

## **A multiethnic GWAS meta-analysis of 585,243 individuals identifies new risk loci associated with cataract and reveals sex-specific effects**

Hélène Choquet<sup>1,\*</sup>, Ronald B. Melles<sup>2</sup>, Deepti Anand<sup>3</sup>, Jie Yin<sup>1</sup>, Gabriel Cuellar-Partida<sup>4</sup>, Wei Wang<sup>4</sup>, 23andMe Research Team<sup>4</sup>, Thomas J. Hoffmann<sup>5,6</sup>, K Saldas Nair<sup>7</sup>, Pirro G. Hysi<sup>8,9,10</sup>, Salil A. Lachke<sup>3,11</sup>, and Eric Jorgenson<sup>1</sup>

<sup>1</sup>Kaiser Permanente Northern California (KPNC), Division of Research, Oakland, CA 94612, USA;

<sup>2</sup>KPNC, Department of Ophthalmology, Redwood City, CA 94063, USA

<sup>3</sup>Department of Biological Sciences, University of Delaware, Newark, DE 19716, USA;

<sup>4</sup>23andMe Inc, Sunnyvale, CA 94086, USA

<sup>5</sup>Institute for Human Genetics, UCSF, San Francisco, CA 94143, USA.

<sup>6</sup>Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA 94158, USA.

<sup>7</sup>Departments of Ophthalmology and Anatomy, School of Medicine, UCSF, San Francisco, CA 94143, USA;

<sup>8</sup>King's College London, Section of Ophthalmology, School of Life Course Sciences, London, UK;

<sup>9</sup>King's College London, Department of Twin Research and Genetic Epidemiology, London, UK;

<sup>10</sup>University College London, Great Ormond Street Hospital Institute of Child Health, London, UK;

<sup>11</sup>Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE 19716, USA;

\*Correspondence should be addressed to H.C. ([Helene.Choquet@kp.org](mailto:Helene.Choquet@kp.org))

**Cataract is the leading cause of blindness among the elderly worldwide and cataract surgery is one of the most common operations performed in the United States<sup>1-3</sup>. The etiology remains largely unclear, and to contribute to its elucidation we conducted a multiethnic genome-wide association meta-analysis of cataract, combining results from the GERA and UK Biobank cohorts, and tested for replication in the research cohort from 23andMe, Inc.. We report 54 genome-wide significant loci, 37 of which were previously unknown. Sex-stratified analyses identified two additional novel loci (*CASP7* and *GSTM2*) specific to women and sex differences in effect sizes and significance of association at five other loci. We show that genes within or near 80% of the cataract-associated loci are significantly expressed and/or enriched-expressed in the mouse lens across various spatiotemporal stages. Further, 32 candidate genes in the associated loci have altered gene expression in 9 different gene perturbation mouse models of lens defects/cataract, suggesting their relevance to lens biology.**

Cataracts are caused by opacification of the crystalline lens, which leads to progressive loss of vision. They can present as a developmental disorder in younger patients (congenital or pediatric cataracts) but, more commonly, occur as a disease of aging<sup>4,5</sup>, and are a leading cause of visual impairment. Cataract formation and cataract surgery are more common in women<sup>6</sup>. Twin and family aggregation studies strongly support an important role for genetic factors in cataract susceptibility with heritability estimates ranging from 35 to 58%<sup>7-12</sup>. A recent study<sup>13</sup> investigated the genetic basis of eye disease, reported 20 genetic loci associated with cataract at a genome-wide level of significance in the UK Biobank European sample, although none of these loci was independently replicated. It is also unclear what proportion of clinical variability these loci help explain, as well as to what contribution they have in populations of diverse ethnic background.

Here, we present the largest and most ethnically diverse genetic study of cataract susceptibility conducted to date to our knowledge. Following a stepwise analytical approach, we conducted a genome-wide association analyses, followed by meta-analysis, including 585,243 individuals (67,844 cases and 517,399 cataract-free controls) from two cohorts: the Genetic Epidemiology Research in Adult Health and Aging (GERA)<sup>14</sup> and the UK Biobank (UKB)<sup>15,16</sup>. We tested the top independently associated SNPs ( $P < 5.0 \times 10^{-8}$ ) at each locus in 3,234,455 participants (347,209 self-reported cataract cases and 2,887,246 controls) from the 23andMe research cohort. Cohorts summary details are presented in **Supplementary Table 1**. We subsequently fine-mapped these associations<sup>17</sup> and examined changes in the expression of candidate genes in associated loci in 9 gene perturbation mouse models of lens defects<sup>18,19</sup>. We then undertook conditional, ethnic-, and sex-specific association analyses (**Supplementary Fig. 1**). Finally, we assessed the genetic correlation between cataract and other disorders and complex traits<sup>20</sup>.

We first undertook GWAS analysis of cataract in the GERA and UKB cohorts, stratified by ethnic group, followed by a meta-analysis across all analytical strata. In the multiethnic meta-analysis, we identified 54 loci ( $P < 5.0 \times 10^{-8}$ ), of which 37 were novel (i.e., not previously reported to be associated with cataract at a genome-wide level of significance) (**Table 1, Fig. 1, and Supplementary Fig. 2**). The effect estimates of 54 lead SNPs were consistent across the 2 studies (**Table 1 and Supplementary Fig. 3**). In 23andMe research cohort, 45 out of 51 lead SNPs available (88.2%) replicated with a consistent direction of effect at a Bonferroni corrected significance threshold of  $9.8 \times 10^{-4}$  ( $P\text{-value} = 0.05/51$ ) and additional 2 SNPs were nominally significant ( $P < 0.05$ ) (**Table 1 and Supplementary Fig. 4**).

To determine whether there were additional signals in individual ancestry groups that did not reach genome-wide significance in the meta-analysis, we conducted ethnic-specific meta-analyses of each ancestry group. We identified three additional novel loci in the European

ancestry (GERA non-Hispanic whites + UKB Europeans) meta-analysis: *EPHA4*, *CD83-JARID2*, and near *EXOC3L2* (**Supplementary Fig. 5.a. and Supplementary Table 2**). Regional association plots of the association signals are presented in **Supplementary Fig. 6**. To identify independent signals within the 44 genomic regions identified in the European-specific meta-analysis (**Supplementary Table 3**), we performed a multi-SNP-based conditional & joint association analysis (COJO)<sup>21</sup>, which revealed 5 additional independent SNPs within 4 of the identified genomic regions, including at known loci (*CDKN2B*, *RIC8A*, and *LOC338694*) and at newly identified *DNMBP* locus (**Supplementary Table 4**). Neither the meta-analysis of East Asian groups nor the meta-analysis combining the GERA African American and UKB African British groups result in the identification of additional novel genome-wide significant findings (**Supplementary Fig. 5.b. and 5.c.**).

Next, we conducted genetic association analyses for interaction between genetic factors and sex, in sex-specific meta-analyses combining data from GERA and UKB. We identified two additional novel loci, *CASP7* and *GSTM2*, in the women-specific meta-analysis (GERA+UKB) (**Fig. 2 and Supplementary Table 5**). *CASP7* rs12777332 and *GSTM2* rs3819350 were significantly associated with cataract in women (*CASP7* rs12777332: OR=1.06,  $P=3.41 \times 10^{-8}$ ; *GSTM2* rs3819350: OR=1.06,  $P=2.10 \times 10^{-8}$ ) but not in men (*CASP7* rs12777332: OR=1.01,  $P=0.25$ ; *GSTM2* rs3819350: OR=1.01,  $P=0.25$ ) (**Supplementary Fig. 7**). Further, among the loci identified in the multiethnic meta-analysis (GERA+UKB), we observed significant differences in the effect sizes and significance of association at five loci: one locus, *DNMBP-CPN1*, was strongly associated with cataract in women but not in men (*DNMBP-CPN1* rs1986500, OR=0.94,  $P=5.04 \times 10^{-11}$  in women, and OR=1.01,  $P=0.40$  in men;  $Z=-5.03$ ,  $P=2.44 \times 10^{-7}$ ) (**Supplementary Fig. 8 and Supplementary Table 5**); and four loci, *QKI*, *SEMA4D*, *RBFOX1*, and *JAG1*, were strongly associated in men than women (*QKI* 6:163840336, OR=0.94,  $P=1.23 \times 10^{-10}$  in men, and OR=0.99,  $P=0.21$  in women;  $Z=-3.95$ ,

$P=3.89 \times 10^{-5}$ ; *SEMA4D* rs62547232, OR=1.15,  $P=1.83 \times 10^{-9}$  in men, and OR=0.98,  $P=0.33$  in women;  $Z=5.03$ ,  $P=2.43 \times 10^{-7}$ ; *RBFOX1* rs7184522, OR=1.07,  $P=9.10 \times 10^{-12}$  in men, and OR=1.03,  $P=0.0020$  in women;  $Z=2.98$ ,  $P=1.43 \times 10^{-3}$ ; *JAG1* rs3790163, OR=0.92,  $P=3.14 \times 10^{-12}$  in men, and OR=0.96,  $P=9.63 \times 10^{-4}$  in women;  $Z=-2.95$ ,  $P=1.59 \times 10^{-3}$ ) (**Fig. 2 and Supplementary Table 6 and Supplementary Fig. 8**). Regional association plots illustrate the sex-specific association signals (**Supplementary Fig. 7**).

We adopted a Bayesian approach (CAVIARBF)<sup>17</sup> to compute variants likelihood to explain the observed association at each locus and derived the smallest set of variants that has a 95% probability to include the causal origin of the signals (95% credible set). Nine sets included a single variant (**Supplementary Table 7**) such as rs62621812 (*ZNF800*), rs1014607 (*BAMBI-LOC100507605*), rs1428885924 (*NEK4*), rs1679013 (*CDKN2B-DMRTA1*), rs1539508 (*LOC100132354*), rs73238577 (*RFC1-KLB*), rs17172647 (*IGFBP3-TNS3*), rs73530148 (*ALDOA*), and rs549768142 (*JAG1*), suggesting that those single variants may be the causal origin of the associations observed in their respective loci.

A gene-based analysis, using the VEGAS2 integrative tool<sup>22</sup> on 22,673 genes, found significant associations with cataract for 8 genes within 4 loci identified in the multiethnic combined (GERA+UKB) meta-analysis, including *EFNA1* and *KRTCAP2* (chr1q22), *CDKN2B* and *CDKN2B-AS1* (chr9p21.3), *MRPL21* and *LOC338694* (chr11q13.3), *HERC2* (chr15q13.2), and *BLVRA* gene (chr7p13) (**Supplementary Table 8**).

We next examined the expression of genes within identified loci potentially associated with cataract in lens tissue using the web-resource tool iSyTE (integrated Systems Tool for Eye gene discovery)<sup>18,19</sup>. iSyTE contains genome-wide expression data, based on microarray or RNA-seq analysis, on the mouse lens at different embryonic and postnatal stages<sup>18,23</sup>. In addition to expression, iSyTE also contains information of “lens-enriched expression” which has proved to be an excellent predictor of cataract-linked genes in humans and animal models<sup>24-27</sup>. The iSyTE-

based lens microarray data on Affymetrix and/or Illumina platforms showed that orthologs of 47 candidates were significantly expressed in the mouse lens ( $>100$  expression units,  $P<0.05$ ) in one or more embryonic/postnatal stages (**Fig. 3**). Over 60% of the expressed genes were found to have high lens-enriched expression ( $>1.5$  fold-change over whole embryonic body (WB) reference dataset,  $P<0.05$ ), suggesting their likely relevance to lens development, homeostasis and pathology (**Supplementary Fig. 9**). This was further supported by iSyTE RNA-seq data that also showed lens-expression of 46 candidates ( $\geq 2.0$  CPM, counts per million,  $P<0.05$ ), 31 out of which (~68%) exhibited high lens-enriched expression in one or more embryonic/postnatal stages ( $>1.5$  fold-change over WB,  $P<0.05$ ) (**Fig. 3 and Supplementary Fig. 9**). Together, this analysis offered strong support for lens expression for total 52 different genes, with at least one candidate gene for 43 of the 54 loci, thus accounting for nearly 80% of the identified loci. Additionally, iSyTE also informs on lens gene expression changes in specific gene perturbation mouse models of lens defects/cataract. This analysis showed that 38 candidate genes had significant differences in gene expression ( $P<0.05$ ) in one or more of the 9 different gene perturbation mouse models of lens defects/cataract (**Supplementary Fig. 10 and Supplementary Table 9**). Together, iSyTE analysis offers independent experimental evidence that support the direct relevance of these candidate genes to lens biology and cataract.

We also conducted a pathway analysis using VEGAS software<sup>22</sup> to assess enrichment in 9,732 pathways or gene-sets derived from the Biosystem's database. We found that the notochord development was the only gene-set significantly enriched in our results, after correcting for multiple testing ( $P<5.14\times 10^{-6}$ ) (**Supplementary Table 10**). In addition, we identified 781 pathways/gene-sets that were nominally enriched ( $P<0.05$ ), with the most significant of which were 'circadian clock' ( $P=2.07\times 10^{-5}$ ), followed by 'lens morphogenesis in camera-type eye' ( $P=2.14\times 10^{-5}$ ), and 'notochord morphogenesis' ( $P=2.81\times 10^{-5}$ ). Our findings

are consistent with early work, demonstrating that mice deficient in circadian clock proteins, such as *BMAL1* and *CLOCK*, display age-related cataract<sup>28,29</sup>.

To estimate the pairwise genetic correlations ( $r_g$ ) between cataract and more than 700 diseases/traits from different publicly available resources/consortia, we compared our GWAS results with summary statistics for other traits by performing an LD score regression using the LD Hub web interface<sup>20</sup>. Genetic correlations were considered significant after Bonferroni adjustment for multiple testing ( $P < 6.48 \times 10^{-5}$  which corresponds to 0.05/772 phenotypes tested). We found significant genetic correlations between cataract and 39 traits, including three of them directly related to eye traits: 'wears glasses or contact lenses' ( $r_g = 0.30$ ,  $P = 2.56 \times 10^{-7}$ ), 'self-reported: glaucoma' ( $r_g = 0.30$ ,  $P = 4.57 \times 10^{-6}$ ), and 'reason for glasses/contact lenses: myopia' ( $r_g = 0.25$ ,  $P = 1.10 \times 10^{-5}$ ) (**Supplementary Table 11**).

A phenome-wide association study (PheWAS) analysis of 43 cataract-associated SNPs, available in the GeneAtlas was run across 776 traits measured and previously analyzed in the UKB<sup>30</sup>. Twenty-three of the most significantly associated cataract-associated variants were significantly associated ( $P < 5.0 \times 10^{-8}$ ) with other traits (**Fig. 4**). Most were associated with disorders of the lens, with the strongest association observed for the intronic variant rs4814857 at *SLC24A3* ( $P = 2.48 \times 10^{-39}$ ) (**Supplementary Table 12**). *SLC24A3* encodes the carrier family 24 member 3 and has been thought to be involved in retinal diseases<sup>31</sup>. Variants at *PLCE1* and *HMG2* were significantly associated with anthropometric traits, such as impedance of whole body, impedance of arm and leg. The *PLCE1* gene encodes the phospholipase C epsilon 1 and common variation at this locus has been shown to be associated with retinal detachment<sup>13</sup> (**Supplementary Table 13**), and *HMG2* loss of function variants were linked with bilateral cataracts in a fetus presenting with growth delay<sup>32</sup>. Our PheWAS analysis also highlighted that variants at *OCA2* and *NPLOC4* were significantly associated with pigmentation phenotypes. The *OCA2* gene encodes the melanosomal transmembrane protein, whose variants determine

iris color and have been linked to corneal and refractive astigmatism, syndromic forms of myopia, refractive error, and type 2 oculocutaneous albinism<sup>33-37</sup> (**Supplementary Table 13**). *NPLOC4* encodes the homolog, ubiquitin recognition factor and has been previously associated with macular thickness and the risk of strabismus and corneal and refractive astigmatism<sup>34,38,39</sup>.

Our study should be interpreted within the context of its limitations. First, the cataract phenotypes were assessed differently across the 3 study cohorts. While our cataract phenotype in GERA was based on electronic health records (EHRs) data and International Classification of Disease, Ninth (ICD9) or Tenth Revision (ICD10) diagnosis codes, most of the cataract cases in UKB, and all of the cataract cases in 23andMe research cohort (our replication sample) were based on self-reported data. This may result in phenotype misclassification, however, our meta-analysis combining GERA and UKB showed consistency of the SNPs effect estimates between cohorts, and the identified associations were well validated in the 23andMe research cohort. Second, subtypes of cataract were not available in the 3 study cohorts, which may result in underestimates of the effects of individual SNPs due to phenotype misclassification. Future studies will determine whether the identified loci contribute to different cataract subtypes (i.e. nuclear, cortical, or subcapsular) and the extent to which these loci display shared effects across subtypes.

In conclusion, we report the results of a large GWAS that identified 47 novel loci (37 from the multiethnic-meta-analysis + 3 European-specific meta-analysis + 5 conditional analysis + 2 from the female-specific meta-analysis) for the development of cataract and that likely contribute to the pathophysiology of this common vision disorder. Several genes within these cataract-associated loci, including *RARB*, *KLF10*, *DNMBP*, *HMGA2*, *MVK*, *BMP4*, *CPAMD8*, and *JAG1*, represent potential candidates for the development of drug targets as previous work supports the relevance of these candidates to cataracts<sup>32,40-46</sup> (**Supplementary Table 13**). We also report three loci that show women-specific effects on cataract susceptibility and 4 others that showed



significant differences in effects between women and men. These loci provide a biological foundation for understanding the etiology of sex-differences in cataract susceptibility, and, may suggest potential targets for the development of non-surgical treatment of cataracts.

## **Acknowledgements**

We are grateful to the Kaiser Permanente Northern California members who have generously agreed to participate in the Kaiser Permanente Research Program on Genes, Environment, and Health. Support for participant enrollment, survey completion, and biospecimen collection for the RPGEH was provided by the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, and Kaiser Permanente Community Benefit Programs. Genotyping of the GERA cohort was funded by a grant from the National Institute on Aging, National Institute of Mental Health, and National Institute of Health Common Fund (RC2 AG036607). H.C. and E.J. were supported by the National Eye Institute (NEI) grant R01 EY027004, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) R01 DK116738 and by the National Cancer Institute (NCI) R01CA2416323. This work was also made possible in part by NIH-NEI EY002162—Core Grant for Vision Research, by the Research to Prevent Blindness Unrestricted Grant (UCSF, Ophthalmology). K.S.N. receives support from NEI grant EY022891, BrightFocus Foundation (G2019360), Marin Community Foundation-Kathlyn McPherson Masneri and Arno P. Masneri Fund, and That Man May See Inc. T.J.H. was supported by National Institutes of Aging (NIA) grant R21 AG046616. S.A.L. was supported by National Institutes of Health / National Eye Institute grants R01 EY021505 and EY029770 and D.A. was supported by a Knights Templar Pediatric Ophthalmology Career Starter Grant Award. We would like to thank the research participants and employees of 23andMe for making this work possible. The following members of the 23andMe Research Team contributed to this study: Michelle Agee, Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Katarzyna Bryc, Emily Bullis, Briana Cameron,

Daniella Coker, Gabriel Cuellar Partida, Devika Dhamija, Sayantan Das, Sarah L. Elson, Teresa Filshtein, Kipper Fletez-Brant, Pierre Fontanillas, Will Freyman, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Katelyn Kukar, Keng-Han Lin, Maya Lowe, Marie K. Luff, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Steven J. Micheletti, Meghan E. Moreno, Joanna L. Mountain, Sahar V. Mozaffari, Priyanka Nandakumar, Elizabeth S. Noblin, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Chao Tian, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton.

### **Author contributions**

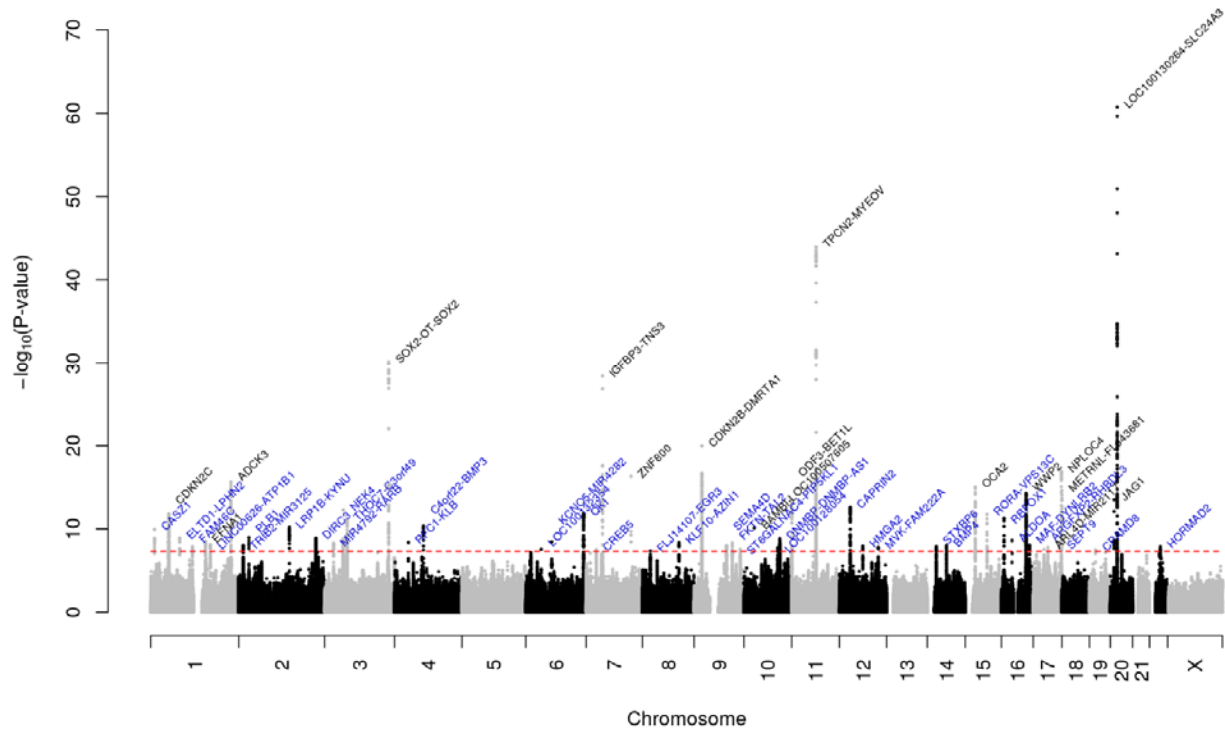
H.C., S.A.L., and E.J. contributed to study conception and design. T.J.H., and E.J. were involved in the genotyping and quality control of the GERA samples. T.J.H. performed the imputation analyses in the GERA cohort. R.B.M. extracted phenotype data for the GERA subjects based on EHRs. J.Y. performed statistical analyses and in silico analyses. D.A., and S.A.L. carried out the gene expression analyses in lens tissue using iSyTE. W.W. and G.C.P. oversaw the replication analyses in the 23andMe research cohort. H.C., S.A.L., and E.J. interpreted the results of analyses and wrote the manuscript with help from R.B.M., K.S.N., P.G.H.

### **COMPETING INTERESTS**

Gabriel Cuellar Partida and Wei Wang are employed by and hold stock or stock options in 23andMe, Inc. The other authors declare no competing financial or non-financial interests.

## Figure Legends

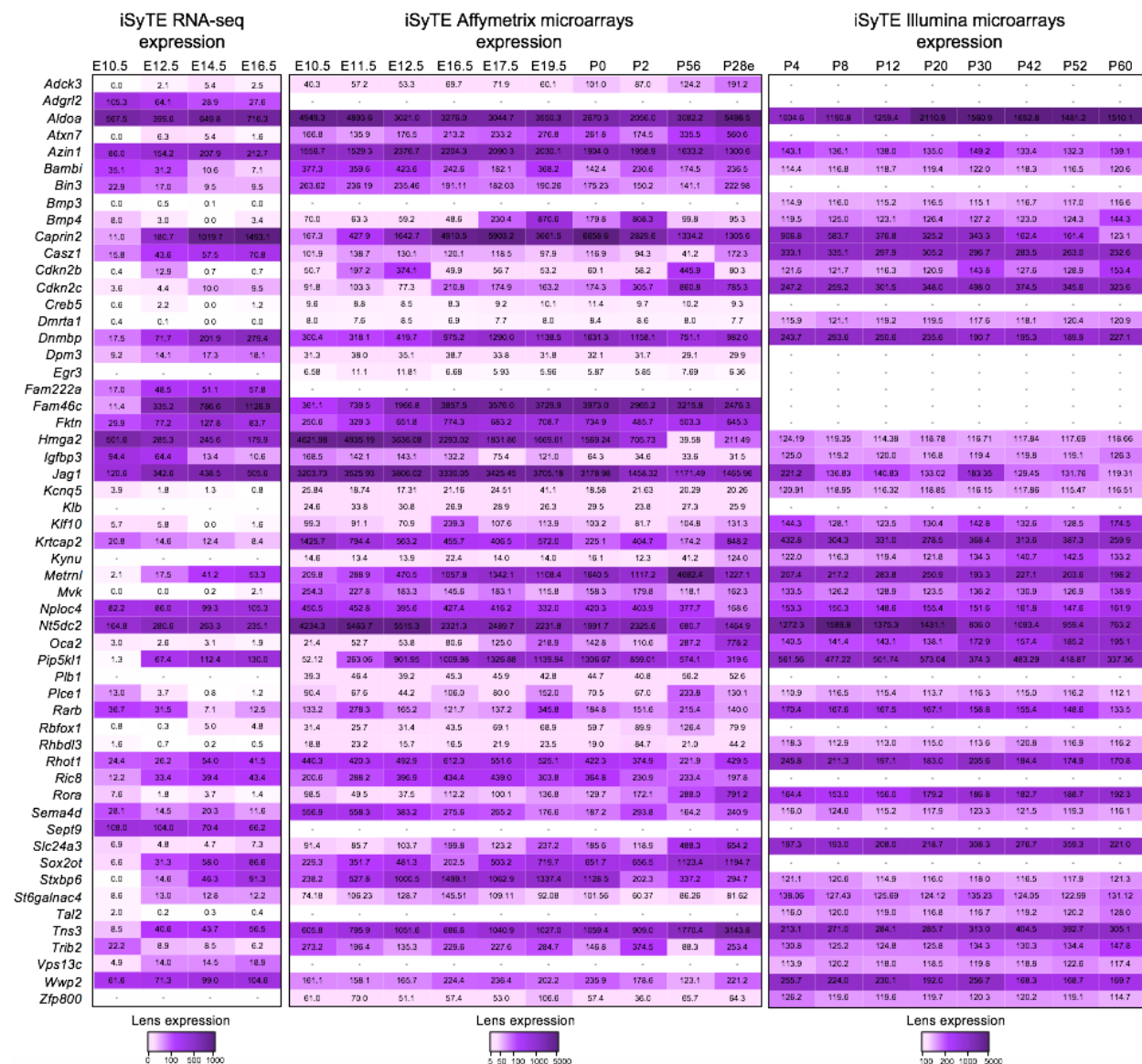
**Figure 1.** Manhattan plot of the multiethnic combined (GERA+UKB) GWAS meta-analysis of cataract



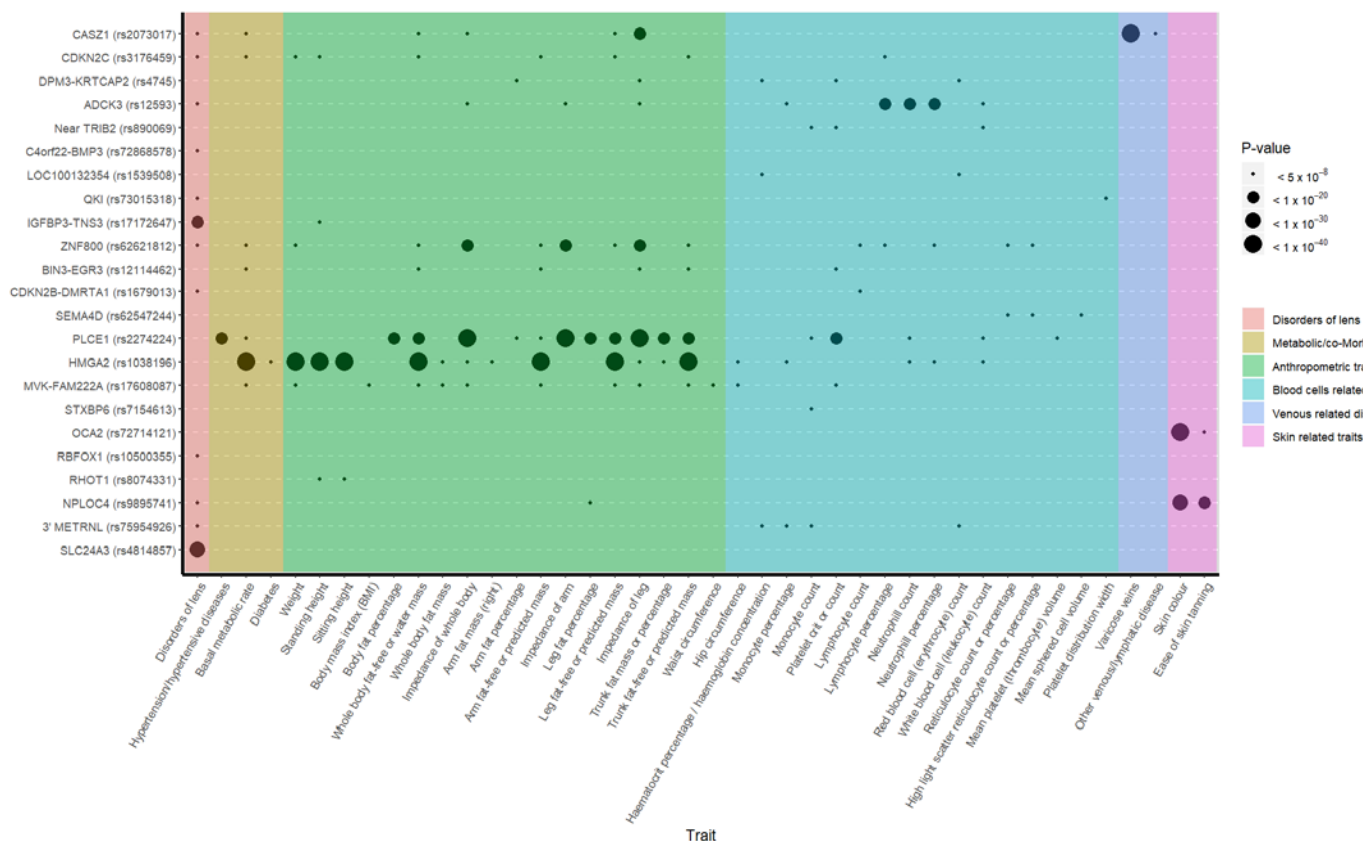
Locus names in blue are for the novel loci and the ones in dark are for the previously reported ones.



**Figure 3. Expression of candidate genes in mouse lens.** Mouse orthologs of the human candidate genes in the 54 loci were examined for their lens expression in the iSyTE database. Analysis of whole lens tissue data on various platforms, microarrays (Affymetrix, Illumina) and RNA-seq indicates expression of 55 genes at different stages indicated by embryonic (E) and postnatal (P) days and ranged from early lens development (*i.e.*, E10.5) through adulthood (*i.e.*, P60). Note: P28 in Affymetrix represents expression data on isolated lens epithelium. The range of expression on each platform is indicated by a specific heat-map. The numbers within individual tiles indicate the level of expression in fluorescence intensity (for microarrays) and in counts per million (CPM) (for RNA-seq).



**Figure 4.** Phenome-wide association matrix of cataract top variants.



PheWAS was carried out for the 54 lead SNPs in our loci of interest identified in the combined (GERA+UKB) multiethnic analysis. SNPs were queried against 776 traits ascertained for UKB participants and reported in the Roslin Gene Atlas<sup>30</sup>, including disorders of the lens, anthropometric traits, hematologic laboratory values, ICD-10 clinical diagnoses and self-reported conditions. Among the 54 lead SNPs, 43 were available in Gene Atlas database. We reported SNPs showing genome-wide significant association with at least one trait (in addition to cataract).

## Tables

**Table 1.** Cataract loci identified in the combined (GERA+UKB) GWAS multiethnic meta-analysis and replication in 23andMe research cohort

SNP	Chr	Pos	Locus	Alleles A1/A2	Combined Meta-Analysis		Replication in 23andMe		Direction of Effect (GERA-UKB- 23andMe)
					OR (SE)	P	OR (SE)	P	
rs2073017	1	10822808	<i>CASZ1</i>	C/T	0.95 (0.0081)	1.21E-10	0.98 (0.0042)	1.40E-06	---
rs3176459	1	51437247	<i>CDKN2C</i>	G/A	0.95 (0.007)	1.63E-12	0.98 (0.0036)	7.71E-10	---
rs71646944	1	82151352	<i>ADGRL2</i>	T/C	1.05 (0.0087)	1.26E-09	1.04 (0.0045)	5.60E-18	+++
rs10633030	1	118160005	<i>FAM46C</i>	CT/C	1.05 (0.0081)	1.48E-08	1.03 (0.0041)	2.56E-11	+++
rs4745	1	155106227	<i>DPM3-KRTCAP2</i>	T/A	0.96 (0.0066)	3.26E-09	1.01 (0.0207)	0.66	---+
rs2982459	1	169028716	<i>LINC00970</i>	G/A	0.96 (0.0071)	7.76E-09	0.97 (0.0037)	1.58E-12	---
rs12593	1	227172290	<i>ADCK3</i>	T/C	1.06 (0.0067)	2.41E-16	1.04 (0.0035)	2.16E-33	+++
rs890069	2	12867504	Near <i>TRIB2</i>	T/C	0.96 (0.0069)	1.03E-08	0.98 (0.0035)	1.45E-8	---
rs10210444	2	28859979	<i>PLB1</i>	A/G	1.05 (0.0084)	1.09E-09	1.04 (0.0043)	1.49E-19	+++
rs7604689	2	143577407	<i>LRP1B-KYNU</i>	C/T	0.95 (0.0082)	6.74E-11	0.97 (0.0043)	5.88E-14	---
rs140625707	2	218529842	<i>DIRC3</i>	C/CT	1.08 (0.0132)	1.35E-09	1.06 (0.007)	2.52E-18	+++
rs62237590	3	25437648	<i>RARB</i>	C/A	1.05 (0.0086)	5.37E-09	1.02 (0.0047)	4.61E-07	+++
rs35173917	3	52760984	<i>NT5DC2</i>	C/CT	0.94 (0.0086)	5.49E-13	NA	NA	--NA
rs35256080	3	63823014	<i>ATXN7</i>	AT/A	1.05 (0.0069)	8.25E-12	1.02 (0.0035)	3.05E-08	+++
rs10663094	3	181363464	<i>SOX2-OT</i>	ACT/A	1.08 (0.0068)	7.30E-31	1.05 (0.0038)	2.22E-32	+++

rs73238577	4	39403760	<i>Near KLB</i>	C/T	1.08 (0.0137)	4.52E-09	1.02 (0.0076)	0.01486	+++
rs72868578	4	81951344	<i>C4orf22-BMP3</i>	A/C	1.12 (0.017)	4.44E-11	1.07 (0.0091)	3.08E-12	+++
rs1539508	6	43868986	<i>LOC100132354</i>	G/A	0.96 (0.007)	3.00E-08	0.98 (0.0119)	0.15	---
rs7744813	6	73643289	<i>KCNQ5</i>	A/C	1.04 (0.0067)	3.68E-09	1.02 (0.0035)	4.38E-10	+++
rs73015318	6	163853539	<i>QKI</i>	A/C	1.07 (0.0089)	1.43E-12	1.05 (0.0047)	8.70E-24	+++
rs10258092	7	28547607	<i>CREB5</i>	C/T	1.05 (0.009)	3.27E-08	1.03 (0.0047)	1.50E-08	+++
rs17172647	7	46214253	<i>IGFBP3-TNS3</i>	G/A	1.09 (0.0074)	4.22E-29	1.07 (0.0038)	5.37E-76	+++
rs62621812	7	127015083	<i>ZNF800</i>	A/G	1.22 (0.0237)	4.93E-17	1.12 (0.0119)	4.86E-22	+++
rs12114462	8	22539641	<i>BIN3-EGR3</i>	C/T	1.04 (0.007)	4.62E-08	1.02 (0.0044)	2.45E-06	+++
rs370102919	8	103676072	<i>KLF10-AZIN1</i>	G/GAA	0.96 (0.0068)	4.72E-09	NA	NA	--NA
rs1679013	9	22206987	<i>CDKN2B-DMRTA1</i>	T/C	1.07 (0.0071)	1.07E-20	1.03 (0.0035)	1.29E-13	+++
rs62547244	9	92050323	<i>SEMA4D</i>	T/C	1.04 (0.0076)	1.08E-08	1.01 (0.004)	0.082	+++
rs4742654	9	108409041	<i>FKTN-TAL2</i>	T/G	1.04 (0.0073)	5.12E-09	1.02 (0.0038)	7.03E-05	+++
rs4837205	9	130681892	<i>ST6GALNAC4-PIP5KL1</i>	C/T	0.95 (0.0094)	2.97E-08	0.98 (0.0050)	1.10E-04	---
rs1014607	10	29024130	<i>BAMBI-LINC01517</i>	A/G	0.95 (0.0073)	7.80E-11	0.96 (0.0039)	2.21E-21	---
rs2274224	10	96039597	<i>PLCE1</i>	C/G	1.04 (0.0066)	1.84E-08	1.03 (0.0034)	5.74E-24	+++
rs779436795	10	101647971	<i>DNMBP</i>	GTTTGTTT TTGTT/G	1.04 (0.0066)	1.57E-09	NA	NA	++NA
rs73386631	11	202017	<i>ODF3-BET1L</i>	T/C	1.13 (0.0155)	2.81E-14	1.06 (0.0086)	3.59E-11	+++



rs150648223	11	68942162	5' LOC338694	ATTT/A	1.17 (0.0109)	1.12E-44	1.1 (0.0059)	7.32E-60	+++
rs17739338	12	30884092	CAPRIN2	T/C	0.91 (0.0126)	2.63E-13	0.94 (0.0063)	2.64E-22	---
rs1038196	12	66343400	HMGA2	C/G	0.96 (0.0066)	1.21E-08	0.99 (0.0034)	0.001763	---
rs17608087	12	110044342	MVK-FAM222A	G/A	1.07 (0.0124)	1.70E-08	1.05 (0.0067)	1.36E-11	+++
rs7154613	14	25472083	STXBP6	T/C	1.06 (0.0097)	1.42E-08	1.03 (0.0051)	1.01E-06	+++
rs2855530	14	54421917	BMP4	C/G	1.04 (0.0066)	1.02E-08	1.03 (0.0034)	7.01E-17	+++
rs72714121	15	28334889	OCA2	T/G	1.1 (0.0118)	9.32E-16	1.05 (0.0063)	3.47E-17	+++
rs12901945	15	61802203	RORA-VPS13C	A/G	1.05 (0.0067)	1.91E-12	1.02 (0.0035)	8.38E-08	+++
rs10500355	16	7459347	RBFOX1	A/T	1.05 (0.0069)	5.47E-12	1.04 (0.0036)	6.33E-26	+++
rs73530148	16	30070540	ALDOA	T/C	1.08 (0.0133)	2.57E-09	1.04 (0.0069)	5.02E-07	+++
rs73568154	16	69884306	WWP2	A/T	1.05 (0.0066)	5.38E-15	1.02 (0.0034)	8.71E-09	+++
rs11150202	16	79835003	LINC01229	A/G	1.04 (0.0076)	9.63E-09	1.00 (0.0038)	0.63	+++
rs8074331	17	30571412	RHOT1 - RHBDL3	A/T	0.96 (0.0073)	3.11E-08	0.98 (0.0038)	7.06E-10	---
rs7207025	17	41520200	near MIR2117HG	A/G	0.96 (0.0071)	1.75E-08	0.98 (0.0036)	8.36E-10	---
rs9038	17	75495397	SEPT9	C/T	0.96 (0.0069)	8.28E-09	0.96 (0.0035)	8.90E-28	---
rs9895741	17	79603831	NPLOC4	G/A	0.93 (0.0084)	1.20E-17	0.95 (0.0037)	1.48E-42	---
rs75954926	17	81061048	3' METRNL	G/A	1.07 (0.0084)	2.61E-15	1.03 (0.0041)	4.23E-13	+++
rs61744414	19	17100550	CPAMD8	T/A	1.15 (0.0256)	3.47E-08	1.1 (0.0135)	5.02E-13	+++
rs549768142	20	10649677	JAG1	GAAAAAAAA AAAT/G	0.94 (0.0087)	3.19E-14	0.96 (0.0045)	3.34E-17	---
rs4814857	20	19457268	SLC24A3	G/A	1.16 (0.0088)	1.84E-61	1.09 (0.0047)	1.58E-70	+++

rs35089120	22	30531101	<i>HORMAD2</i>	C/CAT	1.05 (0.0089)	1.47E-08	1.03 (0.0047)	9.62E-13	+++
------------	----	----------	----------------	-------	---------------	----------	---------------	----------	-----

Highlighted in grey are previously reported loci

## REFERENCES

- 1 Congdon, N. *et al.* Prevalence of cataract and pseudophakia/aphakia among adults in the United States. *Arch Ophthalmol* **122**, 487-494, doi:10.1001/archophth.122.4.487 (2004).
- 2 Liu, Y. C., Wilkins, M., Kim, T., Malyugin, B. & Mehta, J. S. Cataracts. *Lancet* **390**, 600-612, doi:10.1016/S0140-6736(17)30544-5 (2017).
- 3 Davis, G. The Evolution of Cataract Surgery. *Mo Med* **113**, 58-62 (2016).
- 4 Shiels, A. & Hejtmanick, J. F. Biology of Inherited Cataracts and Opportunities for Treatment. *Annu Rev Vis Sci* **5**, 123-149, doi:10.1146/annurev-vision-091517-034346 (2019).
- 5 Shiels, A., Bennett, T. M. & Hejtmanick, J. F. Cat-Map: putting cataract on the map. *Mol Vis* **16**, 2007-2015 (2010).
- 6 Lou, L. *et al.* Association of Sex With the Global Burden of Cataract. *JAMA Ophthalmol* **136**, 116-121, doi:10.1001/jamaophthalmol.2017.5668 (2018).
- 7 Heiba, I. M., Elston, R. C., Klein, B. E. & Klein, R. Genetic etiology of nuclear cataract: evidence for a major gene. *Am J Med Genet* **47**, 1208-1214, doi:10.1002/ajmg.1320470816 (1993).
- 8 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, doi:10.1056/NEJM200006153422404 (2000).
- 9 Hammond, C. J. *et al.* The heritability of age-related cortical cataract: the twin eye study. *Invest Ophthalmol Vis Sci* **42**, 601-605 (2001).
- 10 Congdon, N. *et al.* Nuclear cataract shows significant familial aggregation in an older population after adjustment for possible shared environmental factors. *Invest Ophthalmol Vis Sci* **45**, 2182-2186 (2004).
- 11 Sanfilippo, P. G., Hewitt, A. W., Hammond, C. J. & Mackey, D. A. The heritability of ocular traits. *Surv Ophthalmol* **55**, 561-583, doi:10.1016/j.survophthal.2010.07.003 (2010).
- 12 Yonova-Doing, E. *et al.* Genetic and Dietary Factors Influencing the Progression of Nuclear Cataract. *Ophthalmology* **123**, 1237-1244, doi:10.1016/j.ophtha.2016.01.036 (2016).
- 13 Boutin, T. S. *et al.* Insights into the genetic basis of retinal detachment. *Hum Mol Genet* **29**, 689-702, doi:10.1093/hmg/ddz294 (2020).
- 14 Banda, Y. *et al.* Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1285-1295, doi:10.1534/genetics.115.178616 (2015).
- 15 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
- 16 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779, doi:10.1371/journal.pmed.1001779 (2015).
- 17 Chen, W. *et al.* Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics. *Genetics* **200**, 719-736, doi:10.1534/genetics.115.176107 (2015).
- 18 Kakrana, A. *et al.* iSyTE 2.0: a database for expression-based gene discovery in the eye. *Nucleic Acids Res* **46**, D875-D885, doi:10.1093/nar/gkx837 (2018).

- 19 Lachke, S. A. *et al.* iSyTE: integrated Systems Tool for Eye gene discovery. *Invest Ophthalmol Vis Sci* **53**, 1617-1627, doi:10.1167/iovs.11-8839 (2012).
- 20 Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279, doi:10.1093/bioinformatics/btw613 (2017).
- 21 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-363, doi:10.1038/ng.2213 (2012).
- 22 Mishra, A. & Macgregor, S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet* **18**, 86-91, doi:10.1017/thg.2014.79 (2015).
- 23 Anand, D. *et al.* RNA sequencing-based transcriptomic profiles of embryonic lens development for cataract gene discovery. *Hum Genet* **137**, 941-954, doi:10.1007/s00439-018-1958-0 (2018).
- 24 Anand, D. & Lachke, S. A. Systems biology of lens development: A paradigm for disease gene discovery in the eye. *Exp Eye Res* **156**, 22-33, doi:10.1016/j.exer.2016.03.010 (2017).
- 25 Lachke, S. A. *et al.* Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. *Science* **331**, 1571-1576, doi:10.1126/science.1195970 (2011).
- 26 Siddam, A. D. *et al.* The RNA-binding protein Celf1 post-transcriptionally regulates p27Kip1 and Dnase2b to control fiber cell nuclear degradation in lens development. *PLoS Genet* **14**, e1007278, doi:10.1371/journal.pgen.1007278 (2018).
- 27 Patel, N. *et al.* Novel phenotypes and loci identified through clinical genomics approaches to pediatric cataract. *Hum Genet* **136**, 205-225, doi:10.1007/s00439-016-1747-6 (2017).
- 28 Kondratov, R. V., Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V. & Antoch, M. P. Early aging and age-related pathologies in mice deficient in BMAL1, the core component of the circadian clock. *Genes Dev* **20**, 1868-1873, doi:10.1101/gad.1432206 (2006).
- 29 Dubrovsky, Y. V., Samsa, W. E. & Kondratov, R. V. Deficiency of circadian protein CLOCK reduces lifespan and increases age-related cataract development in mice. *Aging (Albany NY)* **2**, 936-944, doi:10.18632/aging.100241 (2010).
- 30 Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat Genet* **50**, 1593-1599, doi:10.1038/s41588-018-0248-z (2018).
- 31 Schnetkamp, P. P. The SLC24 Na<sup>+</sup>/Ca<sup>2+</sup>-K<sup>+</sup> exchanger family: vision and beyond. *Pflugers Arch* **447**, 683-688, doi:10.1007/s00424-003-1069-0 (2004).
- 32 Raymond, L. *et al.* Complex translocation t(1;12;14)(q42;q14;q32) and HMGA2 deletion in a fetus presenting growth delay and bilateral cataracts. *Eur J Med Genet* **58**, 591-596, doi:10.1016/j.ejmg.2015.09.006 (2015).
- 33 Gao, J. *et al.* Retrospective analysis in oculocutaneous albinism patients for the 2.7 kb deletion in the OCA2 gene revealed a co-segregation of the controversial variant, p.R305W. *Cell Biosci* **7**, 22, doi:10.1186/s13578-017-0149-3 (2017).
- 34 Shah, R. L., Guggenheim, J. A., Eye, U. K. B. & Vision, C. Genome-wide association studies for corneal and refractive astigmatism in UK Biobank demonstrate a shared role for myopia susceptibility loci. *Hum Genet* **137**, 881-896, doi:10.1007/s00439-018-1942-8 (2018).

- 35 Flitcroft, D. I. *et al.* Novel Myopia Genes and Pathways Identified From Syndromic Forms of Myopia. *Invest Ophthalmol Vis Sci* **59**, 338-348, doi:10.1167/iovs.17-22173 (2018).
- 36 Shoji, H. *et al.* A nonsense nucleotide substitution in the oculocutaneous albinism II gene underlies the original pink-eyed dilution allele (Oca2(p)) in mice. *Exp Anim* **64**, 171-179, doi:10.1538/expanim.14-0075 (2015).
- 37 Hysi, P. G. *et al.* Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. *Nat Genet* **52**, 401-407, doi:10.1038/s41588-020-0599-0 (2020).
- 38 Plotnikov, D. *et al.* A commonly occurring genetic variant within the NPLOC4-TSPAN10-PDE6G gene cluster is associated with the risk of strabismus. *Hum Genet* **138**, 723-737, doi:10.1007/s00439-019-02022-8 (2019).
- 39 Gao, X. R., Huang, H. & Kim, H. Genome-wide association analyses identify 139 loci associated with macular thickness in the UK Biobank cohort. *Hum Mol Genet* **28**, 1162-1172, doi:10.1093/hmg/ddy422 (2019).
- 40 Slavotinek, A. M. *et al.* Exome sequencing in 32 patients with anophthalmia/microphthalmia and developmental eye defects. *Clin Genet* **88**, 468-473, doi:10.1111/cge.12543 (2015).
- 41 Ma, X., Jiao, X., Ma, Z. & Hejtmančík, J. F. Polymorphism rs7278468 is associated with Age-related cataract through decreasing transcriptional activity of the CRYAA promoter. *Sci Rep* **6**, 23206, doi:10.1038/srep23206 (2016).
- 42 Ansar, M. *et al.* Bi-allelic Loss-of-Function Variants in DNMBP Cause Infantile Cataracts. *Am J Hum Genet* **103**, 568-578, doi:10.1016/j.ajhg.2018.09.004 (2018).
- 43 Kellner, U., Stohr, H., Weinitz, S., Farmand, G. & Weber, B. H. F. Mevalonate kinase deficiency associated with ataxia and retinitis pigmentosa in two brothers with MVK gene mutations. *Ophthalmic Genet* **38**, 340-344, doi:10.1080/13816810.2016.1227459 (2017).
- 44 Hayashi, S. *et al.* Heterozygous deletion at 14q22.1-q22.3 including the BMP4 gene in a patient with psychomotor retardation, congenital corneal opacity and feet polysyndactyly. *Am J Med Genet A* **146A**, 2905-2910, doi:10.1002/ajmg.a.32519 (2008).
- 45 Hollmann, A. K. *et al.* Morgagnian cataract resulting from a naturally occurring nonsense mutation elucidates a role of CPAMD8 in mammalian lens development. *PLoS One* **12**, e0180665, doi:10.1371/journal.pone.0180665 (2017).
- 46 Chen, X. *et al.* MicroRNA-26a and -26b inhibit lens fibrosis and cataract by negatively regulating Jagged-1/Notch signaling pathway. *Cell Death Differ* **24**, 1431-1442, doi:10.1038/cdd.2016.152 (2017).



