A genome-wide association study identifies a potential novel gene locus for keratoconus, one of the commonest causes for corneal transplantation in developed countries

Xiaohui Li1,2, Yelena Bykhovskaya1,3, Talin Haritunians2, David Siscovick4, Anthony Aldave5, Loretta Szczotka-Flynn6, Sudha K. Iyengar6, Jerome I. Rotter2, Kent D. Taylor2 and Yaron S. Rabinowitz1,3,*

1Cornea Genetic Eye Institute, 2Medical Genetics Institute and 3Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA and 4University of Washington, Seattle, WA, USA, 5The Jules Stein Eye Institute, University of California Los Angeles, Los Angeles, CA, USA and 6Case Western Reserve University, Cleveland, OH, USA

Received April 29, 2011; Revised September 22, 2011; Accepted October 3, 2011

Keratoconus is a condition in which the cornea progressively thins over time, and is a major cause for cornea transplantation. To identify keratoconus susceptibility regions, we performed a comprehensive genome-wide association study (GWAS) using a discovery and replication design. A discovery panel of 222 keratoconus Caucasian patients and 3324 Caucasian controls was genotyped using Illumina 370K beadchips. Further associated and fine-mapping single nucleotide polymorphisms (SNPs) (n = 4905) were genotyped in an independent replication case–control panel of 304 cases and 518 controls and a family panel of 307 subjects in 70 families. Logistic regression models implemented in PLINK were performed to test associations in case–control samples with and without principal component (PC) adjustments. Generalized estimation equation models accounting for familial correlations implemented in GWAF were used for association testing in families. No genome-wide associations were identified in the discovery GWAS panel. From the initial testing without adjustments for PCs, the top three SNPs located at 3p26 (rs6442925), 2q21.3 (rs4954218) and 19q13.3 (rs1428642) were identified with unadjusted P-values of 6.5 × 10⁻⁸, 2.4 × 10⁻⁷ and 3.1 × 10⁻⁷, respectively. After adjustments for PCs, rs1428642 became the most significant through the genome with a P-value of 1.4 × 10⁻⁶, while rs6442925 and rs4954218 were less significant (P = 1.9 × 10⁻⁵ and 2.6 × 10⁻⁴). SNP rs4954218 was confirmed in two independent replication panels with P-values of 0.004 and 0.009, respectively. Meta-analysis revealed a highest association at rs4954218 with adjusted P = 1.6 × 10⁻⁷ (unadjusted P = 1.2 × 10⁻⁴). These findings suggest SNP rs4954218, located near the RAB3GAP1 gene, previously reported to be associated with corneal malformation, is a potential susceptibility locus for keratoconus.

INTRODUCTION

Keratoconus is a clinical term used to describe a condition in which the cornea assumes a conical shape as a result of non-inflammatory thinning and protrusion. It typically commences at puberty and progresses intermittently until the mid-thirties at which point in time it usually arrests. In ~20% of cases, it progresses to the point of legal blindness and which can be treated only by corneal transplantation (1–3). It is a major cause of corneal transplantation in western developed countries (4). It has an estimated prevalence of 1 in 2000 in the general population and may even be higher if new

*To whom correspondence should be addressed at: Cornea Genetic Eye Institute, Cedars-Sinai Medical Center, 50 N La Cienega Bl., Suite 340, Beverly Hills, CA 90211, USA. Tel: +1 3104239640; Fax: +1 3102487485; Email: rabinowitzy@cshs.org

© The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
technologies are used for early diagnosis (5). It can represent up to 5% of individuals who present in screening for refractive corneal surgery, such as LASIK eye surgery (6). Despite intensive biochemical and genetic investigations, its underlying cause remains poorly understood (1.7–12).

Although the etiology of keratoconus is still unknown, genetic predisposition plays an important role as indicated by the association of keratoconus with genetic syndromes, segregation analyses, genetic epidemiological data and gene mapping studies (8,9,11,12). A number of gene loci/chromosomal regions for keratoconus have been mapped by linkage analysis (7,10,13–17); however, susceptibility genes have not been identified in these regions. Mutations in the visual system homeobox gene 1 (VSX1) have been identified through the targeted screening of this gene in keratoconus patients but with conflicting results (18–21).

Results of gene mapping efforts imply the importance of genetic determinants, but the lack of consistent chromosomal loci among different studies suggests genetic heterogeneity and illustrates the complex nature of the genetic contribution to the development of keratoconus. In addition, the sample sizes for most studies were moderate, contributing to difficulty in replication. Moreover, it has been difficult to localize regions precisely using the linkage method. For example, using a four-generation pedigree and a fine-mapping approach, we could only localize a linkage region to a 20 Mb region on chromosome 5, which still contained hundreds of genes (16). Although the candidate gene approach is useful, it is greatly limited by lack of information to adequately select keratoconus susceptibility genes. The recent developments in genotyping technology have made large-scale genetic studies feasible with hundreds of thousands of dense markers, such as genome-wide association studies (GWASs). This hypothesis-free high-throughput approach makes it possible to identify common genomic variations contributing to complex genetic diseases, such as glaucoma, macular degeneration, refractive error (22–24), ocular quantitative traits such as corneal thickness (25–27) and central retinal vessels (28). However, thus far, no GWASs have been performed for keratoconus.

In this first GWAS study of keratoconus, we genotyped three independent case-control and family cohorts of Caucasian keratoconus patients. We identified in the discovery panel and confirmed in both replication panels a potential novel keratoconus susceptibility region located at the 2q21.3. This result suggests the involvement of the gene RAB3GAP1 coding for the Rab3 GTPase-activating protein subunit 1, previously implicated in Warburg Micro Syndrome with corneal malformation. This comprehensive GWAS and replication performed in patients with keratoconus offers long-awaited insight into the genetic susceptibility to keratoconus, one of the common causes of corneal transplantation in the developed world (4).

RESULTS

GWAS design: discovery and replication cohorts

In this report, we have studied three independent cohorts of patients with keratoconus. The first case-control cohort used for the discovery phase of the GWAS consisted of 222 Caucasian keratoconus patients and 3324 Caucasian controls. The second case-control cohort of 304 Caucasian keratoconus patients and 518 Caucasian controls was genotyped in the regions identified in the discovery cohort for the purpose of their replication. We also analyzed a family sample of 307 individuals from 70 families consisting of 146 keratoconus patients and 161 unaffected family members.

The gender distribution was similar between cases and controls in the discovery cohort (males: 56 versus 61%, \(P = 0.20\)). However, significantly more males were present in cases than in controls in the replication cohort (68 versus 52%, \(P < 0.05\)).

Given the sample size, we have moderate power to identify genetic associations of keratoconus. Consequently, we sought to maximize the number of putative genomic regions for study by first applying mathematical methods without formal population structure correction in our primary analysis, then reducing the number of false positive associations and avoiding power loss by testing of multiple replication samples. However, we report association results both without and with population structure adjustments below in order to allow comparisons of the strength of our results with current GWASs.

Genetic association testing of the discovery cohort

From the initial association testing without adjustments for population structure in the discovery cohort, the top three single nucleotide polymorphisms (SNPs), rs6442925 at 3p26, rs4954218 at 2q21.3 and rs1428642 at 19q13.3, were identified with unadjusted \(P\)-values of 6.5 \(\times\) 10\(^{-8}\), 2.4 \(\times\) 10\(^{-7}\) and 3.1 \(\times\) 10\(^{-7}\), respectively (Table 1, Fig. 1A). To control for population stratification, we performed PC analysis (PCA). After adjusting for sex and three significant PCs, the quantile–quantile (QQ) plot had a genomic control of 1.03 (Supplementary Material, Fig. S1). The most significant SNP identified after adjustments was rs1428642 (ranked number 3 without adjustments) with a \(P\)-value of 1.4 \(\times\) 10\(^{-9}\). Other top two ranked SNPs from the initial test, rs6442925 and rs4954218, became less significant with adjusted \(P\)-values of 1.9 \(\times\) 10\(^{-5}\) and 2.6 \(\times\) 10\(^{-5}\), respectively (Table 1, Fig. 1B). No genomic variants reached genome-wide significance level of 5 \(\times\) 10\(^{-8}\) after adjustments. All SNPs with the suggestive genome-wide \(P\)-values \(<\) 5 \(\times\) 10\(^{-6}\) before and after adjusting for population structure are shown in Table 1. PCA testing of the genotyping data of cases and controls in the discovery panel revealed a similar distribution between cases and controls (Supplementary Material, Fig. S2).

Genetic association testing of replication cohorts

Given the moderate power in the discovery cohort as indicated by lack of hits reaching the genome-wide significance level of 5 \(\times\) 10\(^{-8}\), all SNPs identified with a less stringent significant threshold (10\(^{-4}\)) were selected for genotyping and analysis in the replication cohorts. In addition, we included proxy SNPs in strong linkage disequilibrium with our primary associated SNPs and tag SNPs spanning associated regions to provide redundancy for replication and to perform additional fine-mapping of the regions. Minor allele frequencies were
Table 1. SNPs identified in the discovery panel with \( P < 5 \times 10^{-6} \), before and after adjustments for PCs, and genotyped in the confirmation cohorts

<table>
<thead>
<tr>
<th>Position</th>
<th>Closest gene</th>
<th>SNP</th>
<th>Minor allele</th>
<th>Minor allele frequency (in cases)</th>
<th>Minor allele frequency (in controls)</th>
<th>Meta-analysis</th>
<th>GWAS</th>
<th>Replication (case–control)</th>
<th>Replication (family)</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>F_A</td>
<td>F_U</td>
<td></td>
<td>P-value</td>
<td>P-A</td>
<td>P-value</td>
<td>P-value</td>
<td>P-value</td>
</tr>
<tr>
<td>rs6442925</td>
<td>3p26 BHLHB2</td>
<td>T</td>
<td>0.25</td>
<td>0.15</td>
<td>1.85</td>
<td>6.5E-8</td>
<td>1.9E-2</td>
<td>1.7E-5</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>rs4954218</td>
<td>2q21.3 RAB3GAP1</td>
<td>G</td>
<td>0.14</td>
<td>0.25</td>
<td>0.50</td>
<td>2.4E-2</td>
<td>1.6E-2</td>
<td>1.4E-2</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>rs1428642</td>
<td>19q13.3 BIRC8</td>
<td>A</td>
<td>0.33</td>
<td>0.45</td>
<td>0.59</td>
<td>3.1E-2</td>
<td>1.4E-2</td>
<td>1.4E-2</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>rs1328083</td>
<td>13q33.3 –</td>
<td>G</td>
<td>0.24</td>
<td>0.15</td>
<td>1.76</td>
<td>6.7E-4</td>
<td>6.0E-8</td>
<td>6.0E-8</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>rs8111998</td>
<td>19p12 –</td>
<td>T</td>
<td>0.12</td>
<td>0.06</td>
<td>2.10</td>
<td>1.0E-6</td>
<td>1.0E-6</td>
<td>1.0E-6</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>rs4839200</td>
<td>1p13.2 KCND3</td>
<td>A</td>
<td>0.22</td>
<td>0.13</td>
<td>1.79</td>
<td>3.1E-5</td>
<td>3.0E-5</td>
<td>3.0E-5</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>rs1240742</td>
<td>1q44 KIF26B</td>
<td>T</td>
<td>0.21</td>
<td>0.13</td>
<td>1.79</td>
<td>1.8E-2</td>
<td>4.1E-2</td>
<td>4.1E-2</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>rs1328089</td>
<td>13q33.3 –</td>
<td>C</td>
<td>0.34</td>
<td>0.25</td>
<td>0.71</td>
<td>4.7E-4</td>
<td>2.4E-4</td>
<td>2.4E-4</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>rs3749350</td>
<td>3p26.2 LRRN1</td>
<td>T</td>
<td>0.20</td>
<td>0.12</td>
<td>1.77</td>
<td>3.0E-5</td>
<td>2.4E-5</td>
<td>2.4E-5</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>rs2659546</td>
<td>4q23 PPP3CA</td>
<td>A</td>
<td>0.10</td>
<td>0.05</td>
<td>2.16</td>
<td>3.3E-2</td>
<td>7.4E-2</td>
<td>7.4E-2</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>rs1978238</td>
<td>12p13.3 –</td>
<td>C</td>
<td>0.29</td>
<td>0.39</td>
<td>0.64</td>
<td>3.4E-2</td>
<td>3.4E-2</td>
<td>3.4E-2</td>
<td>0.08</td>
<td>0.08</td>
</tr>
</tbody>
</table>

bF_A, minor allele frequency in cases; F_U, minor allele frequency in controls.

cAdjustments for gender and PC variables.

‡ P-value (after adjusting for rs4954218 as a confounding factor in the association test).

Meta-analysis

Meta-analysis of the combined genotyping data of discovery and two replication panels determined SNP rs4954218, located at the 2q21.3 genomic region, to be associated with keratoconus with PC-adjusted \( P \)-value of \( 1.6 \times 10^{-7} \) (unadjusted \( P \)-value of \( 1.2 \times 10^{-8} \)). The SNP rs4954218 was the top ranked SNP identified by meta-analysis both with and without PC adjustments; however, no SNPs reached genome-wide significance level after adjustments for PCs in the meta-analysis. The distributions of minor allele frequencies between cases and controls are similar among three data sets (0.14 versus 0.25 in GWAS; 0.17 versus 0.23 in the case–control replication panel; 0.16 versus 0.22 in the family-based replication panel).

Variation across 2q21.3 and keratoconus

The most significant association signal identified not only in the discovery cohort, but also further confirmed in two replication cohorts, is located at SNP rs4954218 (Fig. 2). It is located 6.4 kb 5′ to \( RAB3GAP1 \), which codes for the catalytic subunit of the RAB3 GTPase-activating protein, and, on the other strand, 21.2 kb 5′ to \( YSK4 \), coding for SPS1/STE20-related protein kinase. In addition, as shown on Figure 2C, two additional SNPs, genotyped only in the confirmation panels, showed positive association in this region: SNP rs4954218 located at the chromosomal region 2q21.3 (\( \approx 2.7 \times 10^{-7} \)), and SNP rs1561277 located \( \approx 100 \) kb away in an intron of \( RAB3GAP1 \), and SNP rs1561277 located \( \approx 290 \) kb away in the intron of \( ZRANB3 \), coding for zinc finger, RAN-binding domain containing three genes (\( P = 2.7 \times 10^{-7} \)). In order to determine whether these signals represent independent associations, we performed conditional analysis by re-analyzing the replication keratoconus cohort after adjusting for rs4954218 genotypes in the association test. Figure 2 shows the regional plots at 2q21.3 before (Fig. 2A, original regional plot; Fig. 2B, regional plot with adjustments for PCs and Fig. 2C, regional plot in replication samples) and after (Fig. 2D) adjustment. After controlling for rs4954218 as a covariate, the associations for the other two SNPs were similar in the discovery and replication cohorts, thus confirming that both samples were well matched.

SNP rs4954218 located at the chromosomal region 2q21.3 was confirmed in both replication panels with \( P \)-values of 0.004 and 0.009 for case–controls and family panels, respectively, with consistent direction of effect in all three panels. Additional SNPs at 2q21.3 also showed significant association, with \( P \)-values ranging from \( 4.0 \times 10^{-5} \) to 0.05. Two regions at 19p12 and 7q31, associated in the discovery panel with unadjusted \( P \)-values of \( 1.0 \times 10^{-6} \) and \( 6.0 \times 10^{-6} \) (adjusted \( P = 0.0058 \) and 0.0064) (Table 1), were significant in one of the two replication panels. The region at 19p12 was associated in the family panel with a \( P \)-value of 0.003; the region at 7q31 in the case–control panel with the \( P \)-value of 0.03 (Table 1). Other top regions identified by GWAS were not associated with keratoconus in either confirmation panel. Although there was a difference in gender distribution between cases and controls, the \( P \)-values remained similar before and after adjustments for gender and other covariates (data not shown).
greatly diminished but remained significant ($P = 0.004$ and $0.02$, respectively) as shown in Figure 2D. Because the signal of rs4954218 is weaker than the one from rs6730157 in the replication cohort where the conditional analysis is performed, there is a strong possibility that there is only one signal and that rs6730157 captures better the causal variant(s) than rs4954218. A final conclusion could be reached through additional fine-mapping and sequencing in the region, and likely will require additional subjects.

**DISCUSSION**

In this study, we identified a potential genetic susceptibility region located at 2q21.3 in the first GWAS of keratoconus and then confirmed this observation in two replication sample sets. Several candidate genes are located at this region. However, based on association results in the region and biological evidence, we propose *RAB3GAP1* as a gene for keratoconus. The most significantly associated SNP rs4954218 is located 5' upstream and within a theoretical possibility of a regulatory effect to two genes: *RAB3GAP1* (6.4 kb away) and *YSK4* (21.2 kb away). The second highest association signal is located at SNP rs6730157 in the *RAB3GAP1* intron, and based on results of the conditional analysis, there is a strong possibility that both associated SNPs represent a single association signal linked to the causal variant(s) located in the *RAB3GAP1* gene. This gene encodes a catalytic subunit of Rab3 GTPase-activating protein (*RAB3GAP*), a direct activator of the Rab3 GTPase, a key regulator of calcium-mediated hormone and neurotransmitter exocytosis (29,30). Multiple mutations in *RAB3GAP1* have been identified in patients with Warburg Micro Syndrome, a rare, autosomal recessive syndrome characterized by a number of neurological and ocular developmental defects, including microcornea, congenital cataracts and optic atrophy (31,32). Congenital cataracts are also frequent in patients with Martsolf syndrome due to mutations in the non-catalytic subunit (*RAB3GAP2*) of *RAB3GAP* (33). Eye abnormalities have been also identified in another Rab-related inherited disorder choroideremia, an inherited choroidal and retinal regeneration, due to mutations of Rab escort protein 1, a cofactor required for Rab protein function (34).

According to a published study, *RAB3GAP1* is ubiquitously expressed in a variety of human tissues; however, no eye
tissue was tested (35). BLAST search of \textit{RAB3GAP1} transcript through EST database identified 15 ESTs from several independent cDNAs from a variety of eye tissues, including human normalized iris, trabecular meshwork, fetal lens, eyeball, retina, fetal eye; and mixed library of fetal lens, eye anterior segment, optic nerve, retina and choroid. No normal corneal ESTs matched to the \textit{RAB3GAP1} transcript; however, most interestingly, a matching EST derived from an unamplified human keratoconus cornea library constructed by our group in collaboration with National Eye Institute (36) was identified in a BLAST search. Search of the Bgee database (37) of gene expression data derived from Affymetrix expression microarrays indicated that individual probes specific to the \textit{RAB3GAP1} transcript are present in a mouse cornea in three independent microarray experiments, thus suggesting its presence in normal human corneas as well.

Recurrent ocular abnormalities in disease phenotypes due to defects in \textit{RAB3GAP1} provide clear indication that Rab3 proteins are required for both function and normal structural development of the eye. However, the precise mechanisms by which loss of \textit{RAB3GAP} function could lead to developmental anomalies in the eye are not known. It is clear that in the human cornea, a number of biologically important pathways will be directly affected by a defect in the vesicular membrane transport due to the \textit{RAB3GAP} defect.

Based on its genomic location, biological function relevant to the function of the eye, involvement in ocular abnormalities and expression pattern, \textit{RAB3GAP1} represents the most plausible candidate gene responsible for the keratoconus association signal. Extensive database searches and data mining did not identify any clear evidence to suggest preferential expression or function of \textit{YSK4} and \textit{ZRANB3} in the eye, as well as no ocular phenotypes linked to these genes’ defects.

Other potentially associated with keratoconus chromosomal regions identified in the discovery phase have not been formally confirmed. The 19p12 and 7q31 genomic regions were only significant in one of the two replication sample sets. The initial top hits at the 3p26 region (unadjusted) and at the 19q13.3 region (adjusted) were not significant in any replication panels.

The keratoconus susceptibility region at 2q21.3 represents a potential novel locus which has not been described in previously published linkage studies of keratoconus families (cited in the Introduction). This comes as no surprise, as linkage and association methods are designed to identify rare and common disease-associated variants for complex multigenic disorders. This locus was not described among major genomic loci associated with population-based variation in central corneal thickness in recently published GWASs (25–27), indicating that other pathways may make a greater contribution to the keratoconus pathogenesis.
Some limitations of our study should be noted. First is our approach to control for population structure and reduce the false positives. Keratoconus is a rare disease with population prevalence of 1/2000, which limits the identification and recruitment of new patients. Given the sample size and resources, we have moderate power to identify the associations. Although adjusting for population structure is an important tool to reduce the number of potentially inflated associations, we identified no SNPs with P-values at the $10^{-7}$ level after adjustments, and only two SNPs with P-values below the $5 \times 10^{-6}$ level: rs1428642 ($P = 1.4 \times 10^{-6}$; ranked #3 without adjustment) and rs1978238 ($P = 3.4 \times 10^{-6}$). Meta P-value of the top SNP, rs1428642, was only 0.014, clearly indicating that it was not confirmed in either of the replication panels. In contrast, SNP rs4954218 in the 2q21.3 region with the unadjusted P-value of $2.4 \times 10^{-7}$ in the discovery panel was much less significant after PC adjustments ($P = 2.6 \times 10^{-8}$). However, it went through successful replication in two replication cohorts, with and without PC adjustments, and retained a high significance level with an adjusted meta P-value of $1.6 \times 10^{-7}$, close to the GWAS significance level. In addition, the distributions of minor allele frequencies are similar in three data sets, which is further evidence that the SNP association is likely real. This suggests a possibility that PC adjustments used in the analysis of a relatively rare complex genetic disease, such as keratoconus, reduce the power for identifying potential disease-associated regions, as has been stated in the paper by Price et al. (38). Thus, using the strategy we used in this paper, i.e. of following-up more possible candidate regions in the replication panels, appears to be a useful approach to reduce the false positive rate and still avoid the power loss. Notably in the past, some GWASs done in a limited number of patients have achieved major successes in identifying genetic associations, e.g. complement factor H in AMD (39). These studies applied a design without population structure adjustments, which were not available at that time.

Another issue is the failure of successful genotyping of several SNPs in the replication cohorts as we noted in the Table 1. Different from the technology for GWAS, we used customized Illumina iSelect panel to genotype the samples in replication cohorts. The total failure rate is ~5% in our study (255/4905) which is much lower than the guarantee from Illumina (~20%). Although the failure rate is not high, it could reduce the probabilities for identifying some true associations.

In summary, in this first comprehensive GWAS in patients with keratoconus, we identified a novel potential keratoconus locus at 2q21.3, containing a candidate gene RAB3GAP1. This novel genetic susceptibility locus should be further investigated in the future genetic research of keratoconus, the most common cause of corneal transplantation in the western developed world.

**MATERIALS AND METHODS**

**Subjects**

Clinically affected keratoconus patients were recruited at three major sites: the Cornea Genetic Eye Institute ($n = 618$) at Cedars-Sinai Medical Center, Los Angeles, CA, USA; the Jules Stein Eye Institute ($n = 26$) at UCLA, Los Angeles, CA, USA; and the University Hospitals Eye Institute ($n = 46$) at University Hospitals Case Medical Center and Case Western Reserve University, Cleveland, OH, USA. Institutional Review Board (IRB) approval was obtained at all clinic sites. Written informed consent was obtained from all subjects. The study was conducted in concordance with the provisions of the Declaration of Helsinki.

GWAS cohort. Clinically affected Caucasian keratoconus cases ($n = 240$) were enrolled into the GWAS as a part of the longitudinal videokeratography and genetic study at the Cornea Genetic Eye Institute (12). After removing samples with poor quality of genotyping, 222 samples were included in the analysis. Caucasian controls ($n = 3324$) were obtained from the Cardiovascular Health Study (CHS), a population-based cohort study of risk factors for cardiovascular disease and stroke in adults 65 years of age or older, recruited at four field centers. A total of 5201 predominantly Caucasian individuals were recruited in 1989–1990 from random samples of Medicare eligibility lists, followed by an additional 687 African-Americans recruited in 1992–1993 (total $n = 5888$). CHS was approved by the Institutional Review Board at each recruitment site, and subjects provided informed consent for the use of their genetic information. African-American CHS participants were excluded from the analysis due to insufficient number of ethnically matched cases (40,41).

**Replication cohort.** An independent group of keratoconus cases and controls was recruited as a part of the collaborative replication study. A total of 232 independent keratoconus cases were recruited at the Cornea Genetic Eye Institute; 26 cases at the Jules Stein Eye Institute; and 46 cases at University Hospitals Eye Institute. This replication panel consisted of a total of 304 keratoconus cases. Five hundred and eighteen normal controls were also recruited at Cedars-Sinai Medical Center in Los Angeles.

**Keratoconus families.** Family members of keratoconus cases diagnosed at the Cornea Genetic Eye Institute were recruited to perform family-based studies. A total of 307 individuals from 70 families consisting of 146 keratoconus patients and 161 unaffected family members were obtained for this study. One hundred and eighty-six individuals in 41 pedigrees were Caucasians, whereas 91 individuals in 20 pedigrees were identified as Hispanics.

**Clinical diagnosis**

The diagnosis of keratoconus was performed by a cornea fellowship trained ophthalmologist or an experienced research optometrist (L.S.-F.) based on clinical examination and videokeratography pattern analysis. Clinical examination included slit-lamp biomicroscopy, cycloplegic retinoscopy and fundus evaluations. Slit-lamp biomicroscopy was used to identify stromal corneal thinning, Vogts’ striae or a Fleischer ring. Retinoscopy examination was performed with a fully dilated pupil to determine the presence or absence of retro
illumination signs of keratoconus, such as the oil droplet sign and scissoring of the red reflex 20 min after phenylephrine 2.5% and cyclopentolate 1% drops had been instilled in the eye. Videokeratography evaluation was performed on each eye using the Topographic Modeling System (Computed Anatomy, New York, NY, USA), Orbscan II (Bausch & Lomb, Rochester, NY, USA), Oculus Pentacam (Oculus, Inc., Lynnwood, WA, USA) or Keratron (Optikon, Rome, Italy).

Subjects were considered having keratoconus if they had at least one clinical sign of keratoconus and a confirmatory videokeratography map with an asymmetric bowtie with skewed radial axis above and below the horizontal meridian (AB/SRAX) pattern (42). Importantly, topography was screened for mimicking disease such as pellucid marginal degeneration which was excluded. Subjects that had bilateral keratoplasty for keratoconus were included if the surgical pathology report confirmed the presence of the disease.

Genotyping

DNA extraction. DNA was extracted from EBV transformed lymphoblastoid cell lines established from peripheral whole blood of each study participant using NucleoSpin Tissue kit (MACHEREY-NAGEL Inc., Bethlehem, PA, USA). Salt extraction protocol was used to tract DNA from buffy coats whenever cell lines were not available (43).

Genome-wide association study. All genotyping was performed at the genotyping laboratory of the General Clinical Research Center and the Medical Genetics Institute at Cedars-Sinai Medical Center. Discovery genotyping was performed using the Illumina Human CNV370 Quad whole genome genotyping beadchip, following the manufacturer’s protocol (Illumina, San Diego, CA, USA) (44,45). The average genotyping rate for samples passing quality control was 99.77%. Samples with poor genotyping quality (n = 18) were removed from the analysis. Procedures for control of genotyping quality included genotyping rate per subject >99%, narrow distribution (theta standard deviation) and wide separation between genotyping clusters (R/theta plots), minor allele frequency >5%, Hardy–Weinberg test (SNPs with P-value < 0.001 were excluded). SNPs were also excluded from analysis for significant differences in missing data between cases and controls (SNPs with P-value in missingness test < 0.01 were excluded). Approximately 290K SNPs were used for the primary GWAS analysis. After the analyses, the cluster plots of the top SNPs identified by GWAS at the 10\(^{-8}\) significance level were carefully reviewed to ensure the genotyping quality.

Replication. Replication genotyping was performed using the custom iSelect Infinium technology (46). Using the software Tagger as implemented in Haploviev (47,48) and data from the CEPH population in the International HapMap project, release 2 (49,50), we designed a panel of 4905 SNPs that were compatible with the Illumina Infinium technology, of which 4650 SNPs remained for analysis following clustering quality control. The average genotyping rate for samples genotyped on the iSelect platform and passing quality control was 99.98%. Genotyping concordance among 20 pairs of blind duplicates 100%.

Statistical analyses

GWAS and replication study. Odds ratios (ORs) and their standard errors (SE) were calculated using genome-wide SNP data under logistic regression models using the PLINK program (v1.07) (http://pngu.mgh.harvard.edu/purcell/plink/) (51). Genome-wide significance threshold of 5 × 10\(^{-8}\) was applied to our GWAS analysis. PCA for population stratification was performed in the discovery GWAS panel using EIGENSTRAT (38). Gender and three significant PC variables identified from PCA were used as covariates. Because of the large age difference between patients and controls in the discovery panel, age was not used as a covariate. Genomic control (\(\lambda_{GC}\)) was calculated to assess the inflation factor of the test statistic. For the replication case–control samples, we additionally adjusted for the site in the analysis. In order to identify additional association signals at the 2q21.3 locus, testing for association was also performed after adjustment for SNP rs4954218 as an additional covariate. Association tests controlling for familial correlations by the Generalized Estimation Equation (GEE) model were performed in families using GWAF package (52). Regional association plots were generated by LocusZoom (http://cgs.sph.umich.edu/locuszoom/) (53).

Meta-analysis. Meta-analysis of associated SNPs obtained from the logistic regression models from discovery and replication cohorts was calculated using an inverse-variance weighting by PLINK.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

This research was also supported in part by the Cedars-Sinai Board of Governors’ Chair in Medical Genetics. The authors thank the other investigators, the staff and the participants of the study for their valuable contributions. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Eye Institute (NEI 09052), the Eye Defect Research Foundation Inc. and in part by the Cedars-Sinai General Research Center Genotyping (NCRR M01-RR00425). The CHS data reported in this article was supported by the National Heart, Lung, and Blood Institute (N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, U01 HL08295 and R01 HL087652), with additional contribution from the National Institute of Neurological Disorders and Stroke.