, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 18. Ten independent variants with the lowest P values for the Extremes of DR Analysis, European participants without accounting for duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs140543605	7	76951640	G	797	594	0.9612	308	0.9941	0.15	8.74 X 10-7
rs76319596	18	74086357	G	797	594	0.9466	308	0.9857	0.20	9.43 X 10-7
rs9765	1	169101526	A	523	435	0.9733	187	1	Undefined	9.77 X 10-7
rs202105116	17	15899617	C/CA	797	594	0.9482	308	0.9857	0.26	1.25 X 10-6
rs58196296	3	74196589	C	797	594	0.8884	308	0.9503	0.40	1.77 X 10-6
rs2993334	13	113570995	C	797	594	0.7944	308	0.6902	1.72	4.00 X 10-6
rs2509462	18	2681138	T	797	594	0.0324	308	0.0051	6.52	4.23 X 10-6
rs147971778	1	159274231	A/AT	797	594	0.9627	308	0.9916	0.22	5.22 X 10-6
rs200035846	3	137896135	A/ATG	797	594	0.7298	308	0.8232	0.57	5.44 X 10-6
rs12326669	18	24065283	G	797	594	0.8398	308	0.7458	1.82	5.56 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = number of subjects, REF FREQ = reference allele frequency, OR = odds ratio

Supplemental Table 19. Ten independent variants with the lowest P values for the Extremes of DR Analysis, European participants with liability threshold modeling of duration of diabetes and glycemic control

					CASES		CONTROLS			
RSID	CHR	POSITION	REF*	NEFF	N	REF FREQ	N	REF FREQ	OR	P VALUE
rs142293996	1	224448059	C	523	187	0.9947	435	0.9874	2.38	2.10 X 10-9
rs17706958	3	73837141	T	797	308	0.8139	594	0.7332	1.58	3.04 X 10-8
rs80117617	2	40855125	T	797	308	0.9838	594	0.9445	3.77	4.04 X 10-8
rs181278228	19	10414126	C	523	187	0.9973	435	0.9874	4.77	2.14 X 10-7
rs17256891	3	55104767	A	797	308	0.9984	594	0.9882	3.1	2.28 X 10-7
rs188425511	14	23749992	A	523	187	0.984	435	0.9678	2.04	2.90 X 10-7
rs202105116	17	15899617	C/CA	797	308	0.9482	594	0.9857	0.26	3.44 X 10-7
rs10932347	2	211533070	A	797	308	0.0178	594	0.0455	0.33	4.22 X 10-7
rs201817087	14	67070639	CAA/C	523	187	1	435	0.9885	Undefined	4.32 X 10-7
rs12174773	6	4695787	C	797	308	0.3236	594	0.4276	0.63	9.63 X 10-7

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, N = number of subjects, FREQ = frequency, OR = odds ratio

Supplemental Table 20. Ten independent variants with the lowest P values for the any DR Analysis, Multiethnic Analysis, without accounting for covariates

RSID	CHR	POSITION	REF*	NEFF	MAF	OR	P VALUE
rs202087767	10	128244090	AAAG/A	3714	0.2361	1.31	3.25 X 10-7
rs60678552	6	116610201	C	3960	0.1303	1.41	1.31 X 10-6
rs17058810	18	73784174	C	3960	0.0367	1.84	2.29 X 10-6
rs5996194	22	42986513	T	3960	0.1662	0.73	2.50 X 10-6
rs10432638	2	24202512	C	3960	0.2652	0.78	2.60 X 10-6
rs12550405	8	39470136	T	3960	0.4288	0.80	2.74 X 10-6
rs10058442	5	7481314	C	3190	0.098	1.53	2.83 X 10-6
rs779139	5	116260113	G	3960	0.497	0.81	3.34 X 10-6
rs201856528	15	82286425	AAAG/A	2254	0.0254	0.42	3.54 X 10-6
rs118030504	12	60837470	C	2500	0.037	0.51	4.13 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, MAF = minor allele frequency, OR = odds ratio

Supplemental Table 21. Ten independent variants with the lowest P values for the PDR Analysis, Multiethnic Analysis, without accounting for covariates

RSID	CHR	POSITION	REF*	NEFF	MAF	OR	P VALUE
rs115523882	3	167876205	A	1452	0.0279	3.10	9.42 X 10-9
rs1601780	13	83557758	G	2550	0.19	0.66	1.95 X 10-7
rs78340493	4	43638196	C	1452	0.1629	1.65	3.01 X 10-7
rs149201869	2	216352519	T	1364	0.0113	5.57	8.62 X 10-7
rs141785043	2	105038245	C	2461	0.1509	0.70	1.08 X 10-6
rs7792156	7	152855875	T	2550	0.375	1.37	1.45 X 10-6
rs73347124	10	54680020	A	1452	0.0226	0.20	1.54 X 10-6
rs200035846	3	137896135	A/ATG	2272	0.124	0.65	1.73 X 10-6
rs35836657	19	12768948	T	1541	0.0828	0.56	1.86 X 10-6
rs116521242	12	68595896	G	1452	0.0528	0.43	2.12 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. PDR= Proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, MAF = minor allele frequency, OR = odds ratio

Supplemental Table 22. Ten independent variants with the lowest P values for the any DR, African Americans, Sensitivity Analysis, without accounting for covariates

RSID	CHR	POSITION	REF	NEFF	MAF	OR	P VALUE
rs115330915	5	106563985	G	1164	0.0348	0.32	5.95 X 10-8
rs4726066	7	151322272	G	1164	0.0247	3.33	1.30 X 10-7
rs184340784	1	4589883	C	641	0.0063	16.42	1.59 X 10-7
rs140572960	2	134144064	T	1164	0.0284	0.28	7.11 X 10-7
rs2104455	6	9368518	A	1164	0.226	1.58	1.67 X 10-6
rs1006396	22	30017000	G	1164	0.0485	2.42	1.69 X 10-6
rs7611735	3	89186994	A	1164	0.2381	1.58	2.57 X 10-6
rs12095420	1	226788952	C	932	0.0098	7.03	2.58 X 10-6
rs75333138	2	38320106	C	1164	0.0903	0.50	2.75 X 10-6
rs13251066	8	59417753	A	1164	0.1637	1.65	3.10 X 10-6

DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, MAF = minor allele frequency, OR = odds ratio

Supplemental Table 23. Ten independent variants with the lowest P values for the any DR Analysis, Europeans, Sensitivity Analysis, without accounting for covariates

RSID	CHR	POSITION	REF*	NEFF	MAF	OR	P VALUE
rs4704979	5	155594567	T	1959	0.2229	1.46	3.22 X 10-6
rs1508244	6	87133646	A	1520	0.0248	0.33	3.74 X 10-6
rs150775408	3	82302346	C	1959	0.0527	2.03	4.59 X 10-6
rs200568615	10	128244089	AAAC/A	1713	0.2749	1.38	5.54 X 10-6
rs76512108	11	5613131	G	1831	0.2731	1.39	6.52 X 10-6
rs113264876	7	113173455	C	1959	0.0359	Undefined	6.66 X 10-6
rs3736388	3	124443522	T	1959	0.3967	1.31	8.23 X 10-6
rs4512907	12	97556471	G	1959	0.4074	0.75	8.23 X 10-6
rs1394919	4	76827946	C	1959	0.284	0.73	8.51 X 10-6

* For insertions/deletions, the first allele listed is the reference allele and the second allele listed is the non-reference allele. DR= Diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, MAF = minor allele frequency, OR = odds ratio

Supplemental Table 24. Ten independent variants with the lowest P values for the sensitivity analysis excluding patients without DR from PDR Analysis (eg. NPDR vs PDR comparison), African Americans, with accounting for covariates

RESULTS FROM SENSITIVITY ANALYSIS: NPDR vs. PDR								RESULTS FROM MAIN PDR ANALYSIS FOR THESE SNPs		
RSID	CHR	POSITION	REF	NEFF	MAF	OR	P VALUE	NEFF	OR	P VALUE
rs7548269	1	21089472	T	151	0.0932	3.22	5.36 X 10-8	1452	1.26	6.10 X 10-2
rs185674586	5	51231691	G	80	0.0124	Undefined	1.66 X 10-7	1220	0.93	5.40 X 10-1
rs12406144	1	21347228	A	151	0.0905	3.62	1.91 X 10-7	1452	1.24	3.26 X 10-2
rs62249672	3	77293697	C	109	0.0123	Undefined	2.72 X 10-7	1406	0.78	6.14 X 10-1
rs138903571	7	75079016	C	112	0.0167	Undefined	2.75 X 10-7	800	Undefined	2.92 X 10-3
rs144802783	10	13512079	G	109	0.0182	Undefined	3.15 X 10-7	1406	1.73	1.75 X 10-2
rs2374588	7	92441941	C	151	0.2096	2.20	3.50 X 10-7	1452	0.93	2.52 X 10-1
rs61781069	1	21130735	T	151	0.081	3.17	4.06 X 10-7	1452	1.20	1.63 X 10-1
rs6662707	1	21415321	G	151	0.0789	3.12	6.12 X 10-7	1452	1.27	6.81 X 10-2
rs139690049	7	92434043	A	151	0.1283	2.58	6.49 X 10-7	1452	0.93	4.11 X 10-1

DR= Diabetic retinopathy, PDR = Proliferative diabetic retinopathy, NPDR = Non-proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, MAF = minor allele frequency, OR = odds ratio

Supplemental Table 25. Ten independent variants with the lowest P values for the sensitivity analysis excluding patients without DR from PDR Analysis (eg. NPDR vs PDR comparison), Europeans, with accounting for covariates

RESULTS FROM SENSITIVITY ANALYSIS: NPDR vs. PDR								RESULTS FROM MAIN PDR ANALYSIS FOR THESE SNPs		
RSID	CHR	POSITION	REF	NEFF	MAF	OR	P VALUE	NEFF	OR	P VALUE
rs828579	9	22166971	A	669	0.4268	1.52	2.17 X 10-8	1097	1.26	2.48 X 10-6
rs80004147	14	26897234	C	590	0.0205	Undefined	3.45 X 10-7	1018	Undefined	1.37 X 10-4
rs1087427	5	99705657	G	669	0.2396	1.80	6.96 X 10-7	1097	1.69	2.11 X 10-6
rs78271579	6	130313725	G	669	0.0802	0.45	1.13 X 10-6	1097	0.58	2.08 X 10-5
rs35174795	10	87399545	G	590	0.0441	0.22	1.61 X 10-6	1018	0.41	5.85 X 10-5
rs13187825	5	140677114	C	669	0.1963	0.61	3.75 X 10-6	1097	0.71	1.75 X 10-4
rs72742809	5	27832607	G	669	0.1997	1.72	3.84 X 10-6	1097	1.37	1.71 X 10-4
rs141100408	4	151508991	G	480	0.0201	0.27	4.26 X 10-6	907	0.88	1.47 X 10-2
rs142352402	10	109637245	G	480	0.0077	0.18	4.68 X 10-6	907	Undefined	1.19 X 10-5
rs77568641	6	68075850	A	480	0.0326	6.00	5.34 X 10-6	907	3.51	1.07 X 10-4

DR= Diabetic retinopathy, PDR = Proliferative diabetic retinopathy, NPDR = Non-proliferative diabetic retinopathy, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, MAF = minor allele frequency, OR = odds ratio

Supplemental Table 26. Case/control sample sizes and allele frequencies per cohort for top DR associations.

Analysis	RSID	CHR	Position	REF	ALT	Cohort	NCASE	NCTRL	FCASE	FCTRL
PDR AA EIG	rs115523882	3	167876205	A	G	All	1105	1119	0.9823	0.9611
						AAPDR	255	75	0.9804	0.92
						FIND-Eye	279	246	0.9857	0.9695
						JHS	12	239	1	0.9728
						MESA	11	348	1	0.9698
						WFU	548	211	0.9808	0.9384
PDR AA LTSOFT	rs115523882	3	167876205	A	G	All	1105	1119	0.9823	0.9611
						AAPDR	255	75	0.9804	0.92
						FIND-Eye	279	246	0.9857	0.9695
						JHS	12	239	1	0.9728
						MESA	11	348	1	0.9698
						WFU	548	211	0.9808	0.9384
PDR EU EIG	rs139205645	2	201949806	T	C	All	309	975	0.9725	0.9959
						AUST	187	770	0.9626	0.9955
						FIND-Eye	122	205	0.9877	0.9976
PDR EU LTSOFT	rs17791488	17	26232732	T	G	All	309	975	0.9871	0.9661
						AUST	187	770	0.9813	0.9714
						FIND-Eye	122	205	0.9959	0.9463
Extremes of DR AA EIG	rs184340784	1	4589883	C	T	All	520	230	0.999	0.9784
						AAPDR	255	56	0.998	0.9375
						FIND-Eye	265	174	1	0.9915
Extremes of DR EU LT	rs142293996	1	224448059	C	A	All	187	435	0.9947	0.9874
						AUST	187	435	0.9947	0.9874
						FIND-Eye	NA	NA	NA	NA
Extremes of DR EU LT	rs17706958	3	73837141	T	C	All	308	594	0.8139	0.7332
						AUST	187	435	0.8048	0.7322
						FIND-Eye	121	159	0.8279	0.7358
Extremes of DR EU LT	rs80117617	2	40855125	T	G	All	308	594	0.9838	0.9445
						AUST	187	435	0.9786	0.9483
						FIND-Eye	121	159	0.9918	0.934

Genotypes of QC-retained SNPs that did not survive filters in a given cohort were set to 'NA' in the given cohort.

REF = reference allele, ALT = alternative allele, NCASE = number of cases, NCTRL =number of controls; FCASE = frequency of REF allele in cases, FCTRL = frequency of REF allele in controls, AAPDR= African American Proliferative Diabetic Retinopathy Study, AUST= Australian Genetics of Diabetic Retinopathy Study, FIND-Eye = Family Study of Nephropathy and Diabetes-Eye, JHS = Jackson Heart Study, MESA = Multiethnic Study of Atherosclerosis, WFU=Wake Forest University.

Supplemental Table 27. Replication results for variants with $P < 5 \times 10^{-8}$ (traditional, nominal threshold for genome wide significance) in the discovery analysis broken down by ethnicity

Phenotype/ Discovery Population/ LT modeling	RSID	REF	All NEFF	All RAF	All OR	All P	ASA Rep NEFF	ASA Rep RAF	ASA Rep OR	ASA Rep P	HIS Rep NEFF	HIS Rep RAF	HIS Rep OR	HIS Rep P	EUR NEFF	EUR RAF	EUR Rep OR	EUR Rep P
PDR/AA/no	rs115523882	A	571	0.9975	0.20	0.13	NA	NA	NA	NA	571	0.9975	0.20	0.13	NA	NA	NA	NA
PDR/AA/yes	rs115523882	A	571	0.9975	0.20	0.18	NA	NA	NA	NA	571	0.9975	0.20	0.18	NA	NA	NA	NA
PDR/EUR/no	rs139205645	T	3431	0.9900	0.74	0.77	NA	NA	NA	NA	476	0.9962	0.99	0.59	2722	0.9889	0.72	0.87
PDR/EUR/yes	rs17791488	T	5883	0.9772	0.82	0.33	NA	NA	NA	NA	839	0.9769	1.26	0.61	4975	0.9772	0.76	0.20
Extremes of DR/AA/no	rs184340784	C	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Extremes of DR /EUR/yes	rs142293996	C	1229	0.9910	3.23	0.16	NA	NA	NA	NA	516	0.9951	6.29	0.36	654	0.9877	2.65	0.27
Extremes of DR /EUR/yes	rs17706958	T	4194	0.9828	1.28	0.02	2322	0.9828	0.94	0.85	760	0.8802	1.69	0.0009	654	0.7787	1.13	0.51
Extremes of DR /EUR/yes	rs80117617	T	3345	0.9726	1.29	0.24	NA	NA	NA	NA	760	0.9860	1.58	0.31	2578	0.9687	1.24	0.39

LT= liability threshold, RSID= rs identifier, CHR= chromosome, REF= reference allele, NEFF= Effective sample size, RAF= reference allele frequency in sample, ALL= all replication cohorts, ASA Rep= Asian replication cohorts, HIS Rep= Hispanic replication cohorts, EUR Rep = European replication cohorts, OR= odds ratio for reference allele, AA= African American, EUR = European, PDR = Proliferative diabetic retinopathy, DR = diabetic retinopathy, NA = not available

* None of the replication cohorts were able to provide data for this SNP.

Supplemental Table 28. Testing robustness of top DR associations using Fisher's exact test on aggregated or separate cohorts.

RSID	Discovery analysis	Nearest Gene	REF	gnomAD MAF§	Cohort	NCase	NCtrl	FCase	FCtrl	Fisher's exact test P-value†	OR	95% CI	Inverse variance-weighted meta-analysis P-value†	Weighted average of effect sizes (OR)‡
rs115523882	PDR, AA, EIG or LTSCORE	*GOLIM4*	A	0.032	All	1105	1119	0.9823	0.9611	**$1.83X10^{-5}$**	2.25	1.54-3.30	**$3.48X10^{-7}$**	2.87
					AAPDR	255	75	0.9804	0.92	$9.71X10^{-4}$	4.35	1.84-10.28	-	-
					FIND-Eye	279	246	0.9857	0.9695	0.091	2.16	0.91-5.14	-	-
					JHS	12	239	1	0.9728	0.552	0.75Φ	0.09-5.94	-	-
					MESA	11	348	1	0.9698	0.546	0.75Φ	0.10-5.79	-	-
					WFU	548	211	0.9808	0.9384	$8.32X10^{-5}$	3.36	1.87-6.04	-	-
rs139205645	PDR, EUR, EIG	*NDUFB3*	T	0.015	All	309	975	0.9725	0.9959	**$3.73X10^{-6}$**	0.15	0.06-0.34	**$1.72X10^{-6}$**	0.13
					AUST	187	770	0.9626	0.9955	$2.89X10^{-6}$	0.12	0.05-0.29	-	-
					FIND-Eye	122	205	0.9877	0.9976	0.149	0.2	0.02-1.90	-	-
rs17791488	PDR, EUR, LTSCORE	*NOS2 / LYRM9*	T	0.022	All	309	975	0.9871	0.9661	$5.42X10^{-3}$	2.67	1.28-5.60	$5.37X10^{-2}$	2.09
					AUST	187	770	0.9813	0.9714	0.371	1.54	0.69-3.45	-	-
					FIND-Eye	122	205	0.9959	0.9463	$3.21X10^{-4}$	13.78	1.85-102.9	-	-
rs184340784	Extremes of DR, AA, EIG	*AJAP1*	C	0.003	All	520	230	0.999	0.9784	**$5.48X10^{-5}$**	23.09	2.95-180.9	**$4.85X10^{-4}$**	14.96
					AAPDR	255	56	0.998	0.9375	$3.57X10^{-5}$	33.93	4.13-278.7	-	-
					FIND-Eye	265	174	1	0.9915	0.084	6.14Φ	0.68-55.15	-	-

rs142293996	Extremes of DR, EUR, LTSCORE	*NVL*	C	0.014	All	187	435	0.9947	0.9874	0.365	2.38	0.53-10.80	0.260	2.38
					AUST	187	435	0.9947	0.9874	0.365	2.38	0.53-10.80	-	-
rs17706958	Extremes of DR, EUR, LTSCORE	*PDZRN3*	T	0.209	All	308	594	0.8139	0.7332	**$1.52X10^{-4}$**	1.59	1.25-2.02	**$2.34X10^{-4}$**	1.57
					AUST	187	435	0.8048	0.7322	$6.36X10^{-3}$	1.51	1.12-2.03	-	-
					FIND-Eye	121	159	0.8279	0.7358	0.014	1.71	1.13-2.59	-	-
rs80117617	Extremes of DR, EUR, LTSCORE	*SLC8A1*	T	0.043	All	308	594	0.9838	0.9445	**$3.45X10^{-5}$**	3.56	1.82-6.98	**$6.43X10^{-4}$**	3.24
					AUST	187	435	0.9786	0.9483	0.014	2.5	1.16-5.35	-	-
					FIND-Eye	121	159	0.9918	0.934	$3.88X10^{-4}$	8.48	1.97-36.55	-	-

RSID= rs identifier, AA= African Americans, EUR = European, CHR= chromosome, REF= reference allele, P_LTSOFT= P-value from meta-analysis in Table 4, OR_LTSOFT= Odds ratio from meta-analysis in Table 4; NCASE= number of cases; NCTRL= number of controls; RAF= reference allele frequency, OR= odds ratio for reference allele, CI= confidence interval of OR, MAF= minor allele frequency.
* The Fisher's exact test is two-tailed.
†The weighted average of effect sizes across cohorts was computed using ln(OR) and variance of ln(OR). Variance of the ln(OR) was computed as 1/(2x(NCASExRAF_CASE)) + 1/(2x(NCASEx(1-RAF_CASE))) + 1/(2x(NCTRLxRAF_CTRL)) 1/(2x(NCTRLx(1-RAF_CTRL))).
‡OR were computed adding 1 to all four group counts, as one of the four group had 0 counts. This enabled to estimate an effect size for the meta-analysis. The inverse variance-weighted meta-analysis P-values were computed using a chi-square cumulative distribution function assuming 1 degree of freedom. P-values in bold pass Bonferroni correction (P<0.0036) correcting for 14 tests. The identity of the cohorts per variant can be found in Supplementary Table 27.

Supplemental Table 29. Replication results for variants with nominal significance (P <0.05) in the replication cohorts broken down by ethnicity.

Discovery Population/ LT modeling	RSID	REF*	ALL OR	ALL P	ASA OR	ASA P	EUR OR	EUR P	HIS OR	HIS P
Variants identified in the Any DR Discovery Analysis										
European (Sens)/no	rs1394919	C	0.91	**0.003**	0.97	0.74	0.88	**0.009**	0.91	0.06
AA (Sens)/no	rs75360147	T	2.65	**0.009**	NaN	NaN	NaN	NaN	2.65	**0.009**
European/no	rs1508244	A	0.92	**0.01**	0.95	0.14	0.74	**0.02**	0.83	0.07
ME/no	rs10432638	C	0.93	**0.01**	0.98	0.93	0.92	0.05	0.89	0.05
EU/no	rs150775408	C	1.27	**0.04**	2.57	0.17	1.21	0.15	1.19	0.30
AA/yes	rs143894698	G	1.45	**0.004**	2.33	0.20	1.40	**0.01**	1.34	0.42
European/yes	rs13006587	G	0.93	**0.006**	0.97	0.76	0.92	**0.003**	0.93	0.77
European/yes	rs73642012	C	0.90	**0.02**	0.91	0.81	0.91	**0.04**	0.87	0.11
European/yes	rs62295141	C	1.14	0.07	0.63	0.25	1.16	**0.04**	1.39	0.61
Variants identified in the PDR Discovery Analysis										
Europeans/no	rs139921826	G	0.66	**0.03**	NaN	NaN	0.66	**0.04**	0.70	0.34
Europeans/no	rs139618558	T	0.62	0.06	NaN	NaN	0.58	**0.04**	2.44	0.34
AA/no	rs1144964	G	0.99	0.61	0.85	**0.01**	1.05	0.48	1.12	0.29
AA/no	rs137949823	G	1.02	0.92	0.81	**0.03**	1.06	0.45	1.22	0.17
AA/no	rs7533141	T	1.00	0.57	1.35	**0.04**	0.94	0.55	1.13	0.23
European and ME/no	rs202105116	C/CA	1.07	0.90	NaN	NaN	1.11	0.71	0.37	**0.006**
AA/yes	rs1414474	C	1.12	**0.01**	NaN	NaN	1.11	0.06	1.23	0.07
AA/yes	rs9998354	T	0.92	**0.04**	0.84	0.08	0.94	0.33	0.88	0.16
European/yes	rs142293996	C	2.40	**0.04**	NaN	NaN	2.24	0.06	5.09	0.42
AA/yes	rs11612613	C	1.08	0.09	1.28	**0.006**	1.04	0.50	0.93	0.68
AA/yes	rs202105116	C/CA	1.00	0.69	NaN	NaN	1.14	0.57	0.37	**0.006**
Variants identified in the NPDR Discovery Analysis										
European/no	rs1508244	A	0.89	**0.005**	0.91	**0.04**	0.81	0.16	0.83	0.10
European/no	rs7944308	G	0.94	**0.02**	0.94	0.07	0.97	0.55	0.88	0.06
European/no	rs12199358	A	1.00	0.97	1.06	0.08	0.91	**0.04**	0.96	0.42
European/no	rs1316491	T	1.00	0.89	1.03	0.29	1.02	0.725	0.86	**0.02**

AA/no	rs2104455	A	0.99	0.85	1.00	0.9049	0.94	0.15	1.17	**0.03**
AA/yes	rs635762	C	0.97	0.78	0.73	**0.02**	1.05	0.53	0.99	0.72
AA/yes	rs13089163	G	1.00	0.61	1.22	0.05	0.99	0.88	0.83	**0.008**
Variants identified in the Extremes of DR Discovery Analysis										
AA/no	rs74161190	A	0.40	**0.03**	NaN	NaN	#N/A	#N/A	0.40	**0.03**
European/no	rs202105116	C/CA	0.70	0.15	NaN	NaN	1.42	0.4652	0.43	**0.03**
AA/no	rs4129798	A	0.99	0.88	1.10	0.18	#N/A	#N/A	0.79	**0.03**
European/yes	rs17706958	T	1.28	**0.02**	0.94	0.85	1.13	0.51	1.69	**0.0009**
European/yes	rs10932347	A	0.64	**0.02**	1.02	0.98	0.61	**0.01**	0.75	0.97
AA/yes	rs2690028	C	0.80	**0.03**	1.06	0.52	0.70	**0.03**	0.92	0.41
European/yes	rs116972715	C	3.62	**0.03**	NaN	NaN	4.55	**0.02**	2.90	0.50
European/yes	rs75167957	C	9.77	**0.04**	NaN	NaN	16.94	**0.01**	0.53	0.65
AA/yes	rs6577631	G	0.89	**0.04**	0.96	0.71	0.93	0.39	0.77	**0.01**
European/yes	rs117752221	T	3.45	0.13	NaN	NaN	6.58	**0.04**	0.81	0.72
European/yes	rs114080992	G	1.00	0.24	0.86	0.25	0.71	0.36	1.69	**0.0004**
AA/yes	rs17059444	C	1.14	0.15	1.84	0.42	1.12	0.57	1.40	**0.03**

* For insertions-deletion, the reference allele is shown first followed by the alternate allele

Sens= Sensitivity Analysis, ME = Multiethnic AA = African American, EUR = European, ASA = Asian, HIS = Hispanic

P-values of nominal significance (P <0.05) are bolded.

Supplemental Table 30. Gene set enrichment analysis of 41 protein-protein interconnected genes.

Case Control Definition	Population	# EXP genes/loci above 95th percentile enrichment cutoff	# OBS genes/loci above enrichment cutoff	Excess # genes/loci above enrichment cutoff (OBS-EXP)†	Enrichment fold (OBS/EXP)	Nominal MAGENTA Enrichment P-value*
PDR	AA	2	33	31	16.5	$<1x10^{-6}$
PDR	EUR	2	1	0	0.5	0.871
Extremes of DR	AA	2	8	6	4	$2x10^{-4}$
Extremes of DR	EUR	2	1	0	0.5	1

†Estimated number of genes that may be true associations with DR based on the MAGENTA analysis. None of the 'most significant variant per gene' for any of the 41 genes exceeded genome-wide significance *(P*$<5x10^{-8}$). *To correct for physical clustering along the genome of subsets of genes within the gene set, genes that shared the same most significant variant were collapsed to one effective gene for the gene set enrichment analysis. AA = African American; EUR = European; OBS = observed; EXP = expected.

Supplemental Table 31. MAGENTA gene association scores and most significant variants for 41 protein-protein interconnected genes based on African American PDR Discovery Analysis.

Gene Symbol	Entrez ID	Gene association P-value	Chr	Gene Start Pos, hg19	Gene End Pos, hg19	Gene Size kb	# Variants per Gene Region	# LD Independent Variants per Gene Region	# Recombination Hotspots per Gene Region	Most Significant (Best) Variant per Gene Region	Best Variant Chr Pos hg19	Best Variant GWAS P-value	Best Variant log(OR)	Best Variant Allele Frequency gnomAD AFR	Best Variant Allele Frequency gnomAD EUR NonFinnish
DDX27	55661	1.06 X 10-05	20	47835831	47860614	25	703	22	0	rs2426139	47780105	2.60 X 10-06	-0.2643	0.4966	0.3035
COMMD1	150684	2.32 X 10-05	2	62132802	62363205	230	1464	53	3	rs7604016	62192135	1.41 X 10-06	-0.9255	0.9954	0.9946
CSE1L	1434	3.72 X 10-05	20	47662837	47713486	51	668	30	0	rs2426125	47700982	5.94 X 10-06	-0.2459	0.5207	0.3109
IMP3	55272	6.27 X 10-05	15	75931425	75932664	1	464	14	3	rs114597692	75936154	5.37 X 10-06	-1.23	0.03127	0.0001999
HIPK3	10114	9.52 X 10-05	11	33279167	33375939	97	720	45	3	rs12293762	33404732	6.74 X 10-06	0.6488	0.04415	0.0002009
ARFGEF2	10564	2.74 X 10-04	20	47538274	47653230	115	1100	36	2	rs201466610	47638974	7.69 X 10-06	-0.2413	NA	NA
HSP90B1	7184	3.97 X 10-04	12	104324188	104341703	18	952	29	3	rs11612613	104260208	5.73 X 10-06	0.6967	0.05254	0.2606
SIN3A	25942	7.59 X 10-04	15	75661719	75748124	86	467	16	0	rs114768415	75803647	4.06 X 10-05	-1.0463	0.02832	0.0002668
POLR2E	5434	9.23 X 10-04	19	1086577	1095391	9	234	50	5	rs4147940	1066738	3.60 X 10-05	0.4863	0.06824	0.3329
TMEM62	80021	1.03 X 10-03	15	43425721	43477341	52	582	27	0	rs79649407	43315951	4.17 X 10-05	0.6893	0.02601	0.0004003
TCEB1	6921	1.52 X 10-03	8	74858633	74884346	26	912	29	3	rs79069049	74882703	1.55 X 10-05	0.8192	0.02876	0
ERP44	23071	2.37 X 10-03	9	102741462	102861330	120	744	30	0	rs138995597	102705139	5.54 X 10-05	-1.1597	0.01031	0
EIF2A	83939	2.77 X 10-03	3	150264573	150303803	39	771	26	2	rs73169821	150251045	3.85 X 10-05	-0.4267	0.1058	0.2003
CFH	3075	3.59 X 10-03	1	196621007	196716634	96	1205	37	2	rs16840419	196651745	2.98 X 10-05	-0.3582	0.2234	0.134
SELL	6402	4.35 X 10-03	1	169659805	169680843	21	1129	32	4	rs56747461	169654798	1.57 X 10-05	0.7781	0.05785	0.03848
UTP20	27340	4.43 X 10-03	12	101673904	101780397	106	1174	41	3	rs146262654	101665156	3.38 X 10-05	-1.6891	0.02431	0.0000672
PLAGL1	5325	4.47 X 10-03	6	144261436	144385735	124	1119	52	5	rs6570599	144319407	3.25 X 10-05	-0.3333	0.7662	0.6885
RBBP8	5932	5.03 X 10-03	18	20513294	20606449	93	1209	54	0	rs192950559	20548086	4.91 X 10-05	0.6514	0.02066	0
SEC61A2	55176	5.34 X 10-03	10	12171639	12211957	40	860	58	6	rs147716326	12235918	3.18 X 10-05	0.3912	0.08708	0.0002686
ALDH1B1	219	6.06 X 10-03	9	38392701	38398657	6	870	48	7	rs11791018	38406458	1.81 X 10-05	-0.5266	0.0466	0.1004
LIMS2	55679	6.24 X 10-03	2	128395995	128439360	43	985	31	2	rs13015157	128381594	4.61 X 10-05	0.4221	0.067	0.2018
CANX	821	6.76 X 10-03	5	179125929	179158639	33	960	50	3	rs13179916	179091092	4.42 X 10-05	-0.3249	0.1942	0.06963
HNRNPH1	3187	6.80 X 10-03	5	179041178	179050722	10	838	40	3	rs13179916	179091092	4.42 X 10-05	-0.3249	0.1942	0.06963
PPP2R2C	5522	7.11 X 10-03	4	6322304	6474326	152	1666	130	8	rs4247203	6351492	2.31 X 10-05	0.23	0.6022	0.7136
EEF1E1	9521	7.93 X 10-03	6	8073592	8102828	29	1050	41	4	rs60791514	8051387	3.28 X 10-05	0.308	0.2936	0.1851
CAP1	10487	9.21 X 10-03	1	40506254	40538321	32	888	32	3	rs6600316	40564905	5.80 X 10-05	0.478	0.07194	0.08137
NUP153	9972	9.69 X 10-03	6	17615268	17706818	92	1254	64	3	rs6925217	17588245	5.26 X 10-05	0.2452	0.4078	0.5126
PPT1	5538	9.74 X 10-03	1	40538381	40563142	25	948	28	2	rs6600316	40564905	5.80 X 10-05	0.478	0.07194	0.08137
SPON1	10418	1.08 X 10-02	11	13983913	14289655	306	2329	72	7	rs140991985	13952047	2.29 X 10-05	1.2544	0.01271	0
IFNG	3458	1.11 X 10-02	12	68548549	68553521	5	997	45	3	rs2193044	68534441	4.26 X 10-05	-0.5161	0.9376	0.7288
MYO1E	4643	1.36 X 10-02	15	59428562	59665071	237	2316	115	10	rs17269587	59592116	2.10 X 10-05	0.3447	0.2642	0.315
ARHGAP15	55843	1.36 X 10-02	2	143886898	144525921	639	2996	130	9	rs41429953	144089893	1.47 X 10-05	1.0488	0.01834	0.00006672
TAF4	6874	1.39 X 10-02	20	60549853	60640866	91	1421	52	2	rs6061931	60543295	5.78 X 10-05	0.7307	0.08257	0.002335
ADAMTS9	56999	1.50 X 10-02	3	64501330	64673365	172	1852	87	8	rs4401354	64649949	3.17 X 10-05	0.2934	0.5194	0.6391
IL22RA1	58985	1.98 X 10-02	1	24446260	24469611	23	915	43	7	rs34654982	24496085	5.02 X 10-05	-0.2501	0.2635	0.03499
CAP2	10486	2.41 X 10-02	6	17393735	17558023	164	1734	64	7	rs6925217	17588245	5.26 X 10-05	0.2452	0.4078	0.5126
THRB	7068	4.63 X 10-02	3	24158644	24536313	378	2915	129	10	rs73823307	24520940	4.95 X 10-05	0.9898	0.02485	0.0002002
NUP50	10762	6.27 X 10-02	22	45559725	45583890	24	1050	68	7	rs115404138	45566422	1.17 X 10-04	0.925	0.02303	0
SEC61B	10952	8.65 X 10-02	9	101984569	101992901	8	401	28	3	rs77676592	102032626	7.43 X 10-04	0.7095	0.01225	0.00006668
WFS1	7466	3.40 X 10-01	4	6271576	6304992	33	756	90	4	rs6849178	6235405	2.24 X 10-03	-0.2338	0.3538	0.3047

ATM	472	9.48 X 10-01	11	108093558	108239826	146	1292	42	0	rs181987872	108036580	2.99 X 10-02	-0.5971	0.018	0.000271

Gene association p-values were computed based on the GWAS p-value of the most significant variant per gene region (+110kb/-40kb) from the African American PDR Discovery Analysis GWAS, correcting for gene size, number of variants per gene region, number of ancestry-specific, LD independent variants per gene region, and number of recombination hotspots spanning the gene region, using step-wise multivariate linear regression analysis. Pos, position; LD, linkage disequilibrium; OR, odds ratio; AFR, African American; EUR, European; Chr, chromosome.

Supplemental Table 32. Correlation between type 2 diabetes and glycemic trait variants and any diabetic retinopathy case status.

Trait	R	Z score	P value
Type 2 diabetes	+0.0458	+2.256	0.024
Fasting glucose	+0.0099	+0.490	0.624
Fasting insulin	+0.0130	+0.638	0.523
HbA1c	-0.0109	-0.539	0.590
Pro-insulin	+0.0213	+1.051	0.293
2-hour glucose	+0.0228	+1.122	0.266

Supplemental Table 33. Results for SNPs previously associated with diabetic retinopathy in other genetic studies of diabetic retinopathy.

Study	Cases #(Definition)	Controls #(Definition)	Diabetes Type	Covariates adjusted for	Ethnicity	Top SNPs	Nearest Gene	Risk allele	Original study OR	Original Study P value	Present study case-control definition	Present study P value European	Present study P value AA
Fu et al 2010*	103 (moderate to severe NPDR and PDR)	183 (no retinopathy to early NPDR)	2	Age, sex, diabetes duration and HbA1c	Mexican American	rs6909083	*TINAG*	NA	NA	2×10^{-5}	Any DR (with covariates)	0.558	0.508
						rs17083119	*TBC1D32*	NA	NA	3×10^{-5}	Any DR	NA	0.562
						rs2300782	*CAMK4*	A	2.64	6×10^{-5}	Any DR	0.429	0.227
Huang et al 2011	174 (NPDR and PDR)	575 diabetic patients without DR and 100 nondiabetic patients	2	Diabetes duration and HbA1c	Chinese	rs17376456	*KIAA0825*	A	3.63	3×10^{-15}	Any DR	0.776	0.141
						rs2038823	*HS6ST3*	C	2.33	5×10^{-11}	Any DR	0.255	0.801
						rs4838605	*ARHGAP22*	C	1.58	2×10^{-9}	Any DR	0.898	0.47
						rs12219125	*PLXDC2*	T	1.62	9×10^{-9}	Any DR	0.476	0.471
						rs4462262	*MIR3924*	C	1.54	$<9 \times 10^{-8}$	Any DR	0.567	0.104
Grassi et al 2011	973 (PDR and DME)	1856 (all others, including NPDR)	1	None	European	rs476141	*LOC339529*	A	1.37	1×10^{-7}	PDR	0.927	0.067
						rs4787008	*RBFOX1*	G	1.47	6×10^{-7}	PDR	0.63	0.124
						rs13064954	*CCNL1*	G	1.02	7×10^{-7}	PDR	0.388	0.525
Sheu et al 2013†	437 (PDR)	570 (no DR with diabetes for at least 8 years)	2	None	Chinese	rs9565164	*TBC1D4*	C	1.7	1×10^{-7}	Extremes of DR	NA	NA
						rs1399634	*LRP2*	A	1.5	2×10^{-6}	Extremes of DR	0.0124	0.781
						rs2380261	*AF279775*	T	1.5	2×10^{-6}	Extremes of DR	0.951	0.867
Awata et al 2014‡	837 (any DR)	1149 (no DR)	2	Sex, duration of diabetes and HbA1c	Japanese	rs9362054	*LINC01611*	T	1.64	1.4×10^{-7}	Any DR	0.668	0.329
Burdon et al 2016	336 (sight-threatening DR)	508 (no or minimal DR)	2	Age, sex, duration of diabetes, HbA1c, blood pressure, kidney disease	European	rs9896052	*GRB2*	A	1.67	4.7×10^{-8}	Extremes of DR	0.006	0.15
Luo et al 2013	383 (PDR)	756 (no DR)	2	BMI	European	rs7903146	*TCF7L2*	T	1.37	1.1×10^{-3}	Extremes of DR	0.815	0.971
Tong et al 2008	1618 (PDR)	954 (no DR)	1 and 2	None	European	rs1617640	*EPO*	T	1.3	5.5×10^{-11}	Extremes of DR	0.364	0.743
Sobrin et al 2011	222 (ETDRS>=14 or ETDRS >=30)	1032 (no DR)	2	None	European	rs6128	*SELP*	T	0.43	0.0001	Any DR	0.0119	0. 367
						rs6856425	*SLC26A1*	C	3.83	0.0001	Any DR	0.371	NA

DR = diabetic retinopathy; PDR = proliferative diabetic retinopathy; NPDR = non-proliferative diabetic retinopathy; ETDRS = Early Treatment Diabetic Retinopathy Study Score; BMI = body mass index; DME = diabetic macular edema; HbA1c = hemoglobin A1c; NA = not available
* Samples from this study overlap partially with the samples from the Starr County Health Studies samples which provided replication data as part of the current study.
† Samples from this study overlap partially with the samples from the Taiwan-US Diabetic Retinopathy study which provided replication data as part of the current study.
‡ Samples from this study overlap partially with the samples from the Rikagaku Kenkyusho - Institute of Physical and Chemical Research (RIKEN) study which provided replication data as part of the current study.

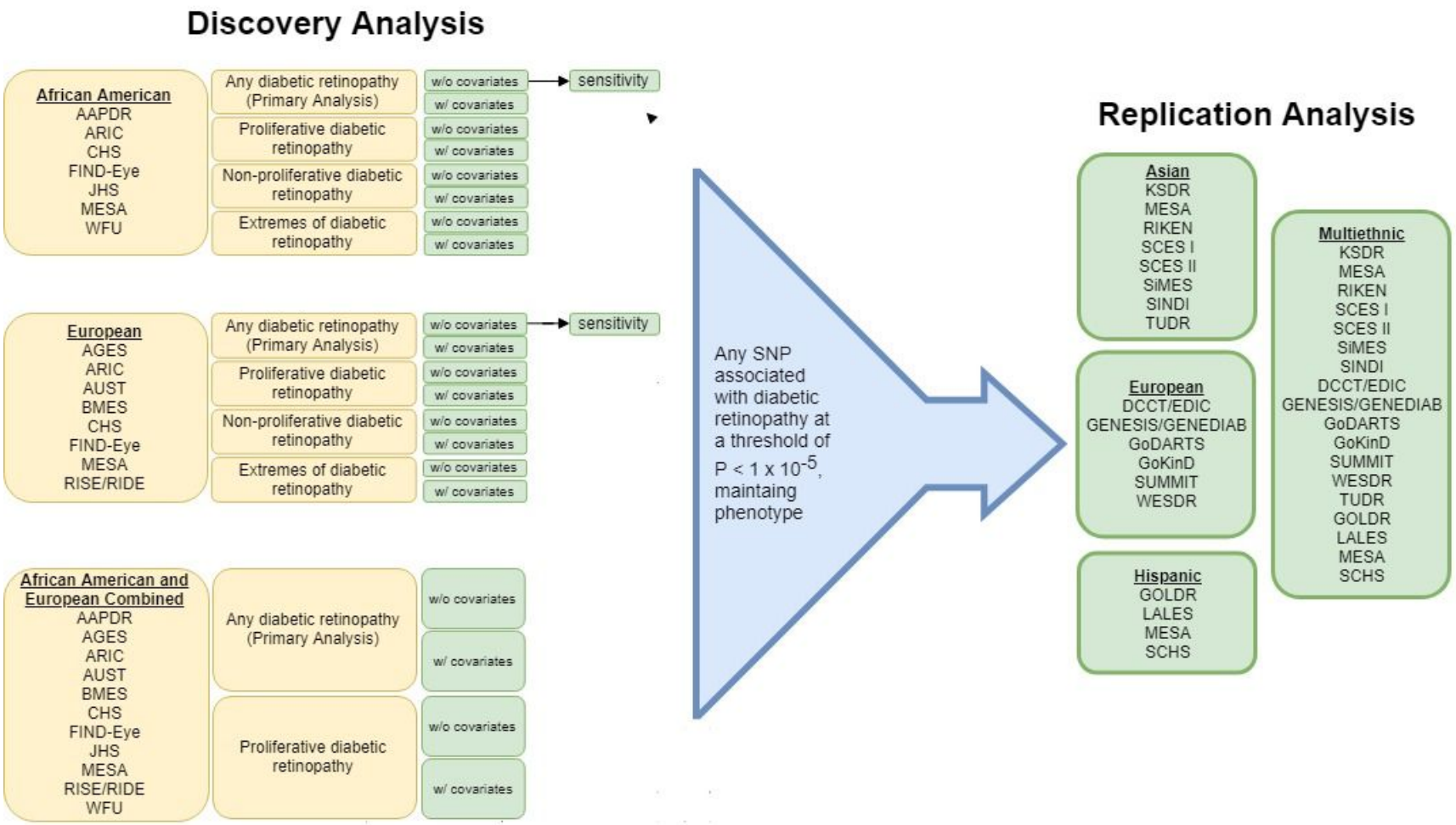
Supplemental Table 34. List of Type 2 Diabetes and Associated Glycemic Traits Variants

SNPs divided by glycemic traits									
2 hour glucose	Type 2 Diabetes			Fasting Glucose		HbA1c		Fasting Insulin*	Pro-Insulin
rs11782386	rs10923931	rs2943640	rs2447090	rs9368222	rs10830963	rs2191349	rs17533903	rs4841132	rs7903146
rs1260326	rs2075423	rs10203174	rs11651052	rs10811661	rs13179048	rs4607517	rs17747324	rs780094	rs11558471
rs6975024	rs17106184	rs243088	rs12970134	rs6048205	rs3783347	rs10830963	rs198846	rs17036328	rs11603334
rs11672660	rs11257655	rs4812829	rs10401969	rs1371614	rs16913693	rs11708067	rs2110073	rs141203811	rs4790333
rs1436958	rs12242953	rs2023681	rs8182584	rs780094	rs10885122	rs2779116	rs2375278	rs731839	rs10501320
rs1019503	rs12571751	rs1801282	rs8108269	rs2293941	rs7903146	rs552976	rs2383208	rs4646949	rs10838687
rs12255372	rs1111875	rs4402960	rs7569522	rs576674	rs3802177	rs1800562	rs2408955	rs4865796	rs4502156
rs11717195	rs2334499	rs17301514	rs13389219	rs10305492	rs11708067	rs1799884	rs267738	rs1421085	rs1549318
rs7651090	rs5215	rs6808574	rs1552224	rs6072275	rs10747083	rs16926246	rs282587	rs459193	rs6235
	rs231360	rs1496653	rs10830963	rs4607517	rs651007	rs1387153	rs3782123	rs1167800	
	rs163184	rs12497268	rs7903146	rs11605924	rs3829109	rs7998202	rs3824065	rs3822072	
	rs1061810	rs6795735	rs3802177	rs2302593		rs1046896	rs4737009	rs35767	
	rs12427353	rs6819243	rs459193	rs7944584		rs855791	rs4783565	rs974801	
	rs4275659	rs6813195		rs1483121		rs10774625	rs4820268	rs2745353	
	rs10842994	rs4458523		rs11715915		rs11154792	rs579459	rs1530559	
	rs11063069	rs702634		rs6943153		rs11224302	rs592423	rs4691380	
	rs2261181	rs6878122		rs2657879		rs11248914	rs6474359	rs10195252	
	rs7955901	rs11759026		rs174550		rs11558471	rs6980507	rs7607980	
	rs1359790	rs3130501		rs11071657		rs11603334	rs7040409	rs2785980	
	rs7177055	rs4299828		rs1552224		rs11619319	rs7616006	rs2943634	
	rs11634397	rs3734621		rs7708285		rs11964178	rs7756992		
	rs2007084	rs9505118		rs10830963		rs12621844	rs8192675		
	rs12899811	rs17867832		rs13179048		rs13134327	rs837763		
	rs9936385	rs11717195		rs3783347		rs1467311	rs857691		
	rs7202877	rs10811661		rs560887		rs1558902	rs9604573		
	rs2925979	rs7756992		rs11920090		rs174577	rs9818758		
	rs780094	rs4502156		rs340874		rs17509001	rs9914988		

*adjusted for body mass index
HbA1c = hemoglobin A1c

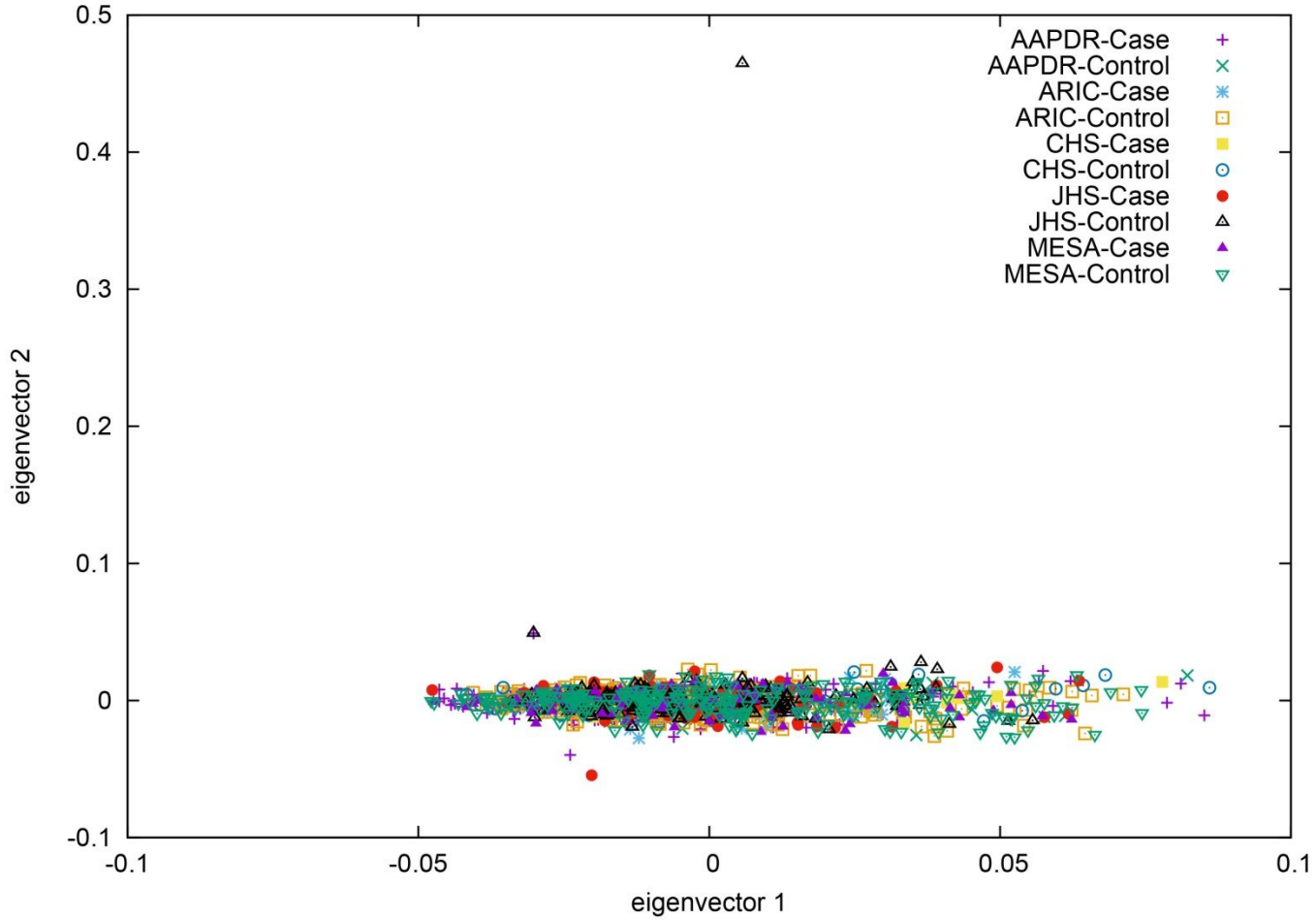
Supplemental Figures

Supplemental Figure 1. Summary of analysis performed. Green boxes represent final analyses and results can be found in main tables and Supplemental tables.

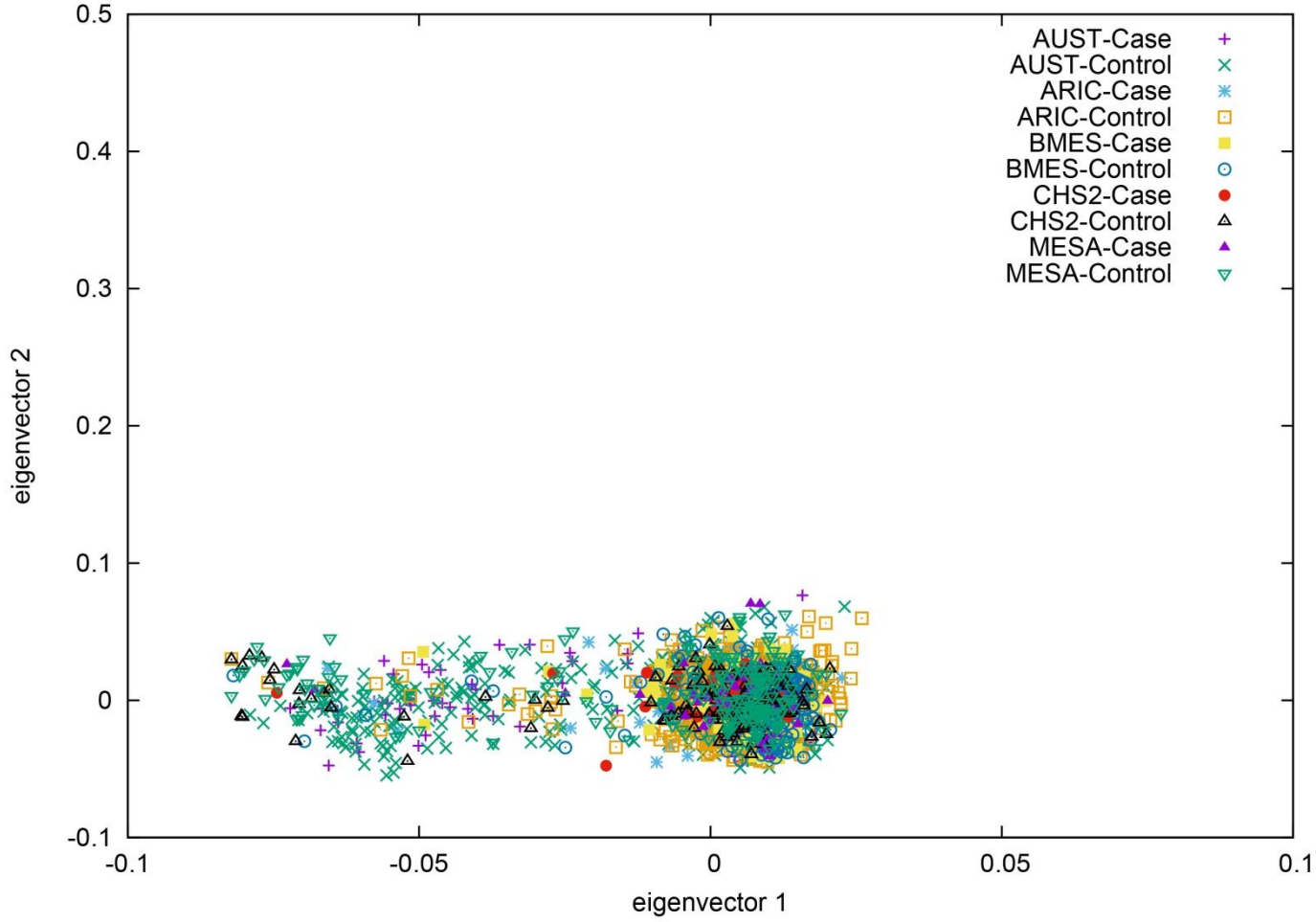


Supplemental Figure 2. Principal components analysis results for cohorts included in the primary (any DR) discovery meta-analysis

A. African Americans

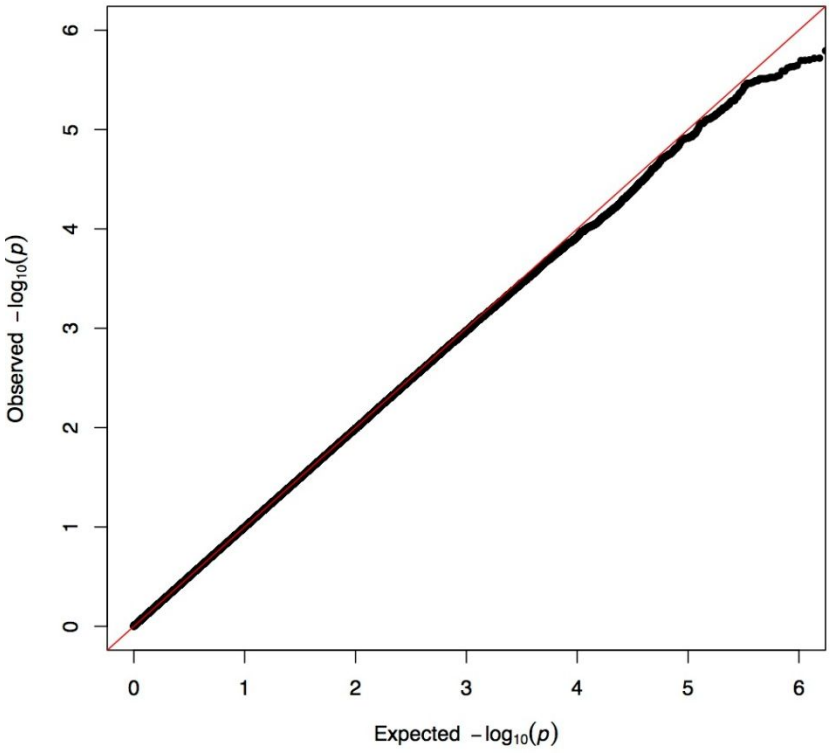


B. Europeans

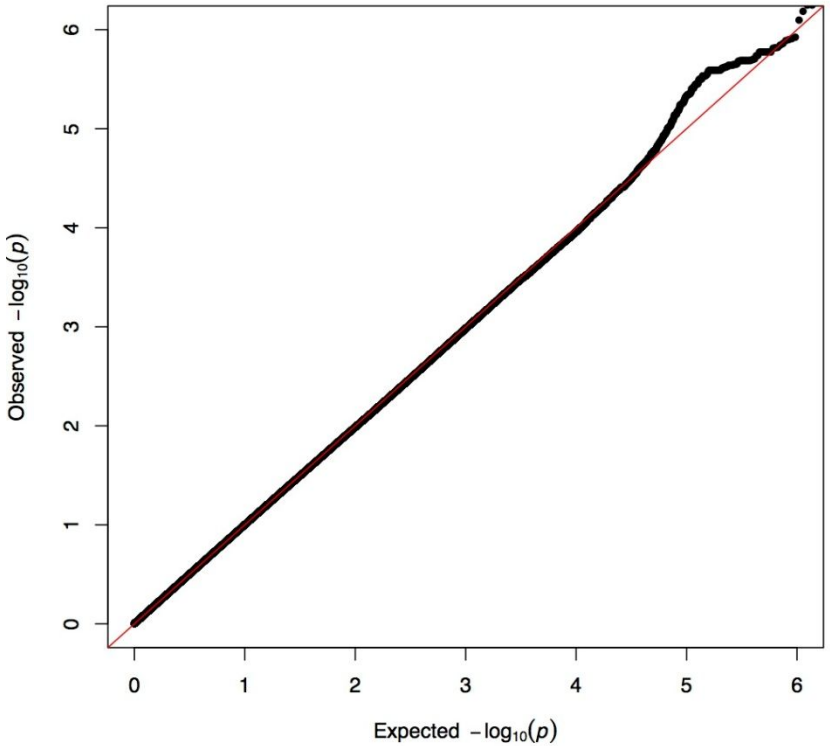

AUST-Case
AUST-Control
ARIC-Case
ARIC-Control
BMES-Case
BMES-Control
CHS2-Case
CHS2-Control
MESA-Case
MESA-Control
eigenvector 2
eigenvector 1
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.1
-0.05
0
0.05
0.1

Supplemental Figure 3. Quantile-quantile plots of the results for the primary (any DR) discovery meta-analyses for (A) African American participants without accounting for duration of diabetes and glycemic control (EIGENSOFT analysis), (B) African American participants with liability threshold modeling of duration of diabetes and glycemic control (LTSCORE analysis), (C) European participants without accounting for duration of diabetes and glycemic control (EIGENSOFT analysis) and (D) European participants with liability threshold modeling of duration of diabetes and glycemic control (LTSCORE analysis). The λ_{gc} value was 1.0009 for both the EIGENSOFT and LTSCORE association analyses for African Americans. The λgc values were 1.0059 and 0.9912 for the EIGENSOFT and LTSCORE association analyses, respectively, for Europeans.

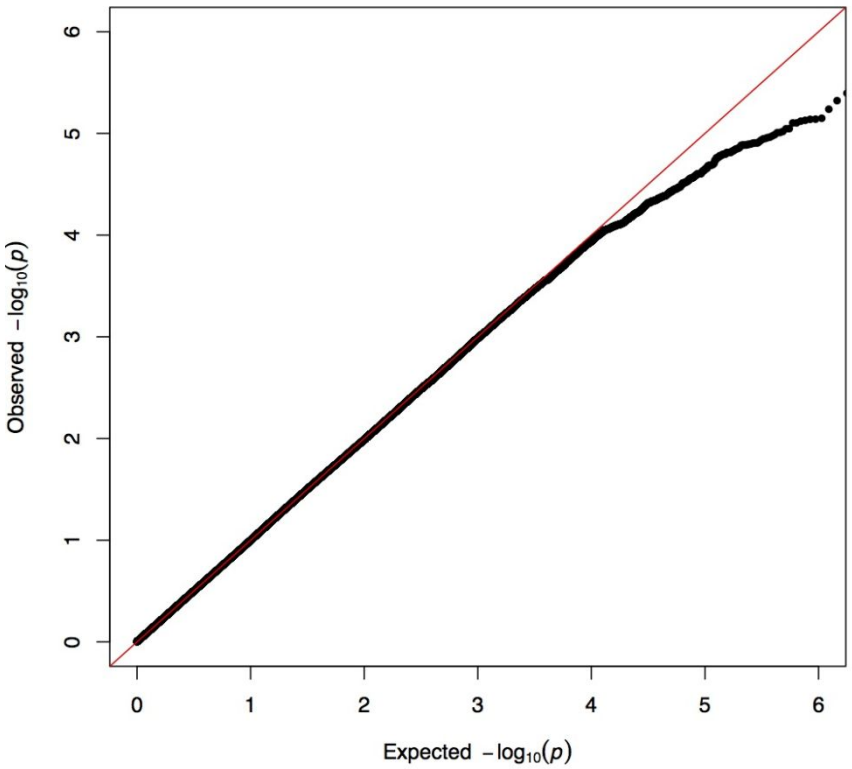
A. African American participants without accounting for duration of diabetes and glycemic control



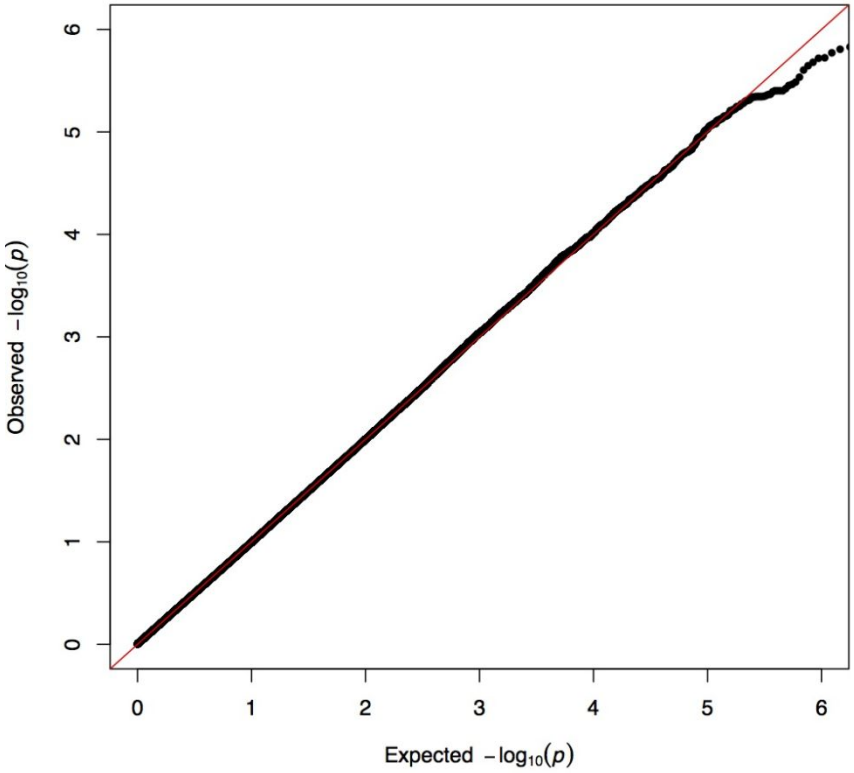
B. African American participants with liability threshold modeling of duration of diabetes and glycemic control



C. European participants without accounting for duration of diabetes and glycemic control

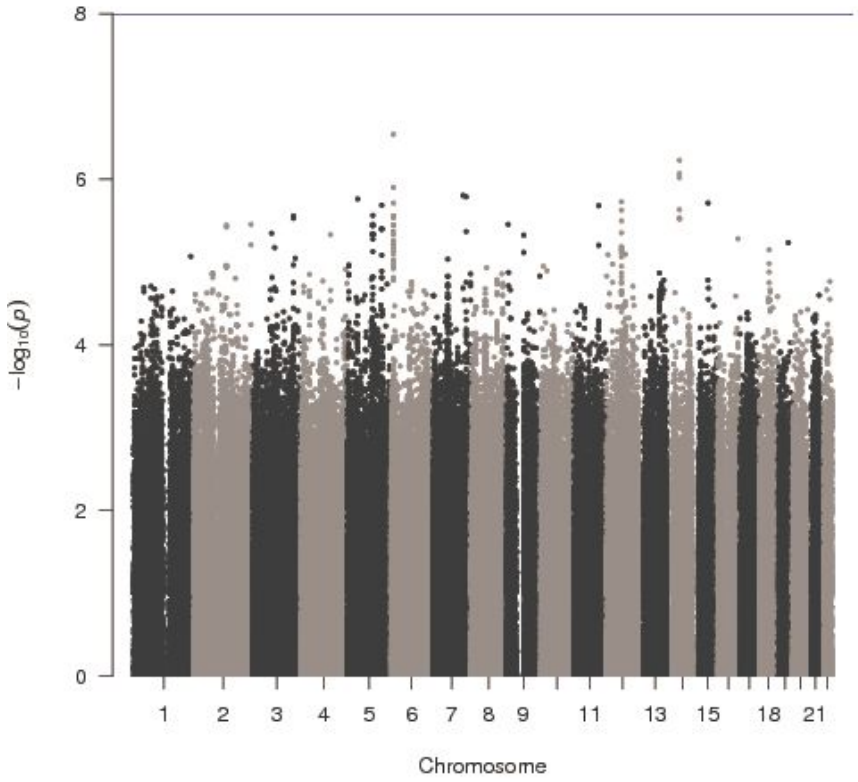


D. European participants with liability threshold modeling of duration of diabetes and glycemic control

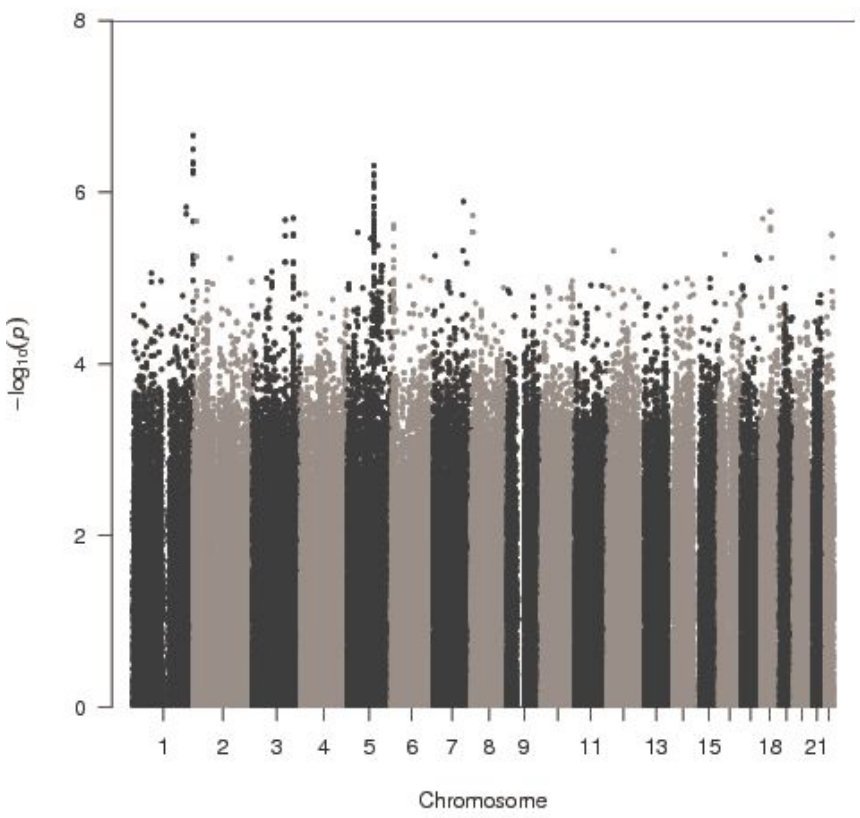


Supplemental Figure 4. Manhattan plots for the primary (any DR) discovery meta-analyses for (A) African American participants without accounting for duration of diabetes and glycemic control, (B) African American participants with liability threshold modeling of duration of diabetes and glycemic control, (C) European participants without accounting for duration of diabetes and glycemic control and (D) European participants with liability threshold modeling of duration of diabetes and glycemic control.

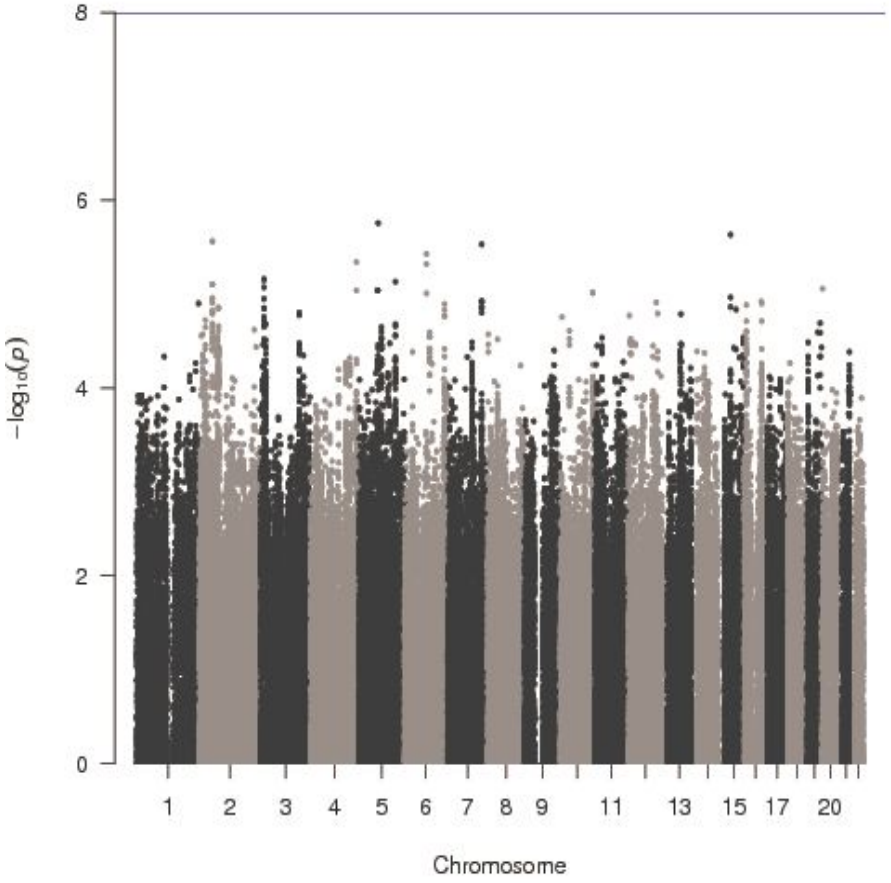
A. African American participants without accounting for duration of diabetes and glycemic control



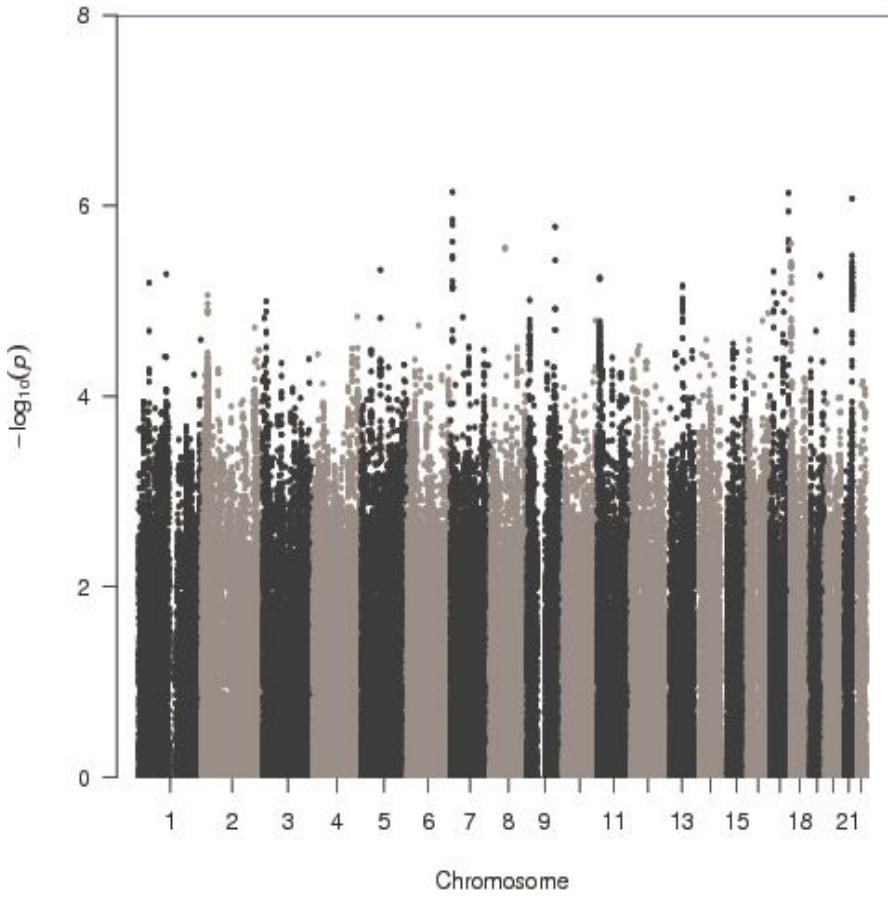
B. African American participants with liability threshold modeling of duration of diabetes and glycemic control



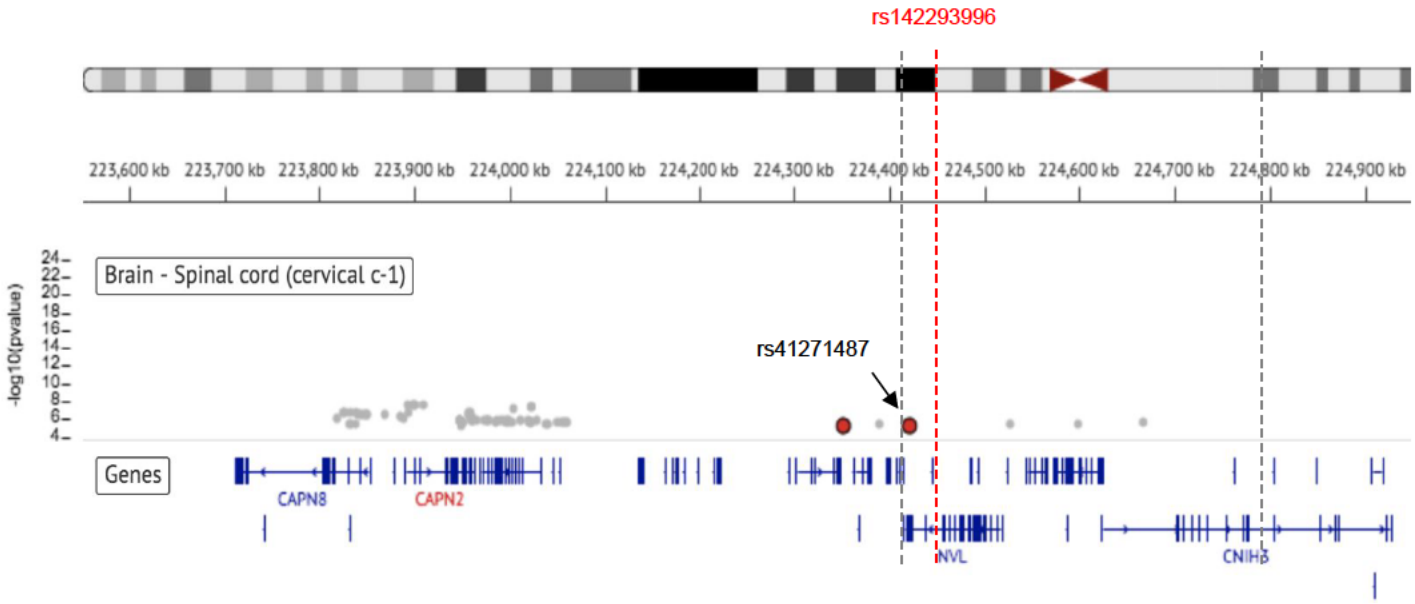
C. European participants without accounting for duration of diabetes and glycemic control



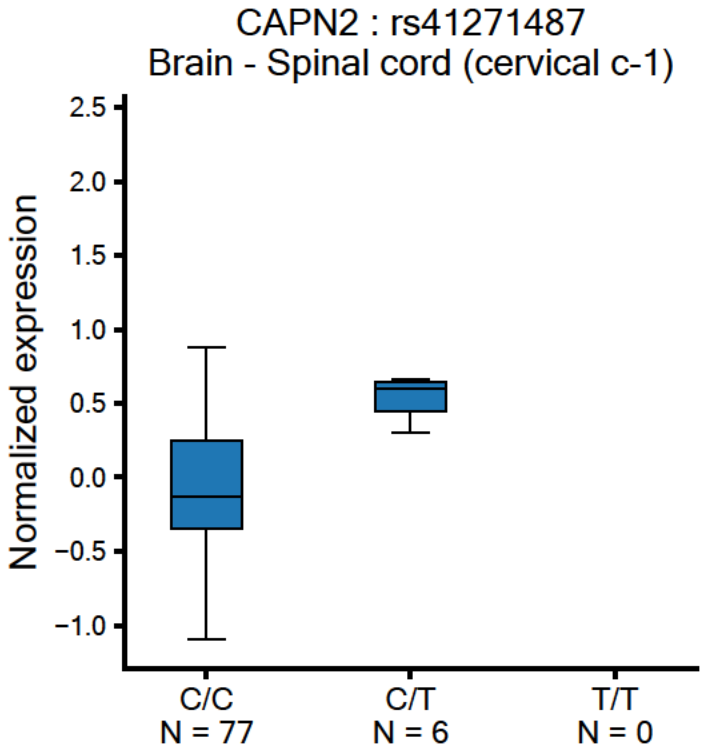
D. European participants with liability threshold modeling of duration of diabetes and glycemic control



Supplemental Figure 5: Genome-wide significant association with diabetic retinopathy in Europeans in linkage disequilibrium with an expression quantitative trait locus (eQTL) targeting CAPN2 in neuronal tissue. This plots shows –log10(P-value) of significant eQTLs [false discovery rate (FDR) ≤0.05] as a function of chromosome position in the genomic region around rs142293996, the one genome-wide significant association found in the extremes of DR analysis discovery analysis in Europeans (marked by a vertical dashed red line), plotted using the Integrated Genomics Viewer (IGV) eQTL browser from the Genotype-Tissue Expression (GTEx) portal (https://www.gtexportal.org/). The red dots represent the eQTL variants acting on CAPN2 and the gray dots represent all other variant-gene pair eQTLs in the region acting on other genes. The vertical dashed gray lines mark the r2>0.5 boundaries around rs142293996.



Supplemental Figure 6: Genome-wide significant association with diabetic retinopathy in Europeans in linkage disequilibrium with an eQTL targeting CAPN2 in neuronal tissue. Box plot demonstrating the correlation between rs142293996 genotype and normalized CAPN2 expression in Brain - Spinal Cord (cervical c-1) (n=83 samples), corrected for covariates used in cis-expression quantitative trait loci (eQTL) analysis release v7 (see https://www.gtexportal.org/). Expression values genome-wide were normalized between samples using trimmed mean of M-values (TMM) implemented in edgeR(1), followed by per-gene normalization across samples using inverse normal transformation. Boxes depict the interquartile range, whiskers 1.5x the interquartile range, and center lines represent the median. The minor T allele, associated with higher expression levels, is associated with decreased DR risk.



Supplemental Figure 7. Distribution of proliferative diabetic retinopathy (PDR) gene association *P* values for DAPPLE significant genes in African Americans (AA) and Europeans. A noncumulative distribution of gene association *P* values, computed with MAGENTA, is shown for 41 genes in top ranked AA PDR Analysis GWAS loci, enriched for protein-protein interactions based on the DAPPLE analysis. The enrichment of top ranked gene *P* values using the AA PDR GWAS (line with arrow) is contrasted against the more uniform distribution of gene *P* values computed with the European (EUR) PDR GWAS (line with boxes). A bin of 0.1 was used**.** The dashed line marks the 95th percentile enrichment cutoff used by MAGENTA in the AA PDR analysis. The vertical lines in the bottom two tracks correspond to the individual gene *P* values based on AA or EUR PDR GWAS.