**Genetic epidemiology of diabetic retinopathy**

Lisa Tom, Samaneh Davoudi, Lucia Sobrin

Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

**Contributions:** (I) Conception and design: All authors; (II) Administrative support: All authors; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

**Correspondence to:** Lucia Sobrin, MD, MPH. Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, 243 Charles Street, 12th floor, 1252E, Boston, MA 02114, USA. Email: Lucia_Sobrin@meei.harvard.edu.

**Abstract:** The disease burden of diabetic retinopathy (DR) is tremendous around the world. While DR is correlated with hemoglobin A1c (HbA1c) and duration of diabetes, genetic differences likely account for variation in susceptibility to DR. DR is a polygenic disorder with demonstrated heritability. However, linkage and admixture analyses, candidate gene association studies, and genome-wide association studies (GWAS) have not identified many loci for DR that can be consistently replicated. Larger, collaborative, multi-ethnic GWAS are needed to identify common variants with small effects. Rigorous defining of controls groups as patients with a long duration of diabetes without DR, and case groups as patients with severe DR will also aid in finding genes associated with DR. Replication in independent cohorts will be key to establishing associated loci for DR. Investigations of mitochondrial DNA and epigenetics in DR are ongoing. Whole exome sequencing presents new opportunities to identify rare variants that might be implicated in DR development. Continued research in the genetic epidemiology of DR is needed, with the potential to elucidate pathogenesis and treatment of an important disease.

**Keywords:** Diabetic retinopathy (DR); genetics; genome-wide association studies; genetic association studies; diabetes complications

Received: 28 May 2017; Accepted: 12 July 2017; Published: 03 August 2017.

doi: 10.21037/aes.2017.07.04

View this article at: http://dx.doi.org/10.21037/aes.2017.07.04

**Introduction**

There are an estimated 93 million people with diabetic retinopathy (DR) and 17 million with proliferative diabetic retinopathy (PDR) worldwide (1). Duration of type 1 diabetes (T1D) and type 2 diabetes (T2D) and hyperglycemia, as measured by hemoglobin A1c (HbA1c), are strongly associated risk factors for retinopathy (2,3). However, as little as approximately 11% of variation in retinopathy risk is attributed to HbA1c and duration of disease (4). Transient hyperglycemia, hypertension, hyperlipidemia, and other environmental and genetic factors have been proposed as additional determinants of the development of DR (4-6). Genetic susceptibility may explain much of the heterogeneity in DR among patients with similar glycemic exposure.

Linkage disequilibrium (LD) studies, candidate gene association studies, admixture analysis, and genome wide association studies have been conducted to try to elucidate the genetic factors that influence progression of DR. These studies have been limited by insufficient sample size, lack of replication, and variation in how case/control groups are defined (7,8). Mitochondrial DNA and epigenetic changes, including miRNA, in DR have also begun to be examined (9-12). This review will summarize the current understanding of the genetics of DR, highlighting key findings and future directions of investigation.

**Heritability and LD**

Studies of twins and families first established a genetic basis for DR. In the 1970’s and 80’s, studies of monozygotic
Candidate gene association studies

Population-based studies that seek to identify common genetic variants are more promising for polygenic diseases like DR. Candidate gene association studies investigate whether a variant of a gene with a hypothesized role in the disease is significantly more common in a group of cases vs. controls. The ideal association study for DR has several characteristics. First, the presence of DR should be determined with fundus photographs and grading with a standard scale such as the Early Treatment Diabetic Retinopathy Study (ETDRS) criteria. Cases should be advanced (PDR or diabetic macular edema) because these are likely the more heritable forms of the disease. Control groups consisting of patients without diabetes can result in the identification of genes that contribute to diabetes in general rather than DR specifically. Controls should ideally be patients who have had diabetes but no or minimal DR. It is best if the controls also have a long duration of diabetes, at least 10–15 years. This minimizes misclassification of cases as controls, as some patients with no DR and short duration of diabetes will go on to develop severe DR with longer diabetes duration. It is not yet clear if DR risk variants will differ between patients with T2D and T1D. DR has more similarities than differences clinically between T1D and T2D patients, so it is likely that there are some shared variants. However, to limit heterogeneity, many genetic studies restrict the discovery phase to one of the diabetes types. Finally, large sample sizes and independent replication cohorts are key to reliably identifying the common variants of modest effect that we hypothesize are involved in DR.

Previous reviews have summarized findings for a number of candidate DR genes (7,8,12,25,26). Genes in the renin-angiotensin system as well as vascular endothelial growth factor (VEGF), erythropoietin (EPO), transcription factor 7-like 2 (TCF7L2), aldose reductase (AKR1B1), receptor for advanced glycation end products (RAGE), nitrous oxide synthase (NOS3), methylenetetrahydrofolate reductase (MTHFR), solute carrier family 19 member 3 (SLC19A3), nuclear factor erythroid 2-like 2 (NFE2L2), CDK5 regulatory subunit associated protein 1 like 1 (CDKAL1), and complement pathway genes have been investigated (7,8,12,25–31). No consistent, rigorously replicated gene has emerged for DR (12,32), likely due to insufficient sample size, lack of comprehensive coverage of genetic variants, or incorrect hypotheses of the candidate genes involved.

The Candidate Gene Association Resource (CARe) conducted one of the best powered candidate gene studies for DR with a large sample size (n=8,040) including discovery and replication cohorts (33). This study was performed with comprehensive coverage of 2,000 genes in inflammatory, metabolic and cardiovascular pathways, with correction for multiple hypothesis testing. This strategy increased the likelihood that the correct variant would be chosen by choosing multiple genes in relevant pathways and maximizing coverage of all the variants in these genes (33). P selectin (SELP) and iduronidase (IDUA) were associated with DR in the discovery sample, but not in replication cohorts. An EPO association was consistent with initial report (34), but was not significant when corrected for multiple hypothesis testing (33).

The P selectin association was further followed up in a study of DR in the Jackson Heart Study (JHS). Among 629 African Americans with T2D in the JHS, Penman et al. showed that higher plasma P selectin levels
were associated with DR [odds ratio (OR) =1.11, 95% confidence interval (CI) =1.02–1.21, P=0.02] and PDR (OR =1.23, 95% CI =0.93–1.64, P=0.02) (35). Subjects with T2D without retinopathy were more likely to be minor allele homozygotes (TT) for the single nucleotide polymorphism (SNP) rs6128 in the P selectin gene than those with retinopathy (P=0.03) (35). This same variant in P selectin was one of those found to be associated with DR in Caucasians with the same direction of effect as in the CARe discovery sample (33,35).

Recently, Porta et al. found that variants of SLC19A3, which encodes a thiamine transporter, were associated with a decreased risk of severe DR. Thiamine regulates intracellular glucose metabolism, supporting an a priori hypothesis for involvement in DR. The study population comprised of T1D patients from the FinnDiane Study: 1,566 cases of severe DR (defined as ETDRS score ≥53 or any retinal laser treatment) and 218 control subjects with no/mild DR (ETDRS score <35, no laser treatment, and diabetes duration >20 years). Two SNPs in SLC19A3 in LD with each other (FinnDiane r²=0.93) were associated with a reduced risk of severe DR and the combined phenotype of end-stage renal disease and severe DR: rs12694743 (P=3.81×10⁻⁶, OR 0.51 (95% CI: 0.38–0.68]) and rs6713116 (P=3.15×10⁻⁶, OR 0.41 (0.28–0.60)). However, the negative association of these two SNPs with DR could not be replicated in two independent cohorts (28), the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) (3) and the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) (36). However, rs12694743 was found to be associated with a decreased risk of the combined phenotype of severe DR and end-stage renal disease in the study's meta-analysis of the FinnDiane and WESDR cohorts, even after adjustment for body mass index and HbA1c (P=2.30×10⁻⁸) (28).

**Admixture mapping**

Admixed individuals inherit chromosomal segments from two distinct continental populations that mixed relatively recently in evolutionary history, within the last 20 generations (37). African Americans are an example of an admixed population with European and African ancestry. Admixed individuals allow for detection of a variant associated with a disease if the variant differs in frequency between the ancestral populations.

Tandon et al. studied participants who self-identified as African American with T2D: 305 DR cases (ETDRS score ≥60) and 1,135 controls (ETDRS score <60) (38). The proportion of African ancestry was not associated with PDR after adjustment for clinical (duration of diabetes, HbA1c, systolic blood pressure), demographic, and socioeconomic (income, education) factors. Admixture analysis failed to identify a genome-wide significant locus (38). Larger sample sizes may be needed for admixture analysis to reveal potential loci for DR genes.

**Genome-wide association studies**

Rather than investigate hypothesized candidate genes, genome-wide association studies (GWAS) agnostically analyze SNPs across the genome that commonly vary between humans. A higher threshold of significance P<5×10⁻⁸ is the accepted standard in the field, and associations should be replicated in independent cohorts (8). In many complex diseases, including age-related macular degeneration (AMD), GWAS have successfully identified genes, providing novel opportunities for understanding pathogenesis and treatment (39-42).

GWAS for DR are summarized in Table 1. The results from these GWAS have been well-summarized in previous reviews (8,12). Two of these GWAS studies found variants that achieved genome-wide significance of P<5×10⁻⁸ in their discovery samples, but neither study corrected for the multiple genetic models tested or had independent replication of their results (12,47). One GWAS has found a genome-wide significant finding including replication in independent cohorts. Burdon et al. found an association of genome-wide significance between rs9896052 and sight-threatening DR (P=4.15×10⁻⁸) in a meta-analysis of three cohorts in the study: Caucasian patients with T2D, Indian patients with T2D, and Caucasian patients with T1D (50). The growth factor receptor bound protein 2 (GRB2) gene downstream of this locus binds phosphorylated insulin receptor substrate 1 to activate the MAPK pathway via Ras in response to insulin (51,52). This gene is also involved in VEGF signaling (52). Burdon et al. showed expression of GRB2 in the human retina and increased expression in a mouse model of retinopathy, supporting a possible association between this locus and DR (50).

Four studies have attempted to replicate reported loci associations with DR in independent cohorts (32,53-55). Grassi et al. first attempted replication of 389 putatively associated SNPs identified in the Genetics of Kidneys in Diabetes (GoKinD) and EDIC discovery cohort (44) in the
WESDR cohort in 2012. The study compared 208 cases (prior laser treatment for either PDR or diabetic macular edema) and 261 controls (all other patients in WESDR). No associations reached genome-wide significance (53).

In 2014, McAuley et al. examined 24 of the most significant SNPs previously reported by Grassi et al. (44) and Huang et al. (45) in a predominantly Caucasian population with T1D and T2D in Australia: 163 cases with severe non-proliferative diabetic retinopathy (NPDR) or PDR and 300 controls with T2D for 5 years or more with no or mild DR (54). McAuley et al. found that rs1073203 was associated in a dominant model (P=0.005). The SNP rs1073203 was first reported to be associated with severe DR (diabetic macular edema or PDR) in the GoKinD and EDIC T1D cohort (P=8.5×10⁻⁶) (44). McAuley et al. also found rs4838605 to be significant in an additive model (P=0.047). A Taiwanese GWAS for T2D had reported an association for rs4838605 with DR (P=1.87×10⁻²⁸) (45), raising the possibility that this locus may affect DR in different ethnicities (12,47,54). The multiple models tested for these variants make it difficult to determine if they truly are significant after correction for multiple hypothesis testing.

In 2015, Hosseini et al. (32) attempted replication for 90 SNPs at 34 independent loci for DR. Top reported signals (P<10⁻⁵, 54 SNPs in 11 loci) came from four previous GWAS studies (43-46) and two large candidate gene association studies with broad coverage (33,56). Hosseini et al. also used top reported signals (P<10⁻⁵, 22 SNPs in 11 loci) (32) from a GWAS of retinopathy in subjects without diabetes (57). SNPs with evidence of association with DR (P<0.05, 18 SNPs in 16 loci) from previous candidate gene association studies were also included (25,58-74). Of these, 87 SNPs (32 loci) were genotyped or imputed in Hosseini et al.’s data and suitable proxies (r²>0.9) were identified for

---

**Table 1 Published GWAS in diabetic retinopathy**

<table>
<thead>
<tr>
<th>Study</th>
<th>Diabetes type</th>
<th>Number of cases, case definition</th>
<th>Number of controls, control definition</th>
<th>Top findings</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fu et al. 2010 (43)</td>
<td>2</td>
<td>103; severe NPDR, PDR</td>
<td>183; no DR or early NPDR</td>
<td>CAMK4</td>
<td>2.6</td>
<td>1.6–4.3</td>
<td>6.0×10⁻⁶</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FMN1</td>
<td>0.3</td>
<td>0.2–0.5</td>
<td>6.2×10⁻⁶</td>
</tr>
<tr>
<td>Grassi et al. 2011 (44)</td>
<td>1</td>
<td>973; laser treatment</td>
<td>1,856; no laser treatment</td>
<td>rs476141</td>
<td>1.3</td>
<td>NR</td>
<td>1.2×10⁻⁷</td>
</tr>
<tr>
<td>Huang et al. 2011 (45)</td>
<td>2</td>
<td>174; NPDR, PDR</td>
<td>575; no DR and non-diabetics</td>
<td>ARHGAP22</td>
<td>1.6</td>
<td>1.0–2.5</td>
<td>1.9×10⁻⁹</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LPLDC2</td>
<td>1.7</td>
<td>1.1–2.7</td>
<td>3.5×10⁻⁷</td>
</tr>
<tr>
<td>Sheu et al. 2013 (46)</td>
<td>2</td>
<td>437; PDR</td>
<td>570; no DR, ≥8 years of T2D</td>
<td>LRP2-BBS5</td>
<td>1.5</td>
<td>NR</td>
<td>2.0×10⁻⁸</td>
</tr>
<tr>
<td>Lin et al. 2013 (47)</td>
<td>2</td>
<td>174; NPDR and PDR</td>
<td>574; no DR</td>
<td>TBC1D4</td>
<td>1.7</td>
<td>1.0–1.3</td>
<td>8.5×10⁻²¹</td>
</tr>
<tr>
<td>Awata et al. 2014 (48,49)</td>
<td>2</td>
<td>837; any DR</td>
<td>1,149; no DR</td>
<td>rs9362054</td>
<td>1.6</td>
<td>NR</td>
<td>1.4×10⁻⁷</td>
</tr>
<tr>
<td>Burdon et al. 2015* (50)</td>
<td>1 and 2</td>
<td>336; sight-threatening DR</td>
<td>508; no DR</td>
<td>rs3805931</td>
<td>0.5</td>
<td>0.4–0.7</td>
<td>2.7×10⁻⁷</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Did not replicate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* study pursued replication. NR, not reported; CAMK4, calcium/calmodulin dependent protein kinase IV; FMN1, formin 1; CCDC101, coiled-coil domain-containing protein 101; ARHGAP22, rho GTPase activating protein 22; PLXDC2, plexin domain containing 2; TBC1D4, TBC1 domain family member 4; LRP2-BBS5, LDL receptor related protein 2-Bardet-Biedl syndrome 5; DR, diabetic retinopathy; GWAS, genome-wide association studies; NPDR, non-proliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy; GRB2, growth factor receptor bound protein 2; T2D, type 2 diabetes.
three more SNPs. Hosseini et al. did not find any genome-wide significant associations among the 90 tested SNPs at 34 independent loci after accounting for multiple hypothesis testing (32). Peng et al. similarly did not find any genome-wide significant associations for 40 SNPs previously reported in three GWAS [Mexican-Americans (43), GoKinD and EDIC (44), and WESDR (43,44,53)] in a large study of Chinese patients with T2D (819 with DR and 1,153 without) (55).

Replication of associations with genome-wide significance has been challenging for DR. One contributing factor is that variants for DR likely have small effects, requiring larger sample sizes for detection. In many other complex diseases, sample sizes upwards of 20,000 participants have been necessary to identify replicable variants (75). Thus far, sample sizes this large have not yet been collected for DR. With larger collaborative efforts, GWAS with larger sample sizes for DR will be possible. To increase the likelihood of success of such efforts, precise phenotyping and strict definitions for cases and controls are needed.

**Mitochondrial DNA and epigenetics**

Exploration of the role of mitochondrial DNA (mtDNA) and epigenetics in DR is beginning. Mishra et al. used extended-length PCR to measure mtDNA damage in peripheral blood mtDNA in rat and mouse models of diabetes (9). Diabetic rats had a significantly reduced ratio of long to short amplicons of mtDNA and decreased mtDNA copy numbers, compared with age-matched normal rats, suggesting higher mtDNA damage in diabetes (P<0.05) (9). Lipoic acid, which prevents retinopathy in diabetic rats, decreased mtDNA damage in this study (P<0.05). Overexpression of superoxide dismutase 2 (SOD2) or suppression of matrix metalloproteinase 9 (MMP-9) prevented retinopathy in diabetic mice, and these populations of mice also did not show the increase in mtDNA seen in diabetic mice (9). Mishra et al. also measured mitochondrial DNA damage as significantly increased in diabetic patients with retinopathy (n=6) vs. without retinopathy (n=5) (P<0.05) (9).

Estopinal et al. showed that mitochondrial haplotypes are associated with severity of DR in Caucasian patients (153 with NPDR and 138 with PDR) (76). The study found that the frequency of PDR differed significantly by mitochondrial haplogroup (P=0.027). In the study, an independent cohort of Caucasian patients with DR (44 NPDR; 57 PDR) confirmed this association (P=0.0064) (76). In the combined cohort, patients from the common haplogroup H were more likely to have PDR [OR =2.0 (95% CI =1.3–3.0), P=0.0012]. Patients from haplogroup UK had a decreased risk of having PDR [OR =0.5 (95% CI =0.3–0.8, P=0.0049)]. These associations with PDR were independent of HbA1c, diabetes duration, and hypertension in multivariate logistic regression analyses haplogroup H [OR =2.1 (95% CI =1.3–3.4)]; haplogroup UK [OR =0.41 (95% CI =0.23–0.73)] (76).

Investigation into epigenetic mechanisms in DR has been underway (77). Diabetic eyes have been shown to have miRNA changes including upregulation of miR200b, a VEGF-regulating miRNA (10). An increase in miR-29b is thought to be protective against apoptosis of retinal ganglion cells in streptozotocin-induced diabetic rats (11). Agardh et al. analyzed DNA methylation genome-wide in 485,577 sites in T1D patients. False discovery rate analysis was used to account for multiple hypothesis testing. Cases had PDR (n=28) and controls (n=30) were defined as having at least 10 years of diabetes with no or mild DR. The study identified differential DNA methylation at 349 CpG sites, representing 233 genes, the majority of which (79%) had decreased methylation in the PDR group. The Natural Killer cell-mediated cytotoxicity pathway was significantly (P=0.006) enriched among these differentially methylated genes (78).

**Conclusions**

Current understanding of the genetics of DR is incomplete. Linkage studies, candidate gene association studies, admixture analysis, and GWAS have not yet achieved consistent replication of many loci or genes associated with DR.

GWAS will require larger international collaborative efforts to assemble multi-ethnic cohorts and increase sample sizes. To maximize the chances of identifying true variants, cases should be defined as those with PDR or diabetic macular edema determined from a standard grading scale of imaging, while control groups should be defined rigorously as patients without DR despite a long duration of diabetes (15–20 years). Patients with mild DR are sometimes classified as cases (48) and sometimes as controls (43,79), biasing the field towards null findings. Future studies also need to correctly account for glycemic control, which is strongly correlated with DR. Replication in independent data sets will also be key to strengthening future GWAS findings (8,12). In addition, examination of rare variation and DR risk has not yet been attempted. Whole exome
sequencing may reveal associated rare variants, particularly if the extremes of phenotype are examined. Epigenetics and mitochondrial DNA also deserve further investigation. Ongoing research in these areas and large collaborations for GWAS have the potential to illuminate the genetic foundations of DR.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References


52. Anselmi F, Orlandini M, Rocchigiani M, et al. c-ABL


doi: 10.21037/aes.2017.07.04
Cite this article as: Tom L, Davoudi S, Sobrin L. Genetic epidemiology of diabetic retinopathy. Ann Eye Sci 2017;2:56.