Check for updates

Multi-tissue transcriptome-wide association study identifies novel candidate susceptibility genes for cataract

Hélène Choquet[1]*, Matthieu Duot[2,3], Victor A. Herrera[1], Sanjaya K. Shrestha[2], Travis J. Meyers[1], Thomas J. Hoffmann[4,5], Poorab K. Sangani[6] and Salil A. Lachke[2,7]

[1]Kaiser Permanente Northern California (KPNC), Division of Research, Oakland, CA, United States, [2]Department of Biological Sciences, University of Delaware, Newark, DE, United States, [3]The National Centre for Scientific Research (CNRS), IGDR (Institut de Génétique et Développement de Rennes) - Joint Research Units (UMR), Univ Rennes, Rennes, France, [4]Institute for Human Genetics, University of California San Francisco (UCSF), San Francisco, CA, United States, [5]Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA, United States, [6]Department of Ophthalmology, KPNC, South San Francisco, CA, United States, [7]Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, United States

Introduction: Cataract is the leading cause of blindness among the elderly worldwide. Twin and family studies support an important role for genetic factors in cataract susceptibility with heritability estimates up to 58%. To date, 55 loci for cataract have been identified by genome-wide association studies (GWAS), however, much work remains to identify the causal genes. Here, we conducted a transcriptome-wide association study (TWAS) of cataract to prioritize causal genes and identify novel ones, and examine the impact of their expression.

Methods: We performed tissue-specific and multi-tissue TWAS analyses to assess associations between imputed gene expression from 54 tissues (including 49 from the Genotype Tissue Expression (GTEx) Project v8) with cataract using FUSION software. Meta-analyzed GWAS summary statistics from 59,944 cataract cases and 478,571 controls, all of European ancestry and from two cohorts (GERA and UK Biobank) were used. We then examined the expression of the novel genes in the lens tissue using the iSyTE database.

Results: Across tissue-specific and multi-tissue analyses, we identified 99 genes for which genetically predicted gene expression was associated with cataract after correcting for multiple testing. Of these 99 genes, 20 (*AC007773*.1, *ANKH*, *ASIP*, *ATP13A2*, *CAPZB*, *CEP95*, *COQ6*, *CREB1*, *CROCC*, *DDX5*, *EFEMP1*, *EIF2S2*, *ESRRB*, *GOSR2*, *HERC4*, *INSRR*, *NIPSNAP2*, *PICALM*, *SENP3*, and *SH3YL1*) did not overlap with previously reported cataract-associated loci. Tissue-specific analysis identified 202 significant gene-tissue associations for cataract, of which 166 (82.2%), representing 9 unique genes, were attributed to the previously reported 11q13.3 locus. Tissue-enrichment analysis revealed that gastrointestinal tissues represented one of the highest proportions of the Bonferroni-significant gene-tissue associations (21.3%). Moreover, this gastrointestinal tissue type was the only anatomical category significantly enriched in our results, after correcting for the number of tissue donors and

imputable genes for each reference panel. Finally, most of the novel cataract genes (e.g., *Capzb*) were robustly expressed in iSyTE lens data.

Discussion: Our results provide evidence of the utility of imputation-based TWAS approaches to characterize known GWAS risk loci and identify novel candidate genes that may increase our understanding of cataract etiology. Our findings also highlight the fact that expression of genes associated with cataract susceptibility is not necessarily restricted to lens tissue.

Introduction

Cataract is the leading cause of blindness among older people worldwide and is a leading cause of vision loss in the United States (U.S.), affecting 22% of Americans aged 40 years and older (1). Cataracts are characterized by the opacification of the crystalline lens, leading to progressive loss of vision. Risk factors for cataract include type 2 diabetes, high blood pressure, high body mass index, myopic refractive error, cigarette smoking, and alcohol consumption (2). However, in a recent Mendelian randomization study, we demonstrated that only genetically determined myopic refractive error and primary open-angle glaucoma were significantly associated with cataract risk (3). In addition, women have a higher cataract burden than men of the same age (4), however it is not clear why this sex difference exists.

Twin and family studies strongly support an important role for genetic factors in cataract risk with heritability estimates up to 58% (5–10). Over the past few years, genome-wide association studies (GWASs) have identified more than 50 genetic susceptibility loci for cataracts in adults (11–13). Although those GWASs revealed many genetic loci associated with cataract susceptibility, the causal genes underlying those associations remain poorly understood. Moreover, the role of potential causal genes in the lens and other tissues and cataract is unknown.

We have previously conducted a multiethnic GWAS meta-analysis of cataract (11), using the Kaiser Permanente Northern California (KPNC) Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, the UK Biobank, and data from the 23andMe research cohort, and identified 55 genetic loci associated at a genome-wide level of significance ($P < 5 \times 10^{-8}$) with cataract (11). Interestingly, one of these loci (*CASP7*) was specific to women (11). However, the number of risk factors associated with cataract specifically in women that may explain the sex difference in disease burden remain limited.

Recently, transcriptome-wide association study (TWAS) approaches have been developed to characterize established GWAS risk loci and uncover additional gene–disease associations

(14–19). These TWAS approaches leverage data from GWAS and expression quantitative trait loci (eQTL) to impute differential expression and test for gene expression associated with the GWAS disease of interest. TWASs have been fruitful in detecting functioning genes regulated by disease-associated variants, thus providing important insight into mechanisms of diseases (19).

In addition to GWAS findings, our previous GWAS meta-analysis of cataract (11) also reported positive genetic correlations between cataract with disorders other than eye disorders, including chronic pulmonary and gastrointestinal diseases. For this reason, we hypothesized that tissues in other anatomical parts of the body than the eye could be relevant to investigate, to better understand the mechanisms underlying cataract.

In this study, we conducted a TWAS of cataract to identify novel associated genes and interpret the transcriptional and disease risk mechanisms for cataract susceptibility genes. We imputed gene expression into GWAS data (59,944 cataract cases and 478,571 controls of European ancestry from GERA and UK Biobank cohorts) from our previous GWAS (11) using eQTL datasets (20) from multiple tissues (54 tissue reference panels). We conducted tissue-specific and multi-tissue TWAS analyses, as well as tissue type-enrichment analysis. Finally, we subsequently fine-mapped those associations and examined the expression of the novel genes identified in the current TWAS in lens tissues using the iSyTE database (21–24). The data sources used for the current TWAS study and TWAS analyses and results are summarized in a flowchart (Supplementary Figure S1).

Methods

Cataract GWAS data

We used summary statistics from our recent GWAS meta-analysis (11). Briefly, we conducted a meta-analysis, including 538,515 individuals of European ancestry (59,944 cataract cases and 478,571 controls) from the GERA (25) and UK Biobank (26,

27) cohorts. The meta-analysis was performed using the R package "meta" (28) and fixed-effects summary estimates were calculated for an additive model. In total, 9,056,148 single nucleotide variations (SNVs) passing quality control were used for the TWAS analyses.

For the GERA cohort, all study procedures were approved by the Institutional Review Board of the Kaiser Permanente Northern California, and written informed consent was obtained from all participants. For the UK Biobank, this research has been conducted using the UK Biobank Resource project #14105.

FUSION eQTL data

Local (*cis*) eQTL datasets for 54 tissue types were downloaded from the FUSION website. These reference data were sourced from the Genotype Tissue Expression (GTEx) Project v8 (N=49 tissue reference panels) (20), the CommonMind Consortium (CMC) (N=2 tissue reference panels) (29), the Metabolic Syndrome in Men Study (METSIM) (N=1 tissue reference panel) (30), the Netherlands Twin Registry (NTR) (N=1 tissue reference panel) (31), and the Cardiovascular Risk in Young Finns Study (YFS) (N=1 tissue reference panel) (32). Supplementary Table S1 reports the datasets sources, number of individuals, and number of imputable genes for each tissue reference panel.

Tissue-specific TWAS analyses

We conducted a TWAS of cataract using FUSION (18), which computes predictive models for eQTLs from reference data, and tests the association between predicted gene expression with a trait from GWAS summary statistics. As previously done (33), we performed tissue-specific TWAS analyses using FUSION default settings and the three following data inputs: 1) the above-mentioned GWAS summary statistics for cataract; 2) FUSION gene expression predictive models for 54 reference tissues; and 3) 1000 Genomes (European ancestry) Phase 3 data from the 1000 Genomes Project (34) as a reference panel for linkage disequilibrium (LD). Model weights for tissue-specific gene expression regressed on SNVs were computed from best linear unbiased predictor (BLUP), Bayesian sparse linear mixed model (BSLMM), least absolute shrinkage and selection operator (LASSO), and elastic net regression, as well as from the model with the top associated SNV.

A total of 325,513 gene-tissue-pairs (representing 37,920 unique genes across 54 tissue reference panels) were tested for associations between imputed gene expression with cataract susceptibility. Associations with a Bonferroni significance p-value less than 1.54×10^{-7} (=0.05/325,513) were considered significant. Novel TWAS genes were defined as those located over 1 Mb apart from any previously reported cataract GWAS loci (i.e., no prior GWAS SNVs within 1 Mb from the start or end of the gene).

Colocalization analyses

To assess whether GWAS SNVs colocalized with eQTLs, we conducted a Bayesian colocalization analysis using the

COLOCv3.2.1 software, which is implemented in FUSION using marginal expression weights, for Bonferroni-significant TWAS associations (35). Thus, we tested the hypothesis that a single variant in each TWAS-significant model was associated with both cataract (from the GWAS) and imputed gene expression. Bayesian posterior probability greater than 0.9 was considered supporting evidence for colocalization.

Conditional and joint analyses

To determine if the TWAS associations were conditionally independent of the GWAS hits, we conducted conditional analyses by adjusting transcriptome-wide associations for SNV-level effects from GWAS. Specifically, we used the COJO software program to adjust the GWAS summary statistics (the meta-analyzed results from the GERA and UK Biobank European samples) by the most statistically significant risk variants within 1 Mb of each TWAS gene (36). Using the marginal TWAS associations from the single-tissue analysis, we performed a FUSION joint analysis for cataract-associated genes located on the same chromosome region within each reference panel.

Tissue enrichment analyses

To identify tissues potentially relevant to cataract, we assigned the 54 tissue reference panels to 12 anatomical categories as per Strunz et al. (2020) (37): adipose (n = 3 reference panels), brain (n = 15), cardiovascular (n = 9), female reproductive (n = 3), gastrointestinal (n = 7), gland (n = 11), lung (n = 1), skeletal muscle (n = 1), skin (n = 2), tibial nerve (n = 1), and transformed fibroblasts (n = 1). Supplementary Table S1 lists the tissue reference panels and their corresponding anatomical categories used for this analysis. We assessed the frequency of Bonferroni-significant TWAS genes in each anatomical category. Because more Bonferroni-significant TWAS genes are expected from eQTL reference panels with more tissue donors and more imputable genes, we used the hypergeometric test to estimate the probability of observing at least as many TWAS-significant genes from all the gene-tissue pairs that we tested in each anatomical category.

Sex-specific TWAS analyses

We also conducted sex-specific TWAS analyses using sex-specific GWAS summary statistics (i.e., women and men analyzed separately) and tissue reference panels (i.e., ovary, uterus, and vagina eQTLs for women; and prostate and testis eQTLs for men).

Multi-tissue TWAS analysis

We conducted an omnibus test in FUSION for associations with cataract across multiple tissues. Specifically, TWAS associations from all 54 tissue reference panels were jointly analyzed accounting for correlation between expression weights across tissues. Two filters were applied to the omnibus test results to consider a

multi-tissue gene expression test significant: 1) using a Bonferroni correction, we divided the α of 0.05 by the effective number of genes tested (n = 13,328), and retained genes with omnibus test p-values less than this value (P < 3.75x10^{-6}); and 2) genes with a minimum tissue-specific p-value suggestive of a significant association (P < 1x10^{-5}) were retained as described by Barbeira et al (17).

Expression of novel cataract-associated genes in lens tissues

The iSyTE 2.0 database, which contains meta-analyzed mouse lens gene expression data across different stages, was used to examine the expression of the novel genes identified in the current study in the lens tissue (21–23). Mouse orthologs of the human candidate genes for the novel cataract genes identified in the current TWAS analyses were examined in iSyTE, which contains meta-analyzed lens transcriptome data generated on microarrays or RNA-sequencing (RNA-seq) (21, 23). Mouse whole lens tissue gene expression datasets at embryonic day (E) stages E10.5, E11.5, E12.5, E16.5, E17.5, E19.5, and postnatal (P) day stages P0, P2, and P56, in addition to isolated lens epithelium at P28 were available on the Affymetrix 430 2.0 platform (GeneChip Mouse Genome 430 2.0 Array and/or 430A 2.0 Array) and were used in this analysis. Further, mouse whole lens tissue gene expression datasets at stages P4, P8, P12, P20, P30, P42, P52, and P60 were available on the Illumina platform (BeadChip MouseWG-6 v2.0 Expression arrays), and were used in this analysis. We also examined RNA-seq data generated on mouse whole lens tissue at E10.5, E12.5, E14.5 and E16.5. Additionally, because lens-enriched expression of a candidate gene has proven to be an effective predictor of its role in the lens (21, 22), the lens-enrichment of these candidate genes was also investigated at these stages. "Lens-enriched expression" is a measure of expression of a candidate gene in the lens compared to that in mouse whole embryonic body (WB) as described (21–23, 38, 39). Microarray expression data is publicly available on several gene-specific perturbation mouse models that exhibit lens defects or cataract, as described (21). We analyzed these datasets to examine potential changes in expression of the novel cataract candidate genes, as done before (11). Additionally, to gain insights into expression of candidate genes specifically in lens epithelial or fiber cells, we examined previously described RNA-seq data from isolated epithelium and fiber cells (40, 41). Gene expression analysis was performed as previously described (11, 23, 42). The University of Delaware Institutional Animal Care and Use Committee (IACUC) reviewed and approved the animal protocol.

Results

Tissue-specific TWAS analysis identified 202 gene-tissue pairs associated with cataract

We found that 202 gene-tissue pairs reached the Bonferroni significance threshold for their associations between imputed gene expression with cataract susceptibility (Supplementary Table S2). While increased predicted expression was associated with cataract risk for 79 of the Bonferroni-significant gene-tissue pairs (e.g., *IGHMBP2*-colon sigmoid, z = 8.25), decreased predicted expression was associated with cataract for 123 Bonferroni-significant gene-tissue pairs (e.g., *MRPL21* - whole blood, z = -7.64) (Figures 1, 2).

These 202 gene-tissue pairs were represented by 27 unique genes across 54 tissue reference panels (Figures 1, 2). Importantly, 2 of the 27 unique genes did not overlap previously identified cataract GWAS loci: *INSRR* on chromosome 1, and *CEP95* on chromosome 17 (Table 1). Furthermore, out of the 27 genes, we found that 9 (33.3%) were located in the 11q13.3 genomic region which was previously identified in our GWAS (11). These include: *TESMIN*, *AP000808.1*, *MRPL21*, *IGHMBP2*, *MRGPRD*, *MRGPRF*, *AP003071.4*, *MRGPRF-AS1*, and *TPCN2* (Figure 1).

Interestingly, 12 genes were Bonferroni-significant in only one tissue reference panel; these included 10 genes within previously reported cataract-associated loci: *ARL4D* (thyroid); *CDC42BPA* (heart atrial appendage); *CDKN2A* (brain cortex); *GSTM2* (whole blood); *OCA2* (brain); *PKD2L1* (lung); *SEMA4D* (spleen); *ST6GALNAC4* (blood); *TESMIN* (adipose subcutaneous); and *TPCN2* (brain cerebellar hemisphere); and 2 genes newly identified in the current study: *CEP95* (skin sun exposed lower leg), and *INSRR* (brain nucleus accumbens basal ganglia) Supplementary Table S2.

To assess whether common genetic variants underly eQTL and GWAS associations with cataract, we conducted a colocalization analysis for the 202 Bonferroni-significant gene-tissue pairs. We found that 128 (63.4%) of the Bonferroni-significant gene-tissue pairs had a colocalized variant associated with both cataract risk (from GWAS) and predicted gene expression based on our TWAS results (column COLOC.PP4 in Supplementary Table S2).

Conditional analyses provide additional support for cataract TWAS associations

To identify TWAS signals for cataract independent of GWAS genome-wide significant risk variants, we repeated the FUSION analysis with GWAS summary statistics conditioned on the top GWAS SNV in each of the 202 Bonferroni-significant gene-tissue pairs. We found that all gene-tissue pairs reached nominal significance (*P* < 0.05) (Supplementary Table S2). Furthermore, we assessed joint TWAS associations in tissue reference panels with more than one Bonferroni-significant gene on the same chromosome region (Supplementary Table S3). Of the four pairwise joint models including six unique genes, all the associations were attenuated but retained marginal significance (P<0.05). All six of these genes (*IGHMBP2*, *TPCN2*, *MRPL21*, *MRGPRF-AS1*, *AP000808.1*, and *MRGPRD*) are located within the 11q13.3 chromosome region, which has been previously identified as a GWAS susceptibility locus for cataract (11, 12).
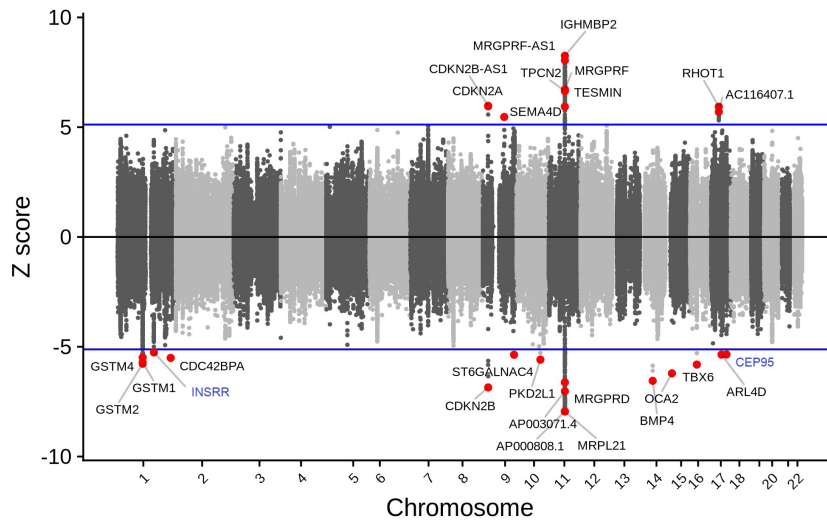
FIGURE 1
Tissue-specific TWAS analysis identified 27 unique genes associated with cataract. While increased predicted expression was associated with cataract risk for 10 genes (i.e., genes with z > 5.0; which corresponds to the results presented on the upper panel), decreased predicted expression was associated with cataract risk for 17 genes (i.e., genes with z < -5.0; which corresponds to the results presented on the lower panel). Genes in blue are novel (i.e., no prior reported GWAS SNV within 1 Mb).

Sex-specific TWAS analyses revealed 9 genes associated with cataract

Because cataracts are more common in women (4) and genetic susceptibility loci specific to women have been previously identified

(11), we evaluated sex-specific TWAS associations. We used sex-specific GWAS summary statistics and tissue reference panels, i.e., GWAS summary statistics from women for TWAS of ovary, uterus, and vagina eQTLs; and GWAS summary statistics from men for TWAS of prostate and testis eQTLs. We found that 22 of the sex-
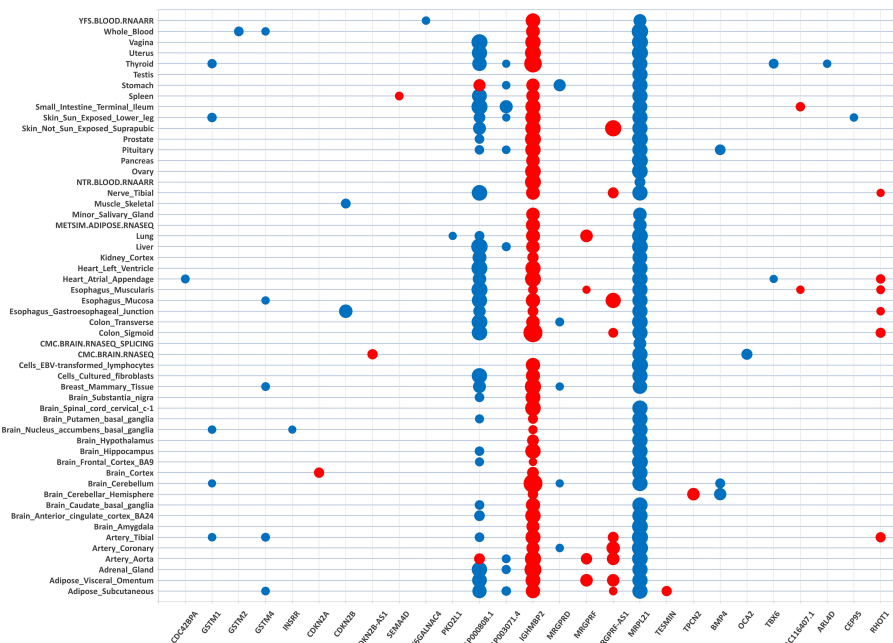
FIGURE 2
Transcriptome-wide association matrix of cataract significant gene-tissue associations. Tissue-specific TWAS analysis identified 202 gene-tissue pairs represented by 27 unique genes (on the x-axis) across 54 tissue reference panels (on the y-axis). The 27 unique genes are listed by chromosome position order (chr 1 on the left side of the matrix; chr 22 on the right side of the matrix). The tissue reference panels are listed by alphabetic order. The size of the dot for each gene-tissue association is proportional -log₁₀ (TWAS.P). Color corresponds to the predicted direction of expression changes: red and blue for increased and decreased expression changes, respectively.

TABLE 1   TWAS analyses of cataract identified 99 unique genes.

GENE	Chr:position (GRCh37/hg19)	TWAS Type analyses	Novel gene (compared to known GWAS loci that reached GW level of significance)
CROCC	**chr1:17248426-17299459**	Multi-tissue TWAS	**novel**
ATP13A2	**chr1:17312453-17338423**	Multi-tissue TWAS	**novel**
CAPZB	**chr1:19665269-19810789**	Multi-tissue TWAS	**novel**
GSTM4	chr1:110198721-110204322	Tissue-specific TWAS	GSTM2 - PMID: 34127677
GSTM2	chr1:110210679-110217908	Tissue-specific TWAS	
GSTM1	chr1:110230439-110236367	Tissue-specific TWAS	
ADAM15	chr1:155023792-155031159	Multi-tissue TWAS	DPM3-KRTCAP2 - PMID: 34127677
EFNA1	chr1:155100352-155107375	Multi-tissue TWAS	
INSRR	**chr1:156809855-156828909**	Tissue-specific TWAS	**novel**
ADCK3	chr1:227127938-227175246	Multi-tissue TWAS	ADCK3 - PMID: 34127677
CDC42BPA	chr1:227177559-227506193	both	
SH3YL1	**chr2:218136-261130**	Multi-tissue TWAS	**novel**
TRIB2	chr2:12857062-12882860	Multi-tissue TWAS	near TRIB2 - PMID: 34127677
TRMT61B	chr2:29072687-29093175	Multi-tissue TWAS	PLB1 - PMID: 34127677
EFEMP1	**chr2:56093102-56150917**	Multi-tissue TWAS	**novel**
CREB1	chr2:208394616-208470284	Multi-tissue TWAS	**novel**
DIRC3	chr2:218148746-218621316	Multi-tissue TWAS	DIRC3 - PMID: 34127677
CXCR2	chr2:218990736-219001976	Multi-tissue TWAS	
PPM1M	chr3: 52279808-52284615	Multi-tissue TWAS	NT5DC2 - PMID: 34127677
GLYCTK	chr3:52321844-52329273	Multi-tissue TWAS	
WDR82	chr3:52288438-52312659	Multi-tissue TWAS	
SEMA3G	chr3:52467268-52479043	Multi-tissue TWAS	
NT5DC2	chr3:52558403-52567873	Multi-tissue TWAS	
PBRM1	chr3:52579383-52719615	Multi-tissue TWAS	
GNL3	chr3:52719936-52728513	Multi-tissue TWAS	
SPCS1	chr3:52739857-52742197	Multi-tissue TWAS	
NEK4	chr3:52742460-52804956	Multi-tissue TWAS	
ITIH1	chr3:52811615-52826078	Multi-tissue TWAS	
PRKCD	chr3:53195225-53226733	Multi-tissue TWAS	
THOC7	chr3:63819546-63849481	Multi-tissue TWAS	ATXN7 - PMID: 34127677
ATXN7	chr3:63884075-63989136	Multi-tissue TWAS	
ANKH	**chr5:14704909-14871887**	Multi-tissue TWAS	**novel**
HLA-DQB1	chr6:32627244-32634434	Multi-tissue TWAS	3' HLA-DQB1 - PMID: 31816047
QKI	chr6:163835675-163999628	Multi-tissue TWAS	QKI - PMID: 34127677
IGFBP3	chr7:45951844-45960871	Multi-tissue TWAS	IGFBP3-TNS3 - PMID: 34127677
NIPSNAP2	**chr7:56032278-56067872**	Multi-tissue TWAS	**novel**
C8orf58	chr8:22457112-22461655	Multi-tissue TWAS	BIN3-EGR3 - PMID: 34127677

(Continued)

TABLE 1 Continued

GENE	Chr:position (GRCh37/hg19)	TWAS Type analyses	Novel gene (compared to known GWAS loci that reached GW level of significance)
BIN3	chr8:22477931-22526634	Multi-tissue TWAS	
CCAR2	chr8:22462270-22477984	Multi-tissue TWAS	
CDKN2B-AS1	chr9:21994790-22077889	Tissue-specific TWAS	CDKN2B-DMRTA1 - PMID: 34127677
CDKN2A	chr9:21967751-21974856	Tissue-specific TWAS	
CDKN2B	chr9:22002902-22009304	Tissue-specific TWAS	
DMRTA1	chr9:22446823-22455739	Multi-tissue TWAS	
SEMA4D	chr9:91975702-92094805	both	SEMA4D - PMID: 34127677
FKTN	chr9:108320411-108403399	Multi-tissue TWAS	FKTN-TAL2 - PMID: 34127677
ST6GALNAC4	chr9:130670165-130679320	Tissue-specific TWAS	ST6GALNAC4-PIP5KL1 - PMID: 34127677
HERC4	chr10:69681656-69835103	Multi-tissue TWAS	**novel**
PLCE1	chr10:95753688-96092580	Multi-tissue TWAS	PLCE1 - PMID: 34127677
NOC3L	chr10:96092988-96122683	Multi-tissue TWAS	
ABCC2	chr10:101542397-101612351	Multi-tissue TWAS	DNMBP - PMID: 34127677
PKD2L1	chr10:102047906-102089985	both	
ENTPD7	chr10:101419266-101470998	Multi-tissue TWAS	
CUTC	chr10:101491991-101515891	Multi-tissue TWAS	
DNMBP	chr10:101635328-101673849	Multi-tissue TWAS	
MRPL21	chr11:68658746-68671300	both	11q13.3 - PMID: 34127677
IGHMBP2	chr11:68671359-68708069	both	
TESMIN	chr11:68474908-68518988	both	
AP000808.1	chr11:68708971-68710320	Tissue-specific TWAS	
MRGPRD	chr11:68747490-68748455	Tissue-specific TWAS	
AP003071.4	chr11:68768233-68769516	Tissue-specific TWAS	
MRGPRF-AS1	chr11:68779822-68785915	Tissue-specific TWAS	
MRGPRF	chr11:68771866-68780714	both	
TPCN2	chr11:68816400-68858065	both	
PICALM	**chr11:85668218-85780126**	Multi-tissue TWAS	**novel**
CAPRIN2	chr12:30862487-30907885	Multi-tissue TWAS	CAPRIN2 - PMID: 34127677
UBE3B	chr12:109915439-109928527	Multi-tissue TWAS	MVK-FAM222A - PMID: 34127677
BMP4	chr14:54416454-54420113	both	BMP4 - PMID: 34127677
COQ6	**chr14:74416629-74430373**	Multi-tissue TWAS	**novel**
ESRRB	**chr14:76837614-76968180**	Multi-tissue TWAS	**novel**
OCA2	chr15:28000021-28344461	both	OCA2 - PMID: 34127677
MVP	chr16:29831715-29859360	Multi-tissue TWAS	ALDOA - PMID: 34127677
TBX6	chr16:30097114-30103245	Tissue-specific TWAS	
NFAT5	chr16:69599869-69738569	Multi-tissue TWAS	WWP2 - PMID: 34127677

(Continued)

TABLE 1 Continued

GENE	Chr:position (GRCh37/hg19)	TWAS Type analyses	Novel gene (compared to known GWAS loci that reached GW level of significance)
NOB1	chr16:69775757-69788871	Multi-tissue TWAS	
WWP2	chr16:69796186-69975644	Multi-tissue TWAS	
CLEC18A	chr16:69984805-69997889	Multi-tissue TWAS	
NPIPB14P	chr16:70010291-70030091	Multi-tissue TWAS	
NQO1	chr16:69743304-69760463	Multi-tissue TWAS	
PDXDC2P	chr16:70,010,201-70,099,851	Multi-tissue TWAS	
PDPR	chr16:70147529-70196440	Multi-tissue TWAS	
DDX19A	chr16:70380806-70407286	Multi-tissue TWAS	
COG4	chr16:70514470-70557457	Multi-tissue TWAS	
SENP3	**chr17:7465236-7475287**	Multi-tissue TWAS	**novel**
UTP6	chr17:30187923-30228727	Multi-tissue TWAS	
RHBDL3	chr17:30592851-30651678	Multi-tissue TWAS	*RHOT1-RHBDL3* - PMID: 34127677
RHOT1	chr17:30469521-30552746	both	
AC116407.1	chr17:30462748-30462833	Tissue-specific TWAS	
CNTNAP1	chr17:40834549-40852011	Multi-tissue TWAS	near *MIR2117HG* - PMID: 34127677
ARL4D	chr17:41476361-41478505	Tissue-specific TWAS	
GOSR2	**chr17:45000526-45014188**	Multi-tissue TWAS	**novel**
DDX5	**chr17:62494374-62502484**	Multi-tissue TWAS	**novel**
CEP95	**chr17:62503095-62534064**	Tissue-specific TWAS	**novel**
AC007773.1	**chr19:32868188-32868273**	Multi-tissue TWAS	**novel**
NECTIN2	chr19:45349554-45382195	Multi-tissue TWAS	near *EXOC3L2* - PMID: 34127677
JAG1	chr20:10618332-10654694	Multi-tissue TWAS	*JAG1* - PMID: 34127677
SLC24A3	chr20:19193286-19703570	Multi-tissue TWAS	*SLC24A3* - PMID: 34127677
EIF2S2	chr20:32676115-32700085	Multi-tissue TWAS	**novel**
ASIP	chr20:32848171-32857148	Multi-tissue TWAS	
MTMR3	chr22:30279163-30426857	Multi-tissue TWAS	*HORMAD2* - PMID: 34127677

Genes in bold are novel (i.e., no prior reported GWAS SNV within 1 Mb).

specific tests reached the Bonferroni significance level that we applied to the main analysis (P<1.54x10^{-7}), including 3, 5, and 4 genes for ovary, uterus, and vagina, respectively, and 7 and 3 genes for prostate and testis, respectively (Supplementary Table S4). Of those 22 sex-specific associations, 9 unique genes were identified, all of these genes (*ITPKB, AC104162.1, AP000808.1, MRPL21, IGHMBP2, CAPRIN2, CLEC18A, LINC01229* and *AC003681.1*) were located nearby previously identified GWAS loci for cataract (11, 12). For instance, while differential gene expression of *MRPL21* at 11q13.3 was associated with cataract in the 5 sex-specific tissue reference panels (i.e., ovary, uterus, vagina, prostate, and testis), differential gene expression of *ITPKB* was associated with cataract in vagina only.

Importance of gastrointestinal tissues in cataract susceptibility

Across the 54 tissue reference panels, the greatest number of Bonferroni-significant gene-tissue pairs was observed for the GTEx adipose subcutaneous, artery tibial, and thyroid datasets (seven gene-tissue pairs for each dataset), followed by the GTEx artery aorta, heart atrial appendage, esophagus muscularis, and skin sun exposed datasets (six gene-tissue pairs for each dataset) (Supplementary Table S1). To identify tissues potentially relevant to cataract, we assigned tissue reference panels to anatomical categories as described above in the Methods (Supplementary Table S1). In the tissue-specific TWAS results, gastrointestinal

tissues represented one of the highest proportion of the 202 Bonferroni-significant genes (43 genes; 21.3%) (Supplementary Figure S2). Interestingly, this gastrointestinal tissue type was the only anatomical category significantly enriched in our results, after accounting for the number of gene-tissue pairs tested per anatomical category (p-value from the hypergeometric test = 0.0055) (Supplementary Table S5).

Multi-tissue TWAS revealed additional novel candidate genes for cataract susceptibility

The multi-tissue TWAS using the omnibus test in FUSION revealed 86 genes for which imputed expression was associated with cataract susceptibility (Bonferroni p-value <0.05/13,328 effective gene tests ≈ 3.75×10^{-6} and minimum tissue-specific p-value < 1×10^{-5}) (Supplementary Table S6). Interestingly, 14 of the 86 multi-tissue associated genes were also associated with cataract in tissue-specific models, including *INSRR* (chr1), *CDC42BPA* (chr1), *CDKN2A* (chr9), *CDKN2B* (chr9), *SEMA4D* (chr9), *ST6GALNAC4* (chr9), *PKD2L1* (chr10), *TESMIN* (chr11), *IGHMBP2* (chr11), *BMP4* (chr14), *OCA2* (chr15), *TBX6* (chr16), *RHOT1* (chr17) and*ARL4D* (chr17). In addition, 18 of the 86 multi-tissue associated genes were located outside of previously described risk loci (Table 1). These included: *CROCC* (chr1), *ATP13A2* (chr1), *CAPZB* (chr1), *SH3YL1* (chr2), *EFEMP1* (chr2), *CREB1* (chr2), *ANKH* (chr5), *NIPSNAP2* (chr7), *HERC4* (chr10), *PICALM* (chr11), *COQ6* (chr14), *ESRRB* (chr14), *SENP3* (chr17), *GOSR2* (chr17), *DDX5* (chr17), *AC007773.1* (chr19), *EIF2S2* (chr20), and *ASIP* (chr20).

Gene expression in the lens tissue

We identified the mouse orthologs for 19 of the 20 novel genes as follows. For *NIPSNAP2*, in the Affy and Illumina microarray data, the gene alias for mouse *Gbas* was used. For AC007773.1, *ZNF507* (mouse ortholog, *Zfp507*) and *DPY19L3* (mouse ortholog, *Dpy19l3*) were considered as candidate genes. We first examined the expression of these genes in the lens tissue across various stages using the iSyTE microarray database (21, 22). While majority of the genes were found to be expressed, several exhibited robust expression (Figure 3A). For example, *Atp13a2*, *Capzb*, *Crocc*, *Efemp1*, *Gbas*, *Gosr2*, *Picalm*, *Senp3* and *Zfp507* had high expression in Affymetrix datasets. When examined for "enriched expression" in the lens, several candidates (*e.g.*, *Atp13a2*, *Capzb*, *Cep95*, *Crocc*, *Dpy19l3*, *Efemp1*, *Esrrb*, *Gbas*, *Gosr2*, *Insrr*, *Picalm*, and *Senp3*) were identified (Figure 3B). Moreover, RNA-seq data from whole lens tissue confirmed 10 of the mouse orthologs to have expression or enriched expression in the lens (Supplementary Figures S3, 4). Further, expression data from isolated lens epithelium and fiber cells at different time-points, spanning embryonic (E14.5 through newborn) through aging stages (3 months through age 2 years) showed that all the novel cataract candidate genes with mouse orthologs exhibit robust expression in the epithelium and/or fiber cells (Supplementary Figure S5).

Interestingly, this cell-specific data also shows that three candidate genes are specifically enriched in the postnatal epithelium (*e.g.*, *Efemp1*, *Esrrb*, *Insrr*) and a subset of these exhibit progressively high expression with aging in the epithelium (*e.g.*, *Efemp1*, *Insrr*). Finally, all novel candidate genes exhibited differential expression in at least one gene-perturbation mouse models of lens defects/cataract (Supplementary Figure S6).

Discussion

By leveraging data from GWAS and eQTL, we identified 99 genes associated with cataract susceptibility (16 from the tissue-specific analysis alone, 69 from the multi-tissue analysis alone, and 14 from both analyses). Of these 99 genes, 20 were novel to the extent they did not overlap known cataract risk loci from GWAS (2 of these, *INSRR* and *CEP95*, were from tissue-specific models, and 18 from the multi-tissue analysis). Mouse orthologs of the vast majority of the human candidate genes were found to be robustly expressed in the lens. We also highlighted the contribution of the 11q13.3 genomic region in cataract susceptibility. Our results implicated a role for gastrointestinal tissues and confirmed the importance of the lens in cataract etiology.

Our tissue-specific TWAS analysis identified *INSRR* (1q23.1) and *CEP95* (17q23.3) as novel cataract-associated genes. *INSRR* encodes the insulin receptor related receptor which is involved in the transmembrane receptor protein tyrosine kinase activity, actin cytoskeleton reorganization, and the protein autophosphorylation, and has an important role in the alkaline pH-dependent activation mechanism (43). *CEP95* encodes the centrosomal protein 95 and belongs to the family of proteins containing coiled-coil domains (CCDCs), which are involved in several functions in cell growth and development, such as regulation of gene expression (44). To date, no mutations in either *INSRR* or *CEP95* have been linked to eye diseases, and further studies are needed to confirm the role of these genes in cataract etiology and determine their precise role in cataract susceptibility.

Our multi-tissue TWAS analysis identified *ATP13A2* (1p36.13), *CAPZB* (1p36.13), *EFEMP1* (2p16.1), and *SENP3* (17p13.1) associated with cataract, all were not previously reported as significant in GWA studies of cataract. *ATP13A2* encodes a member of the P5 subfamily of ATPases which transports inorganic cations as well as other substrates. The *ATP13A2* locus has been previously reported to be associated with age-related cataract in a GWAS conducted in a Chinese cohort (45); however, this association did not reach a genome-wide level of significance (lead SNV rs2871776, $P=4.18 \times 10^{-5}$), possibly due to limited sample size (total of 191 cataract cases and 208 controls) (45). *CAPZB* encodes the beta subunit of the barbed-end actin binding protein, which belongs to the F-actin capping protein family. Interestingly, *CAPZB* is located within the 1p36 chromosome region that was previously linked to congenital cataract in three genetic linkage studies (46–48). However, no segregating mutations that contribute to congenital cataract were identified in this *CAPZB* gene in a six-generation Australian family displaying linkage to chromosome 1p36 (49). *EFEMP1* encodes a
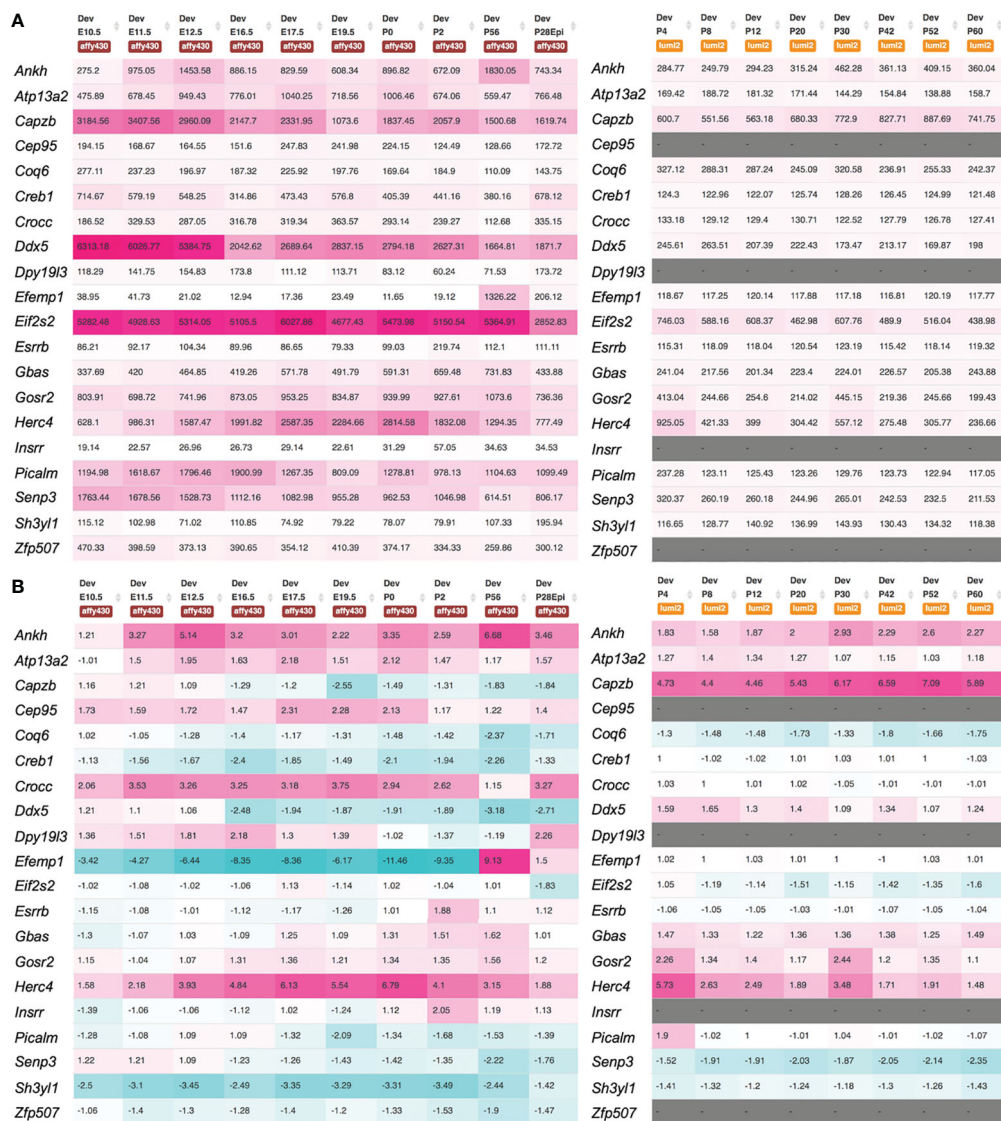
A

Affymetrix (affy430)

	Dev E10.5	Dev E11.5	Dev E12.5	Dev E16.5	Dev E17.5	Dev E19.5	Dev P0	Dev P2	Dev P56	Dev P28Epi
Ankh	275.2	975.05	1453.58	886.15	829.59	608.34	896.82	672.09	1830.05	743.34
Atp13a2	475.89	678.45	949.43	776.01	1040.25	718.56	1006.04	674.06	559.47	766.48
Capzb	3184.56	3407.56	2960.09	2147.7	2331.95	1073.6	1837.45	2057.9	1500.68	1619.74
Cep95	194.15	168.67	164.55	151.6	247.83	241.98	224.15	124.49	128.66	172.72
Coq6	277.11	237.23	196.97	187.32	225.92	197.76	169.64	184.9	110.09	143.75
Creb1	714.67	579.19	548.25	314.86	473.43	576.8	405.39	441.16	380.16	678.12
Crocc	186.52	329.53	287.05	316.78	319.34	363.57	293.14	239.27	112.68	335.15
Ddx5	6313.18	6026.71	5384.75	2042.62	2689.64	2837.15	2794.18	2627.31	1664.81	1871.7
Dpy19l3	118.29	141.75	154.83	173.8	111.12	113.71	83.12	60.24	71.53	173.72
Efemp1	38.95	41.73	21.02	12.94	17.36	23.49	11.65	19.12	1326.22	206.12
Eif2s2	5282.48	4928.63	5314.05	5105.5	6027.66	4677.43	5473.98	5150.54	5364.91	2852.83
Esrrb	86.21	92.17	104.34	89.96	86.65	79.33	99.03	219.74	112.1	111.11
Gbas	337.69	420	464.85	419.26	571.78	491.79	591.31	659.48	731.83	433.88
Gosr2	803.91	698.72	741.96	873.05	953.25	834.87	939.99	927.61	1073.6	736.36
Herc4	628.1	986.31	1587.47	1991.82	2587.35	2284.66	2814.58	1832.08	1294.35	777.49
Insrr	19.14	22.57	26.96	26.73	29.14	22.61	31.29	57.05	34.63	34.53
Picalm	1194.98	1618.67	1796.46	1900.99	1267.35	809.09	1278.81	978.13	1104.63	1099.49
Senp3	1763.44	1678.56	1528.73	1112.16	1082.98	955.28	962.53	1046.98	614.51	806.17
Sh3yl1	115.12	102.98	71.02	110.85	74.92	79.22	78.07	79.91	107.33	195.94
Zfp507	470.33	398.59	373.13	390.65	354.12	410.39	374.17	334.33	259.86	300.12

Illumina (luml2)

	Dev P4	Dev P8	Dev P12	Dev P20	Dev P30	Dev P42	Dev P52	Dev P60
Ankh	284.77	249.79	294.23	315.24	462.28	361.13	409.15	360.04
Atp13a2	169.42	188.72	181.32	171.44	144.29	154.84	138.88	158.7
Capzb	600.7	551.56	563.18	680.33	772.9	827.71	887.69	741.75
Cep95	-	-	-	-	-	-	-	-
Coq6	327.12	288.37	287.24	245.09	320.58	236.91	255.33	242.37
Creb1	124.3	122.96	122.07	125.74	128.26	126.45	124.99	121.48
Crocc	133.18	129.12	129.4	130.71	122.52	127.79	126.78	127.41
Ddx5	245.61	263.51	207.39	222.43	173.47	213.17	169.87	198
Dpy19l3	-	-	-	-	-	-	-	-
Efemp1	118.67	117.25	120.14	117.88	117.18	116.81	120.19	117.77
Eif2s2	746.03	588.16	608.37	462.98	607.76	489.9	516.04	438.98
Esrrb	115.31	118.09	118.04	120.54	123.19	115.42	118.14	119.32
Gbas	241.04	217.56	201.34	223.4	224.01	226.57	205.38	243.88
Gosr2	413.04	244.66	254.6	214.02	445.15	219.36	245.66	199.43
Herc4	925.05	421.33	399	304.42	557.12	275.48	305.77	236.66
Insrr	-	-	-	-	-	-	-	-
Picalm	237.28	123.11	125.43	123.26	129.76	123.73	122.94	117.05
Senp3	320.37	260.19	260.18	244.96	265.01	242.53	232.5	211.53
Sh3yl1	116.65	128.77	140.92	136.99	143.93	130.43	134.32	118.38
Zfp507	-	-	-	-	-	-	-	-

B

Affymetrix (affy430)

	Dev E10.5	Dev E11.5	Dev E12.5	Dev E16.5	Dev E17.5	Dev E19.5	Dev P0	Dev P2	Dev P56	Dev P28Epi
Ankh	1.21	3.27	5.14	3.2	3.01	2.22	3.35	2.59	6.68	3.46
Atp13a2	-1.01	1.5	1.95	1.63	2.18	1.51	2.12	1.47	1.17	1.57
Capzb	1.16	1.21	1.09	-1.29	-1.2	-2.55	-1.49	-1.31	-1.83	-1.84
Cep95	1.73	1.59	1.72	1.47	2.31	2.28	2.13	1.17	1.22	1.4
Coq6	1.02	-1.05	-1.28	-1.4	-1.17	-1.31	-1.48	-1.42	-2.37	-1.71
Creb1	-1.13	-1.56	-1.67	-2.4	-1.49	-1.49	-2.1	-1.94	-2.26	-1.33
Crocc	2.06	3.53	3.26	3.25	3.18	3.75	2.94	2.62	1.15	3.27
Ddx5	1.21	1.1	1.06	-2.48	-1.94	-1.87	-1.91	-1.89	-3.18	-2.71
Dpy19l3	1.36	1.51	1.81	2.18	1.3	1.39	-1.02	-1.37	-1.19	2.26
Efemp1	-3.42	-4.27	-6.44	-8.35	-8.36	-6.17	-11.46	-9.35	9.13	1.5
Eif2s2	-1.02	-1.08	-1.02	-1.06	1.13	-1.14	1.02	-1.04	1.01	-1.83
Esrrb	-1.15	-1.08	-1.01	-1.12	-1.17	-1.26	1.01	1.88	1.1	1.12
Gbas	-1.3	-1.07	1.03	-1.09	1.25	1.09	1.31	1.51	1.62	1.01
Gosr2	1.15	-1.04	1.07	1.31	1.36	1.21	1.34	1.35	1.56	1.2
Herc4	1.58	2.18	3.93	4.84	6.13	5.54	6.79	4.1	3.15	1.88
Insrr	-1.39	-1.06	-1.06	-1.12	1.02	-1.24	1.12	2.05	1.19	1.13
Picalm	-1.28	-1.08	1.09	1.09	-1.32	-2.09	-1.34	-1.68	-1.53	-1.39
Senp3	1.22	1.21	1.09	-1.23	-1.26	-1.43	-1.42	-1.35	-2.22	-1.76
Sh3yl1	-2.5	-3.1	-3.45	-2.49	-3.35	-3.29	-3.31	-3.49	-2.44	-1.42
Zfp507	-1.06	-1.4	-1.3	-1.28	-1.4	-1.2	-1.33	-1.53	-1.9	-1.47

Illumina (luml2)

	Dev P4	Dev P8	Dev P12	Dev P20	Dev P30	Dev P42	Dev P52	Dev P60
Ankh	1.83	1.58	1.87	2	2.93	2.29	2.6	2.27
Atp13a2	1.27	1.4	1.34	1.27	1.07	1.15	1.03	1.18
Capzb	4.73	4.4	4.46	5.43	6.17	6.59	7.09	5.89
Cep95	-	-	-	-	-	-	-	-
Coq6	-1.3	-1.48	-1.48	-1.73	-1.33	-1.8	-1.66	-1.75
Creb1	1	-1.02	-1.02	1.01	1.03	1.01	1	-1.03
Crocc	1.03	1	1.01	1.02	-1.05	-1.01	-1.01	-1.01
Ddx5	1.59	1.65	1.3	1.4	1.09	1.34	1.07	1.24
Dpy19l3	-	-	-	-	-	-	-	-
Efemp1	1.02	1	1.03	1.01	1	-1	1.03	1.01
Eif2s2	1.05	-1.19	-1.14	-1.51	-1.15	-1.42	-1.35	-1.6
Esrrb	-1.06	-1.05	-1.05	-1.03	-1.01	-1.07	-1.05	-1.04
Gbas	1.47	1.33	1.22	1.36	1.36	1.38	1.25	1.49
Gosr2	2.26	1.34	1.4	1.17	2.44	1.2	1.35	1.1
Herc4	5.73	2.63	2.49	1.89	3.48	1.71	1.91	1.48
Insrr	-	-	-	-	-	-	-	-
Picalm	1.9	-1.02	1	-1.01	1.04	-1.01	-1.02	-1.07
Senp3	-1.52	-1.91	-1.91	-2.03	-1.87	-2.05	-2.14	-2.35
Sh3yl1	-1.41	-1.32	-1.2	-1.24	-1.18	-1.3	-1.26	-1.43
Zfp507	-	-	-	-	-	-	-	-

FIGURE 3

Expression of novel candidate genes for cataract in the mouse lens. Mouse orthologs of the human candidate genes were examined for their lens expression in the iSyTE microarray datasets. **(A)** Analysis of whole lens tissue data on the Affymetrix and Illumina microarray platforms at different embryonic (E) and postnatal (P) stages indicates that majority of the candidates are expressed in the lens. The heat-map denotes the range of expression on either the Affymetrix or Illumina platform, while the number represents the mean fluorescence intensity for individual genes. **(B)** Mouse orthologs of the human candidate genes were examined for their "lens enriched" expression in the iSyTE microarray data. "Enriched expression" in the lens is estimated by analyzing the fold-change enrichment of candidate gene expression in the lens compared to that in whole embryonic body as indicated by the number and heatmap. Please note that WEB in the Affymetrix data represents expression in whole embryonic body in **(A)** and P28Epi in the Affymetrix data represents expression data on isolated lens epithelium in **(A, B)**.

member of the fibulin family of extracellular matrix glycoproteins, and mutations in this gene have been shown to cause Doyne honeycomb retinal dystrophy and familial juvenile-onset open-angle glaucoma (50, 51). Recently, *EFEMP1* has been demonstrated to be a potential biomarker for choroidal neovascularization in age-related macular degeneration, and for choroidal thickness change in myopia (52, 53). Interestingly, a transcriptome analysis of neural progenitor cells derived from patients with Lowe syndrome, a multisystem disorder characterized notably by anomalies affecting the eye, including congenital cataracts, identified *EFEMP1* as a candidate gene (54). Future studies will clarify how *EFEMP1* contributes to cataract

susceptibility. Our study also identified *SENP3* as a cataract-associated gene. *SENP3* encodes the SUMO specific peptidase 3 which is a de-sumoylation enzyme (SENP) that plays an important role in regulating eye development and is highly expressed in vertebrate ocular cell lines, including human, mouse, and rabbit lens epithelial cells lines (55, 56). Previous works have demonstrated that sumoylation function plays indispensable roles during lens differentiation (57, 58). Recently, glucose oxidase and UVA irradiation seem to affect the expression patterns of the *SENP*s, including *SENP3*, in the *in vitro* cataract models, providing evidence to link sumoylation function to stress-induced cataractogenesis (59). Moreover, the changing patterns in some *SENP*s levels seem

to act as molecular markers for both senile and complicated cataracts (60). Additional studies investigating the sumoylation functions and the related mechanisms in cataract development and progression will help to understand the role of *SENP3* in cataract susceptibility.

Among the 9 unique genes identified in the tissue-sex specific analyses of cataract, we identified *ITPKB* (1q42.12), which was differentially expressed in the vagina. *ITPKB* encodes the inositol-trisphosphate 3-kinase B that plays an important role in the regulation of the levels of a large number of inositol polyphosphates (61). A *de novo* 5.8-Mb deletion encompassing the chromosome 1q42.12q42.2 region (and *ITPKB*, among other genes) was reported in a 4-year-old child who presented hypoplastic corpus callosum and bilateral cataracts, in addition to other clinical features such as epilepsy (62). Future investigations may determine the implication of *ITPKB* in cataract etiology.

Our study should be interpreted within the context of its limitations. Although GTEx data for the 49 tissues represent the most comprehensive eQTL dataset of human tissues, it does not include ocular tissues and consequently we may have failed to identify the real causal genes in the unsampled ocular tissue. However, we have confirmed using the iSyTE database that novel cataract genes identified in the current study are robustly expressed in lens tissue, which is a cataract relevant eye tissue. Moreover, although cataract is primarily a lens disorder, it has been demonstrated that most complex diseases, including vision disorders such as age-related macular degeneration, might manifest in several tissues across the body (63). Despite the great success in prioritizing gene-trait associations in complex diseases and traits, TWAS may present multiple hits per locus, owing to co-regulation, which remain problematic (14, 19). Thus, future models could consider more complex genetic architecture containing different regulatory effects, and our TWAS results could benefit from subsequent functional assays to indicate the potential targets underlying the identified associations, notably at 11q13.3. Despite these limitations, our TWAS study is based on results from a large GWAS meta-analysis on almost 60,000 cataract cases, enabling the prioritization and the discovery of potential causal genes for cataract. Finally, in the current study, we performed a multi-tissue TWAS analysis which enables increased statistical precision compared to single-tissue approaches (17, 64, 65).

Our study also highlighted the important contribution of gastrointestinal tissues in cataract susceptibility consistent with previous work showing associations between cataracts and gastrointestinal disorders (66, 67). For instance, patients diagnosed with early-onset cataracts have been shown to be at increased risk of peptic ulcer (66). Furthermore, rare syndromic disorders for which patients present early-onset congenital cataracts can present gastrointestinal disorders as additional features (68–71). For instance, patients with Lowe Syndrome (oculocerebrorenal syndrome) can present both dense congenital cataracts and gastroesophageal reflux (68). Similarly, patients with inherited spastic paraplegia can present with bilateral cataracts and gastroesophageal reflux with persistent vomiting (69). A splice

site mutation in *CYP27A1* has been reported to lead to cerebrotendinous xanthomatosis which can be characterized by pulverulent cataracts and gastrointestinal problems such as diarrhea (71). Recently, pathogenic variants in the *WFS1/RP1/NOD2* genes have been shown to cause congenital cataract, retinitis pigmentosa, and Crohn's disease in a five generation British family (70). A comprehensive evaluation of systemic disorders associated with age-related cataract – as previously done for dry eye disease (72)- would help to identify which gastrointestinal disorders are risk factors for cataract. Altogether, expression of genes associated with cataract seems not to be restricted to lens tissue, as could be expected for this lens disorder, and the processes underlying cataract pathology seem to be systemic as observed for other vision disorders, such as age-related macular degeneration and exfoliation syndrome (37, 73).

In conclusion, we identified 99 genes associated with cataract susceptibility, of which 20 did not overlap with known cataract risk loci. Our results provide evidence of the utility of imputation-based TWAS approaches to characterize known GWAS risk loci and identify novel candidate genes that may increase our understanding of cataract etiology.

Data availability statement

The original contributions presented in the study are publicly available. FUSION models trained on the GTEx version 8 data are available here: http://gusevlab.org/projects/fusion/. Gene expression and eQTL data are freely available at https://gtexportal.org/home/datasets. Expression or lens-enriched expression heat-map for candidate genes can be accessed through the iSyTE web-tool (https://research.bioinformatics.udel.edu/iSyTE). The RNA-seq data can be found here: Gene Expression Omnibus (GEO; GSE113887, GSE166619, GSE119596). The microarray data can be found here: Gene Expression Omnibus (GEO; GSE100136, GSE32334, GSE65500, GSE47694, GSE16533, GSE31643, GSE9711, GSE13402, GSE25775, GSE25776, GSE22322, GSE22362, GSE9711). To protect individual's privacy, complete GERA data are available upon approved applications to the KP Research Bank Portal (https://researchbank.kaiserpermanente.org/). A subset of the GERA cohort consented for public use can be found at NIH/dbGaP: phs000674.v3.p3.

Ethics statement

The studies involving humans were approved by The Institutional Review Board of the Kaiser Permanente Northern California. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. The animal study was approved by The University of Delaware Institutional Animal Care and Use Committee (IACUC). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

HC: Conceptualization, Funding acquisition, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. MD: Formal analysis, Visualization, Writing – review & editing. VH: Formal analysis, Visualization, Writing – review & editing. SS: Formal analysis, Visualization, Writing – review & editing. TM: Conceptualization, Methodology, Writing – review & editing. TH: Methodology, Resources, Writing – review & editing. PS: Conceptualization, Writing – review & editing. SL: Conceptualization, Funding acquisition, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fopht.2024.1362350/full#supplementary-material

References

1. Congdon N, Vingerling JR, Klein BE, West S, Friedman DS, Kempen J, et al. Prevalence of cataract and pseudophakia/aphakia among adults in the United States. *Arch Ophthalmol*. (2004) 122:487–94. doi: 10.1001/archopht.122.4.487

2. Liu YC, Wilkins M, Kim T, Malyugin B, Mehta JS. Cataracts. *Lancet*. (2017) 390:600–12. doi: 10.1016/S0140-6736(17)30544-5

3. Jiang C, Melles RB, Sangani P, Hoffmann TJ, Hysi PG, Glymour MM, et al. Association of behavioral and clinical risk factors with cataract: A two-sample mendelian randomization study. *Invest Ophthalmol Vis Sci*. (2023) 64:19. doi: 10.1167/iovs.64.10.19

4. Lou L, Ye X, Xu P, Wang J, Xu Y, Jin K, et al. Association of sex with the global burden of cataract. *JAMA Ophthalmol*. (2018) 136:116–21. doi: 10.1001/jamaophthalmol.2017.5668

5. Heiba IM, Elston RC, Klein BE, Klein R. Genetic etiology of nuclear cataract: evidence for a major gene. *Am J Med Genet*. (1993) 47:1208–14. doi: 10.1002/ajmg.1320470816

6. Hammond CJ, Snieder H, Spector TD, Gilbert CE. Genetic and environmental factors in age-related nuclear cataracts in monozygotic and dizygotic twins. *N Engl J Med*. (2000) 342:1786–90. doi: 10.1056/NEJM200006153422404

7. Hammond CJ, Duncan DD, Snieder H, de Lange M, West SK, Spector TD, et al. The heritability of age-related cortical cataract: the twin eye study. *Invest Ophthalmol Vis Sci*. (2001) 42:601–5.

8. Congdon N, Broman KW, Lai H, Munoz B, Bowie H, Gilber D, et al. Nuclear cataract shows significant familial aggregation in an older population after adjustment for possible shared environmental factors. *Invest Ophthalmol Vis Sci*. (2004) 45:2182–6. doi: 10.1167/iovs.03-1163

9. Sanfilippo PG, Hewitt AW, Hammond CJ, Mackey DA. The heritability of ocular traits. *Surv Ophthalmol*. (2010) 55:561–83. doi: 10.1016/j.survophthal.2010.07.003

10. Yonova-Doing E, Forkin ZA, Hysi PG, Williams KM, Spector TD, Gilbert CE, et al. Genetic and dietary factors influencing the progression of nuclear cataract. *Ophthalmology*. (2016) 123:1237–44. doi: 10.1016/j.ophtha.2016.01.036

11. Choquet H, Melles RB, Anand D, Yin J, Cuellar-Partida G, Wang W, et al. A large multiethnic GWAS meta-analysis of cataract identifies new risk loci and sex-specific effects. *Nat Commun*. (2021) 12:3595. doi: 10.1038/s41467-021-23873-8

12. Boutin TS, Charteris DG, Chandra A, Campbell S, Hayward C, Campbell A, et al. Insights into the genetic basis of retinal detachment. *Hum Mol Genet*. (2020) 29:689–702. doi: 10.1093/hmg/ddz294

13. Yonova-Doing E, Zhao W, Igo RP Jr., Wang C, Sundaresan P, Lee KE, et al. Common variants in SOX-2 and congenital cataract genes contribute to age-related nuclear cataract. *Commun Biol*. (2020) 3:755. doi: 10.1038/s42003-020-01421-2

14. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet*. (2019) 51:592–9. doi: 10.1038/s41588-019-0385-z

15. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PloS Genet*. (2010) 6:e1000895. doi: 10.1371/journal.pgen.1000895

16. Mancuso N, Gayther S, Gusev A, Zheng W, Penney KL, Kote-Jarai Z, et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat Commun*. (2018) 9:4079. doi: 10.1038/s41467-018-06302-1

17. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. *PloS Genet*. (2019) 15:e1007889. doi: 10.1371/journal.pgen.1007889

18. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. (2016) 48:245–52. doi: 10.1038/ng.3506

19. Mai J, Lu M, Gao Q, Zeng J, Xiao J. Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun Biol*. (2023) 6:899. doi: 10.1038/s42003-023-05279-y

20. GTEx project maps wide range of normal human genetic variation: A unique catalog and follow-up effort associate variation with gene expression across dozens of body tissues. *Am J Med Genet A*. (2018) 176:263–4. doi: 10.1002/ajmg.a.38426

21. Kakrana A, Yang A, Anand D, Djordjevic D, Ramachandruni D, Singh A, et al. iSyTE 2.0: a database for expression-based gene discovery in the eye. *Nucleic Acids Res.* (2018) 46:D875–85. doi: 10.1093/nar/gkx837

22. Lachke SA, Ho JW, Kryukov GV, O'Connell DJ, Aboukhalil A, Bulyk ML, et al. iSyTE: integrated Systems Tool for Eye gene discovery. *Invest Ophthalmol Vis Sci.* (2012) 53:1617–27. doi: 10.1167/iovs.11-8839

23. Anand D, Kakrana A, Siddam AD, Huang H, Saadi I, Lachke SA. RNA sequencing-based transcriptomic profiles of embryonic lens development for cataract gene discovery. *Hum Genet.* (2018) 137:941–54. doi: 10.1007/s00439-018-1958-0

24. Aryal S, Anand D, Hernandez FG, Weatherbee BAT, Huang H, Reddy AP, et al. MS/MS in silico subtraction-based proteomic profiling as an approach to facilitate disease gene discovery: application to lens development and cataract. *Hum Genet.* (2020) 139:151–84. doi: 10.1007/s00439-019-02095-5

25. Banda Y, Kvale MN, Hoffmann TJ, Hesselson SE, Ranatunga D, Tang H, 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.* (2015) 200:1285–95. doi: 10.1534/genetics.115.178616

26. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* (2018) 562:203–9. doi: 10.1038/s41586-018-0579-z

27. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, 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.* (2015) 12:e1001779. doi: 10.1371/journal.pmed.1001779

28. Balduzzi S, Rucker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health.* (2019) 22:153–60. doi: 10.1136/ebmental-2019-300117

29. Hoffman GE, Bendl J, Voloudakis G, Montgomery KS, Sloofman L, Wang YC, et al. CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci Data.* (2019) 6:180. doi: 10.1038/s41597-019-0183-6

30. Laakso M, Kuusisto J, Stancakova A, Kuulasmaa T, Pajukanta P, Lusis AJ, et al. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J Lipid Res.* (2017) 58:481–93. doi: 10.1194/jlr.O072629

31. Ligthart L, van Beijsterveldt CEM, Kevenaar ST, de Zeeuw E, van Bergen E, Bruins S, et al. The Netherlands twin register: longitudinal research based on twin and twin-family designs. *Twin Res Hum Genet.* (2019) 22:623–36. doi: 10.1017/thg.2019.93

32. Akerblom HK, Uhari M, Pesonen E, Dahl M, Kaprio EA, Nuutinen EM, et al. Cardiovascular risk in young Finns. *Ann Med.* (1991) 23:35–9. doi: 10.3109/07853899109147928

33. Meyers TJ, Yin J, Herrera VA, Pressman AR, Hoffmann TJ, Schaefer C, et al. Transcriptome-wide association study identifies novel candidate susceptibility genes for migraine. *HGG Adv.* (2023) 4:100211. doi: 10.1016/j.xhgg.2023.100211

34. Genomes Project, C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* (2015) 526:68–74. doi: 10.1038/nature15393

35. Giambartolomei C, Vukcevic D, SChadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PloS Genet.* (2014) 10:e1004383. doi: 10.1371/journal.pgen.1004383

36. Yang J, Ferreira T, Morris AP, Medland SEGenetic Investigation of, A.T.C. and Replication, D.I.G., Meta-analysis, C., et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* (2012) 44:369–375, S361–363. doi: 10.1038/ng.2213

37. Strunz T, Lauwen S, Kiel CInternational, A.M.D.G.C., , Hollander AD, Weber BHF. A transcriptome-wide association study based on 27 tissues identifies 106 genes potentially relevant for disease pathology in age-related macular degeneration. *Sci Rep.* (2020) 10:1584. doi: 10.1038/s41598-020-58510-9

38. Anand D, Agrawal S, Siddam A, Motohashi H, Yamamoto M, Lachke SA. An integrative approach to analyze microarray datasets for prioritization of genes relevant to lens biology and disease. *Genom Data.* (2015) 5:223–7. doi: 10.1016/j.gdata.2015.06.017

39. Anand D, Lachke SA. Systems biology of lens development: A paradigm for disease gene discovery in the eye. *Exp Eye Res.* (2017) 156:22–33. doi: 10.1016/j.exer.2016.03.010

40. Zhao Y, Zheng D, Cvekl A. A comprehensive spatial-temporal transcriptomic analysis of differentiating nascent mouse lens epithelial and fiber cells. *Exp Eye Res.* (2018) 175:56–72. doi: 10.1016/j.exer.2018.06.004

41. Faranda AP, Shihan MH, Wang Y, Duncan MK. The aging mouse lens transcriptome. *Exp Eye Res.* (2021) 209:108663. doi: 10.1016/j.exer.2021.108663

42. Siddam AD, Duot M, Coomson SY, Anand D, Aryal S, Weatherbee BAT, et al. High-throughput transcriptomics of celf1 conditional knockout lens identifies downstream networks linked to cataract pathology. *Cells.* (2023) 12(7):1070. doi: 10.3390/cells12071070

43. Wang L, Hall C, Li J, Choi E, Bai XC. Structural basis of the alkaline pH-dependent activation of insulin receptor-related receptor. *Nat Struct Mol Biol.* (2023) 30:661–9. doi: 10.1038/s41594-023-00974-0

44. Liu S, Jia L, Quan B, Rong G, Li M, Xie R, et al. Coiled-coil domain-containing protein 45 is a potential prognostic biomarker and is associated with immune cell

enrichment of hepatocellular carcinoma. *Dis Markers.* (2022) 2022:7745315. doi: 10.1155/2022/7745315

45. Li JK, Li LL, Li W, Wang ZW, Gao FJ, Hu FY, et al. Panel-based targeted exome sequencing reveals novel candidate susceptibility loci for age-related cataracts in Chinese Cohort. *Mol Genet Genomic Med.* (2020) 8:e1218. doi: 10.1002/mgg3.1218

46. Eiberg H, Lund AM, Warburg M, Rosenberg T. Assignment of congenital cataract Volkmann type (CCV) to chromosome 1p36. *Hum Genet.* (1995) 96:33–8. doi: 10.1007/BF00214183

47. Ionides AC, Berry V, Mackay DS, Moore AT, Bhattacharya SS, Shiels A. A locus for autosomal dominant posterior polar cataract on chromosome 1p. *Hum Mol Genet.* (1997) 6:47–51. doi: 10.1093/hmg/6.1.47

48. McKay JD, Patterson B, Craig JE, Russell-Eggitt IM, Wirth MG, Burdon KP, et al. The telomere of human chromosome 1p contains at least two independent autosomal dominant congenital cataract genes. *Br J Ophthalmol.* (2005) 89:831–4. doi: 10.1136/bjo.2004.058495

49. Burdon KP, Hattersley K, Lachke SA, Laurie KJ, Maas RL, Mackey DA, et al. Investigation of eight candidate genes on chromosome 1p36 for autosomal dominant total congenital cataract. *Mol Vis.* (2008) 14:1799–804.

50. Collantes ERA, Delfin MS, Fan B, Torregosa JMR, Siguan-Bell C, Florcruz NVG, et al. EFEMP1 rare variants cause familial juvenile-onset open-angle glaucoma. *Hum Mutat.* (2022) 43:240–52. doi: 10.1002/humu.24320

51. Stone EM, Lotery AJ, Munier FL, Heon E, Piguet B, Guymer RH, et al. A single EFEMP1 mutation associated with both Malattia Leventinese and Doyne honeycomb retinal dystrophy. *Nat Genet.* (1999) 22:199–202. doi: 10.1038/9722

52. Cheng L, Chen C, Guo W, Liu K, Zhao Q, Lu P, et al. EFEMP1 overexpression contributes to neovascularization in age-related macular degeneration. *Front Pharmacol.* (2020) 11:547436. doi: 10.3389/fphar.2020.547436

53. Shi WQ, Wan T, Li B, Li T, Zhou XD. EFEMP1 is a potential biomarker of choroid thickness change in myopia. *Front Neurosci.* (2023) 17:1144421. doi: 10.3389/fnins.2023.1144421

54. Liu H, Barnes J, Pedrosa E, Herman NS, Salas F, Wang P, et al. Transcriptome analysis of neural progenitor cells derived from Lowe syndrome induced pluripotent stem cells: identification of candidate genes for the neurodevelopmental and eye manifestations. *J Neurodev Disord.* (2020) 12:14. doi: 10.1186/s11689-020-09317-2

55. Liu Y, Zhang L, Tang X, Liu F, Fu JL, Gong XD, et al. Determination of expression patterns of seven de-sumoylation enzymes in major ocular cell lines. *Curr Mol Med.* (2018) 18:584–93. doi: 10.2174/1566524019666190107153440

56. Liu Y, Liu F, Wang L, Fu JL, Luo ZW, Nie Q, et al. Localization analysis of seven de-sumoylation enzymes (SENPs) in ocular cell lines. *Curr Mol Med.* (2018) 18:523–32. doi: 10.2174/1566524019666190112142025

57. Yan Q, Gong L, Deng M, Zhang L, Sun S, Liu J, et al. Sumoylation activates the transcriptional activity of Pax-6, an important transcription factor for eye and brain development. *Proc Natl Acad Sci U.S.A.* (2010) 107:21034–9. doi: 10.1073/pnas.1007866107

58. Gong L, Ji WK, Hu XH, Hu WF, Tang XC, Huang ZX, et al. Sumoylation differentially regulates Sp1 to control cell differentiation. *Proc Natl Acad Sci U.S.A.* (2014) 111:5574–9. doi: 10.1073/pnas.1315034111

59. Xiang JW, Xiao Y, Gan Y, Chen H, Liu Y, Wang L, et al. Glucose oxidase- and UVA-induced changes in the expression patterns of seven de-sumoylation enzymes (SENPs) are associated with cataract development. *Curr Mol Med.* (2019) 19:48–53. doi: 10.2174/1566524019666190311094313

60. Liu FY, Fu JL, Wang L, Nie Q, Luo Z, Hou M, et al. Molecular signature for senile and complicated cataracts derived from analysis of sumoylation enzymes and their substrates in human cataract lenses. *Aging Cell.* (2020) 19:e13222. doi: 10.1111/acel.13222

61. Erneux C, Ghosh S, Koenig S. Inositol(1,4,5)P3 3-kinase isoenzymes: Catalytic properties and importance of targeting to F-actin to understand function. *Adv Biol Regul.* (2016) 60:135–43. doi: 10.1016/j.jbior.2015.09.004

62. Radha Rama Devi A, Ganapathy A, Mannan AU, Sabharanjak S, Naushad SM. 1q42.12q42.2 deletion in a child with midline defects and hypoplasia of the corpus callosum. *Mol Syndromol.* (2019) 10:161–6. doi: 10.1159/000496079

63. Ongen H, Brown AA, Delaneau O, Panousis NI, Nica ACConsortium, G.T., et al. Estimating the causal tissues for complex traits and diseases. *Nat Genet.* (2017) 49:1676–83. doi: 10.1038/ng.3981

64. Grinberg NF, Wallace C. Multi-tissue transcriptome-wide association studies. *Genet Epidemiol.* (2021) 45:324–37. doi: 10.1002/gepi.22374

65. Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, et al. A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat Genet.* (2019) 51:568–76. doi: 10.1038/s41588-019-0345-7

66. Hsia NY, Tsai YY, Lin CL, Chiang CC. Increased risk of peptic ulcer in patients with early-onset cataracts: A nationwide population-based study. *PloS One.* (2018) 13: e0207193. doi: 10.1371/journal.pone.0207193

67. Trumler AA. Evaluation of pediatric cataracts and systemic disorders. *Curr Opin Ophthalmol.* (2011) 22:365–79. doi: 10.1097/ICU.0b013e32834994dc

68. Lewis RA, Nussbaum RL, Brewer ED. Lowe Syndrome. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, . Gripp KW, Amemiya A, editors. *GeneReviews((R))* (1993). Seattle (WA): University of Washington, Seattle.

69. Lo Nigro C, Cusano R, Gigli GL, Forabosco P, Valente M, Ravazzolo R, et al. Genetic heterogeneity in inherited spastic paraplegia associated with epilepsy. *Am J Med Genet A*. (2003) 117A:116–21. doi: 10.1002/ajmg.a.10141

70. Berry V, Ionides A, Georgiou M, Quinlan RA, Michaelides M. Multimorbidity due to novel pathogenic variants in the WFS1/RP1/NOD2 genes: autosomal dominant congenital lamellar cataract, retinitis pigmentosa and Crohn's disease in a British family. *BMJ Open Ophthalmol*. (2023) 8(1):e001252. doi: 10.1136/bmjophth-2023-001252

71. Bourkiza R, Joyce S, Patel H, Chan M, Meyer E, Maher ER, et al. Cerebrotendinous xanthomatosis (CTX): an association of pulverulent cataracts and pseudo-dominant developmental delay in a family with a splice site mutation in CYP27A1–a case report. *Ophthalmic Genet*. (2010) 31:73–6. doi: 10.3109/13816810903584963

72. Vehof J, Snieder H, Jansonius N, Hammond CJ. Prevalence and risk factors of dry eye in 79,866 participants of the population-based Lifelines cohort study in the Netherlands. *Ocul Surf*. (2021) 19:83–93. doi: 10.1016/j.jtos.2020.04.005

73. Hirbo JB, Pasutto F, Gamazon ER, Evans P, Pawar P, Berner D, et al. Analysis of genetically determined gene expression suggests role of inflammatory processes in exfoliation syndrome. *BMC Genomics*. (2023) 24:75. doi: 10.1186/s12864-023-09179-7