Meta-analysis of genome-wide association studies in multiethnic Asians identifies two loci for age-related nuclear cataract

Jiemin Liao1,2,†, Xinyi Su1,2,†, Peng Chen4, Xu Wang4, Liang Xu5, Xiang Li1,6, Lennard Thean2, Clement Tan2, Ava G. Tan7, Wan-Ting Tay1, Gyungah Jun8,9,10, Yingfeng Zheng1, Merwyn Chew1, Ya Xing Wang5, Queenie S. Tan1, Veluchamy A. Barathi1,2,11, Barbara E. Klein12, Seang-Mei Saw1,2,4,11, E-Shyong Tai3,4,11, Sudha K. Iyengar13,14,15,16, Paul Mitchell7, Chiea-Chuen Khor2,17, Tin Aung1,2, Jie Jin Wang6,18, Jost B. Jonas5,19, Yik-Ying Teo4,6,†, Tien Yin Wong1,2,4,† and Ching-Yu Cheng1,2,4,11,†,*

1Singapore Eye Research Institute, Singapore National Eye Center, Singapore 168751, Singapore, 2Department of Ophthalmology, 3Department of Medicine, National University of Singapore and National University Health System, Singapore 119228, Singapore, 4Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore 117597, Singapore, 5Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Science Key Lab, Beijing 100730, China, 6Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore, 7Department of Ophthalmology, Centre for Vision Research, Westmead Millennium Institute, University of Sydney, Sydney, NSW 2145, Australia, 8Department of Medicine (Biomedical Genetics), 9Department of Ophthalmology, Boston University School of Medicine, Boston, MA 02118, USA, 10Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA, 11Duke-National University of Singapore Graduate Medical School, Singapore 169857, Singapore, 12Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI 53705, USA, 13Department of Epidemiology and Biostatistics, 14Department of Genetics and Genome Sciences, 15Center for Clinical Investigation, Case Western Reserve University, Cleveland, OH 44106, USA, 16Department of Ophthalmology and Visual Sciences, Case Western Reserve University and University Hospitals Eye Institute, Cleveland, OH 44106, USA, 17Genome Institute of Singapore, Agency for Science, Technology, and Research, Singapore 138672, Singapore, 18Centre for Eye Research Australia (CERA), University of Melbourne, Royal Victorian Eye and Ear Hospital, Melbourne VIC 3002, Australia and 19Department of Ophthalmology, Medical Faculty Mannheim of the Ruprecht-Karls-University Heidelberg, 68167 Mannheim, Germany

Received January 1, 2014; Revised and Accepted June 16, 2014

Age-related cataract is a leading cause of blindness worldwide, especially in developing countries where access to cataract surgery remains limited. Previous linkage and candidate gene studies suggested genetic influences on age-related nuclear cataract but few genetic markers have been identified thus far. We conducted genome-wide association studies on 4569 Asians (including 2369 Malays and 2200 Indians), and replicated our analysis in 2481 Chinese from two independent cohorts (1768 Chinese in Singapore and 803 Chinese in Beijing). We confirmed two genome-wide significant loci for nuclear cataract in the combined meta-analysis of four cohorts (n = 7140). The first locus was at chromosome 3q25.31 in KCNAB1 (rs7615568, fixed-effect \( P_{\text{meta}} = 2.30 \times 10^{-8} \); random-effect \( P_{\text{meta}} = 1.08 \times 10^{-7} \)). The second locus was at chromosome 21 in the proximity of CRYAA (rs11911275, fixed-effect \( P_{\text{meta}} = 2.77 \times 10^{-7} \); random-effect \( P_{\text{meta}} = 1.98 \times 10^{-9} \)), a major protein component

†These authors contributed equally to this work.

© The Author 2014. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
of eye lens. The findings were further supported by up-regulation and down-regulation of KCNAB1 and CRYAA in human lens capsule, respectively, as the severity of nuclear cataract increases. The results offer additional insights into the pathogenesis of nuclear cataract in Asians.

INTRODUCTION

Age-related cataract is the leading cause of global blindness accounting for >20 million bilaterally blind people, and an additional 94 million visually impaired people (1). Cataract most commonly develops as part of the ageing process in the crystalline lens. This causes opacification of the nuclear, cortical or posterior subcapsular regions of the lens, resulting in reduction of light transmission, thus leading to impaired vision. Of the three subtypes of age-related cataract, nuclear opacification or nuclear cataract is the most common.

In developed nations, although age-related cataract is treatable by surgical removal, cataract extraction surgery accounts for a significant proportion of the expenditure in the eye healthcare system. In the USA, the estimated annual total financial burden to the economy constitutes more than a third of total expenditure of $35.4 billion for major adult vision problems (2). In developing nations, including part of Asia, due to barriers in accessing eye care, un-operated cataracts remain by far the major cause of blindness, making up at least 50% of blindness (3) with the prevalence expected to increase with aging populations (4–6). It has been estimated that if the onset or progression of cataract was delayed by 10 years, the burden of cataract surgeries would be reduced by 50% (7). Therefore, a better understanding of the pathogenesis and genetic etiology of cataract formation is crucial for generating insights into novel strategies for prevention and treatment of cataracts.

Based on family studies (8,9), heritability of nuclear cataract was moderately high, ranging from 36 to 48%. Most of previous linkage and candidate gene studies were limited to congenital cataract, for which >20 loci have been identified, but they may not be relevant to age-related cataract (10–16). On the other hand, genetic studies on age-related cataract were relatively small-scaled and thus may not have been powered sufficiently to identify susceptibility loci.

To uncover genetic loci for age-related cataract, we performed the largest genome-wide association study (GWAS), to date, on age-related nuclear cataract. First, we conducted GWAS in two Asian populations, including the Singapore Malay Eye Study (SiMES) and Singapore Indian Eye Study (SINDI). To validate our results, we took 33 SNPs with P-values of < 10^{-3} from nine distinct genomic regions including the CRYAA locus forward in a replication analysis in two independent Chinese cohorts (SCES and BES). We observed nominally significant association in the replication cohorts with rs7615568 at the KCNAB1 locus (meta-analysis P-values: discovery = 6.00 × 10^{-7}, replication = 0.012) (Table 2). The locus at CRYAA, however, did not replicate in the two Chinese cohorts (P = 0.195).

We then performed genome-wide meta-analysis of all four cohorts (n = 7,140; Fig. 1), which confirmed the association at KCNAB1 (rs7615568, P = 2.30 × 10^{-8}), where it surpassed genome-wide significance level. Presence of each copy of the effect allele (G) was associated with ~0.09 SD increase in severity of nuclear cataract, which was equivalent to ~0.07 unit increase in Wisconsin System grading scale of nuclear cataract.

Despite the absence of replication of CRYAA in the two replication cohorts, the effects of alleles were in the same direction for all four cohorts, and the locus remains genome-wide significant in the meta-analysis of all four cohorts (rs11911275, P = 2.77 × 10^{-8}; Table 2 and Fig. 1). The overall effect size of the risk allele in the CRYAA locus was similar to that in KCNAB1. Regional association plots at the two loci are shown in Figure 2. The two top SNPs in CRYAA and KCNAB together explained 0.4% of the variance of nuclear cataract. There were no other new loci emerged from the meta-analysis of all four cohorts.

Heterogeneity was observed at the two loci (I^2 = 78.4, P_{het} = 0.003 for CRYAA; and I^2 = 66.8, P_{het} = 0.029 for KCNAB1), as reflected by the effect size and allele frequencies of each cohort. We thus also used random effects (RE2) models with genomic control for meta-analysis. The RE2 test statistic could be decomposed into a fixed effect statistic and a test statistic for testing for heterogeneity, and the inflation factors for each statistic were 1.03 and 0.76, respectively (see Materials and Methods). Accounting for random effects improved model fit slightly and the two loci remained genome-wide significant [KCNAB1 (rs7615568), P = 1.08 × 10^{-8};
Conditional analyses were performed to determine the presence of other independent association signal with nuclear cataract apart from the top SNPs on two identified loci. Supplementary Material, Table S2 showed that the leading SNP in each of the loci (rs11911275 in CRYAA and rs7615568 in KCNAB1) was sufficient to account for the individual association signals.

In the combined meta-analysis, there were 14 other loci showing suggestive associations ($P < 10^{-5}$, Supplementary Material, Table S3) with nuclear cataract. Among them, mutations in SLC4A4 were known to cause proximal renal tubular acidosis, accompanied with early onset-cataract (20), and congenital cataract has previously been reported in patients with neurologic disease associated with mutations in COL4A1 (21). The $P$-values for the genotyped SNPs from the overall meta-analysis can be found in Supplementary Data.

Pathway enrichment analysis

We performed gene-based tests using GATES (22) with genome-wide significance declared if $P_{\text{gene-based}} < 0.05 / 21,420 = 2.33 	imes 10^{-9}$ (21,420 genes tested). CRYA4 attain significance ($P_{\text{gene-based}} = 2.79 	imes 10^{-9}$), yet no additional new locus was identified. CRYA4 together with 18 other genes showing suggestive association ($P < 0.001$) were brought forward for pathway enrichment analysis using Gene Relationships Across Implicated Loci (GRAIL). GRAIL applies a statistical text-mining algorithm (23) to its database of published scientific abstracts to explore the relationships between the 19 genes. None of the genes obtained GRAIL scores of $P < 0.05$, suggesting that there was an absence of strong connections among the 19 genes.

To identify potential biological pathways associated with nuclear cataract, we applied gene set enrichment analysis (GSEA) in Meta-Analysis Gene-set Enrichment of variAnt Associations (MAGENTA) (24) to the GWAS results of all four cohorts, which revealed three gene sets showing significant
enrichment. The strongest enrichment was observed for the positive regulation of MAPK cascade \(\text{MAGENTA P}_{\text{GSEA}} = 2.2 \times 10^{-5}\). Enrichment was also observed in the insulin signaling pathway and protein phosphorylation process (both \(\text{MAGENTA P}_{\text{GSEA}} = 1.0 \times 10^{-4}\)).

**Functional annotation**

We investigated the potential enrichment of functional significance of our top SNPs as well as SNPs in linkage disequilibrium with them (LD, \(r^2 > 0.8\)) using the Encyclopedia of DNA Elements (ENCODE) data sets \(^{(25)}\) (Supplementary Material, Table S4). The results indicated that the sentinel SNP on \(\text{KCNAB1}\) was located in a DNase site while LD SNPs altered regulatory motifs. All of the SNPs near \(\text{CRYAA}\) were located in a region with potential regulatory motifs, such as DNase I hypersensitivity, modification of histone marks and binding of transcription factors.

**Gene expression**

We measured and compared the mRNA levels of \(\text{CRYAA}\) and \(\text{KCNAB1}\) in lens capsule from 30 patients with cataract (Fig. 3). The severity of nuclear cataract was assessed under slit-lamp using the Lens Opacities Classification System III (LOCS III) \(^{(26)}\), where a digital score was assigned according to the severity of nuclear opacification (NO1 to NO6), with NO6 being the most severe to NO1 being unaffected. The mean age of the included participants was 68.20 ± 8.8 years. Quantitative PCR showed that \(\text{CRYAA}\) mRNA expression decreased as the severity of nuclear cataract increased (by 24.3-fold in NO5 and NO6 lens, compared with NO3, \(P < 0.01\); \(P\) for trend <0.001 across NO3–NO6). Conversely, \(\text{KCNAB1}\) expression increased as the severity of nuclear opacification increased (84.4-fold increase in NO6 compared with NO3, \(P < 0.01\); \(P\) for trend =0.009 across NO3–NO6).

**DISCUSSION**

We identified two susceptibility loci \((\text{KCNAB1} \text{ and } \text{CRYAA})\) for age-related nuclear cataract. Our findings were substantiated by the presence of potential regulatory motifs at site of SNPs as observed from the functional annotation using ENCODE data. This is in addition to the observed up-regulation and down-regulation in mRNA expression of \(\text{KCNAB1}\) and \(\text{CRYAA}\), respectively, in lens capsular tissue, as the severity of nuclear cataract increases. To the best of our knowledge, this is the first GWAS of age-related nuclear cataract conducted to date.

The first locus, \(\text{KCNAB1}\), encodes a member of the potassium channel, voltage-gated, shaker-related subfamily. It includes three distinct isoforms encoded by three alternatively spliced transcript variants. These isoforms form the beta subunits, which in turn form heteromultimeric α4β4 complexes with alpha subunits, and modulate their activity. Voltage-gated potassium channels play an important role in determining membrane excitability, and are involved in nervous system signal transduction pathways and other physiological functions such as generation of atrial repolarization in cardiac tissue \(^{(27)}\). Interestingly, in patients with type 2 diabetes, another member of the potassium voltage-gated channel protein \(\text{KCNQ1}\) has been found to be associated with cataracts. This has led to the postulation that conduction of K\(^+\) ions across cellular membranes, affects the intracellular concentration of iron, which in turn leads to the denaturation of lens crystalline and accelerated formation of cataracts. This was further confirmed in the observed proportional increase in mRNA expression of...
Figure 2. Regional association plots from the meta-analysis of four cohorts at the (A) KCNAB1 and (B) CRYAA loci. Log_{10} P-values are plotted as a function of genomic position (NCBI Build 37). Data of both directly genotyped and imputed SNPs are presented. Purple diamond indicates the P-value for the lead SNP in each locus. The LD (r^2) between the lead SNP and neighboring SNPs are represented by the shading of the circles by color. Recombination rates of ASN panel from 1000 Genomes March 2012 release are represented by the blue lines. The plots were created using LocusZoom (http://csg.sph.umich.edu/locuszoom/ last accessed 22 April 2014).
**CRYAA** encodes the \( \alpha \)-crystallin chain protein—a major protein component of the human lens, and contributes to the clarity and refractive properties of the lens (32). Studies have shown that \( CRYAA \) is down-regulated in the human lens epithelia of age-related nuclear cataract, due to DNA methylation of \( CRYAA \) promoters, which epigenetically suppresses \( CRYAA \) expression (33). The chaperone-like activity of \( \alpha \)-A-crystallin is considered to be critical for the maintenance of eye lens transparency (33). A decrease in \( CRYAA \) expression is likely to result in lens opacification, which is consistent with our gene expression results where we observed a dose–response relationship between mRNA level and the severity of age-related nuclear cataract. Other diverse roles of alpha-crystallins include regulation of cell growth, modulation of resistance to stress and enhancement of genomic stability. These in vivo functions may also lead to formation of cataract when they are altered (32).

Although GRAIL analysis with our top genes did not produce any candidate pathways based on its text algorithms, our analyses using MAGENTA highlighted the positive regulation of MAPK cascade, insulin signaling and protein phosphorylation as biological pathways and processes related to cataract. The association between MAPK cascade protein upregulation and cataract formation had been previously reported (34,35), while insulin and insulin-like growth factor had been shown to stimulate phosphatidylinositol 3-kinase in the lens and affect lens epithelial cell proliferation and differentiation (36).

One caveat of this study includes significant heterogeneity of associations at \( KCNAB1 \) and \( CRYAA \) across the study cohorts. The strongest association was observed in the Malay cohort, while the association is relatively weaker in the two Chinese cohorts. Inter-ethnic differences in LD pattern observed in the LD and VarLD (37) plots of \( KCNAB1 \) could contribute to this observed heterogeneity (Supplementary Material, Fig. S4). The reference cohorts used were the East Asians and Gujarati Indians cohorts from HapMap 3 (http://hapmap.ncbi.nlm.nih.gov), and 100 whole-genome sequenced Malays from Singapore (38). There was no evidence for differences in LD at the \( CRYAA \) locus, suggesting that the differences in the effect size and association could be attributed to differences in inter-ethnic genetic heterogeneity or phenotype measurement (i.e. cataract grading systems). For \( KCNAB1 \), different LD patterns among ethnic groups may result in the observed heterogeneity in part.

Several cataract grading systems have been developed to measure the severity of cataract and the adoption of a grading system mainly depends on the availability of resources. In our studies, the Wisconsin System is regarded as a more precise measurement than other grading systems, as the grading was performed by a single experienced grader based on subjects’ lens photos through comparison with standard lens photos. Despite performing phenotype standardization to combine the results from multiple grading schemes for current studies, there are potentially varied levels of measurement error across the grading schemes. The application of meta-analysis using standard error weights account for differences in error of measurement between studies by allocating more weight to studies with more precise estimates. The presence of heterogeneity in associations between the Malay and Indian cohorts, which were both graded using the same grading protocol (the Wisconsin System), suggests that the inter-ethnic differences could be a main contributor to the observed heterogeneity, instead of phenotype measurement error introduced through different grading systems.

---

**Figure 3.** \( CRYAA \) and \( KCNAB1 \) expression in the lens capsule of age-related nuclear cataract. No. of lens capsules included were NO3 = 10, NO4 = 12, NO5 = 4, and NO6 = 4. Error bars indicate standard deviations of gene expression levels. (A) Expression of \( KCNAB1 \) increased with increasing severity of cataract. (B) Expression of \( CRYAA \) decreased with increasing severity of cataract. \( GAPDH \) was co-amplified as an internal control.
In conclusion, through a meta-analysis of GWAS on age-related nuclear cataract in multi-ethnic Asian populations, we identified a novel genetic locus KCNAB1, which could be a potential therapeutic target for delaying cataract formation or progression albeit with additional validation studies. Heterogeneity among ethnic groups at KCNAB1 could be attributed to the differences in LD patterns. Inherent ethnic-specific variation in the genetic etiological process under-pining nuclear cataract formation and differences in methods of cataract grading between studies may partly account for the variation despite standardized phenotypes were used in analysis. Identification of the association between CRYAA and nuclear cataract reaffirms the validity of our results.

MATERIALS AND METHODS

Ethics statement

Ethics approval was obtained from the Singapore Eye Research Institute Institutional Review Board and from the Medical Ethics Committee of the Beijing Tongren Hospital for all Singapore studies and BES, respectively. All study participants were provided with written informed consent in adherence to the Declaration of Helsinki.

Study discovery cohorts

Participants in SiMES and SINDI were included as discovery cohorts in the present study, because the two cohorts followed the same cataract grading system, while the two replication cohorts used different grading systems (see ‘Nuclear cataract grading’ below).

Singapore Malay eye study

SiMES is a population-based cross-sectional epidemiological study on eye diseases for Malays aged between 40 and 80 years old residing in Singapore. Details of the SiMES design, sampling plan and methodology have been reported elsewhere (39). In brief, between August 2004 and June 2006, a total of 4168 Malay residents in the southwestern part of Singapore were identified through age-stratified random sampling, of which 3400 participants (75.6% response rate) were recruited. We performed genome-wide genotyping in 2953 SINDI subjects.

Replication cohorts

Singapore Chinese eye study

SCES is a population-based cross-sectional epidemiological study on eye diseases for Chinese aged between 40 and 80+ years old residing in Singapore. Using age-stratified random sampling strategy, 4605 ethnic Chinese residents in the southwestern part of Singapore were eligible from the sampling frame (n = 6752), of which 3353 (72.8% response rate) participants were recruited between February 2009 and December 2012 (40). Genome-wide genotyping was done in a subset of 1952 SCES participants.

Beijing eye study

BES is a population-based cross-sectional study of Chinese aged 40+ years residing in the village area of Yufa in Daxing District, south of Beijing, and in the Haidian urban district, north of Central Beijing. At the time of the survey in 2001, a total of 5324 individuals were eligible to participate in the study, of which 4439 individuals (83.4% response rate) were recruited. In 2006, all participants from the survey in 2001 were re-invited and 3251 participants (73.2% response rate) were recruited, and blood was taken from 2,929 (90.1%) (41). We performed genome-wide genotyping in 988 subjects.

Phenotype grading

In SiMES and SINDI, the severity of nuclear cataract was assessed using the Wisconsin Cataract Grading System (Wisconsin System) (42), based on lens photographs and followed a decimalized system (decimal scores from 0.1 to 5.0). In brief, lens photographs were taken using a digital slit-lamp camera (model DC-1 with DF-21 flash attachment; Topcon, Tokyo, Japan) and grading was performed through comparison with four standard photographs at the University of Sydney by a single experienced grader (AGT), with adjudication by a senior ophthalmologist (P.M.) and a senior researcher (J.J.W.) (43).

In SCES, nuclear cataract was assessed using the LOCS III (26) and followed a decentralized system (decimal scores from 0.1 to 6.9). Participants went through slit lamp bio-microscopy where nuclear cataract was graded with light focused in the center of the nucleus. Trained study ophthalmologists performed comparison with standard slide transparency of the LOCS III scale to assess the grading.

In BES, the degree of nuclear cataract was assessed in 6 grades using the grading system of the Age-Related Eye Disease Study (AREDS), which is an extension of the Wisconsin System (44). Digital photographs of the lens were obtained using the slit lamp. A single grader under the supervision of senior ophthalmologists and researchers performed the lens grading (45). The nuclear cataract grading from the 2006 follow-up study was used.

Genotyping and imputation

For all studies, the Illumina Human610-Quad Beadchips (Illumina Inc.) were used for genotyping. Detailed data quality control (QC) procedures for each study were described elsewhere (46,47). In brief, a first round of SNP QC was performed to obtain a cleaned set of genotypes for sample QC by excluding SNPs with (1) high missingness (>5%); (2) gross departure from
HWE ($P < 10^{-6}$); and (3) were monomorphic. Samples were excluded based on the following conditions: (1) sample call-rates of less than 95% and excessive heterozygosity; (2) cryptic relatedness; (3) gender discrepancies; and (4) discordant ethnic memberships. PLINK software (version 2.0) (48) was used to derive identity-by-state information to identify related samples and EIGENSTRAT program (49) was used to perform principle components analysis. A final round of SNP QC was applied to the remaining samples on autosomal SNPs using the same three criteria earlier, and an additional restriction of minor allele frequency $>5\%$. The final sample size with genotype and phenotype data (i.e. nuclear cataract grading, age and gender) for discovery and replication cohorts were 4569 and 2571, respectively (Table 1).

Genome-wide imputation of SNPs was performed on post-QC set of genotype data, using minimac (50) with 1000 Genome Project (51) Phase 1 version 3 release (http://www.1000genomes.org). The reference panel included 1092 samples from 14 populations and 38 million variants on build 37 (52). The same SNP QC procedure was performed on the imputed SNPs. In addition, for association testing, only imputed SNPs that had information score (info) $> 0.5$ were included.

**Per-SNP association analysis**

Before association testing, the nuclear cataract phenotype was standardized within each cohort by subtracting the mean and dividing by the standard deviation. After standardization, the distribution of nuclear cataract was centered with mean of zero and variance of one for all cohorts. The SNPTEST software (version 2.3.0) was used to test for association between standardized nuclear cataract and SNPs, where $F$-test and score test were performed on genotyped and imputed SNPs, respectively. We assumed an additive genetic model where the dosage of each SNP is a continuous variable ranging from 0 to 2 for minor alleles carried, and standardized nuclear cataract was treated as a quantitative trait, adjusting for age, gender and genetic principal components. The effect of each genotype on the phenotype was assumed to be additive and the genotype was incorporated in the model as a trend test.

Meta-analyses for the discovery and replication cohort were performed using Metasoft software (53). The primary meta-analysis was implemented using the inverse-variance, fixed effects model. SNPs that showed suggestive association ($P \leq 10^{-5}$) for the model were taken forward for replication. Between-study heterogeneity was indicated by $P$-values of $<0.1$ from the Cochran’s $Q$ test (54). For SNPs showing evidence of heterogeneity, Han and Eskin’s random effects (RE2) model was used. The RE2 model increases the power of association testing in the presence of heterogeneity compared with the conventional DerSimonian—Laird random effects (RE) model. Genomic control was applied to the results of RE2 as the model is sensitive to small confounding (53). The statistic under the RE2 model can be decomposed into the conventional fixed effect statistic and the test statistic used for testing heterogeneity. Each statistic was adjusted by an inflation factor to correct for confounding as given in the following equation:

$$S_{RE2} = S_{FE}/\lambda_{FE} + S_{Het}/\lambda_{Het}$$

where $S_{RE2}$, $S_{FE}$ and $S_{Het}$ are test statistics for the RE2 model, fixed effect model and testing of heterogeneity, respectively. $\lambda_{FE}$ and $\lambda_{Het}$ are the inflation factors.

Random-effects meta-analysis assumes that each population has a different underlying allelic effect. However, some ethnic groups with closely related ancestry may have similar allelic effects. To model this, we utilized MANTRA software to perform a trans-ethnic meta-analysis (17). MANTRA clusters populations according to a Bayesian partition model of relatedness, as defined by pairwise genome-wide mean allele frequency differences. Evidence in favor of association of the phenotype at each SNP is measured by a BF, where a log$_{10}$ BF 6 or higher is considered a conservative threshold for genome-wide significance.

**Pathway enrichment analysis**

Gene-based association tests were performed using the extended Simes procedure (GATES) (22) using an open source tool, Knowledge-based mining system for Genome-wide Genetic studies (KG3, version 2.5) (55). Each SNP was annotated to each of 21420 genes according to the gene coordinate information from NCBI hg19 assembly, with 5 kb boundary on either side of gene. To account for heterogeneity across cohorts, gene-based tests were performed within each cohort and the results were meta-analyzed using Fisher sum statistic with $\chi^2$ distribution. The $r^2$ values from the 1000 Genome Phase 1 version 3 release cosmopolitan reference panel with 1092 samples were used to adjust for marker dependency in each study.

We implemented two methods to assess presence of enriched pathways in our GWAS data. (i) Genes reaching a $P$-value of $<0.001$ were analyzed by GRAIL to explore similarities in PubMed articles, published in or before December 2006. GRAIL was run with gene size correction on, utilizing data from HapMap release 22 JPT+CHB samples. Genes with a nominal GRAIL similarity $P < 0.05$ were considered to be significant. (ii) GWAS results of all four cohorts were analyzed by MAGENTA, where each gene was scored based on a single index SNP with the smallest $P$-value within a 110 kb upstream, 40 kb downstream boundary. The gene score was then adjusted for confounding factors such as gene size and LD-related properties. Adjusted $P$-values were ranked and the 75th percentile of association $P$-values from all genes in the genome was used to determine the enrichment cutoff, based on the assumption of high polygenicity of nuclear cataract. The observed number of genes with a ranked score above the cutoff was calculated in each pathway. This observed statistic was then compared with 10000 randomly permuted pathways of identical size to generate an empirical GSEA $P$-value for each pathway. We reported pathways which remained significant under a conservative Bonferroni correction of the GSEA $P$-values.

**Functional annotation**

We annotated SNPs in LD with the sentinel SNP ($r^2 > 0.8$) using HaploReg (56) and ENCODE data and sought potential regulatory evidence of promoter and enhancer histone modification signals, DNase I hypersensitivity, binding of transcription factors or effects on regulatory motifs.
Gene expression in human lens epithelium

Lens epithelium samples were obtained from eyes during cataract surgery at the National University Hospital, Singapore. All samples used in this study were collected with the informed consent of the patients, who were >50 years of age. Ethics approval was obtained from the National Healthcare Group Domain Specific Review Board.

All patients had a complete preoperative ophthalmologic examination. The type and severity of cataracts were assessed and graded according to the modified version of LOCS III by senior ophthalmologists. Intra-operatively, lens epithelium samples were obtained by intact continuous curvilinear capsulorhexis. Care was taken to avoid vascular contact or damage to the iris or other intraocular structures. After removal, lens epithelia were rapidly frozen in liquid nitrogen for storage at the National University Health System Tissue Repository until retrieval for gene expression analysis.

Lenses with a LOCS III NO score of 3 to 6 (n = 30) were used to measure the RNA expression of the identified genes. Lenses of the same grade (NC3 = 10, NC4 = 12, NC5 = 4 and NC6 = 4) were pooled cryogenically and total RNA isolated using TRIzol Reagent (Invitrogen, Carlsbad, CA). The first strand of cDNA was synthesized with 500 ng of total RNA, oligo(dT) 15 primer and AMV reverse transcriptase (Promega, Madison, WI, USA). Real-time PCR reactions were performed with KAPA SYBR FAST Roche Lightcycler 480 2× qPCR Master Mix (Kapa Biosystems, GenomeHoldings, Singapore). The specificity of the PCR amplification products was checked by performing dissociation melting curve analysis and visualization on a 1% agarose gel. Quantification of the identified genes forming dissociation melting curve analysis and visualization of the PCR amplification products was checked by performing linear trend tests on the severity of NO.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

The authors thank the staff and participants of all studies for their important contributions. The Singapore Tissue Network and the Tissue Repository of the National University Health System, Singapore provided services for tissue archival. The Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore provided services for genotyping.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Medical Research Council (NMRC), Singapore (0796/2003, IRG07nov013), IRG09nov014, NMRC 1176/2008, STA/R/0003/2008, CG/SERI/2010, CSA/033/2012 to C.C., CSA/008/2009 to E.T.); Biomedical Research Council (BMRc), Singapore (08/1/35/19/550, 09/1/35/19/616); and the National Natural Science Foundation of China (81170890).

REFERENCES


