Novel Cataract Risk Loci Identified from a Large-Scale Multi-ethnic Genome-Wide Association Meta-Analysis

Dima L. Chaar[1], Chen Jiang[1], Vidhu Choudhary[1], Thomas J. Hoffmann[2,3], Poorab Sangani[4], Salil A. Lachke[5,6], Hélène Choquet[1,7]

[1]Kaiser Permanente Northern California (KPNC), Division of Research (DOR), Pleasanton, CA, USA; [2]Institute for Human Genetics, University of California, San Francisco (UCSF), San Francisco, CA, USA; [3]Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA, USA; [4]KPNC, Department of Ophthalmology, South San Francisco, CA, USA; [5]Department of Biological Sciences, University of Delaware, Newark, DE, USA; [6]Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, USA; [7]Department of Health Systems Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, USA

Background

- Cataract is the leading cause of blindness among the elderly worldwide[1].
- Cataract is a multifactorial disease characterized by the clouding of the lens **(Fig. 1),** resulting in reduced vision, caused by tissue degradation and protein clumping, as well as environmental and genetic risk factors[2,3].
- Twin and family studies highlight the important role of genetic factors to cataract risk, with heritability estimates up to 58%[4-9].
- To date, >100 cataract-associated loci have been identified by genome-wide association studies (GWAS) [9-12].

Figure 1. Medical illustration of cataract

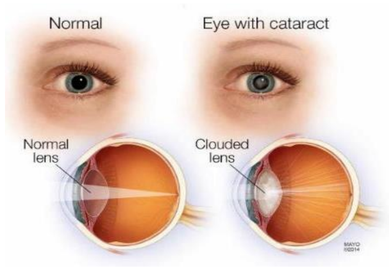


Image source: NEI

Hypothesis

Additional loci remain to be discovered to improve our understanding of the genetic etiology of cataract.

Methods

❖ Study Population

- Individuals were selected from the **Kaiser Permanente Research Bank (KPRB)**, which includes clinical, demographic, and genetic data on ~400,000 adult members of the Kaiser Permanente (KP) Medical Care Plan.
- KPRB which includes high-density genotype data that were generated using the ThermoFisher Precision Medicine Diversity Array (**PMDA v2**)
- The **TOPMed** (Release 3) imputation reference panel was used, containing ~445 million variants.
- **Cataract cases** were defined as patients with pseudophakia were diagnosed by a KP ophthalmologist and were identified in the KP electronic health record (EHR) based on International Classification of Disease (ICD-9 or -10) diagnosis codes: ICD-9:V43.1 and ICD-10: Z96.1; or history of cataract surgery at KP.
- **Control group** included all non-cases and excluded individuals with evidence of cataract.
- In total, **76,380 cataract cases** and **212,135 controls** from 4 race/ethnicity groups were included (**Table 1**).

Table 1. Characteristics of cataract cases and controls from the KPRB cohort

	Cataract Cases	Controls
Total, N (proportion that are cases)	76,380 (26.5%)	212,135
Sex, N (proportion that are cases)		
Female	43,162 (23.9%)	137,709
Male	33,218 (30.9%)	74,426
Race/Ethnicity, N (proportion that are cases)		
non-Hispanic white	63,184 (30.8%)	141,938
African American	3,562 (18.5%)	15,668
East Asian	4,513 (17.0%)	22,069
Hispanic/Latino	5,121 (13.6%)	32,460
Age, Mean ± SD (years)	71.9 ± 8.6	52.8 ± 15.6

❖GWAS and Meta-Analyses

- Logistic model GWAS for cataract was conducted separately in each race/ethnicity group of the KPRB cohort, adjusting for age, sex, and genetic ancestry principal components using **REGENIE v3.6** [13].
- **GWA meta-analysis** combining our KPRB results with publicly available GWAS summary statistics[12] for cataract was conducted, including a total of **198,105 cataract cases** and **1,033,991 controls** (**Fig. 2**).
- To **prioritize genes** within the identified loci and **biological pathways** related to cataract susceptibility, we conducted gene-based and pathways association analyses using Versatile Gene-based Association Study-2 version 2 (**VEGAS2v02**) integrative tool[14].
- **Functional enrichment** analysis was performed on the loci identified in the multi-ethnic GWAS meta-analysis, using the **g:GOSt tool** on the web software g:Profiler[15].
  - Loci were mapped to known functional informational sources (e.g., Gene Ontology (GO): molecular function (MF), GO: biological process (BP), and GO: cellular component (CC)).
  - The g:SCS method (Set Counts and Sizes) in g:Profiler was applied for multiple testing correction and pathway results were presented with an adjusted *P*-value (*P*-adj) <0.05.

Results

Figure 2. Flow Chart of the Study Design

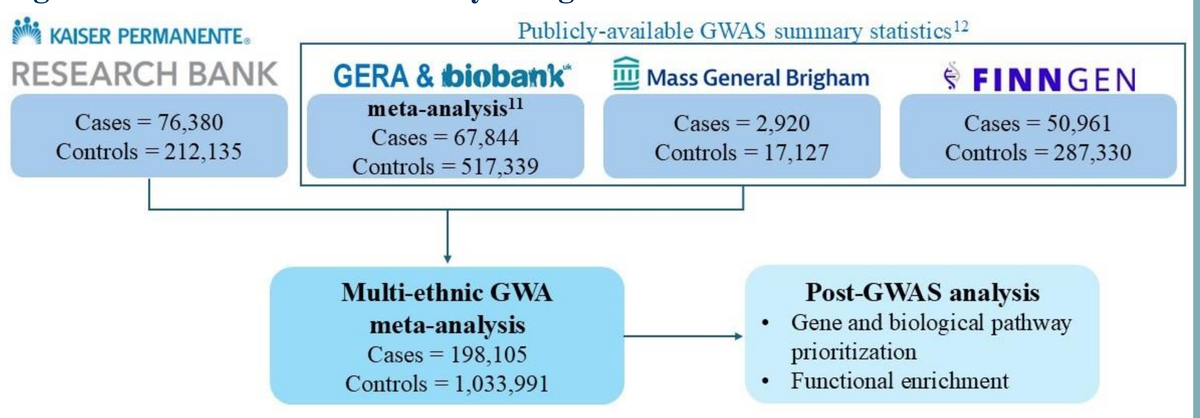
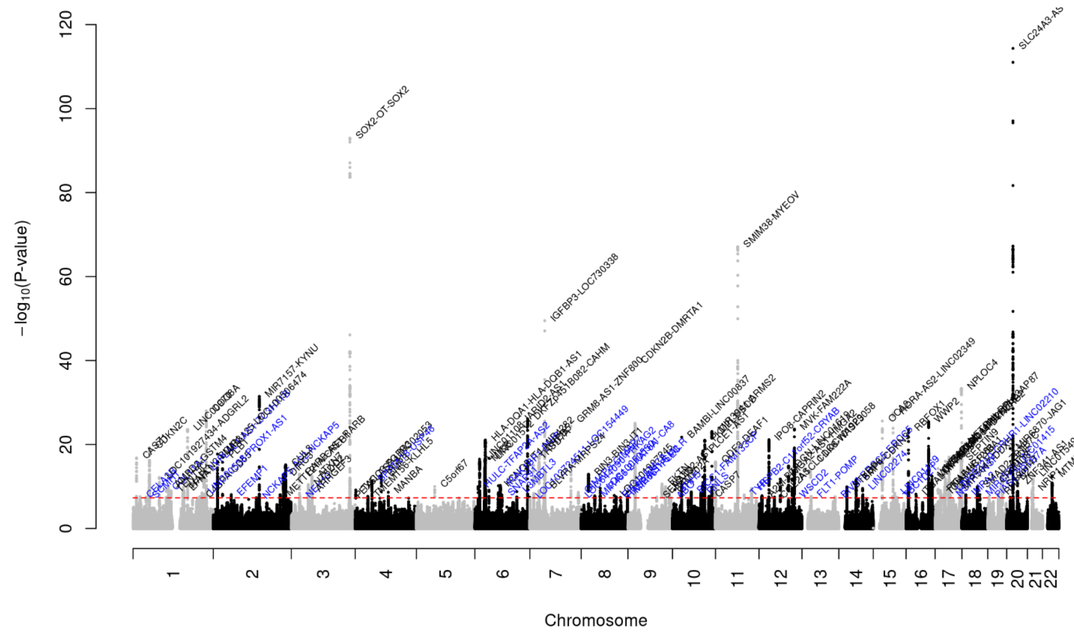


Figure 3. Multi-ethnic GWA meta-analysis of cataract combining KPRB results and publicly-available GWA summary statistics

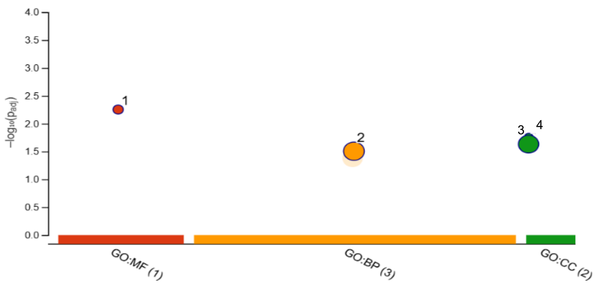


❖ Our multi-ethnic GWA meta-analysis identified **128** genome-wide significant ($P<5.0\times10^{-8}$) loci associated with cataract, of which **38 were novel** (**Fig. 3 –** loci in **blue**), as they were located over 1 Mb apart from any previously reported locus[9-12] and the identified lead genetic variants were not in LD with previously reported variants [9-12].

❖ VEGAS2 gene-based association analysis prioritized **195 genes** within the identified cataract-associated loci ($P<2.3\times10^{-6}$ = 0.05/21,402 genes tested) and **8 gene-set/pathways** significantly enriched ($P<5.1\times10^{-6}$= 0.05/9,734 pathways tested).

❖ Associated loci were enriched in pathways related to **cAMP response element binding** (*P*-adj=5.6×10^{-3}), anatomical structure development (*P*-adj=0.03), collagen type XI trimer (*P*-adj=0.02), and extracellular region (*P*-adj=0.02) (**Fig. 4**).

Figure 4. Functional enrichment analysis prioritized biological pathways underlying genetic associations with cataract



ID	Source	Term Name	*P*-adj
1	GO:MF	cAMP response element binding	5.6×10^{-3}
2	GO:BP	Anatomical structure development	0.03
3	GO:CC	Collagen type XI trimer	0.02
4	GO:CC	Extracellular region	0.02

Conclusions

❖ In this large multi-ethnic GWA meta-analysis of cataract, we identified 128 cataract-associated loci, of which 38 were novel.

❖ The current study represents the largest and most ethnically diverse genetic investigation of cataract conducted to date to our knowledge.

❖ Study findings enhance our understanding of the genetic etiology of cataract risk by uncovering novel loci and prioritizing genes and biological pathways.

References: 1. Pesudovs, K. *et al*. Eye. 2024; **2.** Nizami, A. *et al*. StatPearls 2024; **3.** Shiels, A. *et al.* Exp. Eye Res. 2021; **4.** Hammond, C. *et al.* N. Engl. J. Med. 1993; **5.** Heiba, I. *et al*. Am. J. Med. Genet. 1993; **6.** Hammond, C. *et al.* IOVS. 2001; **7.** Congdon, N. *et al.* IOVS 2004. **8.** Sanfilippo, P. *et al.* Surv. Ophthalmol. 2010. **9.** Yonova-Doing, E. *et al.* Commun. Biol. 2020. **10.** Boutin. T. *et al.* HMG 2020. **11.** Choquet, H. *et al.* Nat. Commun. 2021. **12.** Diaz-Torres, S. *et al.* Nat. Commun. 2024. **13.** Mbatchou, J. *et al.* Nat. Genet. 2021. **14.** Mishra, A. *et al.* Twin Res & Hum Gen. 2015. **15.** Kolberg, L. *et al.* F1000Res 2020. **Funding Support:** This research is supported by the National Eye Institute of the National Institutes of Health Award:R01EY033010.