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

Family Invest Nephropathy Diabet-E

2019-02-01

Family Invest Nephropathy Diabet-E , DCCT EDIC Res Grp , Pollack , S , Groop , L , Toppila , I , Sandholm , N , Groop , P-H & Sobrin , L 2019 , ' Multiethnic Genome-Wide Association Study of Diabetic Retinopathy Using Liability Threshold Modeling of Duration of Diabetes and Glycemic Control ' , Diabetes , vol. 68 , no. 2 , pp. 441-456 . https://doi.org/10.2337/db18-0567

http://hdl.handle.net/10138/311416
https://doi.org/10.2337/db18-0567

acceptedVersion

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[13ǂ], 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], Michel Marre[39,40,41], David-Alexandre Tregouet[42,43], Roberta Mckean-Cowdin[17,44], Rohit Varma[17,44], Mark I. McCarthy[45,46,47], Leif Groop[48], Emma Ahlqvist[48], Valeriya Lyssenko[48,49], Elisabet Agardh[48], Andrew Morris[50], Alex S.F. Doney[51], Helen M. Colhoun[52], Iiro Toppila[53,54,55], Niina Sandholm[53,54,55], Per-Henrik Groop[53,54,55,56], Shiro Maeda[14, 15, 16], Craig L. Hanis[18], Alan Penman[57], Ching J. Chen[58], Heather Hancock[58ǂ], Paul Mitchell[59], Jamie E. Craig[60], Emily Y. Chew[61], Andrew D. Paterson[62,63,64], Michael A. Grassi[65,66], Colin Palmer[67], Donald W. Bowden[7,8], Brian L. Yaspan[68], David Siscovick[69], Mary Frances Cotch[61], Jie Jin Wang[5, 59], Kathryn P. Burdon[70], Tien Y. Wong[5,71], Barbara E. K. Klein[72], Ronald Klein[72], 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 ǂCurrently 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

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. Université Paris Diderot, Sorbonne Paris Cité, Paris, France
40. Diabetology, Endocrinology and Nutrition Department, DHU FIRE, Bichat Hospital, AP-HP, Paris, France
41. INSERM U1138, Centre de Recherche des Cordeliers, Paris, France
42. Sorbonne Université, UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Paris, France

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

68. Genentech Inc., South San Francisco, CA
69. Institute for Urban Health, New York Academy of Medicine, New York City, New York
70. Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia
71. Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
72. 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: 3996
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 for association. 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 in discovery and replication, (2) incorporating DoD and glycemic control via LT modeling and (3) employing techniques to collectively examine 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 (HbA_{1C} or FPG) and DoD.

Statistical Analyses

The genotyping platform and the number of single nucleotide polymorphisms (SNPs) genotyped are summarized in Supplemental Table 2 by cohort. 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 all 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

Twenty replication cohorts (eight European, eight Asian and four Hispanic) 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) 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.(36) It has been shown that top associated loci, despite not being genome-wide significant, tend to cluster in biological networks.(36, 37) 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.(36-38)

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) (39) 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 statistical inflation in the distribution of the association statistics (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.(40) 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 rs41271487in the 24th intron of *NVL*. Rs41271487 is a significant eQTL (P = 6.4 X 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 (41), 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 X 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. This 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.(42) *IL22RA1* encodes the IL-22 receptor and *CFH* encodes complement factor H; both proteins are suspected to play a role in PDR.(43, 44) *SELL* encodes L-selectin, which is expressed at higher levels in lymphocytes from DR patients and associated with increased endothelial adhesion.(45) 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 risk for type 2 diabetes 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 GWAS or large candidate gene studies for DR (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 independent 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 gene set enrichment analysis 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 the a transcription factor.(46) The variant in *GOLIM4* 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 incorporation of covariates. 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.(47) 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 gene set enrichment analyses 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 type 2 diabetes patients.(48) 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.(49)

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 discovery and replication cohorts, they 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 $< 5 \times 10^{-8}$, 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 HHSN2682017000001I, HHSN2682017000002I, HHSN2682017000003I, HHSN2682017000004I and HHSN2682017000005I). 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, HHSN2682008000007C, 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 HHSN2682015000003I, 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. 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; In Press

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. 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

37. 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

38. 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

39. 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

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

41. 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

42. 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

43. 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

44. 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

45. 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

46. 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

47. 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

48. 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

49. 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 $P < 5 \times 10^{-8}$ (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	9.42×10^{-9}	3.10	571	0.9975	0.20	0.13	2.89 (1.97, 4.23)	8.51×10^{-8}
AA/yes	rs115523882	*GOLIM4*	A	1452	0.9721	5.37×10^{-9}	3.10	571	0.9975	0.20	0.18	2.89 (1.99, 4.20)	4.25×10^{-8}
European/no	rs139205645	*NDUFB3*	T	907	0.9907	3.93×10^{-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	7.26×10^{-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	3.52×10^{-8}	NA	*	*	*	*	--	--
European/yes	rs142293996	*NVL*	C	523	0.9895	2.10×10^{-9}	2.38	1229	0.9910	3.23	0.16	2.91 (1.85, 4.57)	4.10×10^{-6}
European/yes	rs17706958	*PDZRN3*	T	797	0.7615	3.04×10^{-8}	1.58	4194	0.9828	1.28	0.02	1.39 (1.24, 1.56)	7.41×10^{-8}
European/yes	rs80117617	*SLC8A1*	T	797	0.9598	4.04×10^{-8}	3.78	3345	0.9726	1.29	0.24	1.71 (1.30, 2.25)	1.35×10^{-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

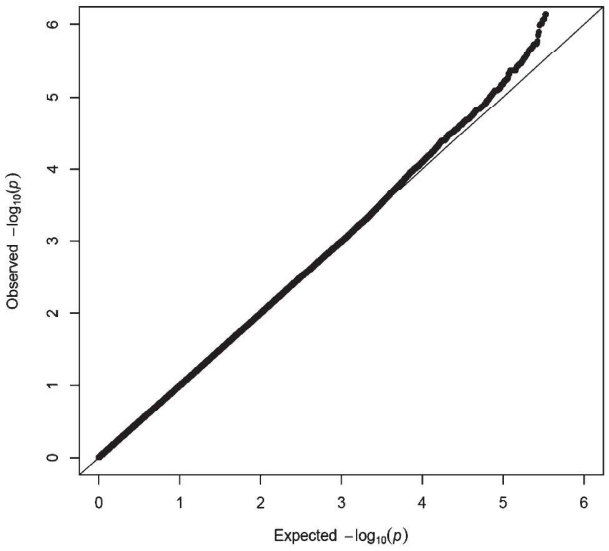
Sens= Sensitivity Analysis, ME = Multiethnic, AA = African American, DR = diabetic retinopathy, PDR = proliferative diabetic retinopathy, NPDR = non-proliferative diabetic retinopathy. * For insertions-deletion, the reference allele is shown first followed by the alternate allele

Figure Legends

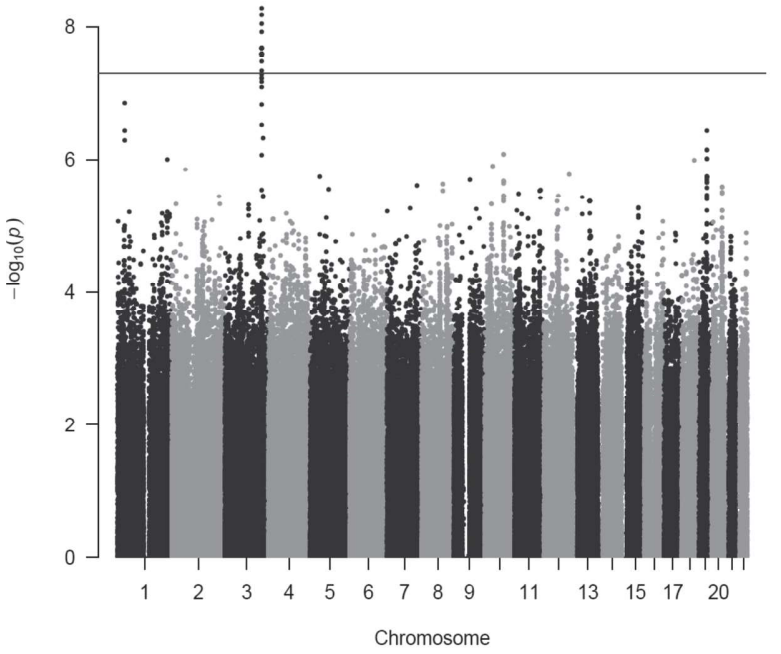
Figure 1. Quantile-quantile and Manhattan plots for the PDR and extremes of DR discovery meta-analyses for: (A and B) PDR analysis in African American participants with liability threshold modeling of duration of diabetes and glycemic control; (C and D) PDR analysis in European participants with liability 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*).

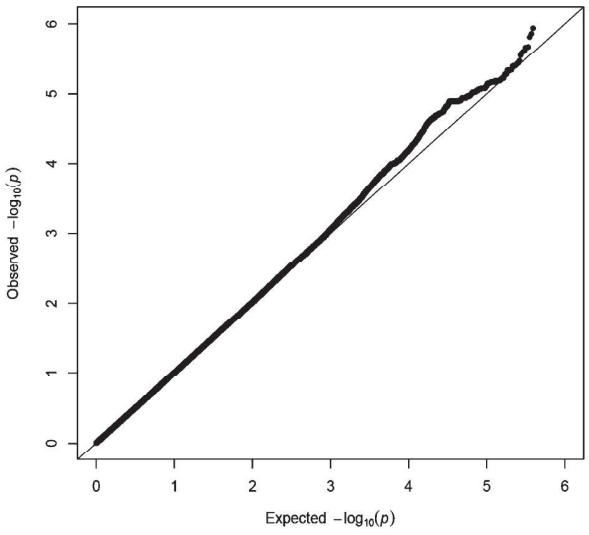
A.


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

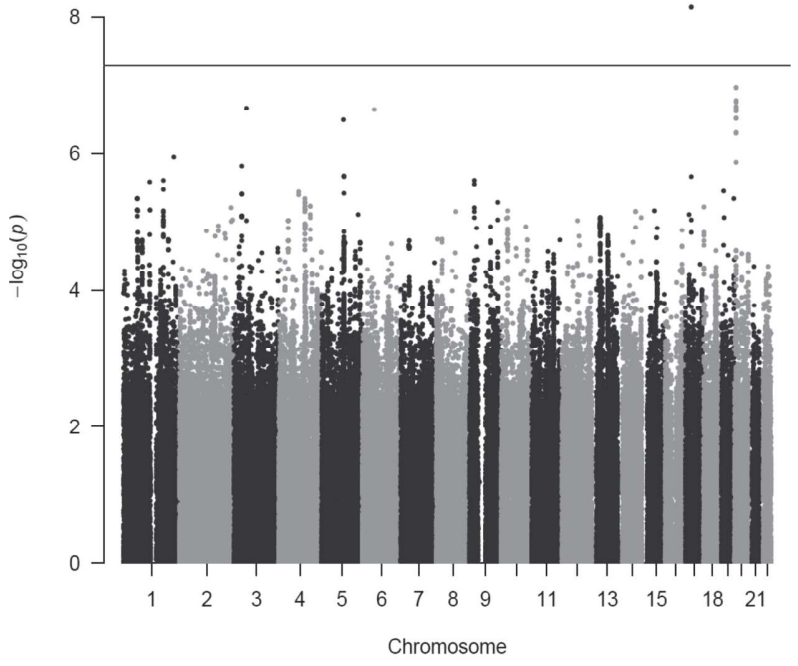
B.


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

C.

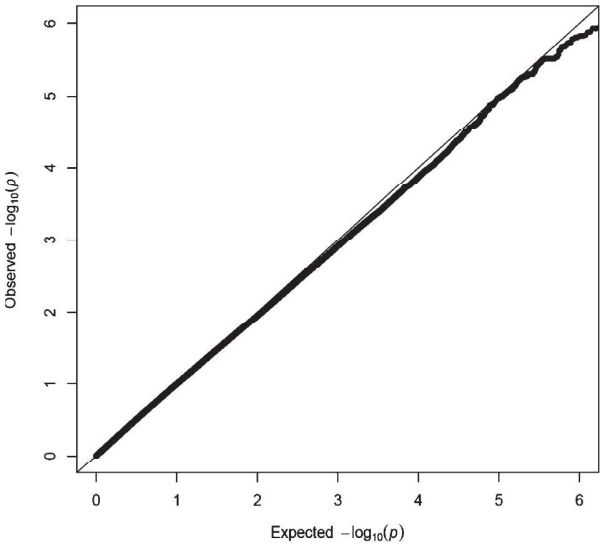

Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
1
2
3
4
5
6

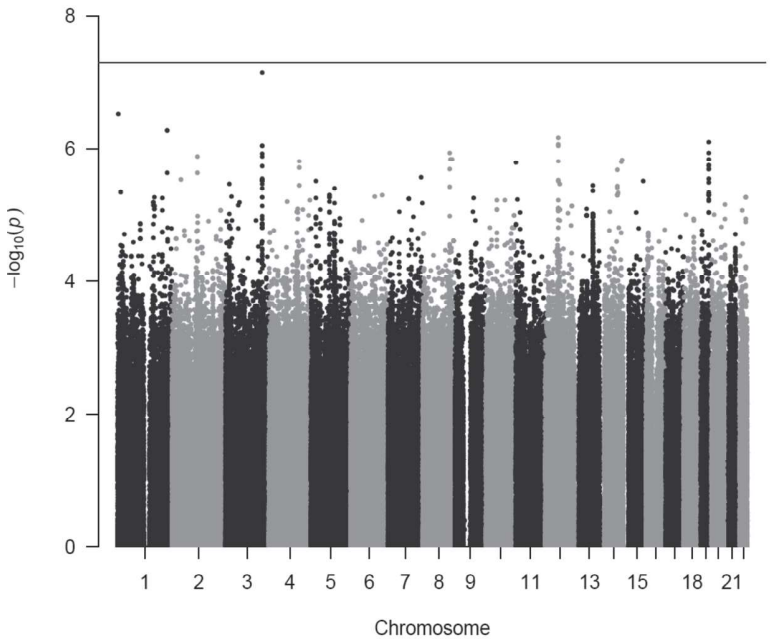
D.


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

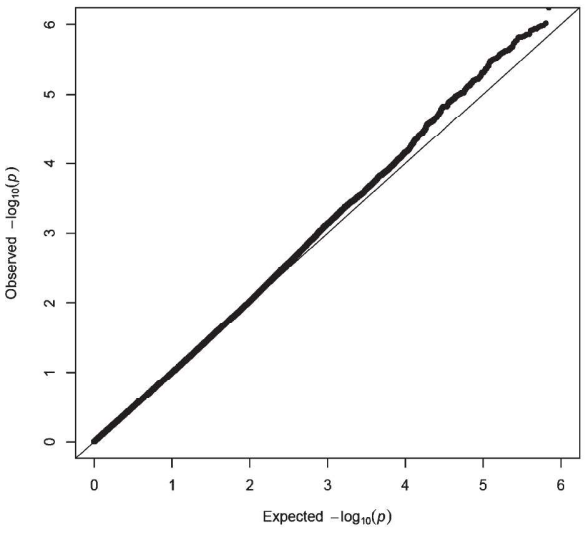
E.

F.

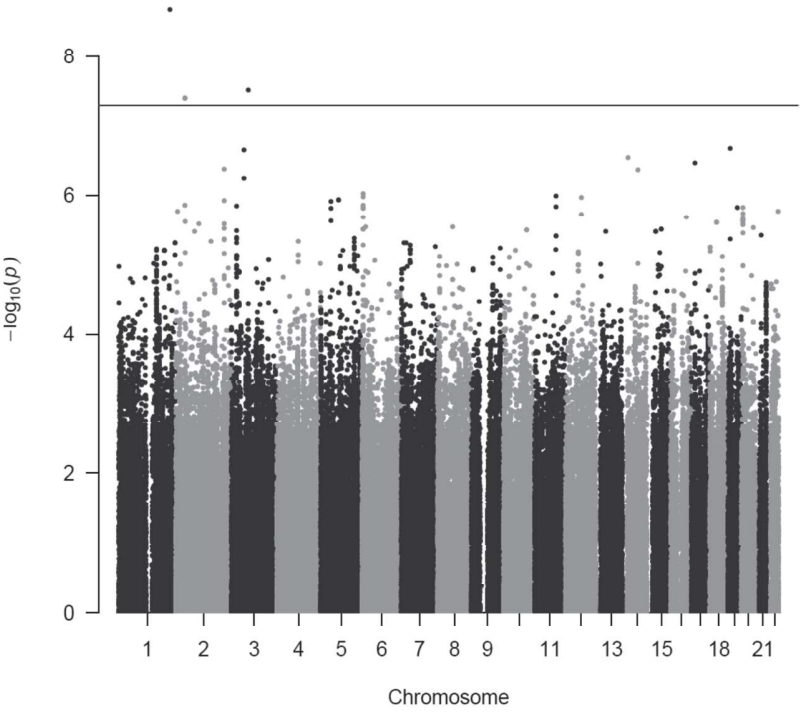

Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0 1 2 3 4 5 6

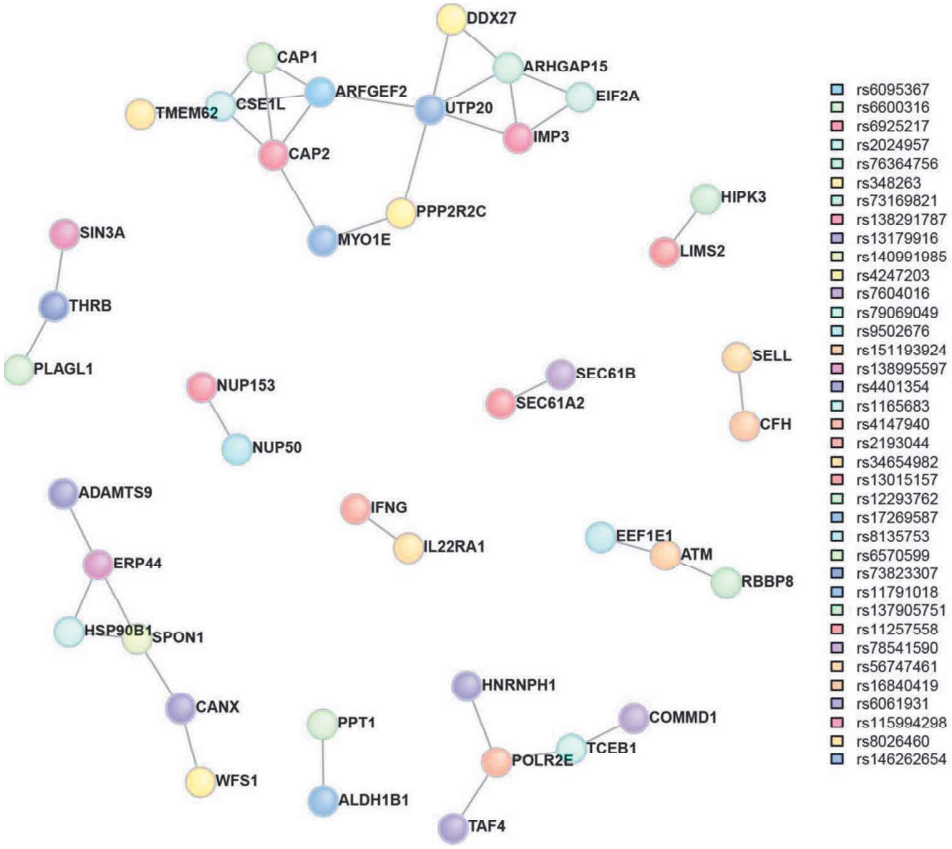

$-\log_{10}(p)$
0 2 4 6 8
1 2 3 4 5 6 7 8 9 11 13 15 18 21
Chromosome

G.


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

H.


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



132x114mm (220 x 220 DPI)