

1 **Genetic variation on chromosome 11 is associated with keratoconus**

2 Bennet J. McComish, PhD^{1,*}; Srujana Sahebjada, PhD^{2,3}; Yelena Bykhovskaya, PhD⁴; Colin E. Willoughby,
3 FRCOphth⁵; Andrea J. Richardson, BSc (Hons)²; Abi Tenen, FRANZCO⁶; Jac C. Charlesworth, PhD¹; Stuart
4 MacGregor, PhD⁷; Paul A. Mitchell, MD, PhD, FRANZCO⁸; Sionne E.M. Lucas, PhD¹; Richard A. Mills,
5 PhD, FRANZCO⁹; David A. Mackey, MD, FRANZCO¹⁰; Xiaohui Li, MD¹¹; Jie Jin Wang, PhD⁸; Richard A.
6 Jensen, PhD¹²; Jerome I. Rotter, MD¹¹; Kent D. Taylor, PhD¹¹; Alex W. Hewitt, PhD, FRANZCO^{1,2}; Yaron
7 S. Rabinowitz, MD⁴; Paul N. Baird, PhD^{2,3}; Jamie E. Craig, DPhil, FRANZCO^{9,1}; Kathryn P. Burdon, PhD^{1,9}

- 8 1. Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia
- 9 2. Centre for Eye Research Australia, Melbourne, Victoria, Australia
- 10 3. Department of Surgery (Ophthalmology), University of Melbourne, Royal Victorian Eye and Ear
11 Hospital, Melbourne, Victoria, Australia
- 12 4. Department of Surgery and Board of the Governors Regenerative Medicine Institute, Cedars-Sinai
13 Medical Center, Los Angeles CA USA; and Cornea Genetic Eye Institute, Beverly Hills, CA USA
- 14 5. Biomedical Sciences Research Institute, Ulster University, Coleraine, Northern Ireland, UK; and
15 Belfast Health and Social Care Trust, Belfast, Northern Ireland, UK
- 16 6. Vision Eye Institute, Melbourne, Victoria, Australia; School of Primary and Allied Health Care,
17 Monash University, Melbourne, Victoria, Australia; and Melbourne Stem Cell Centre, Melbourne,
18 Victoria, Australia
- 19 7. QIMR Berghofer Medical Research Institute, Brisbane, Australia
- 20 8. Department of Ophthalmology and Centre for Vision Research, Westmead Institute for Medical
21 Research, University of Sydney, Sydney, Australia
- 22 9. Department of Ophthalmology, Flinders University, Adelaide, Australia
- 23 10. Lions Eye Institute, University of Western Australia, Perth, Australia
- 24 11. Institute for Translational Genomics and Population Science, Los Angeles Biomedical Research
25 Institute; and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance CA USA
- 26 12. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle,
27 WA, USA

28 *Corresponding author

29 **Corresponding author:**

30 Bennet J McComish

31 Menzies Institute for Medical Research, University of Tasmania, Private Bag 23, Hobart 7000, Tasmania,
32 Australia

33 Tel: +61 3 6226 4285

34 Email: bennet.mccomish@utas.edu.au

35

36 3000 words.

37 **Key points**

38 **Question** Which genetic loci associate with keratoconus?

39 **Findings** In this case-control genome-wide association study, with three independent replication cohorts, a
40 locus containing multiple variants across six protein-coding genes on chromosome 11 was associated with
41 keratoconus. Several of these genes are likely involved in apoptotic pathways.

42 **Meaning** This study of patients with keratoconus demonstrates a potential role of genes involved in
43 apoptotic pathways.

44

45 **Abstract**

46 **Importance** Keratoconus is a condition in which the cornea progressively thins and protrudes in a conical
47 shape, severely affecting refraction and vision. It is a major indication for corneal transplantation. To
48 discover new genetic loci associated with keratoconus and to better understand the causative mechanism of
49 this disease, we performed a genome-wide association study (GWAS) on patients with keratoconus.

50 **Objective** To identify keratoconus genetic susceptibility regions in the human genome.

51 **Design** The discovery cohort was genotyped using the Illumina HumanCoreExome single nucleotide
52 polymorphism (SNP) array. Following quality control and data cleaning, genotypes were imputed against the
53 1000 Genomes Project reference panel (Phase III, version 5) and association analysis was completed using
54 PLINK. SNPs with $P < 1 \times 10^{-6}$ were assessed for replication in three additional cohorts.

55 **Setting** Eye clinics in Australia, the USA, and Northern Ireland.

56 **Participants** The discovery cohort comprised 522 Australian keratoconus cases and 655 Australian
57 controls. Controls were drawn from the Blue Mountains Eye Study and normal individuals previously
58 recruited as controls for a study of glaucoma. Replication cohorts were a previous keratoconus GWAS
59 dataset from the USA (222 cases, 2927 controls); 331 cases and 229 controls from Australia and Northern
60 Ireland (AUS+NI), and an Australian case–control cohort (VIC, 265 cases, 702 controls).

61 **Main Outcomes and Measures** Associations between keratoconus and 6,252,612 genetic variants were
62 estimated using logistic regression, were adjusting for ancestry using the first three principal components.

63 **Results** Two novel loci reached genome-wide significance (defined as $P < 5 \times 10^{-8}$), with $P = 7.46 \times 10^{-9}$ at
64 rs61876744 in *PNPLA2* on chromosome 11 and $P = 6.35 \times 10^{-12}$ at rs138380 2.2kb upstream of *CSNK1E* on
65 chromosome 22. One additional locus was identified with $P < 10^{-6}$, in *MAML2* on chromosome 11. The
66 novel locus in *PNPLA2* reached genome-wide significance in meta-analysis of all four cohorts ($P = 2.45 \times$
67 10^{-8}).

68 **Conclusions and Relevance** In this relatively large keratoconus GWAS, we identified a genome-wide
69 significant locus for keratoconus in the region of *PNPLA2* on chromosome 11.

70

71 **Introduction**

72 Keratoconus is characterized by progressive thinning of the cornea, the clear tissue at the front of the eye.
73 Asymmetrical bulging and conical protrusion of the cornea leads to extreme refractive error (myopia and
74 irregular astigmatism) causing severe visual impairment.¹ Keratoconus is relatively common, with a reported
75 prevalence of around 55 per 100,000 in white populations¹ and up to 229 per 100,000 in Asian populations.²
76 Due to recent advances in diagnostic imaging the true incidence and prevalence of keratoconus may be 5 to
77 10 times higher than previously reported.³ The etiology of keratoconus is poorly understood. Links have
78 been made to eye rubbing and atopy, but no direct causative connection has been established.⁴ Various
79 biochemical pathways may be involved, including oxidative stress, apoptosis, and disruption to extracellular
80 matrix turnover.⁵

81 Although many cases of keratoconus present as sporadic, there is a well-recognised genetic component to the
82 disease. The estimated prevalence in relatives of keratoconus patients is 3.34% (95% CI: 3.22–3.46%),
83 which is 15–67 times higher than in the general population.⁶ In addition, more than 20 syndromes are
84 associated with keratoconus, including Down syndrome, Leber congenital amaurosis, and several connective
85 tissue disorders.⁷ Linkage studies have identified at least 16 loci for keratoconus;^{8,9} however, the causative
86 genes and variants have remained elusive. Common variants in the *DOCK9*¹⁰ and *LOX*¹¹ genes have been
87 implicated as well as rare mutations in *MIR184*,^{12,13} although these loci have not been broadly replicated.
88 Genome-wide association studies (GWAS) have implicated several loci. Variation in the promoter region of
89 the *HGF* gene^{14,15} and upstream of *RAB3GAP1*^{16,17} have both been associated in multiple independent
90 studies. Further, a GWAS for central corneal thickness (CCT) identified loci that are also associated with
91 keratoconus, including *RXRA–COL5A1*, *FOXO1* and *FNDC3B*,¹⁸ and more recently a suggestive association
92 at *DCN*.¹⁹

93 We present findings from a GWAS of 522 patients with keratoconus, a relatively large sample for this
94 complex disease. We show association and independent replication at the *PNPLA2* region with keratoconus.

95

96 **Methods**

97 **Study design**

98 We report a study of 4 independent Caucasian cohorts of patients with keratoconus. The first case–control
99 cohort used for the discovery phase comprised 522 keratoconus patients and 655 controls. All single
100 nucleotide polymorphisms (SNPs) with $P < 1 \times 10^{-6}$ were looked up in imputed genotypes from a previously
101 published GWAS study of 222 keratoconus patients and 3324 controls from the USA. In addition, 27 SNPs
102 were genotyped in an independent replication cohort of 331 cases and 229 controls from Australia and
103 Northern Ireland (AUS+NI), and in an additional 265 cases and 702 controls from Victoria, Australia (VIC).
104 The demographics for each cohort are given in Table 1.

105 **Discovery cohort**

106 The protocol was approved by the Southern Adelaide Clinical Human Research Ethics Committee (HREC)
107 and the HREC of the Royal Victorian Eye and Ear Hospital and the Health and Medical HREC of the
108 University of Tasmania. All participants gave written informed consent and the study conformed to the
109 tenets of the Declaration of Helsinki. Participants with keratoconus were ascertained through the eye clinic
110 of Flinders Medical Centre, Adelaide; optometry and ophthalmology clinics in Adelaide and Melbourne; or
111 an Australia-wide invitation to members of Keratoconus Australia, a community-based support group for
112 patients. Clinical data were obtained from the participants' eye care practitioner.

113 The diagnosis of keratoconus was based on both clinical examination and videokeratography pattern analysis
114 as described previously.²⁰ Clinical examination included slit-lamp biomicroscopy, cycloplegic retinoscopy,
115 and fundus evaluation. Slit-lamp biomicroscopy was used to identify stromal corneal thinning, Vogts' striae,
116 or a Fleischer ring. Retinoscopy examination was performed on a fully dilated pupil to determine the
117 presence or absence of retro-illumination signs of keratoconus, such as the oil droplet sign and scissoring of
118 the red reflex. Videokeratography evaluation was performed on each eye using the Orbscan (Orbtek/Bausch
119 & Lomb, Salt Lake City, UT, USA). Patients were classified as having keratoconus if they had at least one
120 clinical sign of keratoconus and a confirmatory videokeratography.²¹ A history of penetrating keratoplasty
121 performed because of keratoconus was also sufficient for inclusion as a case. Patients with syndromic forms

122 of keratoconus were excluded, and if multiple individuals from the same family presented, only one was
123 included.

124 Controls were obtained from the Australian cohort previously described in a GWAS for age-related macular
125 degeneration (AMD) from the International AMD Genomics Consortium²² and have been described in detail
126 previously. For the current analysis, data from 676 Australian unaffected controls (including 465 from the
127 Blue Mountains Eye Study²³ and 211 normal individuals previously recruited as controls for a study of
128 glaucoma²⁴) were combined as controls for keratoconus. Related individuals and those who did not pass all
129 sample QC for the AMD GWAS were excluded.

130 DNA for cases and controls was extracted from whole blood using the QiaAMP DNA Maxi kit (Qiagen,
131 Hilden, Germany).

132 **Genotyping and data quality control**

133 Cases were genotyped for 551,839 variants using the HumanCoreExome array (HumanCoreExome-24v1-
134 1_A, Illumina, San Diego, CA, USA). For the controls, genotypes of 569,645 variants were generated with a
135 customized Illumina HumanCoreExome array (“HumanCoreExome_Goncalo_15038949_A”) as described
136 previously.²²

137 Quality control was carried out according to the protocol described by Anderson et al²⁵, modified as follows.
138 Reverse and ambiguous strand SNPs were detected using snpflip (<https://github.com/biocore-ntnu/snpflip>,
139 accessed March 24, 2017) and flipped or excluded. Only SNPs common to both arrays were included.
140 Individuals with discordant sex information, missing genotype rate >0.05, or heterozygosity more than 3
141 standard deviations from the mean were excluded. Related individuals were detected by calculating pairwise
142 identity by descent (IBD), and the individual with the lower genotyping rate in any pair with IBD >0.185
143 was removed. Ancestry outliers were identified by principal component analysis (PCA) using
144 EIGENSTRAT,²⁶ and removed. Markers were excluded if they had missing genotype rate >3%, significantly
145 different missing data rates between cases and controls, minor allele frequency (MAF) <0.01, or deviated
146 significantly ($P < 10^{-5}$) from Hardy–Weinberg equilibrium. Following all exclusions, there were 522 cases
147 and 655 controls genotyped for 264,115 common platform SNPs.

148 **Genomic imputation and association analysis**

149 We phased autosomal genotype data using Eagle (version 2.3.5)²⁷ and then imputed genotypes on the basis
150 of the EUR subset of the 1000 Genomes Project reference panel (Phase III, version 5)²⁸ using Minimac3
151 (version 2.0.1).²⁹ We excluded indels, SNPs within 5bp of an indel, rare variants (MAF <0.01), and variants
152 with poor imputation quality ($R^2 < 0.8$). This filtering yielded a total of 6,252,612 quality-controlled variants,
153 including 250,964 genotyped variants. Association analysis was performed on most-likely genotypes under a
154 logistic regression model using PLINK (version 1.90)³⁰ using the first three principal components as
155 covariates. A P value less than 5×10^{-8} was considered significant.

156 **Replication and meta-analysis**

157 The USA cohort has been previously described.¹⁸ Briefly, clinically affected Caucasian keratoconus cases (n
158 = 240) were enrolled into the GWAS³¹ as a part of the longitudinal videokeratography and genetic study at
159 the Cornea Genetic Eye Institute.⁶ After removing samples with poor genotyping quality, 222 samples were
160 included in the analysis. Caucasian controls ($n = 3324$) were obtained from the Cardiovascular Health Study
161 (CHS), a population-based cohort study of risk factors for cardiovascular disease and stroke in adults 65
162 years of age or older, recruited at 4 field centres.^{32,33} CHS was approved by the Institutional Review Board at
163 each recruitment site, and subjects provided informed consent for the use of their genetic information.
164 African-American CHS participants were excluded from analysis due to insufficient number of ethnically-
165 matched cases. Participants did not have an eye examination to exclude keratoconus. The samples included
166 in the analysis were from self-reported Caucasians. Outliers detected by PCA were excluded, and the
167 analysis was adjusted for the top 3 principal components.

168 IMPUTE version 2.3.0 was used to perform imputation of the genotyping data from 370K BeadChip arrays
169 (Illumina) in keratoconus patients and CHS Caucasian controls using 1000 Genomes Phase I data as the
170 reference panel. All SNPs with $P < 10^{-6}$ identified in the discovery analysis were looked up and extracted,
171 with the exception of 4 SNPs at the *CSNK1E* locus that were not imputed in the USA study.

172 SNPs with $P < 10^{-6}$ in the discovery cohort were selected for genotyping in additional replication cohorts
173 using the MassARRAY® System (Agena Bioscience, San Diego, CA, USA) by the Australian Genome
174 Research Facility. SNPs were chosen from each locus that were compatible with the assay design, with

175 preference given to SNPs with the smallest P value in the discovery cohort. Twenty-seven SNPs were
176 genotyped in 186 additional cases recruited under the same protocol as the discovery cohort as well as in 145
177 patients from Northern Ireland, described previously,¹⁴ for a total of 331 cases. Cases were compared to 229
178 unaffected examined controls, consisting of 84 individuals from the Blue Mountains Eye Study not included
179 in the discovery cohort and 145 older individuals recruited from nursing homes in the Launceston area of
180 Tasmania, Australia. All controls underwent a thorough ocular examination and keratoconus was excluded.
181 These SNPs were also genotyped in a replication cohort from Melbourne, Victoria, consisting of 265 cases
182 and 702 examined controls described previously.¹⁵

183 Association was assessed in each cohort individually using logistic regression. Meta-analysis of results from
184 discovery and replication cohorts was performed using METAL.³⁴

185 **Functional annotation of associated variants**

186 The lead SNP at the novel locus on chromosome 11, rs61876744, was queried in HaploReg V4.1
187 (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>; accessed 28 May 2019),³⁵ including data
188 from the Genotype–Tissue Expression (GTEx) pilot analysis.³⁶ Genes in the associated region were assessed
189 for ocular tissue expression using the Ocular Tissue Database (<https://genome.uiowa.edu/otdb/>)³⁷ and for
190 differential expression between corneas from keratoconus and myopia patients in a previously published
191 study.³⁸

193 **Results**

194 **Genome-wide association testing of the discovery cohort**

195 Genome-wide association analysis was conducted in the discovery cohort (Figure 1, eFigure 1 in the
196 Supplement). The genomic inflation factor was $\lambda_{1000} = 1.023$ and all included samples were of European
197 ancestry (eFigure 2 in the Supplement). Two novel loci reached genome-wide significance (defined as $P < 5$
198 $\times 10^{-8}$), with $P = 7.46 \times 10^{-9}$ at rs61876744 in the *PNPLA2* gene on chromosome 11 and $P = 6.35 \times 10^{-12}$ at
199 rs138380 2.2kb upstream of the *CSNK1E* gene on chromosome 22. One additional locus was identified with
200 $P < 10^{-6}$, in the *MAML2* gene on chromosome 11 (rs10831500, $P = 3.91 \times 10^{-7}$). Association results for all

201 SNPs with $P < 10^{-6}$ are shown in eTable 1 in the Supplement. Locus-specific plots for all 3 loci are shown in
202 eFigure 3 in the Supplement.

203 The locus on chromosome 11 showing genome-wide significance included 25 SNPs with $P < 10^{-6}$ spanning
204 6 protein-coding genes (*CEND1*, *SLC25A22*, *PANO1*, *PIDD1*, *RPLP2* and *PNPLA2*) and multiple RNA
205 genes.

206 Because previously-reported keratoconus risk loci were initially identified in GWAS for CCT, we assessed
207 each of the loci reaching suggestive significance in the current analysis in our previously reported meta-
208 analysis for CCT. As shown in Table 2, only the *MAML2* locus shows nominal association with CCT. We
209 also looked up the lead SNPs from our CCT meta-analysis in the results from our discovery cohort (eTable 2
210 in the Supplement). SNPs at *FNDC3B*, *MPDZ*, and *SMAD3* show nominal association in our analysis.

211 **Association testing of the replication cohorts**

212 All SNPs with $P < 10^{-6}$ identified in the discovery cohort were analysed in the previously generated GWAS
213 data for the USA cohort, and 27 SNPs compatible with a single assay design were genotyped in the AUS+NI
214 and VIC replication cohorts. (Table 3 and eTable 3 in the Supplement).

215 Multiple SNPs in both novel loci at *PNPLA2* and *MAML2* showed association with $p < 0.05$ in the USA
216 cohort. None of the SNPs in the *CSNK1E* locus reached significance, although all showed the same direction
217 of association as in the discovery cohort. None of the 3 loci reached significance in either the AUS+NI or the
218 VIC replication cohort, but most SNPs showed the same direction of association as the discovery cohort.
219 MAFs for cases and controls in each cohort are given in eTable 3 in the Supplement.

220 **Meta-analysis**

221 Meta-analysis of the combined data from the discovery and all 3 replication cohorts found 12 SNPs at the
222 novel *PNPLA2* locus to be associated with keratoconus at genome-wide significance ($P < 5 \times 10^{-8}$). The
223 *MAML2* locus on chromosome 11 showed suggestive association with $P = 3.83 \times 10^{-6}$ at rs10831500. The
224 *CSNK1E* locus on chromosome 22 reached $P = 3.18 \times 10^{-4}$ at rs138378. For all 3 loci, there were other SNPs
225 with smaller P values on meta-analysis, but these did not include data for all replication cohorts (eTable 3 in
226 the Supplement).

227 **Functional annotations of novel associated loci**

228 At the significant locus on chromosome 11, the lead SNP rs61876744 is located in the second intron of the
229 *PNPLA2* gene (NM_020376.3). The associated region extends for around 40kbp encompassing multiple
230 transcripts. The *PNPLA2* gene is highly expressed in all eye tissues assessed in the ocular tissue database,
231 including cornea, as are other protein-coding genes at this locus (eTable 4 in the Supplement). A recent study
232 compared gene expression in corneal epithelium from keratoconus and myopia patients,³⁸ and found that
233 *PNPLA2* and *PIDD1* were differentially expressed with a false discovery rate (FDR) < 0.05 (eTable 4 in the
234 Supplement) while *RPLP2* and *CSNK1E* are significant at FDR < 0.1.

235 HaploReg identified 7 SNPs in strong linkage disequilibrium with the lead SNP (rs61876744), and all report
236 an eQTL for an antisense RNA transcript *AP006621* (multiple transcripts 1–8) in multiple tissues assessed in
237 GTex (eTable 5 in the Supplement) where the more common allele, C, is associated with increased transcript
238 levels (eFigure 4 in the Supplement). This RNA gene is not represented in the ocular tissue database. A
239 similar trend is seen in GTex for the *PIDD1* gene in sun exposed skin (eFigure 4 in the Supplement).

240 **Discussion**

241 This study has identified a candidate locus for keratoconus on chromosome 11 that shows replication in the
242 USA data and consistent direction of association in the other cohorts. The lead SNP is located in an intron of
243 *PNPLA2*. This gene encodes Patatin-like phospholipase domain-containing protein 2, which catalyzes the
244 initial step in triglyceride hydrolysis. The relevance of this pathway to keratoconus is not obvious, but it is
245 well known that the closest gene to an association signal is not necessarily the causative gene. At least 4
246 other protein coding genes at this locus are also expressed in the cornea, and RNA coding genes are also
247 annotated in the region. There is a strong eQTL signal of the lead SNP rs61876744 for an antisense RNA
248 gene, *AP006621.8* located on the opposite strand to the protein coding *PNPLA2* gene. The antisense RNA
249 *AP006621* transcripts may have a role in regulating *PNPLA2* or other genes at this locus and elsewhere. The
250 minor allele at rs61876744, T, is associated with reduced risk of keratoconus and with reduced expression of
251 *AP006621* in many tissues. This suggests that overexpression of *AP006621* may destabilise corneal
252 structures. Oxidative stress and apoptosis have been suggested as part of the pathogenesis of keratoconus⁵
253 and sun (or UV light) exposure is known to trigger oxidative stress and DNA damage pathways.³⁹ Several

254 genes at this locus likely play a role in apoptotic pathways, including *PPID* (p53-induced death domain
255 protein 1) and *PANO1* (proapoptotic nucleolar protein 1).

256 The chromosome 11 locus overlaps with a previously reported (although not genome-wide significant)
257 association signal for Fuchs Endothelial Corneal Dystrophy (FECD).⁴⁰ The lead SNP in the FECD GWAS is
258 rs12223324 in the *PDDC1* gene, upstream of *PNPLA2*. This SNP does not reach significance in our
259 keratoconus GWAS, ($P = 2.02 \times 10^{-5}$). It is not known how this locus might lead to FECD, but the overlap
260 of genetic association with keratoconus is intriguing, given both diseases affect the cornea. Although rare,
261 there are reports in the literature of patients with both FECD and keratoconus.^{41,42} The participants in the
262 current study do not have FECD and thus this disease does not account for the association observed here.

263 Although we observed the strongest association in the discovery cohort at the *CSNK1E* locus, this result was
264 not replicated. The signal appears to be driven by a single genotyped SNP that has influenced the imputation
265 of a surrounding LD block (see eTable 1; eFigure 3C in the Supplement). The signal at the *MAML2* locus is
266 supported by the USA replication cohort. Further replication of these loci is required before any firm
267 conclusions can be drawn.

268 All control cohorts used in this study were older than the case cohorts, thus it will be important to assess
269 these loci for age-related effects in future studies. Batch effects are a potential problem in an analysis where
270 cases and controls are genotyped separately; however, the low inflation factor seen in our analysis reassures
271 us that batch effects are unlikely to be having a major impact.

272 Previous GWAS for keratoconus (using a subset of cases involved in the current study) reported
273 *RAB3GAP1*³¹ and a region upstream of the *HGF*,¹⁴ although neither study reached genome-wide
274 significance. The lead SNPs at these loci reach P values of 0.066 and 0.038 respectively in the current
275 discovery cohort, suggesting these loci are not major contributors in this better powered study. Previously
276 reported genome-wide significant keratoconus associated loci (*RXRA–COL5A1*, *FOXO1* and *FNDC3B*) are
277 also associated with CCT. The current findings suggest that mechanisms other than susceptibility to a thinner
278 cornea may also be at play in the genetic risk of keratoconus.

279 In summary, we have identified a locus for keratoconus on chromosome 11. The lead SNP is in an intron of
280 the *PNPLA2* gene and is an eQTL for a long non-coding RNA *AP006621.8*. We also suggest loci near

281 *MAML2* and *CSNK1E* that require further replication. It is very likely that additional risk loci exist for
282 keratoconus, and larger studies will be needed to identify them.

283 **Acknowledgements**

284 **Funding/Support:** The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund
285 of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA,
286 NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the
287 GTEx Portal on 05/27/2019.

288 This study was supported by the Australian National Health and Medical Research Council (NHMRC)
289 project grant GNT1104700 and Senior Research Fellowships (1138585 to PN Baird and 1059954 to KP
290 Burdon). The Centre for Eye Research Australia (CERA) receives Operational Infrastructure Support from
291 the Victorian Government. The discovery case cohort was funded by an NHMRC Centre for Research
292 Excellence (1023911). Control genotype data for the discovery cohort were provided by the international
293 AMD Genetics Consortium genotyped under The Center for Inherited Diseases Research (CIDR) Program
294 contract number HHSN268201200008I.

295 The USA replication cohort is supported in part by the National Eye Institute contract R01 EY009052. The
296 provision of genotyping data was supported in part by the National Center for Advancing Translational
297 Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney
298 Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes
299 Endocrinology Research Center. The CHS (controls) was supported by contracts HHSN268201200036C,
300 HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080,
301 N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114
302 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National
303 Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629
304 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be
305 found at CHS–NHLBI.org. The content is solely the responsibility of the authors and does not necessarily
306 represent the official views of the National Institutes of Health.

307 **Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection,
308 management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and
309 decision to submit the manuscript for publication.

310 **Author Contributions:** Drs McComish and Burdon had full access to all the data in the study and take
311 responsibility for the integrity of the data and the accuracy of the data analysis.

312 *Study concept and design:* Burdon, Craig, Baird, Bykhovskaya, MacGregor, Willoughby.

313 *Acquisition, analysis, or interpretation of data:* McComish, Sahebjada, Bykhovskaya, Willoughby,
314 Richardson, Tenen, Charlesworth, Mitchell, Lucas, Mills, Mackey, Li, Wang, Jensen, Rotter, Taylor, Hewitt,
315 Rabinowitz, Baird, Craig, Burdon.

316 *Drafting of the manuscript:* McComish, Burdon.

317 *Critical revision of the manuscript for important intellectual content:* McComish, Burdon, Lucas, Mackey,
318 Bykhovskaya, Willoughby, Tenen, Mills, MacGregor.

319 *Statistical analysis:* McComish, Li.

320 *Obtained funding:* Burdon, Craig, Baird, Rabinowitz, Hewitt, Rotter.

321 *Study supervision:* Craig, Burdon, Mills.

322 **Conflict of Interest:**

323 None of the authors has any conflicts of interest to disclose.

324

325 **References**

- 326 1. Rabinowitz YS. Keratoconus. *Surv Ophthalmol*. 1998;42:297-319.
- 327 2. Pearson AR, Soneji B, Sarvananthan N, Sandford-Smith JH. Does ethnic origin influence the
328 incidence or severity of keratoconus? *Eye*. 2000;14:625-628.
- 329 3. Godefrooij DA, de Wit GA, Uiterwaal CS, Imhof SM, Wisse RP. Age-specific incidence and
330 prevalence of keratoconus: A nationwide registration study. *Am J Ophthalmol*. 2017;175:169-172.
- 331 4. Zadnik K, Barr JT, Edrington TB, et al. Baseline findings in the Collaborative Longitudinal
332 Evaluation of Keratoconus (CLEK) Study. *Invest Ophthalmol Vis Sci*. 1998;39:2537-2546.
- 333 5. Wojcik KA, Blasiak J, Szaflik J, Szaflik JP. Role of biochemical factors in the pathogenesis of
334 keratoconus. *Acta Biochim Pol*. 2014;61:55-62.
- 335 6. Wang Y, Rabinowitz YS, Rotter JI, Yang H. Genetic epidemiological study of keratoconus:
336 Evidence for major gene determination. *Am J Med Genet*. 2000;93:403-409.
- 337 7. Nowak DM, Gajecka M. The genetics of keratoconus. *Middle East Afr J Ophthalmol*. 2011;18:2-6.
- 338 8. Abu-Amero KK, Al-Muammar AM, Kondkar AA. Genetics of keratoconus: Where do we stand? *J*
339 *Ophthalmol*. 2014;2014:641708.
- 340 9. Burdon KP, Vincent AL. Insights into keratoconus from a genetic perspective. *Clin Exp Optom*.
341 2013;96:146-154.
- 342 10. Czugala M, Karolak JA, Nowak DM, et al. Novel mutation and three other sequence variants
343 segregating with phenotype at keratoconus 13q32 susceptibility locus. *Eur J Hum Genet*.
344 2012;20:389-397.
- 345 11. Bykhovskaya Y, Li X, Epifantseva I, et al. Variation in the lysyl oxidase (*LOX*) gene is associated
346 with keratoconus in family-based and case-control studies. *Invest Ophthalmol Vis Sci*. 2012;53:4152-
347 4157.
- 348 12. Hughes AE, Bradley DT, Campbell M, et al. Mutation altering the miR-184 seed region causes
349 familial keratoconus with cataract. *Am J Hum Genet*. 2011;89:628-633.
- 350 13. Lechner J, Bae HA, Guduric-Fuchs J, et al. Mutational analysis of *MIR184* in sporadic keratoconus
351 and myopia. *Invest Ophthalmol Vis Sci*. 2013;54:5266-5272.

- 352 14. Burdon KP, Macgregor S, Bykhovskaya Y, et al. Association of polymorphisms in the hepatocyte
353 growth factor gene promoter with keratoconus. *Invest Ophthalmol Vis Sci.* 2011;52:8514-8519.
- 354 15. Sahebjada S, Schache M, Richardson AJ, Snibson G, Daniell M, Baird PN. Association of the
355 hepatocyte growth factor gene with keratoconus in an Australian population. *PloS one.*
356 2014;9:e84067.
- 357 16. Bae HA, Mills RA, Lindsay RG, et al. Replication and meta-analysis of candidate loci identified
358 variation at *RAB3GAP1* associated with keratoconus. *Invest Ophthalmol Vis Sci.* 2013;54:5132-
359 5135.
- 360 17. Li X, Bykhovskaya Y, Haritunians T, et al. A genome-wide association study identifies a potential
361 novel gene locus for keratoconus, one of the commonest causes for corneal transplantation in
362 developed countries. *Hum Mol Genet.* 2012;21:421-429.
- 363 18. Lu Y, Vitart V, Burdon KP, et al. Genome-wide association analyses identify multiple loci
364 associated with central corneal thickness and keratoconus. *Nat Genet.* 2013;45:155-163.
- 365 19. Iglesias AI, Mishra A, Vitart V, et al. Cross-ancestry genome-wide association analysis of corneal
366 thickness strengthens link between complex and Mendelian eye diseases. *Nat Commun.*
367 2018;9:1864.
- 368 20. Burdon KP, Coster DJ, Charlesworth JC, et al. Apparent autosomal dominant keratoconus in a large
369 Australian pedigree accounted for by digenic inheritance of two novel loci. *Hum Genet.*
370 2008;124:379-386.
- 371 21. Rabinowitz YS. Videokeratographic indices to aid in screening for keratoconus. *J Refract Surg.*
372 1995;11:371-379.
- 373 22. Fritsche LG, Igl W, Bailey JN, et al. A large genome-wide association study of age-related macular
374 degeneration highlights contributions of rare and common variants. *Nat Genet.* 2016;48:134-143.
- 375 23. Burdon KP, Mitchell P, Lee A, et al. Association of open-angle glaucoma loci with incident
376 glaucoma in the Blue Mountains Eye Study. *Am J Ophthalmol.* 2015;159:31-36.e31.
- 377 24. Dimasi DP, Burdon KP, Hewitt AW, et al. Genetic investigation into the endophenotypic status of
378 central corneal thickness and optic disc parameters in relation to open-angle glaucoma. *Am J*
379 *Ophthalmol.* 2012;154:833-842.e832.

- 380 25. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality
381 control in genetic case-control association studies. *Nat Protoc.* 2010;5:1564-1573.
- 382 26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components
383 analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904-909.
- 384 27. Loh P-R, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference
385 Consortium panel. *Nat Genet.* 2016;48:1443-1448.
- 386 28. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature.*
387 2015;526:68-74.
- 388 29. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat*
389 *Genet.* 2016;48:1284-1287.
- 390 30. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:
391 rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
- 392 31. Li X, Bykhovskaya Y, Haritunians T, et al. A genome-wide association study identifies a potential
393 novel gene locus for keratoconus, one of the commonest causes for corneal transplantation in
394 developed countries. *Hum Mol Genet.* 2012;21(2):421-429.
- 395 32. Fried LP, Borhani NO, Enright P, et al. The cardiovascular health study: Design and rationale. *Ann*
396 *Epidemiol.* 1991;1:263-276.
- 397 33. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for heart and aging research in genomic
398 epidemiology (CHARGE) consortium: Design of prospective meta-analyses of genome-wide
399 association studies from 5 cohorts. *Circ Cardiovasc Genet.* 2009;2:73-80.
- 400 34. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association
401 scans. *Bioinformatics.* 2010;26:2190-2191.
- 402 35. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and
403 regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*
404 2012;40:D930-934.
- 405 36. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene
406 regulation in humans. *Science.* 2015;348:648-660.
- 407 37. Wagner AH, Anand VN, Wang WH, et al. Exon-level expression profiling of ocular tissues. *Exp Eye*
408 *Res.* 2013;111:105-111.

- 409 38. You J, Corley SM, Wen L, et al. RNA-Seq analysis and comparison of corneal epithelium in
410 keratoconus and myopia patients. *Sci Rep.* 2018;8:1-13.
- 411 39. Birch-Machin MA, Russell EV, Latimer JA. Mitochondrial DNA damage as a biomarker for
412 ultraviolet radiation exposure and oxidative stress. *Br J Dermatol.* 2013;169 Suppl 2:9-14.
- 413 40. Afshari NA, Igo RP, Jr., Morris NJ, et al. Genome-wide association study identifies three novel loci
414 in Fuchs endothelial corneal dystrophy. *Nat Commun.* 2017;8:14898.
- 415 41. Jurkunas U, Azar DT. Potential complications of ocular surgery in patients with coexistent
416 keratoconus and Fuchs' endothelial dystrophy. *Ophthalmology.* 2006;113(12):2187-2197.
- 417 42. Vira S, Abugo U, Shih CY, et al. Descemet stripping endothelial keratoplasty for the treatment of
418 combined fuchs corneal endothelial dystrophy and keratoconus. *Cornea.* 2014;33(1):1-5.
- 419
- 420

421 **Figure 1:** Manhattan plot of association results in the discovery cohort. Results of logistic
422 regression with the first three PCs as covariates ($-\log_{10} p$ -values) are plotted for each chromosome.
423 The red and blue lines represent the genome-wide significance threshold of $p = 5 \times 10^{-8}$ and our
424 threshold for follow-up of $p = 10^{-6}$, respectively.

425

426 **Table 1:** Demographics of the Australian discovery cohort and the three replication cohorts.

427

Cohort	Cases			Controls		
	N	% male	Mean age [SD]	N	% male	Mean age [SD]
Discovery: Australia	522	56	45 [15.2]	655	47	65 [10.6]
Replication: USA	222	55	44 [13.3]	2927	39	72 [5.4]
Replication: AUS+NI	331	61	41 [15.9]	229	37	75 [11.5]
Replication: VIC	265	60	35 [14.9]	702	38	52 [15.2]

428

429 **Table 2:** Lead SNPs at all three loci with $p < 1 \times 10^{-6}$ in discovery cohort.

430

Locus ^a	Chr	Lead SNP	Base pair ^b	A1/2	F_A ^c	F_U ^c	OR ^d [95% CI]	P	P (CCT) ^e
<i>PNPLA2</i>	11	rs61876744	820754	T/C	0.341	0.447	0.59 [0.49–0.71]	7.46×10^{-9}	0.553
<i>MAML2</i>	11	rs10831500	95982642	G/T	0.422	0.330	1.59 [1.33–1.91]	3.91×10^{-7}	0.014
<i>CSNK1E</i>	22	rs138380	38796629	G/A	0.384	0.524	0.49 [0.40–0.60]	6.35×10^{-12}	0.790

431

432 ^aLocus assigned to the RefSeq protein-coding gene within or near the association signal interval.

433 ^bGenomic positions are based on hg19.

434 ^cF_A, minor allele frequency in cases; F_U, minor allele frequency in controls.

435 ^dOdds ratios with respect to A1.

436 ^eP-values in our previously reported analysis for CCT.¹⁹

437

438 **Table 3:** Replication and meta-analysis of association results at lead SNPs.

439

Locus ^a	Chr	Lead SNP ^b	Base pair ^c	A1/2	Discovery		Replication:USA		Replication: AUS+NI		Replication: VIC		Meta-analysis	
					OR [95% CI]	<i>P</i>	OR	<i>P</i>	OR [95% CI]	<i>P</i>	OR [95% CI]	<i>P</i>	<i>P</i>	direction
<i>PNPLA2</i>	11	rs61876744	820754	T/C	0.59 [0.49–0.71]	7.46×10^{-9}	0.77 [0.63–0.95]	8.88×10^{-4}	0.84 [0.65–1.08]	0.179	0.97 [0.79–1.19]	0.763	2.45×10^{-8}	----
<i>MAML2</i>	11	rs10831500	95982642	G/T	1.59 [1.33–1.91]	3.91×10^{-7}	1.35 [1.11–1.65]	5.69×10^{-3}	1.09 [0.85–1.40]	0.515	1.03 [0.84–1.26]	0.792	3.83×10^{-6}	++++
<i>CSNK1E</i>	22	rs138378	38796159	A/G	0.50 [0.41–0.61]	1.77×10^{-11}	0.95 [0.78–1.16]	0.790	0.97 [0.76–1.24]	0.818	0.93 [0.76–1.13]	0.448	3.18×10^{-4}	----

440

441 ^aLocus assigned to the RefSeq protein-coding gene within or near the association signal interval.

442 ^bThe lead SNP at *CSNK1E* (rs138380) was not successfully genotyped in the AUS+NI and VIC cohorts, and a proxy (rs138378, $R^2=0.938$) was used for replication
443 instead.

444 ^cGenomic positions are based on hg19.

445

