Review Article

Progress and challenges in genome-wide studies to understand the genetics of diabetic retinopathy

Kathryn P. Burdon, Bennet J. McComish, Jac C. Charlesworth

Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia

Contributions: (I) Conception and design: KP Burdon; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: KP Burdon; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: A/Prof Kathryn P. Burdon. Menzies Institute for Medical Research, University of Tasmania, 17 Liverpool St, Hobart, TAS 7000, Australia. Email: Kathryn.burdon@utas.edu.au.

Abstract: There are many advantages to understanding the genetics of human disease. Genetic markers can be used to calculate the risk of developing a disease, and elucidation of genetic risk factors can pinpoint the molecular aetiology of disease, which can facilitate the development of targeted therapies. Diabetic retinopathy (DR) is a common complication of diabetes that has a significant impact on quality of life. It has a clear genetic component, but determination of the genetic risk factors has proven difficult. To date, genome-wide studies for DR have been conducted on relatively small patient cohorts compared to other complex eye diseases and replication of genetic findings has been limited. The disease is highly heterogeneous, confounding attempts to classify patients into appropriate groups for genetic analysis and making direct comparisons between studies challenging. Future studies to determine the genetic causes of DR will need to focus on larger sample sizes, detailed phenotyping and appropriate classification of patients. Global co-operation and meta-analyses combining data from multiple studies will be critical to the discovery of genetic risk loci for DR.

Keywords: Genome-wide association study; diabetic complications; genetic risk factors; diabetic retinopathy (DR)

Received: 23 May 2018; Accepted: 29 July 2018; Published: 27 August 2018.
doi: 10.21037/aes.2018.08.04
View this article at: http://dx.doi.org/10.21037/aes.2018.08.04

Diabetic retinopathy (DR) is a heterogeneous disease

DR is a common microvascular complication of diabetes. It is a leading cause of blindness and visual impairment, affecting 30–40% of diabetic patients over the age of 40 years (1,2). Despite its frequency, little is known about the molecular pathogenesis of this condition. DR is a complex, progressive and heterogeneous disease. The early stages are labelled as non-proliferative DR (NPDR) and are characterised by microaneurysms along with dot and blot haemorrhages and hard exudates in the retina (3). As the disease progresses, retinal vessel occlusions lead to retinal ischaemia and nerve fibre layer infarcts or 'cotton-wool spots'. Retinal ischaemia leads to proliferative DR (PDR). This neovascularisation of the retina by poorly formed new blood vessels and subsequent complications leads to significant visual impairment and blindness (3).

Diabetic patients may also develop diabetic macular oedema (DMO), characterised by vessel leakage and the accumulation of fluid in the macular region (3). DMO can occur in patients with no other signs of DR but can also accompany both NPDR and PDR (3,4). Due to the involvement of the macula, DMO causes much of the visual impairment attributed to DR.

Risk factors for DR

There are well-described risk factors for the development of DR. The most clearly recognised are the duration of

diabetes and the degree of glycaemic control measured as glycosylated haemoglobin (HbA1c) (5). In addition, hypertension, dyslipidaemia and smoking have all been reported as risk factors for ocular diabetic complications in multiple ethnic groups across both types of diabetes (2,5-8). DR also correlates with diabetic nephropathy (9,10) and quantitative measures of renal function (11), possibly suggesting similar aetiologies for these microvascular diabetic complications. Clinical risk factors are important considerations in the monitoring and treatment of DR, but even when taken together they do not account for all disease (5-8). Genetic differences have been hypothesised to account for some of the unexplained heterogeneity. Understanding the genetic risk factors can lead to more accurate predictions of risk as well as clearer insights into the molecular pathophysiology, potentially facilitating development of novel medical therapies.

Grading of DR for genetic research

The phenotypic heterogeneity of DR and DMO makes standardised disease classification difficult. Under the hypothesis that different subtypes of DR have differing genetic aetiologies, accurate phenotyping and classification is essential so that homogeneous groups can be studied and findings can be compared between studies.

There have been several attempts to standardise phenotyping for DR for the benefit of both research and clinical practise. The earliest attempt was the Airlie House classification scheme developed in the 1960s, which graded the presence and severity of 14 lesion types each on a three-point scale (12). This proved to be too insensitive for accurate representations of severity and was modified for use in the Early Treatment in Diabetic Retinopathy Study (ETDRS) (12). The revised scale scores each quadrant of the retina on a variety of features and provides an overall score ranging from 0 to 60 indicating the severity of NPDR, with 60+ representing PDR. The scale was further modified and used by the Wisconsin Epidemiological Study of Diabetic Retinopathy (WESDR) (13). These two large studies provided a significant evidence base for the diagnosis, classification and treatment of DR, and the grading system was widely adopted by research studies for accurate phenotyping of patients.

This detailed scheme, although useful in a research setting, proved overly complex for broad adoption in clinical practise. It had too many levels and relied on fundus photography, which was not always available or practical in the clinical setting. In 2003, an international group of experts led by the American Academy of Ophthalmology developed a simpler standardised classification system (14). The interdisciplinary consensus project team agreed on a five-point scale: no apparent retinopathy, mild NPDR, moderate NPDR, severe NPDR and PDR. Clear descriptors were developed for each category, based on the ETDRS and WESDR evidence and scales, but not relying on detailed scoring of each quadrant of the fundus photograph. The scheme also provides classification for DMO (present or absent) with a more detailed scale for use when appropriate training and equipment are available for the operator. This simplification of the ETDRS scheme was named the International Clinical DR Severity (ICDRS) scale and was quickly adopted, both clinically and for research. The strength of the scheme is that patients can be classified on the basis of the clinical examination without reference to standardised fundus photographs, and that the categories are based on the well-understood ETDRS grading system.

The ETDRS scale and subsequent ICDRS scale have been broadly adopted by most research studies. This has allowed basic comparisons between studies; however, major limitations still apply in the different ways in which studies combine severity categories. Some studies group all stages of DR together (usually called 'any DR') and compare them to the No DR group. Others restrict analysis to those with severe NPDR or worse, while others use PDR alone. Controls may be classified as patients with no DR, or those with a category less severe than the case definition (e.g., mild or moderate NPDR when cases are defined as severe NPDR or PDR). In addition to this, different studies handle DMO differently. Many studies ignore it altogether under the hypothesis that it has a different aetiology than NPDR/PDR. Others group clinically significant DMO together with severe NPDR/PDR to create a 'sight-threatening DR' category under the hypothesis of overlapping genetic aetiologies for all forms of diabetic retinal disease. This need to collapse the well-defined categories into bigger groups is driven predominantly by statistical power requiring larger numbers of patients and fewer categories. As the size of research cohorts for genetic studies of DR increases, this requirement will be mitigated, allowing the full range of the DR scale to be used.

Evidence for genetic component to DR

The contribution of genetics to DR is difficult to quantify

Table 1 Studies investigating heritability and familial clustering of diabetic retinopathy

Cohort	Ethnicity	N	Diabetes type	DR phenotype classification	Odds ratio*	Broad sense heritability	Familial severity correlation	Reference
UK twins	UK	37 twin pairs	2	None, mild, severe[c]			91%	(15)
		31 twin pairs	1				68%	
DCCT	USA	219 families	1[b]	ETDRS: severe DR *vs.* none or mild DR[d]	3.12 (1.12–8.76)		0.327[i]	(16)
South Indian	Indian	322 families	2	Any DR *vs.* no DR[e]	3.37 (1.56–7.29)			(17)
Starr County	MexAm	282 families	2	ETDRS: severe *vs.* mild DR[f]	1.71 (1.03–2.84)			(18)
Pima	NatAm	211 families	2	ETDRS: all categories[g]		0.18		(19)
FinnDiane	Finnish	188 families	1	ETDRS: PDR *vs.* no DR	2.76 (1.25–6.11)	0.52		(10)
FIND-Eye[a]	USA-mixed	767 families	2	ETDRS: all categories[h]		0.27	0.1358	(9)
	MexAm	370 families	2			0.24	0.1224	

*, odds ratio for DR in family members of affected proband, adjusted for covariates (vary by study); [a], probands ascertained on the basis of diabetic nephropathy. All other studies ascertained for diabetes; [b], probands had type 1 DM, family members could have either T1 or T2; [c], severe retinopathy included maculopathy or proliferative disease; [d], severe DR = ETDRS >47 (severe NPDR or worse), CSME or laser treatment; [e], any DR = NPDR or PDR or DMO. Statistics also described for sub-classifications of NPDR ± DMO and PDR; [f], severe DR = severe NPDR or PDR, mild DR = no DR or early NPDR; [g], participants categorised as no DR, any DR, moderate NPDR and PDR; [h], participants categorised as no DR, mild NPDR, moderate NPDR, severe NPDR or PDR; [i], analysis of parent-offspring pairs. Other familial relationships were also reported. UK, United Kingdom; DCCT, Diabetes Control and Complications Trial; FinnDiane, Finnish Diabetic Nephropathy Study; FIND, Familial Investigation of Nephropathy and Diabetes; MexAm, Mexican Americans; NatAm, Native American USA; United States of America; ETDRS, Early Treatment in Diabetic Retinopathy Study standardized grading scheme for DR.

for several reasons. Heritability, or the portion of disease risk due to genetic contributions, is best estimated using families. To determine the heritability of DR, multiple family members need to first develop diabetes and then DR. Combined with the typical age of onset of both diabetes and DR, this makes ascertainment of appropriate families challenging.

The genetic contribution to DR has been investigated using a variety of study designs and statistical approaches (*Table 1*). As early as 1982, Leslie and Pike (15) described concordance for DR in twins with type 2 diabetes, with 35 out of 37 twin pairs (91%) having both twins in the same severity category for DR. They also observed concordance amongst twins with type 1 diabetes, albeit to a lesser extent, with 68% of twin pairs being in the same category. The Diabetes Control and Complications Trial (16) evaluated the families of 372 trial participants and undertook more formal analyses, showing clustering of the severity of DR in families. Of note, probands in this study had type 1 diabetes, but family members could have either type 1 or type 2. The significant familial correlation of DR regardless of the type of underlying diabetes suggests susceptibility to retinopathy in the context of diabetes, independently of the cause of diabetes itself. A study of South Indian families with at least two type 2 diabetic siblings found an increased prevalence of DR amongst siblings of affected probands, at all grades of DR, independent of hypertension, glycaemic control and duration of diabetes (17). Similarly, a study of Mexican-American families with type 2 diabetes from Texas, USA (18) and of Finnish families with type 1 diabetic probands (10) both reported significant familial clustering with an increased risk of DR (including NPDR and PDR)

Table 2 Genomic regions potentially containing DR susceptibility alleles identified by linkage studies

Cohort	Ethnicity	N families	N individuals	Resolution	Chr, position (cM), peak marker	LOD	Reference
Pima	NatAm	57	136	6.4 cm	chr3, 188cM, D3S3053-D3S2427	1.36	(21)
					chr9, 89cM, D9S1120-D9S910	1.46	
Pima	NatAm	211	607	6.4 cm	chr1, 34.2cM D1S3669	3.01	(19)
Starr County	MexAm	393	794	9.38 cm	chr3, 117cM, GATA68D03	2.41	(22)
					chr12, 15.5cM, GATA49D12	2.47	

NatAm, Native American; MexAm, Mexican American.

amongst siblings of probands with DR. The FIND-Eye (9) study of multiple ethnic groups from the USA reports a broad sense heritability of 27% in the study overall and of 24% in the subset of Mexican-American participants. Although significant, this is substantially lower than the heritability of 52% reported in the FinnDiane study of type 1 diabetic siblings (10), although this study limited the analysis to participants with proliferative DR, capturing only the extreme end of the DR phenotype. Probands in the FIND-Eye study were recruited on the basis of advanced diabetic nephropathy, whereas other studies required only diabetes in the proband. This design may have biased the outcomes towards families with a propensity for severe microvascular complications. The approach of extreme phenotype enrichment is often taken in genetic studies and can be powerful in increasing the ability of the study to identify the genetic effects (20).

Linkage studies for gene discovery in DR

Early attempts to locate genes for DR utilised the genome-wide linkage scan approach. These were low resolution scans by today's standards but were designed to identify regions of the genome shared by family members with the same disease, or 'linked' to disease. While the approach was commonly applied to mapping genes for diabetes, only three genome-wide linkage scans were reported for DR before higher resolution technologies became more common (*Table 2*). Two of these studies were in the same cohort of Pima Indians, with the second scan based on updated and more detailed retinopathy information and including more individuals. The first analysis in 57 Pima Indian families identified suggestive linkage on chromosomes 3 and 9 (21), but these loci were not consistent in the expanded study of 211 families (19). The larger analysis identified a novel locus on chromosome 1, just reaching the threshold of statistical significance of LOD >3.0. The third and largest study was in Mexican Americans from Starr County in Texas, USA, and identified suggestive linkage on chromosomes 3 and 12, but neither region overlapped with those reported in the Pima Indians (22). Although linkage analysis of sibling pairs proved a successful approach to mapping genes for single gene disorders with clear Mendelian inheritance patterns, there are very few examples of success in mapping susceptibility loci for complex disease. Linkage studies for DR have not yet identified candidate genes or risk alleles for DR. The extreme heterogeneity and small effect sizes of individual loci mean that linkage analysis is not well suited to gene mapping for complex disease, including DR. Advances in genotyping technology have since driven a move towards association analysis for gene mapping in complex disease.

Genome-wide association studies for gene discovery

More recent approaches to gene mapping for DR have focused on association statistics. In the context of genetics, this method looks for statistical differences in the frequency of genetic variants between cases and controls. The advantage of the association approach is that families are not required, but conversely very large cohorts of well-phenotyped patients and controls are necessary. The typical approach is to select single nucleotide polymorphisms (SNPs) of interest and compare the allele frequencies between cases (those with DR or one of the subtypes) and controls (typically diabetic patients without the designated subtype of DR). The ability to measure (or 'genotype') hundreds of thousands of SNPs in parallel using SNP arrays has led to the development of the genome-wide association study (GWAS). GWAS use SNP arrays to assess 'tag' SNPs throughout the genome. Tag SNPs are chosen for their ability to provide information about the genotype at nearby variants, through linkage disequilibrium. This

characteristic means that GWAS can feasibly be conducted by genotyping tag SNPs, rather than every variable position in the genome. GWAS cannot pinpoint a causative variant, but can flag the region where one exists, in linkage disequilibrium with the genotyped tag SNP. This approach has been highly successful for mapping genetic risk loci for many ophthalmic diseases, including most notably age-related macular degeneration (23,24) and primary open angle glaucoma (25-28).

GWAS for DR

Six GWAS for DR have been reported to date (*Table 3*). The studies are across multiple ethnicities including Mexican Americans (29), Chinese (Taiwanese) (30,32), Japanese (33) and Caucasians from USA (31) and Australia (34), with replication cohorts also including Hispanic (32) and Indian (34) participants. Most of the studies have been conducted in patients with type 2 diabetes, the exception being a meta-analysis of the GoKinD (Genetics of Kidney in Diabetes) and EDIC (Epidemiology of Diabetes Intervention and Control Trial) (31) studies conducted in type 1 diabetic patients. This meta-analysis is also notable for its definition of DR, which was based on laser treatment for PDR or DME (self-reported in GoKinD), whereas all other studies have used ETDRS or the related ICDRS severity scales. It is also evident that there is very little consensus about the grouping of participants into cases and controls, with some studies evaluating PDR and others a combination of PDR and NPDR, with or without DMO.

Table 3 clearly shows that to date, only small numbers of patients have been analysed, with a total of only 4,506 cases and 5,902 controls included across all discovery and replication cohorts, and fewer than 2,500 DR patients genotyped by genome-wide SNP array. The largest single discovery cohort was the meta-analysis of GoKinD and EDIC, with a total of 973 cases, but that study lacked a replication cohort to confirm the findings. The TUDR study of Taiwanese patients with replication in a US-based Hispanic population and the Australian GSDR include larger numbers of DR cases, but only around one third of the participants in each study were included in the genome-wide discovery stages, limiting the power of the discovery phase.

Findings from GWAS for DR

Table 4 outlines the main findings from each of the six studies, including the replication studies presented in the initial report. None of the studies provided replicated evidence of association at a genome-wide level of statistical significance ($P<5\times10^{-8}$) and there is effectively no overlap in the genetic regions identified as harbouring risk alleles in each of the six studies.

The study in Starr County Mexican Americans (29), the first study in Taiwanese Chinese (30) and the meta-analysis of type 1 DM studies (31) did not include replication cohorts, relying instead on future studies to replicate findings. The Taiwanese study of Huang *et al.* (30) reported several highly significant loci with p-values well below the genome-wide significant threshold of $P<5\times10^{-8}$; however, these are not replicated in the discovery cohort of a subsequent study by Sheu *et al.* (32) in the same ethnic group. Although both studies highlighted loci on chromosome 13, the two lead SNPs, rs2038823 and rs9565164, are around 20 Mb apart, well beyond the reach of linkage disequilibrium and representing independent loci. The reasons for lack of replication may be attributed to slightly different case cohorts (Huang *et al.* include NPDR in the discovery while Sheu *et al.* is limited to PDR), but is most likely due to the small sample sizes leading to false positive findings (*Table 3*).

The primary findings of Sheu *et al.* (32) in the Taiwanese discovery cohort were also unable to be replicated, with the Hispanic replication cohort reducing the overall significance of the loci when combined in a meta-analysis of all stages of the study. Ideally, replication studies would initially be carried out in a population comparable to the discovery sample, particularly in terms of ethnicity, type of diabetes, definitions of DR and statistical approaches. Differences in several of these variables between discovery and replication in the TUDR study of Sheu *et al.* may explain the lack of replication observed.

The GWAS of Japanese type 2 DM patients by Awata *et al.* (33) was conducted in three stages, with stages 2 and 3 using participant samples very similar to the stage 1 cohort, but only genotyped on SNPs reaching thresholds of suggestive association in the stage 1 discovery cohort. A similar approach was utilised in our own study of an Australian cohort with sight-threatening DR in type 2 patients (34). This multi-stage approach can reduce overall study costs due to a reduction in genotyping requirements, whilst capitalising on the power of the full cohort by the final stage (35). This approach was commonly implemented when the cost of SNP array genotyping was relatively high, but as the cost of genotyping has reduced and the ability

Table 3 Details of GWAS studies for diabetic retinopathy

Reference	Study	Study stages	Ethnicity	DM type	Cases		Controls		Grading scale	Genotyping	Imputation reference
					N	Phenotype	N	Phenotype			
(29)	Starr county	Discovery	MexAm	2	102	ETDRS >43	183	ETDRS 10–37	ETDRS	Affymetrix 100K	HapMap III
(30)	Taiwan	Discovery	Chinese	2	174	NPDR + PDR	575	No DR	ICDRS	HumanHap550K	none
(31)	GoKinD	Meta-analysis	Caucasian	1	815	Self-report laser	803	No laser		Affymetrix 5.0	HapMap II
	EDIC				158	Laser	1,053		ETDRS	HumanHap550	
(32)	TUDR	Discovery	Chinese	2	437	PDR	570	No DR	ICDRS	OmniExpress	HapMap
		Extension-cases	Chinese	2	479	NPDR			ICDRS	Targeted	
		Replication	Hispanic	2	329	ETDRS ≥14	256	ETDRS <14	ETDRS	Targeted	
(33)	Japanese	Stage 1 (discovery)	Japanese	2	205	NPDR + PDR	241	No DR	ICDRS	Affymetrix 6.0	None
		Stage 2		2	335		288			Targeted	
		Stage 3		2	297		620			Targeted	
(34)	Australian	Discovery	Caucasian	2	336	NPDR or PDR or CSME	502	No DR	ICDRS	OmniExpress	None
	GSDR	Replication 1		2	263		320			Targeted	
		Replication 2		1	242		126			OmniExpress	
		Replication 3	Indian	2	334		365			Targeted	

GoKinD, Genetics of Kidney in Diabetes; EDIC, Epidemiology of Diabetes Intervention and Control Trial; TUDR, Taiwan-US Diabetic Retinopathy Study; GSDR, Genetic Study of Diabetic Retinopathy; MexAm, Mexican American; ETDRS, grading as per the Early Treatment in Diabetic Retinopathy Study scale; ICDRS, grading as per the International Clinical DR Severity scale.

Table 4 Results of GWAS for DR. Association statistics for top ranked SNPs as selected by original authors are given, along with P values in replication cohorts and overall meta-analyses included in original publications

Reference	Chr	SNP	P value, discovery	Replication and meta-analysis				Nearest gene(s)
				Rep 1 P	Rep 2 P	Rep 3 P	Meta P	
(29)[a]	5	rs2300782	6.04×10^{-5}					*CAMK4*
	15	rs10519765	6.21×10^{-5}					*FMN1*
	6	rs6909083	1.80×10^{-5}					*TINAG*
	6	rs17083119	2.76×10^{-5}					*C6orf170*
	1	rs1033465	4.50×10^{-5}					*TNFSF18*
	1	rs11583330	5.35×10^{-5}					*GNAI3*
(30)	5	rs17376456	2.99×10^{-15}					*FAM172A*
	13	rs2038823	4.68×10^{-11}					*HS6ST3*
	9	rs4838605	1.87×10^{-9}					*ARHGAP22*
	10	rs12219125	9.29×10^{-9}					*PLXDC2*
	10	rs4462262	9.21×10^{-8}					*Gene Desert*
	1	rs2811893	3.09×10^{-7}					*MSYM1*
	4	rs4470583	4.25×10^{-7}					*FSTL5*
(31)	1	rs476141	1.20×10^{-7}					*AKT-ZNF238*
	1	rs512825	6.20×10^{-7}					*AKT-ZNF238*
	16	rs4787008	6.40×10^{-7}					*A2BP1*
	3	rs13064954	7.10×10^{-7}					*LERK1-CCNL1*
	1	rs1711347	8.50×10^{-7}					*AKT-ZNF238*
	3	rs9866141	8.80×10^{-7}					*KRT18P34-VEPH1*
(32)[b]	13	rs9565164	4.40×10^{-7}	rs9543976[f]	0.41		7.40×10^{-6}	*TBC1D4-COMMD6-UCHL3*
	2	rs1399634	4.20×10^{-6}	rs4668142[f]	0.57		3.60×10^{-4}	*LRP2-BBS5*
	2	rs2380261	4.70×10^{-6}		0.82		2.00×10^{-4}	*ARL4C-SH3BP4*
(33)[c]	6	rs9362054[e]	1.20×10^{-3}	1.20×10^{-3}	0.015		1.30×10^{-6}	*LINC01611*
			3.30×10^{-8}	When all three stages combined				
(34)[d]	6	rs3805931	2.66×10^{-7}	0.870	0.097			*PTK7*
	6	rs1537638	3.11×10^{-7}	0.770	0.025			*PTK7*
	17	rs9896052[e]	6.55×10^{-5}	0.035	0.041	0.016	4.12×10^{-8}	*GRB2*

[a], adjusted for age, sex, DM duration and HbA1c; [b], adjusted for age and sex; meta-analysis P value includes the discovery and extension samples combined; [c], adjusted for sex, DM duration and HbA1c; [d], adjusted for age, sex, DM duration, hypertension, nephropathy and HbA1c; [e], top ranked SNP overall once all stages and replication cohorts combined; [f], replaced discovery SNP in replication and meta-analysis as rs9565164 is rare in Hispanics and rs1399639 failed targeted genotyping assay design.

to impute data against maps of ever-increasing density has improved, future studies will likely aim to genotype all samples using SNP arrays wherever possible. Both these studies reported genome-wide significant association with the inclusion of all samples in the final stage of analysis. The Japanese study identified a locus in the long non-coding RNA gene *LINC01611*, which reached statistical significance when all samples were combined, but fell

short under formal meta-analysis of the three stages. In this case, the combined analysis is likely appropriate as the participants included in each stage were recruited and assessed under the same protocols, although differences in clinical characteristics of participants at each stage can be seen (33).

The main finding of our own study (Burdon *et al.*) (34) is the locus upstream of the *GRB2* gene on chromosome 17. Similar to the experience in the Japanese study, the lead SNP at this locus, rs9896052, was not among the top few SNPs following the discovery analysis, but was the only stage 2 SNP showing at least nominal association in both Caucasian replication cohorts. This SNP reached genome-wide significance with the inclusion of an Indian cohort in the final stage. Of note, one of the replication cohorts had type 1 diabetes, making this one of the first loci to be reported showing some level of association across ethnic groups and in both types of diabetes. The other locus reported by this study on chromosome 6 was nominally associated in the type 1 cohort, but not in the type 2 cohort which was the best match for the discovery sample, recruited under the same protocol and displaying similar clinical characteristics. We concluded that this locus was not robustly associated with sight-threatening DR.

Relatively standardised SNP array genotyping methods are used in all the studies, and technical differences in specimen handling or genotyping are unlikely to account for the lack of replication to date. It is most likely due to a combination of underpowered studies and different phenotype definitions. These issues have also been seen in the GWAS for other diabetic complications including nephropathy and cardiovascular disease, with many SNPs reported that have not yet been replicated; however, larger studies and meta-analyses have identified reproducibly associated genetic risk loci for these diseases (36,37).

Follow-up of GWAS findings in independent cohorts

Replication of findings in an independent cohort is required to confirm any association. There have been a number of reports attempting to replicate previously published GWAS results. Grassi *et al.* (38) attempted to replicate their own previous findings from the GoKinD + EDIC meta-analysis in type 1 diabetic participants of the Wisconsin Epidemiological Study of Diabetic Retinopathy (WESDR) cohort. This cohort has very similar demographics to the original study and comparable disease definitions were used. They genotyped 389 top ranked SNPs from GoKinD and EDIC and the meta-analysis of those two studies (31). The authors did not present data for the WESDR cohort alone, but conducted a meta-analysis of all three cohorts. None of the evaluated SNPs provided robust evidence for association, with the overall p-values decreasing compared to the original study.

McAuley *et al.* (39) also followed up the GoKinD + EDIC study as well as the early Taiwanese GWAS of Huang *et al.* (30) by typing the top 24 SNPs from the two reports in 163 cases and 300 controls from the Diabetic Management Project (DMP) cohort, consisting of Australian patients with type 1 and type 2 diabetes. They stratified the analyses by diabetes type as well as combining all the data, but were not able to find robust associations surviving correction for multiple testing. The most encouraging results were at rs1073203 from the GoKinD + EDIC study and rs4838605 in the intron of the *ARHGAP22* gene from the Taiwanese study.

Peng *et al.* (40) evaluated 40 SNPs reported in the first two published GWAS (30,41) plus the WESDR replication study (38) in a Chinese cohort of 789 cases and 1,110 controls. Once again, no statistically significant study-wide associations were observed. Nominal association was reported at rs17684886 near *ZNRF1*, originally highlighted in the GoKinD + EDIC GWAS. A nominal association was also reported at rs599019 from the initial Mexican American GWAS (29).

Hosseini *et al.* (42) assessed the top SNPs from the first four published GWAS in a meta-analysis of WESDR, EDIC and the Diabetes Control and Complications Trial (DCCT) cohorts. The DCCT is similar to the other cohorts in that it consists of predominantly European Americans with type 2 diabetes. To prevent overlap with the earlier reports, they only considered SNPs from the GoKinD cohort of the original GoKinD + EDIC GWAS. Of note, they report nominal association at the *PLXDC2* gene originally reported in the Taiwanese study, but again, the P value did not reach the required threshold for statistical significance in the context of testing 34 loci. It is notable that neither WESDR nor DCCT were able to replicate the findings from the only GWAS of DR in type 1 diabetes.

Cheung *et al.* (43) also evaluated findings from the first four GWAS, but in a Hong Kong Chinese type 2 diabetes cohort consisting of 567 patients with sight-threatening DR (severe NPDR or PDR) and 1,490 controls. The most promising finding from this study was association of SNP rs2115386 near the *INSR* gene. This association does

survive multiple testing correction; however, it did not reach genome-wide significance in the original GoKinD + EDIC GWAS analysis, nor do any subsequent replication studies highlight this variant. Thus, although this is an attractive candidate gene for a role in diabetic complications, further evidence is required for a definitive conclusion.

Our Australian GWAS also reported association statistics in the type 2 diabetes discovery cohort for all SNPs reported in the five earlier studies (or equivalent tag SNPs) (34). We were not able to replicate any of the loci with statistical significance when accounting for multiple testing, but did report nominal association at the *LRP2* locus (34) initially reported by Sheu *et al.* in the Taiwanese cohort, as well as at a locus near the *PCSK2* gene on chromosome 20, reported as having suggestive association in the GoKinD + EDIC meta-analysis (31).

Why has replication of GWAS findings proven so difficult?

Failure to replicate findings suggests either false positive findings in the initial studies, differences in the underlying genetic architecture between cohorts of different ethnicities, differences in phenotype classification or consideration of covariates between studies, or under-powered replication studies. It is apparent that small size has plagued DR GWAS to date and is very likely to be the driving force behind false positive associations and subsequent lack of replication. Ethnic and clinical differences are also important; it is evident that genetic architecture and disease classification varies between populations, and this affects the outcomes of genetic studies. For example, common variation at the *LOXL1* gene is strongly and reproducibly associated with pseudoexfoliation syndrome (44,45); however, risk and protective alleles are reversed in European populations compared with Asian and South African populations (46). When this 'flipped' association was first reported in Japanese populations, it was tempting to hypothesise technical errors and false positive results; however, it is now clear through a wealth of evidence that this is not the case and the locus is important in disease in all populations. The mechanism of disease and the reason for the flipped association remains elusive (47). Although ethnicity can have an effect on association, it should be noted that, more often than not, the SNPs of largest effect are associated across ethnicities. For example, large studies of primary open angle glaucoma have replicated findings across multiple ethnic groups (48) and this is also proving to be the case in other diabetic complications including nephropathy and cardiovascular disease where some robust associations have been reported (36). Further elucidation of the genetics of all diabetic complications, including DR, are required before the genetic overlap between them can be determined. It seems likely that each complication will have unique factors, but with some overlap between conditions.

Difficulties in replication may also be caused by the limitations of the replication study itself. For example, statistical significance at the *LINC01611* locus in the Japanese GWAS (33) was not seen until all 837 cases were included, but replication was assessed only in the initial discovery phase of the Australian study with 336 patients with sight-threatening DR (34) with genome-wide SNP array data available. Thus the replication study itself is likely underpowered. Several other attempted replications also indicate they were underpowered for SNPs with lower minor allele frequencies (39,43).

The use of endophenotypes for mapping disease loci

In other complex diseases, endophenotypes have been employed extensively to aid mapping of genetic risk loci for the disease of interest. An endophenotype is a trait that confers risk of disease, but must be measurable in the whole population (not just in patients with the disease of interest) and has a genetic basis. GWAS for endophenotypes have highlighted loci which have subsequently been shown to also be associated with the disease of interest. For example, optic disc parameters (49,50) and intraocular pressure (51) in glaucoma and central corneal thickness (52) in keratoconus. As glycaemic control, measured by HbA1c, is an important risk factor for the development of DR, it follows that risk alleles for this trait may also be associated with DR (53). GWAS for HbA1c amongst non-diabetic individuals have identified a number of loci associated with this trait in Caucasian (54-56) and Asian (57) cohorts but very little evaluation of these loci in DR has been reported. Chen *et al.* (57) conducted a GWAS in a collection of cohorts from Singapore but did not find any replicating loci beyond those reported in the earlier European based studies. They further evaluated the European-based loci (54-56) for association with DR in the Asian samples, but were unable to identify strong associations accounting for multiple testing, using 599 cases and 1,423 controls with type 2 diabetes. Paterson *et al.* (58) undertook a GWAS for HbA1c in type 1 diabetes patients from the DCCT

and further showed association of the top ranked loci with a number of diabetic complications, including the *BNC2* locus with NPDR and DMO in the same cohort. This work would be further strengthened by assessment of this locus in an independent cohort for association with DR. Of note, this region is not reported in any of the GWAS for DR published to date.

Future directions in genetic risk locus identification in DR

GWAS is a generic technique which has predominantly been applied to the assessment of common variation. It has long been hypothesised that rare variation makes a major contribution to the overall genetic risk profiles of complex diseases (59); this has been demonstrated in age-related macular degeneration (23) and is also likely to be the case for DR. The GWAS conducted to date have not been powered to assess even common variation of moderate effect size and imputation has not yet been able to infer genotypes for very rare variants from SNP array data. Thus, the contribution of rare variants has not yet been assessed to any degree in DR.

Much attention of late has been given to 'next generation' sequencing studies including whole exome and whole genome sequencing. The advantages of this technology for the discovery of rare alleles in particular are clear; however, the size and power of the cohort and quality of phenotyping remain paramount to achieving reproducible associations, no matter what genotyping technology or statistical techniques are used. Shtir *et al.* (60) describe a very small study of 64 type 2 diabetes patients with DR and 43 without DR from Saudi Arabia, using exome sequencing and gene-based statistics to search for rare variants contributing to DR. They report three genes apparently enriched for protective variants in the patients without DR (*NME3*, *LOC728688* and *FASTK)*, but without a replication cohort, this study suffers from the same likelihood of false positive results as the prior GWAS using SNP arrays for common variants. While sequencing technologies have revolutionised detection of variants, the underlying study design (to associate the variation with disease risk) remains the most important factor to successful gene mapping. Although exome and genome sequencing have the ability to detect rare variation, this technology represents an alternative genotyping method to a SNP array and the data analysis relies on similar statistics. It does not negate the need for large, well-powered and well-phenotyped cohorts with appropriate replication.

Conclusions

Despite the challenges in phenotype grading and comparing studies with differing methodologies across multiple ethnic groups, it is clear that the size of the current studies is a major limitation to gene discovery. Studies from several other ophthalmic complex diseases including pseudoexfoliation syndrome (45), age-related macular degeneration (24) and Fuch's endothelial corneal dystrophy (61) have discovered common variants with large effect sizes, detectable in cohorts of under 100 cases. The lack of replicated associations in the DR GWAS published to date, however, strongly indicates that there are no such common variants with large effect sizes responsible for DR risk. The genetic architecture of this disease is highly complex and confounded by numerous environmental risk factors and gene-environment interactions. Larger, well-phenotyped cohorts will be required, along with concerted efforts for cross study meta-analysis. This will require co-operation between studies and a co-ordinated effort to bring together well-characterised cohorts with dense phenotypic information for combined and meta-analyses, as has been successful for other diseases. This process is already underway [e.g., (62)] and the efforts of many will overcome the issues highlighted by this review. Appropriate consideration for covariates in the analysis will be required and GWAS for endophenotypes of DR are likely to be fruitful. Genetic studies of DMO are also very limited in the current literature and consideration of this sub-type as a separate disease entity may also facilitate gene discovery. Genomics has historically focused on cohorts of Western European descent, so care will need to be taken to continue to appropriately analyse and consider cohorts from multiple ethnicities. There are thousands of DR patients described in the literature across all studies combined, but to date, only a handful of studies have used genome-wide SNP array data for comprehensive GWAS in fewer than 2,500 cases in total. As the number of cases and controls with genotype data available increases, the required meta-analyses will become feasible.

Acknowledgments

KP Burdon is supported by a Senior Research Fellowship

from the National Health and Medical Research Council of Australia.
Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editor (Jay M. Stewart) for the series "Diabetic Retinopathy" published in *Annals of Eye Science*. The article has undergone external peer review.

Conflicts of Interest: The authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/aes.2018.08.04). The series "Diabetic Retinopathy" was commissioned by the editorial office without any funding or sponsorship. KB reports grants from National Health and Medical Research Council of Australia, during the conduct of the study. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Kempen JH, O'Colmain BJ, Leske MC, et al. The prevalence of diabetic retinopathy among adults in the United States. Arch Ophthalmol 2004;122:552-63.
2. Tan GS, Gan A, Sabanayagam C, et al. Ethnic Differences in the Prevalence and Risk Factors of Diabetic Retinopathy: The Singapore Epidemiology of Eye Diseases Study. Ophthalmology 2018;125:529-36.
3. Nathan DM. Long-term complications of diabetes mellitus. N Engl J Med 1993;328:1676-85.
4. Kaidonis G, Abhary S, Daniell M, et al. Genetic study of diabetic retinopathy: recruitment methodology and analysis of baseline characteristics. Clin Experiment Ophthalmol 2014;42:486-93.
5. Klein R, Klein BE, Moss SE, et al. Relationship of hyperglycemia to the long-term incidence and progression of diabetic retinopathy. Arch Intern Med 1994;154:2169-78.
6. Srinivasan S, Dehghani C, Pritchard N, et al. Ophthalmic and clinical factors that predict four-year development and worsening of diabetic retinopathy in type 1 diabetes. J Diabetes Complications 2018;32:67-74.
7. Klein R, Klein BE, Moss SE, et al. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: XVII. The 14-year incidence and progression of diabetic retinopathy and associated risk factors in type 1 diabetes. Ophthalmology 1998;105:1801-15.
8. Klein R, Klein BE, Moss SE, et al. The Beaver Dam Eye Study. Retinopathy in adults with newly discovered and previously diagnosed diabetes mellitus. Ophthalmology 1992;99:58-62.
9. Arar NH, Freedman BI, Adler SG, et al. Heritability of the severity of diabetic retinopathy: the FIND-Eye study. Invest Ophthalmol Vis Sci 2008;49:3839-45.
10. Hietala K, Forsblom C, Summanen P, et al. Heritability of proliferative diabetic retinopathy. Diabetes 2008;57:2176-80.
11. López M, Cos FX, Alvarez-Guisasola F, et al. Prevalence of diabetic retinopathy and its relationship with glomerular filtration rate and other risk factors in patients with type 2 diabetes mellitus in Spain. DM2 HOPE study. J Clin Transl Endocrinol 2017;9:61-5.
12. The Diabetic Retinopathy Study Research Group. Diabetic retinopathy study. Report Number 7. A modification of the Airlie House classification of diabetic retinopathy. . Invest Ophthalmol Vis Sci 1981;21:1-26.
13. Klein R, Klein BE, Magli YL, et al. An alternative method of grading diabetic retinopathy. Ophthalmology 1986;93:1183-7.
14. Wilkinson CP, Ferris FL, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology 2003;110:1677-82.
15. Leslie RD, Pyke DA. Diabetic retinopathy in identical twins. Diabetes 1982;31:19-21.
16. The Diabetes Control and Complications Trial Research Group. Clustering of long-term complications in families with diabetes in the diabetes control and complications trial. Diabetes 1997;46:1829-39.
17. Rema M, Saravanan G, Deepa R, et al. Familial clustering of diabetic retinopathy in South Indian Type 2 diabetic patients. Diabet Med 2002;19:910-6.

18. Hallman DM, Huber JC Jr, Gonzalez VH, et al. Familial aggregation of severity of diabetic retinopathy in Mexican Americans from Starr County, Texas. Diabetes Care 2005;28:1163-8.
19. Looker HC, Nelson RG, Chew E, et al. Genome-wide linkage analyses to identify Loci for diabetic retinopathy. Diabetes 2007;56:1160-6.
20. Guey LT, Kravic J, Melander O, et al. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. Genet Epidemiol 2011;35:236-46.
21. Imperatore G, Hanson RL, Pettitt DJ, et al. Sib-pair linkage analysis for susceptibility genes for microvascular complications among Pima Indians with type 2 diabetes. Pima Diabetes Genes Group. Diabetes 1998;47:821-30.
22. Hallman DM, Boerwinkle E, Gonzalez VH, et al. A genome-wide linkage scan for diabetic retinopathy susceptibility genes in Mexican Americans with type 2 diabetes from Starr County, Texas. Diabetes 2007;56:1167-73.
23. Fritsche LG, Igl W, Bailey JN, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat Genet 2016;48:134-43.
24. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385-9.
25. Bailey JN, Loomis SJ, Kang JH, et al. Genome-wide association analysis identifies TXNRD2, ATXN2 and FOXC1 as susceptibility loci for primary open-angle glaucoma. Nat Genet 2016;48:189-94.
26. Burdon KP. Genome-wide association studies in the hunt for genes causing primary open-angle glaucoma: a review. Clinical and Experimental Ophthalmology 2012;40:358-63.
27. Gharahkhani P, Burdon KP, Fogarty R, et al. Common variants near ABCA1, AFAP1 and GMDS confer risk of primary open-angle glaucoma. Nat Genet 2014;46:1120-5.
28. Shiga Y, Akiyama M, Nishiguchi KM, et al. Genome-wide association study identifies seven novel susceptibility loci for primary open-angle glaucoma. Hum Mol Genet 2018;27:1486-96.
29. Fu YP, Hallman DM, Gonzalez VH, et al. Identification of Diabetic Retinopathy Genes through a Genome-Wide Association Study among Mexican-Americans from Starr County, Texas. J Ophthalmol 2010;2010:861291.
30. Huang YC, Lin JM, Lin HJ, et al. Genome-wide Association Study of Diabetic Retinopathy in a Taiwanese Population. Ophthalmology 2011;118:642-8.
31. Grassi MA, Tikhomirov A, Ramalingam S, et al. Genome-wide Meta-analysis for Severe Diabetic Retinopathy. Hum Mol Genet 2011;20:2472-81.
32. Sheu WH, Kuo JZ, Lee IT, et al. Genome-wide association study in a Chinese population with diabetic retinopathy. Hum Mol Genet 2013;22:3165-73.
33. Awata T, Yamashita H, Kurihara S, et al. 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.
34. Burdon KP, Fogarty RD, Shen W, et al. Genome-wide association study for sight-threatening diabetic retinopathy reveals association with genetic variation near the GRB2 gene. Diabetologia 2015;58:2288-97.
35. Skol AD, Scott LJ, Abecasis GR, et al. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 2006;38:209-13.
36. Ma RC. Genetics of cardiovascular and renal complications in diabetes. J Diabetes Investig 2016;7:139-54.
37. Regele F, Jelencsics K, Shiffman D, et al. Genome-wide studies to identify risk factors for kidney disease with a focus on patients with diabetes. Nephrol Dial Transplant 2015;30 Suppl 4:iv26-34.
38. Grassi MA, Tikhomirov A, Ramalingam S, et al. Replication analysis for severe diabetic retinopathy. Invest Ophthalmol Vis Sci 2012;53:2377-81.
39. McAuley AK, Wang JJ, Dirani M, et al. Replication of genetic loci implicated in diabetic retinopathy. Invest Ophthalmol Vis Sci 2014;55:1666-71.
40. Peng D, Wang J, Zhang R, et al. Common variants in or near ZNRF1, COLEC12, SCYL1BP1 and API5 are associated with diabetic retinopathy in Chinese patients with type 2 diabetes. Diabetologia 2015;58:1231-8.
41. Fu W, O'Connor TD, Jun G, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 2013;493:216-20.
42. Hosseini SM, Boright AP, Sun L, et al. The association of previously reported polymorphisms for microvascular complications in a meta-analysis of diabetic retinopathy. Hum Genet 2015;134:247-57.
43. Cheung CY, Hui EY, Lee CH, et al. 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.
44. Hewitt AW, Sharma S, Burdon KP, et al. Ancestral LOXL1 variants are associated with pseudoexfoliation in Caucasian Australians but with markedly lower penetrance than in Nordic people. Hum Mol Genet 2008;17:710-6.
45. Thorleifsson G, Magnusson KP, Sulem P, et al. Common

sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. Science 2007;317:1397-400.
46. Aboobakar IF, Johnson WM, Stamer WD, et al. Major review: Exfoliation syndrome; advances in disease genetics, molecular biology, and epidemiology. Exp Eye Res 2017;154:88-103.
47. Pasutto F, Zenkel M, Hoja U, et al. Pseudoexfoliation syndrome-associated genetic variants affect transcription factor binding and alternative splicing of LOXL1. Nat Commun 2017;8:15466.
48. Ng SK, Casson RJ, Burdon KP, et al. Chromosome 9p21 primary open-angle glaucoma susceptibility locus: a review. Clinical and Experimental Ophthalmology 2014;42:25-32.
49. Springelkamp H, Hohn R, Mishra A, et al. Meta-analysis of genome-wide association studies identifies novel loci that influence cupping and the glaucomatous process. Nat Commun 2014;5:4883.
50. Springelkamp H, Mishra A, Hysi PG, et al. Meta-analysis of Genome-Wide Association Studies Identifies Novel Loci Associated With Optic Disc Morphology. Genet Epidemiol 2015;39:207-16.
51. Hysi PG, Cheng CY, Springelkamp H, et al. Genome-wide analysis of multi-ancestry cohorts identifies new loci influencing intraocular pressure and susceptibility to glaucoma. Nat Genet 2014;46:1126-30.
52. Lu Y, Vitart V, Burdon KP, et al. Genome-wide association analyses identify multiple loci associated with central corneal thickness and keratoconus. Nat Genet 2013;45:155-63.
53. Paterson AD, Bull SB. Does familial clustering of risk factors for long-term diabetic complications leave any place for genes that act independently? J Cardiovasc Transl Res 2012;5:388-98.
54. Franklin CS, Aulchenko YS, Huffman JE, et al. The TCF7L2 diabetes risk variant is associated with HbA(1)(C) levels: a genome-wide association meta-analysis. Ann Hum Genet 2010;74:471-8.
55. Paré G, Chasman DI, Parker AN, et al. Novel association of HK1 with glycated hemoglobin in a non-diabetic population: a genome-wide evaluation of 14,618 participants in the Women's Genome Health Study. PLoS Genet 2008;4:e1000312.
56. Soranzo N, Sanna S, Wheeler E, et al. Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. Diabetes 2010;59:3229-39.
57. Chen P, Ong RT, Tay WT, et al. A study assessing the association of glycated hemoglobin A1C (HbA1C) associated variants with HbA1C, chronic kidney disease and diabetic retinopathy in populations of Asian ancestry. PLoS One 2013;8:e79767.
58. Paterson AD, Waggott D, Boright AP, et al. A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. Diabetes 2010;59:539-49.
59. Schork NJ, Murray SS, Frazer KA, et al. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 2009;19:212-9.
60. Shtir C, Aldahmesh MA, Al-Dahmash S, et al. 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.
61. Baratz KH, Tosakulwong N, Ryu E, et al. E2-2 protein and Fuchs's corneal dystrophy. N Engl J Med 2010;363:1016-24.
62. Sobrin L, Chong YH, Fan Q, et al. Genetically Determined Plasma Lipid Levels and Risk of Diabetic Retinopathy: A Mendelian Randomization Study. Diabetes 2017;66:3130-41.

doi: 10.21037/aes.2018.08.04
Cite this article as: Burdon KP, McComish BJ, Charlesworth JC. Progress and challenges in genome-wide studies to understand the genetics of diabetic retinopathy. Ann Eye Sci 2018;3:46.

