

**EXOME ARRAY ANALYSIS IDENTIFIES NOVEL
SUSCEPTIBILITY LOCI FOR INTRAOCULAR PRESSURE**

by

Fei Chen

A thesis submitted to Johns Hopkins University in conformity with
the requirements for the degree of Master of Science

Baltimore, Maryland

April, 2014

ABSTRACT

Intraocular pressure (IOP) is an important physiologic characteristic in maintaining the structure and function of the eye. Elevated IOP is a major risk factor of primary-open angle glaucoma and the only target for current glaucoma therapy. The goal of this study is to identify rare and low frequency variants in regulation of IOP.

Using the exome array data, we performed a genome-wide association study for IOP in 1,661 unrelated individuals from a population-based cohort, the Beaver Dam Eye Study (BDES). IOP was adjusted for age, sex, systolic blood pressure, IOP treatment and population stratification. We conducted single-variant analysis on 35,091 low frequency and common variants and gene-based analysis on 9,650 gene regions with at least two protein-altering variants.

IOP was significantly associated with rs52809447, located in *GGA3* at 17q25 (MAF=1.11%, $p = 1.3 \times 10^{-6}$) and with rs757702, located in *OGDH* at 7p14-p13 (MAF = 1.60%, $p = 1.5 \times 10^{-6}$). Two novel loci, the *HAP1* (*huntingtin-associated protein 1*) at 17q21 ($p = 1.1 \times 10^{-6}$) and *TNR* (*tenascin R*) at 1q24 ($p = 3.0 \times 10^{-6}$) reached genome-wide significance in gene-based analysis. The inhibitory role of *TNR* gene in retinal ganglion cell survival and axonal regeneration suggests a potential involvement in the glaucomatous progression. The implications of *GGA3*, *OGDH* and *HAP1* with Alzheimer's disease and IOP, lend support to the shared pathways of neurodegeneration in the brain and in the eye.

Primary Reader: Dr. Priya Duggal

Secondary Reader: Dr. Alison Klein

ACKNOWLEDGEMENTS

My most sincere thanks go to my advisor and mentor, Dr. Priya Duggal. I thank her for her continuous guidance, encouragement and support during the development of this work. Without her supervision and constant help this thesis would not have been possible.

I would like to thank my secondary reader, Dr. Alison Klein, for her precious insights and comments.

I thank my fellow colleagues Meg Parker, Genevieve Wojcik, Man Li and Shan Andrews, for their generous help.

Last but not the least, to my parent Yangfu Chen and Xiaobing Zhao, thank you for encouraging me in all of my pursuits and inspiring me to follow my dreams. Special thanks go to my husband Yukon Zhang, for the unconditional love and support throughout the entire process.

TABLE OF CONTENTS

ABSTRACT.....	II
ACKNOWLEDGEMENTS	III
TABLE OF CONTENTS	IV
LIST OF FIGURES	VI
LIST OF TABLES	VII
INTRODUCTION	1
GENETIC COMPONENTS FOR IOP AND POAG.....	2
STUDY OBJECTIVES	3
STATISTICAL APPROACHES	4
MOTIVATIONS.....	4
SINGLE-VARIANT AND MULTIPLE-VARIANT APPROACH	4
COLLAPSING/SUM METHODS.....	5
CMC METHOD AND VARIABLE-THRESHOLD APPROACH	7
WEIGHTED SUM TEST	8
C-ALPHA TEST	9
SKAT AND OPTIMAL SKAT	10
MATERIALS AND METHODS	13
STUDY POPULATION	13
CLINICAL EVALUATION	14
DNA EXTRACTION AND GENOTYPING.....	14
QUALITY CONTROL PROCEDURES	15
<i>Quality Control for Samples</i>	15

<i>Quality Control for Markers</i>	16
<i>Assessment of Population Stratification</i>	17
STATISTICAL ANALYSIS	17
<i>Quantitative Trait Adjusted for Covariates</i>	17
<i>Annotations</i>	17
<i>Single-Variant Analysis</i>	18
<i>Gene-Based Analysis</i>	19
<i>Replication of Published GWAS Loci</i>	20
RESULTS	21
STUDY PARTICIPANTS	21
SINGLE-VARIANT ASSOCIATIONS	21
<i>GGA3 Loci</i>	22
<i>OGDH Loci</i>	23
GENE-BASED ASSOCIATIONS	24
<i>HAP1 Region</i>	25
<i>TNR Region</i>	26
REPLICATED ASSOCIATIONS OF PUBLISHED GWAS LOCI	27
DISCUSSION	29
FIGURES AND TABLES	32
BIBLIOGRAPHY	45
CURRICULUM VITA	55

LIST OF FIGURES

Figure 1	Quality Control Procedures for Exome Array Data	32
Figure 2	Probability of IBD Sharing for BDES Samples (N = 1,908)	33
Figure 3	Principal Component Analysis for BDES Samples and HapMap Controls	34
Figure 4	Distributions of Baseline IOP in Analyzed Cohort	35
Figure 5	Manhattan Plot for Single-Variant Associations	36
Figure 6	QQ Plot of Single-Variant Associations	37

LIST OF TABLES

Table 1	Review of Statistical Approaches for Rare Variant Associations	38
Table 2	Characteristics of 1,661 Beaver Dam Eye Study Participants in Analyzed Cohort	40
Table 3	Single-Variant Associations Reached Genome-Wide Significance of $p < 1.5 \times 10^{-6}$	41
Table 4	Gene-Based Associations Reached Genome-Wide Significance of $p < 5.2 \times 10^{-6}$	42
Table 5	Coverage of the GWAS Loci for IOP and POAG in our Exome Array Data	43
Table 6	Replication of GWAS Locus rs4236601 in <i>CAVI/CAV2</i> Region for IOP and POAG	44

INTRODUCTION

Glaucoma is a leading cause of visual impairment and irreversible blindness globally ¹. As the population ages, the number of people with glaucoma is increasing rapidly. It is estimated that by the year of 2020, there will be around 80 million people with glaucoma worldwide ¹. Primary open-angle glaucoma (POAG) is the most common form of glaucoma, accounting for over 70% of all glaucoma cases ². Except for juvenile-onset POAG, which follows a Mendelian inheritance, adult-onset POAG is considered a complex trait with a multifactorial etiology involves both genetic and environmental risk factors.

Intraocular pressure (IOP) is an important physiologic characteristic in maintaining the structure and function of the eye ³. As one of the three ophthalmologic vital signs, the level of IOP provides essential information in the evaluation of ocular health. Normal IOP is regulated by the balance between the production and the drainage of aqueous humor through the trabecular meshwork in the anterior chamber angle of human eyes. Clinically, IOP ranging from 10 mmHg and 21 mmHg is considered normal ⁴.

Elevated IOP is a major risk factor of glaucomatous optic neuropathy, and the only target for current glaucoma therapy. A one mmHg increase in baseline IOP above the population average was associated with 10-14% greater risk of developing POAG in the following 5-9 years ^{5,6,7}. Results from clinical trials have repeatedly showed that IOP lowering treatments in eyes with ocular hypertension are effective both in preventing the glaucoma onset and delaying its progression, and that eyes with normal tension glaucoma

also suffer less progression of field loss ^{8,9}. Elevated IOP contributes to the accelerated death of retinal ganglion cells (RGCs) and the damage of the optic nerve, and ultimately results in progressive visual field loss and eventually blindness in glaucoma patients ¹⁰.

Genetic Components for IOP and POAG

Intraocular pressure is a heritable polygenetic trait with environmental influences ¹¹. Heritability for IOP is consistently estimated in the range of 0.30 – 0.42 in population-based studies of European ancestry ^{11,12,13}. Based on the familial correlations in the Beaver Dam Eye Study (BDES) population, Duggal et al. ¹¹ and Klein et al. ¹³ previously reported 30% - 36% of the total variation of IOP explained by the additive genetic effects.

Family-based linkage studies in populations of European ancestry have located two regions on chromosome 10q22 ¹⁴ and chromosome 19p ¹⁵ with significant linkage to IOP. Other linkage studies across different ethnicity groups also reported a number of suggestive linkage regions on chromosome 5, 6, 12, 13, and 14, including a region that harbored *WDR36*, a susceptibility locus for POAG ^{16,17,18}. However, no causal gene in these regions has yet been identified for IOP.

Genome-wide association studies (GWAS) have revealed evidence of multiple loci associated with IOP. Novel associations in *GAS7* region on chromosome 17p13 and *TMC6I* region on chromosome 1q24 were first discovered in a large GWAS from Netherlands ¹⁹, and both associations were replicated in a meta-analysis of three US population-based cohorts ²⁰. The associations for IOP in *TMC6I* region overlap with those for advanced POAG ²¹, suggesting that loci in this region can produce IOP changes that lead to glaucomatous optic neuropathy later in life. Other significant genetic

associations with IOP include the *GLCCII/ICAI* region on 7p21²², and the *MVB12B* region on 9q33²³. However, neither of these recently identified loci has been replicated.

GWAS have yielded a number of loci associated with POAG. Significant association of common variants between *CAVI* and *CAV2* genes on chromosome 7q31 were discovered²⁴ and replicated in independent populations of European descent²⁵. Four GWAS across different ethnicity groups each reported significant associations between POAG and the *CDKN2BAS* region on 9p21^{26,27,28,29}, and two of the GWAS identified the *SIX1/SIX6* region on 14q23 as a susceptibility locus for POAG^{26,29}.

Study Objectives

Genomics studies of IOP may elucidate key biological pathways and improve our understanding of the pathogenesis of disease.

Traditional genetic approaches including family based linkage analysis and genome wide association studies may localize a gene region or gene, but do not have the resolution to identify causal alleles. Next generation sequencing provides an opportunity to investigate associations with rare and low frequency variants but can be costly for large studies. The exome array is a cost-effective option for the identification of rare and low frequency variants that may contribute to complex diseases.

The primary goal of this study is to identify novel variants associated with IOP and to replicate previously reported genetic variants in the following gene regions *CAVI/CAV2*, *CDKN2BAS*, *SIX1/SIX6*, *TMC01*, *GAS7*, *GLCCII/ICAI* and *MVB12B*.

STATISTICAL APPROACHES

Motivations

Genome-wide association studies have identified hundreds of genetic variants that are reproducibly associated with complex human diseases and traits, and have provided valuable insights into the complexities of their genetic architecture. Although a substantial proportion of individual differences in disease susceptibility is known due to genetic factors, most common variants identified confer relatively small increments in risk, and explain only a small portion of overall heritability³⁰. Both evolutionary studies and empirical data suggest that for complex disease, extreme allelic heterogeneity is caused collectively by multiple rare and low frequency variants with moderate to high penetrances³¹, thus highlighting the importance of identifying less common variants.

While in principle sequencing studies can reveal the role of rare and common variants in complex diseases and traits, there are analytic challenges. Table 1 outlines the features and limitations of statistical approaches in detection of rare variant associations.

Single-Variant and Multiple-Variant Approach

The conventional single-variant approach assesses individual variants for association with a disease or trait, by performing a standard univariate statistical test, such as Pearson χ^2 test, Fisher's exact test, Cochran-Armitage test for trend, or a likelihood ratio test (LRT) for regression analysis. When applied to rare variants, the power of the single-variant approach is limited due to the small number of observations

for any given variant and a more stringent multiple-test corrections as compared to common variants^{32,33}.

Given that many different mutations collectively contribute to a complex disease or trait, it may be beneficial to focus on the group rather than on each variant individually. An alternative to single-variant approach is to employ a multiple-variant analysis, which tests a number of variants in a defined functional unit (e.g. gene, pathway) simultaneously with the use of multivariate methods, such as Fisher's product method, Hotelling's T^2 test, or LRT in linear/logistic regression³⁴. The multiple-variant approach usually involves a large number of degrees of freedom that results in substantial power reduction. As suggested by simulation data, the type I error rate of multiple-variant test is inflated when rare variants were analyzed in regression models³⁴. Although permutation can be used to control for type I error, its application in the whole genome scale can be computationally intensive.

Collapsing/Sum Methods

As an extension to the multiple-variant approach, burden tests assess the cumulative effects of multiple variants in a functional unit (e.g. gene, pathway) by collapsing or summarizing the rare variants within the region into a single variable, which is then tested for association with the disease or trait.

There are various ways to aggregate the information of multiple loci within a region. In the cohort allelic sums test (CAST), the count of minor alleles is summed over all variant sites in a region for cases and controls, and a statistical test is applied to determine if the difference in these sums is greater than would be expected by chance³⁵.

A standard univariate test (e.g. Pearson's χ^2 test or LRT in logistic regression) can then be applied to determine its statistical significance^{34,36}.

Similar sum methods can be generalized to test quantitative trait associations where the phenotype is modeled in a linear regression framework as a function of either the direct count of minor alleles over all variant sites^{33,37}, or the proportion of variant sites that a minor allele presents³⁶. The genotype can also be represented in a dichotomous indicator variable taking the value 1 if a minor allele is present at any of the variant sites in a region, or 0 otherwise. For each of these three sum methods, the likelihood ratio test is constructed to make statistical inference of the cumulative association of rare variants with the quantitative trait.

Independent simulation studies demonstrate that the collapsing/sum method performs considerably better than a single-variant and multiple-variant approach for qualitative traits across a variety of scenarios that differ in sample size, variant allele frequencies, proportion of causal variants in a region and total number of variants being collapsed^{34,37}. The results for quantitative traits are similar³⁶. Comparing the collapsing method based on dichotomous variable to the sum method on a count/proportion variable, studies found that the sum method always had equivalent or better power than the collapsing method across different scenarios^{36,38}.

In general, burden tests based on collapsing/sum methods can enrich the association signal and at the same time reduce the number of test's degrees of freedom, and hence improve its power for the detection of rare variants. However, this approach relies greatly on a key assumption that variants within the region have individual effects in the same direction and magnitude, and each contributes to the disease susceptibility

independently. In the presence of both protective and detrimental variants in the same region, the power of the burden test is diminished as these counteracted effects nullify the association signals. Moreover, misclassification of variants resulting in exclusion of causal variants and/or inclusion of non-causal variants could adversely affect the power^{34,37}.

CMC Method and Variable-Threshold Approach

A number of strategies have been adopted to address the issue of misclassification. One common approach is to collapse/sum variants according to a minor allele frequency (MAF) threshold. It is suggested that the variants below the specified MAF threshold are more likely to be functional than the more common variants. Thus this approach is aimed to increase the proportion of causal variants being tested in each collapsed region.

In the Combined Multivariate and Collapsing (CMC) method, variants in a region are divided into subgroups on the basis of predefined criteria of MAF, and within each group variant are collapsed/summed. A multivariate test (e.g. Hotelling's T^2 test, LRT) is then applied for analysis of the groups of variants^{33,34}. Compared to the original collapsing/sum methods where all variants regardless of their functional consequence were aggregated into a single group, the CMC method is more robust to misclassification when non-causal variants are grouped with causal variants^{34,37}.

The application of the CMC method to an empirical sequencing data shows that optimal power of this method can only be achieved when true causal variants have an allele frequency consistent with the proposed threshold³³. However, different genes may have very different relationships between allele frequency and functional effect. This

motivates an extension of the CMC method to a variable-threshold approach that can adapt to properties of individual genes³³. Using this variable-threshold approach, rare variants in each region are grouped together by an optimized allele-frequency threshold that maximizes the test statistics in the data. An exact p value is then obtained from permutation to control for type I error³³. Although the variable-threshold approach performs well relative to original CMC method, the optimizing procedure and the required permutation make implementation of this approach challenging in genome-wide studies.

Weighted Sum Test

Another approach to minimize the power reduction due to misclassification is to incorporate weight that reflects the likelihood of a variant to be functional. Decreasing the weight of non-causal variants and increasing the weight of causal variants can improve the power of burden tests. In a case-control study, a weighted sum test collapses a set of variants into a single weighted average of the number of minor alleles. The weight for each variant is defined as a function of the expected variance of minor allele counts in the control population. Thus a mutation that is rare among unaffected individuals has a higher weight than one that is common among the unaffected individuals³⁷. The phenotype is permuted (e.g. affected/unaffected status) to control for type I error^{33,37}. Similar weight schemes can be applied to the burden test for quantitative traits, in which case the statistical significance can be assessed via a Monte Carlo p value

³².

In addition to weighting variants as a function of allele frequency, incorporating computational predictions of the functional consequence for rare variants (e.g. PolyPhen-2 score) can also boost the power of the original burden test in which rare variants are weighted equally ³³.

Several independent simulation and empirical data analyses conclude that the weighted sum method has the best power when all variants in the targeted region affect the phenotype in the same direction ^{38,39}. Both the CMC method and the weighted sum method, are more robust relative to other burden tests in the presence of misclassification, ^{33,34,37,39}. However, all forms of burden tests in general are subjected to the power reduction due to presence of variants with opposite effects in the same region.

C-alpha Test

In contrast to burden tests that seek to determine if the mean of allele count is biased from what is expected under the null hypothesis, the C-alpha test is based on the variance of allele counts to determine whether alleles are randomly distributed across cases and controls as expected under the null hypothesis of no association, or follow a mix distribution as a result of detrimental, neutral and protective variants presenting at the same region ³⁸. From the simulation results, C-alpha shows comparable or slightly better power than burden tests when the effects of rare alleles are uniformly in one direction and much greater power when protective and detrimental variants are mixed ³⁸.

The C-alpha test is robust to the direction and magnitude of effect. However, as with other burden-style tests, C-alpha requires permutation for accurate significance estimation, especially when the number of variants in the region is small or when variants

are strongly correlated³⁸. Furthermore, the C-alpha test requires comparisons between two populations (i.e. cases versus controls) and thus has very limited application to analyze quantitative traits³².

SKAT and Optimal SKAT

The sequence kernel association test (SKAT) is proposed to overcome the limitations of burden test and C-alpha test. SKAT uses a multiple regression model as in the multiple-variant approach to directly regress the qualitative or quantitative phenotype on genetic variants in a region. Since no collapsing/summarizing is required, SKAT allows different variants to have different direction and magnitude of effects, including no effects. The regression framework of SKAT accommodates for easy adjustment of covariates, such as principal components, to account for population stratification³².

Due to the large number of degrees of freedom, the likelihood ratio test has limited power, particularly for rare variants. Therefore, SKAT compares the extended model including the variant effects to the null model of covariates alone, and uses a variance-component score test to examine whether the coefficient for each variant in the region follows a distribution under the null hypothesis of no association³².

The correct weighting scheme can improve the power of SKAT. The default weight in SKAT is calculated from a beta distribution density function with pre-specified shape parameters evaluated at the sample MAF ($\sqrt{\omega_j} = \text{Beta}(\text{MAF}_j; a_1 = 1, a_2 = 25)$). In practice, this weighting scheme increases the weight for rare variants while putting sufficient weights for variants with MAF 1%-5%. Under the null hypothesis, the variance-component score statistic follows a mixture of chi-square distributions. A

computationally efficient algorithm is used to approximate this mixed distribution and estimate the p value for significance⁴⁰.

SKAT has improved power over burden tests in the presence of causal variants with opposite effects, and is generally the most powerful method across all scenarios with one exception. SKAT assumes that variants act independently to affect disease risk. Without accounting for correlations between variants, SKAT is less powerful than burden tests when a large number of variants in a region are causal and in the same direction. Since in practice the correlation between variants is unknown and can vary from one region to another across the genome, a data-adaptive optimal test (SKAT-O) incorporating the correlation parameter was developed. SKAT-O consists of a class of tests including burden test and SKAT as special cases, and derives an optimal test within the class that maximize the power⁴¹.

A simulation study in SKAT-O suggests that when the sample size is moderate or large (i.e. 2000 or 5000), type I error is accurately controlled at moderate α levels, but is slightly inflated at very stringent α levels (i.e. 2×10^{-6}). If the sample sizes are small, the precise control of type I error is necessary. Permutation-based procedures can be used to generate a Monte Carlo p values for significance, though this can be computationally expensive.

SKAT-O has comparable power to burden tests in the presence of a large proportion of causal variants and is most powerful when the proportion of causal variants is low. Meanwhile, both SKAT and SKAT-O outperform burden tests when the causal variant coefficients are in mixed directions⁴¹.

In conclusion, SKAT-O is a computationally efficient regression approach that allows flexible adjustment for covariates. Comparing to burden tests, SKAT-O performs very well under broader circumstances. Although SKAT and SKAT-O are more robust to misclassification, they still suffer a substantial power reduction as the number of non-causal variants included in the model increases. Prior knowledge in the functional consequences of rare variants can be utilized to guide the variant selection to optimize the power of SKAT analysis.

MATERIALS AND METHODS

Study Population

The Beaver Dam Eye Study (BDES) is a population-based cohort study established in Beaver Dam, Wisconsin in 1987⁴². At the time the BDES was initiated, Beaver Dam, Wisconsin was a well-defined community with a primarily white population of approximately 17,000 individuals. From 1987 to 1988, the Wisconsin Survey Research Laboratory conducted a complete census of the city and township of Beaver Dam and identified 3,715 households with at least one occupant aged from 43 to 84 years. Individuals in the targeted age range were contacted by telephone and invited to participate in the examination phase of the study. Of the 5,924 individuals eligible to enroll, 4,926 (83.14%) individuals underwent a complete ocular examination and a personal history questionnaire during 1988 - 1990. For the 2,336 participants with known familial relationships in the catchment area of the study, pedigrees were constructed and the familial relationships were confirmed at their five-year follow-up visits.

We selected individuals from the parent BDES cohort with extreme baseline IOP or refractive error measurements with available covariate information. The informed consent was obtained from all study participants and the institutional review board at the University of Wisconsin approved all protocols. The Johns Hopkins University institutional review board also approved the statistical analysis of these existing data.

Clinical Evaluation

Detailed medical histories and eye examinations were performed for each participant at the baseline evaluation ¹¹. To maintain consistency and validity of measurements, all examinations were carried out by trained observers. IOP was measured with a Goldmann applanation tonometer after instilling a drop of fluorescein combined with a topical anesthetic (Fluores; Barnes-Hind Armour Pharmaceutical Co., Kankakee, IL) in each eye. Before measuring the IOP in one eye, the tonometer was set to 10 mm Hg and the pressure was recorded only after the tonometer was moved back from the cornea. The procedure was repeated for the other eye. Any unreliable measurement marked by the observers was excluded from the analysis. Blood pressure was measured according to the Hypertension Detection and Follow-up Program (HDFP) protocol. A detailed medical history, including information about hypertension, diabetes, and other medical conditions and history of medication use, was obtained from each participant. Although longitudinal eye measurements were collected, only baseline examination of IOP measurements was used in this analysis.

DNA Extraction and Genotyping

For 66% of participants, DNA was extracted from buffy coat separated at the time of the blood draw from the second or third Beaver Dam visit and stored at -80° C. For 20% of participants, DNA was extracted from frozen whole blood cells from the first examination. The remaining 14% individuals had their DNA extracted from both sources.

All participants were genotyped on Illumina HumanExome-12 V1-1 BeadChip at the Genetic Resources Core Facility (GRCF) at Johns Hopkins Institute of Genetic

Medicine. Genotyping calling was carried out using Illumina's GenTrain version 1.0 clustering algorithm in GenomeStudio version 2011.1.

The Illumina HumanExome-12 V1-1 BeadChip targets 242,901 variants, the majority of which are protein-altering variants (i.e. non-synonymous, splice and stop altering variants) selected from ~ 12,000 sequenced genomes representing multiple ethnicities and complex traits. The exome array also includes previously reported GWAS associations, ancestry informative markers (AIMs), HLA tags, randomly selected synonymous variants and variants for identity-by-descent (IBD) estimation. Details about exome array content and selection strategies can be found at the exome array design webpage (http://genome.sph.umich.edu/wiki/Exome_Chip_Design).

Quality Control Procedures

We applied both sample-level and marker-level quality control filters. Figure 1 illustrates the quality control procedures for our exome array data.

Quality Control for Samples

In total, 1,973 of 2,098 (94.0%) attempted samples were successfully genotyped and passed quality control metrics (i.e. GenCall score > 0.15 and call rate $> 98\%$). We identified 36 sex discrepancies of which 21 had a sex chromosome abnormality. The remaining 15 individuals were excluded from further analysis. Since this is a population-based study that included families, we evaluated cryptic relatedness. Using pairwise IBD sharing and 4,801 common variants across the autosomes we identified 156 close relatives with IBD sharing > 0.20 (Figure 2). Based on the pedigree data, we further confirmed the relationships of these 156 individual pairs as 15 parent-child pairs, 50 full-

sibling pairs, 6 first-cousin pairs, 73 avuncular/grandparent-child pairs and 12 unexpected duplicate pairs. In addition, the estimated genome-wide IBD sharing identified a parent-child pair with a Mendelian inconsistency. Thus, we excluded 127 samples with the lower sample call rate from the confirmed relative pairs with cryptic relatedness or Mendelian error.

All BDES participants had baseline IOP measurements in both eyes. Potential confounding variables were included as covariates in the analyses including age, sex, systolic blood pressure, and IOP treatment. 105 samples were missing one or more of these values for the specified covariates and were excluded, leaving 1,661 individuals for the association analysis (Figure 1).

Quality Control for Markers

In total, 236,546 of 242,901 (97.4%) attempted variants were successfully genotyped and had call rate > 95%. The HapMap concordance rate was 99.82% and the genotype concordance among 21 blind duplicate sample pairs was 99.99%. We further restricted the variants to those on autosomal chromosomes with a call rate > 98%. A total of 158,595 variants were excluded because they were monomorphic (N = 132,202) or extremely rare with only one minor allele in the study sample (MAF<0.05%) (N = 26,389), resulting in 75,669 variants for association analysis. Since departures from Hardy-Weinberg equilibrium (HWE) can be a result of evolutionary forces, genotyping errors, population admixture and marker-trait associations, we evaluated the HWE on the final set of variants, and reported the p value estimated from the HWE exact test for significant associations⁴³.

Assessment of Population Stratification

The BDES population self-reports as primarily European-American. To evaluate the genetic ancestry we performed principal component analysis (PCA) using SMARTPCA in EIGENSTRAT 4.2⁴⁴. There are 2,736 ancestry informative makers (AIMs) designed to distinguish Native Americans and Africans from European ancestry in the genotyped array. Figure 3 shows the plot of PCA clusters for BDES samples and 44 HapMap controls from seven ancestry groups, and there is no apparent population substructure in the study samples. The first two principle component scores from the PCA analysis were incorporated as covariates for IOP adjustment.

Statistical Analysis

Quantitative Trait Adjusted for Covariates

For the quantitative trait analysis, the higher baseline IOP measurement from either eye was used. The distribution of intraocular pressure was approximately normal. (Figure 4). Prior to the association analysis, the trait was first adjusted using a linear regression model as a function of age, sex, systolic blood pressure, IOP treatment, and the first two principal components. Then for each participant, the variation of IOP due to the individual's specific values of covariates was calculated and subtracted from the observed IOP measurements to generate an adjusted IOP value. The adjusted IOP values were used in the association analysis.

Annotations

Annotation was performed using SeattleSeq Annotation Server 138 version 9.01 under CRCh37/hg19 (see [URLs](#)). Except for 57 insertions and deletions, nearly all

variants (N = 75,612) that passed quality control filters were annotated. This server recognizes each variant by its CRCh37/hg19 position, and returns a variety of annotations derived from the Genome Variation Server (GVS) database, which integrates information from dbSNP, UCSC, NCBI and several other sources.

Single-Variant Analysis

Single-variant analysis was conducted to test the association of individual variants with IOP assuming additive effects. The analysis was performed using linear regression models for quantitative trait implemented in PLINK's '--assoc' option, which estimates the significant values from the Wald test assuming a normal distribution of the trait⁴⁵. We restricted this analysis to low frequency and common variants with MAF $\geq 1\%$ (N = 35,091), corresponding to approximately 30 copies of the minor allele in the analyzed cohort. Given the sample size of the study, the lower limit of 1% MAF allows stable estimates of sampling errors and provides adequate statistical power to detect low frequency variants with effect size (β) greater than 2.5. We set the genome-wide significance threshold to $p < 1.5 \times 10^{-6}$, corresponding to a Bonferroni correction for 35,091 tests.

Permutation tests were performed for variants with significant associations. The empirical p-values for these variants were obtained via 10,000,000 permutations using PLINK's '--mperm' option, applying a label-swapping approach. This option assumes carriers and non-carriers of a given variant are interchangeable under the null hypothesis of no association⁴⁵. Both nominal and empirical p values are reported for variants with significant associations.

Gene-Based Analysis

We performed gene-based analysis using SKAT-O implemented in R^{46,47}. The adjusted IOP was regressed on rare and low frequency variants (MAF 0.05% - 5%) in a gene region, allowing them to have different directions and magnitude of effects. Both simulation studies and real data analysis suggest that SKAT-O suffers a considerable power reduction when a large proportion of non-causal variants in the region are included in the analysis. Therefore, we limited the gene-based analysis to protein-altering variants, as they are more likely to result in functional alternations. A total of 58,043 variants were annotated to have a missense, nonsense, splice site, or frameshift consequence. We grouped these variants into their gene regions and allowed them to be analyzed in multiple regions if they were located in overlapping genes.

To maximize the statistical power, we applied the default weight $\sqrt{w_j} = \text{Beta}(MAF_j; a_1 = 1, a_2 = 25)$, which up-weights rarer variants while still putting nonzero weight for low frequency variants. After limiting the SKAT-O analysis to gene regions with at least two variants meeting the pre-specified criteria, we tested 9,650 genes consisting of 45,867 variants across autosomal chromosomes. For each gene region, the statistical significance of the variance-component score statistic was estimated from a mixed χ^2 distribution under null hypothesis of no association⁴¹. The genome-wide significance threshold was set to $p < 5.2 \times 10^{-6}$, corresponding to a Bonferroni correction for 9,650 tests. To ensure the accurate inference, we filtered the gene-based results with a cumulative MAF < 1% for SKAT-O analysis, corresponding to approximately 30 copies of the minor allele across the analyzed cohort.

Replication of Published GWAS Loci

To replicate the association of published GWAS loci, we manually identified 30 variants in or near the reported *CAVI/CAV2*, *CDKN2BAS*, *SIX1/SIX6*, *TMC01*, *GAS7* and *GLCCI1/ICAI* regions. Given that some of these loci were originally identified for POAG, we examined their single variant associations on IOP with and without adjustment for the glaucoma status in the model. Empirical p values were obtained for these variants via 10,000 permutations. Associations of these variants were considered as replicated if the nominal p-value was < 0.01 .

RESULTS

Study Participants

The characteristics of the 1,661 Beaver Dam Eye Study participants are summarized in Table 2. Females constitute 59.8% of the analyzed cohort and the mean age was 60 years for both men and women. The distribution of IOP in our study sample had a mean 15.6 mmHg with standard deviation 4.1 mmHg, which is comparable to other large population-based cohorts of European descent. The mean IOP of similar aging populations in the Blue Mountain Eye Study (BMES) and the NEIGHBOR consortium was 16 mmHg⁴⁸ and 15.4 mm Hg⁴⁹, respectively. The mean and median of IOP measurements in this study (15.0 mm Hg vs. 15.6 mm Hg) were very close, supporting a normal distribution for this quantitative trait. Nine-percent of the individuals had elevated IOP defined as > 21 mm Hg, and about three-percent of the analyzed cohort had reported IOP medical treatment. The mean systolic blood pressure was 130 mm Hg and the mean diastolic blood pressure was 77 mm Hg. The IOP values after adjusting for age, sex, systolic blood pressure, treatment for IOP and the first two principal components follows an approximately normal distribution with mean of 10.3 mm Hg and standard deviation of 3.8 mmHg (Figure 4).

Single-Variant Associations

A total of 35,091 low frequency and common variants ($MAF \geq 1\%$) were tested individually for association with IOP. We identified two novel associations at *GGA3* and *OGDH* (Figure 5, Figure 6).

GGA3 Loci

The most significant association with IOP was a missense variant on chromosome 17 (rs52809447, nominal $p = 1.3 \times 10^{-6}$, empirical $p = 3.7 \times 10^{-5}$) in the exonic region of the golgi-associated gamma adaptin ear containing ARF binding protein 3 (*GGA3*) gene. Each copy of minor allele C of variant rs52809447 results in a 3.05 mm Hg increase in IOP (95% CI = [1.77,4.33]). The A to G polymorphism substitutes a glutamic acid residue with a glycine residue at amino acid 97 (c.290A>G, [p.Glu97Gly]).

The products of GGA family include the ubiquitous coat proteins that regulate the trafficking of proteins between the trans-Golgi network and the lysosome. This gene family plays a significant role in Alzheimer' disease (AD) pathogenesis where GGA3 acts as a central regulator of β -Site APP-cleaving enzyme (BACE1) trafficking and degradation. The cerebral accumulation of amyloid-beta ($A\beta$) protein is a key neuro-pathological event in AD and the BACE1 cleaves the amyloid precursor protein (APP) at the β -secretase site to initiate the production of $A\beta$ peptides⁵⁰. Depletion of GGA3 during apoptosis stabilizes BACE1 by impairing its sorting to lysosome where it is normally degraded, and hence increases the levels of $A\beta$ protein in animal models⁵¹. The inverse correlation between GGA3 and BACE1 protein levels in AD and ischemic brain samples suggest that GGA3 is responsible for enhanced BACE1 levels and $A\beta$ accumulations during ischemia and in AD brain^{51,52}.

Although there is no direct implication of *GGA3* on IOP, POAG or any ocular disorders, an increasing number of studies indicate a potential participation of amyloid beta peptide ($A\beta$) in the retinal ganglion cell (RGCs) death following a long-term exposure to elevated IOP. Comparing to normal rat and monkey eyes, the glaucomatous

eyes with chronic ocular hypertension had significantly higher concentration of amyloid precursor protein (APP) and A β peptides in the optic nerves and RGCs^{53,54}. As in Alzheimer's disease, the increased level of APP and excess production of A β peptides in ocular hypertensive eyes were a result of the caspase activation, especially caspase-3 and caspase-8⁵⁵. Several studies have reported that inhibition of caspases enhances the survival of RGCs^{56,57}. These molecule and mechanistic similarities suggest a potential involvement of *GGA3* in the pathological event of RGCs death through interactions with caspases and APP, and provide evidence of common etiology between AD and glaucoma at the cellular level.

OGDH Loci

The second discovery was a novel association of a low frequency intronic variant (rs757702, MAF = 1.6%, nominal $p = 1.5 \times 10^{-6}$, empirical $p = 3.4 \times 10^{-5}$) on chromosome 7p14-p13 the 2-oxoglutarate dehydrogenase (*OGDH*) gene. One copy of the minor allele A of rs757702 was associated with 2.50 mm Hg increase in IOP (95% CI = [1.46, 3.54]).

2-Oxoglutarate dehydrogenase (*OGDH*), also designated as α -ketoglutarate dehydrogenase (α -*KGDH* or *Elk*), encodes one subunit of the mitochondrial 2-oxoglutarate dehydrogenase complex (OGDHC or KGDHC). OGDHC is a key regulatory enzyme in the tricarboxylic acid cycle (TCA cycle), participating directly in the bioenergetics processes⁵⁸. OGDHC also plays an essential role in the degradation of glucose and glutamate, the excess of which are neurotoxic and have specific significance in the brain⁵⁹. Diminished OGDHC activity may predispose to oxidative stress and disturb the brain mitochondrial function through impaired energy production and glucose metabolism, which ultimately leads to the neurodegeneration in age-related disorders like

Alzheimer's disease, Huntington's disease, and Parkinson's disease⁵⁹. Animal studies also corroborate that deficits in glucose and oxidative metabolism were not the result of neurodegeneration but the early events that precede the overexpression of APP and the formations of A β plaque in the neurodegeneration process⁵⁹.

Interestingly, many ophthalmologic disorders involved age-related neurodegeneration are increasingly recognized to have an origin of mitochondrial dysfunction⁶⁰. The impaired cerebral glucose utilization and deranged bioenergetics in AD were also observed in animal models of glaucoma^{61,62}. Elevated IOP induces the degenerative process of RGCs via mechanism of mitochondrial oxidation. Novel non-synonymous mitochondrial DNA changes paralleled with reduction in respiratory activities in POAG patients indicate that oxidative stress and mitochondrial dysfunction may interactively contribute to the pathogenesis of POAG⁶³.

A microarray analysis of retinal gene expression found differential regulation of *OGDH* gene over postnatal development period in mice⁶⁴. Although the role of OGDHC enzyme in the retina is unclear, the changes of its expression could potentially affect the metabolic rate of retinal cells, and through similar pathways as in the brain predispose risk of elevated IOP and glaucoma.

Gene-Based Associations

Gene-based test offers an alternative to single-variant analysis, which is often underpowered to detect associations with rare variants. In gene-based analyses, only protein-altering variants with rare and low minor allele frequencies (MAF 0.05% - 5%)

are included. We evaluated 9,650 genes having at least two such variants using the SKAT-O test.

We filtered the gene-based results with a cumulative MAF < 1% and identified the *HAP1* region on chromosome 17q21 and the *TNR* region on chromosome 1q24 to be significantly associated with IOP.

HAP1 Region

Novel association with huntingtin-associated protein 1 (*HAP1*) on chromosome 17q21.2-q21.3 reached the genome-wide significance (CMAF = 2.05%, $p = 1.13 \times 10^{-6}$). Four missense variants contributed to the gene-based analysis, including one low frequency variant rs34853043 that showed a suggestive association with IOP in single-variant analysis (MAF = 1.11%, nominal $p = 3.72 \times 10^{-5}$, empirical $p = 2.1 \times 10^{-4}$).

HAP1 encodes a protein interacting with huntingtin (Htt), the Huntington's disease (HD) protein. HAP1 is expressed predominantly in the neurons of the brain. Htt and HAP1 both participate in intracellular trafficking and vesicular transport⁶⁵. Together, they regulate the level of a polyglutamine expanded Htt (polyQ-htt). Disruption of this regulation causes aggregation of polyQ-htt, which leads to the neurodegeneration and cell death in HD⁶⁶.

HAP1 may also participate in the pathophysiological change of the Alzheimer's disease. In the neurodegenerative process, amyloid precursor protein (APP) is axonally transported, endocytosed and sorted to different cellular compartments to initiate the production of amyloid beta (A β). A study of the HAP1 protein knock-down mice with Alzheimer's disease observed a decreased rate of APP transport and increased A β levels,

suggesting that HAP1 may control APP subcellular trafficking and negatively regulate A β in the neurons⁶⁷.

Given the implication of amyloid beta peptide (A β) in the apoptotic death of retinal ganglion cells, the regulatory role of HAP1 on APP and A β peptides in brain might be extended to a common pathway in the eye, contributing to the physiologic changes in glaucoma.

TNR Region

Tenascin R (*TNR*) on chromosome 1q24 (CMAF = 1.33%, $p = 3.01 \times 10^{-6}$) was also significantly associate with higher IOP. Six rare missense variants in the *TNR* region contributed to the analysis.

The *TNR* gene encodes a member of the tenascin family of extracellular matrix glycoproteins. The encoded protein is exclusively expressed in the central nervous system. It affects cell migration, adhesion and differentiation, although no remarkable clinical consequences have been shown in knock-out animal models⁶⁸.

It is known that retinal ganglion cells (RGCs) death is a characteristic pathological event that precedes the glaucoma onset. Study of lesioned optic nerve of adult mice revealed that overexpression of tenascin-R was responsible for the failure of the regeneration of adult RGCs axons after injury⁶⁹. The role of tenascin-R as a repellent guidance molecule for newly growing and regenerating axons has been observed for different neuronal cell types across different species in vitro^{69,70}. The inhibitory effect of *TNR* in retinal ganglion cells survival and axonal regeneration suggests a potential role in glaucomatous progression, particularly under stress of elevated IOP.

Replicated Associations of Published GWAS Loci

Table 5 summarizes the coverage of each published GWAS loci for IOP and POAG in our exome array data. Except for *MVB12B* region, our exome array contains at least one variant for six published GWAS loci, among which the *TMCO1*, *CAV1/CAV2*, *CDKN2B-AS1* and *SIX1/SIX6* region have the same variants previously associated with IOP and/or POAG.

Of the 30 variants in six GWAS loci, we successfully replicated the association of a previously published POAG variant rs4236601 on chromosome 7q31 that contains the caveolin genes *CAV1* and *CAV2* (Table 6). This association was first discovered with POAG in an Icelandic population (OR = 1.36 [1.23 – 1.50], $p = 5.0 \times 10^{-10}$)²⁴, and was later replicated in two independent studies in populations of European ancestry^{25, 71}. The detected association in our study ($\beta = 0.46$, nominal $p = 1.9 \times 10^{-3}$, empirical $p = 1.8 \times 10^{-3}$) is consistent in direction with the POAG studies, namely, the minor allele A of variant rs4236601 was associated with elevated IOP and also higher risk of POAG²⁴. In all three POAG studies, cases were consisted of both normal-tension and high-tension POAG patients, suggesting that rs4236601 could increase the risk of POAG through affecting IOP levels. However, in one of the POAG studies, analysis on the subgroup of high-tension POAG individuals revealed no association with rs4236601²⁵.

Our result is also in agreement with two meta-analyses of IOP in Caucasian populations (Koolwijk et al.¹⁹, $\beta = 0.19$, $p = 1.1 \times 10^{-4}$; Ozel et al.²⁰, $\beta = 0.38$, $p = 2.8 \times 10^{-4}$). The adjustment of glaucoma status in our analysis did not affect the association of rs4236601 with IOP ($\beta = 0.46$, nominal $p = 2.0 \times 10^{-3}$, empirical $p = 2.8 \times 10^{-3}$), which may be reflective of the small number of self-reported glaucoma cases.

The proteins encoded by caveolin-1 (*CAV1*) and caveolin-2 (*CAV2*) gene are both major components of the caveolar plasma membranes. Caveolin-1 is expressed in several retinal cell types, including photoreceptor, retinal vascular endothelia cells, Müller galia, and retinal pigment epithelium (RPE) cells. Recent evidence links caveolin-1 to ocular pathologies including autoimmune uveitis, diabetic retinopathy, and POAG, but its role in the retina is largely unknown. Genetic ablation of caveolin-1 causes retinal functional deficits due to disruptions in micro-environmental homeostasis⁷² and blood-retinal barrier (BRB) breakdown⁷³. A cytotoxic agent that increases intraocular pressure and aqueous outflow resistance in mice can also cause increased expression of caveolin-1 in the treated human trabecular meshwork cells⁷⁴. Further investigations are required to elucidate the role of caveolin-1 and caveolin-2 in controlling IOP.

DISCUSSION

The contribution of rare and low frequency variants to complex diseases and traits has been long debated. In this study, we performed the first exome array analysis on 1,661 BDES participants to identify rare and low frequency variants influencing intraocular pressure. We discovered two low-frequency variants with large effects on IOP in the single-variant analysis: variant rs52809447 ($\beta = 3.05$ mmHg, nominal $p = 1.3 \times 10^{-6}$) in the *GGA3* region on chromosome 17 and variant rs757702 ($\beta = 2.50$ mmHg, nominal $p = 1.5 \times 10^{-6}$) in the *OGDH* region on chromosome 7. The gene-based analysis reported two genome-wide significant associations with IOP in the *HAPI* region on chromosome 17q21 and *TNR* region on chromosome 1q24,

In our discussion of the biological plausibility and the potential contribution of these four identified regions for IOP, we noticed an interesting connection between three of the genes and the Alzheimer's disease. Both *GGA3* gene and *HAPI* gene on chromosome 17 contribute to AD through indirect regulation of the A β peptide production, the excess of which is neurotoxic and has strong implication in both AD and glaucoma. The *OGDH* gene on chromosome 7 linked to AD through mechanisms involves interactions between mitochondrial dysfunction and oxidative stress, a mechanism that is also associated with a number of ocular disorders including elevated IOP and glaucoma.

Glaucoma and Alzheimer's disease are two chronic neurodegenerative diseases with a strong age-related incidence. Higher prevalence of glaucoma has been reported among AD patients in a number of population studies⁷⁵. Experimental studies in animal

models and human cell lines recognized several common pathogenetic pathways associated with both disorders, including mitochondrial dysfunction and A β peptide accumulation. These two mechanisms may act separately or synergistically in the pathogenesis of both glaucoma and AD. Although still not fully understood, research and clinical findings suggest a possible common causal relationship between AD and glaucoma.

There are limitations to this study. The Exome Array is constrained to the protein-altering variants including the non-synonymous coding and splice site variation observed in the ~ 12,000 individuals who comprised the initial exome sequencing discovery set. Thus majority of the exome array content have a low or rare minor allele frequency. In our analyzed cohort of 1,661 individuals, almost 70% of the variants in the genotyping assay are monomorphic or singletons. To detect the burden of such vary rare mutations to elevated IOP will require an extraordinarily large sample size.

Gene-based analysis was performed on the regions with at least two variants in the targeted MAF range. After excluding gene regions containing only one variant, on average, each analyzed region had six variants contributing to the gene-based analysis. About 75% of these regions had no more than seven variants. In some cases, theses variants span across the entire gene region but in others they cluster at a relatively small range. This raises a concern of adequate coverage of gene regions with such a small number of variants.

In addition, for a gene region with a small number of variants, the test is more likely to be driven by a single more common variant. This relatively common variant can be truly causal but when it is observed in only a handful of individuals, SKAT is less

competent to provide accurate significance estimations for the gene region. Therefore, we filtered our gene-based associations to two regions, as the variants contributing to the significant associations in both *HAPI* and *TNR* region had cumulative MAF > 1%. To detect the associations in other gene regions, a larger sample size is necessary.

In summary, we identified two low frequency variants in novel genetic loci and two new genes with rare and low frequency coding variants associated with intraocular pressure. The novel associations of three AD related genes, *GGA3*, *OGDH* and *HAPI*, lend support to the shared pathways of neurodegeneration in Alzheimer's disease and glaucoma. The inhibitory role of *TNR* gene in retinal ganglion cell survival and axonal regeneration has a potential implication in the glaucomatous progression induced by high intraocular pressure. Given the limited coverage of exome array data for many rare variants, a larger sample size is needed for comprehensive assessment of the variants associated with IOP.

FIGURES AND TABLES

Figure 1 Quality Control Procedures for Exome Array Data

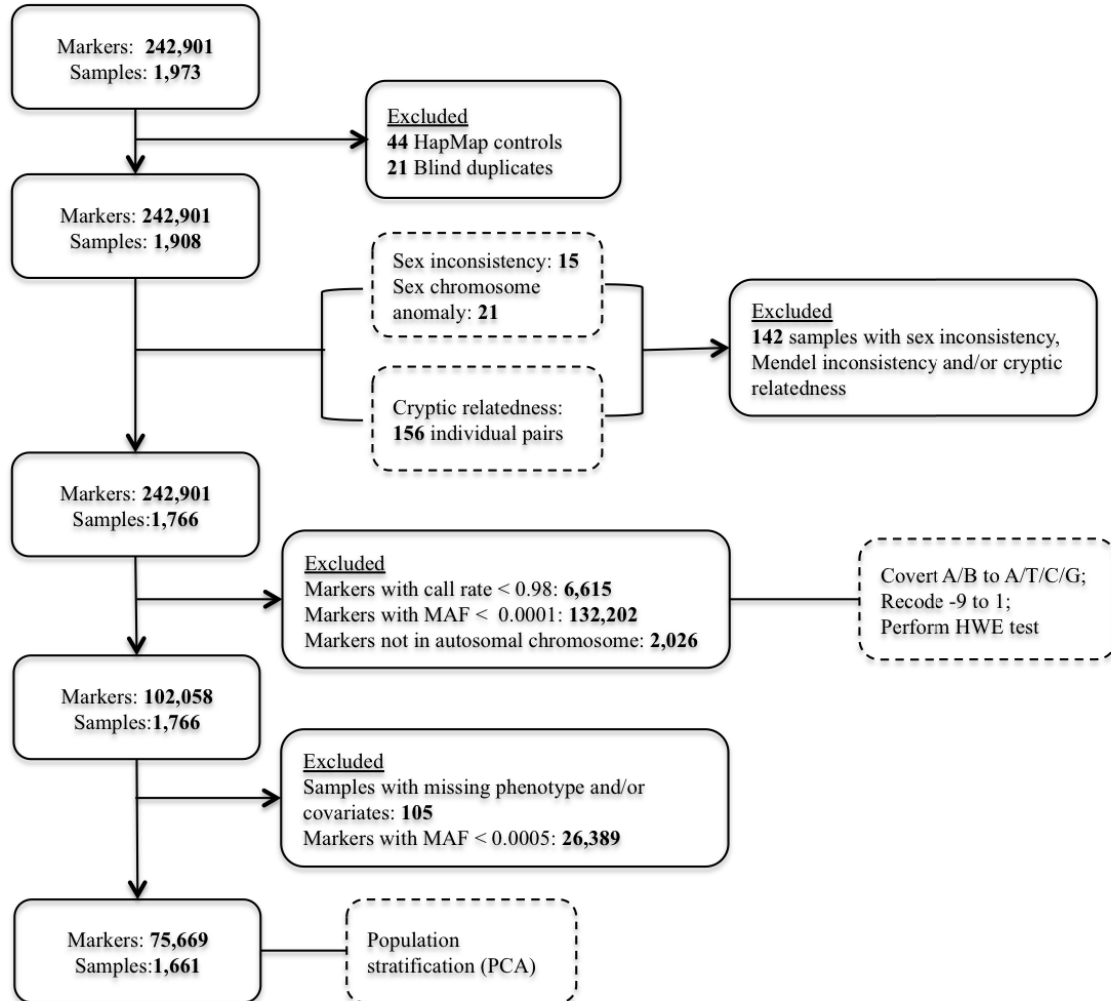
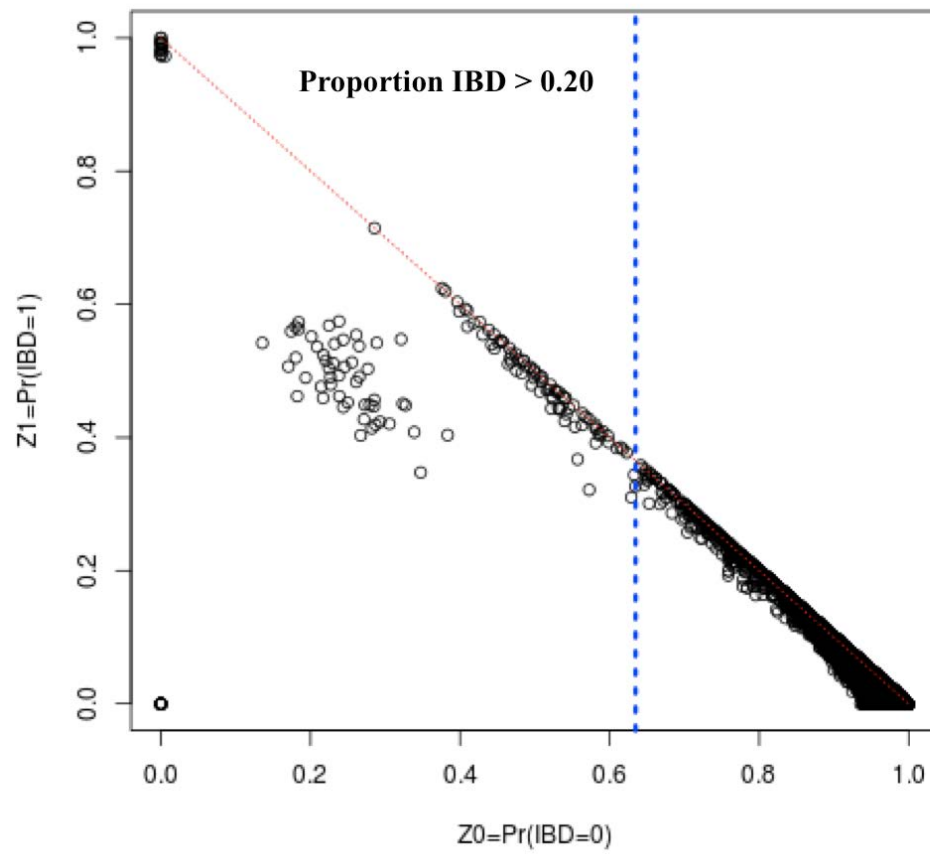


Figure 2 Probability of IBD Sharing for BDES Samples (N = 1,908)



**Figure 3 Principal Component Analysis for BDES Samples and
HapMap Controls**

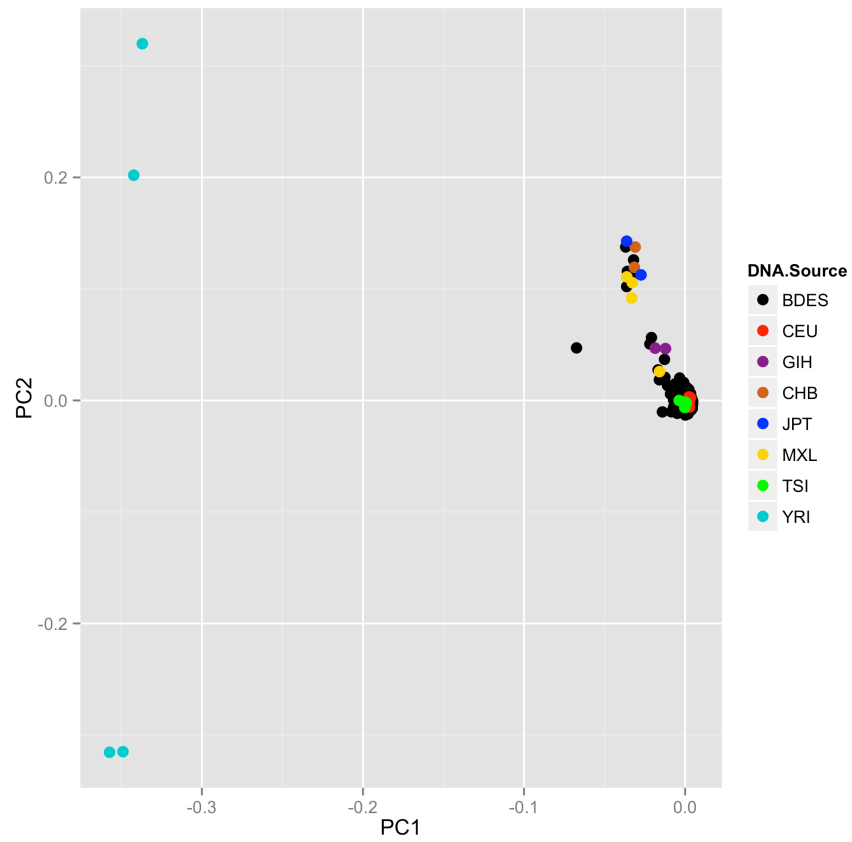


Figure 4 Distributions of Baseline IOP in Analyzed Cohort

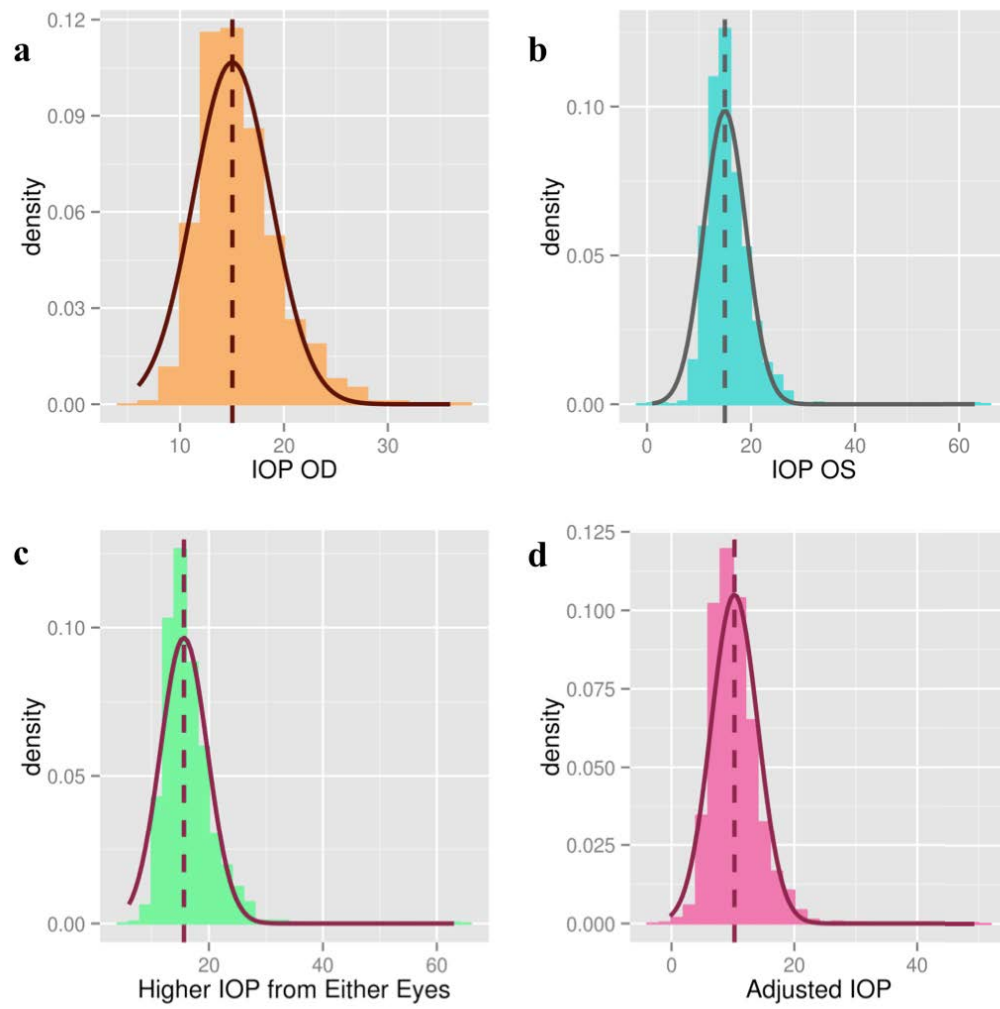


Figure 5 Manhattan Plot for Single-Variant Associations

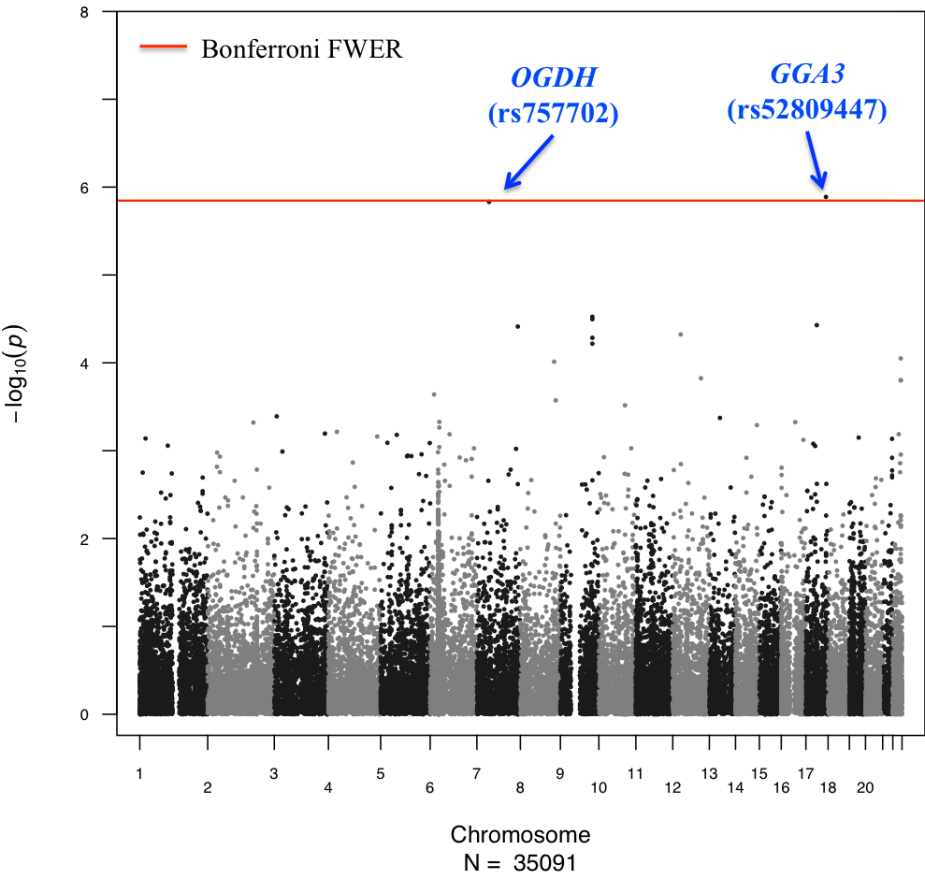


Figure 6 QQ Plot of Single-Variant Associations

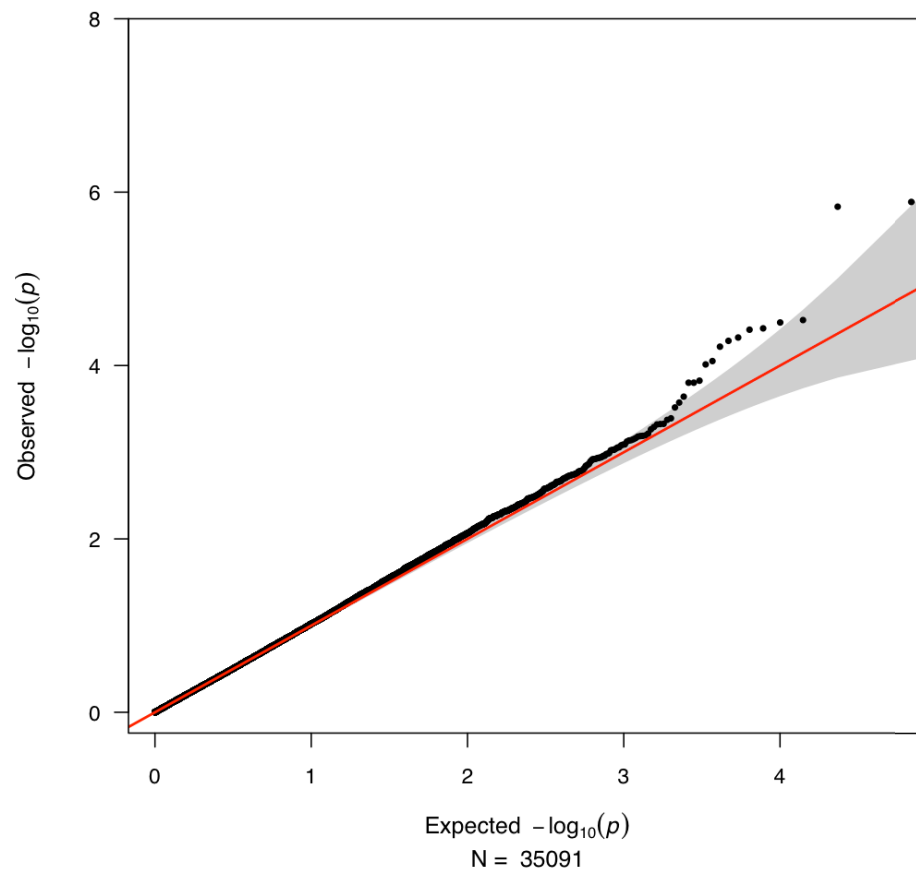


Table 1 Review of Statistical Approaches for Rare Variant Associations

Methods	Unit of Variants	Traits	Test Statistics	Robustness to Mixed Effects^a	Strategies for Misclassification^b	Account for Correlation^c
Single-Variant	Individual	Qualitative	Pearson χ^2 test	Robust	<i>NA</i>	No
		Quantitative	Fisher's exact test			
			Cochran-Armitage trend			
			Likelihood ratio test			
Multiple-Variant	Individual	Qualitative	Fisher's product method	Robust	<i>NA</i>	No
		Quantitative	Hotelling's T^2 test			
			Likelihood ratio test			
Collapsing/Sum	Group	Qualitative	Pearson χ^2 test	Not- Robust	<i>NA</i>	Yes
		Quantitative	Fisher's exact test			
			Cochran-Armitage trend			
			Likelihood ratio test			
CMC	Group	Qualitative	Fisher's product method	Not- Robust	Group by fixed thresholds	Yes
		Quantitative	Hotelling's T^2 test			
			Likelihood ratio test			
Variable-Threshold	Group	Qualitative	Likelihood ratio test	Not- Robust	Group by variable thresholds	Yes
		Quantitative				

Methods	Unit of Variants	Traits	Test Statistics	Robustness to Mixed Effects ^a	Strategies for Misclassification ^b	Account for Correlation ^c
Weighted Sum	Group	Qualitative Quantitative	Likelihood ratio test	Not- Robust	Weighted	Yes
C-alpha	Group	Qualitative	C-alpha score test	Robust	Mixed distributions	Yes
SKAT	Individual	Qualitative Quantitative	Variance component score statistic	Robust	Weighted	No
SKAT-O	Individual	Qualitative Quantitative	Variance component score statistic	Robust	Weighted	Yes
^a Mix effects are defined as the association of variants with IOP are in different directions and magnitude.						
^b Misclassification refer to inclusion of non-causal variants and exclusion of causal variants.						
^c Correlation between tested variants.						

Table 2 Characteristics of 1,661 Beaver Dam Eye Study Participants in Analyzed Cohort

Characteristic	Value
Age, mean \pm SD (range)	60.1 \pm 10.9 (43 – 86)
Female gender, % (N)	59.8% (994)
IOP (mm Hg) ^a , mean \pm SD (range)	15.6 \pm 4.1 (6 – 63)
IOP > 21 mm Hg ^a , % (N)	8.6% (143)
Participants with IOP Lowering Treatment, % (N)	2.7% (45)
Participants with Glaucoma ^b , % (N)	2.9% (48)
Systolic BP (mmHg), mean \pm SD (range)	130.2 \pm 20.6 (89 – 248)
Diastolic BP (mmHg), mean \pm SD (range)	77.4 \pm 10.9 (44 – 123)

^a Higher baseline IOP measurements from either eye.
^b Self-reported diagnosis of glaucoma.

Table 3 Single-Variant Associations Reached Genome-Wide Significance of $p < 1.5 \times 10^{-6}$

Gene	Chr:Position ^a	dbSNP ID	Mutations ^b	MAF (allele)	Beta (SE)	Nominal p Value	Empirical p Value ^c	HWE p Value
<i>GG43</i>	17: 73238508	rs52809447	p.Glu97Gly	1.11% (C)	3.05 (0.63)	1.3×10^{-6}	3.7×10^{-5}	1.0000
<i>OGDH</i>	7: 44686497	rs757702	Intron	1.60% (A)	2.50 (0.52)	1.5×10^{-6}	3.4×10^{-5}	0.0879

^a Chr:Position is reported in GRCh37.

^b Functional consequence is annotated in dbSNP.

^c Empirical p-value is calculated from 10,000,000 permutations.

Table 4 Gene-Based Associations Reached Genome-Wide Significance of $p < 5.2 \times 10^{-6}$

Gene	Location ^a	P value	CMAF ^b	Functional Consequences (N) ^c
<i>HAP1</i>	17q21.2-q21.3	1.1×10^{-6}	2.05%	Missense (4)
<i>TNR</i>	1q24	3.0×10^{-6}	1.33%	Missense (6)

^a Gene location is reported in GRCh37.

^b CMAF = cumulative minor allele frequency sum over all variants in the gene region.

^c Functional consequences in dbSNP. N = no. of variants.

Table 5 Coverage of the GWAS Loci for IOP and POAG in our Exome Array Data

Chr	GWAS Loci	Associated Trait	Reported Variant ^a	Correlated Variant (r ²)	No. Variant ^b
1q24	<i>TMCO1</i>	IOP, POAG	rs4656461	rs4656461 (1.000)	2
7p21	<i>GLCCII/ICAI</i>	IOP	rs59072263	<i>NA</i>	4
7q31	<i>CAVI/CAV2</i>	POAG	rs4236601	rs4236601 (1.000)	2
9p21	<i>CDKN2B-AS1</i>	POAG	rs1063192	rs7865618 (0.967) / rs2157719 (0.935) / rs564398 (0.873) / rs1412829 (0.873)	14
9q33	<i>MVB12B</i>	IOP	rs2286885	<i>NA</i>	0
14q23	<i>SIX1/SIX6</i>	POAG	rs10483727	rs10483727 (1.000) / rs33912345 (1.000)	5
17p13	<i>GAS7</i>	IOP	rs11656696	<i>NA</i>	3
			rs12150284	rs9897130 (0.702)	

^a Variants that reported in published GWAS.

^b Variants correlated (r²>0.7) with reported locus in our exome array data. Correlation is estimated from 1000 Genomes Pilot 1.

^c No. of variants tested for replication.

Table 6 Replication of GWAS Locus rs4236601 in *CAVI/CAV2* Region for IOP and POAG

Reference	dbSNP ID	Chr: Position ^a	MAF [allele]	Effect Size	Significance ^c	Associated GWAS Trait ^d
<i>Thorleifsson et al.</i> ²⁴	rs4236601	7: 116162729	23% [A] ^b	OR = 1.36	5.0×10^{-10}	POAG
<i>Koolwijk, et al.</i> ¹⁹	rs4236601	7: 116162729	29% [A]	$\beta = 0.19$	1.1×10^{-4}	IOP
Current	rs4236601	7: 116162729	26% [A]	$\beta = 0.46$	1.9×10^{-3}	IOP

^a Chr:Position is reported in GRCh37.

^b In published case-control studies for POAG, MAF in controls is reported.

^c Nominal p value from the discovery cohort in published studies and from the current study.

^d IOP = intraocular pressure; POAG = primary open-angle glaucoma.

BIBLIOGRAPHY

- 1 Quigley, Harry A., and Aimee T. Broman. "The number of people with glaucoma worldwide in 2010 and 2020." *British Journal of Ophthalmology* 90.3 (2006): 262-267.
- 2 Vajaranant, Thasarat S., et al. "The changing face of primary open-angle glaucoma in the United States: demographic and geographic changes from 2011 to 2050." *American journal of ophthalmology* 154.2 (2012): 303-314.
- 3 Klein, B. E., Ronald Klein, and K. L. Linton. "Intraocular pressure in an American community. The Beaver Dam Eye Study." *Investigative ophthalmology & visual science* 33.7 (1992): 2224-2228.
- 4 Glaucoma Overview. (n.d.). *eMedicineHealth*. Retrieved April 20, 2014, from http://www.emedicinehealth.com/glaucoma_overview/article_em.htm
- 5 Le, Anhchuong, et al. "Risk factors associated with the incidence of open-angle glaucoma: the visual impairment project." *Investigative ophthalmology & visual science* 44.9 (2003): 3783-3789.
- 6 Leske, M. Cristina, et al. "Risk factors for incident open-angle glaucoma: the Barbados Eye Studies." *Ophthalmology* 115.1 (2008): 85-93.
- 7 de Voogd, Simone, et al. "Is diabetes mellitus a risk factor for open-angle glaucoma?: The Rotterdam Study." *Ophthalmology* 113.10 (2006): 1827-1831.
- 8 Leske, M. Cristina, et al. "Factors for glaucoma progression and the effect of treatment: the early manifest glaucoma trial." *Archives of Ophthalmology* 121.1 (2003): 48-56.

- 9 AGIS Investigators. "The Advanced Glaucoma Intervention Study (AGIS): 4. Comparison of treatment outcomes within race: seven-year results." *Ophthalmology* 105.7 (1998): 1146-1164.
- 10 Garcia-Valenzuela, E., et al. "Programmed cell death of retinal ganglion cells during experimental glaucoma." *Experimental eye research* 61.1 (1995): 33-44.
- 11 Duggal, Priya, et al. "A genetic contribution to intraocular pressure: the beaver dam eye study." *Investigative ophthalmology & visual science* 46.2 (2005): 555-560.
- 12 van Koolwijk, Leonieke ME, et al. "Genetic contributions to glaucoma: heritability of intraocular pressure, retinal nerve fiber layer thickness, and optic disc morphology." *Investigative ophthalmology & visual science* 48.8 (2007): 3669-3676.
- 13 Klein, Barbara EK, Ronald Klein, and Kristine E. Lee. "Heritability of risk factors for primary open-angle glaucoma: the Beaver Dam Eye Study." *Investigative ophthalmology & visual science* 45.1 (2004): 59-62.
- 14 Charlesworth, Jac C., et al. "Linkage to 10q22 for maximum intraocular pressure and 1p32 for maximum cup-to-disc ratio in an extended primary open-angle glaucoma pedigree." *Investigative ophthalmology & visual science* 46.10 (2005): 3723-3729.
- 15 Duggal, Priya, et al. "Identification of novel genetic loci for intraocular pressure: a genomewide scan of the Beaver Dam Eye Study." *Archives of ophthalmology* 125.1 (2007): 74-79.
- 16 Rotimi, Charles N., et al. "Genomewide scan and fine mapping of quantitative trait loci for intraocular pressure on 5q and 14q in West Africans." *Investigative ophthalmology & visual science* 47.8 (2006): 3262-3267.

- 17 Lee, Mi Kyeong, et al. "Replication of a glaucoma candidate gene on 5q22. 1 for intraocular pressure in mongolian populations: the GENDISCAN Project." *Investigative ophthalmology & visual science* 51.3 (2010): 1335-1340.
- 18 Axenovich, Tatiana, et al. "Linkage and association analyses of glaucoma related traits in a large pedigree from a Dutch genetically isolated population." *Journal of medical genetics* 48.12 (2011): 802-809.
- 19 van Koolwijk, Leonieke ME, et al. "Common genetic determinants of intraocular pressure and primary open-angle glaucoma." *PLoS genetics* 8.5 (2012): e1002611.
- 20 Ozel, A. Bilge, et al. "Genome-wide association study and meta-analysis of intraocular pressure." *Human genetics* 133.1 (2014): 41-57.
- 21 Burdon, Kathryn P., et al. "Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1." *Nature genetics* 43.6 (2011): 574-578.
- 22 Strange, Amy, et al. "Genome-wide association study of intraocular pressure identifies the GLCCI1/ICA1 region as a glaucoma susceptibility locus." *Human Molecular Genetics* 22.22 (2013): 4653-4660.
- 23 Nag, Abhishek, et al. "A genome-wide association study of intra-ocular pressure suggests a novel association in the gene FAM125B in the TwinsUK cohort." *Human Molecular Genetics* (2014): ddu050.
- 24 Thorleifsson, Gudmar, et al. "Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma." *Nature genetics* 42.10 (2010): 906-909.

- 25 Wiggs, Janey L., et al. "Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma in Caucasians from the USA." *Human molecular genetics* 20.23 (2011): 4707-4713.
- 26 Osman, Wael, et al. "A genome-wide association study in the Japanese population confirms 9p21 and 14q23 as susceptibility loci for primary open angle glaucoma." *Human molecular genetics* 21.12 (2012): 2836-2842.
- 27 Nakano, Masakazu, et al. "Common variants in CDKN2B-AS1 associated with optic-nerve vulnerability of glaucoma identified by genome-wide association studies in Japanese." *PloS one* 7.3 (2012): e33389.
- 28 Burdon, Kathryn P., et al. "Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1." *Nature genetics* 43.6 (2011): 574-578.
- 29 Wiggs, Janey L., et al. "Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma." *PLoS genetics* 8.4 (2012): e1002654.
- 30 Manolio, Teri A., et al. "Finding the missing heritability of complex diseases." *Nature* 461.7265 (2009): 747-753.
- 31 Gibson, Greg. "Rare and common variants: twenty arguments." *Nature Reviews Genetics* 13.2 (2012): 135-145.
- 32 Wu, Michael C., et al. "Rare-variant association testing for sequencing data with the sequence kernel association test." *The American Journal of Human Genetics* 89.1 (2011): 82-93.

- 33 Price, Alkes L., et al. "Pooled association tests for rare variants in exon-resequencing studies." *The American Journal of Human Genetics* 86.6 (2010): 832-838.
- 34 Li, Bingshan, and Suzanne M. Leal. "Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data." *The American Journal of Human Genetics* 83.3 (2008): 311-321.
- 35 Morgenthaler, Stephan, and William G. Thilly. "A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST)." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 615.1 (2007): 28-56.
- 36 Morris, Andrew P., and Eleftheria Zeggini. "An evaluation of statistical approaches to rare variant analysis in genetic association studies." *Genetic epidemiology* 34.2 (2010): 188-193.
- 37 Madsen, Bo Eskerod, and Sharon R. Browning. "A groupwise association test for rare mutations using a weighted sum statistic." *PLoS genetics* 5.2 (2009): e1000384.
- 38 Neale, Benjamin M., et al. "Testing for an unusual distribution of rare variants." *PLoS genetics* 7.3 (2011): e1001322.
- 39 Han, Fang, and Wei Pan. "A data-adaptive sum test for disease association with multiple common or rare variants." *Human heredity* 70.1 (2010): 42-54.
- 40 Davies, Robert B. "Algorithm AS 155: The distribution of a linear combination of χ^2 random variables." *Applied Statistics* (1980): 323-333.
- 41 Lee, Seunggeun, Michael C. Wu, and Xihong Lin. "Optimal tests for rare variant effects in sequencing association studies." *Biostatistics* 13.4 (2012): 762-775.

- 42 Linton, Kathryn LP, Barbara EK Klein, and Ronald Klein. "The validity of self-reported and surrogate-reported cataract and age-related macular degeneration in the Beaver Dam Eye Study." *American journal of epidemiology* 134.12 (1991): 1438-1446.
- 43 Wang, Jian, and Sanjay Shete. "Testing Departure from Hardy–Weinberg Proportions." *Statistical Human Genetics*. Humana Press, 2012. 77-102.
- 44 Price, Alkes L., et al. "Principal components analysis corrects for stratification in genome-wide association studies." *Nature genetics* 38.8 (2006): 904-909.
- 45 Purcell, Shaun, et al. "PLINK: a tool set for whole-genome association and population-based linkage analyses." *The American Journal of Human Genetics* 81.3 (2007): 559-575.
- 46 Seunggeun Lee, Larisa Miropolsky and Micheal Wu (2014). SKAT: SNP-set (Sequence) Kernel Association Test. R package version 0.95. <http://CRAN.R-project.org/package=SKAT>
- 47 R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- 48 Rochtchina, Elena, Paul Mitchell, and Jie Jin Wang. "Relationship between age and intraocular pressure: the Blue Mountains Eye Study." *Clinical & experimental ophthalmology* 30.3 (2002): 173-175.
- 49 Wiggs, Janey L., et al. "The NEIGHBOR Consortium Primary Open-Angle Glaucoma Genome-wide Association Study: Rationale, Study Design, and Clinical Variables." *Journal of glaucoma* 22.7 (2013): 517-525.

- 50 Zhao, Jie, et al. " β -Site amyloid precursor protein cleaving enzyme 1 levels become elevated in neurons around amyloid plaques: implications for Alzheimer's disease pathogenesis." *The Journal of neuroscience* 27.14 (2007): 3639-3649.
- 51 Tesco, Giuseppina, et al. "Depletion of GGA3 stabilizes BACE and enhances β -secretase activity." *Neuron* 54.5 (2007): 721-737.
- 52 Walker, Kendall R., et al. "Depletion of GGA1 and GGA3 mediates postinjury elevation of BACE1." *The Journal of Neuroscience* 32.30 (2012): 10423-10437.
- 53 Kipfer-Kauer, Anna, et al. "Distribution of amyloid precursor protein and amyloid- β in ocular hypertensive C57BL/6 mouse eyes." *Current eye research* 35.9 (2010): 828-834.
- 54 Ito, Yasushi, et al. "Induction of amyloid- β 1-42 in the retina and optic nerve head of chronic ocular hypertensive monkeys." *Molecular vision* 18 (2012): 2647.
- 55 McKinnon, Stuart J., et al. "Caspase activation and amyloid precursor protein cleavage in rat ocular hypertension." *Investigative ophthalmology & visual science* 43.4 (2002): 1077-1087.
- 56 Chaudhary, P., et al. "Caspase inhibitors block the retinal ganglion cell death following optic nerve transection." *Molecular brain research* 67.1 (1999): 36-45.
- 57 Kermer, Pawel, et al. "Inhibition of CPP32-like proteases rescues axotomized retinal ganglion cells from secondary cell death in vivo." *The Journal of neuroscience* 18.12 (1998): 4656-4662.
- 58 Bunik, Victoria, et al. "Novel isoenzyme of 2-oxoglutarate dehydrogenase is identified in brain, but not in heart." *FEBS journal* 275.20 (2008): 4990-5006.

- 59 Gibson, Gary E., et al. "The α -ketoglutarate-dehydrogenase complex." *Molecular neurobiology* 31.1-3 (2005): 43-63.
- 60 Schrier, Samantha A., and Marni J. Falk. "Mitochondrial disorders and the eye." *Current opinion in ophthalmology* 22.5 (2011): 325.
- 61 Schober, Michael S., et al. "Bioenergetic-based neuroprotection and glaucoma." *Clinical & experimental ophthalmology* 36.4 (2008): 377-385.
- 62 Hayreh, Sohan Singh, Anders Bill, and Goran Olof Sperber. "Effects of high intraocular pressure on the glucose metabolism in the retina and optic nerve in old atherosclerotic monkeys." *Graefe's archive for clinical and experimental ophthalmology* 232.12 (1994): 745-752.
- 63 Lee, Shanjean, et al. "Mitochondrial dysfunction in glaucoma and emerging bioenergetic therapies." *Experimental eye research* 93.2 (2011): 204-212.
- 64 Schippert, Ruth, Frank Schaeffel, and Marita Pauline Feldkaemper. "Microarray analysis of retinal gene expression in Egr-1 knockout mice." *Molecular vision* 15 (2009): 2720.
- 65 Rong, Juan, Shi-Hua Li, and Xiao-Jiang Li. "Regulation of intracellular HAP1 trafficking." *Journal of neuroscience research* 85.14 (2007): 3025-3029.
- 66 Wong, Yvette C., and Erika LF Holzbaur. "The Regulation of Autophagosome Dynamics by Huntingtin and HAP1 Is Disrupted by Expression of Mutant Huntingtin, Leading to Defective Cargo Degradation." *The Journal of Neuroscience* 34.4 (2014): 1293-1305.

- 67 Yang, Gui-Zhi, et al. "Huntingtin associated protein 1 regulates trafficking of the amyloid precursor protein and modulates amyloid beta levels in neurons." *Journal of neurochemistry* 122.5 (2012): 1010-1022.
- 68 Anlar, Banu, and Ayşen Gunel-Ozcan. "Tenascin-R: Role in the central nervous system." *The international journal of biochemistry & cell biology* 44.9 (2012): 1385-1389.
- 69 Becker, Thomas, et al. "Tenascin-R inhibits regrowth of optic fibers in vitro and persists in the optic nerve of mice after injury." *Glia* 29.4 (2000): 330-346.
- 70 Becker, Catherina G., et al. "Tenascin-R as a repellent guidance molecule for newly growing and regenerating optic axons in adult zebrafish." *Molecular and Cellular Neuroscience* 26.3 (2004): 376-389.
- 71 Loomis, Stephanie J., et al. "Association of "CAV1/CAV2" Genomic Variants with Primary Open-Angle Glaucoma Overall and by Gender and Pattern of Visual Field Loss." *Ophthalmology* 121.2 (2014): 508-516.
- 72 Li, Xiaoman, et al. "Loss of caveolin-1 impairs retinal function due to disturbance of subretinal microenvironment." *Journal of Biological Chemistry* 287.20 (2012): 16424-16434.
- 73 Gu, Xiaowu, et al. "Loss of Caveolin-1 Causes Blood–Retinal Barrier Breakdown, Venous Enlargement, and Mural Cell Alteration." *The American journal of pathology* (2013).
- 74 Giovingo, Michael, et al. "sCD44 overexpression increases intraocular pressure and aqueous outflow resistance." *Molecular vision* 19 (2013): 2151.

75 Tsolaki, Fani, et al. "Alzheimer's disease and primary open-angle glaucoma: is there a connection?" *Clinical ophthalmology (Auckland, NZ)* 5 (2011): 887.

CURRICULUM VITA

FEI CHEN

Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
Email: fchen@jhsph.edu

EDUCATION

09/2012–05/2013	Johns Hopkins Bloomberg School of Public Health Sc.M. in Genetic Epidemiology	Baltimore, MD
09/2007–05/2009	Columbia University M.A. in Biomedical Informatics	New York, NY
09/2003–07/2007	Fudan University School of Life Science B.S. in Biology	Shanghai, China

HONORS AND AWARDS

Master's Tuition Scholarship, Johns Hopkins Bloomberg School of Public Health	2013 – 2014
3 rd Class People's Scholarship, Fudan University	2004 – 2005

RESEARCH EXPERIENCE

07/2013–Present	Genetic Epidemiology	Baltimore MD
	Johns Hopkins Bloomberg School of Public Health	

Research Assistant

- **Thesis:** *“Exome array analysis identifies novel susceptibility loci for intraocular pressure”*
Conducted single-variant and gene-based analysis on exome array data to examine the association of rare and low frequency variants with intraocular pressure and refractive error in a population-based cohort from the Beaver Dam Eye Study.
Advisor: Dr. Priya Duggal
Secondary Reader: Dr. Alison Klein
- Served as second reviewer in a meta-analysis for XRCC2-Arg188His polymorphism as a susceptibility variant for breast, thyroid, endometrial, colorectal, pancreatic and lung cancers.

09/2007–08/2009	Translational Research Lab	New York, NY
	Columbia University	

Research Assistant

- **Thesis:** *“The role of Electronic Medical Records in public health research: opportunities and*

challenges.”

- Conducted a systematic literature review to characterize the value of and issues associated with using clinical data for public health research purposes and to provide informatics solutions to minimize the data quality issues in the EHR system.

Advisor: Dr. Chunhua Weng

Secondary Reader: Dr. James Cimino

- Natural Language Processing: Trained in-house developed software to conduct semantic analysis of clinical trial eligibility criteria aimed at facilitating subject recruitment in clinical research.

05/2005–05/2007 Biodiversity Science and Ecological Engineering Lab Shanghai, China
Fudan University School of Life Science

Research Assistant

- **Thesis:** “*Analysis of the genetic diversity of *Amygdalus Communis* in Xinjiang Province using SSR biomarkers.*”

Worked independently to evaluate the genetic diversity of *Amygdalus Communis*, a species endemic to Xinjiang Province, China, based on the molecular biomarker SSR

Advisor: Dr. Yang Zhong

- Applied molecular techniques to detect the genetic structure and phylogeography of a cushion plant *Androsace* from Qinghai-Tibet, western China using the molecular marker AFLP

PUBLICATIONS

Botsis T, Hartvigsen G, **Chen F**, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. *AMIA Summits Transl Sci Proc.* 2010; 2010: 1–5.

Chen F, Lee K, Klein B, Klein R, Klein A, Duggal P. Exome Array Analysis Identifies Novel Susceptibility Loci with Intraocular Pressure. (In preparation)

Chen F, Lee K, Klein B, Klein R, Duggal P, Klein A. Exome Array Analysis Identifies Novel Susceptibility Loci with Refractive Error. (In preparation)

TEACHING EXPERIENCE

Teaching Assistant Principals of Genetic Epidemiology I 1st Term, 2013–2014
Johns Hopkins Bloomberg School of Public Health

Teaching Assistant Principals of Genetic Epidemiology III 3rd Term, 2013–2014
Johns Hopkins Bloomberg School of Public Health

EMPLOYMENT

09/2009 – 06/2012 Unidose Systems, Inc. Rancho Cucamonga, CA
Senior Scientist - Quality Control and Compounding Laboratory

- Led efforts as Senior Scientist to manage and direct the day-to-day quality control functions to ensure compliance with FDA cGMP guidelines and USP/NF standards
- Supervised and executed compounding activities of cosmetic and OTC products
- Conducted R&D projects in order to achieve quality or financial outcomes

LANGUAGE AND SKILLS

- Language: Chinese (native fluency)
- Computer Skills: SAS, STATA, PLINK, EIGENSTRAT, Java programming language, R programming language, Perl programming language
- Laboratory Techniques: HPLC, FTIR, PCR, electrophoresis, DNA extraction