COMMUNICATIONS BIOLOGY

ARTICLE

https://doi.org/10.1038/s42003-020-01421-2 OPEN

Common variants in *SOX-2* and congenital cataract genes contribute to age-related nuclear cataract

Ekaterina Yonova-Doing et al.[#]

Nuclear cataract is the most common type of age-related cataract and a leading cause of blindness worldwide. Age-related nuclear cataract is heritable ($h^2 = 0.48$), but little is known about specific genetic factors underlying this condition. Here we report findings from the largest to date multi-ethnic meta-analysis of genome-wide association studies (discovery cohort $N = 14{,}151$ and replication $N = 5299$) of the International Cataract Genetics Consortium. We confirmed the known genetic association of *CRYAA* (rs7278468, $P = 2.8 \times 10^{-16}$) with nuclear cataract and identified five new loci associated with this disease: *SOX2-OT* (rs9842371, $P = 1.7 \times 10^{-19}$), *TMPRSS5* (rs4936279, $P = 2.5 \times 10^{-10}$), *LINC01412* (rs16823886, $P = 1.3 \times 10^{-9}$), *GLTSCR1* (rs1005911, $P = 9.8 \times 10^{-9}$), and *COMMD1* (rs62149908, $P = 1.2 \times 10^{-8}$). The results suggest a strong link of age-related nuclear cataract with congenital cataract and eye development genes, and the importance of common genetic variants in maintaining crystalline lens integrity in the aging eye.

[#]A list of authors and their affiliations appears at the end of the paper.

Age-related cataract is the leading cause of blindness, accounting for more than one-third of blindness worldwide[1,2]. Cataract is an opacification of the lens of the eye, resulting in reduced vision, glare and decreased ability to perform daily activities. Although surgery is often effective in restoring vision, its costs to health-care systems are considerable[3]. The prevalence of cataract and the number of cataract surgeries is projected to rise globally, as the population ages[4,5], and so will the costs of cataract to society.

The most frequent form of age-related cataract, nuclear cataract (15 year cumulative incidence of 49.6% in individuals aged 65–74 years) affects the lens nucleus[6]. Susceptibility to age-related nuclear cataract (ARNC) was conferred by a mixture of genetic and environmental risk factors: up to half on nuclear cataract variation is due to genetic risk factors[7], while smoking[8], obesity[9] and diet[10] are potentially modifiable exposures associated with ARNC.

Despite the public health significance of ARNC, relatively little is known about its underlying genetic factors. To date, genome-wide association studies (GWAS) have not been very successful in the identifying common genetic variants for age-related cataract, partly due to the difficulties in objectively phenotyping ARNC. Studies using cataract surgery (either self-reported or based on information from electronic health record) as a proxy for the presence of cataract has been challenging, as the severity of cataract when cataract surgery is done varies greatly among individuals[11,12]. On top of this, there are three major subtypes of age-related cataract (i.e., nuclear, cortical and subcapsular cataract); each of them may have different pathophysiology. To date, the only reported GWAS of ARNC with objective phenotyping via lens photos and detailed cataract grading was done in Asian cohorts, where two genetic loci (*CRYAA*, *KCNAB1*) were found associated with ARNC[13]. The *CRYAA* gene encodes for most abundant structural protein present in the lens and mutations in this gene cause congenital cataracts[14,15]. *KCNAB1* encodes voltage-gated potassium channel, previously linked to ageing bone phenotypes[16]. However, a more recent exome array analysis of ~1500 Europeans failed to find any variants associated at genome-wide significance[17]. Previous analysis of poorly defined (self-report) cataract phenotypes from the UK Biobank (http://www.nealelab.is/uk-biobank, https://www.leelabsg.org/resources) found no common variant associations. A GWAS of retinal detachment in UK Biobank found 20 loci associated with cataract surgery, likely reflecting several age-related cataract subtypes[18]. We are not aware of any other GWAS studies of cataract subtypes, other than for age-related diabetic cataract: a small Taiwanese study found several suggestive loci and a recent larger European-ancestry GWAS identified *CACNA1C* gene at GWAS significance[19,20].

Given the potential of appropriately powered genetic studies to reveal aetiologies and pathways of ARNC, we aimed to identify additional genomic regions associated with the susceptibility to ARNC via a meta-analysis of GWAS of 12 well-phenotyped studies from the International Cataract Genetics Consortium. We replicated genetic association of *CRYAA* (rs7278468, $P = 2.8 \times 10^{-16}$) with nuclear cataract and identified six new loci associated with this disease. The results suggest a strong link of ARNC with genes linked to congenital cataract and eye development, as well as and the importance of common genetic variants in maintaining crystalline lens integrity during ageing.

Results

The results from the meta-analysis of 8.5 million variants in eight studies (Supplementary Fig. 1 and Supplementary Tables 1–3) followed a polygenic model with no evidence of population structure (meta-analysis genomic inflation factor $\lambda = 1.009$, Supplementary Table 4 and Supplementary Fig. 2). In the discovery stage we found three loci to be associated at genome-wide significance (Fig. 1) and this number increased to six after the all-data meta-analysis stage (Supplementary Figs. 3–6). As expected for a common age-related trait, the majority of associated variants or variants in LD with those were situated outside of coding regions and we observed suggestive depletion of intronic variants and enrichment in ncRNA and upstream variants (Supplementary Fig. 4).

We confirmed the *CRYAA* genomic region previously found significantly associated with ARNC score at a GWAS-significant level. The strongest evidence for association was found for rs7278468 ($\beta = 0.08$; $P = 3.6 \times 10^{-17}$), just upstream of the *CRYAA* gene transcript. However, *KCNAB1* variants that were previously reported in association with ARNC[13] were rare in Europeans (MAF = 0.03) and were not significantly associated in this meta-analysis ($\beta = 0.04$; $P = 0.02$ for *KCNAB1* rs55818638). In addition, we identified two novel susceptibility regions that at this stage were significantly associated with ARNC (Table 1 and Supplementary Figs. 3 and 5). Markers located on chromosome 3q26.33, in proximity of the *SOX2* gene and within its regulator, *SOX2-OT*, were significantly associated with the ARNC score ($\beta = 0.07$; $P_{discovery} = 2.6 \times 10^{-12}$ for rs9842371). The *SOX2* locus has not previously been associated with nuclear cataract but was associated with cataract surgery in UK Biobank[18].

A second novel susceptibility genetic locus significantly associated with ARNC score was located on chromosome 11.q23.2 and overlapped with the genomic sequence of the *TMPRSS5* gene ($\beta = 0.06$; $P_{discovery} = 4.2 \times 10^{-11}$ for rs4936279). Furthermore, a third locus, overlapping with the *COMMD1* gene-coding region, also approached genome-wide significance in this meta-analysis ($\beta = -0.06$; $P_{discovery} = 6.5 \times 10^{-8}$ for rs62149908). Among the genes that were associated at suggestive, but not GWAS-significant levels overall, ancestry-specific significant associations were observed at chromosome 13q12.11 in Asians ($\beta = 0.07$; $P_{Asians} = 2.7 \times 10^{-8}$ for rs4769087) within the *GJA3* genomic sequence, and on chromosome 11q23.1 in Europeans upstream of *CRYAB* ($\beta = 0.07$; $P_{Europeans} = 2.5 \times 10^{-5}$ for rs10789852).

Genome-wide associated SNPs showing suggestive association ($P < 10^{-6}$) in the discovery phase were taken forward to the replication stage of this study (Table 1). Despite the smaller sample size for replication, four out of nine markers tested showed nominal replication ($P < 0.05$, Supplementary Fig. 7).

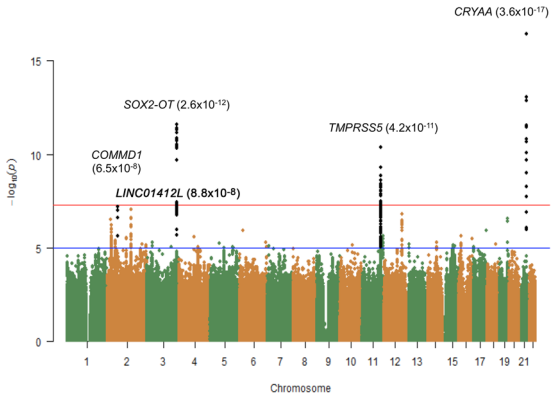

Fig. 1 Manhattan plot of the GWAS meta-analysis for age-related nuclear cataract in the combined analysis ($N = 14,151$). The plot shows −log10-transformed *P* values for all SNPs; the upper horizontal line represents the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$; the lower line indicates a *P* value of 10^{-5}. Data of both directly genotyped and imputed SNPs are presented in the Manhattan plot.

Table 1 Genome-wide significant associations for age-related nuclear cataract.

					Discovery					Replication		Meta-analysis	
SNP	Chr.	Pos.	Nearest gene	A1	A2	EAF European	EAF Asian	β (SE)	$P_{discovery}$	P_{het}	OR (SE)	$P_{replication}$	Z-score	$P_{combined}$
rs61185326	2	20747778	intragenic	T	A	0.04	0.02	0.02 (0.04)	3.0×10^{-7}	0.09	1.00 (0.21)	0.99	4.60	4.2×10^{-6}
rs13021828	2	24439276	*ITSN2*	C	G	0.38	0.39	−0.05 (0.01)	6.1×10^{-7}	0.24	1.07 (0.05)	0.17	−3.54	4.0×10^{-4}
rs62149908[a]	2	62191878	*COMMD1*	T	C	0.22	0.32	−0.06 (0.04)	6.5×10^{-8}	0.04	0.89 (0.06)	0.04	−5.70	1.2×10^{-8}
rs16823886	2	145341259	*LINC01412, ZEB2*	A	G	0.13	0.18	−0.06 (0.01)	8.8×10^{-8}	0.02	0.86 (0.05)	3.9×10^{-3}	−6.07	1.3×10^{-9}
rs9842371	3	181346937	*SOX2*	T	C	0.35	0.54	0.07 (0.01)	2.6×10^{-12}	0.11	1.31 (0.05)	4.4×10^{-9}	9.03	1.7×10^{-19}
rs4936279[a]	11	113566207	*TMPRSS5*	A	C	0.30	0.48	0.06 (0.01)	4.2×10^{-11}	0.35	1.07 (0.05)	0.18	6.33	2.5×10^{-10}
rs11067211[a]	12	109988214	*MMAB*	G	A	0.26	0.16	0.06 (0.01)	1.6×10^{-7}	0.76	1.09 (0.06)	0.14	5.24	1.6×10^{-7}
rs1005911	19	48206092	*GLTSCR1*	G	T	0.25	0.36	−0.05 (0.01)	2.8×10^{-7}	0.84	0.87 (0.05)	9.5×10^{-3}	−5.73	9.8×10^{-9}
rs72784684[a]	21	44588757	*CRYAA*	G	T	0.69	0.37	0.08 (0.01)	3.6×10^{-17}	0.27	1.13 (0.07)	0.06	8.18	2.8×10^{-16}

This table summarises the SNPs that were associated at genome-wide significance level ($P < 5 \times 10^{-8}$) with age-related nuclear cataract in the combined analysis (discovery phase) and/or after the replication phase. *SNP* single-nucleotide polymorphism, *Chr.* chromosome, *pos* position (NCBI build 37), *A1* reference allele 1, *A2* the other allele, *EAF* effect allele frequency, *Beta* effect size on standardised nuclear cataract scores based on the effect allele in all discovery cohorts meta-analysis, *SE* standard errors of the effect size, P_{het}, *P* value for heterogeneity, *OR* odds ratio estimated from the case–control collections in the replication phase, *Z* Z-score derived from the overall meta-analysis combining the discovery and replication phases. [a]These variants were not available in the INDEYE(S) study due to probe design issues and the following variants in high linkage disequilibrium with the lead SNP were genotyped instead: rs55785307 (*COMMD1*, $R^2 = 0.84$; $D' = 0.99$), rs11601037 (*TMPRSS5*, $R^2 = 0.90$; $D' = 1.0$), rs74486178 (*MMAB*, $R^2 = 0.83$; $D' = 0.99$) and rs870137 (*CRYAA*, $R^2 = 0.48$; $D' = 0.98$).

Another three of the SNPs failed to achieve significance, but the association in the replication meta-analysis was in the same direction as that in the discovery phase (Table 1). Notably association was replicated for markers in the *SOX2* locus (OR = 1.31; $P = 4.4 \times 10^{-9}$ for rs9842371), but the replication results were not statistically significant for the markers in the *TMPRSS5* locus, nor in the previously established *CRYAA* locus (OR = 1.13; $P = 5.6 \times 10^{-2}$ for rs7278468). Nevertheless, we observed that the direction of allele's effects was the same between the discovery stage and replication stage all SNPs (i.e., the allele associated with higher ARNC score in the discovery stage also had a odds ratio of >1 for ARNC in the replication stage), except *ITSN2* rs13021828.

An all-inclusive meta-analysis of all leading SNPs of regions associated at or close to GWAS-significance levels using both the discovery and replication loci was performed (Table 1 and Supplementary Fig. 4). In addition to the loci of *SOX2*/*SOX2-OT* ($Z = 9.03$; $P = 1.7 \times 10^{-19}$ for rs9842371), *CRYAA*, ($Z = 8.18$; $P = 2.8 \times 10^{-16}$ for rs7278468) and *TMPRSS5* ($Z = 6.33$; $P = 2.5 \times 10^{-10}$ for rs4936279), novel genome-wide significant associations were found for rs16823886 upstream of the *ZEB2* gene ($Z = -6.07$; $P = 1.3 \times 10^{-9}$), rs62149908 ($Z = -5.70$; $P = 1.2 \times 10^{-8}$), within the Copper Metabolism Domain Containing 1 (*COMMD1*) gene; for rs1005911 within the *GLTSCR1* gene ($Z = -5.73$; $P = 9.8 \times 10^{-9}$). At those loci, the following genes are expressed in lens (Supplementary Table 6): *ZEB2*, *GLTSCR1*, *NAPA*, but the eQTL and regulatory sequence analysis (Supplementary Fig. 6, Supplementary Tables 5, Supplementary Data 1) did not provide conclusive evidence on how those genes may exert their effects on ARNC formation. The eQTL analysis (Supplementary Data 1), however, found a strong association between the following SNPs and transcript levels: rs7278468 and the *CRYAA* ($P = 1.3 \times 10^{-7}$, liver tissue); rs11067211 and *MMAB* ($P = 5.3 \times 10^{-8}$, brain); rs61185326 and *RHOB* ($P = 3.0 \times 10^{-7}$, muscle); and rs10789852 and *CRYAB* ($P = 7.6 \times 10^{-22}$; fat). It is possible that similar effects are present for other genes, but at tissues and developmental stages that are not captured in the available GTEx or TwinsUK tissues.

The common variants associated at GWAS-levels with ARNC in our discovery stage analysis explained ~3% of heritability. A conditional analysis of SNPs identified from discovery phase loci (Supplementary Table 7) and a gene-based test (Supplementary Table 8) was performed on the results of the discovery stage meta-analysis, but they did not yield any additional association beyond those already reported above. Pathway analysis were a few pathways associated with ARNC (Supplementary Table 9), with the strongest enrichment observed for cholesterol biosynthesis ($P_{permuted} = 0.01$), whose importance in cataract is not clearly known.

An LDscore systematic analysis of genetic correlations suggested that the ARNC genetic risk was correlated with the following eye-related traits measured in UK Biobank: cataract (0.48), diabetes-related eye diseases (0.27) and glaucoma (0.20). In addition, there was correlation with the genetic risks of (Supplementary Fig. 8) hip (0.34) and waist (0.30) circumference, different classes of circulating lipids (median = 0.26) and age at menarche (−0.12). However, none of the correlations survived correction for multiple testing. Similarly, the Open Targets SNP and gene co-localisation results point to sharing of signals with astigmatism-related traits (*CRYAA*, *SOX2* and *GLTSCR1* loci), cardio-metabolic traits, anthropometric and blood cell traits (Supplementary Fig. 9). Of note, there was also co-localisation with smoking-related GWAS signals at the *ZEB2* and *ITSN2* loci (Supplementary Fig. 9).

Multiple variants in proximity to 47 genes linked to congenital cataract were nominally associated with ARNC in our analysis (Fig. 2 and Supplementary Data 2), but only 5 survived correction

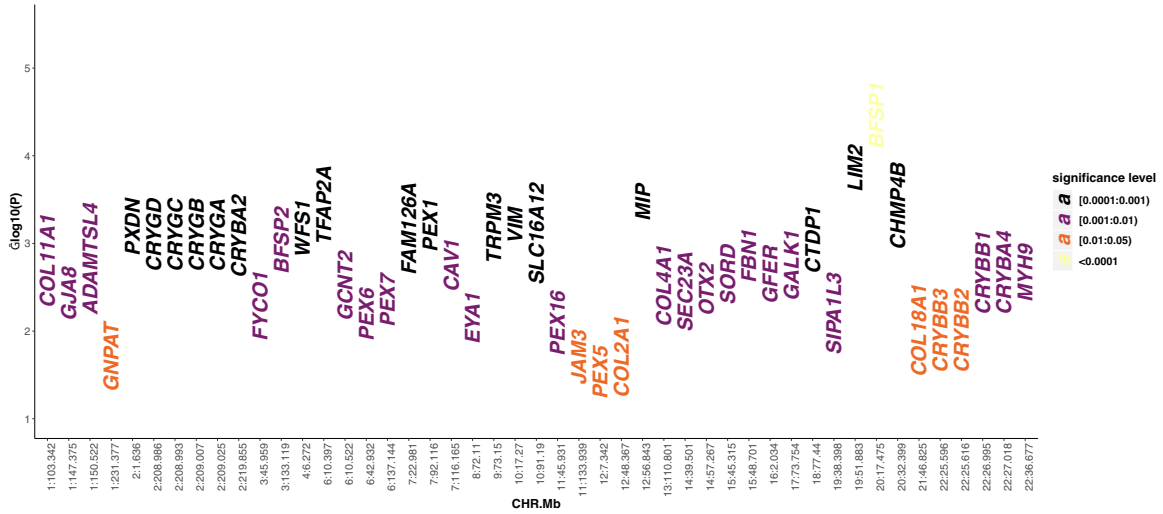

Fig. 2 Common variants in congenital cataract genes. This Manhattan plot shows the association results for the congenital cataract genes. The −log10 (*P* value) of the most strongly associated variant per gene is plotted against the gene location (in chromosome followed by mega base format: CHR.Mb). The colour code represents the strength of association in terms of *P* value.

for multiple testing ($\alpha = 5 \times 10^{-4}$) in *BFSP1* ($\beta = 0.08$; $P = 3.5 \times 10^{-5}$), *LIM2* ($\beta = 0.04$; $P = 1.4 \times 10^{-4}$), *MIP* ($\beta = 0.02$; $P = 3.4 \times 10^{-4}$), *TFAP2A* ($\beta = 0.05$; $P = 3.7 \times 10^{-4}$) and *CHMP4B* ($\beta = 0.08$; $P = 3.8 \times 10^{-4}$).

Discussion

Here we report the results of a GWAS on nuclear cataract, conducted on 14,151 participants with detailed ARNC severity phenotypes and replicated in 5299 samples. Apart from confirming association at the *CRYAA*, we increased the number of known associations by reporting five additional ARNC genetic loci.

The SOX2 Overlapping Transcript (*SOX2-OT*) encodes for a highly conserved long noncoding RNA, which overlaps and regulates *SOX2* expression. SOX2 is a single exon transcription factor, previously associated with anophthalmia[21] and coloboma[22]. Sox2 is involved in crystallin regulation in murine[23] and avian models[24] and in humans, and *SOX2* mutations cause microphthalmia and cataract[25–27].

ZEB2 is a still uncharacterised member of the Zinc Finger E-Box Binding Homeobox family. However a structurally similar member of the same family, ZEB1 is associated to Fuch's[28] and posterior[29] corneal dystrophy, while COMMD1 is involved in copper homoeostasis[30] and metabolism and in Wilson's disease[31]. Mutations in the *UBE3* gene are known causes of the Kaufman oculocerebrofacial syndrome[32], a severe malformation in the newborn with numerous ocular manifestations.

We observe association for genetic variants near the *GJA3* locus, as previously reported[33,34]; however, this association was ethnicity-specific and could not be replicated in Europeans or in the smaller cohort of nuclear cataract case–control replication panel. This gene encoding for a gap-junction connexin (Connexin-46, CXA46) can induce cataract in animal models[35] and some of its mutations cause congenital cataracts in humans[36,37]. Given the evidence for association and its biological properties, variants at the *GJA3* locus need to be better characterised in future studies.

Variants in proximity to the *CRYAA* and *CRYAB* gene, encoding for the two forms of α-crystallin, were associated with ARNC. The α-crystallins contribute to the clarity and refractive properties of the lens, may prevent protein damage and protect against oxidative stress[33,34,38]. The common variants that we identified appear to affect transcription and expression of these genes, as suggested by previous studies where both proteins were down-regulated in lenses with ARNC[13,39,40].

Most of the genes located nearest to our association signals have functional properties that suggest an involvement in eye morphogenesis in general and crystallin expression and regulation. This together with the signals from the genes linked to congenital cataract point to overlap in mechanisms between the congenital and late-onset forms. In that respect, the genetic architecture of ARNC likely does not differ from other common complex conditions where deleterious coding variants cause congenital forms while common variants regulating gene expression are associated with increased risk of developing age-related forms. Given that smoking is an established risk factor for ARNC, it is also interesting that two of the loci co-localised with signals from GWAS of smoking. What is intriguing and would merit further research is the suggested systemic involvement in the disease. Both the Open Targets colocalization analysis and LDscore results suggest genetic sharing with metabolic syndrome components[41], age at menarche and other hormonal factors[42] in the pathogenesis of cataract. Systemic risk factors are known to influence other age-related cataract forms, such as cortical and diabetic cataracts, and when well-phenotyped and well-powered GWAS for these phenotypes become available, it will be interesting to see if there is any genetic overlap between those and loci identified here.

This work has several strengths, such as the use of the largest sample to date for genetic analysis of ARNC and, more importantly in the discovery phase, of precisely and quantitatively phenotyped cohorts. It also provides evidence of genetic mechanisms shared between congenital and age-related cataract and shows the importance of common genetic variants in maintaining crystalline lens integrity in the aging eye.

This study also has some limitations. The GWAS used in this study employed different grading systems, and despite phenotypic standardisation before the analyses, some residual heterogeneity between the studies may not be fully excluded. This study also sought to maximise the discovery power at the expense of increasing heterogeneity. We believe that replication was constrained by the power in the replication sample: a combined panel of 2807 cases and 2492 controls would afford sufficient (≥0.7) power only to the most common and strongest genetic effects (Fig. S5), which in our case are only encountered in the *SOX2* locus.

However, our conservative approach at dealing with the ethnic heterogeneity may have uneven power across the regions where there are significant differences the LD structure between the two main ancestral groups (European and Asian), or whenever there are significant differences in the minor allele frequency at certain loci. These circumstances, however, would have not affected the specificity of our findings.

Notwithstanding imperfections arising from sample and phenotypic availability, this study has doubled the number of loci positively associated with cataract and improved the proportion of phenotypic variance explained by them. The remaining heritability gap will be reduced by future with more powered, well-phenotyped studies and cohorts to further confirm association of known loci with ARNC and improve our understanding of the genetic architecture of this age-related cataract type.

Methods

Meta-analyses of summary statistics from GWAS were performed in four cohorts of European ($N = 7352$) and four of Asian ($N = 6799$) ancestry. Genetic variants associated with ARNC at GWAS ($P < 5 \times 10^{-8}$) or suggestive levels of statistical significance ($P < 1 \times 10^{-6}$) were carried forward for replication in the four additional cohorts.

Subjects and phenotyping. The following population-based cohort studies were included in the meta-analyses: Age-related Eye Diseases Study (AREDS), Blue Mountains Eye Study (BMES)[43], Rotterdam Study I, Rotterdam Study phase III (RSI-III)[44] and TwinsUK[45] all of European ancestry, as well as Singapore Malay Eye Study (SiMES)[46], Singapore Indian Eye Study (SINDI)[47] and two separate subsets of the Singapore Chinese Eye Study (SCES)[47]. Detailed demographic information and phenotyping methods are shown in the Supplementary Notes and Supplementary Tables 1 and 2. All studies were conducted with the approval of their local Research Ethics Committees, and written informed consent was obtained from all participants, in accordance with the Declaration of Helsinki.

All participants underwent detailed eye examination, including lens photography after pupil dilation for quantitative assessment of cataract severity in the discovery phase. Nuclear cataract was graded using standard grading systems from lens photographs (Supplementary Tables 1 and 2 and Supplementary Note: Grading systems) and, when scores for both eyes were available, the higher of the two scores was used in the analyses. Individuals who had undergone cataract surgery in both eyes were excluded.

In the replication phase, a dichotomous nuclear cataract status (presence or absence) was used as phenotypic outcome for the association models. This categorical binary trait was used as only semi-quantitative grading was available from these study populations, either from slit-lamp grading by clinician or from lens photography. In the replication phase we used two population-based cohorts of Asian ancestry, the Beijing Eye Study (BES)[48] and India Eye Study-South India (INDEYE(S)[49] as well as two European cohorts, the population-based (Beaver Dam Eye Study or BDES[50]) and a clinic-based case–control study (South London Case Control Study or SLCCS). The definition of cataract cases is shown in Supplementary Tables 1 and 2; the criteria included AREDS grade 3 or more for BES[48], LOCS III grade 4 or higher for INDEYE(S)[49], Wisconsin grade 3 or higher for BDES[50] and LOCS III grade 3 or higher for SLCCS. Controls were individuals with no significant nuclear opacity at the time of recruitment and no prior history of cataract surgery.

Genotyping and imputation. Different platforms were used for the genotyping of each cohort (Supplementary Table 3). All GWAS datasets were imputed against the 1000 Genomes Phase 1, with either IMPUTE2 (ref. [51]) or Minimac[52].

Statistical analysis. In the discovery phase, we included only cohorts where ARNC phenotyping was conducted according to an objective, standardised grading system of nuclear cataract severity. The details of each cohort and ARNC phenotyping can be found in Supplementary Tables 1–3, Supplementary Fig. 1 and Supplementary Notes. The distribution of quantitative ARNC scores was normalised whenever needed, and subsequently standardised within each cohort (mean 0 and standard deviation 1). The distribution of the transformed phenotypes is shown in Supplementary Fig. 1. For the replication, we used four cohorts of nuclear cataract patients and cataract-free controls (Supplementary Tables 2 and 3), not included in the quantitative, discovery phase (due to unavailability of genome-wide genotyping or quantitative nuclear cataract information).

Each cohort was ancestrally homogeneous: ethnic outliers were identified through Principal Component Analysis clustering and excluded from subsequent analyses. Genome-wide association analyses were performed in each cohort separately by building additive linear regression models, with the standardised ARNC score as the dependent variables and the number of alleles at each genetic locus as the explanatory variables, adjusting for age, sex and, when appropriate, principal components. In TwinsUK, linear mixed models with a kinship matrix as a random effect term (GEMMA)[53] were used to account for non-independence of observations due to familial relationships.

Fixed-effect inverse-variance meta-analyses using METAL[54] were performed on the GWAS summary statistics provided by each study for all variants with MAF >1%, genotyping call rate >0.97 and imputation quality >0.3 (the '*RSQ*' parameter in MACH[55] or '*info*' for IMPUTE[51]) that were present in at least three of the European or at least three of the Asian cohorts. Additionally, variants showing high heterogeneity ($I^2 > 0.75$) were excluded.

Gene-based analyses were performed using GATES[56] and gene set enrichment analysis using PASCAL[57]. The proportion of genetic variance explained by associated SNPs was calculated using individual-level data using GCTA[58]. Shared heritability between ARNC and other traits, for which GWAS results were available through the LDscore Hub website, was calculated using linkage disequilibrium score regression[59], taking Europeans as a reference.

Genome-wide associated SNPs showing suggestive association ($P < 10^{-6}$) in the discovery phase were taken forward to the replication stage of this study. We performed logistic regressions within each replication cohort, followed by an inverse-variance meta-analysis. Finally, SNPs that were identified through discovery and were genotyped in replication cohorts were meta-analysed together through a sample size-weighted *P* value analysis using METAL[54].

Gene expression in publicly available databases. Gene expression data in human and mouse lens were obtained using publicly available databases: iSyte[60], Ocular tissue database and the Mouse Genome informatics (MGI) gene expression database. Expression patterns were examined not only for the gene closest to the most strongly associated variant in each associated region, but also for all other genes in in the same LD block with them.

eQTL analysis. Lens tissue eQTLs are not currently available, but as eQTL effects are often shared between tissues[61,62], we assessed whether SNPs associated with nuclear cataract ($P < 1 \times 10^{-5}$) regulate gene expression of adjacent genes (i.e. have eQTL effects) by searching publicly available data (GTEx)[63] and the available literature[64].

Regulatory elements. The most significantly associated variant at each locus was annotated for regulatory functions (enhancer histone modification signals, DNase I hypersensitivity, binding of transcription factors or effects on regulatory motifs), using HaploReg[65] and ENCODE data track in the UCSC genome browser.

Additional annotation and data integration. Additional annotation and data integration were performed using FUMA (https://fuma.ctglab.nl, SNP2GENE and GENE2FUNCTION) and Open Targets Genetics (https://genetics.opentargets.org/, sentinel-variant PheWAS and candidate gene co-localisation).

Congenital cataract genes. Given the significant associations of markers within or in the proximity of congenital cataract genes such as *GJA3* and *CRYAA*, we enquired whether other common variants within genomic regions hosting additional known congenital cataract loci[66,67] were associated with ARNC. We explored association for a list of genes linked to congenital cataract by an extensive literature search and by using following databases: Online Mendelian Inheritance in Man (OMIM), Cataract Map (Cat-Map) and Clinical Variants (ClinVar). Each database was queried for variants within a 100 kb window and within the same LD block as the strongest associated SNP.

Web resources. http://genome.ucsc.edu/
http://ldsc.broadinstitute.org/
https://genome.uiowa.edu/otdb/
http://Supplemental.informatics.jax.org/
http://Supplemental.gtexportal.org/home/
http://omim.org/
https://cat-map.wustl.edu/
https://Supplemental.ncbi.nlm.nih.gov/clinvar/
https://fuma.ctglab.nl
https://genetics.opentargets.org/

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The GWAS summary statistics are available in Supplementary Data 3. Individual-level data can be requested by contacting the participating studies.

Received: 22 January 2020; Accepted: 1 October 2020;
Published online: 11 December 2020

References

1. Bourne, R. R. et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Glob. Health* **1**, e339–349 (2013).
2. Pascolini, D. & Mariotti, S. P. Global estimates of visual impairment: 2010. *Br. J. Ophthalmol.* **96**, 614–618 (2012).
3. Foster, A. Cataract and "Vision 2020-the right to sight" initiative. *Br. J. Ophthalmol.* **85**, 635–637 (2001).
4. Murthy, G., Gupta, S. K., John, N. & Vashist, P. Current status of cataract blindness and Vision 2020: the right to sight initiative in India. *Indian J. Ophthalmol.* **56**, 489–494 (2008).
5. Kessel, L. Can we meet the future demands for cataract surgery? *Acta Ophthalmol.* **89**, e289–290 (2011).
6. Klein, B. E., Klein, R., Lee, K. E. & Gangnon, R. E. Incidence of age-related cataract over a 15-year interval the Beaver Dam Eye Study. *Ophthalmology* **115**, 477–482 (2008).
7. Hammond, C. J., Snieder, H., Spector, T. D. & Gilbert, C. E. Genetic and environmental factors in age-related nuclear cataracts in monozygotic and dizygotic twins. *N. Engl. J. Med.* **342**, 1786–1790 (2000).
8. Chang, J. R. et al. Risk factors associated with incident cataracts and cataract surgery in the Age-related Eye Disease Study (AREDS): AREDS report number 32. *Ophthalmology* **118**, 2113–2119 (2011).
9. Hiller, R. et al. A longitudinal study of body mass index and lens opacities. The Framingham Studies. *Ophthalmology* **105**, 1244–1250 (1998).
10. Mares, J. A. et al. Healthy diets and the subsequent prevalence of nuclear cataract in women. *Arch. Ophthalmol.* **128**, 738–749 (2010).
11. Ritchie, M. D. et al. Electronic medical records and genomics (eMERGE) network exploration in cataract: several new potential susceptibility loci. *Mol. Vis.* **20**, 1281–1295 (2014).
12. See, C. W., Iftikhar, M. & Woreta, F. A. Preoperative evaluation for cataract surgery. *Curr. Opin. Ophthalmol.* **30**, 3–8 (2019).
13. Liao, J. et al. Meta-analysis of genome-wide association studies in multiethnic Asians identifies two loci for age-related nuclear cataract. *Hum. Mol. Genet* **23**, 6119–6128 (2014).
14. England, S. K., Uebele, V. N., Kodali, J., Bennett, P. B. & Tamkun, M. M. A novel K+ channel beta-subunit (hKv beta 1.3) is produced via alternative mRNA splicing. *J. Biol. Chem.* **270**, 28531–28534 (1995).
15. Litt, M. et al. Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Hum. Mol. Genet.* **7**, 471–474 (1998).
16. Lunetta, K. L. et al. Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study. *BMC Med. Genet.* **8**(Suppl 1), S13 (2007).
17. Loomis, S. J. et al. Exome array analysis of nuclear lens opacity. *Ophthalmic Epidemiol.* **25**, 215–219 (2018).
18. Boutin, T. S. et al. Insights into the genetic basis of retinal detachment. *Hum. Mol. Genet.* **29**, 689–702 (2020).
19. Chang, C. et al. A genome-wide association study provides new evidence that CACNA1C gene is associated with diabetic cataract. *Invest Ophthalmol. Vis. Sci.* **57**, 2246–2250 (2016).
20. Lin, H. J. et al. Novel susceptibility genes associated with diabetic cataract in a Taiwanese population. *Ophthalmic Genet.* **34**, 35–42 (2013).
21. Ragge, N. K. et al. SOX2 anophthalmia syndrome. *Am. J. Med. Genet. A* **135**, 1–7 (2005). discussion 8.
22. Wang, P., Liang, X., Yi, J. & Zhang, Q. Novel SOX2 mutation associated with ocular coloboma in a Chinese family. *Arch. Ophthalmol.* **126**, 709–713 (2008).
23. Cvekl, A., McGreal, R. & Liu, W. Lens development and crystallin gene expression. *Prog. Mol. Biol. Transl. Sci.* **134**, 129–167 (2015).
24. Shimada, N., Aya-Murata, T., Reza, H. M. & Yasuda, K. Cooperative action between L-Maf and Sox2 on δ-crystallin gene expression during chick lens development. *Mechanisms Dev.* **120**, 455–465 (2003).
25. Donner, A. L., Episkopou, V. & Maas, R. L. Sox2 and Pou2f1 interact to control lens and olfactory placode development. *Dev. Biol.* **303**, 784–799 (2007).
26. Fantes, J. et al. Mutations in SOX2 cause anophthalmia. *Nat. Genet.* **33**, 461–463 (2003).
27. Kondoh, H., Uchikawa, M. & Kamachi, Y. Interplay of Pax6 and SOX2 in lens development as a paradigm of genetic switch mechanisms for cell differentiation. *Int J. Dev. Biol.* **48**, 819–827 (2004).
28. Riazuddin, S. A. et al. Missense mutations in TCF8 cause late-onset Fuchs corneal dystrophy and interact with FCD4 on chromosome 9p. *Am. J. Hum. Genet.* **86**, 45–53 (2010).
29. Moroi, S. E. et al. Clinicopathologic correlation and genetic analysis in a case of posterior polymorphous corneal dystrophy. *Am. J. Ophthalmol.* **135**, 461–470 (2003).
30. Kodama, H. & Fujisawa, C. Copper metabolism and inherited copper transport disorders: molecular mechanisms, screening, and treatment. *Metallomics* **1**, 42–52 (2009).
31. Yu, C. H., Lee, W., Nokhrin, S. & Dmitriev, O. Y. The dtructure of metal binding domain 1 of the copper transporter ATP7B reveals mechanism of a singular Wilson disease mutation. *Sci. Rep.* **8**, 581 (2018).
32. Basel-Vanagaite, L. et al. Deficiency for the ubiquitin ligase UBE3B in a blepharophimosis-ptosis-intellectual-disability syndrome. *Am. J. Hum. Genet.* **91**, 998–1010 (2012).
33. Boelens, W. C. Cell biological roles of alphaB-crystallin. *Prog. Biophys. Mol. Biol.* **115**, 3–10 (2014).
34. Christopher, K. L. et al. Alpha-crystallin-mediated protection of lens cells against heat and oxidative stress-induced cell death. *Biochim Biophys. Acta* **1843**, 309–315 (2014).
35. Liu, K. et al. Altered ubiquitin causes perturbed calcium homeostasis, hyperactivation of calpain, dysregulated differentiation, and cataract. *Proc. Natl Acad. Sci. USA* **112**, 1071–1076 (2015).
36. Li, Y., Wang, J., Dong, B. & Man, H. A novel connexin46 (GJA3) mutation in autosomal dominant congenital nuclear pulverulent cataract. *Mol. Vis.* **10**, 668–671 (2004).
37. Zhang, X., Wang, L., Wang, J., Dong, B. & Li, Y. Coralliform cataract caused by a novel connexin46 (GJA3) mutation in a Chinese family. *Mol. Vis.* **18**, 203–210 (2012).
38. Andley, U. P. Effects of alpha-crystallin on lens cell function and cataract pathology. *Curr. Mol. Med.* **9**, 887–892 (2009).
39. Zhou, H. Y. et al. Quantitative proteomics analysis by iTRAQ in human nuclear cataracts of different ages and normal lens nuclei. *Proteomics Clin. Appl.* **9**, 776–786 (2015).
40. Zhou, P., Luo, Y., Liu, X., Fan, L. & Lu, Y. Down-regulation and CpG island hypermethylation of CRYAA in age-related nuclear cataract. *FASEB J.* **26**, 4897–4902 (2012).
41. Sabanayagam, C. et al. Metabolic syndrome components and age-related cataract: the Singapore Malay eye study. *Invest. Ophthalmol. Vis. Sci.* **52**, 2397–2404 (2011).
42. Younan, C. et al. Hormone replacement therapy, reproductive factors, and the incidence of cataract and cataract surgery: the Blue Mountains Eye Study. *Am. J. Epidemiol.* **155**, 997–1006 (2002).
43. Mitchell, P., Cumming, R. G., Attebo, K. & Panchapakesan, J. Prevalence of cataract in Australia: the Blue Mountains eye study. *Ophthalmology* **104**, 581–588 (1997).
44. Hofman, A. et al. The Rotterdam Study: objectives and design update. *Eur. J. Epidemiol.* **22**, 819–829 (2007).
45. Spector, T. D. & Williams, F. M. The UK Adult Twin Registry (TwinsUK). *Twin Res. Hum. Genet.* **9**, 899–906 (2006).
46. Foong, A. W. et al. Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol.* **14**, 25–35 (2007).
47. Lavanya, R. et al. Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. *Ophthalmic Epidemiol.* **16**, 325–336 (2009).
48. Jonas, J. B., Xu, L. & Wang, Y. X. The Beijing Eye Study. *Acta Ophthalmol.* **87**, 247–261 (2009).
49. Vashist, P. et al. Prevalence of cataract in an older population in India: the India study of age-related eye disease. *Ophthalmology* **118**, 272–278 e271-272 (2011).
50. Klein, B. E., Klein, R. & Linton, K. L. Prevalence of age-related lens opacities in a population. The Beaver Dam Eye Study. *Ophthalmology* **99**, 546–552 (1992).
51. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
52. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
53. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
54. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
55. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
56. Li, M. X., Gui, H. S., Kwan, J. S. & Sham, P. C. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am. J. Hum. Genet.* **88**, 283–293 (2011).
57. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
58. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
59. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

60. Lachke, S. A. et al. iSyTE: integrated systems tool for eye gene discovery. *Invest Ophthalmol. Vis. Sci.* **53**, 1617–1627 (2012).
61. Aguet, F. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
62. Mele, M. et al. Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
63. Carithers, L. J. & Moore, H. M. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv. Biobank* **13**, 307–308 (2015).
64. Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
65. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–934 (2012).
66. Deng, H. & Yuan, L. Molecular genetics of congenital nuclear cataract. *Eur. J. Med. Genet.* **57**, 113–122 (2014).
67. Santana, A. & Waiswo, M. The genetic and molecular basis of congenital cataract. *Arq. Bras. Oftalmol.* **74**, 136–142 (2011).

Acknowledgements

The authors thank the staff and participants of all studies for their important contributions. Complete funding information and acknowledgements are provided in the Supplementary Information.

Author contributions

C.-Y.C., C.J.H., J.J.W., S.K.I., A.E.F., B.E.K.K. and E.Y.-D. conceived the project. E.Y.-D., W.Z., R.P.I.Jr, C.W., K.E.L., G.R.J., X.C., H.L., A.E.F., Y.S., Q.F., J.L., X. Su, K.E.L., Y.S., J. Chung, W.L., P.G.H. and A.C.A. performed analyses. Y.X.W., C.-Y.C., Y.-Y.T., T.A., K.S.S., P.M., J.B.J., T.Y.W., C.C.K., B.E.K.K., C.C.W.K., S.-P.C., Q.S.T., P.G., X. Sim, P.S., A.F., A.G.T., J. Chua, M.L.C., E.Y.C., M.C.L., A.S.Y.C., E.N.V., Z.L., J.M.C., K.P.B., L.G.F., M.T., P.W.M.B., M.K.S., M.H.H., R.D.R. and Y.-C.T. were responsible for collecting clinic data and performing genotyping in each study. E.Y.-D., C.J.H. and C.-Y.C. drafted the paper. P.G.H., J.J.W., B.E.K.K., S.K.I. and T.Y.W. critically reviewed the manuscript.

Competing interests

C.C.K. is an Editorial Board Member for *Communications Biology*, but was not involved in the editorial review of, nor the decision to publish, this article. All authors declare no additional competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s42003-020-01421-2.

Correspondence and requests for materials should be addressed to C.J.H. or C.-Y.C.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ekaterina Yonova-Doing[1,34], Wanting Zhao[2,3,34], Robert P. Igo Jr[4,34], Chaolong Wang[5,6,34], Periasamy Sundaresan[7], Kristine E. Lee[8], Gyungah R. Jun[9], Alexessander Couto Alves[1], Xiaoran Chai[2], Anita S. Y. Chan[2,10,11], Mei Chin Lee[2,10], Allan Fong[2], Ava G. Tan[12], Chiea Chuen Khor[2,13], Emily Y. Chew[14], Pirro G. Hysi[1,15], Qiao Fan[2,3], Jacqueline Chua[2,10], Jaeyoon Chung[10], Jiemin Liao[2], Johanna M. Colijn[16,17], Kathryn P. Burdon[18,19], Lars G. Fritsche[20,21], Maria K. Swift[8], Maryam H. Hilmy[22], Miao Ling Chee[2], Milly Tedja[16,17], Pieter W. M. Bonnemaijer[16,17], Preeti Gupta[2], Queenie S. Tan[23], Zheng Li[13], Eranga N. Vithana[2,10], Ravilla D. Ravindran[24], Soon-Phaik Chee[2,10,25], Yuan Shi[2], Wenting Liu[13], Xinyi Su[2,23,25], Xueling Sim[26], Yang Shen[5], Ya Xing Wang[27], Hengtong Li[2], Yih-Chung Tham[2], Yik Ying Teo[26,28], Tin Aung[2,10,25], Kerrin S. Small[1], Paul Mitchell[12], Jost B. Jonas[27,29], Tien Yin Wong[2,10,25], Astrid E. Fletcher[29,30], Caroline C. W. Klaver[16,17,31,32], Barbara E. K. Klein[8], Jie Jin Wang[12,33], Sudha K. Iyengar[4], Christopher J. Hammond[1,15,35]✉ & Ching-Yu Cheng[2,10,25,35]✉

[1]Department of Twin Research and Genetic Epidemiology, The School of Life Course Sciences, King's College London, London SE1 7EH, UK. [2]Singapore Eye Research Institute, Singapore National Eye Center, 168751 Singapore, Singapore. [3]Center for Quantitative Medicine, Duke-NUS Medical School, 169857 Singapore, Singapore. [4]Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA. [5]Computational and Systems Biology, Genome Institute of Singapore, 138672 Singapore, Singapore. [6]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, 430030 Wuhan, China. [7]Department of Genetics, Aravind Medical Research Foundation, Madurai, Tamil Nadu 625020, India. [8]Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI 53726, USA. [9]Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA 02118, USA. [10]Histology, Department of Pathology, Singapore General Hospital, 169856 Singapore, Singapore. [11]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, 169857 Singapore, Singapore. [12]Centre for Vision Research, Westmead Institute for Medical Research, University of Sydney, Sydney, NSW 2145, Australia. [13]Division of Human Genetics, Genome Institute of Singapore, 138672 Singapore, Singapore. [14]National Eye Institute, National Institutes of Health, Bethesda, MD 20814, USA. [15]Department of Ophthalmology, King's College London, London SE5 9RS, UK. [16]Department of Epidemiology, Erasmus Medical Centre, 3015 GD Rotterdam, The Netherlands. [17]Department of Ophthalmology, Erasmus Medical Centre, 3015 GD Rotterdam, The Netherlands. [18]Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7000, Australia. [19]Department of Ophthalmology, Flinders University, 5042 Adelaide, SA, Australia. [20]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor,

MI 48109, USA. [21]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, Norwegian University of Science and Technology, 7491 Trondheim, Norway. [22]Department of Anatomical Pathology and Cytology, Singapore General Hospital, 169608 Singapore, Singapore. [23]Institute of Molecular and Cell Biology, 138673 Singapore, Singapore. [24]Aravind Eye Hospital, Madurai, Tamil Nadu 625020, India. [25]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, 117597 Singapore, Singapore. [26]Saw Swee Hock School of Public Health, National University of Singapore, 117549 Singapore, Singapore. [27]Beijing Institute of Ophthalmology, Beijing Ophthalmology and Visual Science Key Lab, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, WC29+WV Beijing, China. [28]Department of Statistics and Applied Probability, National University of Singapore, 119077 Singapore, Singapore. [29]Department of Ophthalmology, Medical Faculty Mannheim of the Ruprecht-Karls-University Heidelberg, Seegartenklinik Heidelberg, 69115 Heidelberg, Germany. [30]Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. [31]Department Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands. [32]Institute of Molecular and Clinical Ophthalmology, Basel, Basel, Switzerland. [33]Health Services and Systems Research, Duke-NUS Medical School, 169857 Singapore, Singapore. [34]These authors contributed equally: Ekaterina Yonova-Doing, Wanting Zhao, Robert P. Igo, Chaolong Wang. [35]These authors jointly supervised this work: Christopher J. Hammond, Ching-Yu Cheng. ✉email: chris.hammond@kcl.ac.uk; chingyu.cheng@duke-nus.edu.sg