The Genetics of Primary Open-Angle Glaucoma: A Complex Human Disease

by

Megan Rebecca Ulmer Carnes

University Program in Genetics and Genomics
Duke University

Date:_____

Approved:

_____
Michael A. Hauser, Supervisor

_____
Allison E. Ashley-Koch

_____
Elizabeth Hauser

_____
Beth Sullivan

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in the University Program in Genetics and Genomics
in the Graduate School of Duke University

2014

ABSTRACT

The Genetics of Primary Open-Angle Glaucoma: A Complex Human Disease

by

Megan Rebecca Ulmer Carnes

University Program in Genetics and Genomics
Duke University

Date:_____

Approved:

_____
Michael A. Hauser, Supervisor

_____
Allison E. Ashley-Koch

_____
Elizabeth Hauser

_____
Beth Sullivan

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in the University Program in Genetics and Genomics
in the Graduate School of Duke University

2014

Abstract

Glaucoma is a chronic ocular neuropathy and a leading cause of blindness worldwide. Primary open-angle glaucoma (POAG) is the most common subtype with an estimated 2 million affected individuals in the Unites States. POAG is a heritable complex trait. Understanding the genetics of POAG may increase our ability to predict disease onset and help elucidate the underlying biological mechanisms responsible for the development of the disease. With this overall goal, three different approaches are presented here.

First, the genetics of an important POAG-associated trait, central corneal thickness (CCT), was investigated using genome-wide single nucleotide polymorphism (SNP) data available from the NEIGHBOR and GLAUGEN consortia to identify novel POAG candidate genes. Twenty previously published CCT-associated SNPs were tested for association with both CCT (N = 1,117) and POAG (N = 6,470). While several of these variants were significantly associated with CCT in our dataset (top SNP = rs12447690, near *ZNF469* (beta = -5.08 μm/allele; p = 0.001), none were associated with POAG. A CCT genome-wide association study was conducted. Using a p-value threshold of $1X10^{-4}$, 50 candidate SNPs were tested for association with POAG. One SNP, rs7481514, within the *NTM* gene was significantly associated with POAG in a low tension subset of cases (odds ratio (OR) = 1.28; p = 0.001). Additionally, SNPs in the *CNTNAP4* gene

showed suggestive evidence of association with POAG (top SNP = rs1428758; OR = 0.84; p = 0.018). A gene expression analysis showed evidence of *NTM* and *CNTNAP4* gene expression in relevant ocular tissues. This study suggests previously reported CCT loci do not increase POAG susceptibility. However, by using a two-step gene mapping approach, the cell adhesion molecules, *NTM* and *CNTNAP4*, were identified as potential POAG candidate genes in a subset of cases.

The second study aimed to identify functional alleles within a POAG candidate gene. Previous association studies identified a significant association between POAG and the *SIX6* locus (top SNP = rs10483727, OR = 1.32, $p = 3.87X10^{-11}$). *SIX6* plays a role in ocular development and has been associated with the morphology of the optic nerve. Sequencing of the *SIX6* coding and regulatory regions in 262 POAG cases and 256 controls identified six nonsynonymous coding variants. Of these six, five were rare (minor allele frequency (MAF) ≤ 0.002) and one, Asn141His (rs33912345), was a common variant that showed a significant association with POAG (OR = 1.27, $p = 4.2X10^{-10}$) in the NEIGHBOR/GLAUGEN dataset. These variants were tested in an *in vivo* zebrafish complementation assay to evaluate ocular metrics. Five of the six alleles had a functional effect on the protein. These five variants, found primarily in POAG cases, were hypomorphic or null, while a sixth variant, found only in controls, was benign. One variant in the *SIX6* enhancer increased expression of the *SIX6* gene and disrupted its regulation. Using optical coherence tomography, the retinal thickness of POAG with and

without the common *SIX6* risk allele, Asn141His (rs33912345), was measured. Patients who are homozygous for the *SIX6* risk allele (His141) have a statistically thinner retinal nerve fiber layer than patients homozygous for the *SIX6* non-risk allele (Asn141). These results in combination with previous SIX6 work, leads us to hypothesize that *SIX6* risk variants disrupt the development of the neural retina and result in a reduced number of retinal ganglion cells generated during development, thereby increasing the risk of glaucoma-associated vision loss later in life.

Next, the transcriptional landscape of three POAG-related tissues; the trabecular meshwork, the cornea, and the ciliary body; were evaluated using an RNA sequencing (RNA-seq) approach. Tissues were selected from two fetal and four adult human donor samples with no known history of ocular disease. Deep RNA-seq was performed, and the total number of paired reads per sample ranged from 32,137,380 to 59,784,117. A descriptive analysis was conducted and included the identification of the top most expressed genes in each tissue and the distribution of gene expression values. Additionally, gene expression of selected POAG candidate genes (*CDKN2B, CDKN2A, CDKN2B-AS, SIX6*, *SRBD1, ATOH, CAV1*, *CAV2*, *ELOVL5,* and *TMCO1*) was evaluated. Most of these genes showed high expression values in the trabecular meshwork and cornea. *ATOH* was only found to be expressed in the fetal TM and, interestingly, *SIX6* was shown to be highly expressed in the adult and fetal ciliary body. *CDKN2B-AS* was not found to be expressed in any of the tissues evaluated. Finally, the RNA-seq data was

used to identify potential novel isoforms of these candidate genes. Using a stringent threshold, five novel isoforms were identified in *CDKN2B*, *SRBD1*, and *SIX6*. The data generated as part of this study can be used to develop novel hypotheses, guide future work, and is broadly applicable for ocular research because the tissues included in this analysis are essential for normal vision and play important roles in ocular diseases.

In this dissertation, three different approaches (assessment of a quantitative risk factor, candidate gene functional analysis, and the assessment of the transcriptional landscape of relevant ocular tissues) were used to study the common blinding disorder, primary open-angle glaucoma. Continued research in this field is essential. There is a need for increased functional follow-up of genetic association studies in order to identify true causal susceptibility genes and improve our understanding of POAG biology. Additionally, researchers should focus on building and implementing accurate prediction models to increase POAG diagnosis rates and preemptive treatment.

Dedication

I dedicate this dissertation to Drs. Michael Hauser and Allison Ashley-Koch, whose unparalleled mentorship and support made this work possible. This dissertation is also dedicated to my best friend, my husband Michael Carnes, who is always by my side and can always make me smile.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to express deep gratitude to Dr. Lydia Kwee for providing everything from statistical advising, code, manuscript editing, and career advice to moral support and laughter. Thank you to the members of the Hauser laboratory and everyone at the Duke Center for Human Genetics for creating an amazing atmosphere for my education and providing a lot of help along the way. I would like to thank the members of my thesis committee, Drs. Michael Hauser, Allison Ashley-Koch, Elizabeth Hauser, and Beth Sullivan, who are amazing teachers and encouraged me throughout graduate school. Also, thanks to my colleagues and friends, Drs. Christina Markunas and Kristin McDonald Gibson, who taught me a lot about both genetics and life.

Thank you to the members of the NEIGHBOR and GLAUGEN consortia and the funding institution, the National Eye Institute. I was privileged to work with this team and learned an enormous amount from the opportunity. I would also like to sincerely thank Dr. Rand Allingham for spending hours explaining the details of ocular biology to me and for always encouraging me to persevere.

1. Introduction

1.1 Primary open-angle glaucoma (POAG) epidemiology

Glaucoma is a chronic, bilateral optic neuropathy characterized by the progressive loss of retinal ganglion cells and visual field. It is the second leading cause of blindness with an estimated 60 million affected individuals worldwide (Resnikoff, Pascolini et al. 2004; Quigley and Broman 2006). Glaucoma is divided into subgroups based on clinical presentation. The most common subtype, primary open-angle glaucoma (POAG), is classified by the absence of a known secondary cause and an open iridocorneal angle (Allingham, Liu et al. 2009; Kwon, Fingert et al. 2009).

There are over 2 million POAG patients in the United States alone, and it is the leading cause of blindness in African Americans (Tielsch, Sommer et al. 1991; Racette, Wilson et al. 2003; Quigley and Broman 2006; Kwon, Fingert et al. 2009). The mean prevalence of POAG in individuals over the age of 40 from eight geographic regions are estimated to be the following: Africa 4.16%; Japan 3.31%; Latin American 3.16%; Europe and the U.S 1.97%; India 1.75%; China 1.40%; Middle East 1.31%; and South East Asia 1.18% (Quigley and Broman 2006). There are 4.47 million individuals with bilateral blindness due to POAG, and this number is estimated to grow as high as 5.6 million by 2020 (Quigley and Broman 2006). POAG onset and progression typically goes unnoticed

until visual impairment and irreversible optic nerve damage has occurred. The diagnosis rates are only 34% and 8% in developed and developing countries respectively, highlighting the importance of research aimed at increasing our understanding of POAG etiology (Quigley and Broman 2006).

1.2 POAG biology and treatment

Tissues from both the anterior and posterior portions of the eye are involved in the development of POAG, although ultimately POAG is defined by visual field loss due to retinal ganglion cells (RGC) death. The axons of these cells form the optic nerve, which is responsible for transmitting visual information to the brain's visual cortex. Optic nerve deformation caused by elevated ocular pressure is believed to be a primary contributor to POAG. Intraocular pressure is regulated by the intricate balance of aqueous humor inflow and outflow. Aqueous humor, generated from the ciliary body, is a clear, watery fluid used for circulation in the eye. It flows from the ciliary body, through the anterior chamber, and out through the trabecular meshwork, a porous and dynamic tissue. Elevated ocular pressure is the result of aqueous humor outflow occlusion (Kwon, Fingert et al. 2009). Blockage of this circulation pattern is observed in POAG patients and causes ocular hypertension, resulting in RGC loss via mechanical stress to the retina and optic nerve (Ray and Mookherjee 2009; Janssen, Gorgels et al. 2013). Other purposed mechanisms include retina and optic nerve head blood perfusion

deficiency, oxidative stress, glutamate excitotoxicity, autoimmunity and inflammation, among others (Kuehn, Fingert et al. 2005; Ray and Mookherjee 2009; Janssen, Gorgels et al. 2013). In actuality, it is likely that POAG is caused by a combination of these factors along with environmental and genetic effects. Although the field of POAG research is dense, the underlying biological mechanisms remain poorly understood due to the complex and heterogeneous nature of the disease. More work is needed to fully elucidate the true processes responsible for POAG development.

POAG treatments have traditionally focused on the reduction of intraocular pressure, considered the only treatable risk factor. Patients may either be prescribed pressure reducing eye drops, undergo ocular surgery, or both. Topical medications are often associated with systemic side effects and suffer from patient compliance issues (Kulkarni et al.) Surgical techniques aimed at improving aqueous humor drainage are often used as a second line of therapy. The most aggressive surgical technique, a trabeculectomy, is effective for the short term reduction in intraocular pressure (Boland, Ervin et al. 2013). However, there is a risk of complications associated with ocular surgery, patients often require repeat treatment, and the long-term effectiveness is unclear (Wilensky and Chen 1996; Kulkarni, Damji et al. 2008; Allingham and Shields 2011). Treatment methods reduce intraocular pressure by either decreasing aqueous humor production or by enabling increased outflow capabilities. It should be noted that

not all POAG cases present with elevated ocular pressure, and currently it is unclear if normal tension glaucoma (POAG with an absence of ocular hypertension) shares the same underlying biological mechanism as high tension glaucoma. However, studies show that ocular pressure reducing therapies mitigate disease progression even in normal tension glaucoma cases (Collaborative Normal-Tension Glaucoma Study Group 1998; Heijl, Leske et al. 2002; Miglior, Pfeiffer et al. 2007).

1.3 POAG risk factors

Elevated intraocular pressure is the most well established risk factor for the development and progression of POAG. Other commonly accepted risk factors include age, family history, race, thinner corneal thickness, and optic nerve cupping parameters (measurements commonly used to diagnosis and track POAG progression) (Gordon, Beiser et al. 2002; Leske, Heijl et al. 2007; Miglior, Pfeiffer et al. 2007; Kwon, Fingert et al. 2009; Janssen, Gorgels et al. 2013). Sex, diabetes mellitus, decreased diastolic profusion pressure, cardio vascular disease, and myopia have all been considered as potential risk factors, but the results have been inconsistent (Miglior, Pfeiffer et al. 2007; Kwon, Fingert et al. 2009; Gibson, Griffiths et al. 2012). Additionally, there is growing evidence that intracranial pressure may play a role in POAG pathogenicity. It is hypothesized that intracranial pressure balances the opposing intraocular pressure to sustain the integrity of the optic nerve head. Thus, a reduced intracranial pressure results in an increased

sensitivity to elevated or fluctuating intraocular pressure (Fleischman and Allingham 2013; Janssen, Gorgels et al. 2013).This emphasizes the importance of intraocular pressure regulation in POAG. Individuals with intraocular pressure above the normal range (10 to 21 mm Hg) have a 3 -39 times increased risk of developing POAG, depending on the level of ocular hypertension (Sommer, Tielsch et al. 1991; Kwon, Fingert et al. 2009). It is estimated that 30-40% of POAG cases have an intraocular pressure in the normal range, and this is estimated to be even higher in Hispanic and Japanese populations (Dielemans, Vingerling et al. 1994; Iwase, Suzuki et al. 2004; Varma, Ying-Lai et al. 2004; Mudumbai 2013). A decreased intracranial pressure may play a larger role in this subset of POAG.

POAG is generally considered a late age disease, occurring after the age of 40. Alternative forms include primary congenital glaucoma (birth to 3 years age of onset) and juvenile open-angle glaucoma (3 to 20 years age of onset), although these are generally considered to be separate diseases (Kwon, Fingert et al. 2009). Increased age is known to increase the risk of developing POAG by 20-40% for each decade after the age of 40 (Coleman and Miglior 2008).

Many quantitative ocular measurements are considered POAG risk factors including corneal thickness and optic nerve head measurements, particularly vertical cup-disc ratio. Central corneal thickness (CCT) is consistently shown to be reduced in

POAG cases compared to controls; there is a 1.7 times increased risk per 40 μm decrease in CCT (Gordon, Beiser et al. 2002; Racette, Wilson et al. 2003; Leske, Heijl et al. 2007; Miglior, Pfeiffer et al. 2007; Coleman and Miglior 2008). While the Ocular Hypertension Treatment study showed CCT is independently associated with POAG, others question its true relationship with POAG due to confounding factors. CCT is known to influence intraocular pressure measurements, and the effect of prescribed topical medications for POAG treatment on CCT remains unclear (Janssen, Gorgels et al. 2013). Nevertheless, the strong association observed between CCT and POAG has warranted many research studies, discussed in greater detail in Chapter 2. Similarly, vertical cup-disc ratio (VCDR) has been associated with POAG and is considered a predictive risk factor due to the fact that VCDR is used to measure POAG disease progression (Gordon, Beiser et al. 2002; Coleman and Miglior 2008). A genetic locus at Chromosome 14q22-23 is associated with both VCDR and POAG (Macgregor, Hewitt et al. 2010; Ramdas, van Koolwijk et al. 2010; Fan, Wang et al. 2011; Osman, Low et al. 2012; Wiggs, Yaspan et al. 2012). A functional analysis of this locus is presented in Chapter 3.

African ancestry has been consistently shown to increase the risk of developing POAG (Tielsch, Sommer et al. 1991; Gordon, Beiser et al. 2002; Racette, Wilson et al. 2003; Quigley and Broman 2006; Rudnicka, Mt-Isa et al. 2006). The Ocular Hypertension Treatment study found that African Americans have a 59% increased risk of developing

POAG (Gordon, Beiser et al. 2002). However, there are racial differences in many ocular parameters and, on average, individuals of African ancestry have thinner corneas, larger optic cups, larger cup-disc ratios, and thinner retinal nerve fiber layers—all parameters associated with POAG risk and progression (Racette, Wilson et al. 2003). The Ocular Hypertension Treatment study showed that race is no longer significantly associated with POAG in multivariate models adjusted for corneal thickness or cup-disc ratio, suggesting African ancestry may increase risk of POAG via inherent, quantitative ocular parameters (Gordon et al).

1.4 POAG genetics

POAG is a heritable, complex disease; both sporadic and familial forms exist. Previous studies have reported that approximately 50% of POAG cases have a family history of the disease, and the risk of developing POAG is 3-9 times greater in individuals with an affected first degree relative ((Tielsch, Katz et al. 1994; Wolfs, Klaver et al. 1998; Weih, Nanjan et al. 2001; Wu, Hewitt et al. 2006). Congenital and juvenile forms of POAG are often inherited in either an autosomal dominant or autosomal recessive manner, while adult-onset POAG is considered a complex trait with multigenic inheritance (Liu and Allingham 2011; Janssen, Gorgels et al. 2013). Generally, familial forms of POAG are more severe and have an earlier average age of onset (Wu,

Hewitt et al. 2006). Taken together, these data suggest POAG likely has a genetic component.

The search for glaucoma genes began with linkage studies, an approach that relies on the collection of nuclear families or extended families with multiple affected individuals. These family-based studies identified 20 loci linked to POAG, and within these loci several likely disease causing genes have been identified including *MYOC*, *OPTN*, *WDR36*, and *CY1B1* (for primary congenital glaucoma). Detailed reviews of these loci are available (Allingham, Liu et al. 2009; Liu and Allingham 2011; Gemenetzi, Yang et al. 2012; Janssen, Gorgels et al. 2013). For the most part, mutations in these genes are limited to early-onset POAG cases and only explain 3-5% of adult-onset cases. The genetics of the majority of POAG cases remain poorly understood and are likely due to multigenic and environmental effects, consistent with a complex etiology (Allingham, Liu et al. 2009; Liu and Allingham 2011; Janssen, Gorgels et al. 2013).

As an alternative to linkage analysis, genome-wide association studies (GWAS) have also been used for POAG disease gene identification. This approach, like linkage mapping, is unbiased and does not rely on an *a priori* hypothesis, but unlike linkage mapping, is applicable for case-control datasets. Many successful POAG GWAS have been conducted, and these studies have been performed in POAG cases with normal and high ocular pressure from various populations (Nakano, Ikeda et al. 2009; Meguro,

Inoko et al. 2010; Thorleifsson, Walters et al. 2010; Burdon, Macgregor et al. 2011; Burdon 2012; Gibson, Griffiths et al. 2012; Osman, Low et al. 2012; Wiggs, Yaspan et al. 2012; Janssen, Gorgels et al. 2013). The most commonly identified and replicated loci include *CAV1*/*CAV2* (Chr. 7q31.1), *CDKN2B* (Chr. 9p21), *SIX1*/*SIX6* (Chr. 14p22-23), and *TMCO1* (Chr. 1q24) (Burdon 2012; Janssen, Gorgels et al. 2013). In addition, other commonly used disease gene identification approaches have been used in POAG studies including a candidate gene approach (where genes are selected for analysis based on predefined, biological driven hypotheses) and endophenotype analysis (where the genetics of potential endophenotypes are used to identify disease genes). For an extensive review of genes associated with POAG and related endophenotypes, please see the recently published review by Janseen et al. (Janssen, Gorgels et al. 2013). Despite the success in the field of POAG genetics, the functional role of genes identified using these methods remains unclear. This fact highlights the current need for focused functional follow-up of POAG genetic association studies.

Complex diseases such as POAG exhibit substantial etiologic and genetic heterogeneity. Therefore, identifying POAG susceptibility alleles requires the use of a variety of approaches and techniques including family-based linkage analysis, case/control association analysis, and bioinformatics approaches. Three methods are presented here: (Chapter 2) the genetic assessment of a quantitative POAG risk factor;

(Chapter 3) the functional analysis of a POAG candidate gene; and (Chapter 4) the transcriptome analysis of ocular tissues- all with the shared goal of increasing our understanding of the biology and genetic mechanisms responsible for the development and progression of POAG.

2. Central corneal thickness (CCT) analysis: a potential POAG endophenotype[1]

2.1 Introduction

The cornea is responsible for the majority of the eye's refractive power, and its structure is therefore an important part of normal vision. The cornea is a highly collagenous, transparent tissue that is a continuation of the sclera, which posteriorly supports the optic nerve. A reduction in central corneal thickness (CCT) has been observed in a number of systemic and ocular diseases including Ehlers-Danlos syndrome, osteogenesis imperfecta, keratoconus, and brittle cornea syndrome (Dimasi, Chen et al. 2010; Lu, Dimasi et al. 2010; Vitart, Bencic et al. 2010; Vithana, Aung et al. 2011). Reduced CCT has also been associated with an increased risk of primary open-angle glaucoma (POAG), the second leading cause of blindness worldwide (Gordon, Beiser et al. 2002; Dimasi, Burdon et al. 2010). In POAG, CCT has been shown to be a risk factor independent of other known risk factors including advanced age and elevated intraocular pressure, and has been linked with increased disease severity and rapid disease progression (Gordon, Beiser et al. 2002; Leske, Heijl et al. 2007; Miglior, Pfeiffer et al. 2007; Dimasi, Burdon et al. 2010). Elucidating the genetics of CCT may lead to

[1] This chapter is a modification of an earlier publication which appeared as Ulmer, M., J. Li, et al. (2012). "Genome-wide analysis of central corneal thickness in primary open-angle glaucoma cases in the NEIGHBOR and GLAUGEN consortia." Invest Ophthalmol Vis Sci **53**(8): 4468-4474. The copyright of the original article is held by the Association for Research in Vision and Ophthalmology.

identification of novel POAG risk genes and could improve our understanding of the molecular mechanism driving this association.

CCT has been shown to be a normally distributed quantitative trait with heritability estimates as high as 95% (Dimasi, Burdon et al. 2010). Several studies have been successful in identifying CCT loci through candidate gene and genome-wide approaches. An Australian and UK cohort study first identified the *ZNF469* gene, an uncharacterized zinc-finger protein, and *FOXO1*, a member of the forkhead family of transcription factors, in a genome-wide analysis of CCT (Lu, Dimasi et al. 2010). The *ZNF469* locus has been replicated in follow-up studies in Croatian, Scottish, and Asian populations (Vitart, Bencic et al. 2010; Cornes, Khor et al. 2011; Vithana, Aung et al. 2011). Consistent with the belief that CCT is a complex quantitative trait with multigenic inheritance, these studies have also identified numerous other loci in genomic regions containing collagen V alpha-1 (*COL5A1*), collagen VIII alpha-2 (*COL8A2*), A-kinase anchor protein 13 (*AKAP13*), inhibitor of Bruton's tyrosine kinase (*IBTK*), leucine-rich repeat kinase 1 (*LRRK1*), and multiple intergenic regions (Vitart, Bencic et al. 2010; Cornes, Khor et al. 2011; Vithana, Aung et al. 2011).

Due to the strong association observed between CCT and POAG, CCT has been considered as a possible POAG endophenotype (Gordon, Beiser et al. 2002; Leske, Heijl et al. 2007; Gemenetzi, Yang et al. 2012). Identification of complex disease genes through

a quantitative endophenotype has many advantages and has been successful in POAG with other quantitative risk factors (Charlesworth, Kramer et al. 2010; Desronvil, Logan-Wyatt et al. 2010; Ramdas, van Koolwijk et al. 2010; Ramdas, van Koolwijk et al. 2011). Disease genes can be identified by first studying the genetics of the endophenotype based on the assumption that candidate genes will be associated with both traits. Using this methodology, we investigate the effects of previously reported CCT-associated genetic variants on POAG risk. Although these genome-wide studies have successfully identified genetic variants associated with CCT, they have been performed in primarily non-glaucomatous population cohorts. It is possible that POAG cases have additional variants associated with CCT compared to normal populations and that these variants may explain the underlying molecular mechanism driving the association between reduced CCT and POAG risk. Therefore, we then performed a genome-wide CCT analysis in a primarily POAG case population to identify a novel list of candidate variants and tested the effects of these single nucleotide polymorphisms (SNPs) using a larger POAG case-control population.

2.2 Materials and methods

2.2.1 Study population

Data were available from the NEI Glaucoma Human Genetics Collaboration (NEIGHBOR) and the Glaucoma Genes and Environment (GLAUGEN) consortia

(Wiggs, Hauser et al. 2012; Wiggs, Yaspan et al. 2012). These studies were approved by the Institutional Review Boards of the Massachusetts Eye and Ear Infirmary, the Harvard University School of Public Health, the Brigham and Woman's Hospital, and Duke University. All human subject research adhered to the tenets of the Declaration of Helsinki. POAG cases for both studies were defined as individuals with glaucomatous optic neuropathy with reproducible visual field tests or a cup-disc ratio of ≥ 0.7 in at least one eye. POAG cases were grouped into two subsets based on intraocular pressure (IOP) for stratified analyses. A low tension subset was defined by an IOP of ≤ 21 mm Hg at the time of sample collection with no history of IOP greater than 21 mm Hg (either by self-report or by chart review). A total of 705 POAG cases met the low tension definition while 1,629 POAG cases with a history of any IOP > 21 mm Hg were included in the high tension subset. POAG controls had normal IOP (≤ 21 mm Hg), no evidence of visual field loss, and a normal cup-disc ratio (< 0.7).

2.2.2 CCT association analysis

CCT measurements were available for 851 individuals from the NEIGHBOR study and 325 individuals from the GLAUGEN study. For the CCT analysis, we used the mean CCT from both eyes as our analysis outcome variable because we observed a strong correlation between right eye CCT and left eye CCT ($R^2 = 0.92$ in GLAUGEN; $R^2 = 0.83$ in NEIGHBOR). We removed 10 outliers in GLAUGEN and 16 outliers in

NEIGHBOR, which were defined by large left-right eye CCT differences (beyond 2.5 standard deviation of the overall distribution of left-right differences). A lack of left-right eye correlation could be due to true right and left eye CCT differences, measurement artifacts, or unknown corneal defects. Additionally, we removed individuals with known corneal diseases, non-Caucasian ancestry, and one distribution outlier with an extremely small average CCT value (< 374 μm). Population substructure was assessed by principal components analysis performed using Eigenstrat (Price, Patterson et al. 2006) and outliers were removed from the analysis (Appendix A). After data cleaning, our CCT dataset consisted of 1,117 individuals including 668 POAG cases and 144 POAG controls from the NEIGHBOR study and 305 POAG cases from the GLAUGEN study (Table 1).

We assessed CCT distribution differences among a variety of clinical variables including POAG case status, CCT measurement device, DNA source (blood versus cheek), DNA extraction method (Qiagen, DNAzol, GENTRA), history of laser trabeculoplasty, and incisional glaucoma filtration surgery using the Student's t-test calculated in SAS (SAS Institute Inc. 2002-2008). Only POAG case status and CCT measurement device were associated with CCT and were included as covariates in the final model. Genotyping was performed using the Illumina Human660W-Quad v1 BeadChip array and was previously filtered using quality control metrics, as described

(Wiggs, Yaspan et al. 2012). The final dataset contained 480,304 cleaned SNPs shared between the two studies. SNPs were tested for association with average CCT using linear regression performed in PLINK (Purcell, Neale et al. 2007), adjusted for age, sex, instrument used to measure CCT, and POAG case/control status. A Manhattan plot and a quantile-quantile plot (Appendix B) of the resulting p-values were generated in R (http://www.R-project.org) using code provided by Getting Genetics Done (http://gettinggeneticsdone.com/). Linkage disequilibrium was calculated and visualized using the Haploview program (Barrett 2009).

2.2.3 POAG association analysis

Candidate SNPs were tested for association with POAG risk in the full NEIGHBOR and GLAUGEN datasets. After quality control filtering including removal of sex mismatches, duplicate samples, and population outliers, the final dataset contained 2125 POAG cases and 2251 POAG controls from the NEIGHBOR study and 964 POAG cases and 1130 POAG controls from the GLAUGEN study[2]. In the NEIGHBOR study, logistic regression models included age, sex, collection site, and 4 principal components as covariates. In the GLAUGEN study, the method of DNA extraction, the specimen type (blood vs. cheek), and site were found to be independently

[2] Sample collection, genotyping, and initial quality control filtering (removal of failed SNPs) was conducted by members of the NEIGHBOR and GLAUGEN consortia, as described in Wiggs, J. L., B. L. Yaspan, et al. (2012). "Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma." PLoS Genet **8**(4): e1002654.

associated with both DNA quality and POAG case/control status and were included in logistic regression models along with age, sex, and 6 principal components. A formal meta-analysis of the two studies was conducted in PLINK (Purcell, Neale et al. 2007). POAG risk was assessed in the overall dataset, the low tension subset, and the high tension subset.

2.2.4 Gene expression analysis

To determine the presence of candidate gene expression in relevant tissues, cDNA was prepared from human adult and fetal brain and ocular tissue samples[3]. Adult and fetal brain cDNA libraries were generated from the Human Total RNA Master Panel II from Clontech Laboratories, Inc. (California, USA). A fetal ocular sample from a 24 week gestation period was obtained from Advanced Biosciences Resources (California, USA) and was preserved in RNA*later* (Qiagen, Limburg, Netherlands) within minutes of collection. An adult eye was provided by the North Carolina Eye Bank. This sample was collected 5 hours and 10 minutes post-mortem and was preserved in RNA*later*. We dissected the cornea, optic nerve, and sclera wall from each eye. The cornea was isolated using a 5 mm biopsy punch at the center of the cornea, optic nerve samples were collected using dissection scissors, and the scleral tissue was isolated using a 7 mm

[3] Ocular tissue dissections and cDNA synthesis was performed by Dr. Felicia Hawthorne at the Duke Center for Human Genetics. The subsequent gene expression assay and analysis was performed by Megan Ulmer Carnes.

biopsy punch centered behind the fovea. To reduce retinal contamination of the scleral sample, a second 5 mm biopsy was taken from the center of the 7 mm punch after complete removal of the retina. Messenger RNA (mRNA) was extracted from each tissue sample independently using the *mir*Vana total RNA extraction kit (Ambion, Texas, USA). Quality control for each sample included measuring RNA concentration and RNA subunit ratios at 260/280 nm using the NanoDrop technology (Invitrogen, California, USA). We generated cDNA from each mRNA library using the SuperScript III First-Strand Synthesis kit (Invitrogen).

We then used the brain, cornea, optic nerve, and sclera cDNA libraries to perform polymerase chain reaction (PCR) assays to determine the presence or absence of contactin-associated protein-like 4 (*CNTNAP4)* and neurotrmin *(NTM)* gene expression. The PCR assays contained 2 uL cDNA, 1 X PCR buffer, 200 μmol/ each dNTP, 1.5 mM $MgCl_2$, 200 ng of the forward primer, 200 ng of the reverse primer, and 3 U of Platinum Taq DNA polymerase (Invitrogen). Primers were designed using Primer3 (Rozen and Skaletsky 2000) and spanned intron-exon boundaries. The glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene was used as a control to check for cDNA quality. *CNTNAP4*, *NTM*, and *GAPDH* were amplified (primer sequences are available in Appendix C) using a touchdown PCR protocol with the following conditions: 94°C for 3 minutes; 2 cycles of 94°C for 5 seconds, 61°C for 30 seconds, and 72°C for 30 seconds

repeated using an incremental 2°C decrease in annealing temperature until 51°C was reached; followed by 30 cycles with a 51°C annealing temperature; and a final extension at 72°C for 3 minutes. The PCRs were purified using Invitrogen's PureLink Quick Gel Extraction Kit, and the products were verified by sequencing at Eton Bioscience, Inc. (North Carolina, USA). The results were visualized on an ethidium bromide stained, 1.5% agarose gel.

2.3 Results

2.3.1 Dataset characteristics

We compared the CCT between POAG cases and controls. The two groups were significantly different ($p = 0.0002$), with a mean (standard deviation) of 557.7 (33.80) μm in POAG controls and 545.9 (35.85) μm in POAG cases (Table 1). We also observed a significant difference ($p = 0.008$) in the distribution of CCT in the low tension POAG cases compared to high tension cases (Table 1). The high tension subset had a mean CCT of 548.4 (35.87) μm, while the low tension POAG cases had a mean CCT of 541.8 (36.10) μm. We found that the instrument used to measure CCT (DGH Technology, Inc. ultrasound table unit versus a portable handheld unit) had a significant effect ($p < 0.0001$, Table 1). Within our CCT dataset, history of laser trabeculoplasty and glaucoma filtration surgery was available for 422 POAG cases. We did not see a difference in CCT

when we compared individuals with or without a history of laser trabeculoplasty (p = 0.86) or glaucoma filtration surgery (p = 0.62).

Table 1: Population characteristics of the CCT dataset

CCT Dataset	N	Mean (SD)
Total	1117	
Age (years)	1117	61.6 (12.58)
Sex (% male)	1117	48.9%
CCT (μm)	1117	
*POAG Control	144	557.7 (33.80)
*POAG Case	973	545.9 (35.85)
*Standard pachymeter	1010	545.6 (35.33)
*Portable pachymeter	107	565.0 (35.60)
*POAG cases, low tension	307	541.8 (36.10)
*POAG cases, high tension	625	548.4 (35.87)

2.3.2 Population-based CCT associated variants and their effect on POAG risk

We looked for evidence of association for 20 previously published CCT SNPs (Table 2) in the NEIGHBOR/GLAUGEN CCT dataset (N = 1,117). Using a Bonferroni multiple testing correction significance threshold of 0 .0025 (0.05/20), one SNP, rs12447690 upstream of *ZNF469* (β = -5.08 μm/minor allele, p = 0.001) showed significant evidence of association with CCT. Three additional SNPs showed suggestive evidence of

association: rs7044529 in *COL5A1* (β = -6.58 μm/allele, p = 0.003), rs1536478 in the *RXRA*/*COL5A1* intergenic region (β = 4.47 μm/allele, p = 0.004), and rs1538138 near *IBTK* (β = -4.98 μm/allele, p = 0.004) (Table 2). The direction and the effect size estimated for these four variants are consistent with previous reports (Lu, Dimasi et al. 2010; Vitart, Bencic et al. 2010; Cornes, Khor et al. 2011; Vithana, Aung et al. 2011). We assessed the association of these 20 SNPs with POAG case/control status in the larger NEIGHBOR/GLAUGEN meta-analysis in the overall dataset (N = 6,470), low tension (N = 4,086), and high tension (N = 4,966) subsets. Using the same significance threshold as above (0.0025), none of these SNPs showed any evidence of association with POAG (Table 2).

Table 2: Association results from the CCT and POAG analyses for 20 previously reported CCT-associated variants

The association results of 20 SNPs previously found to be associated with CCT are shown. The beta (μm) and odds ratio (OR) represent the effect size for each copy of the minor allele.

			CCT Analysis (N = 1,117)		POAG Analysis (N = 6,470)	
SNP	**Chr.**	**Locus**	**Beta**	**P**	**OR**	**P**
rs3767703	1	Col8A2	-2.37	0.37	0.95	0.51
rs7550047	1	Col8A2	-2.78	0.27	0.98	0.78
rs96067	1	Col8A2	-0.83	0.66	0.94	0.20
rs11694554	2	Col4A3	3.82	0.10	1.14	0.25
rs1538138	6	IBTK	-4.98	0.004	0.96	0.42
rs1324183	9	9p23	-3.66	0.04	1.09	0.09
rs1409832	9	RXRA/COL5A1	-3.31	0.06	1.09	0.07
rs1536478	9	RXRA/COL5A1	4.47	0.004	0.99	0.87
rs1536482	9	RXRA/COL5A1	-1.69	0.29	1.04	0.38
rs7044529	9	COL5A1	-6.58	0.003	0.99	0.86
rs1034200	13	AVGR8	2.9	0.08	0.94	0.31
rs2721051	13	FOXO1	-4.69	0.04	1.02	0.75
rs2755237	13	FOXO1	-1.66	0.4	1.06	0.28
rs1828481	15	AKAP13	0.11	0.94	0.96	0.29
rs4965359	15	LRRK1	-2.82	0.07	0.99	0.81
rs6496932	15	PDE8A/AKAP13	-2.35	0.22	1.05	0.30
rs7172789	15	AKAP13	0.23	0.88	0.95	0.25
rs930847	15	LRRK1	2.15	0.23	0.97	0.47
rs12447690	16	ZNF469	-5.08	0.001	1.04	0.33
rs9938149	16	ZNF469	-3.84	0.01	1.05	0.24

2.3.3 CCT genome-wide association study

Next, we performed a CCT GWAS to identify novel candidate SNPs using the CCT dataset. No SNP met the genome-wide significance threshold of 1 X 10^{-7} (Figure 1), either in the overall dataset or when the analysis was confined to only POAG cases.

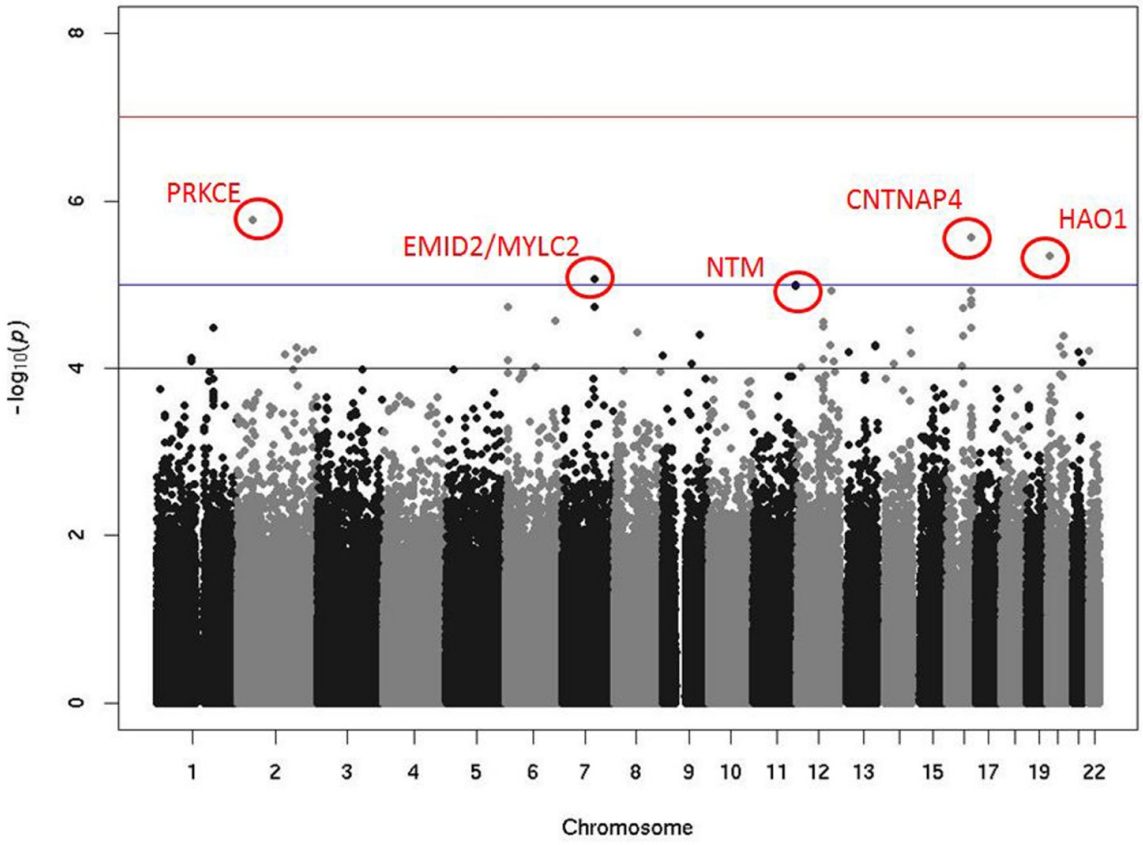


Figure 1: Manhattan plot showing the results of the CCT GWAS from the NEIGHBOR/GLAUGEN CCT dataset

The red line (upper) indicates the genome-wide significance threshold (p-value of 1 X 10-7), the blue line (middle) indicates the suggestive significance threshold (p-value of 1 X 10^{-5}), and the black line (lower) indicates the threshold level selected for a follow-up POAG association analysis (p-value of 1 X 10^{-4}).

The top five most significant SNPs (Table 3, Figure 1) were located on chromosome (Chr.) 2 at *PRKCE* (rs10205024, β = -7.49 μm/allele, p = 1.68 X 10^{-6}); on Chr. 16 at the *CNTNAP4* gene (rs9939043, β = 7.15 μm/allele, p = 2.70 X 10^{-6}); on Chr. 20 at the *HAO1* gene (rs2423322, β = -9.03 μm/allele, p = 4.48 X 10^{-6}); in an intergenic region on Chr. 7 at the *EMID2/MYL10* locus (rs17135662, β = 8.31 μm/allele, p = 8.69 X 10^{-6}); and on Chr. 11 at the *NTM* gene (rs7108536, β = -6.78 μm/allele, p = 1.00 X 10^{-5}).

Table 3: Top ten most significant SNPs from the CCT GWAS

The top ten SNPs from the CCT GWAS are shown. The beta (μm) represents the effect size for each copy of the minor allele, given in the allele column.

Chr.	SNP	Locus	Allele	Beta	p
2	rs10205024	PRKCE	A	-7.49	1.68 X 10^{-06}
16	rs9939043	CNTNAP4	A	7.15	2.70 X 10^{-06}
20	rs2423322	HAO1	A	-9.03	4.48 X 10^{-06}
7	rs17135662	EMID2/MYL10	A	8.31	8.69 X 10^{-06}
11	rs7108536	NTM	G	-6.78	1.00 X 10^{-05}
11	rs7481514	NTM	A	-6.89	1.03 X 10^{-05}
16	rs2866710	CNTNAP4	G	6.84	1.17 X 10^{-05}
12	rs9669708	LOC253724	A	9.84	1.19 X 10^{-05}
16	rs1428759	CNTNAP4	A	6.78	1.53 X 10^{-05}
16	rs1428758	CNTNAP4	G	6.61	1.74 X 10^{-05}

2.3.4 Novel CCT candidate variants and their effect on POAG risk

We selected 50 SNPs most strongly associated with CCT from the GWAS results based on a p-value cut off of 1×10^{-4} and tested these for association with POAG risk using the overall NEIGHBOR/GLAUGEN meta-analysis and meta-analyses stratified by intraocular pressure . Using a significance threshold of $p \leq 0.001$ (Bonferroni correction of 0.05/50), one SNP at the *NTM* gene locus (rs7481514, odds ratio (OR) = 1.28, p = 0.001) was significantly associated with POAG in the low tension subset (Table 4). As expected, each copy of the minor allele is associated with reduced CCT and an increased risk of POAG. A second SNP at the *NTM* locus was marginally associated in the low tension subset (rs7108536, OR = 1.25, p = 0.002). The two *NTM* SNPs are in high linkage disequilibrium ($r^2 = 0.87$). An additional five SNPs at the *CNTNAP4* gene showed suggestive association, although they did not meet the multiple testing corrected significance level (Table 4).

Table 4: Top ten SNPs associated with POAG in the low tension subset

The top ten SNPs most associated with POAG out of the 50 candidate SNPs selected from the CCT GWAS are shown. The beta (μm) and odds ratio (OR) represent the effect size for each copy of the minor allele, given in the allele column.

				CCT association (N = 1,117)		POAG association (N = 4,086)	
SNP	**Locus**	**Chr.**	**Allele**	**Beta**	**p**	**OR**	**p**
rs7481514	NTM	11	A	-6.89	1.03×10^{-5}	1.28	0.001
rs7108536	NTM	11	G	-6.78	1.00×10^{-5}	1.25	0.002
rs1428758	CNTNAP4	16	G	6.61	1.74×10^{-5}	0.84	0.018
rs9939043	CNTNAP4	16	A	7.15	2.70×10^{-6}	0.85	0.022
rs6701037	TNN	1	C	-6.11	3.31×10^{-5}	0.85	0.024
rs4959388	LY86-AS1	6	G	7.07	1.85×10^{-5}	1.18	0.031
rs1109739	16q12	16	A	-8.48	4.13×10^{-5}	0.81	0.038
rs2866710	CNTNAP4	16	G	6.84	1.17×10^{-5}	0.86	0.051
rs2052866	CNTNAP4	16	A	6.51	3.27×10^{-5}	0.86	0.054
rs1428759	CNTNAP4	16	A	6.78	1.53×10^{-5}	0.87	0.072

The most significant *CNTNAP4* SNP was rs1428758 (OR = 0.84, p = 0.018). The strong linkage disequilibrium in this region (Figure 2) suggests these signals are probably not independent. The effects of these five SNPs are also in the expected direction; each copy of the minor allele is associated with an increase in CCT and a protective effect on POAG risk. No SNP was associated with POAG in the overall dataset or the high tension subset.

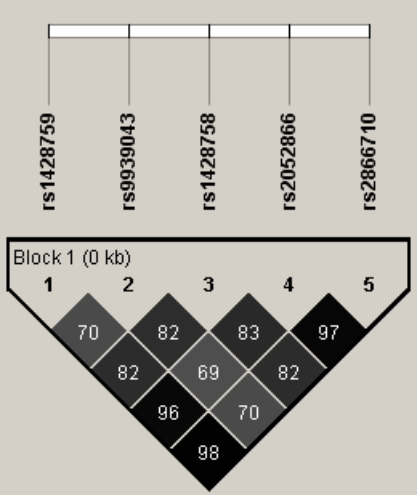


Figure 2: Linkage disequilibrium of the five *CNTNAP4* SNPs identified in the low tension subset

Linkage disequilibrium is given in r^2 calculated in Haploview.

2.3.5 Gene expression analysis of candidate genes

We evaluated *CNTNAP4* and *NTM* gene expression in human adult and fetal brain, optic nerve, cornea and sclera tissues (Figure 3). *CNTNAP4* was expressed in adult brain, optic nerve, and cornea, but we found no evidence of *CNTNAP4* expression in the adult sclera sample. Similarly, we found evidence of *CNTNAP4* expression in the fetal brain, optic nerve, and cornea tissues, but not in the fetal sclera sample. We showed *NTM* expression in all of the tissues examined in both the adult and fetal stage tissues (Figure 3).

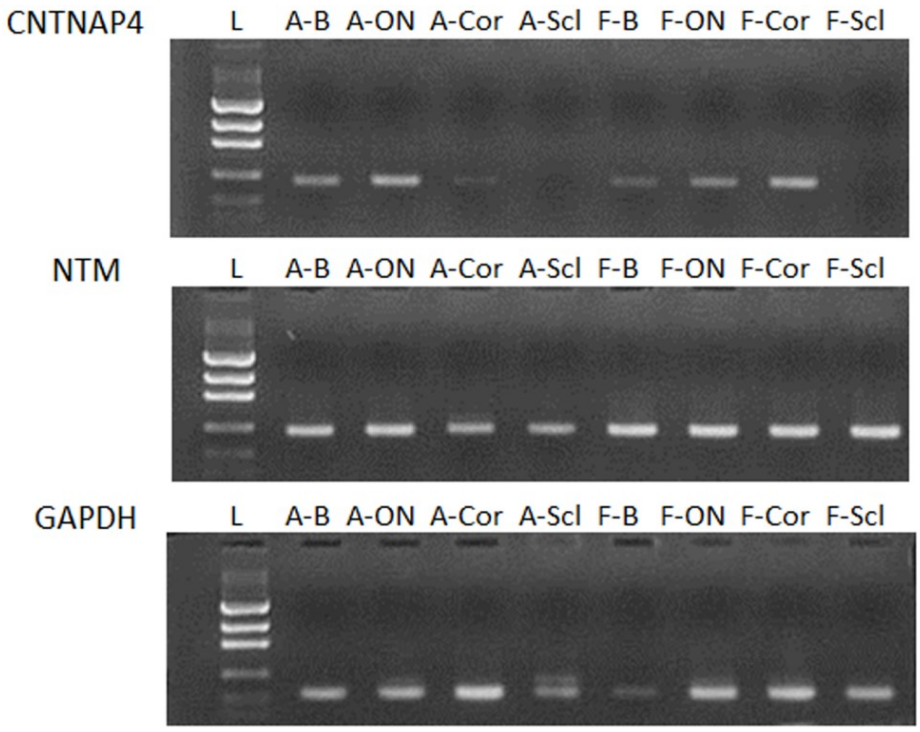


Figure 3: Evidence of *CNTNAP4* and *NTM* gene expression in adult and fetal ocular tissue samples

The figure shows the resulting agarose gel image from the gene expression assays of *CNTNAP4* (top), *NTM* (middle), and *GAPDH* (lower) expression in brain and ocular tissue samples. The lanes from left to right contain size standard ladder (L; band sizes 100, 200, 400, 800, 1200, and 2000 base pairs), adult brain (A-B), adult optic nerve (A-ON), adult cornea (A-Cor), adult sclera (A-Scl), fetal brain (F-B), fetal optic nerve (F-ON), fetal cornea (F-Cor), and fetal sclera (F-Scl).

2.4 Discussion

CCT is an important quantitative risk factor for POAG, but whether CCT directly influences POAG susceptibility remains unknown. One way to improve our understanding of this relationship is to identify genes that are associated with both traits. In this study, we replicated the association of previously published CCT genetic variants in a POAG population and showed that despite the well-established genetic association observed between these loci and CCT, there was no evidence that they influence POAG risk. We have also identified two POAG candidate genes, *NTM* and *CNTNAP4*, first by association with CCT and second by association analysis with POAG risk. These genes are good biological candidates for future research in the etiology of POAG.

In our dataset, POAG cases have significantly lower mean CCT compared to controls and within the POAG cases, the low tension subset have significantly lower mean CCT compared to the high tension subset. Consistent with our results, several studies have suggested CCT is reduced in normal tension glaucoma as compared with POAG (Kniestedt, Lin et al. 2006; Sullivan-Mee, Halverson et al. 2006; Kaushik, Pandav et al. 2012). One possible explanation for this observation is that thinner corneas may be representative of a thinner, weaker posterior eye structure, which may increase

susceptibility to optic nerve stress caused by elevated or fluctuating intraocular pressure.

In this study, we investigated the effects of previously published CCT SNPs and replicated the association of rs12447690 near the *ZNF469* gene, an uncharacterized zinc-finger protein (Lu, Dimasi et al. 2010; Vitart, Bencic et al. 2010; Vithana, Aung et al. 2011). Mutations in this gene are known to cause brittle cornea syndrome, a systemic connective tissue disorder characterized by a high risk of corneal tearing (Abu, Frydman et al. 2008). The association of this SNP with CCT has been replicated in multiple populations, and there is strong evidence that this gene is involved with normal CCT variation (Lu, Dimasi et al. 2010; Vitart, Bencic et al. 2010; Vithana, Aung et al. 2011). SNPs within two genes, *COL5A1* and *IBTK,* were previously identified in multiple population-based CCT studies (Vitart, Bencic et al. 2010; Cornes, Khor et al. 2011; Vithana, Aung et al. 2011). We also found evidence that these SNPs are associated with CCT in our primarily POAG case population. *COL5A1* is a logical candidate gene for cornea thickness variation as the corneal stroma is largely composed of regularly arranged collagen fibers. Mutations in this gene have been identified in Ehlers-Danlos syndrome patients who have reduced cornea thickness (Segev, Heon et al. 2006). *IBTK* was recently identified in a CCT meta-analysis conducted in a multi-ethnic Asian population cohort study (Cornes, Khor et al. 2011). This group hypothesized that *IBTK*

may regulate corneal thickness during developmental stages and embryogenesis through its negative regulation of BTK kinase activity or its effect on nuclear factor-kappa-B (NF-κB) activation.

We were unable to replicate association of the remaining 16 published CCT SNPs with CCT. This may be due to the limited sample size of our CCT dataset (N = 1,117). We had an estimated 66% power, assuming a minor allele frequency of 0.25 and an effect of -6.00 Δ CCT (μm/allele) (Gauderman WJ 2006), to detect association with CCT. Thus, based on this estimate, we cannot rule out false negative findings for those 16 loci. However, we also examined the potential association of all 20 CCT-associated loci with POAG in our dataset and found no evidence that any of these well-established CCT genes results in increased POAG susceptibility. We would expect that the overall POAG dataset has greater power to detect genotypic effects because of the increased sample size (N = 6,470), although the true power of this dataset is unknown because the putative effect sizes of these loci on POAG risk remain unknown.

In our CCT GWAS, no SNPS met genome-wide significance thresholds. As with the replication analysis, this may be due to a lack of statistical power resulting from sample size limitations or due to the fact that our study population consists mainly of POAG cases which may have different genetic determinants of CCT than population controls.

To determine if any of the top SNPs identified in our CCT analysis also influence POAG risk, we tested selected candidates SNPs for association with POAG in the overall dataset and two subsets, low tension and high tension, stratified by a history of elevated intraocular pressure. Stratification by intraocular pressure may provide a more clinically homogenous population resulting in reduced genetic heterogeneity within the dataset. In the low tension subset of POAG cases, neurotrimin (*NTM*) was significantly associated with increased POAG risk and, as expected, also a decrease in CCT. The size of these effects is consistent with published reports suggesting a decrease of 40 μm of CCT is associated with a 1.7-fold increased risk of POAG (Gordon, Beiser et al. 2002). NTM is a member of the immunoglobulin domain-containing IgLON superfamily of glycosyl-phosphatidylinositol (GPI)-anchored cell adhesion molecules. NTM can be found in both membrane-bound and soluble forms and has known roles in mediating cell-cell interactions and regulating neurite outgrowth (Struyk, Canoll et al. 1995; Gil, Zanazzi et al. 1998; Lodge, McNamee et al. 2001). It has high homology with the well characterized neuronal cell adhesion molecule (*NCAM*), which has also been shown to have a role in neurite outgrowth. *NCAM* was found to be constitutively expressed in human corneal endothelium cells, although the role of neuronal cell adhesion molecules within the cornea remains unknown (Foets, van den Oord et al. 1992). Previous reports found *NTM* to be expressed during development, continuing into adulthood, and

abundant expression has been observed in the ganglion cell layer of the retina (Struyk, Canoll et al. 1995). It is possible that *NTM* plays a role in glaucoma susceptibility by influencing neuron outgrowth during development. Therefore, we examined the presence of *NTM* expression in relevant tissues from a developing fetus and found *NTM* is expressed in the brain, optic nerve, cornea, and sclera with similar expression observed in the adult sample. We included an adult brain sample as a positive control since it has been shown that *NTM* is expressed primarily in the central nervous system (Struyk, Canoll et al. 1995). This expression work provides further evidence that *NTM* is a viable candidate for future POAG research.

We have also identified contactin-associated protein-like 4 (*CNTNAP4*) as a candidate gene for glaucoma risk in the low tension POAG subset. Although the statistical evidence for association with this gene is modest, like *NTM*, it is a neural cell adhesion molecule and is important in cell-cell interaction within the nervous system (Spiegel, Salomon et al. 2002). *CNTNAP4* is located within a region linked to keratoconus, a disease defined by the progressive thinning of the cornea (Tyynismaa, Sistonen et al. 2002). In addition, SNPs within *CNTNAP2*, a gene closely related to *CNTNAP4*, has previously been associated with pseudoexfoliation glaucoma (Krumbiegel, Pasutto et al. 2011; Liu and Allingham 2011). We have shown that *CNTNAP4* is expressed in the brain, optic nerve, and cornea during development and in

adulthood. Together, *NTM* and *CNTNAP4* suggest that neural cell adhesion molecules could constitute a novel pathway relating reduced CCT to POAG risk.

As part of an international effort to better understand the genetics of CCT and its effect on POAG risk, a large CCT meta-analysis was performed on > 20,000 individuals and utilized the CCT genetic data generated as part of this study (Lu, Vitart et al. 2013). The meta-analysis (conducted using two non-glaucomatous population-based datasets with 13,057 European individuals and 6,963 Asian individuals) resulted in the identification of over 20 loci significantly associated with CCT at the genome-wide significance threshold. Similar to the approach used in this study, CCT variants identified in the international meta-analysis were tested for association with CCT within the NEIGHBOR/GLAUGEN datasets and then tested for association with POAG risk. The results of that analysis were consistent with the results presented here; CCT-associated variants identified by the international meta-analysis were, in general, associated with CCT in the POAG dataset, suggesting that variants that influence CCT are the same in diseased (POAG) and non-diseased populations. Likewise, these variants were not associated with POAG risk. Since the cutoffs for candidate SNP selection used in these studies are somewhat arbitrary, we applied a polygenic modeling approach which calculates a risk score given a range of p-value cutoffs, as previously described (Purcell, Wray et al. 2009). The goal of this approach is to determine if the

combined effects of SNPs identified in the CCT meta-analysis can predict POAG risk[4]. However, there was little evidence that the aggregate effects of CCT accurately predicted POAG risk. This calls into question the true POAG endophenotype status of CCT. In agreement with these findings, Charlesworth et al. observed a significant phenotypic correlation between CCT and POAG, but no genotypic correlation, and determined CCT is not likely a true POAG endophenotype (Charlesworth, Kramer et al. 2010).

However, this highlights the importance of disease specific datasets and proper stratification by relevant clinical variables, such as intraocular pressure. CCT and POAG may have a shared genetic relationship, as suggested by the potential association of genes identified in a CCT GWAS restricted to POAG cases. Like many other quantitative traits, it is likely to be mediated by many common loci of modest effect sizes. Rare variants within the CCT loci may result in profound structural changes in the cornea, but more common variants might produce the CCT variation that is observed on a population level. Our study suggests that these common variants are not associated with POAG. Nonetheless, CCT genomics research may offer promise to reveal new insight into the genetic architecture of a subset of POAG cases, as suggested by the potential

[4] The polygenic modeling analyses for the NEIGHBOR and GLAGUEN datasets was conducted by Megan Ulmer Carnes at the Duke Center for Human Genetics as described in the original manuscript, Lu, Y., V. Vitart, et al. (2013). "Genome-wide association analyses identify multiple loci associated with central corneal thickness and keratoconus." Nat Genet **45**(2): 155-163.

association of genes within neuronal cell adhesion pathways. However, it may be more immediately advantageous to use CCT in the clinical prediction of POAG development, since its association with POAG is well-established and it is an easily measureable trait.

3. Discovery and functional annotation of *SIX6* variants in POAG[1]

3.1 Introduction

The NEI Glaucoma Human Genetics Collaboration (NEIGHBOR) and the Glaucoma Genes and Environment (GLAUGEN) consortia identified a significant genetic association at the *SIX1*/*SIX6* locus (rs10483727, OR = 1.32, $p = 3.87X10^{-11}$) in a recently published POAG genome-wide association study (Wiggs, Yaspan et al. 2012). Variants in the *SIX1*/*SIX6* locus were first associated with quantitative optic nerve parameters in non-diseased populations including vertical cup-disc ratio (VCDR), which is used clinically to diagnose and monitor POAG progression (Macgregor, Hewitt et al. 2010; Ramdas, van Koolwijk et al. 2010). Several studies have independently confirmed the association of the *SIX1*/*SIX6* locus with both VCDR and POAG (Fan, Wang et al. 2011; Ramdas, van Koolwijk et al. 2011; Osman, Low et al. 2012).

The human *SIX* gene family consists of six members (*SIX1*-*SIX6*), all of which contain two shared protein domains, a DNA binding homeobox domain and a SIX domain, which binds downstream effector molecules (Kawakami, Sato et al. 2000; Kumar 2009). Members of this highly conserved gene family function as transcription factors, regulating key developmental steps through a complex regulatory network,

---

[1] A modified version of this chapter is published as M Ulmer Carnes, et al. (2014). "Discovery and functional annotation of SIX6 variants in primary open-angle glaucoma." PLOS Genetics. *In print.*

were originally identified through homology to the *Drosophila* sine oculis (*so*) gene, which is required for proper eye development (Kawakami, Sato et al. 2000; Kumar 2009). During embryonic development, *SIX1* is broadly expressed in multiple tissues including the otic vesicle, limb mesenchyme, and others. However, expression of *SIX6* is primarily restricted to regions of the retina and the pituitary (Kawakami, Sato et al. 2000; Conte, Marco-Ferreres et al. 2010). *Drosophila* with null *so* alleles have restricted retinal development, and Iglesias et al. recently demonstrated that morpholino knockdown of *six6b* in zebrafish embryos results in a small eye phenotype (Kawakami, Sato et al. 2000; Iglesias, Springelkamp et al. 2013). In humans, a large deletion on chromosome 14q22.3-q23 that includes *SIX6* causes bilateral anophthalmia, the absence of both eyes, demonstrating the importance of *SIX* gene family members in ocular development and human disease (Cheyette, Green et al. 1994; Gallardo, Lopez-Rios et al. 1999; Kawakami, Sato et al. 2000; Kenyon, Yang-Zhou et al. 2005; Kumar 2009).

In this study, we have extended the current understanding of the molecular contributions of *SIX6* to POAG risk in several ways. First, we identified potential POAG risk alleles by sequencing the *SIX6* gene in a case-control dataset. We found both common and rare coding changes within the *SIX6* gene in POAG cases, as well as sequence variants in the *SIX6* enhancer. We then used the zebrafish system to demonstrate that these human coding variants have functional consequences in eye

development. This analysis of 2-3 day old zebrafish embryos is not intended to fully recapitulate the glaucomatous phenotype; however, it provides valuable *in vivo* data about the functional effects of human genetic variation. We next used luciferase reporter assays to show that a sequence variant found in the *SIX6* enhancer of glaucoma patients may increase *SIX6* expression. We also demonstrate, for the first time, that POAG cases homozygous for the *SIX6* risk allele rs33912345 have a statistically significant thinner global retina nerve fiber layer.

3.2 Material and methods

3.2.1 Subjects

Study subjects were unrelated patients from the Duke Eye Center and, after a comprehensive eye examine, were classified as either POAG cases or controls[2]. POAG cases presented with glaucomatous optic neuropathy, defined as a cup-disc ratio greater than 0.7 and visual field loss in at least one eye. Patients with secondary forms of glaucoma or a history of ocular trauma were excluded from the study. POAG controls had no evidence of optic neuropathy, normal intraocular pressure (less than 22 mm Hg in both eyes), and normal visual fields, assessed using standard automated perimetry. This research was approved by the Institutional Review Board of Duke University Medical Center and adheres to the tenets of the Declaration of Helsinki.

[2] POAG patient diagnosis was performed at the Duke Eye Center by Dr. Rand Allingham.

3.2.2 DNA sequencing

Genomic DNA was extracted from patient blood samples using the PureGene chemistry following the manufacturer's standard protocol (Gentra, Minnesota, USA). The coding portions (2 exons) of the *SIX1* and *SIX6* genes were sequenced in 518 Caucasian POAG cases and controls (262 cases, 256 controls) using a polymerase chain-reaction (PCR) containing 1 X Qiagen PCR buffer (Tris·Cl, KCl, (NH4)2SO4, 15 mM $MgCl_2$; pH 8.7); 200 μM each of dATP, dCTP, dGTP, and dTTP; 0.4 μM forward PCR primer; 0.4 μM reverse PCR primer; 3 μL of betaine, 10 ng genomic DNA; and 0.5 U HotStarTaq DNA polymerase (Qiagen, Limburg, Netherlands) to a final volume of 25 μL[3]. Primer sequences are available in Appendix C. The PCR was performed using a touchdown protocol (incremental lowering of annealing temperature) using the following thermocycler conditions: 94˚C for 30 s, 65˚C for 30 s, 72˚C for 30 s with a 2˚C decrease in the annealing temperature every two cycles until a final annealing temperature of 55˚C was reached. The retinal specific *SIX6* enhancer, previously described (Conte, Marco-Ferreres et al. 2010), was amplified a using a touchdown protocol with a final annealing temperature of 57˚C. PCR products were sequencing

[3] DNA was extracted at the Duke DNA Bank. The coding portions of *SIX1* and *SIX6* were amplified by Dr. Benjamin Whigham and the *SIX6* enhancer was amplified by Megan Ulmer Carnes. Sequencing was performed by Beckman Coulter Genomics (Danvers, MA, USA).

using the BigDye chemistry on a 3730 DNA Analyzer (Life Technologies, New York, USA).

3.2.3 Genotyping and imputation

The common missense SNP, rs33912345, was genotyped in the Duke POAG case-control dataset consisting of 482 POAG cases and 433 POAG controls using a TaqMan allelic discrimination assay according to the standard protocols from the manufacturer[4] (Life Technologies). For quality control purposes, the following criteria were met: > 95% genotyping efficiency, matching sample duplicates (two Centre d'Etude du Polymorphisme Humain samples per 96-well plate whose genotype data matched across all plates), and Hardy-Weinberg equilibrium assumptions. We tested for association of rs33912345 with POAG using an additive logistic regression model adjusted for age and sex using SAS (SAS Institute Inc. 2002-2008).

Genome-wide genotype data were available from the NEIGHBOR and GLAUGEN consortia (Wiggs, Yaspan et al. 2012). Chromosome 14 was imputed[5] using IMPUTE2 (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html) with a global 1000 Genomes reference panel. We tested SNPs at the *SIX1*/*SIX6* locus for association with POAG using an additive logistic regression model adjusted for age, sex, and four

[4] Genotyping was performed by Dr. Benjamin Whigham at the Duke Center for Human Genetics.
[5] Imputation was completed by Melanie Garrett at the Duke Center for Human Genetics.

principal components (NEIGHBOR) or age, sex, study site, DNA extraction method, DNA specimen type and principal components 1-6 (GLAUGEN) implemented in PLINK and visualized using LocusZoom (http://csg.sph.umich.edu/locuszoom/) (Purcell, Neale et al. 2007; Pruim, Welch et al. 2010). A meta-analysis was performed in Plink using a random effects model. Linkage disequilibrium was calculated and visualized using Haploview (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview) (Barrett, Fry et al. 2005).

3.2.4 Retinal nerve fiber layer thickness analysis

Optical coherence tomography (OCT) measurements of retinal nerve fiber layer (RNFL) by Spectralis (Heidelberg Engineering, California, USA) spectral domain) and fundus photography were available from the Duke Eye Center[6]. OCT images are not routinely performed in patients without ocular disease, so there was limited data available for controls. Therefore, the analysis was restricted to POAG cases homozygous for rs33912345. Thirty patients had both OCT measurements and *SIX6* genotype data. RNFL thickness, age at POAG diagnosis, and the age at OCT measurement were compared between individuals homozygous for the risk or non-risk allele using a Student's t-test. Analyses were performed in SAS (SAS Institute Inc. 2002-2008).

[6] OCT images were taken by clinicians at the Duke Eye Center and were compiled by Dr. Shane Havens.

3.2.5 Microinjection of morpholino and mRNA

A vector containing human *SIX6* was purchased from the CCSB Human ORFeome Collection that uses the Gateway technology system (Open Biosystems and Life Technologies). *SIX6* alleles identified by sequencing (Glu93Gln, Glu129Lys, Asn141His, Leu205Arg, Thr212Met, and Ser242IIe) were created using the QuikChange II site-directed mutagenesis kit and protocols provided by the manufacturer (Agilent Technologies, California, USA). *SIX6* mRNA was *in vitro* transcribed using the mMESSAGE mMACHINE SP6 Kit (Life Technologies). Translation blocking (TB) morpholinos against *six6a* and *six6b* (Appendix C) were purchased from Gene Tools, LLC (Oregon, USA). Morpholino (2 ng) and mRNA (12.5 pg) were mixed and a volume of 0.5 nL was microinjected into each wild-type zebrafish embryo at one- to eight-cell stage, as previously described (Stuart, McMurray et al. 1988)[7].

3.2.6 *In vitro* luciferase assay

SIX6 enhancer alleles were tested using a dual-luciferase reporter assay system (Promega, Wisconsin, USA). An experimental construct containing a minimal promoter (pGL4.23, firefly luciferase, Promega) was used to test the functional effect of the enhancer alleles identified by sequencing (Chr14:60974363_C, Chr14:60974373_T, Chr14:60974378_T, Chr14:60974400_A, Chr14:60974449_G) in the POAG case/control

[7] The *in vivo* zebrafish assay (morpholino design, microinjection, and ocular measurements) were performed by Drs. Edwin Oh and Yangfan Liu at the Duke Center for Human Disease Modeling.

dataset. The experimental constructs (pGL4.23+Enhancer) were generated using a nested PCR protocol (primers given in Appendix C); the XhoI and HindII enzymes; the Quick Ligation kit (New England BioLabs, Massachusetts, USA); and the QuikChange II site-directed mutagenesis kit (Agilent Technologies), following protocols provided by the manufacturers. Constructs were confirmed to be correct by sequencing. Hek293 cells were cultured according to the supplier's suggestions (ATCC, Virginia, USA). As described by Conte et al., co-transfection with *NeuroD* and *E47* is required for *SIX6* enhancer activation (Conte, Marco-Ferreres et al. 2010). Therefore, cells were co-transfected with an experimental vector (pGL4.23+Enhancer), a control vector (pGL4.74, renilla luciferase, Promega), and vectors containing *NeuroD* and *E47* (provided by the Center for Human Disease Modeling, Duke University) using a standard calcium phosphate transfection protocol. The experiment was performed three times in triplicate, and the results were analyzed using the dual luciferase reporter (DLR) ratio (firefly luciferase sum: renilla luciferase sum) normalized by the reference *SIX6* enhancer included on every plate. The data were analyzed using an ANOVA, adjusted for batch, and linear contrasts were used to determine the effect of each vector. Statistical analyses were performed in SAS (SAS Institute Inc. 2002-2008).

3.3 Results

3.3.1 Identification of *SIX6* risk alleles in POAG cases

Sequencing of the *SIX1* and *SIX6* genes in Caucasian POAG cases and controls (262 cases, 256 controls) revealed the presence of 23 SNPs (Appendix D). Nine SNPs were identified in *SIX1*, but no nonsynonymous SNPs were identified in the POAG cases. Sequencing of *SIX6* revealed the presence of 14 variants including five rare nonsynonymous SNPs in POAG cases and controls, one common nonsynonymous SNP located in the homeobox of *SIX6* (rs33912345, Asn141His), and five sequence variants within the *SIX6* enhancer (Table 5).

Table 5: *SIX6* variants identified by sequencing POAG cases and controls

Locus	Coordinates	SNP ID	MAF POAG Cases	MAF POAG Controls	Base Change	AA Change
Enhancer	Chr14:60974363	----	0.002	0	G>C	----
Enhancer	Chr14:60974373	rs8004739	0.008	0	C>T	----
Enhancer	Chr14:60974378	----	0	0.002	C>T	----
Enhancer	Chr14:60974400	----	0.002	0.002	C>A	----
Enhancer	Chr14:60974449	----	0.002	0	A>G	----
SIX6	Chr14:60976393	rs78954112	0	0.002	G>C	Glu93Gln
SIX6	Chr14:60976501	rs146737847	0.008	0.002	G>A	Glu129Lys
SIX6	Chr14:60976537	rs33912345	0.47	0.37	A>C	Asn141His
SIX6	Chr14:60977843	rs45549246	0.002	0.002	T>G	Leu205Arg
SIX6	Chr14:60977864	rs202029915	0.002	0	C>T	Thr212Met
SIX6	Chr14:60977954	rs139302405	0.002	0	G>T	Ser242Ile

MAF = minor allele frequency; AA = amino acid. Coordinates are given based on the Hg19 reference.

Genotyping of rs33912345 in the Duke POAG case-control dataset (482 cases, 433 controls) resulted in a significant association (OR = 1.40, p = 0.0005, POAG case minor allele frequency (MAF) = 0.47, POAG control MAF = 0.38) with POAG. This SNP is in high linkage disequilibrium ($r^2 = 0.95$) with the intergenic SNP originally identified in POAG and VCDR genome-wide association studies (rs10483727) (Macgregor, Hewitt et al. 2010; Ramdas, van Koolwijk et al. 2010; Wiggs, Yaspan et al. 2012; Iglesias, Springelkamp et al. 2013). As expected, meta-analysis of the imputed genotype data from the NEIGHBOR and GLAUGEN studies confirmed a highly significant association between POAG status and rs33912345 (OR = 1.27, $p = 4.2X10^{-10}$) and other strongly linked SNPs in the region (Appendix E). A closer look at this locus shows that the association signal includes both upstream and downstream regions of the *SIX6* gene, while remaining entirely downstream of the *SIX1* gene (Figure 4).

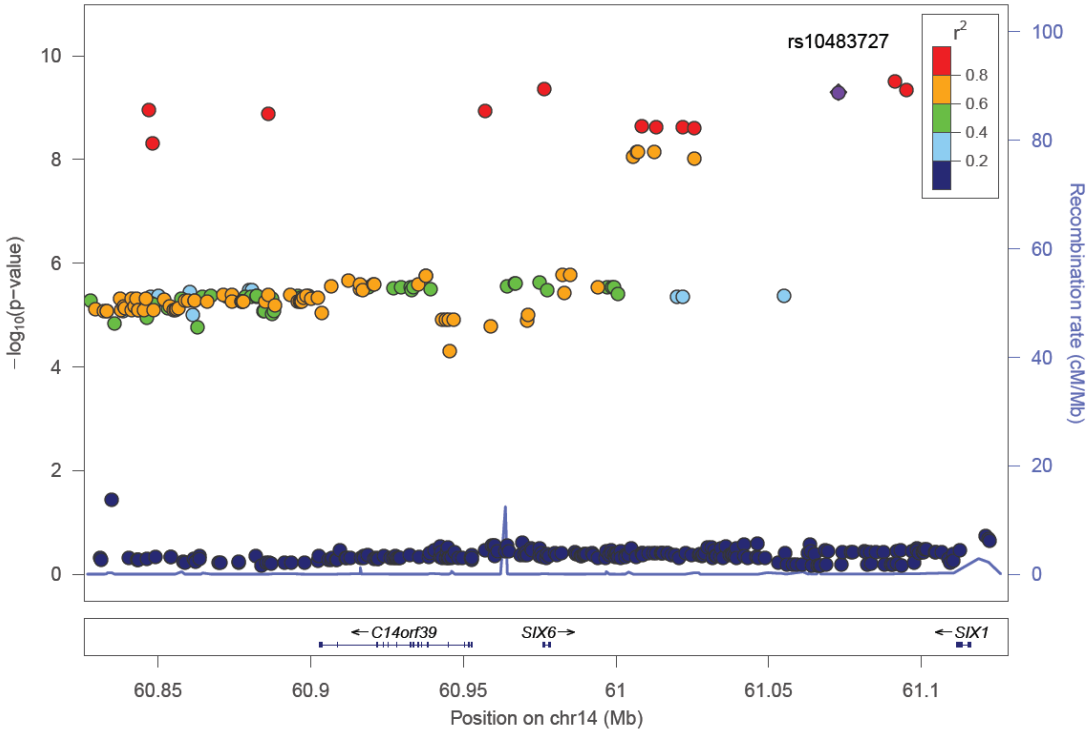


Figure 4: Plot of the association signal observed at the SIX1/SIX6 locus

Plot shows the association results calculated using chromosome 14 imputed genotype data color coded by the r^2 value with the index SNP, rs10483727 (designated in purple).

An analysis of retinal characteristics of POAG cases possessing the *SIX6* risk and non-risk allele was performed using optical coherence tomography (OCT) (Table 6). OCT images were only available for POAG cases with the common SNP, rs33912345; no data were available for individuals with the rare *SIX6* variants. We assessed retinal nerve fiber layer (RNFL) thickness in thirty POAG cases homozygous for the rs33912345 risk allele (C) or the non-risk allele (A) first by comparing age at disease diagnosis and age at OCT across the two genotypes, because age is known to influence retinal thickness and is a potential confounder. No significant difference in age was observed ($p = 0.11$, $p = 0.14$, respectively). Next, RNFL thickness was evaluated. The overall thickness (global RNFL) was significantly reduced in cases homozygous for the risk allele compared to cases with the non-risk allele ($p = 0.03$; mean (SD): C = 58.3 (8.2) μm, A = 67.9 (12.4) μm; Table 6), consistent with the hypothesis that *SIX6* may increase POAG susceptibility via changes in the neural retina. To determine which quadrants were driving this observation, we performed an exploratory, post-hoc comparison of RNFL thickness in the temporal, nasal, inferior, and superior regions. Interestingly, RNFL thickness was significantly reduced in the inferior ($p = 0.03$) and superior ($p = 0.04$) quadrants, the two regions directly affecting VCDR measurements.

Table 6: OCT measurements from POAG cases homozygous for rs33912345

Variable	Allele	N	Mean (SD)	p
Age at POAG diagnosis (years)	A	18	63.9 (9.1)	0.11
	C	12	69.7 (10.2)	
Age at OCT measurement (years)	A	18	71.2 (9.7)	0.13
	C	12	76.9 (10.3)	
Nerve fiber thickness (µm): Global	A	18	67.9 (12.4)	0.03
	C	12	58.3 (8.2)	
Temporal	A	18	57.2 (15.2)	0.85
	C	12	56.2 (12.8)	
Nasal	A	18	57.4 (13.4)	0.06
	C	12	47.9 (12.8)	
Inferior	A	18	79.6 (18.7)	0.03
	C	12	65.2 (12.6)	
Superior	A	18	77.0 (22.2)	0.04
	C	12	63.9 (10.8)	

3.3.2 *In vivo* functional interrogation of *SIX6* missense variants

Given 1) the observed association signal pattern; 2) the lack of coding changes identified in *SIX1*; 3) the presence of rare missense variants and a common, highly associated missense SNP in *SIX6*; 4) retinal nerve fiber layer thickness changes observed in POAG cases homozygous for the *SIX6* risk allele; and 5) the localized expression of *SIX6* in ocular tissues, we concluded *SIX6* is likely the better candidate gene in this region and evaluated the functional relevance of *SIX6* using an *in vivo* zebrafish complementation assay.

First, we performed a reciprocal BLAST analysis and identified two validated orthologs of SIX6 in the zebrafish genome, Six6a and Six6b, both with 91% homology at

the protein level. Previous overexpression and loss of function studies of SIX6 in mouse and frog models reveal a role in regulating the proliferative state of retinal progenitor cells and the size of the eye (Zuber, Perron et al. 1999; Li, Perissi et al. 2002); therefore, to evaluate whether the identified *SIX6* variants are pathogenic and potentially relevant to POAG, we sought to ask whether 1) morpholino-induced suppression of *six6a* or *six6b* leads to a reduced eye size; 2) expression of the human *SIX6* non-risk allele rescues the morphant eye phenotype; and 3) expression of *SIX6* alleles containing POAG risk variants rescues the morphant eye phenotype.

Using translation-blocking morpholinos targeting zebrafish *six6a* and *six6b*, we injected 1-8 cell stage embryos (N = 50-150) and analyzed live embryos at 3 days post fertilization (dpf). A translation-blocking morpholino was used because *SIX6* is a two exon gene—splice-blocking morpholinos are generally not recommended for two exon genes as the targeted transcript will not be subject to nonsense mediated decay, possibly leading to the expression of a truncated protein and potential dominant-negative effects (described by the manufacturer: http://www.gene-tools.com/node/18). Masked scoring[7] of both *six6a* and *six6b* morphants revealed ocular phenotypes consistent with loss of function, including a reduction in eye size in more than 80% of embryos ($p < 0.001$, Figure 5). The specificity of the morpholino reagents was tested by co-injection of 12.5

[7] Zebrafish morpholino optimization, mask scoring, and comparative analyses were performed by Drs. Edwin Oh and Yangfan Liu at the Duke Center for Human Disease Modeling.

pg of the human *SIX6* non-risk allele mRNA and significant ($p < 0.001$) rescue was observed in *six6a* but not *six6b* morphant embryos (90% vs. 10% of embryos, respectively). Together, these data indicate that Six6a is the functional ortholog of human SIX6 and prompted subsequent evaluation of *SIX6* variants using the *six6a* morpholino.

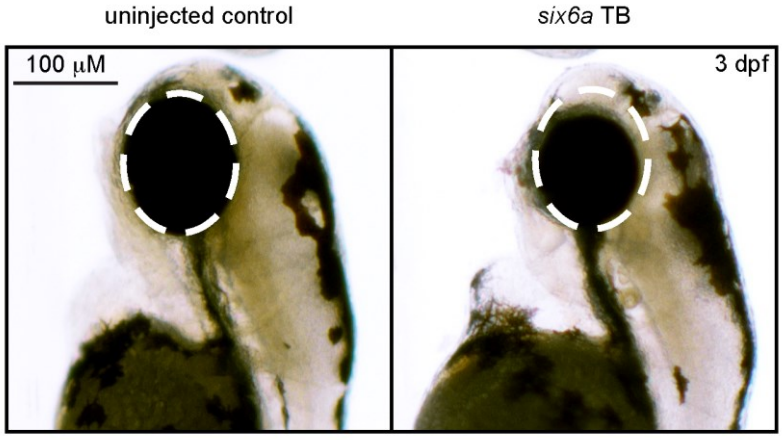


Figure 5: Morpholino knockdown of *six6a*

The figure highlights the reduction in eye size observed in zebrafish microinjected with a *six6a* translation blocking morpholino. Lateral images, taken 3 days post fertilization (3 dpf), of a wild-type zebrafish (left) and a morpholino injected zebrafish (right) are shown, highlighting the small eye phenotype (dashed circle).

To investigate the pathogenic potential of all *SIX6* variants, we used total eye size and the rescue of the morphant phenotype as the assay's phenotypic readout. We injected a mixture containing *six6a* morpholino and the various human *SIX6* alleles containing the coding variants identified via sequencing (Table 5). These results were compared to the rescue condition of the human *SIX6* non-risk allele. Five of the six variants tested were unable to fully rescue the small eye phenotype. Four of these alleles (Glu129Lys, Asn141His, Thr212Met, and Ser242Ile) resulted in an average eye size larger than the morpholino alone ($p < 0.001$), but smaller than the rescue with the non-risk allele ($p < 0.001$), indicating that these alleles are hypomorphic (Figure 6, Table 7). We also observed one variant (Leu205Arg) with an average eye size smaller than the morpholino alone ($p = 0.002$; mean (SD): MO = 34,042 (5,763) μm, Leu205Arg = 31,568 (6,485) μm; Figure 6), suggesting that it may be functionally null. One allele (Glu93Gln) resulted in an eye size similar to the rescue with the non-risk allele ($p = 0.37$) and was determined to be benign. The benign allele was identified in one control individual, while the remaining hypomorphic and null alleles were identified either exclusively or primarily in POAG cases (Table 5). Injection of 12.5 pg of the human *SIX6* risk mRNA into non-morphant zebrafish provided no evidence of a toxic gain of function compared to injection with the non-risk allele.

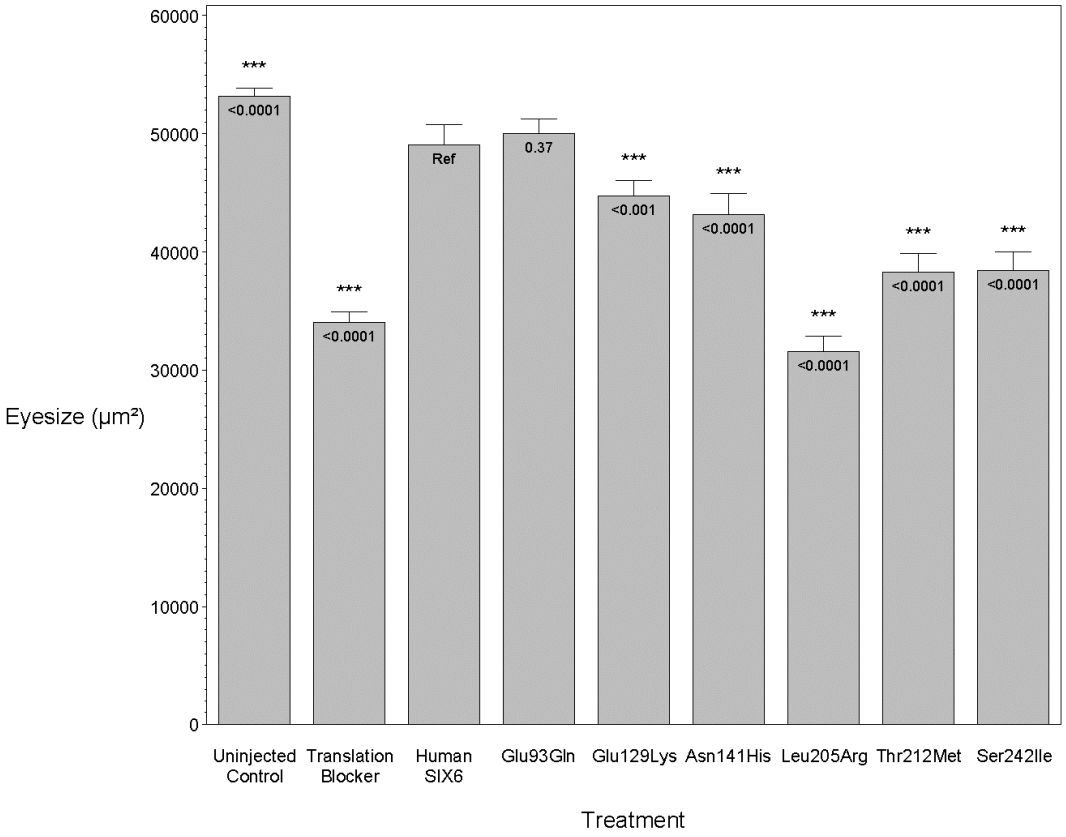


Figure 6: Results from an *in vivo* zebrafish morpholino assay showing the effects of *SIX6* nonsynonymous variants indentified in POAG cases and controls

The plot shows that compared to the uninjected controls, morphants have a significant reduction in eye size (µm²). Zebrafish were co-injected with a translation blocking morpholino designed to target *six6a* and human *SIX6* alleles (Glu93Gln, Glu129Lys, Asn141His, Leu205Arg, Thr212Met, or Ser242IIe). Results of each allele were compared to the SIX6 non-risk allele (Ref). P-values are provided below the mean of each treatment.

Table 7: Results from the *in vivo* zebrafish *six6a* morpholino assay showing changes in total eye size

The effect on eye size, given as mean μm² (standard deviation (SD)), was measured for wild-type zebrafish (Control) and zebrafish injected with a *six6a* morpholinos and SIX6 alleles identified by sequencing POAG cases and controls. Each treatment was compared to the zebrafish injected with the human *SIX6* non-risk allele (Ref).

Treatment		N	Mean eye size, μm² (SD)	p
Uninjected	Control	146	53,140 (4,536)	<0.0001
Translation Blocker	Morpholino	153	34,042 (5,763)	<0.0001
Human *SIX6*	Non-risk allele	177	49,040 (11,652)	Ref
Glu93Gln	Variant allele	92	50,010 (5,893)	0.37
Glu129Lys	Variant allele	100	44,719 (6,791)	<0.001
Asn141His	Variant allele	129	43,150 (10,660)	<0.0001
Leu205Arg	Variant allele	101	31,568 (6,485)	<0.0001
Thr212Met	Variant allele	84	38,278 (7,244)	<0.0001
Ser242Ile	Variant allele	88	38,420 (7,310)	<0.0001

3.3.3 *In vitro* functional interrogation of *SIX6* enhancer variants

We hypothesized that POAG risk may be mediated not only by defects in SIX6 protein function, but also by the level of *SIX6* gene expression. To test this, we sequenced the *SIX6* retinal specific enhancer element in 262 POAG cases and 256 POAG controls and identified 5 variants (Chr14:60974363_C, Chr14:60974373_T, Chr14:60974378_T, Chr14:60974400_A, Chr14:60974449_G) (Table 5). The effect of these variants on expression was tested using an *in vitro* luciferase assay. Of the five variants tested, one (Chr14:60974449_G) resulted in a significant increase in expression compared to the reference enhancer (Figure 7). Activation of the *SIX6* enhancer requires two cofactors, NeuroD and E47 (Figure 8) (Conte, Marco-Ferreres et al. 2010).

Overexpression was observed with the Chr14:60974449_G variant even in the absence of these cofactors (Figure 9), suggesting variants within the enhancer region may result in dysregulated protein expression.

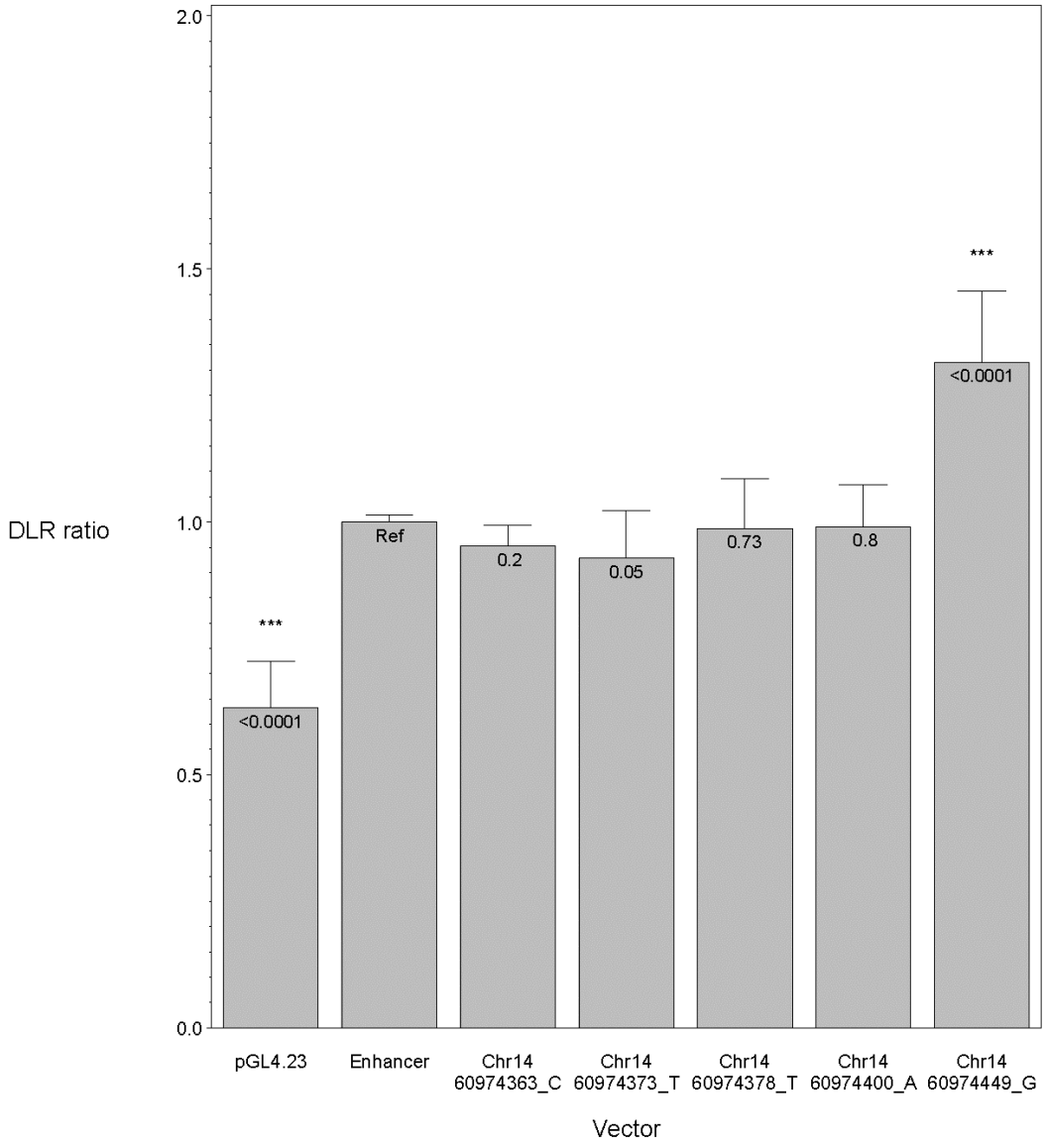


Figure 7: *In vitro* luciferase assay results showing the effects of *SIX6* enhancer variants identified in POAG cases and controls

SIX6 enhancer alleles were tested using a dual luciferase assay and the ratio of the experimental luciferase: control luciferase was calculated (DLR ratio). All vectors were co-transfected with NeuroD and E47. In this context, the *SIX6* enhancer is functioning to increase expression compared to the empty vector (pGL4.23), driven by a minimal promoter. Compared to the reference enhancer (Ref), one variant (Chr14:60974449_G) significantly increases the enhancer's activity. Coordinates are based on the Hg19 reference. P-values are provided below the mean of each vector.

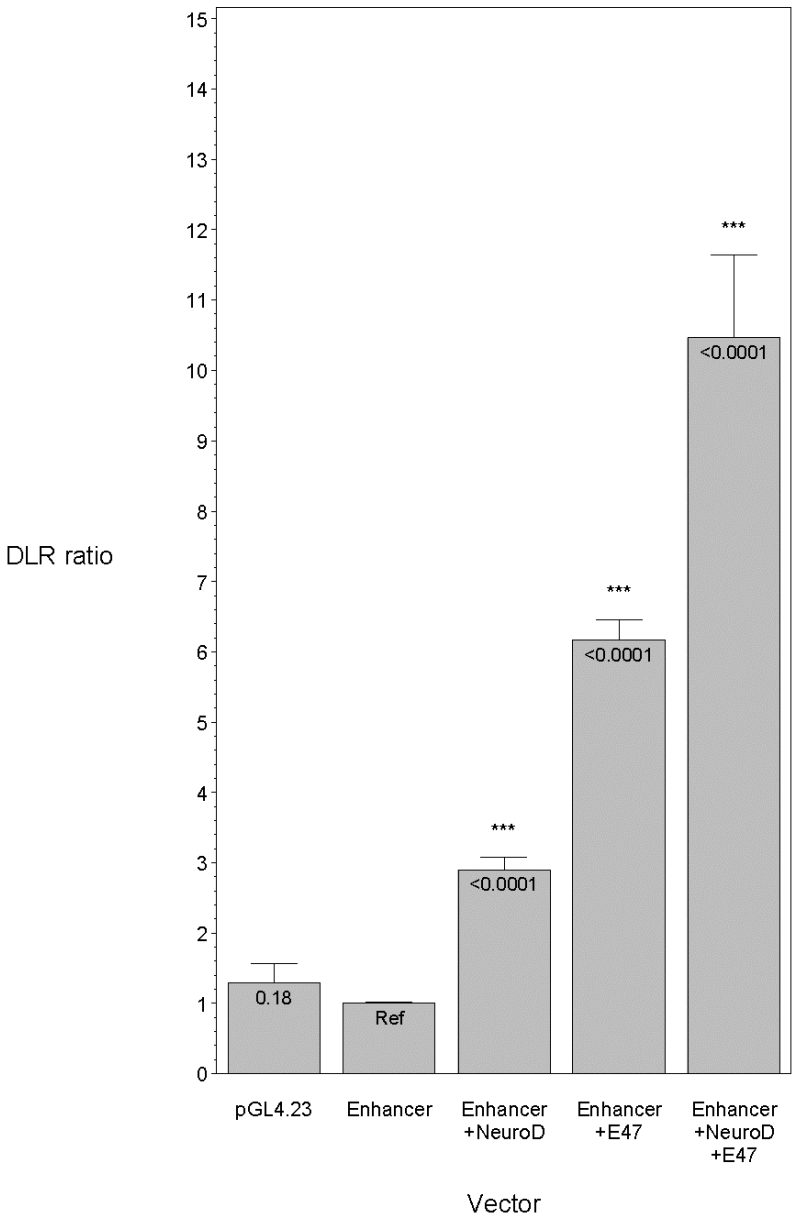


Figure 8: *In vitro* luciferase assay results from the *SIX6* enhancer co-transfected with *NeuroD* and *E47*

The *SIX6* enhancer alone was inactive compared to the empty vector (pGL4.23). Co-transfection of both *NeuroD* and *E47* significantly increased the enhancer's activity. The highest activity level was reached with co-transfection of an equal amount of both, as previously described (Conte, Marco-Ferreres et al. 2010). P-values are provided below the mean of each vector.

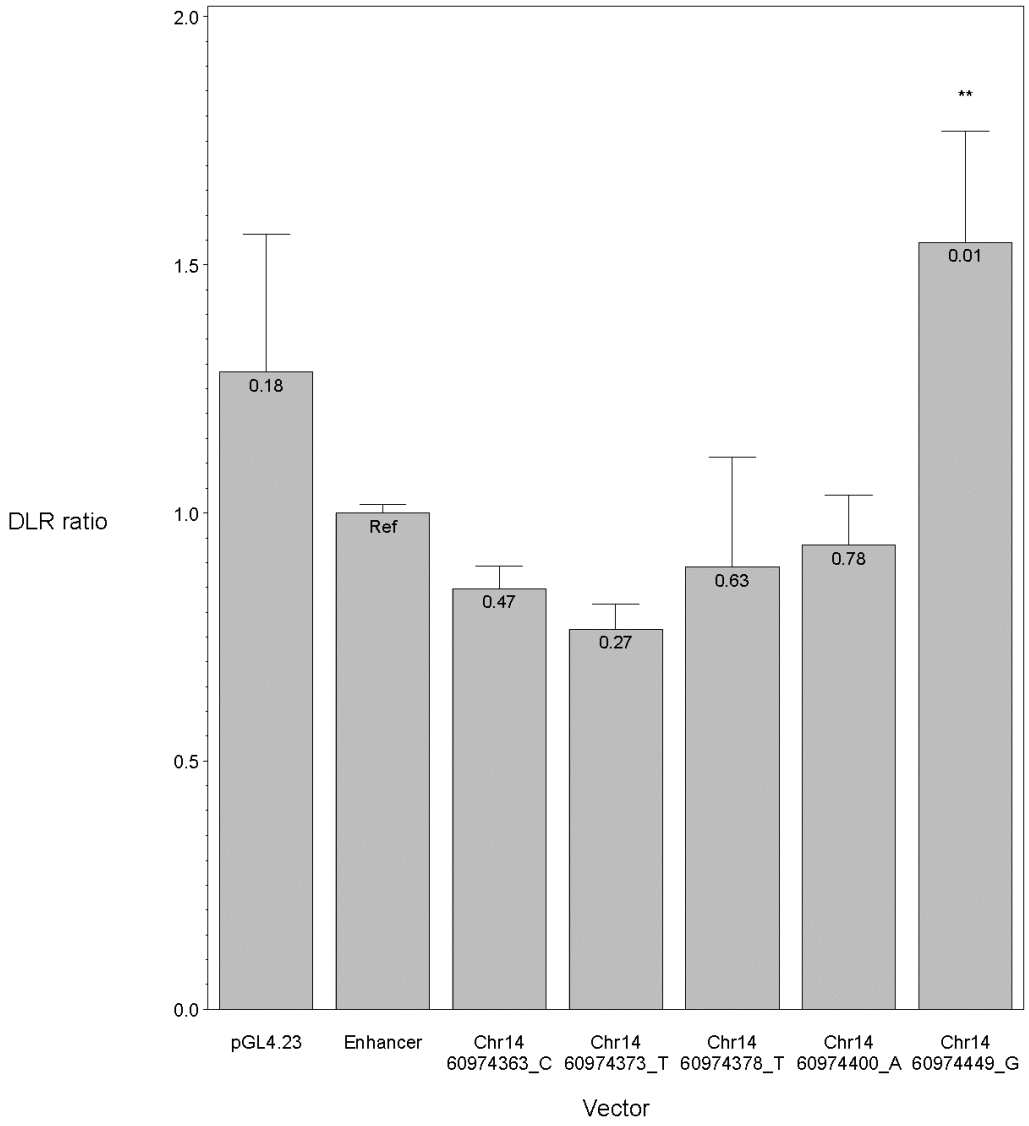


Figure 9: *In vitro* luciferase assay results from the *SIX6* enhancer variants without co-transfection with *NeuroD* and *E47*

In the absence of NeuroD and E47, the *SIX6* enhancer is not active (Ref compared to the empty vector, pGL4.23). However, the Chr14:60974449_G variant still shows increased enhancer activity. Coordinates are based on the Hg19 reference. P-values are provided below the mean of each vector.

3.4 Discussion

The *SIX1*/*SIX6* locus has been shown to be significantly associated with POAG in several independent studies; however, the causal variant(s) driving this association have remained unknown (Ramdas, van Koolwijk et al. 2010; Fan, Wang et al. 2011; Ramdas, van Koolwijk et al. 2011; Osman, Low et al. 2012; Wiggs, Yaspan et al. 2012; Iglesias, Springelkamp et al. 2013). We have demonstrated through several lines of evidence—the tissue specificity of the *SIX* gene family described in the literature, the identification of *SIX6* missense variants in our POAG dataset, and the results of the *in vivo* and *in vitro* assays— that *SIX6* is the most likely POAG susceptibility gene in this region. We identified common and rare coding variants, which alter the function of the SIX6 protein. We also identified a novel variant within the *SIX6* enhancer that appears to disrupt the normal regulation of SIX6 expression, suggesting both regulatory and coding variants may influence POAG susceptibility at this locus. Finally, we are the first to identify clinical features in POAG patients that may be dependent upon *SIX6* genotype: patients who are homozygous for the *SIX6* risk allele have a statistically thinner retinal nerve fiber layer than patients homozygous for the *SIX6* non-risk allele.

The common variant, rs33912345 (Asn141His), which we showed has significantly reduced function in an *in vivo* model, was associated with POAG in our dataset and is in high linkage disequilibrium with the originally identified GWAS SNP, rs10483727 (Wiggs, Yaspan et al. 2012). This highly conserved amino acid is located

within the alpha helix structure of the DNA homeobox domain of *SIX6*. Interestingly, the ancestral allele (C, His) is associated with POAG risk. The allele frequency of this variant differs greatly among populations (C allele frequency: YRI (0.99), ARF (0.96), ASN (0.76), and CEU (0.42); from 1000 Genomes release 14) (Abecasis, Auton et al. 2012). This locus is significantly associated with an increase in vertical cup-disc ratio (VCDR) in population controls, indicating that it may be involved in the normal development of the optic nerve. VCDR is also a clinical measure used to track disease progression in POAG patients (Charlesworth, Kramer et al. 2010; Macgregor, Hewitt et al. 2010; Ramdas, van Koolwijk et al. 2010; Ramdas, van Koolwijk et al. 2011; Iglesias, Springelkamp et al. 2013). Interestingly, African populations have larger VCDRs and an increase in overall POAG prevalence (Girkin 2008; Girkin, Sample et al. 2010; Knight, Girkin et al. 2012). In a recent population-based study, the prevalence of POAG in Ghana exceeded 17%, the highest rate observed anywhere in the world (Budenz, Bandi et al. 2012). While rs33912345 is not associated with POAG risk in a West African POAG case/control dataset from Ghana, the frequency of the ancestral (risk) allele is 99% in both cases and controls (Liu, Hauser et al. 2013). We hypothesize that differences in the structure of the optic nerve and the higher risk of POAG in individuals with African ancestry may, in part, arise from the fact that essentially all individuals in this population are homozygous for the rs33912345 ancestral risk allele.

Given the association of the *SIX6* locus with neural retinal measurements such as VCDR, it is reasonable to hypothesize causal variants may function by inducing quantitative changes in cell populations in the retina. There is extensive evidence that SIX6 regulates early retinal progenitor cell proliferation during eye development (Zuber, Perron et al. 1999; Bernier, Panitz et al. 2000; Li, Perissi et al. 2002). Li et al. showed that *Six6*$^{-/-}$ mice display varying degrees of retinal hypoplasia that is due to a decrease in retinal ganglion cell proliferation resulting from an early exit from the cell cycle during development (Li, Perissi et al. 2002). This is consistent with functional studies of XOptx2, the Xenopus ortholog of SIX6 (Zuber, Perron et al. 1999). We have shown a reduction in eye size upon morpholino knockdown of zebrafish *six6a,* and we were able to rescue this phenotype with co-injection of the human *SIX6* non-risk allele, demonstrating that the zebrafish *six6a* gene is the likely functional ortholog to human *SIX6.* We identified five alleles that could not fully rescue the small eye phenotype, and we observed a reduction in retinal nerve fiber layer thickness in POAG patients homozygous for the His141 SIX6 risk allele. Taken together with previously published findings, our results suggest that risk variants in the human *SIX6* gene increase POAG susceptibility by negatively affecting retinal ganglion cell development, leading to a reduction in the number of retinal ganglion cells in adulthood. Retinal ganglion cells are lost during the normal aging process, and this rate of loss could be increased by the presence of additional glaucoma risk alleles or other risk factors such as increased intraocular pressure (Calkins

2013). The development of glaucomatous optic neuropathy and associated visual field loss would be hastened by a reduction in the initial number of retinal ganglion cells that an individual possesses.

While the role of *SIX6* in development is well established, it remains unclear what function *SIX6* may have in adult tissues. *SIX6* may increase POAG susceptibility through an unknown role in neuron maintenance. Consistent with this hypothesis, our study results show that rare and common variants within the *SIX6* gene result in a loss of protein function and that changes in the *SIX6* regulatory region result in aberrant expression. This and other SIX6 variants could increase the risk of glaucoma by increasing the rate of RGC death, as well as by reducing the initial number of retinal ganglion cells.

In this study, we have identified multiple common and rare *SIX6* sequence variants in POAG cases, and used *in vivo* and *in vitro* assays to demonstrate that these variants have functional consequences on *SIX6* gene expression and protein function. While other risk factors may be required for the onset of POAG, we hypothesize that changes to SIX6 protein function increase an individual's susceptibility to developing the disease via changes to retinal development. Additional work is needed, possibly through the use of transgenic animal model studies, to fully understand the role of *SIX6* in POAG. However, this is the first study to demonstrate clinical differences in POAG

patients carrying *SIX6* risk alleles and to perform a functional analysis of alleles identified in POAG cases.

4. The transcriptional landscape of POAG-related tissues: the trabecular meshwork, cornea, and ciliary body

4.1 Introduction

Normal vision requires the proper development of and a synergistic relationship between various tissues. The trabecular meshwork (TM) and ciliary body (CB) tissues are of particular interest to the field of glaucoma due to their role in intraocular pressure regulation, the primary glaucoma risk factor (Kwon, Fingert et al. 2009). Additionally, the cornea is an extraordinarily important tissue for normal vision, as it is responsible for the majority of light refraction, and for glaucoma, because a reduction in corneal thickness is associated with an increased risk of developing POAG, discussed in detail in Chapter 2. There are many corneal diseases including keratoconus, brittle cornea syndrome, and Fuchs corneal dystrophy. Therefore, understanding corneal biology is fundamentally important.

Some of the most basic biological questions can be answered by looking at the transcriptome, or the overall collection of RNA molecules within a cell or tissue. By global gene expression profiling, transcriptome analysis provides information about the inner workings of the cell type or tissue being studied. To date, a number of studies have investigated gene expression profiles of ocular tissues. However, previous studies have primarily been limited to microarray-based technologies (Liton, Luna et al. 2006; Turner, Budak et al. 2007; Janssen, Gorgels et al. 2012; Janssen, Gorgels et al. 2013; Liu,

Allingham et al. 2013). Microarray analysis is a hybridization-based approach that requires the transcriptional targets to be predefined. Transcriptional analysis via RNA sequencing (RNA-seq) is becoming increasingly popular because, unlike standard microarrays, there is a reduced need for *a priori* biological knowledge. RNA-seq can be used to identify the gene expression level of both known and novel transcripts.

In this study, we performed deep RNA sequencing of human adult and fetal ocular tissues samples to develop the first unbiased glimpse into the transcriptional landscape of the trabecular meshwork, cornea, and ciliary body tissues. We provide a basic descriptive analysis that gives insight into the related biology. We also use this data to investigate the gene expression levels and the presence of novel isoforms in POAG candidate genes identified by genome-wide association studies. This could easily be extended to additional ocular disease genes. Thus, this dataset is broadly applicable for ocular research.

4.2 Materials and methods

4.2.1 Sample selection

Ocular samples were available from the North Carolina Eye Bank (https://www.nceyebank.org/) (Winston-Salem, North Carolina, USA). De-identified medical records were available and included age, race, sex, postmortem delay, primary and secondary cause of death, and a history of ocular disease (yes or no). Limited ocular medical records were available. Without the ability to properly phenotype individual

samples, we chose to select only samples with no known history of ocular disease. Two fetal samples and four adult samples were selected for transcriptome analysis (Table 8) based on the following criteria: no known history of ocular disease, matching age and sex, and Caucasian ancestry.

4.2.2 RNA purification and sequencing

The North Carolina Eye Bank procured the tissue samples and treated whole globes with RNA*later* (Qiagen, Limburg, Netherlands) in order to preserve the RNA integrity. The globes were then dissected by an ophthalmologist[1], and the individual tissues were re-treated with RNA*later* by incubation at 4°C for 24 hours. The samples were then removed and stored at -80°C until the RNA extraction was performed.

Tissues were homogenized using pre-sterilized tubes containing hard tissue grinding mix (Omni International, California, USA) in an Omni Bead Ruptor 24 using the following settings: 5.5 m/s, 15 sec, 2 cycles, D = 15 sec. RNA was then extracted from each tissue sample using the *mir*Vana total RNA extraction kit (Ambion, Texas, USA), following the manufacturer's protocol. The ciliary body is highly pigmented and melanin co-segregates with RNA during the isolation process, inhibiting downstream reactions (Dorrie, Wellner et al. 2006). Therefore, the ciliary body tissues went through a melanin removal processes modified from a previously published protocol (Satyamoorthy, Li et al. 2002). P-60 exclusion chromatography beads (Bio-Gel,

[1] Dr. Rand Allingham (Duke University Eye Center) performed the tissue dissections from whole eyes.

California, USA) were prepared as described. Then, 800 μL of the prepared beads were placed in a Micro Bio-Spin chromatography column (Bio-Rad, California, USA). RNA was applied to the column, incubated for 10 minutes on ice, and centrifuged at 1000 rpm for 1 minute, keeping the RNA containing flow-through. To elute the remaining RNA, this process was repeated using 100 μL of diethylpyrocarbonate (DEPC) treated 10 mM sodium acetate.

RNA from each sample was DNAase treated using the DNA-*free* kit (Life Technologies, New York, USA). Subsequently, the quality was assessed using a chip-based technology (RNA 6000 pico kit) on a 2100 Bioanalyzer (Agilent Technologies, California, USA) and was determined to be high quality (mean RNA integrity number (RIN) score = 8.5; min. RIN = 7.3; max. RIN = 8.7).

RNA libraries were prepared using the TruSeq RNA sample preparation kit (Illumina, California, USA) and sequenced on four lanes of an Illumina HiSeq at the Duke Genome Sequencing and Analysis Core[2]. Samples were randomized across library preparation batches, barcoded, and randomized across sequencing lanes to reduce technical variability. The sequencing was determined to be high quality; all 12 samples had reads with mean quality scores (Q) > 37 and 94.5% or more reads with a $Q \geq 30$. The total number of paired reads per sample ranged from 32,137,380 to 59,784,117.

[2] The TruSeq RNA library preparation and sequencing was performed by members of the Duke Genome Sequencing and Analysis Core.

4.2.3 Transcriptome analysis

First, sequencing reads were checked for quality using the fastqc program, specifically looking at per base sequencing quality and overrepresented sequences (a list of programs used in the transcriptome analysis can be found in Appendix F). Based on these results, Illumina adapter sequences were trimmed using the cutadpat program (Martin 2011). The transcriptome was assembled using the Tuxedo package pipeline (Trapnell, Pachter et al. 2009). Pre-indexed human genome reference sequences (Ensembl build GRCH37) were downloaded from the TopHat website ((http://tophat.cbcb.umd.edu/igenomes.shtml) and were used in all subsequent analyses. Reads were mapped using TopHat using the default parameters with the following adjustments: the library distribution mean, based on trace files provided by the sequencing center, was set to 250 bp (--mate-inner-dist) with a standard deviation of 100 bp (--mate-std-dev 100) and 3 mismatching base pairs per read were allowed (--read-mismatches, --read-edit-dist). The reference transcriptome was used in the initial alignment step (-G). The alignment was quality control checked by calculating the overall alignment rate and the percentage of properly paired mates using samtools (Li, Handsaker et al. 2009). The TM tissues had an average alignment rate of 97.0% (min. 90.5%, max. 98.6%) and a properly paired rate of 72.0% (min. 67.6%, max. 74.2%). The cornea tissues had an average alignment rate of 96.5% (min. 89.2%, max. 98.4%) and a properly paired rate of 71.6% (min 69.8%, max. 73.0%). The CB tissues had an average

alignment rate of 97.3% (min. 96.6%, max. 98.0%) and a properly paired rate of 74.3% (min 69.9%, max. 74.4%).

The resulting alignment files were merged based on tissue and age (6 transcriptomes total: adult and fetal TM, cornea, and CB). Pooling samples improves overall read depth, resulting in more accurate isoform expression estimates. The transcriptomes were assembled using the Cufflinks program in two different ways (Trapnell, Williams et al. 2010). For both, a reference transcriptome was supplied. The first (assembly 1) did not allow for the identification of novel transcripts (-G) and only called transcripts previously annotated in the human transcriptome reference file supplied by Ensembl (build GRCH37). The second (assembly 2) allowed for the identification of novel transcripts, but used the known transcriptome as a guide (--GTF-guide). The goal of assembly 1 is to accurately estimate gene expression levels. The goal of assembly 2 is to identify novel isoforms of candidate genes. The Cufflinks program was run using the default parameters modified to allow for an increased number of reads per locus (--max-bundle-frags) and a larger genomic length per bundle (--max-bundle-length).

Post assembly, output files were annotated using Ensembl build GRCH37 annotation files. Genes with the GENCODE biotype (http://www.gencodegenes.org/gencode_biotypes.html) of Mt_rRNA, Mt_tRNA, Mt_tRNA_pseudogene, rRNA, rRNA_pseudogene, or tRNA pseudogene, and deprecated genes (predicted protein

coding or non-coding RNAs from previous version of the human genome reference that have been removed from the current version) were removed from the descriptive analysis. Genes that failed the analysis (given in the Cufflinks output files as 'lowdata' and 'failed') were also removed. A gene or transcript was determined to be expressed if the fragments per kilobase of transcript per million mapped reads (FPKM) were equal to or great than 1. Downstream descriptive analyses were performed in SAS (SAS Institute Inc. 2002-2008). Post assembly transcriptome files were viewed in RNAseqViewer (Roge and Zhang 2013).

4.3 Results

4.3.1 Ocular tissue sample characteristics

Sixteen ocular tissues samples were selected from a collection of tissues available through the North Carolina Eye Bank. Sample characteristics are shown in Table 8.

Table 8: Ocular tissue sample characteristics

Sample	Age	Race	Sex	Post Mortem Delay	Cause of Death	Tissue
1	28 week gestation	Caucasian	M	4:40	Prematurity	TM, Cornea, CB
2	23 week gestation	Caucasian	F	13:28	Prematurity	TM, Cornea
3	73	Caucasian	M	6:46	Sepsis	TM, Cornea, CB
4	76	Caucasian	F	4:14	Myocardial infarction	TM, Cornea, CB
5	66	Caucasian	M	3:56	Small bowel obstruction	TM, Cornea
6	66	Caucasian	F	4:00	Sepsis	TM, Cornea, CB

4.3.2 Characteristics of the trabecular meshwork, cornea, and ciliary body transcriptomes

4.3.2.1 Initial filtering and annotation

Utilizing the output from assembly 1 (requiring transcripts to match the annotated human transcriptome), the total number of expressed genes was evaluated. Gene expression output is given in fragments per kilobase of transcript per million mapped reads (FPKM), a measure of gene expression normalized by gene length and reads/sample. Before any filters were applied, the average number of genes with an FPKM value > 0 expressed in each tissue ranged from 27,523 to 36,685. To improve confidence in the downstream analyses, an FPKM threshold of 1, generally accepted to be above background noise, was used. After applying this filter, the average number of genes expressed in each tissue ranged from 12,628 to 16,684 (Appendix G). The expression value of several genes (an average of 57/sample) could not be determined by the calling algorithm (flagged as failed and low data), and these were removed. Next, genes were annotated based on the GENCODE biotype provided with the Ensembl reference annotation files (Appendix H). Genes with the following biotypes were removed: Mt_rRNA, Mt_tRNA, Mt_tRNA_pseudogene, rRNA, rRNA_pseudogene, and tRNA_pseudogene. Looking at the distribution and top most expressed genes before additional filters were applied revealed that the gene expression distribution is heavily skewed (Appendix I) and that there is an overrepresentation of deprecated genes in the top most expressed genes (Appendix J). These were excluded from the final analysis,

although this does not ensure other deprecated or undefined genes do not remain in the dataset due to complications in genome annotation.

4.3.2.2 Distribution of gene expression in six ocular tissue RNA-seq datasets

After the initial filter, the number of genes expressed was determined by looking at the FPKM values. An average of 15,099 genes were expressed in each tissue (Table 9). The distribution is heavily skewed; interquartile ranges are provided in Table 9. The adult CB sample had the smallest number of expressed genes (N = 12,101), but the highest gene expression value (FPKM = 40,456). The adult TM had the largest number of expressed genes (N = 17,175).

Table 9: Distribution of gene expression values in six ocular tissue RNA-seq datasets

Sample	N	Q1-25%	Median	Q3-75%	FPKM Max
Adult TM	17,175	3.39	8.98	21.99	23,949
Fetal TM	15,980	3.28	8.22	20.00	21,479
Adult Cornea	15,934	3.29	9.11	23.76	15,484
Fetal Cornea	16,157	3.33	8.49	20.13	13,533
Adult CB	12,101	2.66	8.15	30.70	40,456
Fetal CB	13,247	2.52	7.34	27.92	7,438
Average	15,099	3.08	8.38	24.08	20,390

4.3.3 Most highly expressed genes in adult and fetal trabecular meshwork, cornea, and ciliary body tissues

The most highly expressed genes per tissue type were determined by ranking the calculated gene expression level by the FPKM. Table 10 shows the top ten most expressed genes in the overall dataset. Most tissues showed high gene expression levels for mitochondrial genes, particularly mitochondrially encoded ATP synthases, mitochondrially encoded cytochrome c oxidases, and mitochondrially encoded NADH dehydrogenases. Excluding these, the top twenty most expressed genes in each tissue are given in Tables 11-16.

Table 10: Top ten most expressed genes overall from trabecular meshwork, cornea, and ciliary body ocular tissue samples

Sample	Ensemble ID	Locus	Gene Name	FPKM
Adult CB	ENSG00000228253	MT:8365-8572	MT-ATP8	40,456
Adult TM	ENSG00000228253	MT:8365-8572	MT-ATP8	23,949
Adult CB	ENSG00000107317	9:139871955-139879887	PTGDS	22,151
Fetal TM	ENSG00000113140	5:151040656-151066726	SPARC	21,479
Adult CB	ENSG00000198899	MT:8526-9207	MT-ATP6	15,848
Adult Cornea	ENSG00000228253	MT:8365-8572	MT-ATP8	15,484
Fetal TM	ENSG00000108821	17:48260649-48278993	COL1A1	14,647
Adult Cornea	ENSG00000226958	X:108297360-108297792	CTD-2328D6.1	13,931
Fetal Cornea	ENSG00000113140	5:151040656-151066726	SPARC	13,533

Table 11: Top twenty most expressed genes from adult trabecular meshwork

Adult Trabecular Meshwork		
Gene ID	**Gene Name**	**FPKM**
RNU1-4	RNA, U1 small nuclear 4	10,704
RNU1-2	RNA, U1 small nuclear 2	10,678
APOD	apolipoprotein D	10,052
B2M	beta-2-microglobulin	5,198
RNY4	RNA, Ro-associated Y14	4,570
RNY1	RNA, Ro-associated Y1	4,383
EEF1A1	eukaryotic translation elongation factor 1 alpha 1	3,962
PTGDS	prostaglandin D2 synthase 21kDa (brain)	3,912
ANGPTL7	angiopoietin-like 7	3,693
RNU1-5	RNA, variant U1 small nuclear 18	3,630
RNU1-6	RNA, variant U1 small nuclear 9	3,622
RNU1-9	RNA, variant U1 small nuclear 7	3,622
RNU1-3	RNA, U1 small nuclear 3	3,603
RNU1-1	RNA, U1 small nuclear 1	3,596
RNU1-7	RNA, U1 small nuclear 27, pseudogene	3,570
RNU1-8	RNA, U1 small nuclear 28, pseudogene	3,569
TPT1	tumor protein, translationally-controlled 1	3,183
S100A9	S100 calcium binding protein A9	3,169
CLU	clusterin	2,991
TMSB4X	thymosin beta 4, X-linked	2,441

Of the top twenty most highly expressed genes in the adult trabecular meshwork, 12 were mitochondrial encoded proteins (not shown). Pseudogenes and ribosomal proteins are not shown.

Table 12: Top twenty most expressed genes from fetal trabecucular meshwork

Fetal Trabecular Meshwork		
Gene ID	**Gene Name**	**FPKM**
SPARC	secreted protein, acidic, cysteine-rich (osteonectin)	21,479
COL1A1	collagen, type I, alpha 1	14,648
COL1A2	collagen, type I, alpha 2	10,909
ANGPTL7	angiopoietin-like 7	5,962
EEF1A1	eukaryotic translation elongation factor 1 alpha 1	5,309
TPT1	tumor protein, translationally-controlled 1	2,730
MGP	matrix Gla protein	2,609
DCN	decorin	2,324
FMOD	fibromodulin	2,213
IGFBP2	insulin-like growth factor binding protein 2, 36kDa	1,774
PCOLCE	procollagen C-endopeptidase enhancer	1,743
EEF1A1P5	eukaryotic translation elongation factor 1 alpha 1	1,364
FTL	ferritin, light polypeptide	1,295
COL3A1	collagen, type III, alpha 1	1,278
BGN	biglycan	1,263
GNB2L1	guanine nucleotide binding protein (G protein), beta	1,182
ELN	elastin	1,162
TMSB4X	thymosin beta 4, X-linked	1,160
NACA	nascent polypeptide-associated complex alpha subunit	1,116
MYL6	myosin, light chain 6, alkali, muscle and non-muscle	1,107

Of the top twenty most highly expressed genes in the fetal trabecular meshwork, 9 were mitochondrial encoded proteins (not shown). Pseudogenes and ribosomal proteins are not shown.

Table 13: Top twenty most expressed genes from adult cornea

Adult Cornea		
Gene ID	**Gene Name**	**FPKM**
CLU	clusterin	12,074
KRT5	keratin 5	8,213
FTH1	ferritin, heavy polypeptide 1	6,377
KRT12	keratin 12	5,838
S100A4	S100 calcium binding protein A4	5,829
S100A9	S100 calcium binding protein A9	4,644
S100A6	S100 calcium binding protein A6	4,570
TMSB4X	thymosin beta 4, X-linked	4,111
B2M	beta-2-microglobulin	3,789
TGFBI	transforming growth factor, beta-induced, 68kDa	3,268
EEF1A1	eukaryotic translation elongation factor 1 alpha 1	3,141
ENO1	enolase 1, (alpha)	3,129
TPT1	tumor protein, translationally-controlled 1	3,083
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	2,981
S100A8	S100 calcium binding protein A8	2,835
MT1X	metallothionein 1X	2,737
KRT13	keratin 13	2,465
HSPB1	heat shock 27kDa protein 1	2,432
KRT14	keratin 14	2,242
GSTP1	glutathione S-transferase pi 1	1,939

Of the top twenty most highly expressed genes in the adult cornea, 10 were mitochondrial encoded proteins (not shown). Pseudogenes and ribosomal proteins are not shown.

Table 14: Top twenty most expressed genes from fetal cornea

Fetal Cornea		
Gene ID	**Gene Name**	**FPKM**
SPARC	secreted protein, acidic, cysteine-rich (osteonectin)	13,533
COL1A1	collagen, type I, alpha 1	11,759
COL1A2	collagen, type I, alpha 2	11,582
EEF1A1	eukaryotic translation elongation factor 1 alpha 1	5,123
DCN	decorin	5,026
LUM	lumican	4,413
KERA	keratocan	3,448
KRT5	keratin 5	3,122
H19	H19, imprinted maternally expressed transcript	2,522
FMOD	fibromodulin	2,481
TPT1	tumor protein, translationally-controlled 1	2,196
KRT12	keratin 12	2,181
COL5A1	collagen, type V, alpha 1	2,176
TMSB4X	thymosin beta 4, X-linked	1,873
MGP	matrix Gla protein	1,591
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	1,575
COL5A2	collagen, type V, alpha 2	1,442
EEF1A1P5	Euk. translation elongation factor 1 alpha 1 pseudogene 5	1,364
COL6A1	collagen, type VI, alpha 1	1,273
CLU	clusterin	1,235

Of the top twenty most highly expressed genes in the fetal cornea, 8 were mitochondrial encoded proteins (not shown). Pseudogenes and ribosomal proteins are not shown.

Table 15: Top twenty most expressed genes from adult ciliary body

Adult Ciliary Body		
Gene ID	**Gene Name**	**FPKM**
PTGDS	prostaglandin D2	22,151
GPX3	glutathione peroxidase 3 (plasma)	7,507
FTH1	ferritin, heavy polypeptide 1	6,339
IFITM3	interferon induced transmembrane protein 3	4,500
ALDOA	aldolase A, fructose-bisphosphate	3,488
OPTC	opticin	3,169
TRAJ49	T cell receptor alpha joining 49	2,769
EEF2	eukaryotic translation elongation factor 2	2,733
MYH11	myosin, heavy chain 11, smooth muscle	2,359
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	2,097
PKM2	pyruvate kinase, muscle	2,072
SERPING1	serpin peptidase inhibitor, clade G, member 1	2,005
MYL9	myosin, light chain 9, regulatory	1,933
AES	amino-terminal enhancer of split	1,887
CKB	creatine kinase, brain	1,887
CRYAB	crystallin, alpha B	1,873
IGHJ5	immunoglobulin heavy joining 5	1,733
FTL	ferritin, light polypeptide	1,563
GSN	gelsolin	1,550
APOD	apolipoprotein D	1,541

Of the top twenty most highly expressed genes in the adult ciliary body, 12 were mitochondrial encoded proteins (not shown). Pseudogenes and ribosomal proteins are not shown.

Table 16: Top twenty most expressed genes from fetal ciliary body

Fetal Ciliary Body		
Gene ID	**Gene Name**	**FPKM**
PTGDS	prostaglandin D2 synthase 21kDa (brain)	6,800
CLU	clusterin	4,954
FTH1	ferritin, heavy polypeptide 1	4,232
EEF2	eukaryotic translation elongation factor 2	4,141
H19	H19, imprinted maternally expressed transcript	3,583
ELN	elastin	3,398
COL9A2	collagen, type IX, alpha 2	2,620
VIM	vimentin	2,511
EEF1A1	eukaryotic translation elongation factor 1 alpha 1	2,257
GNAS	GNAS complex locus	2,049
COL9A3	collagen, type IX, alpha 3	1,991
SPARC	secreted protein, acidic, cysteine-rich (osteonectin)	1,974
PMEL	premelanosome protein	1,780
ALDOA	aldolase A, fructose-bisphosphate	1,616
BSG	basigin (Ok blood group)	1,533
ACTB	actin, beta	1,512
COL6A2	collagen, type VI, alpha 2	1,507
ACTG1	actin, gamma 1	1,479
DES	desmin	1,456
COL6A1	collagen, type VI, alpha 1	1,441

Of the top twenty most highly expressed genes in the fetal ciliary body, 6 were mitochondrial encoded proteins (not shown). Pseudogenes and ribosomal proteins are not shown.

4.3.4 Gene expression of POAG GWAS candidate genes and identification of novel isoforms

The gene expression levels of ten well-established POAG candidate genes, identified by genome-wide association studies, were determined for each tissue type (Table 17) (Janssen, Gorgels et al. 2013). Based on the overall gene expression distribution, genes with an FPKM $\geq$ 1, an FPKM $\geq$ 4.3 (33rd percentile), and an FPKM $\geq$ 16.3 (67th percentile) were determined to be lowly (highlighted in yellow), moderately (highlighted orange), or highly expressed (highlighted red), respectively.

Table 17: Gene expression of POAG candidate genes

GENE	Adult TM	Fetal TM	Adult Cornea	Fetal Cornea	Adult CB	Fetal CB
ATOH7	0.06	1.21	0.14	0.29	0.03	0.52
CAV1	34.44	35.47	25.13	43.07	0.87	0.41
CAV2	32.57	14.05	27.43	14.86	2.45	1.67
CDKN2A	3.87	0.02	10.86	0.44	0.58	0.11
CDKN2B	14.66	0.05	27.19	4.32	0.11	0.10
CDKN2B-AS	0.67	0.10	0.67	0.07	0.02	0.00
ELOVL5	49.05	29.79	14.50	19.63	1.28	0.87
SIX6	0.02	0.96	0.01	1.95	13.77	18.22
SRBD1	4.53	4.51	4.41	5.97	0.10	0.18
TMCO1	54.34	45.51	66.41	56.79	1.42	0.65

Of the genes investigated, most showed some evidence of expression. The *CDKN2B*, *CDKN2A*, and *CDKN2B-AS* genes represent a highly replicated POAG-associated locus, although currently the causal gene(s) remains unknown. Interestingly, *CDKN2B-AS* was not found to be expressed in any of the tissues studied, suggesting

CDKN2B and *CDKN2A* may be better candidate genes for follow-up studies. *SIX6*, which is generally considered to function in posterior structures of the eye during development, was found to be highly expressed in the human adult and fetal ciliary body.

We also looked for the presence of novel isoforms in the POAG candidate gene loci. Using a FPKM cutoff of 1, three loci (*CDKN2B, SIX6,* and *SRBD1*) contained five potentially novel isoforms. The remaining loci (*ATOH*, *CAV1*/*CAV2*, *ELOVL5*, *TMCO1*) contained evidence of novel isoforms (FPKM > 0), but these isoforms had expression values below the set threshold (FPKM < 1), reducing confidence in their existence (Table 18).

Table 18: Evidence of novel isoforms in POAG candidate genes
All seven of the loci investigated contained evidence of novel isoform expression. Four of these loci had FPKM values below the threshold (designated by an FPKM < 1 in the novel isoforms column). Five novel isoforms above the FPKM threshold were identified.

Locus	Coordinates	Novel Isoforms	Sample
ATOH7	10:69988385-69993871	FPKM < 1	--
CAV1/CAV2	7:115925433-116203233	FPKM < 1	--
CDKN2B	9:21965750-22123096	2 Novel	Adult TM, Adult Cornea
ELOVL5	6:53130195-53215947	FPKM < 1	--
SIX6	14:60973668-60981568	1 Novel	Fetal CB
SRBD1	2:45613818-45841304	2 Novel	Adult TM, Adult CB
TMCO1	1:165694031-165740417	FPKM < 1	--

Four novel isoforms with FPKM values > 1 were identified at the *CDKN2B* and *SRBD1* loci in the adult TM, CB, and cornea tissues. SIX6, which showed high expression in the ciliary body, contained a potentially novel isoform in the fetal ciliary body tissue (Table 19). The two most expressed novel isoforms were identified in two genes highly associated with POAG in the NEIGHBOR/GLAUGEN meta-analysis (*CDKN2B* and *SIX6*). The novel isoforms at the *CDKN2B* locus were similar to the transcript ENST00000276925, with modified end points (Figure 10). However, the novel *SIX6* isoform is significantly different from any known transcript (Figure 11).

Table 19: Expression of novel isoforms at the *CDKN2B*, *SIX6* and *SRBD1* loci

Sample	Locus	Transcript Coordinates	FPKM
Adult TM	CDKN2B	9:22002954-22009337	3.31
Adult Cornea	CDKN2B	9:22002901-22009371	4.44
Fetal CB	SIX6	14:60975411-60978231	32.06
Adult TM	SRBD1	2:45615913-45776211	1.31
Adult CB	SRBD1	2:45673652-45673918	1.65

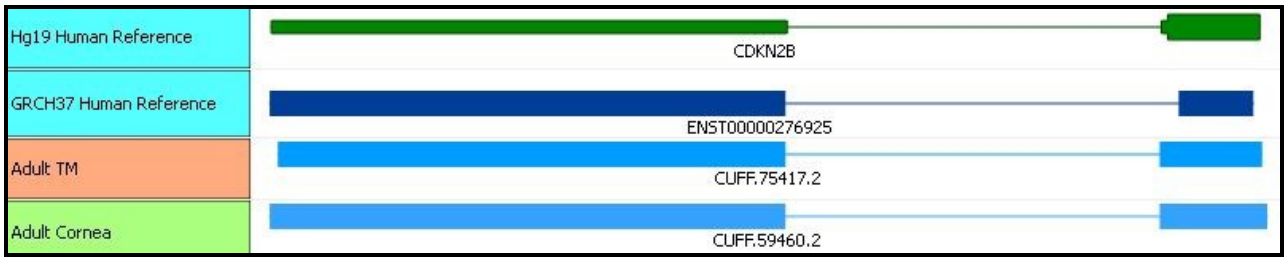


Figure 10: Novel isoforms at the *CDKN2B* locus identified in human adult trabecular meshwork and cornea samples

Potentially novel isoforms are shown in light blue and are assigned a unique identifier by the Cufflinks program (CUFF.xxx). Two novel isoforms at the *CDKN2B* locus were identified in the adult TM and cornea. The closest related isoforms from the human genome references, Hg19 and GRCH37, are shown above.

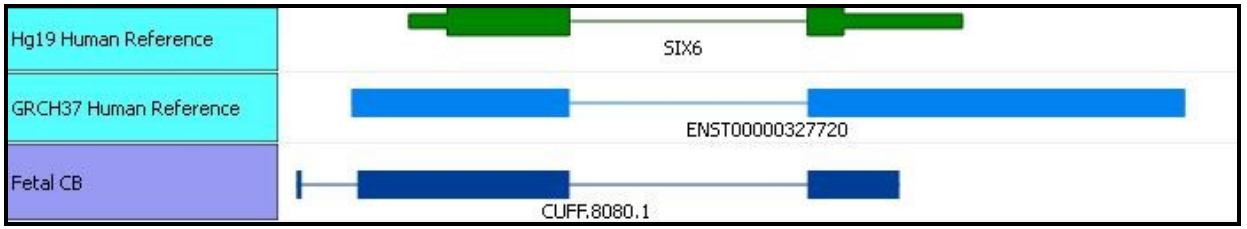


Figure 11: Novel isofrom at the *SIX6* locus identified in the fetal ciliary body sample

A potentially novel isoform at the *SIX6* locus identified in the fetal CB sample (bottom) and the most closely related isoforms from the human genome references, Hg19 and GRCH37, are shown above.

4.4 Discussion

RNA sequencing (RNA-seq) is an enormously informative approach for transcriptome analysis and does not rely entirely on *a priori* knowledge (Wang, Gerstein et al. 2009). It has been used extensively to study healthy and diseased human tissues. However, ocular tissues, especially anteriorly located tissues like the trabecular meshwork, cornea, and ciliary body, have not previously been interrogated with this approach. These tissues are essential for normal ocular function and play important roles in common diseases like glaucoma. In this study, the transcriptional landscape of the trabecular meshwork, cornea, and ciliary body was described using RNA-seq, and the resulting data will be directly applicable in the future for a wide range of ocular research fields.

On average, we identified over 15,000 expressed genes in each tissue. This is consistent with the findings of a recent study that used a similar approach for analyzing the human limbus, a tissue that connects the cornea and the sclera (Bath, Muttuvelu et al. 2013) . While descriptive statistics, such as the number of expressed genes and the distribution of gene expression levels, provide essential tissue specific insight into the transcriptome, the data generated in this study may be used in a variety of additional ways to propel ocular research forward. As we have shown, it can be used to explore the expression of candidate genes in disease relevant tissues, which can help with the prioritization of candidate genes identified in association studies. Genetic association

studies can identify novel regions of interest, but are limited in their ability to directly identify causal genes due to the high correlation that is often locally observed at genetic loci. Using RNA-seq data to identify expression in relevant tissues helps narrow the search. We used this approach and showed *CDKN2A* and *CKDN2B* are highly expressed in POAG-related tissues, while *CDKN2B-AS* was not found to be expressed in any of the tissues examined. This locus is associated with POAG, but the exact causal gene(s) remains unknown (Burdon, Macgregor et al. 2011; Osman, Low et al. 2012; Wiggs, Yaspan et al. 2012; Janssen, Gorgels et al. 2013). The RNA-seq data suggest *CDKN2A* and *CKDN2B* could be given priority over *CDKN2B-AS* in future POAG functional analyses. This, of course, is not conclusive; *CDKN2B-AS* may function in other relevant tissues, like the retina and optic nerve, or may be expressed in diseased tissues. But, the RNA-seq data has provided a starting point for functional follow-up and aided in hypothesis generation.

In addition to evidence of candidate gene expression, RNA-seq data can be used to investigate transcript level information. The specific isoforms of a protein may vary by tissue and identifying specific transcript expression can help with the design of functional follow-up studies. This can be enhanced by using an analytical approach, like the one presented in this study, which allows for the identification of completely novel transcripts. The tissues included in this study were predominately selected based on their relevance to glaucoma biology. So, we specifically looked at transcript level

expression of candidate genes identified via POAG genome-wide association studies. We identified several potentially novel isoforms in POAG candidate genes including the *SIX6* gene. This gene is of particular interest due to the recent identification of POAG specific, functional alleles (Chapter 3). Interestingly, *SIX6* was found to be highly expressed in the adult and fetal ciliary body. While Aijaz et al. previously showed the presence of *SIX6* expression in the human ciliary body, this gene is primarily considered a retinal specific development gene (Kawakami, Sato et al. 2000; Aijaz, Allen et al. 2005; Kumar 2009). We identified a novel *SIX6* transcript in the fetal ciliary body that is remarkably different than any currently known. While this could be an artifact of the program's calling algorithm due to the complicated nature of the sequencing data, it is possible *SIX6* has a novel function in ciliary body development and POAG susceptibility that remains unknown. More work is needed to confirm the presence of novel isoforms in these tissues and determine what role they may have in ocular function.

While RNA-seq is undoubtedly useful, there are some limitations that can hamper downstream applications. In theory, high-throughput sequencing of the entire mRNA population within a cell or tissue eliminates the need for a complete, annotated transcriptome. This is especially important in less studied tissues, like those included in this analysis, because the transcriptome can be highly tissue specific. However, interpretability without the annotation can very difficult. In our dataset, we identified highly expressed genes that are currently labeled as deprecated genes, meaning they

were once considered possible transcripts in previous versions of the reference transcriptome, but have now been removed. The current annotation makes it difficult to infer what role these genes may play in the biology of these tissues. This highlights the need for functional validation, which will become increasingly important as this technique is more frequently applied.

The overall goal of this study was to increase our understanding of the transcriptional landscape of POAG-related tissues and to generate a dataset that could help guide future research. We have shown the benefits of an RNA-seq approach by the identification of potentially novel isoforms and, by looking at a selection of POAG candidate genes, several novel hypotheses have emerged. This approach could be extended to other ocular diseases involving the trabecular meshwork, cornea, and ciliary body, and is therefore broadly applicable to many areas of ocular research.

5. Conclusion

5.1 Project specific considerations and future directions

Primary open-angle glaucoma (POAG) is a common blinding disease with a complex genetic component. The overall goal of studying the genetics of POAG is to better understand who may be at an increased risk of developing the disease (prediction) and the underlying biological mechanisms (biology) in order to better treat patients in the future. The work presented in Chapters 2-4 directly contribute to the field in different ways: Chapter 2) increasing the knowledge-base of a commonly studied POAG quantitative risk factor (prediction); Chapter 3) the characterization of a POAG candidate gene and the identification of functional variants from a genetically associated locus (prediction and biology); and Chapter 4) a description of the transcriptional landscape of essential POAG-related tissues (biology).

5.1.1 POAG endophenotype analyses

As described in Chapter 2 (Central corneal thickness (CCT) analysis: a potential POAG endophenotype), the identification of disease susceptibility alleles via quantitative endophenotypes is an attractive approach. An endophenotype is defined here as a heritable quantitative trait that is state independent (present within the normal population), associated with a disease, and is assumed to lie within the causal pathway (Flint and Munafo 2007; Kendler and Neale 2010). Endophenotypes are of particular importance in diseases like POAG because often the endophenotype is easier to measure

than the disease itself. If a trait is a true endophenotype, there may be SNPs that directly affect the quantitative trait (e.g. reduction in CCT) which may then lead to or increase susceptibility to the disease. If this is the case, the causal SNP(s) would be associated with both traits (the endophenotype and the disease).

Often, studies of potential endophenotypes are conducted as secondary trait analyses, where a genome-wide dataset is re-analyzed for the secondary trait of interest. While secondary trait analyses are appealing because they are cost effective and can provide a unique dataset (e.g. CCT analysis in the context of POAG cases), they can also be greatly affected by biases. This is especially true when the secondary trait is associated with the primary trait, which can lead to an inflated Type I error rate (Monsees, Tamimi et al. 2009; Kendler and Neale 2010). Of course the level of the Type I error is dependent on the effect of the associated endophenotypes and SNP(s) (i.e. if the secondary trait has only a marginal effect on the disease, the bias will be small). An example of this issue is observed in genetic studies of intraocular pressure (van Koolwijk, Ramdas et al. 2012; Ozel, Moroi et al. 2014). Using the NEIGHBOR and GLAUGEN POAG case-control dataset, we identified SNPs in *GAS7* and *TMCO1*, known to be associated with POAG, which were also significantly associated with intraocular pressure (Ozel, Moroi et al. 2014). Because the distribution of intraocular pressure is significantly different in POAG cases compared to controls, any SNPs associated with POAG are going to consequentially appear to be associated with

differences in intraocular pressure, making it difficult to postulate if the original hypothesis is correct (causal SNP directly increases intraocular pressure which then directly increases POAG risk).

So, how does one determine if the observed association signal from a secondary analysis is due to study design bias or a true endophenotype status? While there is no clear answer, certain approaches can be used to help clarify the true relationship. Population-based data are less likely, depending on the disease prevalence, to be susceptible to these biases. In other words, if a SNP is associated with the secondary trait in a population-based cohort, it is less likely to be due to allele frequency differences caused by an underlying association to a disease. This increases the confidence that the SNP may directly affect the tested trait. Looking for an overlap of genetic associations from population-based studies provides increased evidence that a trait may be a true endophenotype. This is the case with the association of intraocular pressure and the *GAS7* and *TMCO1* genes and the association of vertical cup-disc ratio and the *SIX6* gene, discussed in Chapter 3. In both cases, the genetic associations of the quantitative risk factors were observed in non-diseased populations and independently shown to increase disease risk in POAG case-control datasets, suggesting intraocular pressure and vertical cup-disc ratio may be true POAG endophenotypes (Charlesworth, Kramer et al. 2010; Macgregor, Hewitt et al. 2010; Ramdas, van Koolwijk et al. 2010; van Koolwijk, Ramdas et al. 2012; Wiggs, Yaspan et al. 2012; Ozel, Moroi et al. 2014). This, however,

was not observed in central corneal thickness analyses, and there is mounting evidence suggesting CCT may not be a true POAG endophenotype (discussed more in Chapter 2) (Charlesworth, Kramer et al. 2010; Dimasi, Burdon et al. 2012).

These studies, which focus on identifying the shared genetics of an endophenotype and a trait, are unable to decipher between pleiotropy, where SNP(s) cause both the quantitative trait and the disease, and mediation, where the quantitative risk factor lies within the causal pathway. Mediation analyses may help decipher between the two in some circumstances. This analytical approach directly tests the hypothesis of pleiotropy versus mediation using statistical methods (MacKinnon, Lockwood et al. 2002; Kendler and Neale 2010). However, to fully understand the relationship between SNPs associated with quantitative risk factors and the disease of interest, functional laboratory-based assays are ultimately needed. An example of this is discussed in Chapter 3 (Discovery and functional annotation of *SIX6* variants in POAG), where we showed that the *SIX6* locus, associated with both vertical cup-disc ratio and POAG, likely functions via quantitative changes to the retina.

Finally, elucidating the exact endophenotypic status of quantitative risk factors may not be essential if the overall goal is disease prediction. Instead, building prediction models that account for both the genetic and quantitative trait effects and that can accurately assess the potential for disease development would result in earlier disease diagnosis. Individuals could then be preemptively treated with intraocular pressure

reducing topical medications, which have been shown to be effective in slowing POAG progression even in the absence of ocular hypertension (Collaborative Normal-Tension Glaucoma Study Group 1998; Heijl, Leske et al. 2002; Miglior, Pfeiffer et al. 2007). In combination with continued basic research, an emphasis on developing clinically accessible predictive models could have an enormous impact on improving the rate of blindness caused by glaucoma. A predictive model for POAG was published 2007 (Gordon, Torri et al. 2007). However, this model does not account for newly discovered genetic markers, may not be applicable to individuals under the age of 40, and is limited to ocular hypertensives, which represent only a subset of POAG cases (Dielemans, Vingerling et al. 1994; Iwase, Suzuki et al. 2004; Varma, Ying-Lai et al. 2004; Coleman and Miglior 2008; Mudumbai 2013).

One of the additional challenges with predictive modeling is that many of the well-established quantitative POAG risk factors are not routinely measured in undiagnosed individuals. Therefore, for predictive modeling to be successful, known quantitative risk factors like central corneal thickness and vertical cup-disc ratio need to be more commonly measured. Predictive modeling could dramatically improve the prognosis for POAG patients. There is much room for growth in this area, and it should be considered a priority by POAG researchers in the future.

5.1.2 The role of *SIX6* in POAG

The study of the discovery and functional annotation of *SIX6* variants in POAG (Chapter 3) is the first true functional follow-up to a POAG genome-wide association study. It is unique in the fact that the alleles tested were directly identified in POAG cases and showed that rare and common variants influence the protein's functionality. Interestingly, *SIX6* is traditionally thought of as a developmental gene, while POAG has a late age of onset. This study revealed that developmental processes, retinal ganglion cells development in this case, may affect POAG onset decades later. There is still much to learn about this gene and its role in POAG.

Understanding more about the normal functional of the *SIX6* gene could teach us about the biology of POAG and may result in the identification of additional candidate genes or pathways. This can be challenging since *SIX6* has a restricted range of gene expression (primarily expressed in posterior ocular tissues and the pituitary). To learn more about the gene's basic function, pituitary cell lines may be of use since they are easily available, well-established, and should endogenously express the *SIX6* gene. *SIX6* is a known transcription factor, but a comprehensive characterization of its downstream targets has not been completed. An appropriate approach would be to perform chromatin immunoprecipitation-sequencing (ChIP-seq) using anti-bodies targeting *SIX6* in a pituitary cell line. While this approach does not directly answer questions related to the role of *SIX6* in POAG, it does provide a wealth of novel information which could be

used to direct future POAG studies. Immediate downstream targets of *SIX6* could be considered as high priority candidate genes because they may also function to increase POAG susceptibly. This could be tested using a genetic association analysis or performing a zebrafish morpholino assay, similar to approach described in Chapter 3. In addition, understanding intricate gene networks is useful in complex disease research as downstream targets of any POAG causal gene could be considered as potential targets for drug therapy.

An experiment looking at the gene expression or protein level changes of a retinal ganglion cell biomarker (*POU4F2*) in morpholino or cell culture-based knockdown assays would be a more direct approach for studying the role of *SIX6* in POAG and for testing our primary hypothesis (*SIX6* regulates the number of retinal ganglion cells created during development). The *POU4F2* gene is expressed almost exclusively in retinal ganglion cells, and therefore can be used to extrapolate the total number of retinal ganglion cells present (Xiang, Zhou et al. 1993). Correlating changes in *POU4F2* after knockdown, rescue, and co-injection of POAG specific alleles would provide further evidence that *SIX6* increases POAG susceptibility via a direct effect on retinal ganglion cell development. In a similar manner, other potential candidate genes could be analyzed. Cell cycle control genes are of particular interest. There are several lines of evidence that suggest members of the *SIX* gene family directly bind and regulate the expression of cyclin-dependent kinase inhibitor (*CDKN*) genes (Li, Perissi et al. 2002;

Del Bene and Wittbrodt 2005; Iglesias, Springelkamp et al. 2013). Interestingly, a locus at Chr.9p21, that houses *CDKN2A, CDKN2B,* and *CDKN2B-AS1*, is highly associated with POAG (Gibson, Griffiths et al. 2012; Osman, Low et al. 2012; Wiggs, Yaspan et al. 2012; Janssen, Gorgels et al. 2013). The interaction of these two loci would be a novel finding with major implications in the field of POAG research.

In Chapter 3, we showed that coding and regulatory variants at the *SIX6* locus had a functional effect. We hypothesized regulatory variants may influence the level of *SIX6* gene expression, which could in turn influence POAG susceptibility. We focused our efforts on a small region upstream of the *SIX6* gene containing a well-defined enhancer element. The current study could be expanded by extending our approach to include the *SIX6* promoter region. Because the promoter has not been defined, a promoter bashing experiment should be performed first to define the region. Then, repeating the experiments presented in Chapter 3, POAG cases and controls should be sequenced, constructs containing disease alleles generated, and a luciferase assay should be used to test the effects of the identified variants. The enhancer variants previously identified were extraordinarily rare and, due to the small sample size used in our sequence study, we were unable to perform association testing at this region. Therefore, this experiment could also be extended by sequencing additional individuals and performing a rare variant association test to determine if the presences of rare regulatory

variants are more common in POAG cases than controls, providing further evidence the regulatory region may play a causal role in disease development.

5.1.3 The transcriptional landscape of POAG-related tissues

One goal of the work described in Chapter 4 (The transcriptional landscape of POAG-related tissues: the trabecular meshwork, cornea, and ciliary body) was to learn about the basic biology of essential ocular tissues. This was completed by the descriptive analysis of the transcriptome of each tissue at two different developmental stages (fetal and adult). There are many follow-up analyses that can be conducted with the RNA-seq data generated during this project. Of particular interest, a differential gene expression analysis comparing the developmental stages could be performed. A related study was recently published (Young, Hawthorne et al. 2013). In this study, global gene expression profiles from fetal and adult ocular tissues (retina, choroid, sclera, optic nerve, and cornea) were compared. As described by Young et al., comparing gene expression profiles at these two time points aids in the identification of genes involved in normal ocular development as well as genes potentially responsible for diseases characterized by ocular growth abnormalities, such as myopia. The RNA-seq data generated in Chapter 4 could be analyzed using a similar approach. This analysis differs from the previous study in two important ways: 1) our dataset contains two unique tissues types (trabecular meshwork and ciliary body) and 2) our study was conducted with RNA-seq rather than a microarray-based platform. Because of these differences, our analysis could

yield novel results that may not have been identified in the original study. Additionally, this analysis could serve to replicate the original findings and improve the overall confidence in the results.

Using the RNA-seq data, we showed that many POAG candidate genes are expressed in relevant tissues. We also showed that an important POAG candidate gene (*CDKN2B-AS*) was not expressed in the tissues examined. Finally, we identified several novel isoforms. This analysis could easily be extended to include additional POAG candidate genes or to include genes from additional phenotypes such as POAG-related quantitative traits. The same approach, determining the level of gene expression and looking for evidence of novel isoforms, could be applied to these additional candidate genes. Genes identified in central corneal thickness studies, discussed in Chapter 2, would be particularly interesting candidates for follow-up analyses. Similarly, it would be valuable to look at the trabecular meshwork and ciliary body RNA-seq data for genes previously associated with intraocular pressure because of the known role these tissues have in intraocular pressure regulation.

While the identification of novel isoforms in POAG candidate genes is intriguing, these need to be validated. Any inferences made from the RNA-seq data could be incorrect due to random error, errors caused by flaws in the program's calling algorithm, or individual variability. Therefore, validation of the results is essential. Unfortunately, the fetal ocular tissues were extremely small and were consumed during

the RNA sequencing process. Therefore, the first step for validation is to collect additional ocular samples and extract RNA. To confirm the presence of the novel isoforms identified, a combination of real-time PCR and 5 prime or 3 prime RACE (rapid amplification of cDNA ends) will need to be conducted in the proper tissues types (Sambrook and Russell 2006). Additionally, any differentially expressed genes identified in a fetal versus adult comparative study will need to be validated using a real-time PCR or a similar approach in order to ensure the accuracy of the results.

The ultimate goal is to propel the field forward with the biology learned from the RNA-seq analysis. This dataset is broadly applicable to a wide range of ocular studies, so it should be shared with the research community. One possibility is to publish this data through the UCSC EyeBrowse genome browser (http://eyebrowse.cit.nih.gov/), which contains eye specific tracks. This includes cataracts, glaucoma, myopia, cornea, and retina disease candidate gene loci, expressed sequence tag (EST) data from a variety of tissues, and serial analysis of gene expression (SAGE) data from two tissues, retina and trabecular meshwork. Including our RNA-seq data in this browser allows other researchers to benefit because it is a unique dataset that is unlike those currently available.

5.2 Challenges and special considerations in POAG genetic studies

5.2.1 The importance of consistent clinical definitions: a challenge facing consortia

For a genomic-based study to be successful, there are many costly and time consuming steps that must be performed. This includes ascertainment of patient samples (this step alone can take years), collection of clinical data, integration of databasing systems and their management, large scale DNA extractions, DNA quality control checking , and organization of sample manifests. In addition, a large number of samples are required to ensure adequate study power, often leading to cross institutional collaborations. In the field of human genetics, consortia-based studies (large scale collaborations that share resources for the purpose of completing a common goal) are an entity that is here to stay. It is often too challenging for one investigator or one institution to gather all of the resources needed for a large genomic project, and the synergy provided from pooling resources has been effective in the field of human genetics.

However, consortia-based studies are accompanied by their own set of unique challenges including, but not limited to, variations in clinical definitions. Depending on the disease being studied, this could even include differences in case-control phenotyping, where one institution uses a different disease definition than another. There can also be cross-institutional differences in clinical data collection methods, the

type of clinical data available, their means of measurement, and the time period of measure. Consider POAG and intraocular pressure, the most common risk factor. This clinical variable is almost always available because it is measured during routine ocular examinations. Despite its availability, combining intraocular pressure data in large consortia can be challenging due to the following: 1) an institution may have repeat measurements for an individual and may have different measurement selection criteria; 2) there are numerous methods for measuring intraocular pressure, which can increase variability when pooling data; 3) measurements may be confounded because POAG patients take intraocular pressure reducing medication; and others. There may also be issues of missing clinical data, which can vary from site to site, making it impossible to adjust for essential covariates or to assess the effect of these traits on the disease of interest. Inconsistency in clinical data can make downstream analyses more challenging and can introduce error. Clinical variables should be assessed during model building to ensure the most accurate conclusions are being drawn. The accurate assessment of clinical traits in large genomic studies is essential because POAG, like other complex human diseases, is likely caused by a combination of environmental and genetic factors.

Due to the cost and the enormous amount of work required to perform a well powered genome-wide study, many consortia are driven to utilize the generated data in secondary analyses. There are many instances of this throughout the literature (Monsees, Tamimi et al. 2009). An example study is discussed in Chapter 2 where the

NEIGHBOR and GLAUGEN consortia data, collected initially for POAG analysis, was used to study the genetics of central corneal thickness. The motivation behind these secondary analyses is not only financial; many times the secondary trait is of particular interest to the original trait, as it is in the POAG and central corneal thickness study. While the motivation behind these secondary analyses is warranted, they present certain challenges that should be considered *a priori*. Like the primary analysis, they may suffer from differences in clinical definitions and data availability across sites, which can be exacerbated by the fact that the dataset may not have been collected with the secondary trait in mind. Again, this can make evaluating the effect of clinical traits difficult. As an example based on the analysis presented in Chapter 2, the effect of glaucoma surgery on central corneal thickness was evaluated to check for potential confounding within the dataset. Importantly, there was no significant difference observed. However, the glaucoma surgery data was only available for 38% of the dataset. Complete and systematically collected data would be ideal and would further our confidence in our analysis of the effect of glaucoma surgery on corneal thickness.

While these challenges can be overcome, these simple examples highlight the need for standard operation procedures, meticulous planning, and forethought. These issues, and others, can be avoided by using a prospective study design. In this approach, a researcher determines the clinical traits of interest, these are measured at baseline, and then participants are followed for years to determine if the disease develops. The Ocular

Hypertension Treatment study is an example of a prospective study, and this study successfully identified most of the commonly accepted POAG risk factors known to date (Gordon, Beiser et al. 2002). This type of approach in combination with genetic data would be immensely informative. It would help clarify the effects of clinical variables that have been inconsistently associated with POAG because a prospective study design allows for an increased confidence of inferences made on causation (Coleman and Miglior 2008). However, this is generally unrealistic as prospective studies tend to be timely, costly, and do not take advantage of preexisting data. When consortia are formed using a traditional case-control study design, like the NEIGHBOR study, data consistency issues can be circumvented by careful planning. A thorough review of the literature should be performed in order to identify newly discovered clinical risk factors that may become of interest. Ideally, as researchers begin to pool data a definitive list of clinical variables and their definitions would be predefined and each institution would simply extract this data from an in-house database. This can be complicated if an institution does not have a well maintained databasing system or if the desired clinical variables were not originally collected, leading to *post hoc* patient chart reviews. This is often unavoidable and the cost of such a venture is frequently underestimated. Therefore, proper budgeting when a large scale consortia-based project is initiated is essential. Recommendations for improving data consistency in a consortia environment

include 1) using a prospective study design whenever possible; 2) preplanning of designed clinical traits and their definition; and 3) proper budgeting.

5.2.2 Challenges and methods for dealing with heterogeneity in POAG

POAG is genetically and clinically heterogeneous (Allingham, Liu et al. 2009; Janssen, Gorgels et al. 2013). The most obvious example of clinical heterogeneity in POAG is observed by the presence and absence of elevated intraocular pressure. Despite elevated intraocular pressure being a major POAG risk factor, an estimated 30-40% of POAG cases never display pressures above the normal range, as discussed in Chapter 1 (Dielemans, Vingerling et al. 1994; Iwase, Suzuki et al. 2004; Varma, Ying-Lai et al. 2004; Mudumbai 2013). The presence of clinical heterogeneity may be indicative of genetic heterogeneity. Sub-phenotyping can be used for stratified or restricted analyses and, if based on biologically meaningful traits, reduces genetic heterogeneity within a sample and thereby increases statistical power in a genetic analysis. This is another reason, in addition to the concerns described above, that clinical data consistency and accuracy is immensely important in POAG genetic studies. In the context of intraocular pressure, sub-phenotyping approaches have proven successful. In the NEIGHBOR and GLAUGEN POAG meta-analysis, Wiggs et al. identified a region of interest that only reached genome-wide significance in a subset of normal tension POAG cases (Wiggs, Yaspan et al. 2012).

Stratified analysis can be a powerful approach for reducing genetic heterogeneity if based on prior biological or statistical evidence (i.e. formal interaction analysis). However, researchers should be careful not to randomly stratify a dataset. Inappropriately restricted datasets could lead to spurious associations and incorrect conclusions (Patsopoulos, Tatsioni et al. 2007). Not properly restricting analysis in the presence of clinical heterogeneity can also lead to biases. As described by Burdon et al., an example of this was observed in the first POAG genome-wide association study, which included a combination of POAG and pseudoexfoliation glaucoma samples (Thorleifsson, Magnusson et al. 2007; Burdon 2012). This study identified a significant association with variants in the *LOXL1* gene and, after performing a stratified analysis, determined the signal is restricted to pseudoexfoliation glaucoma cases only (Thorleifsson, Magnusson et al. 2007). Without properly reducing the clinical heterogeneity within the dataset (subtypes of glaucoma), an incorrect conclusion regarding the association of *LOXL1* and POAG would have been made. Despite removing a large portion (45%) of their cases, the statistical power was increased. Researchers must balance the need for increased samples sizes and reducing genetic heterogeneity to have the greatest statistical power and detect true genetic association signals.

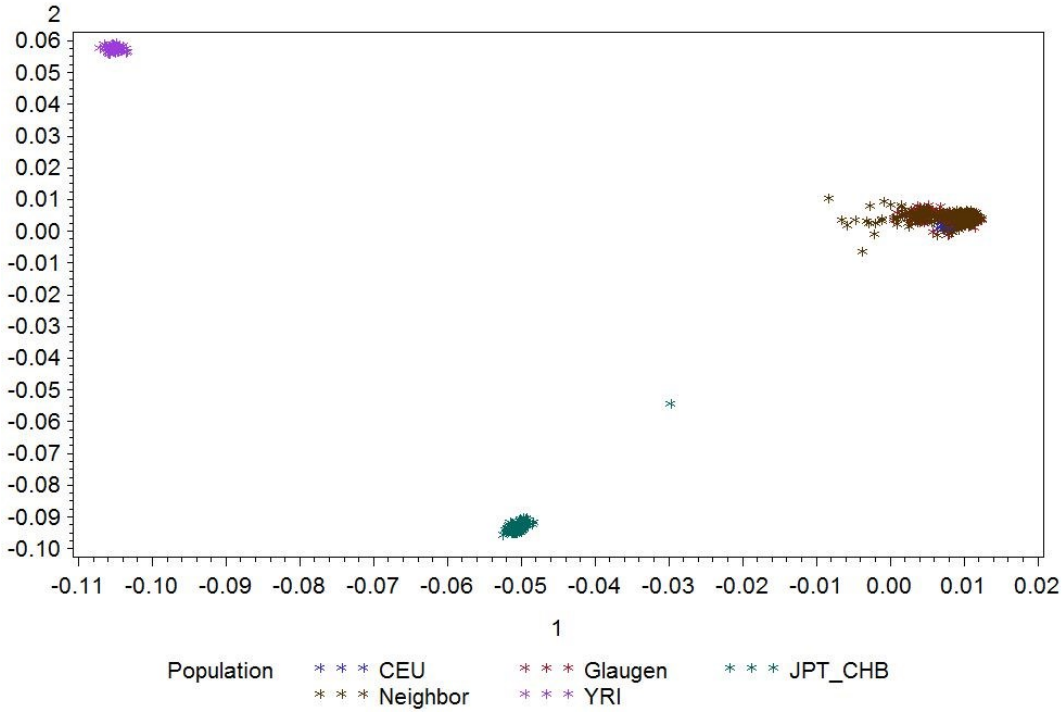
While the presence or absence of elevated intraocular pressure is the most commonly used clinical trait for performing stratified analyses in POAG, other traits

should be considered in the future. Although not currently feasible, intracranial pressure differences could potentially represent different subtypes of the disease. Due to the challenges of measuring intracranial pressure, it is not fully understood how clinically variable this trait is in the context of POAG, but it has been suggested that the effect of intraocular pressure is directly dependent on the opposing force generated by an individual's intracranial pressure (Fleischman and Allingham 2013; Janssen, Gorgels et al. 2013). In addition, there may be other clinical traits that could be used to define more homogeneous subsets. As more insight is gained about the biology, genetics, and environmental factors involved in POAG etiology, true subtypes of the disease can be defined.

5.3 Future directions summary

Glaucoma affects a large number of individuals in the United States and is a leading cause of blindness world-wide (Tielsch, Sommer et al. 1991; Resnikoff, Pascolini et al. 2004; Quigley and Broman 2006). While the work described here directly contributes to the field of POAG research, there is still much to be learned about the biology, prediction, prevention, and treatment of this disease. Care should be taken when collecting related clinical data as better phenotyping is needed for increasing statistical power in genetic analyses and our understanding of POAG biology. In future POAG research, a high priority should be given to the functional follow-up of association signals and to building clinically applicable prediction models.

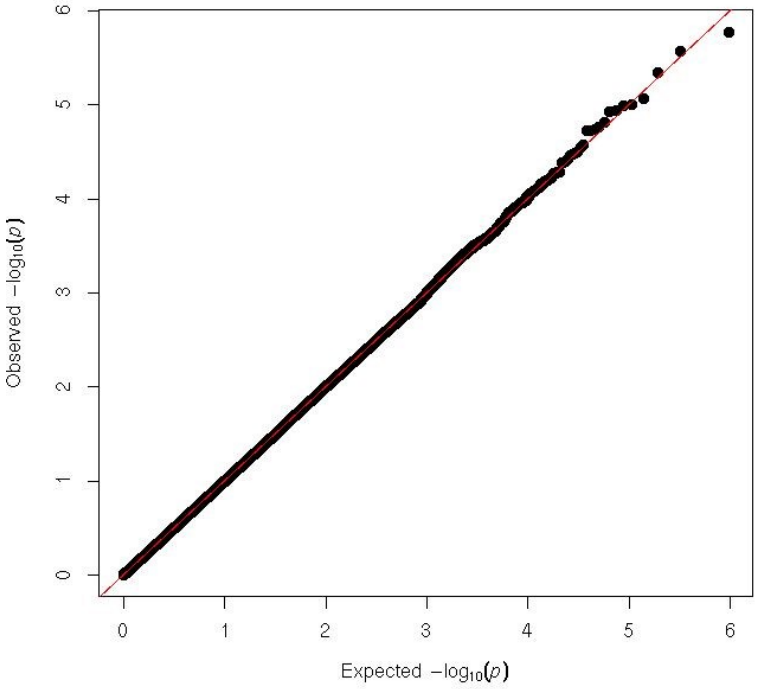
Appendix A



Principal components plot

Plot of the principal components analysis (PC1 vs PC2) resulting from the combined NEIGHBOR/GLAUGEN CCT datasets consisting of 1,117 individuals. Caucasian (CEU), Asian (JPT_CHB), and African (YRI) HapMap samples are included as a reference.

Appendix B



Quantile-quantile plot

Q-Q plot of p-values resulting from the CCT genome-wide association study conducted using a linear regression model that included sex, age, CCT measurement device, and POAG case/control status as covariates.

Appendix C

Primers used in Chapters 2 and 3

Chapter Reference	Amplicon	Forward Primer	Reverse Primer
2	*CNTNAP4*	5′ AGGATACTGCACTGGCAGGT 3′	5′ TGCTGATAAATGCGAACAGC 3′
2	*NTM*	5′ TGGCTTTGTGAGTGAAGACG 3′	5′ AGGCCACGCAAGTGTAGTTC 3′
2	*GAPDH*	5′ CGACCACTTTGTCAAGCTCA 3′	5′ AGGGGTCTACATGGCAACTG 3′
3	*SIX1* exon1	5′ TTGCAAAGCCTAAGGAGGAG 3′	5′ AGGACTTGGTGGCTGGTG 3′
3	*SIX1* exon2	5′ TTTGGGTTGGTGACAGATTG 3′	5′ GTCCACCATTCCTTTATGCG 3′
3	*SIX6* exon1	5′ ATAGTCCTGGCGTGCTGATT 3′	5′ CAGAACGCAGGGCTCTTAAC 3′
3	*SIX6* exon2	5′ TCCCAAAAGTGCACAACAAA 3′	5′ TTCCGAAGGAGACTTTGCAG 3′
3	*SIX6* enhancer1	5′ CGAGTGAACTGTGAAGATCTGTG 3′	5′ AGGTGAGAACGTTCACAGCCGA 3′
3	*SIX6* enhancer2	5′ GGGGCTCGAGCGAGTGAACTGTGAAG 3′	5′ GGGGAAGCTTAGGTGAGAACGTTCAC 3′
3	*six6a* morpholino	5′ CTGGAACATGGAGACTGTAATGTCT 3′	
3	*six6b* morpholino	5′ AATTGGCAACTGAAACATGAAGGCT 3′	

Appendix D

Coding variants identified by sequencing *SIX1* and *SIX6* in 518 POAG cases and controls

Gene	Coordinates	SNP ID	Function	Base Change	AA Change
SIX1	Chr14:61115933	–	5′ UTR	G>T	–
SIX1	Chr14:61115932	–	5′ UTR	C>T	–
SIX1	Chr14:61115506	rs151189392	Synonymous	C>T	Lys134Lys
SIX1	Chr14:61115323	rs200511291	Intronic	C>G	–
SIX1	Chr14:61115322	rs374507275	Intronic	G>C	–
SIX1	Chr14:61113320	–	Intronic	delT	–
SIX1	Chr14:61113316	rs183396626	Intronic	G>A	–
SIX1	Chr14:61113278	rs142301715	Missense	A>T	Asn193Ile
SIX1	Chr14:61113110	rs368974927	Missense	G>T	Pro249Leu
SIX6	Chr14:60976053	rs148591528	5' UTR	C>T	–
SIX6	Chr14:60976137	rs61746410	Synonymous	G>A	Leu7Leu
SIX6	Chr14:60976290	rs45549246	Synonymous	C>T	Ala58Ala
SIX6	Chr14:60976393	rs78954112	Missense	G>C	Glu93Gln
SIX6	Chr14:60976501	rs146737847	Missense	G>A	Glu129Lys
SIX6	Chr14:60976537	rs33912345	Missense	A>C	Asn141His
SIX6	Chr14:60977618	rs56098605	Intronic	G>A	–
SIX6	Chr14:60977756	–	Intronic	C>A	–
SIX6	Chr14:60977774	–	Intronic	T>C	–
SIX6	Chr14:60977843	rs45549246	Missense	T>G	Leu205Arg
SIX6	Chr14:60977864	rs202029915	Missense	C>T	Thr212Met
SIX6	Chr14:60977954	rs139302405	Missense	G>T	Ser242Ile
SIX6	Chr14:60977955	rs143366401	Synonymous	C>T	Ser242Ser
SIX6	Chr14:60978071	rs1061108	3′ UTR	G>C	–

Coordinates are based on the Hg19 reference.

Appendix E

Most significant SNPs from the imputed chromosome 14 POAG association analysis

BP	SNP	P	OR
Chr14:61091401	rs34935520	3.07E-10	1.27
Chr14:60976537	rs33912345	4.20E-10	1.27
Chr14:61095174	rs35155027	4.39E-10	1.27
Chr14:61072875	rs10483727	5.02E-10	1.26
Chr14:60811999	rs8015152	8.61E-10	1.27
Chr14:60847001	rs1254276	1.10E-09	1.26
Chr14:60957279	rs2093210	1.12E-09	1.26
Chr14:60813416	rs10151339	1.25E-09	1.27
Chr14:60886150	rs1272131	1.29E-09	1.26
Chr14:61008596	rs7159392	2.22E-09	1.25
Chr14:61013237	rs2351174	2.36E-09	1.25
Chr14:61021891	rs1955695	2.36E-09	1.25
Chr14:61025617	rs4442732	2.44E-09	1.25
Chr14:60848224	rs1313237	4.78E-09	1.25
Chr14:61006889	rs12883754	7.05E-09	1.26
Chr14:61007104	rs10146342	7.05E-09	1.26
Chr14:61012559	rs12589689	7.05E-09	1.26
Chr14:61005625	rs1010053	8.66E-09	1.25
Chr14:60789176	rs4901977	9.27E-09	1.25
Chr14:61025791	rs1555211	9.34E-09	1.25

Coordinates are based on the Hg19 reference.

Appendix F

Programs used in the ocular transcriptome analysis:

Fastqc: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Cutadapt: https://code.google.com/p/cutadapt/

TopHat: http://tophat.cbcb.umd.edu/index.shtml

Samtools: http://samtools.sourceforge.net/

Cufflinks: http://cufflinks.cbcb.umd.edu/index.html

RNAseqViewer: http://bioinfo.au.tsinghua.edu.cn/software/RNAseqViewer/

Appendix G

The number of genes expressed in the trabecular meshwork, cornea, and ciliary body ocular tissue samples

Counts shown are before filtering using two different FPKM thresholds.

Sample	Number of expressed genes FPKM > 0	Number of expressed genes FPKM ≥ 1
Adult TM	36,685	17,998
Fetal TM	32,284	16,648
Adult Cornea	35,494	16,684
Fetal Cornea	33,611	16,843
Adult CB	31,544	12,628
Fetal CB	27,523	13,784

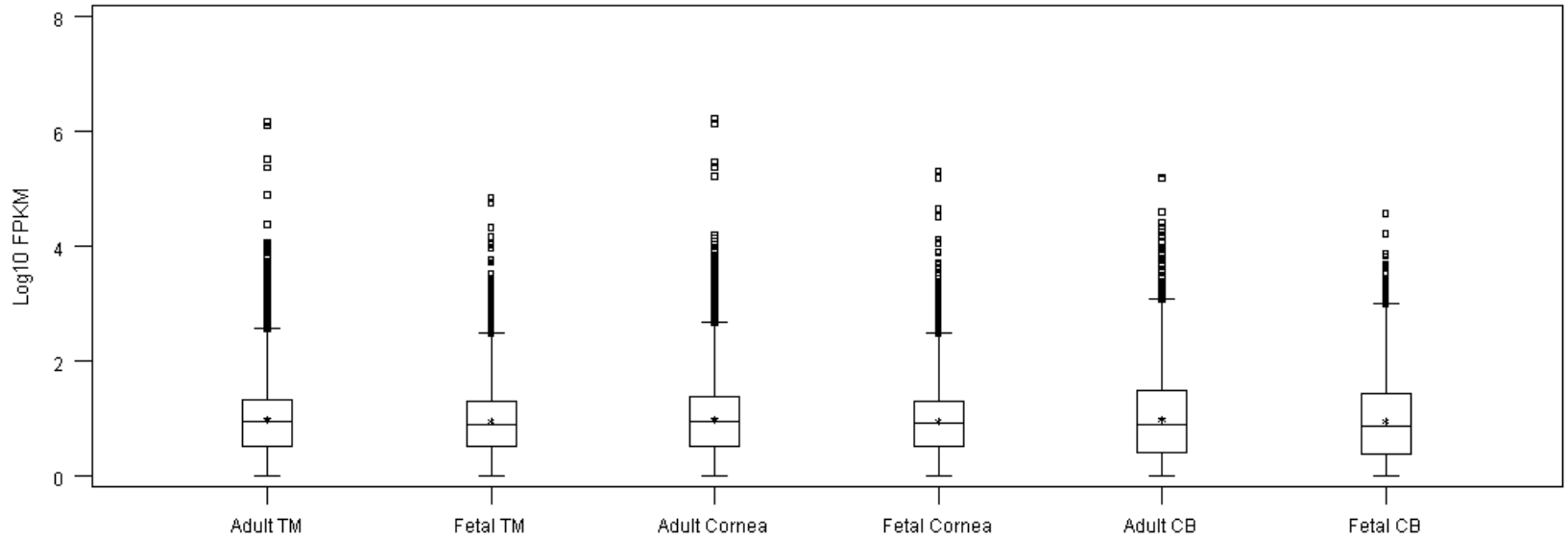
Appendix H

The average distribution of gene biotypes overall in six ocular tissue RNA-seq datasets

Explanation of gene biotype annotation is provided at:
http://www.gencodegenes.org/gencode_biotypes.html

Gene Biotype	Average Frequency	Gene Biotype	Average Frequency
3prime_overlapping_ncrna	5.67	polymorphic_pseudogene	7.50
antisense	546.50	processed_transcript	461.33
IG_C_gene	9.75	protein_coding	12,361.67
IG_C_pseudogene	1.75	pseudogene	1,336.17
IG_J_gene	5.00	rRNA	46.50
IG_J_pseudogene	1.00	rRNA_pseudogene	47.17
IG_V_gene	36.67	scRNA_pseudogene	39.67
IG_V_pseudogene	6.33	sense_intronic	19.17
lincRNA	273.67	sense_overlapping	28.17
miRNA	175.83	snoRNA	96.00
miRNA_pseudogene	1.17	snoRNA_pseudogene	3.00
misc_RNA	50.83	snRNA	99.50
Mt_rRNA	2.00	snRNA_pseudogene	1.60
Mt_tRNA	14.67	TR_C_gene	1.50
Mt_tRNA_pseudogene	17.33	TR_J_gene	2.20
ncrna_host	3.83	TR_V_gene	1.50
non_coding	11.33	TR_V_pseudogene	1.00
tRNA_pseudogene	4.80	None	514.98

Appendix I



Box plot depicting the distribution of gene expression values in six ocular tissue RNA-seq datasets

Gene expression values, measured in FPKM, were $\log_{10}$ transformed. The box shows the 25th, 50th (line), and 75th percentiles. Anything greater than 1.5 times the interquartile range is shown as a square above the plot, indicating extreme values.

Appendix J

Top 20 most expressed genes overall from trabecular meshwork, cornea, and ciliary body ocular tissue samples after removal of selected gene biotypes

There is an enrichment of deprecated genes, as annotated by Ensemble (GRCH37).

Sample	Ensemble ID	Gene Name	LOCUS	FPKM
Adult Cornea	ENSG00000251948	AC092279.1	19:24184074-24184165	1,659,680
Adult TM	ENSG00000251948	AC092279.1	19:24184074-24184165	1,514,330
Adult Cornea	ENSG00000252229	AC098691.1	1:91852861-91852949	1,403,690
Adult TM	ENSG00000252229	AC098691.1	1:91852861-91852949	1,277,510
Adult TM	ENSG00000252318	AC097532.1	2:133038646-133038738	337,063
Adult Cornea	ENSG00000252318	AC097532.1	2:133038646-133038738	289,844
Adult Cornea	ENSG00000252197	AC091047.1	8:70602343-70602417	243,812
Adult TM	ENSG00000252197	AC091047.1	8:70602343-70602417	236,462
Fetal Cornea	ENSG00000252229	AC098691.1	1:91852861-91852949	207,419
Adult Cornea	ENSG00000252248	AC093693.1	7:68527370-68527457	168,987
Fetal Cornea	ENSG00000251948	AC092279.1	19:24184074-24184165	161,281
Adult CB	ENSG00000251948	AC092279.1	19:24184074-24184165	161,006
Adult CB	ENSG00000252229	AC098691.1	1:91852861-91852949	153,737
Adult TM	ENSG00000252248	AC093693.1	7:68527370-68527457	79,925
Fetal TM	ENSG00000252229	AC098691.1	1:91852861-91852949	70,744
Fetal TM	ENSG00000251948	AC092279.1	19:24184074-24184165	58,878
Fetal Cornea	ENSG00000252318	AC097532.1	2:133038646-133038738	46,195
Adult CB	ENSG00000228253	MT-ATP8	MT:8365-8572	40,456
Fetal CB	ENSG00000251948	AC092279.1	19:24184074-24184165	38,281

References

Abecasis, G. R., A. Auton, et al. (2012). "An integrated map of genetic variation from 1,092 human genomes." Nature **491**(7422): 56-65.

Abu, A., M. Frydman, et al. (2008). "Deleterious mutations in the Zinc-Finger 469 gene cause brittle cornea syndrome." Am J Hum Genet **82**(5): 1217-1222.

Aijaz, S., J. Allen, et al. (2005). "Expression analysis of SIX3 and SIX6 in human tissues reveals differences in expression and a novel correlation between the expression of SIX3 and the genes encoding isocitrate dehyhrogenase and cadherin 18." Genomics **86**(1): 86-99.

Allingham, R. R., Y. Liu, et al. (2009). "The genetics of primary open-angle glaucoma: a review." Exp Eye Res **88**(4): 837-844.

Allingham, R. R. and M. B. Shields (2011). Shields' textbook of glaucoma. Philadelphia, Wolters Kluwer/Lippincott Williams & Wilkins Health.

Barrett, J. C. (2009). "Haploview: Visualization and analysis of SNP genotype data." Cold Spring Harb Protoc **2009**(10): pdb ip71.

Barrett, J. C., B. Fry, et al. (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-265.

Bath, C., D. Muttuvelu, et al. (2013). "Correction: Transcriptional Dissection of Human Limbal Niche Compartments by Massive Parallel Sequencing." PLoS One **8**(11).

Bernier, G., F. Panitz, et al. (2000). "Expanded retina territory by midbrain transformation upon overexpression of Six6 (Optx2) in Xenopus embryos." Mech Dev **93**(1-2): 59-69.

Boland, M. V., A. M. Ervin, et al. (2013). "Comparative effectiveness of treatments for open-angle glaucoma: a systematic review for the U.S. Preventive Services Task Force." Ann Intern Med **158**(4): 271-279.

Budenz, D. L., J. R. Bandi, et al. (2012). "Blindness and Visual Impairment in an Urban West African Population: The Tema Eye Survey." Ophthalmology.

Burdon, K. P. (2012). "Genome-wide association studies in the hunt for genes causing primary open-angle glaucoma: a review." Clin Experiment Ophthalmol **40**(4): 358-363.

Burdon, K. P., S. Macgregor, et al. (2011). "Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1." Nat Genet **43**(6): 574-578.

Calkins, D. J. (2013). "Age-related changes in the visual pathways: blame it on the axon." Invest Ophthalmol Vis Sci **54**(14): ORSF37-41.

Charlesworth, J., P. L. Kramer, et al. (2010). "The path to open-angle glaucoma gene discovery: endophenotypic status of intraocular pressure, cup-to-disc ratio, and central corneal thickness." Invest Ophthalmol Vis Sci **51**(7): 3509-3514.

Cheyette, B. N., P. J. Green, et al. (1994). "The Drosophila sine oculis locus encodes a homeodomain-containing protein required for the development of the entire visual system." Neuron **12**(5): 977-996.

Coleman, A. L. and S. Miglior (2008). "Risk factors for glaucoma onset and progression." Surv Ophthalmol **53 Suppl1**: S3-10.

Collaborative Normal-Tension Glaucoma Study Group (1998). "Comparison of glaucomatous progression between untreated patients with normal-tension glaucoma and patients with therapeutically reduced intraocular pressures. Collaborative Normal-Tension Glaucoma Study Group." Am J Ophthalmol **126**(4): 487-497.

Conte, I., R. Marco-Ferreres, et al. (2010). "Proper differentiation of photoreceptors and amacrine cells depends on a regulatory loop between NeuroD and Six6." Development **137**(14): 2307-2317.

Cornes, B. K., C. C. Khor, et al. (2011). "Identification of four novel variants that influence central corneal thickness in multi-ethnic Asian populations." Hum Mol Genet.

Del Bene, F. and J. Wittbrodt (2005). "Cell cycle control by homeobox genes in development and disease." Semin Cell Dev Biol **16**(3): 449-460.

Desronvil, T., D. Logan-Wyatt, et al. (2010). "Distribution of COL8A2 and COL8A1 gene variants in Caucasian primary open angle glaucoma patients with thin central corneal thickness." Mol Vis **16**: 2185-2191.

Dielemans, I., J. R. Vingerling, et al. (1994). "The prevalence of primary open-angle glaucoma in a population-based study in The Netherlands. The Rotterdam Study." Ophthalmology **101**(11): 1851-1855.

Dimasi, D. P., K. P. Burdon, et al. (2010). "The genetics of central corneal thickness." Br J Ophthalmol **94**(8): 971-976.

Dimasi, D. P., K. P. Burdon, et al. (2012). "Genetic investigation into the endophenotypic status of central corneal thickness and optic disc parameters in relation to open-angle glaucoma." Am J Ophthalmol **154**(5): 833-842 e832.

Dimasi, D. P., J. Y. Chen, et al. (2010). "Novel quantitative trait loci for central corneal thickness identified by candidate gene analysis of osteogenesis imperfecta genes." Hum Genet **127**(1): 33-44.

Dorrie, J., V. Wellner, et al. (2006). "An improved method for RNA isolation and removal of melanin contamination from melanoma tissue: implications for tumor antigen detection and amplification." J Immunol Methods **313**(1-2): 119-128.

Fan, B. J., D. Y. Wang, et al. (2011). "Genetic variants associated with optic nerve vertical cup-to-disc ratio are risk factors for primary open angle glaucoma in a US Caucasian population." Invest Ophthalmol Vis Sci **52**(3): 1788-1792.

Fleischman, D. and R. R. Allingham (2013). "The role of cerebrospinal fluid pressure in glaucoma and other ophthalmic diseases: A review." Saudi J Ophthalmol **27**(2): 97-106.

Flint, J. and M. R. Munafo (2007). "The endophenotype concept in psychiatric genetics." Psychol Med **37**(2): 163-180.

Foets, B. J., J. J. van den Oord, et al. (1992). "In situ immunohistochemical analysis of cell adhesion molecules on human corneal endothelial cells." Br J Ophthalmol **76**(4): 205-209.

Gallardo, M. E., J. Lopez-Rios, et al. (1999). "Genomic cloning and characterization of the human homeobox gene SIX6 reveals a cluster of SIX genes in chromosome 14 and associates SIX6 hemizygosity with bilateral anophthalmia and pituitary anomalies." Genomics **61**(1): 82-91.

Gauderman WJ, M. J. (2006). "QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies." http://hydra.usc.edu/gxe.

Gemenetzi, M., Y. Yang, et al. (2012). "Current concepts on primary open-angle glaucoma genetics: a contribution to disease pathophysiology and future treatment." Eye (Lond) **26**(3): 355-369.

Gibson, J., H. Griffiths, et al. (2012). "Genome-wide association study of primary open angle glaucoma risk and quantitative traits." Mol Vis **18**: 1083-1092.

Gil, O. D., G. Zanazzi, et al. (1998). "Neurotrimin mediates bifunctional effects on neurite outgrowth via homophilic and heterophilic interactions." J Neurosci **18**(22): 9312-9325.

Girkin, C. A. (2008). "Differences in optic nerve structure between individuals of predominantly African and European ancestry: Implications for disease detection and pathogenesis." Clin Ophthalmol **2**(1): 65-69.

Girkin, C. A., P. A. Sample, et al. (2010). "African Descent and Glaucoma Evaluation Study (ADAGES): II. Ancestry differences in optic disc, retinal nerve fiber layer, and macular structure in healthy subjects." Arch Ophthalmol **128**(5): 541-550.

Gordon, M. O., J. A. Beiser, et al. (2002). "The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma." Arch Ophthalmol **120**(6): 714-720; discussion 829-730.

Gordon, M. O., V. Torri, et al. (2007). "Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension." Ophthalmology **114**(1): 10-19.

Heijl, A., M. C. Leske, et al. (2002). "Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial." Arch Ophthalmol **120**(10): 1268-1279.

Iglesias, A. I., H. Springelkamp, et al. (2013). "Exome sequencing and functional analyses suggest SIX6 is a gene involved in an altered proliferation-differentiation balance early in life and optic nerve degeneration at old age." Hum Mol Genet.

Iwase, A., Y. Suzuki, et al. (2004). "The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study." Ophthalmology **111**(9): 1641-1648.

Janssen, S. F., T. G. Gorgels, et al. (2012). "Gene expression and functional annotation of the human ciliary body epithelia." PLoS One **7**(9): e44973.

Janssen, S. F., T. G. Gorgels, et al. (2013). "The vast complexity of primary open angle glaucoma: disease genes, risks, molecular mechanisms and pathobiology." Prog Retin Eye Res **37**: 31-67.

Kaushik, S., S. S. Pandav, et al. (2012). "Relationship Between Corneal Biomechanical Properties, Central Corneal Thickness, and Intraocular Pressure Across the Spectrum of Glaucoma." Am J Ophthalmol.

Kawakami, K., S. Sato, et al. (2000). "Six family genes--structure and function as transcription factors and their roles in development." Bioessays **22**(7): 616-626.

Kendler, K. S. and M. C. Neale (2010). "Endophenotype: a conceptual analysis." Mol Psychiatry **15**(8): 789-797.

Kenyon, K. L., D. Yang-Zhou, et al. (2005). "Partner specificity is essential for proper function of the SIX-type homeodomain proteins Sine oculis and Optix during fly eye development." Dev Biol **286**(1): 158-168.

Kniestedt, C., S. Lin, et al. (2006). "Correlation between intraocular pressure, central corneal thickness, stage of glaucoma, and demographic patient data: prospective analysis of biophysical parameters in tertiary glaucoma practice populations." J Glaucoma **15**(2): 91-97.

Knight, O. J., C. A. Girkin, et al. (2012). "Effect of race, age, and axial length on optic nerve head parameters and retinal nerve fiber layer thickness measured by Cirrus HD-OCT." Arch Ophthalmol **130**(3): 312-318.

Krumbiegel, M., F. Pasutto, et al. (2011). "Genome-wide association study with DNA pooling identifies variants at CNTNAP2 associated with pseudoexfoliation syndrome." Eur J Hum Genet **19**(2): 186-193.

Kuehn, M. H., J. H. Fingert, et al. (2005). "Retinal ganglion cell death in glaucoma: mechanisms and neuroprotective strategies." Ophthalmol Clin North Am **18**(3): 383-395, vi.

Kulkarni, S. V., K. F. Damji, et al. (2008). "Medical management of primary open-angle glaucoma: Best practices associated with enhanced patient compliance and persistency." Patient Prefer Adherence **2**: 303-314.

Kumar, J. P. (2009). "The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease." Cell Mol Life Sci **66**(4): 565-583.

Kwon, Y. H., J. H. Fingert, et al. (2009). "Primary open-angle glaucoma." N Engl J Med **360**(11): 1113-1124.

Leske, M. C., A. Heijl, et al. (2007). "Predictors of long-term progression in the early manifest glaucoma trial." Ophthalmology **114**(11): 1965-1972.

Li, H., B. Handsaker, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, X., V. Perissi, et al. (2002). "Tissue-specific regulation of retinal and pituitary precursor cell proliferation." Science **297**(5584): 1180-1183.

Liton, P. B., C. Luna, et al. (2006). "Genome-wide expression profile of human trabecular meshwork cultured cells, nonglaucomatous and primary open angle glaucoma tissue." Mol Vis **12**: 774-790.

Liu, Y. and R. R. Allingham (2011). "Molecular genetics in glaucoma." Exp Eye Res **93**(4): 331-339.

Liu, Y., R. R. Allingham, et al. (2013). "Gene expression profile in human trabecular meshwork from patients with primary open-angle glaucoma." Invest Ophthalmol Vis Sci **54**(9): 6382-6389.

Liu, Y., M. A. Hauser, et al. (2013). "Investigation of known genetic risk factors for primary open angle glaucoma in two populations of African ancestry." Invest Ophthalmol Vis Sci **54**(9): 6248-6254.

Lodge, A. P., C. J. McNamee, et al. (2001). "Identification and characterization of CEPU-Se-A secreted isoform of the IgLON family protein, CEPU-1." Mol Cell Neurosci **17**(4): 746-760.

Lu, Y., D. P. Dimasi, et al. (2010). "Common genetic variants near the Brittle Cornea Syndrome locus ZNF469 influence the blinding disease risk factor central corneal thickness." PLoS Genet **6**(5): e1000947.

Lu, Y., V. Vitart, et al. (2013). "Genome-wide association analyses identify multiple loci associated with central corneal thickness and keratoconus." Nat Genet **45**(2): 155-163.

Macgregor, S., A. W. Hewitt, et al. (2010). "Genome-wide association identifies ATOH7 as a major gene determining human optic disc size." Hum Mol Genet **19**(13): 2716-2724.

MacKinnon, D. P., C. M. Lockwood, et al. (2002). "A comparison of methods to test mediation and other intervening variable effects." Psychol Methods **7**(1): 83-104.

Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads." EMBnet.journal.

Meguro, A., H. Inoko, et al. (2010). "Genome-wide association study of normal tension glaucoma: common variants in SRBD1 and ELOVL5 contribute to disease susceptibility." Ophthalmology **117**(7): 1331-1338 e1335.

Miglior, S., N. Pfeiffer, et al. (2007). "Predictive factors for open-angle glaucoma among patients with ocular hypertension in the European Glaucoma Prevention Study." Ophthalmology **114**(1): 3-9.

Monsees, G. M., R. M. Tamimi, et al. (2009). "Genome-wide association scans for secondary traits using case-control samples." Genet Epidemiol **33**(8): 717-728.

Mudumbai, R. C. (2013). "Clinical update on normal tension glaucoma." Semin Ophthalmol **28**(3): 173-179.

Nakano, M., Y. Ikeda, et al. (2009). "Three susceptible loci associated with primary open-angle glaucoma identified by genome-wide association study in a Japanese population." Proc Natl Acad Sci U S A **106**(31): 12838-12842.

Osman, W., S. K. Low, et al. (2012). "A genome-wide association study in the Japanese population confirms 9p21 and 14q23 as susceptibility loci for primary open angle glaucoma." Hum Mol Genet **21**(12): 2836-2842.

Ozel, A. B., S. E. Moroi, et al. (2014). "Genome-wide association study and meta-analysis of intraocular pressure." Hum Genet **133**(1): 41-57.

Patsopoulos, N. A., A. Tatsioni, et al. (2007). "Claims of sex differences: an empirical assessment in genetic associations." JAMA **298**(8): 880-893.

Price, A. L., N. J. Patterson, et al. (2006). "Principal components analysis corrects for stratification in genome-wide association studies." Nat Genet **38**(8): 904-909.

Pruim, R. J., R. P. Welch, et al. (2010). "LocusZoom: regional visualization of genome-wide association scan results." Bioinformatics **26**(18): 2336-2337.

Purcell, S., B. Neale, et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.

Purcell, S. M., N. R. Wray, et al. (2009). "Common polygenic variation contributes to risk of schizophrenia and bipolar disorder." Nature **460**(7256): 748-752.

Quigley, H. A. and A. T. Broman (2006). "The number of people with glaucoma worldwide in 2010 and 2020." Br J Ophthalmol **90**(3): 262-267.

Racette, L., M. R. Wilson, et al. (2003). "Primary open-angle glaucoma in blacks: a review." Surv Ophthalmol **48**(3): 295-313.

Ramdas, W. D., L. M. van Koolwijk, et al. (2010). "A genome-wide association study of optic disc parameters." PLoS Genet **6**(6): e1000978.

Ramdas, W. D., L. M. van Koolwijk, et al. (2011). "Common genetic variants associated with open-angle glaucoma." Hum Mol Genet **20**(12): 2464-2471.

Ramdas, W. D., L. M. van Koolwijk, et al. (2011). "Common genetic variants associated with open-angle glaucoma." Hum Mol Genet **20**(12): 2464-2471.

Ray, K. and S. Mookherjee (2009). "Molecular complexity of primary open angle glaucoma: current concepts." J Genet **88**(4): 451-467.

Resnikoff, S., D. Pascolini, et al. (2004). "Global data on visual impairment in the year 2002." Bull World Health Organ **82**(11): 844-851.

Roge, X. and X. Zhang (2013). "RNAseqViewer: visualization tool for RNA-Seq data." Bioinformatics.

Rozen, S. and H. Skaletsky (2000). "Primer3 on the WWW for general users and for biologist programmers." Methods Mol Biol **132**: 365-386.

Rudnicka, A. R., S. Mt-Isa, et al. (2006). "Variations in primary open-angle glaucoma prevalence by age, gender, and race: a Bayesian meta-analysis." Invest Ophthalmol Vis Sci **47**(10): 4254-4261.

Sambrook, J. and D. W. Russell (2006). "Rapid Amplification of 5' cDNA Ends (5'-RACE)." CSH Protoc **2006**(1).

SAS Institute Inc. (2002-2008). "SAS 9.2 Software." Cary, NC: SAS Institute Inc.

Satyamoorthy, K., G. Li, et al. (2002). "A versatile method for the removal of melanin from ribonucleic acids in melanocytic cells." Melanoma Res **12**(5): 449-452.

Segev, F., E. Heon, et al. (2006). "Structural abnormalities of the cornea and lid resulting from collagen V mutations." Invest Ophthalmol Vis Sci **47**(2): 565-573.

Sommer, A., J. M. Tielsch, et al. (1991). "Relationship between intraocular pressure and primary open angle glaucoma among white and black Americans. The Baltimore Eye Survey." Arch Ophthalmol **109**(8): 1090-1095.

Spiegel, I., D. Salomon, et al. (2002). "Caspr3 and caspr4, two novel members of the caspr family are expressed in the nervous system and interact with PDZ domains." Mol Cell Neurosci **20**(2): 283-297.

Struyk, A. F., P. D. Canoll, et al. (1995). "Cloning of neurotrimin defines a new subfamily of differentially expressed neural cell adhesion molecules." J Neurosci **15**(3 Pt 2): 2141-2156.

Sullivan-Mee, M., K. D. Halverson, et al. (2006). "Central corneal thickness and normal tension glaucoma: a cross-sectional study." Optometry **77**(3): 134-140.

Thorleifsson, G., K. P. Magnusson, et al. (2007). "Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma." Science **317**(5843): 1397-1400.

Thorleifsson, G., G. B. Walters, et al. (2010). "Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma." Nat Genet **42**(10): 906-909.

Tielsch, J. M., J. Katz, et al. (1994). "Family history and risk of primary open angle glaucoma. The Baltimore Eye Survey." Arch Ophthalmol **112**(1): 69-73.

Tielsch, J. M., A. Sommer, et al. (1991). "Racial variations in the prevalence of primary open-angle glaucoma. The Baltimore Eye Survey." JAMA **266**(3): 369-374.

Trapnell, C., L. Pachter, et al. (2009). "TopHat: discovering splice junctions with RNA-Seq." Bioinformatics **25**(9): 1105-1111.

Trapnell, C., B. A. Williams, et al. (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotechnol **28**(5): 511-515.

Turner, H. C., M. T. Budak, et al. (2007). "Comparative analysis of human conjunctival and corneal epithelial gene expression with oligonucleotide microarrays." Invest Ophthalmol Vis Sci **48**(5): 2050-2061.

Tyynismaa, H., P. Sistonen, et al. (2002). "A locus for autosomal dominant keratoconus: linkage to 16q22.3-q23.1 in Finnish families." Invest Ophthalmol Vis Sci **43**(10): 3160-3164.

Ulmer, M., J. Li, et al. (2012). "Genome-wide analysis of central corneal thickness in primary open-angle glaucoma cases in the NEIGHBOR and GLAUGEN consortia." Invest Ophthalmol Vis Sci **53**(8): 4468-4474.

van Koolwijk, L. M., W. D. Ramdas, et al. (2012). "Common genetic determinants of intraocular pressure and primary open-angle glaucoma." PLoS Genet **8**(5): e1002611.

Varma, R., M. Ying-Lai, et al. (2004). "Prevalence of open-angle glaucoma and ocular hypertension in Latinos: the Los Angeles Latino Eye Study." Ophthalmology **111**(8): 1439-1448.

Vitart, V., G. Bencic, et al. (2010). "New loci associated with central cornea thickness include COL5A1, AKAP13 and AVGR8." Hum Mol Genet **19**(21): 4304-4311.

Vithana, E. N., T. Aung, et al. (2011). "Collagen-related genes influence the glaucoma risk factor, central corneal thickness." Hum Mol Genet **20**(4): 649-658.

Wang, Z., M. Gerstein, et al. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.

Weih, L. M., M. Nanjan, et al. (2001). "Prevalence and predictors of open-angle glaucoma: results from the visual impairment project." Ophthalmology **108**(11): 1966-1972.

Wiggs, J. L., M. A. Hauser, et al. (2012). "The NEIGHBOR Consortium Primary Open-Angle Glaucoma Genome-wide Association Study: Rationale, Study Design, and Clinical Variables." J Glaucoma.

Wiggs, J. L., B. L. Yaspan, et al. (2012). "Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma." PLoS Genet **8**(4): e1002654.

Wilensky, J. T. and T. C. Chen (1996). "Long-term results of trabeculectomy in eyes that were initially successful." Trans Am Ophthalmol Soc **94**: 147-159; discussion 160-144.

Wolfs, R. C., C. C. Klaver, et al. (1998). "Genetic risk of primary open-angle glaucoma. Population-based familial aggregation study." Arch Ophthalmol **116**(12): 1640-1645.

Wu, J., A. W. Hewitt, et al. (2006). "Disease severity of familial glaucoma compared with sporadic glaucoma." Arch Ophthalmol **124**(7): 950-954.

Xiang, M., L. Zhou, et al. (1993). "Brn-3b: a POU domain gene expressed in a subset of retinal ganglion cells." Neuron **11**(4): 689-701.

Young, T. L., F. Hawthorne, et al. (2013). "Whole genome expression profiling of normal human fetal and adult ocular tissues." Exp Eye Res **116**: 265-278.

Zuber, M. E., M. Perron, et al. (1999). "Giant eyes in Xenopus laevis by overexpression of XOptx2." Cell **98**(3): 341-352.

Biography

Megan Rebecca Ulmer Carnes was born in Stuart, Florida on July 25th, 1984. She received her Bachelors of Science degree in Biological Sciences and a minor in Genetics from North Carolina State University in 2006. She worked at the North Carolina State University Genomic Sciences Laboratory until she joined the Duke University Program in Genetics and Genomics in 2009. Her thesis work was conducted in the laboratory of Dr. Michael Hauser at the Duke Center for Human Genetics.

Publications:

1) **M Ulmer Carnes**, et al. (2014). "Discovery and functional annotation of SIX6 variants in primary open-angle glaucoma." PLOS Genetics. *In print*.

2) A Ozel,…, **M Ulmer**, et al. (2014). Genome-wide association study and meta-analysis of intraocular pressure. Human Genetics. **133**:41-57.

3) Y Lu,…, **M Ulmer**, et al. (2013). Genome-wide association analyses identify multiple loci associated with central corneal thickness and keratoconus. Nature Genetics. 45(2):155-163.

4) **M Ulmer**, et al. (2012). Genome-wide analysis of central corneal thickness in primary open-angle glaucoma cases in the NEIGHBOR and GLAUGEN consortia. Invest Ophthalmol Vis Sci. **53**, 4468-74.

5) C Linnertz,…, **M Ulmer**, et al. (2012). "Characterization of the poly-T variant in the TOMM40 gene in diverse populations." PLoS One **7**, e30994.

6) J Cervantes-Flores,…, **M Ulmer**, et al. (2010). "Identification of quantitative trait loci for dry-matter, starch, and b-carotene content in sweet potato." Molecular Breeding. **28**(2):201-216.

Professional Memberships:

1) The American Society of Human Genetics member	2012, 2013
2) The Association for Research in Vision and Ophthalmology member	2012
3) European Journal of Ophthalmology manuscript reviewer	2012

Awards:

1) Duke Translational Medicine Institute voucher grant	2012
2) Duke University conference travel award	2012
3) National Eye Institute travel award	2012
4) UAB Statistical Genetics Short Course travel grant	2011
5) The Rockefeller University Advanced Gene Mapping Course travel award	2010