Genome-wide association study for sight threatening Diabetic Retinopathy reveals association with genetic variation near the *GRB2* gene

Kathryn P Burdon[1,2], Rhys D Fogarty[1], Weiyong Shen[3], Sotoodeh Abhary[1], Georgia Kaidonis[1], Binoy Appukuttan[1], Alex W Hewitt[4], Shiwani Sharma[1], Mark Daniell[5], Rohan W Essex[6], John H Chang[7,8], Sonja Klebe[9], Stewart R Lake[1], Bishwanath Pal[8], Alicia Jenkins[10], Gowthaman Govindarjan[11], Periasamy Sundaresan[11], Ecosse L Lamoureux[4,12], Kim Ramasamy[13], Maria Pefkianaki[8], Philip G Hykin[8], Nikolai Petrovsky[14], Matthew A Brown[15], Mark C Gillies[3], Jamie E Craig[1].

1. Department of Ophthalmology, Flinders University, Flinders Medical Centre, Adelaide, 5042, South Australia, Australia.
2. Menzies Research Institute Tasmania, University of Tasmania, Hobart, 7001, Tasmania, Australia
3. Save Sight Institute, Clinical Ophthalmology and Eye Health, the University of Sydney, Sydney, 2001, New South Wales, Australia.
4. Centre for Eye Research Australia, University of Melbourne, East Melbourne, 3002, Victoria, Australia.
5. Department of Ophthalmology, Royal Melbourne Hospital, Parkville, 3050, Victoria, Australia
6. Academic Unit of Ophthalmology, Australian National University, Canberra, 0200, Australia
7. School of Medical Sciences, University of NSW, Sydney, 2052, New South Wales,
8. Medical Retina Service, Moorfields Eye Hospital, London, EC1V 2PD, UK
9. Department of Anatomical Pathology, Flinders University, Flinders Medical Centre, Adelaide, 5042, South Australia, Australia.
10. St Vincent's Hospital, Fitzroy, 3065, Victoria, Australia
11. Department of Genetics, Aravind Medical Research Foundation, Madurai, 625 020, Tamilnadu, India
12. Singapore Eye Research Institute, Singapore, 168751
13. Retina clinic, Aravind Eye Hospital, Madurai, 625 020, Tamilnadu, India
14. Department of Endocrinology, Flinders University, Flinders Medical Centre, Adelaide, 5042, South Australia, Australia.
15. The University of Queensland Diamantina Institute, Translational Research Institute Princess Alexandra Hospital, Brisbane, 4102, Australia

Joint corresponding authors

A/Prof Kathryn P Burdon
Menzies Research Institute, Tasmania
Private bag 23, Hobart, TAS 7000, Australia
Kathryn.burdon@utas.edu.au
Ph: +61 3 62264288

Prof Jamie E Craig
Dept Ophthalmology, Flinders University
GPO Box 2100, Adelaide, SA, 5001
Jamie.craig@flinders.edu.au
Ph +61 8 82044899

Word Counts: Abstract= 353 Main Text=3951

Abstract

Aims: Diabetic Retinopathy is a serious complication of diabetes mellitus and can lead to blindness. A genetic component, in addition to traditional risk factors, has been well described although strong genetic factors have not yet been identified. Here we aimed to identify novel genetic risk factors for sight threatening diabetic retinopathy using a genome-wide association study.

Methods: Retinopathy was assessed in Caucasian Australians with type 2 diabetes mellitus. Genome-wide association analysis was conducted for comparison of cases with sight threatening diabetic retinopathy (n=336) to diabetic controls with no retinopathy (n=508). Top ranking SNPs were typed in a type 2 diabetes replication cohort and a type 1 diabetes cohort and an Indian type 2 cohort. A mouse model of proliferative retinopathy was used to assess differential expression of the nearby candidate gene GRB2 by immunohistochemistry and quantitative Western blot.

Results: The top ranked variant was rs3805931 with $p=2.66 \times 10^{-7}$, but it was not associated in the replication cohort. Only rs9896052 ($p=6.51 \times 10^{-5}$) also showed association in both the type 2 ($p=0.035$) and the type 1 ($p=0.041$) replication cohorts and was also associated in the Indian cohort ($p=0.016$). The study-wide meta-analysis reached genome-wide significance ($p=4.15 \times 10^{-8}$). The *GRB2* gene is located downstream of this variant and a mouse model of retinopathy showed increased GRB2 expression in the retina.

Conclusion: Genetic variation near *GRB2* on chromosome 17q25.1 is associated with sight threatening diabetic retinopathy. Several genes in this region are promising candidates and in particular *GRB2* is up regulated during retinal stress and neovascularisation.

Keywords
Blinding retinopathy, diabetic complications, genetic risk factors, genome-wide association study, Muller cell, proliferative retinopathy,

Abbreviations

CSMO Diabetic macular oedema

GWAS Genome-wide association study

NPDR Non-proliferative diabetic retinopathy

PDR Proliferative diabetic retinopathy

SNP Single nucleotide polymorphism

VEGF Vascular endothelial growth factor

Introduction

Diabetic Retinopathy (diabetic retinopathy) is the most common cause of blindness in developed countries in patients under 75 years of age [1]. It is a common complication of both type 1 and type 2 diabetes mellitus. The overall prevalence of any retinopathy among persons with diabetes is 40.3%, with a prevalence of 8.2% for sight threatening diabetic retinopathy [2]. The prevalence of diabetes itself is predicted to double between the year 2000 and 2030 [3], and thus the health burden of diabetic retinopathy is also expected to increase dramatically.

Diabetic retinopathy is defined by the presence of retinal microvascular lesions including microaneurysms, retinal haemorrhages, lipid exudates, nerve fibre layer infarcts, intraretinal microvascular abnormalities and venous beading, collectively termed non-proliferative diabetic retinopathy [4]. The growth of abnormal new blood vessels is termed proliferative diabetic retinopathy [4]. Macular oedema is also a major cause of vision loss in diabetes and may occur in both non-proliferative and proliferative diabetic retinopathy [4].

It is well known that poor glycemic control is a risk factor for diabetic retinopathy [5, 6], however, tight control does not fully prevent the disease and intensive diabetic control can result in increased mortality [7]. Several studies have demonstrated familial clustering of diabetic retinopathy independent of glycemic control [8-11]. The FIND-EYE study reported a significant heritability of 0.27 for diabetic retinopathy in families [12], and a study of a large cohort of American families with type 1 diabetes confirmed a strong familial risk for diabetic retinopathy, independent of duration of disease [13]. Numerous potentially associated genes have been reported in the literature [14] but few consistent findings have been reported.

Three recent genome-wide association studies (GWAS) for diabetic retinopathy have reported loci of interest that did not meet the statistical thresholds for genome-wide significance [15-17]. A study in a cohort of Taiwanese type 2 diabetes patients and another in a meta-analysis of Japanese patients identified several loci reaching significance however, these findings have not been validated in an independent cohort [18, 19]. None of the top loci in these studies appear to be replicated by other studies. Thus, no GWAS has yet identified robust genetic associations for diabetic retinopathy.

Here we aimed to identify novel genetic risk loci for sight threatening diabetic retinopathy using a genome-wide association study in Caucasians with type 2 diabetes and demonstrate replication of the association in Caucasians with type 1 and type 2 diabetes mellitus and an Indian type 2 diabetes mellitus cohort.

Methods

Participant Recruitment

This study was approved Human Research Ethics Committees (HREC) of the relevant institutions as listed in the Electronic Supplementary Material (ESM). Written informed consent was obtained from each participant and the project conformed to the tenets of the Declaration of Helsinki.

Australian participants in the discovery cohort were recruited from the locations listed in the ESM between 2006 and 2011 as previously described [20]. Replication cohorts were recruited between 2011 and 2013 from these sites as well as additional sites listed in the ESM.

Discovery cohort

Patients of at least 18 years of age on medical treatment for type 2 diabetes mellitus for at least 5 years were invited to participate as described previously[20]. Retinopathy status was determined from direct ophthalmic examination by the treating ophthalmologist and was graded according to modified Early Treatment in Diabetic Retinopathy Study criteria [21]. Cases were defined as participants with sight-threatening diabetic retinopathy (severe non-proliferative diabetic retinopathy, proliferative diabetic retinopathy or clinically significant macular oedema) and were compared to controls with no diabetic retinopathy or minimal diabetic retinopathy. All participants meeting inclusion criteria for case or control status with DNA available were included in the current analysis. Medical histories were obtained via a questionnaire. Biochemistry results were collected including HbA_{1c}, serum lipids and renal function tests. A detailed description of all clinical variables collected has been previously published [20]. DNA was extracted from whole blood using QIAamp Blood DNA Maxi Kits (Qiagen, Chadstone Centre, VIC, Australia).

Clinical traits were assessed for association with diabetic retinopathy in SPSS (v19.0, SPSS, Chicago, IL, USA) using a chi-squared test for dichotomous variables and binary logistic regression for quantitative variables.

Replication cohorts

The type 1 and type 2 diabetes mellitus Caucasian replication cohorts were recruited from Australian hospitals and Moorfields Eye Hospital, London, under an identical protocol to the discovery cohort. The cohort consisted of 263 cases and 320 controls under the identical definition as the discovery cohort. Type 1 diabetes mellitus participants were genotyped on the OmniExpress arrays in parallel to the discovery cohort. Cases were defined as participants with

any severity of retinopathy in order to include as large a cohort as possible for this analysis. In total, 242 cases and 126 controls were included. The Indian replication cohort was recruited from Aravind Eye Hospital in Tamil Nadu, India. Participants all had type 2 diabetes mellitus of at least 10 years duration. Clinical data were collected and diabetic retinopathy was classified according to ETDRS criteria on examination. Cases were included if they had sight threatening diabetic retinopathy under the same definition as the Caucasian discovery cohort. In total 334 cases and 365 controls were included. DNA was extracted from whole blood using the salt precipitation method[22].

SNP genotyping and association analysis

Genome-wide genotyping at single nucleotide polymorphisms (SNPs) of the discovery cohort was conducted using Human OmniExpress Beadchips (Illumina Inc, San Diego, CA, USA) on a HiScan System (Illumina Inc.). Genotyping data were filtered to remove SNPs with genotyping rates < 99%, minor allele frequencies < 1% or Hardy-Weinberg equilibrium in controls $p < 1 \times 10^{-6}$. Relatedness between individuals was assessed using PLINK[23] and one of each pair of individuals with a PI-HAT > 0.3 was excluded from further analysis (10 individuals). Principal components analysis was conducted using EIGENSOFT[24]. Ancestral outliers were identified by EIGENSOFT as any eigenvalue >6SD from the mean, and from visualising plots of principal components 1 and 2 and were excluded from further analysis.

Association analyses were conducted using PLINK. Logistic regression under an additive model with adjustment for covariates (age, sex, duration of diabetes, HbA_{1c}, hypertension, nephropathy and the first three principal components) was used to test association of the 602,755 SNPs. Individuals with missing clinical data were excluded from the fully adjusted logistic regression model. The Manhattan plot was generated using R (www.r-project.org) with

code adapted from https://raw.github.com/stephenturner/qqman/master/qqman.r. The detailed locus specific plot was generated using LocusZoom[25] with recombination rates taken from the Caucasian component of the 1000 Genomes Project.

The type 2 diabetes replication cohort was genotyped at SNPs of interest using iPlex Gold chemistry (Sequenom Inc, San Diego, CA, USA) visualised on a MassArray Analyzer (Sequenom Inc.). The type 1 replication cohort was typed on the Human OmniExpress Beadchips ((Illumina Inc,) in parallel with the discovery cohort. All SNPs with a $p<1x10^{-5}$ and a selection with $p<1x10^{-4}$ in the discovery cohort that were compatible with multiplexing on the MassArray platform were assessed in these two cohorts. Association analysis in the replication cohorts was conducted in PLINK under an allelic model as well as under logistic regression with adjustment for covariates. The type 1 replication cohort was subjected to the identical quality control procedures as the discovery cohort typed on the same SNP array. This cohort had an overall lambda of 0.994, thus the principal components were not included as covariates in the analysis. Genotyping of the SNP of interest in the Indian replication cohort was conducted using a pre-designed TaqMan Assay (Life Technologies, Mulgrave, VIC, Australia) according to manufacturer's protocols on an ABI 7900 HT Fast Real Time PCR System and analysis was conducted in PLINK under logistic regression with adjustment for all clinical covariates..

Meta-analyses of multiple cohorts were undertaken in METAL with weighting for the number of people in each cohort[26] using the fully adjusted association statistics.

Sequencing of candidate genes

Seventy-two cases with the most severe retinopathy were selected from the discovery and type 2 replication cohorts for direct sequencing of the coding regions *GRB2* (Genbank accession NM_002086.4) and *MIR3678* (Genbank accession NR_037449.1) using standard Sanger sequencing as described in the ESM.

RNA extraction and RT-PCR for *GRB2* and *MIR3678*

Cadaveric human retina was obtained from the Eye Bank of South Australia with approval from the Southern Adelaide Clinical Human Research Ethics Committee. RNA was extracted from homogenised tissue in Trizol (Life Technologies) and DNA was removed using Turbo DNA-free kit (Life Technologies). cDNA was generated by reverse-transcription and transcripts amplified by PCR and visualised on 1% agarose as described in the ESM.

Immunohistochemical labelling in human

Eye tissue was fixed in buffered formalin and embedded in paraffin. For immunolabelling, sections were blocked with 5% normal goat serum and incubated with the rabbit anti-GRB2 primary antibody (1:2000, Abcam#32111; Abcam, Melbourne, VIC, Australia) at +4°C overnight. The bound GRB2 antibody was detected with the Novolink Polymer detection kit (Leica Microsystems, North Ryde, NSW, Australia) and Chromogen substrate coloration (Dako, North Sydney, NSW, Australia). Sections were counterstained with haematoxylin and mounted in dePeX (Merck, Frenchs Forest, NSW, Australia). Light microscopy was performed on Olympus BX50 brightfield upright microscope attached with a Q-Imaging colour CCD camera; images were taken using the Q-Capture software (Q-Imaging Corporate, Surry, BC, Canada).

Transgenic mice with selective Müller cell ablation

This study was performed in accordance with the Association for Research in Vision and Ophthalmology statement and approved by The University of Sydney Animal Ethics Committee. The principles of laboratory animal care were followed. Transgenic mice (Rlbp1-CreER-lacZ) for selective and inducible Müller cell ablation were produced as described previously [27]. Müller cell ablation was induced by daily intraperitoneal injection of tamoxifen (TMX, 3mg/day, dissolved in sunflower oil) for 4 consecutive days at 10-12 weeks of age [27]. Mice not carrying the *Rlbp1* promoter but carrying the *DTA176* gene and receiving TMX treatment in the same way were used as controls.

Western Blotting

Proteins were extracted from retinas and their concentrations determined by detergent compatible protein assay. Equal amounts of protein were subjected to SDS-polyacrylamide gel electrophoresis then transferred to a PVDF membrane. Membranes were probed with the rabbit anti-GRB2primary antibody (1:1000, Abcam #32111) or to glial fibrillary acidic protein (GFAP, goat polyclonal, 1:500, Abcam #ab53554) and then incubated with a secondary antibody conjugated with horseradish peroxidise. Protein bands were visualised using the G:Box BioImaging system. α/β-tubulin (rabbit polyclonal, 1:2000; Cell Signalling Technology #2148; Genesearch Pty Ltd, Arundel. QLD, Australia), served as a loading control. Quantitative analysis was conducted after normalisation of the protein band densitometry to α/β-tubulin.

Fluorescence immunolabelling in mouse

Mouse eyes were fixed in 4% paraformaldehyde and embedded in optimal cutting temperature compound as described previously[27]. For immunolabelling, frozen sections were blocked

with 5% normal goat serum and incubated with the GRB2 primary antibody (1:100, Abcam#32111) and GFAP (rabbit polyclonal, 1:250, Dako #Z0334) at +4°C overnight. After nuclear counterstaining with Hoechst, the bound GRB2 antibody was detected with an Alexa Fluor 488-conjugated secondary antibody diluted 1:1000 (Invitrogen Life Technologies, Mulgrave, VIC, Australia).

Results

Genome-wide genotyping was completed for 442 cases and 582 controls. Principal components analysis identified 69 participants (46 cases and 23 controls) as ethnic outliers and these were excluded from further analysis. A further 111 individuals had missing clinical covariate measures and were also excluded leaving 336 cases and 508 controls in the final analysis. Clinical and demographic details for the discovery cohort are given in Supplementary Online Table S1. As expected, duration of diabetes, HbA_{1c}, nephropathy and hypertension were all significantly associated with the presence of sight threatening diabetic retinopathy, as were age and sex. Genetic association analysis was conducted in plink using a logistic regression adjusted for age, sex, duration of diabetes, hypertension, nephropathy HbA_{1c} and three principal components. The genomic inflation factor (lambda) was 1.002 and the Q-Q plot is shown in Supplementary Figure S1.

We first assessed loci reported in previous GWAS of diabetic retinopathy [15-19] for association in the current study in an attempt to replicate previous findings. A *p*-value of $p<0.002$ was required to account for the 29 loci assessed. No previously reported SNPs showed replication at this level of significance, although several loci demonstrate nominal associations ($p<0.05$) and may be worthy of further investigation. (Supplementary Online Table S2).

No SNP reached genome-wide significance in the GWAS analysis (Figure 1B). The top ranked SNP is rs3805931 on chromosome 6 ($p=2.66x10^{-7}$, OR [95%CI] = 0.50 [0.39-0.65]), with a second SNP at the same locus, rs1537638, showing similar significance (Table 1). Eight SNPs at five loci demonstrated $p<1x10^{-5}$ (Table 1).

All SNPs with a $p<1x10^{-5}$ and a selection with $p<1x10^{-4}$ were assessed in the Caucasian type 2 replication cohort comprising 263 cases and 320 controls (Supplementary Online Table S3). Neither of the top ranked chromosome 6 SNPs showed association in the type 2 cohort and thus replication of the primary study findings was not achieved (Table 2). Of the other SNPs included, rs9896052 (*p* in discovery=$6.55x10^{-5}$, OR [95%CI]=1.67 [1.30-2.15]), was the only SNP associated in the type 2 replication cohort (p=0.035 OR [95%CI]=1.45 [1.03-2.18]) (Table 2). This SNP was assessed in a Caucasian type 1 cohort and an Indian type 2 cohort (Supplementary Online Table S3). SNP rs9896052 showed association in the type 1 cohort (p=0.041, OR [95%CI]=1.56 [1.02-2.38]) in the same direction as in the discovery and type 2 replication cohorts (Table 2). In the Indian cohort, rs9896052 was also associated with sight-threatening diabetic retinopathy (p=0.016, OR [95%CI] = 1.47[1.07-2.02]) (Table 3). Meta-analysis of all four cohorts gave a *p*-value of $4.15x10^{-8}$ (Table 3), meeting the criteria for genome-wide significance ($p<5x10^{-8}$).

Although the top ranked locus on chromosome 6 did not show replication in the type 2 cohort, the type 1 cohort did show a trend towards association at this locus (rs3805931 *p*=0.097, Table 2) and this region remains a candidate for further study in relation to diabetic retinopathy. Meta-analysis of the discovery cohort with the type 1 cohort gave a *p*-value of $1.69x10^{-8}$ at this SNP and it was the top ranked SNP in the full meta-analysis of all SNPs typed on the array. However,

as noted, it did not show association in the type 2 replication cohort. No other SNPs reached genome-wide significance in the meta-analysis (Supplementary Online Table S4).

SNP rs9896052 on chromosome 17 is within a 150kb block of linkage disequilibrium, which contains several genes: *KIAA0195*, *CASKIN2*, *TSEN54* and *LLGL2* (Figure 1C). All these transcripts are expressed in the retina (The Ocular Tissue Database, accessed October 2013, https://genome.uiowa.edu/otdb/), although very little is known about their function. According to ENCODE data (available at the UCSC genome browser, accessed October 2013), the SNP rs9896052 is located between two H3K27Ac marks at 73,401Mbp and 73,425Mbp. 17kb downstream of rs9896052, just outside this large LD block are two additional genes *GRB2* (Figure 1C) and the microRNA gene *MIR3678*, located immediately 5' to *GRB2*. Both these genes are excellent functional candidates for diabetic retinopathy and thus we have evaluated them in more detail. RT-PCR of human retina shows expression of mature *MIR3678* as well as both reported isoforms of *GRB2*, with the full length isoform 1 (NM_002086.4) more abundant than isoform 2 *(NM_203506.2)* which lacks exon 4 (Figure 2). To further investigate the retinal expression pattern we undertook immunohistochemistry for GRB2 in normal human retina. The protein is expressed throughout all layers of the retina including in blood vessel wall (Figure 2).

Direct sequencing of the coding regions of *GRB2* and *MIR3678* in 72 cases (Supplementary Online Table S5) identified known intronic SNPs (rs200248899 in 1 individual and rs12946365 with allele frequency of 0.257 as expected), but no coding variations, suggesting that coding mutations are unlikely to account for disease.

The response of GRB2 to retinal stress was explored in retina from a transgenic model of selective Müller cell ablation. In these mice, patchy loss of Müller cells leads to photoreceptor degeneration, blood-retinal barrier breakdown and the development of intra-retinal neovascularization, mimicking the features of diabetic retinopathy [27]. Retinal stress is demonstrated by a significant increase in GFAP labelling (Figure 3). Increased GRB2 labelling was qualitatively observed in the retinal ganglion cell and both inner and outer nuclear layers of the Müller cell knockout mice, particularly at 3 months post Müller cell ablation, and this difference was significantly different as demonstrated by quantitative Western blot (Figure 3).

Discussion

This study is the first to report association of this 17q25.1 locus with diabetic retinopathy. We show association of SNP rs9896052 with sight threatening diabetic retinopathy in Caucasian and Indian patients with type 2 diabetes as well as patients with type 1 diabetes. A genome-wide significan *p*-value of 4.15×10^{-8} was achieved on meta-analysis of all cohorts. This suggests that the association with retinopathy is independent of the type of diabetes and is a risk factor for retinopathy itself. The minor allele frequencies for rs9896052 for controls in both the Caucasian and Indian cohorts are similar to those in relevant HapMap populations (HapMap CEU Caucasian: 0.321, HapMap Gujarati Indians in Houston: 0.431) and the odds ratio is similar between all three replication cohorts (ranging from 1.45 to 1.56) demonstrating the robustness of this finding. Several additional SNPs at this locus were associated with diabetic retinopathy in the discovery cohort, however, no association was seen in the replication cohorts. Evaluation of the linkage disequilibrium (LD) patterns in each data set revealed that the discovery cohort demonstrates reasonably strong LD between rs9896052 and the two adjacent SNPs rs6501801 and rs9892171 (r^2=0.46 and 0.33 respectively). However, the type 2

replication cohort has less LD (r^2=0.29 and 0.20 respectively), explaining why the association is not seen at these nearby SNPs in this cohort (Supplementary Online Figure S2). It is not clear why the LD patterns differ, however, the association statistics observed are consistent with the LD patterns.

Several other GWAS for diabetic retinopathy have been published [15-19], however, none report a genome-wide significant signal at 17q25. Of note, four of the previously published studies were conducted in non-Caucasian populations (Chinese [17], Taiwanese [18], Japanese [19] and Mexican American [15]) while one was predominantly in patients with type 1 diabetes mellitus [16]. These fundamental differences may explain the lack of replication observed not only with the current study, but also with each other. As our study demonstrated consistent association at 17q25 across multiple types of diabetes and in different ethnic groups, there may be some genetic factors that are independent of diabetes type and ethnicity and others that are specific. A limitation of this and previously published studies possibly affecting reproducibility is the size of the cohort and larger studies and meta-analyses are required to provide a clearer picture of the genetic architecture of diabetic retinopathy.

We restricted the current study to severe forms of diabetic retinopathy and compared them to contemporaneously recruited diabetic patients with no or minimal diabetic retinopathy. The inclusion of minimal diabetic retinopathy in the control group recognises that at this end of the spectrum differences in the grading have limited meaning for clinical decision making as even a single microaneurysm could constitute minimal diabetic retinopathy. The use of clinical grading rather than retinal photography could be considered a weakness of this study, however, the use of the extreme phenotype negates the need for detailed grading as there is strong

agreement between clinical assessments and photographic grading for categorising diabetic retinopathy [28].

It is yet to be determined which of the genes in this region are involved in the pathogenesis of diabetic retinopathy, however, several annotated transcripts, notably *MIR3678* and *GRB2*, make compelling candidates. Although our data indicate that coding mutations in these genes are unlikely to account for disease, it is possible that SNP rs9896052 or a SNP in linkage disequilibrium with it, is in a regulatory region that alters the expression of these genes.

Little is known specifically about *MIR3678*, although amongst many predicted targets (miRbase.org) is *EGLN3,* which is a HIF prolyl hydroxylase important in oxygen sensing and the response to hypoxia, a pathway highly relevant to diabetic retinopathy. *MIR3678* is not annotated in rodent genomes adding complexity to assessing the role of this transcript in animal models of diabetic retinopathy. Further work in human tissues will be required to assess its role in this disease.

Functionally, *GRB2* is a promising candidate for diabetic retinopathy susceptibility. It binds phosphorylated insulin receptor substrate 1 (IRS1) and subsequently activates the MAPK pathway via Ras in response to insulin [29]. It is also involved in VEGF signalling leading to angiogenesis which is characteristic of proliferative diabetic retinopathy [30]. *GRB2* encodes two transcripts in humans [31]. Isoform 1 is expressed in most human tissues and isoform 2 is either absent or expressed at considerably lower levels relative to isoform 1 [31]. Isoform 2, known as GRB3-3 lacks a portion of the SH2 domain and thus does not bind to the phosphorylated tyrosine kinase receptor [31]. It appears to act as a negative regulator of isoform 1 and may be involved in inducing apoptosis [31].This study is the first to show expression of

both isoforms in the human retina. The GRB2 protein was observed in all layers of the human and mouse retina and we demonstrate that retinal stress leading to neovascular retinopathy in a rodent model results in increased expression of the protein consistent with a role in diabetic retinopathy.

Further work assessing all the genes in this region is required to determine if *GRB2*, *MIR3678* or one of the other less well characterised genes at this locus is the primary contributor to diabetic retinopathy and to determine the mechanism of action of the genetic predisposition identified in this study.

Acknowledgements

We thank the following people for technical assistance with patient recruitment, DNA extraction and genotyping: Emmanuelle Souzeau, Kathleen Dowell, Abraham Kuot and Stacey Thorpe (Department of Ophthalmology, Flinders University, Australia), Liudmila Kolmogorova (Save Sight Institute, University of Sydney, Sydney, Australia), Jennfier Haariko (Ophthalmology, Australian National University, Canberra, Australia), Fiona Richter (Department of Ophthalmology, Royal Melbourne Hospital, Victoria, Australia) and Johanna Hadler (The University of Queensland, Diamantina Institute, Brisbane, Australia).

Funding

This work was supported by the National Health and Medical Research Council (NHMRC) of Australia [#595918] and the Ophthalmic Research Institute of Australia. KPB is supported by a Senior Research Fellowship from the NHMRC and JEC by a Practitioner Fellowship from the NHMRC. Research conducted at Moorfields Eye Hospital was funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital

NHS Foundation Trust (www.nihr.ac.uk) and UCL Institute of Ophthalmology. Research at the Centre for Eye Research Australia was funded by the Australian Research Council (www.arc.gov.au) grant LP0884108.

Duality of Interest Statement

The authors have no duality of interest to report

Author contributions

KPB, SA, NP, MAB, JEC designed the study; SA, GK, AWH, MD, RWE, JHC, SRL, BP, AJ, ELL, KR, MP, PGH, NP, MCG recruited patients; RDF, WS, GK, BA, SS, SK, GG conducted laboratory experiments; KPB, RDF, GK, MAB conducted statistical analysis; KPB, RDF, WS, SS wrote the paper; all authors reviewed and critiqued the paper and contributed to data interpretation. KPB is responsible for the integrity of the work as a whole.

References

1. National Diabetes Data Group, Diabetes in America, 2nd Ed. 1995, National Institutes of Health; National Institute of Diabetes and Digestive and Kidney Diseases.
2. Kempen JH, O'Colmain BJ, Leske MC, et al. (2004) The prevalence of diabetic retinopathy among adults in the United States. Arch Ophthalmol. 122:552-63.
3. Wild S, Roglic G, Green A, Sicree R, King H (2004) Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care. 27:1047-53.
4. National Health and Medical Research Council, Clinical practice guidelines for the management of diabetic retinopathy. 1 ed., Canberra, Editor. 1997, .
5. The Diabetes Control and Complications Trial Research Group (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. . N Engl J Med. 329:977-86.
6. UK Prospective Diabetes Study (UKPDS) Group (1998) Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). . Lancet. 352:837-53.
7. Gerstein HC, Miller ME, Byington RP, et al. (2008) Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med. 358:2545-59.
8. The Diabetes Control and Complications Trial Research Group (1997) Clustering of long-term complications in families with diabetes in the diabetes control and complications trial. . Diabetes. 46:1829-39.
9. Rema M, Saravanan G, Deepa R, Mohan V (2002) Familial clustering of diabetic retinopathy in South Indian Type 2 diabetic patients. Diabet Med. 19:910-6.

10. Hallman DM, Huber JC, Jr., Gonzalez VH, Klein BE, Klein R, Hanis CL (2005) Familial aggregation of severity of diabetic retinopathy in Mexican Americans from Starr County, Texas. Diabetes Care. 28:1163-8.

11. Hietala K, Forsblom C, Summanen P, Groop PH (2008) Heritability of proliferative diabetic retinopathy. Diabetes. 57:2176-80.

12. Arar NH, Freedman BI, Adler SG, et al. (2008) Heritability of the severity of diabetic retinopathy: the FIND-Eye study. Invest Ophthalmol Vis Sci. 49:3839-45.

13. Monti MC, Lonsdale JT, Montomoli C, Montross R, Schlag E, Greenberg DA (2007) Familial risk factors for microvascular complications and differential male-female risk in a large cohort of American families with type 1 diabetes. J Clin Endocrinol Metab. 92:4650-5.

14. Abhary S, Hewitt AW, Burdon KP, Craig JE (2009) A systematic meta-analysis of genetic association studies for diabetic retinopathy. Diabetes. 58:2137-47.

15. Fu YP, Hallman DM, Gonzalez VH, et al. (2010) Identification of Diabetic Retinopathy Genes through a Genome-Wide Association Study among Mexican-Americans from Starr County, Texas. Journal of ophthalmology. 2010.

16. Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL (2011) Genome-wide Meta-analysis for Severe Diabetic Retinopathy. Human molecular genetics.

17. Sheu WH, Kuo JZ, Lee IT, et al. (2013) Genome-wide association study in a Chinese population with diabetic retinopathy. Hum Mol Genet. 22:3165-73.

18. Huang YC, Lin JM, Lin HJ, et al. (2011) Genome-wide Association Study of Diabetic Retinopathy in a Taiwanese Population. Ophthalmology. 118:642-8.

19. Awata T, Yamashita H, Kurihara S, et al. (2014) A genome-wide association study for diabetic retinopathy in a Japanese population: potential association with a long intergenic non-coding RNA. PLoS One. 9:e111715.

20. Kaidonis G, Abhary S, Daniell M, et al. (2014) Genetic study of diabetic retinopathy: recruitment methodology and analysis of baseline characteristics. Clin Experiment Ophthalmol. 42:486-93.

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

22. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16:1215.

23. Purcell S, Neale B, Todd-Brown K, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81:559-75.

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

25. Pruim RJ, Welch RP, Sanna S, et al. (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 26:2336-7.

26. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 26:2190-1.

27. Shen W, Fruttiger M, Zhu L, et al. (2012) Conditional Mullercell ablation causes independent neuronal and vascular pathologies in a novel transgenic model. J Neurosci. 32:15715-27.

28. Gangaputra S, Lovato JF, Hubbard L, et al. (2013) Comparison of standardized clinical classification with fundus photograph grading for the assessment of diabetic retinopathy and diabetic macular edema severity. Retina. 33:1393-9.

29. Skolnik EY, Lee CH, Batzer A, et al. (1993) The SH2/SH3 domain-containing protein GRB2 interacts with tyrosine-phosphorylated IRS1 and Shc: implications for insulin control of ras signalling. The EMBO journal. 12:1929-36.

30. Anselmi F, Orlandini M, Rocchigiani M, et al. (2012) c-ABL modulates MAP kinases activation downstream of VEGFR-2 signaling by direct phosphorylation of the adaptor proteins GRB2 and NCK1. Angiogenesis. 15:187-97.

31. Fath I, Schweighoffer F, Rey I, et al. (1994) Cloning of a Grb2 isoform with apoptotic properties. Science. 264:971-4.

Table 1. Top ranked Single Nucleotide Polymorphisms (SNPs) associated with sight threatening diabetic retinopathy in the discovery cohort. Results are shown for SNP with $p<1.0x10^{-5}$ and odds ratios are calculated with respect to the minor allele. *P*-values are adjusted for age, sex, duration of diabetes, HbA_{1c}, hypertension, nephropathy and three principal components. MAF=minor allele frequency, OR (95%CI)=odds ratio with 95% confidence interval.

Chr	SNP	Position in hg18 (bp)	Minor allele	MAF cases	MAF controls	OR (95% CI)	*p*-value	Nearest Gene
6	rs3805931	43197407	A	0.312	0.387	0.50 (0.39-0.65)	$2.66x10^{-7}$	PTK7
6	rs1537638	43204177	C	0.356	0.443	0.51 (0.40-0.66)	$3.11x10^{-7}$	PTK7
20	rs6128541	57351084	C	0.308	0.216	1.95 (1.48-2.58)	$2.43x10^{-6}$	EDN3
21	rs6516749	27279870	T	0.184	0.252	0.50 (0.38-0.679)	$2.76x10^{-6}$	ADAMTS5
17	rs9038	73006992	C	0.460	0.374	1.79 (1.40-2.29)	$3.41x10^{-6}$	SEPT9
20	rs6128542	57351252	T	0.330	0.236	1.85 (1.42-2.41)	$4.41x10^{-6}$	EDN3
21	rs10482979	27238153	C	0.191	0.268	0.53 (0.40-0.70)	$8.47x10^{-6}$	ADAMTS5
21	rs11702862	27167220	G	0.207	0.291	0.53 (0.41-0.71)	$9.34x10^{-6}$	ADAMTS1

Table 2: Analysis of top ranked GWAS SNPs in the discovery and replication cohorts. The primary replication cohort consists of Australian patients with type 2 diabetes and sight threatening diabetic retinopathy and the secondary cohort of patients with type 1 diabetes with any grade of diabetic retinopathy. Odds ratios (OR) are calculated with respect to the minor allele. Chr=chromosome, SNP position is given in hg18

				Discovery cohort		Type 2 replication cohort		Type 1 replication cohort	
Chr	SNP	Position	Allele 1	*p*-value[a]	OR (95%CI)	*p*-value[b]	OR (95%CI)	*p*-value[b]	OR (95%CI)
1	rs1334802	37168926	G	$4.76x10^{-5}$	2.11 (1.47-3.02)	0.987	1.00 (0.60-1.69)	0.683	1.14 (0.61-2.12)
1	rs6687106	69949821	A	$1.45x10^{-5}$	0.54 (0.41-0.71)	0.462	1.15 (0.79-1.67)	0.167	1.39 (0.87-2.20)
1	rs1340773	69958624	T	$2.21x10^{-5}$	0.55 (0.42-0.73)	0.398	1.17 (0.81-1.70)	0.231	1.33 (0.84-2.11)
3	rs6550133	32447284	C	$3.60x10^{-5}$	2.00 (1.44-2.77)	0.804	0.93 (0.54-1.62)	0.961	1.01 (0.55-1.87)
4	rs1353854	30767663	G	$1.25x10^{-5}$	0.36 (0.23-0.57)	0.835	0.93 (0.48-1.81)	0.157	0.61 (0.30-1.21)
5	rs17343615	25063375	T	$5.30x10^{-5}$	2.12 (1.47-3.05)	0.629	0.87 (0.49-1.53)	0.898	1.04 (0.60-1.78)
5	rs10520899	25103129	G	$4.93x10^{-5}$	2.05 (1.45-2.91)	0.471	1.21 (0.72-2.05)	0.434	0.82 (0.50-1.35)
6	rs3805931	43197407	A	$2.66x10^{-7}$	0.50 (0.39-0.65)	0.464	0.87 (0.61-1.25)	0.097	0.70 (0.46-1.07)
6	rs1537638	43204177	C	$3.11x10^{-7}$	0.51 (0.40-0.66)	0.151	0.77 (0.54-1.10)	**0.025**	**0.63 (0.43-0.95)**
7	rs3095031	112000000	A	$2.56x10^{-5}$	0.48 (0.35-0.68)	0.159	0.73 (0.46-1.14)	0.626	0.88 (0.53-1.46)

11	rs11228739	56239803	A	$5.35x10^{-5}$	2.81 (1.70-4.64)	0.246	0.63 (0.29-1.38)	0.907	1.05 (0.47-2.34)
11	rs12289069	56615226	C	$6.97x10^{-5}$	3.22 (1.81-5.74)	0.976	1.01 (0.50-2.04)	0.224	1.78 (0.70-4.50)
16	rs480727	87399730	G	$2.96x10^{-5}$	1.66 (1.31-2.11)	0.362	0.86 (0.62-1.19)	0.547	0.89 (0.60-1.31)
17	rs9896052	70930457	A	$6.55x10^{-5}$	1.67 (1.30-2.15)	**0.035**	1.50 (1.03-2.18)	**0.041**	**1.56 (1.02-2.38)**
17	rs6501801	70958625	C	$4.26x10^{-5}$	1.83 (1.37-2.43)	0.147	1.34 (0.90-1.98)	0.845	1.05 (0.64-1.72)
17	rs9892171	70964124	C	$7.26x10^{-5}$	1.93 (1.40-2.67)	0.381	1.22 (0.78-1.89)	0.992	1.00 (0.58-1.73)
17	rs9038	73006992	C	$3.41x10^{-6}$	1.79 (1.40-2.29)	0.459	0.88 (0.63-1.23)	0.817	1.05 (0.69-1.59)
19	rs8105903	51979990	C	$2.25x10^{-5}$	1.67 (1.32-2.12)	0.968	1.01 (0.73-1.40)	0.412	0.85 (0.58-1.25)
20	rs6128541	57351084	C	$2.43x10^{-6}$	1.95 (1.48-2.58)	0.281	0.80 (0.54-1.20)	0.128	0.71 (0.45-1.11)
20	rs6128542	57351252	T	$4.41x10^{-6}$	1.85 (1.42-2.41)	0.222	0.78 (0.53-1.16)	0.216	0.76 (0.49-1.18)
21	rs11702862	27167220	G	$9.34x10^{-6}$	0.53 (0.41-0.71)	0.566	0.89 (0.59-1.33)	0.716	0.92 (0.60-1.43)
21	rs13048083	27208724	T	$2.20x10^{-5}$	0.55 (0.42-0.73)	0.995	1.00 (0.68-1.46)	0.702	0.92 (0.61-1.40)
21	rs965135	27211535	C	$4.34x10^{-5}$	0.55 (0.42-0.73)	0.709	0.92 (0.61-1.40)	0.981	1.01 (0.65-1.55)
21	rs1444269	27213326	G	$2.20x10^{-5}$	0.55 (0.42-0.73)	0.850	0.96 (0.66-1.41)	0.702	0.92 (0.61-1.40)
21	rs3746836	27214452	A	$2.20x10^{-5}$	0.55 (0.42-0.73)	0.887	0.97 (0.66-1.43)	0.702	0.92 (0.61-1.40)
21	rs11700721	27215795	T	$4.34x10^{-5}$	0.55 (0.42-0.73)	0.984	1.00 (0.66-1.50)	0.981	1.01 (0.65-1.55)
21	rs10482979	27238153	C	$8.47x10^{-6}$	0.53 (0.40-0.70)	0.712	0.93 (0.62-1.39)	0.652	0.90 (0.58-1.40)

21	rs6516749	27279870	T	$2.76x10^{-6}$	0.50 (0.38-0.67)	0.752	0.94 (0.62-1.41)	0.639	1.11 (0.71-1.75)
21	rs4566449	44043952	C	$2.68x10^{-5}$	1.81 (1.37-2.40)	0.947	1.01 (0.69-1.49)	0.268	1.31 (0.81-2.11)

a) adjusted for age, sex, duration of diabetes, hypertension, nephropathy, HbA_{1c} and 3 principal components

b) adjusted for age, sex, duration of diabetes, hypertension, nephropathy, HbA_{1c}

Table 3. Replication and meta-analysis of rs9896052 in multiple cohorts. Replication analysis was conducted in Caucasian replication cohorts with type 2 or type 1 diabetes mellitus and Indian cohort with type 2 diabetes mellitus. Meta-analysis was conducted on the adjusted *p*-values for each cohort with respect to the minor allele (and risk allele) A. MAF=minor allele frequency, OR (95%CI)=Odds Ratio (95% confidence interval) Het=heterogeneity

Cohort	**Cohort**	**MAF**		***p*-value**		**OR (95%CI)**	**Meta-analysis**		
ID	**Description**	**cases**	**controls**	**unadjusted**	**adjusted**	**adjusted**	**Cohorts**	***p*-value**	**Het *p*-val**
1	Discovery	0.410	0.294	$1.35x10^{-7}$	$6.55x10^{-5}$ [a]	1.67 (1.30-2.15)	-		
2	Type 2	0.416	0.365	0.077	0.035[b]	1.45(1.03-2.18)	1+2	$6.79x10^{-7}$	0.702
3	Type 1	0.353	0.270	0.023	0.041[b]	1.56(1.02-2.38)	1+2+3	$8.21x10^{-7}$	0.886
4	Indian	0.489	0.424	0.015	0.016[b]	1.47 (1.07-2.02)	1+2+3+4	$4.15x10^{-8}$	0.962

a) adjusted for age, sex, duration of diabetes, hypertension, nephropathy, HbA_{1c} and 3 principal components

b) adjusted for age, sex, duration of diabetes, hypertension, nephropathy, HbA_{1c}

Figures Legends

Figure 1. Results of Genome-wide Association Study for sight threatening diabetic retinopathy in type 2 diabetes mellitus. A. Manhattan plot showing association results for sight-threatening diabetic retinopathy. The black dashed horizontal line indicates genome-wide significance level of $p=5x10^{-8}$ and the grey line represents suggestive association at $p=1x10^{-6}$. **B.** The associated chromosome 17 locus at SNP rs9896052 showing $-\log_{10}(p$-values) in the discovery cohort (dots, left hand axis), average recombination rates in Caucasians (solid lines, right hand axis) and annotated protein coding transcripts in the region below.

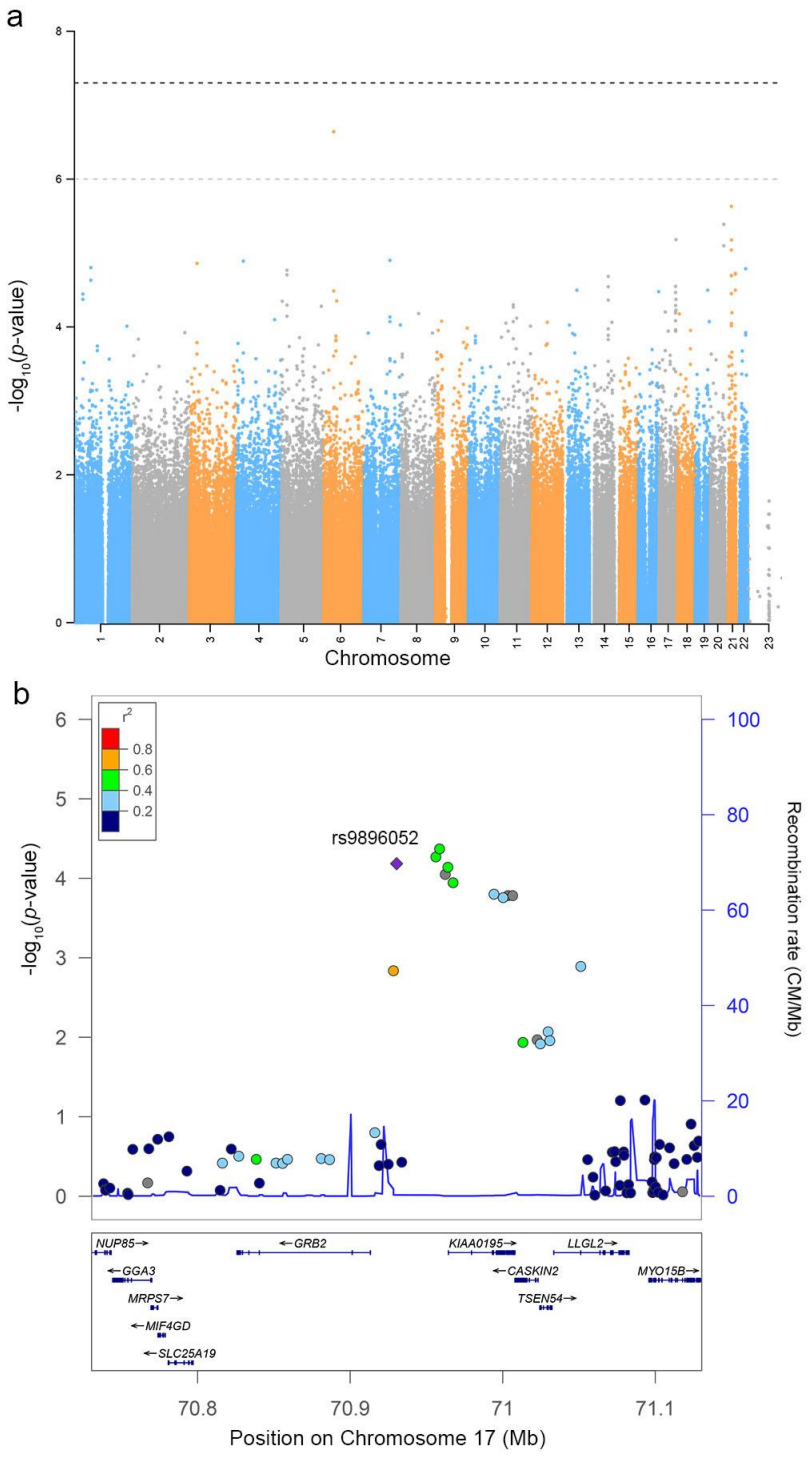

Figure 2: Expression and distribution of candidate genes in normal human retina A. RT-PCR for *MIR3678* **B.** RT-PCR for two transcripts of *GRB2*. **C.** Immunohistochemistry with the anti-GRB2 antibody showing ubiquitous positive immune-labelling through the retinal tissue and in retinal blood vessel endothelium (arrow). **D**. Negative control. GCL, ganglion cell layer; IPL, inner plexiform layer; INL, inner nuclear layer; OPL, outer plexiform layer; ONL, outer nuclear layer; IS, inner segment; OS, outer segment. Scale bar=20 μm.

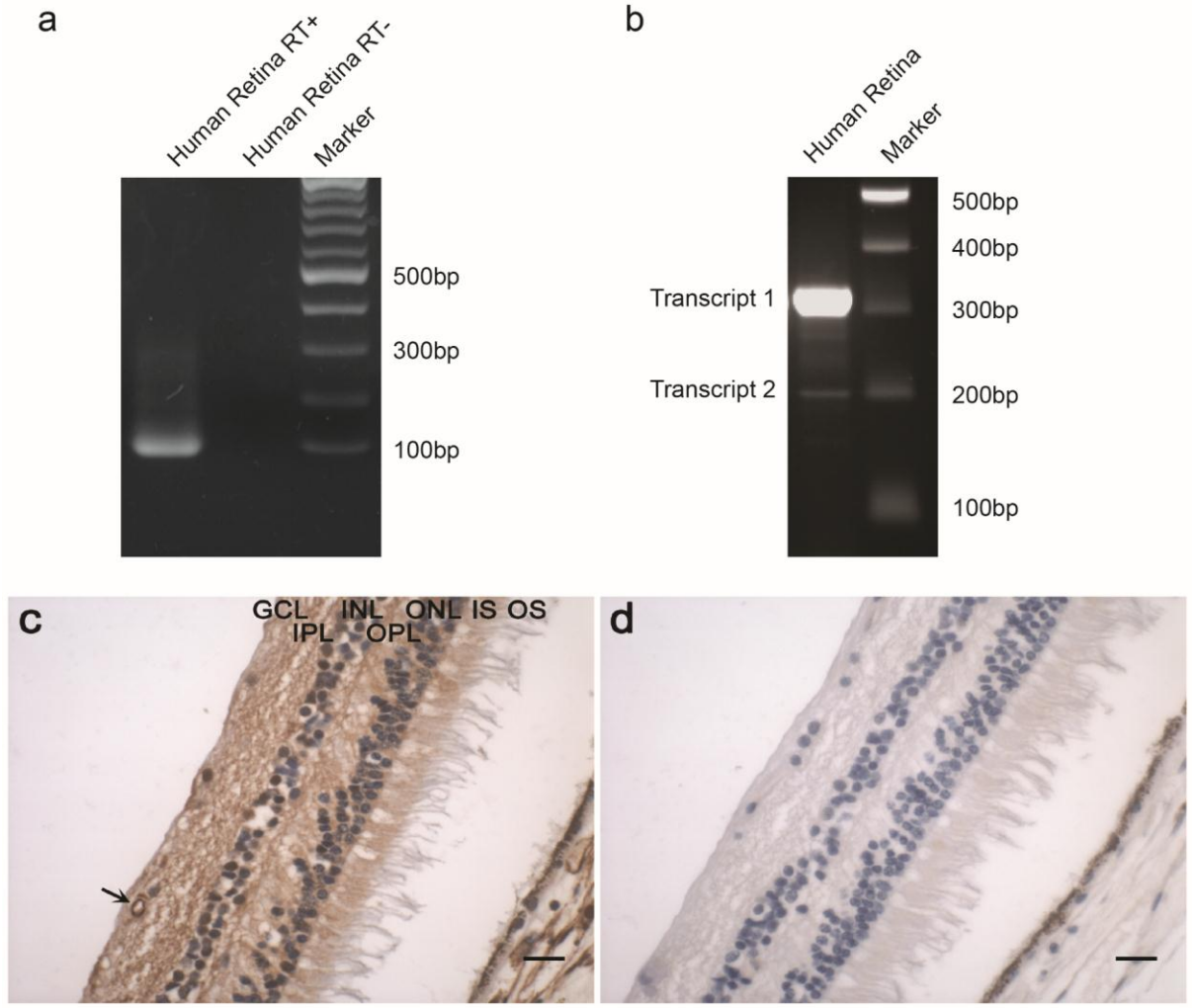

Figure 3: **Differential expression of GRB2 in retinae of transgenic mice after selective Müller cell ablation. (A, C, E):** GRB2 immunolabelling (green) with Hoechst nuclear counterstaining (red). (**B, D, F):** GFAP immunolabelling (green) with Hoechst nuclear counterstaining (red). **(A, B):** control mice. **(C-F):** transgenic mice 2 weeks (C, D) and 3 months (E, F) after Muller cell ablation. INL=inner nuclear layer, ONL=outer nuclear layer. Arrows in C and D point to areas of protrusion of degenerated photoreceptors after patchy Müller cell ablation. Note: profound activation of surviving Muller cells after patchy loss of Müller cells in (D and F). scale bars in (A-F): 50µm. **G.** Western blots for GRB2 and GFAP at 3 months after Müller cell ablation. α/β Tubulin is included as a loading control. **H** shows significant up-regulation of GRB2 in transgenic mice compared to controls and **I** confirms upregulation of GFAP as expected. TG=transgenic, Ctl=control. **p<0.01