Collagen-related genes influence the glaucoma risk factor, central corneal thickness

Eranga N. Vithana1,3,∗, Tin Aung1,2,3,†, Chiea Chuen Khor6,7,‡, Belinda K. Cornes1, Wan-Ting Tay1, Xueling Sim7, Raghavan Lavanya1, Renyi Wu1, Yingfeng Zheng1, Martin L. Hibberd6, Kee Seng Chia4,7, Mark Seielstad8, Liang Kee Goh9, Seang-Mei Saw1,3,4,7, E. Shyong Tai4,5,7,§ and Tien Y. Wong1,3,4,10,¶

1Singapore Eye Research Institute and 2Singapore National Eye Centre, Singapore, Singapore, 3Department of Ophthalmology, 4Department of Epidemiology and Public Health and 5Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, 6Infectious Diseases, Genome Institute of Singapore, A∗STAR, Singapore, Singapore, 7NUS-GIS Center for Molecular Epidemiology, Singapore, Singapore, 8Department of Laboratory Medicine, Institute for Human Genetics, University of California, San Francisco, USA, 9Duke-NUS Graduate Medical School, Singapore, Singapore and 10Centre for Eye Research Australia, University of Melbourne, Victoria, Australia

Received September 13, 2010; Revised November 10, 2010; Accepted November 18, 2010

Central corneal thickness (CCT) is a risk factor of glaucoma, the most common cause of irreversible blindness worldwide. The identification of genetic determinants affecting CCT in the normal population will provide insights into the mechanisms underlying the association between CCT and glaucoma, as well as the pathogenesis of glaucoma itself. We conducted two genome-wide association studies for CCT in 5080 individuals drawn from two ethnic populations in Singapore (2538 Indian and 2542 Malays) and identified novel genetic loci significantly associated with CCT (COL8A2 rs96067, \(p_{\text{meta}} = 5.40 \times 10^{-13}\), interval of RXRA-COL5A1 rs1536478, \(p_{\text{meta}} = 3.05 \times 10^{-9}\)). We confirmed the involvement of a previously reported gene for CCT and brittle cornea syndrome (ZNF469) [rs9938149 (\(p_{\text{meta}} = 1.63 \times 10^{-16}\)) and rs12447690 (\(p_{\text{meta}} = 1.92 \times 10^{-14}\))]. Evidence of association exceeding the formal threshold for genome-wide significance was observed at rs7044529, an SNP located within COL5A1 when data from this study (\(n = 5080, \ P = 0.0012\)) were considered together with all published data (reflecting an additional 7349 individuals, \(P_{\text{Fisher}} = 1.5 \times 10^{-9}\)). These findings implicate the involvement of collagen genes influencing CCT and thus, possibly the pathogenesis of glaucoma.

INTRODUCTION

Glaucoma is the leading cause of irreversible blindness worldwide affecting more than 60 million people (1). Although raised intraocular pressure (IOP) has traditionally been thought to be the major risk factor for glaucoma, emerging evidence suggests that central corneal thickness (CCT) could also be a predictor of glaucoma development, particularly in individuals with higher baseline IOP (2,3). Interestingly, this relationship is independent of the known confounding effect of CCT has on intra-ocular pressure measurements (3). Studies show that persons with thinner corneas are more likely to develop primary open-angle glaucoma (POAG), independent of age, IOP and other factors (2,4–8). Thus, it has been proposed that CCT may represent an intermediate trait or endophenotype for POAG, and CCT may share genes with POAG that may be more easily identified through investigations of CCT as a continuous trait rather than POAG disease as a binary trait (9). However, a genetic correlation

∗To whom correspondence should be addressed at: Singapore Eye Research Institute, 11 Third Hospital Avenue, Singapore 168751, Singapore. Tel: +65 63224542; Fax: +65 63224599; Email: eranga.n.v@seri.com.sg
†The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors.
‡These authors are joint last authors.

© The Author 2010. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
between CCT and glaucoma has not yet been clearly demonstrated. A recent bivariate genetic analysis could not show a substantial overlap of genes that jointly influence variation in CCT and POAG disease risk, likely due to a lack of sample power in this study (10). The inconsistencies observed thus far between CCT and POAG highlight the importance of identifying the genes underlying CCT that may help elucidate the relationship between glaucoma and CCT that has been suggested to exist beyond the influence CCT exerts over IOP measurement.

Thin corneas are also an important clinical feature of other visually debilitating diseases, such as keratoconus (MIM 148300), another complex and heterogeneous disorder that is one of the leading indications for corneal transplantation worldwide (11). Keratoconus normally develops by progressive dissolution of Bowman’s layer, which lies between the corneal epithelium and stroma, leading to the bulging and scarring of the cornea that are characteristic of the disorder. Therefore, identifying genes responsible for normal variation in CCT may also provide insights into pathogenesis of corneal diseases such as keratoconus.

Based on twin and familial studies, CCT, which is a normally distributed quantitative trait within the general population, displays high heritability with a coefficient as high as 0.95 (12–14). Despite the strong evidence of its genetic etiology, the genetic determinants of CCT remained largely unknown until recent genome-wide association studies (GWAS) conducted on European populations led to the identification of several genes (15,16). Prior to this, only a few candidate gene studies based on rare genetic diseases such as osteogenesis imperfecta (MIM 259420), Marfan syndrome (MIM 154700) and aniridia (MIM 106210) with extreme values of association with CCT remained largely unknown until recent genome-wide association studies (GWAS) conducted on European populations led to the identification of several genes (15,16). Prior to this, only a few candidate gene studies based on rare genetic diseases such as osteogenesis imperfecta (MIM 259420), Marfan syndrome (MIM 154700) and aniridia (MIM 106210) with extreme values of association with CCT remained largely unknown until recent genome-wide association studies (GWAS) conducted on European populations led to the identification of several genes (15,16).

As with any quantitative trait, CCT is likely to be determined by multiple genes of various effect sizes. Furthermore, it is not known whether these genes identified in largely Caucasian populations are relevant to Asians, in which the prevalence of glaucoma and possibly keratoconus is more common than in individuals of Caucasian ancestry. Given the clear evidence of ethnic-dependent differences in CCT (20–22), we aimed to identify novel loci that may explain the variations observed for this important ocular parameter in Asians. We therefore conducted two population-based GWAS on 5080 individuals drawn from two Asian ethnic populations in Singapore: 2538 Indian subjects aged 40 years and older [Singapore Indian Eye Study (SINDI)], part of the Singapore Indian Chinese Cohort Eye Study (SiCC)] (23) and 2542 Malay subjects also aged 40 years and older [the Singapore Malay Eye Study (SiMES)] (24–26).

RESULTS

Study samples

The GWAS cohorts described here—SINDI and SiMES—included 2538 and 2542 participants passing quality control (QC) checks, respectively. Table 1 summarizes the descriptive statistics for both cohorts. The SINDI GWAS included a total of 559 119 autosomal SNPs which passed QC. Supplementary Material, Figure S1, shows the quantile–quantile (Q–Q) and Manhattan plots for the directly genotyped SNPs in the SINDI cohort. Similarly, the SiMES GWAS had a total of 557 824 SNPs passing QC. Supplementary Material, Figure S2, shows the Q–Q plot and the accompanying Manhattan diagram for directly genotyped SNPs in the SiMES cohort. The Q–Q plots for both SINDI and SiMES show no deviation of the observed P-values against their expected distribution, except at the extreme tail, suggesting that these deviations at the tail end of the distribution could reflect true associations with CCT. The SINDI cohort was used as a main ‘discovery’ sample set, and the SiMES cohort as the ‘replication’ sample for looking up significant associations (P < 5.00 × 10−8) from SINDI. Principal component (PC) analysis of SINDI and SiMES show no deviation of the observed P-values against their expected distribution, except at the extreme tail, suggesting that these deviations at the tail end of the distribution could reflect true associations with CCT. The SINDI GWAS, as well as results obtained from the replication SiMES cohort and meta-analysis of both cohorts, are provided in Table 2.

Single-locus analysis in SINDI revealed the strongest evidence of association with CCT with a cluster of SNPs on chromosome 16q24 near the ZNF469 gene, with the most significant SNP being rs9938149 (P = 8.77 × 10−11). Regional association and linkage disequilibrium (LD) plots at the chromosome 16 ZNF469 locus are shown in Figure 2. The presence of each copy of the minor allele of rs9938149 was associated with 6.29 mmHg decrease in CCT. This SNP, located ~108 kb upstream of the ZNF469 gene, was also found to associate with CCT (Pmeta = 1.08 × 10−07) in the discovery sample of the two twin cohorts from Australia and the UK, but was not subsequently replicated in the Australian population-based samples with extreme CCT values (15). Notably, we also confirmed association with the previously reported SNP rs12447690 (SINDI P = 5.23 × 10−09), located ~140 kb upstream of the ZNF469 gene, with each...
copy of the minor allele associated with a 5.33 μm decrease of CCT. This SNP was associated with an overall meta-analysis P-value of 8.95 \times 10^{-11} in the Australian–UK study (15). Associations with both these SNPs (rs9938149 and rs12447690) near ZNF469 were replicated in the SiMES meta-analysis finding (\(P = 5.1 \times 10^{-8}\)) after meta-analysis with the Australian–UK study (combined \(n = 7349\)). Data from both SINDI and SiMES further support association for this SNP (\(P = 5.1 \times 10^{-8}\)) at the intergenic interval between RXRA and COL5A1, which supports the genome-wide significant observation by Vitart et al. (16), who observed the same trend toward association in their separate study (per-allele effect size: \(\sim 3.3 \mu m, P = 0.019\), as well as highly suggestive evidence of association (per-allele effect size: \(\sim 4.3 \mu m, P = 5.1 \times 10^{-8}\)).

As there appeared to be two independent quantitative trait loci (QTL) within the chromosome 9q34 region (one in the interval between RXRA and COL5A1, and the other within COL5A1), we performed conditional analysis on both loci. The signals observed at both loci did not change significantly when modeled mutually for the effects of allele dosage, suggesting both QTL to be independent of one another (Supplementary Material, Table S5).

Supplementary Material, Tables S6 and S7, tabulates data from the conditional logistic regression in the pooled Singaporean cohort (\(n = 5080\)) and reports P-values of remaining SNPs after the conditioning of the top chromosome 1 SNP rs96067 and the top chromosome 16 SNP rs9938149, respectively. For both loci (COL5A2 and ZNF469) identified in our report, conditional logistic regression demonstrates that the leading SNPs at each locus (rs96067 and rs9938149, respectively) are sufficient to account for the individual association signals, further confirmed by haplotype analysis which did not show evidence of statistical significance beyond that observed with single markers (data not shown).

On further scrutiny of the SINDI and SiMES GWAS data, rs768787, mapping to AKAP6 on chromosome 14, was found to be marginally associated with CCT at \(P = 2.45 \times 10^{-7}\) (per-allele effect size: \(5.315 \mu m\) in SINDI). A much weaker trend was seen in SiMES (\(P = 0.11\), per-allele effect size: \(2.038 \mu m\)). We then proceed to examine GWAS data for evidence of association with CCT at three previously described loci (FOXO1, AVGR8 and AKA13; Supplementary Material, Table S8). We were able to observe association at AKA13 rs6496932 (per-allele effect size: \(2.2 \mu m, P_{meta} = 0.0013\)), which supports the genome-wide significant observation by Vitart et al. at the same SNP. No evidence of association was observed at FOXO1 and AVGR8 (\(P > 0.05\) for all SNPs across all cohorts).
Expression of COL8A2

Figure 4 shows the reverse transcription polymerase chain reaction (RT-PCR) results for COL8A2 in a variety of normal human ocular tissues. The expected 61 bp RT-PCR product, indicating COL8A2 expression, was observed in the cornea, sclera, retina and retinal pigment epithelium (RPE), iris, lens capsule and in the optic nerve. The presence of the ubiquitously expressed beta-actin (ACTB) gene, seen as a 97 bp RT-PCR fragment, was also confirmed for each sample, thus verifying RNA quality from tissues sampled from a post-mortem human eye. This result adds to the COL8A2 expression data (Hs.353001) in UniGene (http://www.ncbi.nlm.nih.gov/UniGene/), which indicate COL8A2 to be expressed in the RPE, choroid, retina, eye anterior segment as well as in a variety of non-ocular tissues.

DISCUSSION

In this study, we identified two novel genetic loci (1p34.3 near COL8A2, and 9q34 at the interval between RXRA and COL5A1) associated with CCT and confirmed association at the previously identified ZNF469 gene. Our data also provide additional support for the direct involvement of COL5A1, where an SNP within the gene proper was observed to show suggestive evidence of association with CCT in the Australian–UK study (15), as well as in a meta-analysis of three Croatian and a separate UK cohort (16). Pooling all available data (overall n = 12 429), evidence of association exceeding the formal threshold of genome-wide significance was observed between a COL5A1 SNP (rs7044529) and CCT. Conditional regression analysis suggests the presence of two independent association signals at the chromosome 9q34 locus, thus necessitating further fine-mapping, sequencing and molecular functional studies to elucidate the functional variant(s) causally related to CCT.

The combined data across Asians and individuals of European descent support ZNF469 as an important locus for CCT across diverse ethnic groups, with a very consistent per-allele effect size in Europeans and Asians (Supplementary Material, Table S9). ZNF469 is the causative gene of brittle cornea syndrome [BCS (MIM 229200)], an autosomal recessive disorder characterized by a thin cornea with a tendency to perforate, causing progressive visual loss and blindness (27). The protein encoded by ZNF469 has 30% homology to the helical parts of highly expressed corneal collagens, COL1A2 (MIM 120160), COL4A1 (MIM 120130) and COL1A1 (MIM 120150) and is proposed to function either as a nuclear transcription factor or as an extra-nuclear regulatory molecule involved in the synthesis and/or organization of these collagen fibers (27). These purported functions within the cornea make ZNF469 a good biological candidate for further study in the control of the normal variation of CCT.

For the chromosome 1 hit region, the strongest signals of association with CCT were found in the area \(\approx 1.5–6\) kb upstream of the short-chain collagen gene, COL8A2.
Table 2. Directly genotyped SNPs that exceed genome-wide significance in the meta-analysis of SINDI and SIMES GWAS

<table>
<thead>
<tr>
<th>SNP Chr Position</th>
<th>SINDI</th>
<th>SIMES</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>p-value</td>
<td>Z-value</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>C</td>
<td>SE</td>
</tr>
<tr>
<td>rs9393419</td>
<td>16</td>
<td>86859014</td>
<td>0.17</td>
</tr>
<tr>
<td>rs2470790</td>
<td>16</td>
<td>86859014</td>
<td>0.28</td>
</tr>
<tr>
<td>rs9930231</td>
<td>16</td>
<td>86859014</td>
<td>0.45</td>
</tr>
<tr>
<td>rs3500479</td>
<td>16</td>
<td>86859014</td>
<td>0.21</td>
</tr>
<tr>
<td>rs5597702</td>
<td>5</td>
<td>28769010</td>
<td>0.23</td>
</tr>
<tr>
<td>rs6305697</td>
<td>16</td>
<td>86859014</td>
<td>0.23</td>
</tr>
<tr>
<td>rs2300782</td>
<td>16</td>
<td>86859014</td>
<td>0.23</td>
</tr>
<tr>
<td>rs1409383</td>
<td>9</td>
<td>13656982</td>
<td>0.23</td>
</tr>
</tbody>
</table>

(a relatively small gene spanning 5 kb), dropping by a few orders of magnitude scanning from 5′ to 3′ of COL8A2 (Table 2, Fig. 3). We are mindful of the extensive LD across the SNPs genotyped across COL8A2 (r² > 0.75), rendering it challenging to discern the true functional variants from this GWAS exercise alone. Statistical evidence of this study, however, points to the 5′ regulatory region of COL8A2 as the suggestive area harboring the actual CCT-modulating variant(s).

Posterior polymorphous corneal dystrophy [PPCD (MIM 122000)] is a rare disease involving metaplasia and overgrowth of corneal endothelial cells that increase the risk of glaucoma, due to such aberrant growth occurring at times over the iris and nearby structures of the anterior segment. It is thus not surprising that rare mutations in the novel CCT locus, COL8A2, have been reported in patients with PPCD as well as in another posterior corneal layer dystrophy, Fuchs’ endothelial corneal dystrophy (MIM 136800) (28–31). The transcription factor 8 gene (TCF8, also known as AREG6 or ZEB1) which encodes a two-handed zinc-finger homeodomain transcription factor, is also responsible, independently, for nearly half of the cases of PPCD (32). Interestingly, a search of the 20 kb of sequence, upstream of the transcription start of COL8A2, identified two TCF8 core binding sites [(T/G)(T/G)CACCT(T/G)T] (33) in COL8A2 (32), suggesting the potential for these two genes to interact and regulate each other’s activity. This information taken together with our novel finding that the 5′ regulatory region of COL8A2 may harbor CCT-modulating variants immediately suggests the possibility of transcriptional control of COL8A2 expression by TCF8 and thus CCT modulation. This strongly argues for detailed functional analysis of TCF8-COL8A2 interaction to evaluate the biological significance of all these observations.

There is a large body of evidence to support COL8A2 and COL5A1 as candidates for regulating CCT. COL5A1 mutations are found in patients with Ehlers–Danlos syndrome (EDS) type I (MIM 130000) and type II (MIM 130010), which have thin corneas as part of their clinical presentation (34,35). Moreover, mouse models with alterations in genes Col8a2 and Col5a1 also have thinner corneas (36,37). In the haploinsufficient Col5a1−/− mouse cornea, type V collagen content was reduced by ~49%, and cornea stromal thickness was reduced by ~26%. In the Col8a2−/− mice, a heterozygous G257D mutation resulted in increased anterior chamber depth, extended axial length and reduced thickness of corneal layers. This lends credence to the idea that rare severe/deleterious mutations in these genes result in severe/extreme phenotypes, whereas milder variants in the same genes (whose surrogates are detected by the GWAS methodology) are associated with normal variations of CCT.

Our GWAS has managed to identify and/or confirm the presence of genetic loci that appear to modulate CCT independently of ethnicity (e.g. ZNF469, COL5A1, the interval between RXRA and COL5A1, and AKAP13). Other loci (e.g. FOXO1 (13q14.1) and AVGR8 (13q12.11)) (15,16) appear to be more confined within individuals of European descent, suggesting that distinct genetic mechanisms governing CCT may exist for individuals of different ethnic groups, similar to what has been observed for other diseases (38,39).
Other reasons for observing these ethnic-specific effects include sampling variance and possible differences in LD patterns between Europeans and Asians. As the GWAS methodology does not usually allow for direct testing of causal variants, non-homogenous findings across cohorts of distinct ethnicity could reflect differences in local LD patterns. Nevertheless, some clues to the possible ‘ethnic specific’ effects are being shown in **COL8A2** and **RXRA-COL5A1**, where effect sizes are seen to differ between Indians and Malays. Though significant in both cohorts, the association signal observed for **COL8A2** appeared substantially stronger in Indians compared with Malays (P-value for homogeneity = 0.017; Table 2). This is likely to be due to ethnic-specific differences in genetic contribution toward CCT, as the difference in effect size was observed despite similar minor allele frequencies in Indians and Malays. Similar observations were also made for SNPs at the chromosome 9q34 **RXRA-COL5A1** locus, where the effect size was larger in Malays compared with the Indians, although the difference was not statistically significant. As the genetic effect exerted by all of the identified CCT loci appears to be modest in all ethnic groups studied thus far (15,16), it is very likely that yet-unknown genetic variants and environmental exposures may play a role in CCT determination.

One of the strongest rationales for identifying genes associated with CCT is for the opportunity to investigate the mechanisms underlying the association between CCT and POAG. The potential role for collagens in glaucomatous optic neuropathy has been clearly demonstrated in a monkey model with laser-induced glaucoma (40). In these glaucomatous monkey eyes, alterations in the three-dimensional organization of collagen fibrils were observed in the optic nerve head, suggesting that these architectural changes may affect the flexibility and resilience required of the lamina cribrosa in supporting optic nerve fibers. One can therefore speculate that such alterations in the organization of collagen fibrils may also occur due to aberrantly expressed collagens in the lamina cribrosa, driven this time by genetic variant(s). Expression elevation of extracellular matrix collagen gene, **COL8A2**, has also been observed in trabecular meshwork (TM) cells in response to dexamethasone treatment, and this has been suggested to be part of the pathological process in steroid induced-glaucoma (41,42). We also show here that **COL8A2** is expressed in other ocular tissues involved in glaucoma pathology, such as the optic nerve (Fig. 4). Notably, rare **COL8A2** mutations have also been found in some glaucoma patients with CCT <530 μm (43). These data, taken together with our GWAS, suggest that POAG and CCT may indeed
have a true biological link and common genetic determinants, such as those involved in collagen metabolism and regulation. Collectively evidence thus far supports further investigations into the genetic architecture of CCT to identify more candidate genes to study as causal genes for POAG. Such studies may eventually help ophthalmologists to develop a clinically useful genetic profile to be used in the identification, risk stratification and thus treatment of POAG patients in the future (17).

In summary, we report the identification of one novel CCT genetic locus, and highly suggestive ($P < 5 \times 10^{-7}$) association for a second, in population-based GWA studies conducted in 5080 individuals in two Asian ethnic populations in Singapore. Polymorphisms near the COL8A2 gene (rs96067, $P_{\text{meta}} = 5.40 \times 10^{-13}$) and COL5A1 (rs7044529, $P_{\text{meta}} = 2.48 \times 10^{-7}$) were associated with normal variation in CCT. We managed to confirm the previously reported association at ZNF469 as well as at RXRA-COL5A1. This study highlights collagen-related genes expressed within the cornea as novel candidates for the study of the causation of POAG. We provide strong evidence that a true biological link could exist between CCT and POAG via corneal collagen-related genes that modulate not only CCT but perhaps also the pathological changes in ocular tissues and structures pertinent to the development of glaucoma, most notably the optic disc.

**MATERIALS AND METHODS**

**Study population**

Both the SiMES and SICC adhered to the Declaration of Helsinki. Ethics approval was obtained from the Singapore Eye Research Institute (SERI) Institutional Review Board (IRB). All participants were given a choice to provide written informed consent in either Tamil, Mandarin, Malay or English using bilingual interviewers. Both versions of study information sheet and informed consent form were approved by the SERI IRB before the study commenced.

SiMES is a population-based cross-sectional study of 3280 Malay adults aged 40–79 years. Details of the SiMES design, sampling plan and methods have been reported elsewhere (24). In brief, an age-stratified random sampling of all Malay adults, aged 40–80 years, residing in 15 residential neighborhoods...
districts in the southwestern part of Singapore was drawn from the computer-generated random list of 16,069 Malay names provided by the Ministry of Home Affairs. A total of 1,400 names from each decade of age (40–49, 50–59, 60–69 and 70–79 years), or 5,600 names, were selected. Of these, 4,168 individuals (74.4%) were determined to be eligible to participate. A person was considered ineligible if he or she had moved from the residential address, had not lived there in the past 6 months, was deceased or was terminally ill. Of the 4,168 eligible individuals, 3,280 participants (78.7%) took part in the study. The study was conducted from August 2004 through to June 2006.

The SICC is designed to complement the SiMES in ethnic Indians and ethnic Chinese residents of Singapore. This study was further divided into the SINDI and Singapore Chinese Eye Study (SCES). Further information about the SICC has been published elsewhere (23). Similar to SiMES, the SINDI study is a population-based cross-sectional epidemiological study, but of ethnic Indian adults aged between 40 and 80+ years residing in Singapore. As with SiMES, the Ministry of Home Affairs provided an initial computer-generated list of 12,000 ethnic Indian names derived from a simple random sampling of all ethnic Indian adults aged 40–80+ years residing in 15 residential districts in Southwestern Singapore. From this list, a final sampling frame of 6,350 ethnic Indian residents was derived using an age-stratified random sampling strategy. SINDI was conducted from March 2007 to May 2007 and recruited 3,400 (75% response rate) participants.

**Measurements of CCT**

Five CCT measurements were obtained from each eye with an ultrasound pachymeter (Advent, Mentor O & O, Norwell, MA, USA) and the median reading was taken (25,44). As there was good correlation between the measurements in both eyes, only the readings from the right eye were used for analysis.

**Genotyping and data QC**

Study participants were genotyped using the Illumina Human610-Quad BeadChips, which assays 620,901 SNPs across the genome, according to manufacturer’s protocols. SNPtest (45) was used for QC statistics. QC criteria included a first round of SNP QC. SNPs that had missingness >5%, gross departure from Hardy–Weinberg equilibrium (HWE) ($P$-value < $10^{-5}$) or were monomorphic were excluded from subsequent analysis. Sample QC was then performed. Samples with an overall call rate <95% were excluded from analysis. Samples were subjected to biological relationship verification by using the principle of variability in allele sharing according to the degree of relationship. Identity-by-state information was derived by PLINK (46). Those individuals who showed evidence of cryptic relatedness (possible either due to duplicated or biologically related samples) were removed before PC analysis was performed. PC analysis was undertaken using EIGENSTRAT (47) to account for spurious associations resulting from ancestral differences of individual SNPs. PCs showing significant effect on univariate analysis (PC1, PC2 and PC3 for SINDI, and PCs 1 and 2 for SiMES) were used to correct for any underlying population substructure. Lastly, samples showing gender discrepancies between the clinical gender and genetically inferred gender were removed.

In SiMES, 170 individuals showed evidence of admixture and were consequently excluded. This confirmed that participants were drawn from the same, single population. Biological relationship verification revealed a total of 279 samples with cryptic relatedness and 37 samples with impossible biological sharing or heterogeneity, probably because of contamination or high missingness. In addition, 44 individuals were removed due to gender discrepancies. Overall, 530 samples were excluded, leaving a total of 2,542 individuals for statistical analysis. Likewise in SINDI, a total of 415 samples were removed from future analysis: 39 for population structure, 16 for gender discrepancies, 39 for evidence of admixture, 326 samples for evidence of cryptic relatedness, 34 samples with impossible biological sharing or heterogeneity and 17 individuals due to gender discrepancies. This left a total of 2,538 individuals for statistical analysis.

After the removal of the sample, SNP checks were performed only on the autosomes, yielding 600,450 SNPs, of which 20 were first removed in both studies, as they could not be synchronized to the forward strand. In SiMES, SNPs were excluded from the analysis when (i) they were either monomorphic or had an MAF of missingness ($<5\%$, (ii) they were monomorphic or had an MAF $<5\%$ ($n = 13,230$); or (iii) their genotype frequencies deviated from HWE ($P < 5 \times 10^{-6}$) ($n = 3,053$). This left a total of 557,824 SNPs for association analysis. In SINDI, (i) 26,602 SNPs were removed due to high rates of missingness ($>5\%$), (ii) as well as 11,771 SNPs that were either monomorphic or had MAF $<5\%$, (iiii) and 2,956 SNPs for gross departures from HWE ($P < 5 \times 10^{-6}$). Hence, we were left with a total of 559,119 SNPs for association analysis in SINDI.

In both cohorts, genotyping clusters were directly visualized for SNPs showing suggestive evidence of association with CCT at $P < 1 \times 10^{-5}$, and confirmed to be good before inclusion for statistical analysis. Supplementary Material, Figure S4, includes all the raw genotyping cluster plots (Illumina 610K) for all genome-wide significant SNPs in the SINDI discovery cohort.
Statistical analysis

The PLINK software (version 1.06) (46) was used for primary association testing, as well as for modeling within a conditional logistic regression framework. Individual genotypes were coded according to the number of copies of the variant allele present: 0 for the wild-type genotype, 1 for heterozygotes, and 2 for homozygote variants. A trend test incorporated within a linear regression model was used for primary association testing between genotypes and CCT as a quantitative trait, adjusting for age and gender.

Manhattan and LD plots were created with the use of Haploview (version 3.2) (48). Q–Q and regional association plots were created using the software R (www.r-project.org) (49).

Meta-analysis

The pooling of significant results was performed using the inverse-variance method, as described previously for quantitative traits (50). This method weighs each study according to effective sample size and cohort-specific minor allele frequency of the associated variants.

Expression analysis of COL8A2 in ocular tissues

Expression of COL8A2 was assessed by semi-quantitative reverse transcription PCR (RT-PCR) using COL8A2-specific primers (forward 5′-CAAAGGCCAGTACCTGGGAA-3′ and reverse 5′-GGGCTCTCCCTCAGGTC-3′) on total RNA extracted from a variety of ocular tissues (cornea, sclera, retina and RPE, iris, lens capsule and optic nerve) with TRIzol® Reagent (Invitrogen, Carlsbad, CA, USA) in accordance with the manufacturer’s protocol. First-strand cDNA synthesis was performed with SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen). Semi-quantitative RT-PCR was performed according to the manufacturer’s protocol, with the SYBR® Green Master Mix (Invitrogen) using the above COL8A2 primers. The resulting PCR products were electrophoresed on a 2% agarose gel and visualized by ethidium bromide staining. The ubiquitously expressed ACTB gene was amplified using specific primers (forward 5′-CCACACCGCGAGAAGATGA-3′ and reverse 5′-CCAGAGGCGTACAGGGATAG-3′) and used as amplification and normalizing control.

WEB RESOURCES

Genome browser: http://genome.ucsc.edu/.
Software R: http://www.r-project.org.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

The authors would like to thank Wee-Yang Meah, Suo Chen, Jun Liu, William Shei, Divya Venkataraman, Hong-Boon Toh, Khai-Koon Heng and Xiaoyin Chen for technical and genotyping support.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by grants from the National Medical Research Council, Singapore (NMRC 0796/2003, STA/R/0003/2008) and the Biomedical Research Council (BMRC 09/135/19/616 and 08/135/19/550).

REFERENCES