npj | genomic medicine

Article

Published in partnership with CEGMR, King Abdulaziz University

Article in Press

https://doi.org/10.1038/s41525-026-00612-w

Deep-learning-derived glaucoma-related endophenotypes enable novel genome-wide genetic and functional discovery

Check for updates

Liyin Chen[1,2], Yan Zhao[1], Saber Kazeminasab Hashemabad[1], Tobias Elze[1], Mohammad Eslami[1], Mengyu Wang[1], Janey L. Wiggs[1,2], Ayellet V. Segre[1,2,3] & Nazlee Zebardast[1,3]

The genetic architecture of primary open-angle glaucoma (POAG) remains largely unexplained due to the reliance of previous genome-wide association studies (GWAS) on imprecise phenotypes from electronic health records. We overcome this with a disease-trained, task-transfer machine learning (ML) framework that learns glaucoma-related damage patterns from a clinical repository of 18,985 optical coherence tomography (OCT) scans from 8323 glaucoma patients. By applying the derived endophenotypes to 39,146 UK Biobank participants, we performed GWAS in European, African, and Asian ancestral groups, followed by cross-ancestry meta-analyses. We identified 36 and 43 LD-independent GWAS loci that passed genome-wide significance in the EUR and cross-ancestry meta-analyses, respectively. About two-thirds of the identified loci overlapped with previously reported POAG-related associations, demonstrating the validity of our approach. Importantly, more than a third of the loci were novel to glaucoma. Extensive functional analyses, including Bayesian colocalization analysis, gene-based association tests, Mendelian randomization, and single-cell enrichment analysis, converged on 11 putative causal genes, five of which were novel to glaucoma. These genes implicate Wnt-mediated outflow dysfunction and retinal ganglion cell vulnerability as candidate mechanisms in POAG. Our findings expanded POAG genetic associations, provided mechanistic insights at cell-type resolution, and proposed plausible putative causal genes.

Glaucoma is a leading cause of irreversible blindness worldwide, affecting over 80 million people worldwide and is projected to reach >110 million by 2040[1,2]. Because current treatments primarily reduce intraocular pressure (IOP), deeper molecular insight is essential to shift glaucoma care from damage limitation to neuroprotection and prevention. As glaucoma is a highly heritable disease, especially its most common form, primary open-angle glaucoma (POAG), numerous genome-wide association studies (GWAS) have been carried out for this disease over the last two decades[3–11]. Despite these efforts, currently known POAG-associated loci only explain about 9.4% of the disease heritability[5,12].

Accurate disease phenotyping is necessary for successful GWAS, and a limitation of prior glaucoma GWAS is their reliance on diagnostic codes and case-control definitions. These conventional phenotyping methods suffer from noise inherent to diagnostic codes in electronic health records[13,14]. By contrast, studies based on clinically examined cases and controls are often underpowered due to smaller sample sizes. Moreover, the case-control phenotyping approach fails to capture the full spectrum of disease heterogeneity. To address these limitations, intermediate quantitative endophenotypes are often used, because they are measured on a continuous scale, offering greater statistical power than dichotomous labels. Using a multi-trait analysis, prior work has demonstrated the utility of endophenotypes such as IOP[15] and vertical cup-to-disc-ratio (vCDR) for genome-wide discovery for POAG[6]. However, while these endophenotypes significantly improved GWAS power, they are indirect and imperfect surrogates for glaucomatous damage[16,17].

An objective indicator of glaucomatous change includes thinning of the retinal nerve fiber layer (RNFL) and the ganglion cell complex (GCC)[17] as measured by optical coherence tomography (OCT) scans, routinely used in clinical practice. Several studies have performed GWAS on average retinal layer thicknesses calculated from OCT scans[18–20]. The association signals from these studies spanned both loci previously implicated in ocular or systemic disease and loci that modulate normal inter-individual variation in retinal morphology, demonstrating that retinal layer thicknesses capture functional and structural variations in the retina. However, the use of global

[1]Department of Ophthalmology, Mass Eye and Ear, Harvard Medical School, Boston, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]These authors contributed equally: Ayellet V. Segre, Nazlee Zebardast. e-mail: nazlee_zebardast@meei.harvard.edu

average thicknesses, which compress information from 2-D thickness maps to a single value, misses spatially heterogeneous patterns of damage that are typical of glaucomatous eyes. For example, Currant et al.[19] attempted to use the average inner retinal layer thickness as an endophenotype for POAG but only replicated one known POAG locus.

With the recent advances in machine learning (ML) methods, nuanced endophenotypes from medical images, including OCT-derived retinal thickness maps, have been extracted for enhanced GWAS[21–25]. For example, GWAS performed directly on deep-learning embeddings of retinal fundus images[21] and on self-supervised, contrastive-learning representations of such images[22] have each recovered genetic signals not captured by conventional phenotypes, while deep-learning quantification of the retinal vasculature has uncovered numerous novel loci linked to ocular and cardiometabolic traits[26,27]. In addition, Alipanahi et al.[28] showed that ML-derived optic nerve head morphology phenotypes can substantially expand the catalog of glaucoma-associated loci, underscoring the value of disease-relevant ML phenotypes for genetic discovery. More recently, GWAS has been performed on latent representations of retinal structural variability extracted from OCT-derived thickness maps[29,30]. These studies extract richer features than global average layer thicknesses, but because they learn representations entirely from a largely healthy UKBB population, they may be less sensitive to variation specific to glaucomatous damage. We therefore hypothesize that such representations emphasize general inter-individual morphological variation more than the spatially heterogeneous, disease-relevant changes that characterize glaucoma.

To allow for disease-specific genome-wide discovery, here, we developed a task-transfer approach to improve quantitative phenotyping for POAG. We used unsupervised ML and a large clinical repository of OCT scans from 8323 open-angle glaucoma (OAG) patients to capture clinically derived glaucoma-related damage patterns that are not captured in largely healthy population-based cohorts. We subsequently mapped our trained ML endophenotypes onto the OCT scans from 39,146 participants in the UKBB, a large biobank where genotype data is available. This approach leverages the large number of OAG patients in our clinical cohort while bridging both the paucity of glaucoma cases in UKBB and the absence of genotypes in the clinical cohort, enabling well-powered, disease-specific genetic studies. Our approach is conceptually related to recent unsupervised representation-learning-then-GWAS strategies such as REGLE[31], which learn embeddings from high-dimensional clinical data in a population biobank prior to genetic association. A key distinction is that our representation is trained on a disease-enriched cohort (8323 glaucoma patients) and then transferred to the population biobank, with the aim of biasing the learned features toward disease-relevant structural variation rather than general inter-individual morphology.

We identified 36 and 43 GWAS-specific LD-independent loci with genome-wide significance from the EUR and cross-ancestry meta-analysis, respectively. Importantly, 21 loci have not been previously reported to be associated with glaucoma or well-established glaucoma-related traits. We further explored the functional significance of these previously unidentified loci by performing gene-based association tests, colocalization analysis with expression and splicing quantitative trait loci (eQTL and sQTL), Mendelian randomization, and cell type enrichment analysis. By integrating multiple lines of genetic evidence, we highlight 11 gene associations, including 5 novel ones, for their functional relevance to glaucoma, which can be further investigated as clinically interpretable therapeutic leads. Overall, by uniting disease-focused deep representation learning with cross-ancestry genetics and functional single-cell omics, we validated and extended the putative causal variants and genes for POAG. Beyond ophthalmology, our study establishes a scalable, generalizable framework for ML-driven phenotyping of complex diseases and novel genome-wide discovery.

Results

Study design

We previously developed an ML method to extract quantitative endophenotypes from GCC and RNFL thickness maps derived from macular OCT scans[32]. Briefly, two deep learning models, an unsupervised autoencoder (AUTO) and a self-supervised momentum contrastive learning-based method (MOCO), were used to learn the latent representations of the GCC and the RNFL thickness maps. The latent representations then underwent dimensionality reduction and Gaussian mixture modeling for clustering. This method allows for the identification of distinct retinal layer endophenotypes for subsequent genetic studies. We have previously shown the clinical correlations of these endophenotypes.

The study design of this work is illustrated in Fig. 1. Specifically, we trained the two models described above (MOCO and AUTO) using 18,985 macular OCT scans of 8323 glaucoma patients (99% with OAG) treated at Massachusetts Eye and Ear (MEE). The trained model was then mapped to 86,115 OCT scans from 35,739 European (EUR), 1677 Asian (ASN), and 1730 African (AFR) participants with genotype array data in the UKBB (1074 [2.74%] were diagnosed with glaucoma by ICD code and self-reports). Twenty-one endophenotypes, 14 from the AUTO model and 7 from the MOCO model, were found in UKBB. These are the latent dimensions that took non-zero values in UKBB (the remaining model dimensions from MEE were zero for all UKBB participants and were not analyzed), each rank-based inverse normal transformed prior to GWAS (see "Methods").

The genotype arrays in the UKBB were imputed and underwent quality controls, yielding an average of 8,590,260 common variants (minor allele frequency [MAF] > 0.01) in each population for genetic association tests. Because of the lack of a separate cohort with both macular OCT scans and genotyping data for validation, to identify a high-confidence set of variants, we divided the UKBB dataset from each population into a discovery and replication subset in a 3:1 ratio. Related individuals with high cryptic relatedness (PI_HAT > 0.1875) were excluded prior to splitting (see "Methods"). GWAS was performed on each endophenotype and study subset using a fixed-effects linear regression model, adjusting for age at recruitment, sex, refractive error (measured as spherical equivalent), and top 10 genotype principal components (PCs), separately for each ancestral group, followed by meta-analyses of the discovery and replication subsets. A final cross-ancestry meta-analysis was conducted to enhance statistical power and identify shared genetic risk factors across populations. Post-GWAS functional analytical approaches were then applied to prioritize the genes and potential underlying causal mechanisms of the genetic findings for these endophenotypes. To establish the clinical relevance of these ML-OCT-derived endophenotypes, we tested each of the 21 endophenotypes against glaucoma case status and a panel of glaucoma outcomes ascertained independently of the macular OCT scan, including hospital ICD-coded POAG, fundus-photograph expert grading, glaucoma medication, any glaucoma treatment, and self-reported glaucoma, using logistic regression in UKBB European-ancestry participants and adjusting for age and sex with Benjamini-Hochberg FDR correction within each outcome. Ten of the 21 endophenotypes were FDR-significant for a combined glaucoma phenotype, and fifteen of 20 testable endophenotypes tracked a clinical severity gradient from control to untreated to medicated to laser or surgery. The associations were coherent across the independently ascertained outcomes (Figs. S7 and S8). The endophenotypes also improved cross-validated discrimination of treated glaucoma beyond age and sex and beyond a glaucoma polygenic risk score, indicating that they carry disease-relevant information not contained in known genetic risk[5]. Because none of these outcomes are derived from the macular OCT image, this provides non-circular evidence that the endophenotypes capture clinically meaningful glaucomatous variation. The endophenotypes that do not discriminate case status capture complementary heritable structural axes, consistent with our aim of expanding genetic discovery beyond what binary case-control status captures. Full results are reported in SI File 11.

GWAS and meta-analysis on ML-OCT-derived glaucoma endophenotypes

We performed GWAS of 21 total endophenotypes derived using our two ML methods in the EUR subsets. Throughout, LD-independent loci are defined by PLINK clumping at $r^2 < 0.2$ within ±500 kb windows using

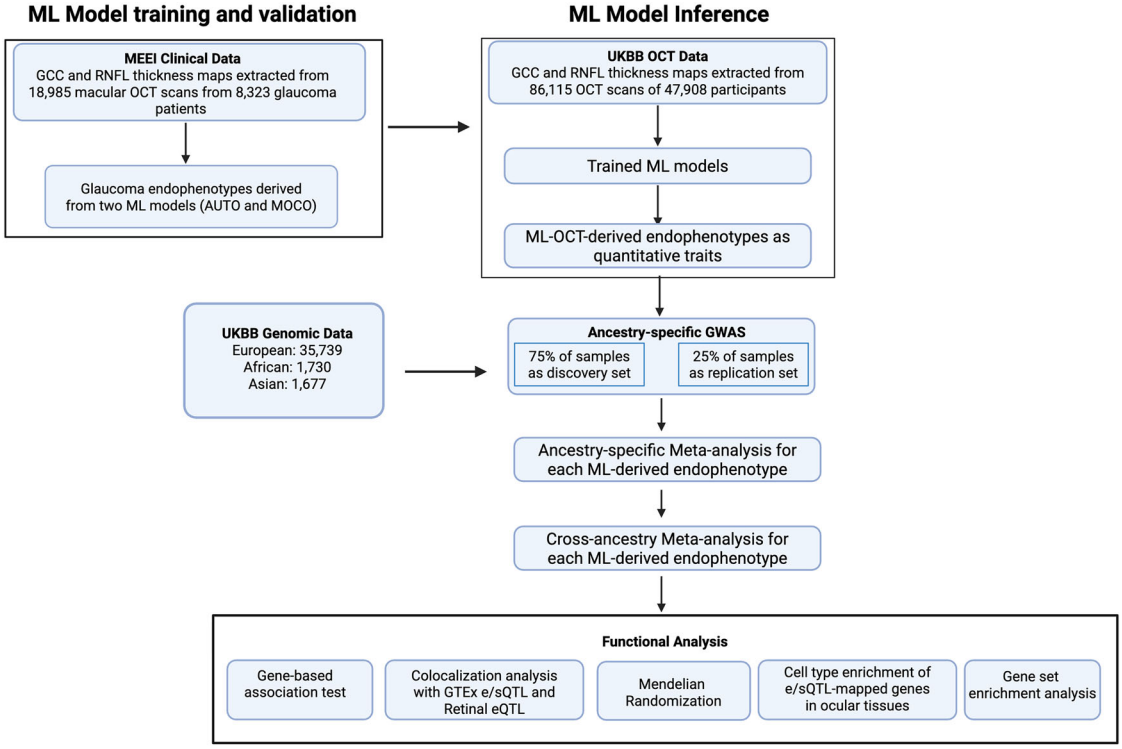


Fig. 1 | Overview of the study design. The flowchart illustrates the main steps of the study. First, ML models were trained on 18,985 OCT scans from 8323 glaucoma patients at MEE to derive glaucoma-related endophenotypes. These trained models were then applied to 86,115 OCT scans from 47,908 UKBB participants to generate quantitative traits for genetic analysis. GWAS were performed separately on participants of European, African, and Asian ancestry using a 75% discovery and 25% replication set. The results were combined through ancestry-specific and subsequent cross-ancestry GWAS meta-analyses for each ML-OCT-derived endophenotype. Finally, a suite of functional analyses, including gene-based tests, colocalization, Mendelian randomization, and cell-type enrichment, was conducted to interpret the genetic findings and prioritize causal genes. The 47,908 participants shown here exceed the 39,146 summed across the three ancestry groups (35,739 EUR, 1677 ASN, 1730 AFR) because participants whose ancestry fell outside these groups, or who were excluded during ancestry-stratified genotype QC, are counted in the total OCT-imaging set but did not contribute to the ancestry-specific GWAS.

ancestry-matched UKBB reference samples. In the EUR discovery GWAS, 637 and 23 unique LD-independent ($r^2 < 0.2$) variants passed suggestive ($P < 1 \times 10^{-5}$) and genome-wide significance ($P < 5 \times 10^{-8}$), respectively (SI File 1). Of these, 63 and 14 unique variants passed nominal significance ($P < 0.05$) and Bonferroni correction in the replication GWAS, respectively. In the EUR meta-analysis of the discovery and replication subsets, 36 GWAS-specific LD-independent variants passed genome-wide significance across the 21 endophenotypes (Fig. 2a, SI File 2). Among them, 17 passed the Bonferroni-corrected genome-wide significance adjusting for the number of endophenotypes ($P < 5 \times 10^{-8}/21$). The effect sizes of these variants show high concordance ($R^2 = 0.93$) between the EUR discovery and replication subsets (Fig. 3a). We note that this 75:25 within-UKBB design is not equivalent to independent external replication and was adopted because, to our knowledge, no separately ascertained cohort with paired macular OCT scans and genotype data of comparable scale was available at the time of analysis. We discuss this limitation in detail in the Discussion.

Due to the small sample sizes of the AFR and ASN groups, only 4 and 3 LD-independent loci were identified, respectively, for these population meta-analyses (SI File 2). Nevertheless, the cross-ancestry meta-analysis identified 43 unique LD-independent genome-wide significant loci across the 21 endophenotypes, with 20 passing the Bonferroni-corrected genome-wide significance ($P < 5 \times 10^{-8}/21$) (Fig. 2b) (SI File 2). Of the 36 unique LD-independent variants identified in the EUR meta-analysis, 49%[18] replicated in the cross-ancestry meta-analysis at genome-wide significance.

Because many genome-wide significant variants reached significance across multiple endophenotypes (Fig. 2), we examined the pairwise Pearson correlation structure of the 21 endophenotypes in UKBB EUR participants (Fig. S5, SI File 12). Hierarchical clustering revealed a clear block structure, in which most AUTO GCC and several RNFL MOCO sectors form a strongly co-varying group (with multiple pairwise $|r| > 0.5$), while a substantial subset of orthogonal latent dimensions shows near-zero pairwise correlation, consistent with the contrastive-learning objective producing independent latent axes. The mean absolute pairwise correlation was 0.32 (median 0.31, maximum $|r| = 0.89$), with 42 of the 210 unique off-diagonal pairs exceeding $|r| > 0.5$. PC analysis of the 21 traits gave an effective number of independent traits of 11[33]. We therefore caution that genome-wide significance shared across correlated endophenotypes should not be read as independent supporting evidence. The 36 EUR and 43 cross-ancestry counts reported above are the union of the LD-independent loci identified across the 21 endophenotype GWAS, with variants shared across endophenotypes counted once. A sensitivity analysis restricted to a decorrelated subset of endophenotypes (greedy pruning at $r^2 < 0.5$, 16 representative traits) recovered all 21 novel loci, of which 12 retained their lead endophenotype and remained genome-wide significant, and the remaining 9 retained a correlated representative endophenotype carrying at least 85% of the peak per-allele effect, confirming that these signals do not arise from correlation among the endophenotypes (SI File 12).

Effect-size heterogeneity at the genome-wide significant loci was not statistically detectable in either meta-analysis. Using Cochran's Q test as the primary measure, 0 of 34 estimable EUR loci and 0 of 43 cross-ancestry loci showed significant heterogeneity after Benjamini-Hochberg or Bonferroni correction, and only one cross-ancestry locus (rs7816434, a previously reported locus) reached nominal Cochran's Q $P = 0.036$. Median I^2 was 0% in both meta-analyses (in EUR, mean $I^2 = 3.1\%$, and in the cross-ancestry meta-analysis, mean $I^2 = 13.3\%$). Because I^2 is upwardly biased and unstable when the number of contributing studies is small ($k = 2$ for the EUR discovery and replication, and $k = 6$ across the ancestry-stratified cross-ancestry meta-analysis), we report Cochran's Q-based heterogeneity tests as

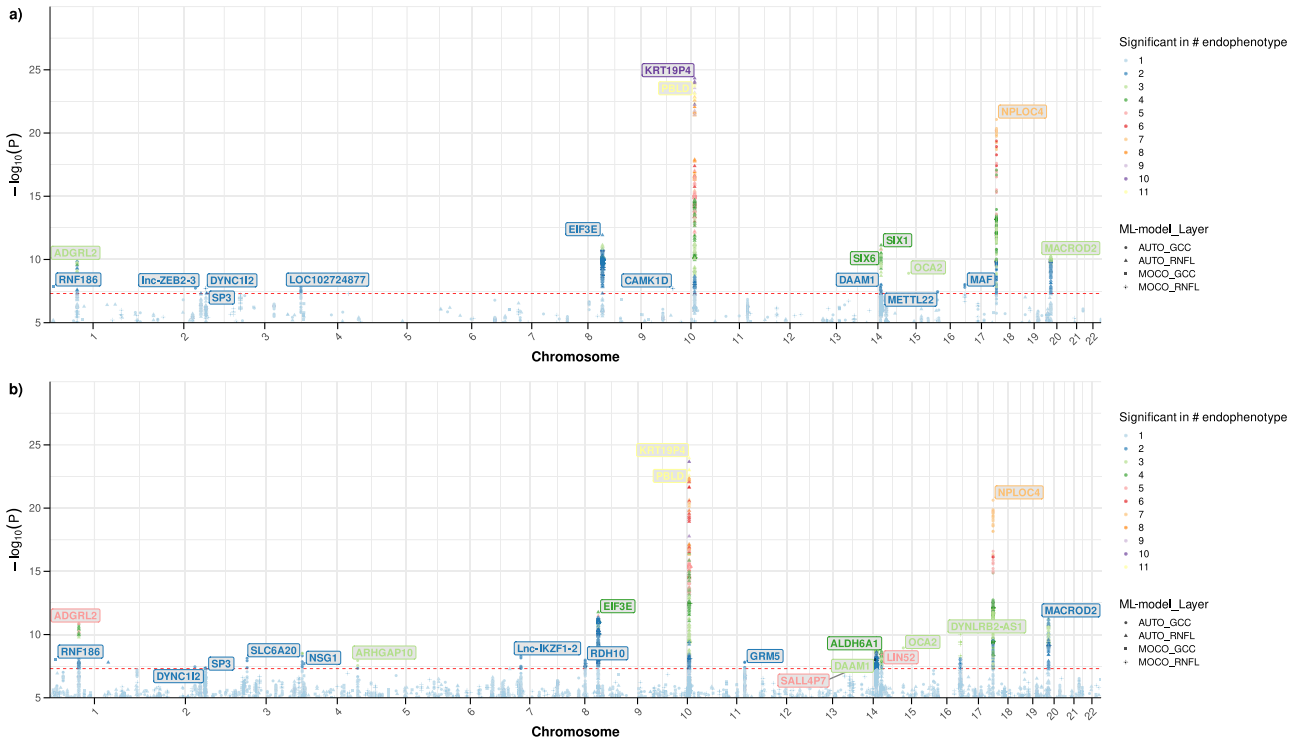


Fig. 2 | Overlayed Manhattan plots of GWAS meta-analyses for 21 ML-OCT-derived endophenotypes. a displays the results from the EUR meta-analysis, while **b** shows the results from the cross-ancestry meta-analysis. In both plots, the *y*-axis represents the -log10(*P*-value) of association, and the *x*-axis shows the chromosomal position. The red dashed line indicates the threshold for genome-wide significance ($P \leq 5 \times 10^{-8}$). Each point represents a genetic variant, with colors indicating the number of endophenotypes for which the variant reached significance. The shape of the points corresponds to the ML model and retinal layer for the most significant association at that locus. Nearest genes of the genome-wide significant loci are labeled.

the primary statistic and treat I^2 descriptively[34,35]. Directional concordance of effect estimates was 100% (EUR) and 90.3% (cross-ancestry). We visualized the consistency of effect sizes across EUR and cross-ancestry meta-analyses using forest plots (Fig. 3b, SI File 3). All loci demonstrated highly consistent effect sizes between EUR and cross-ancestry analyses (Pearson correlation coefficient $r = 0.998$, $P = 1.341 \times 10^{-52}$). Most loci also exhibited consistency within AFR and ASN populations, although these effect estimates had larger standard errors and did not reach statistical significance, as expected due to the smaller sample sizes. None of the GWAS or meta-analyses showed any evidence of population structure inflation (=1.000–1.028) (SI File 4).

Genetic associations for glaucoma were replicated and expanded

We systematically compared the genome-wide significant loci identified in our EUR and cross-ancestry meta-analyses to previously reported associations with POAG and established glaucoma-related traits, including IOP, RNFL thickness and macular thickness. Of the LD-independent genome-wide significant loci, 23 out of 36 (64%) in the EUR meta-analysis and 27 out of 43 (63%) in the cross-ancestry meta-analysis replicated known glaucoma-related genome-wide significant genetic associations (including LD proxies with $r^2 > 0.1$) (Fig. 3c, d) (SI File 5). The novel glaucoma loci found in this study were listed in Table 1. Multiple of them were overlapped between the EUR and the cross-ancestry meta-analyses, yielding 21 unique novel loci. We adopted $r^2 > 0.1$ as a lenient LD-proxy threshold to maximize sensitivity in identifying overlap with previously reported associations, acknowledging that a proxy at $r^2 = 0.1$ tags only a small fraction of the original signal and that "overlap" at this threshold should be interpreted as suggestive rather than definitive replication. In sensitivity analyses using stricter thresholds, 19 EUR loci and 20 cross-ancestry loci overlapped previously reported associations at $r^2 \geq 0.5$, and 15 EUR and 12 cross-ancestry loci overlapped at $r^2 \geq 0.8$, and correspondingly the number of "novel" loci increased to 31 and 40 at these stricter thresholds (SI File 13). The set of high-confidence novel loci that remain at $r^2 \geq 0.8$ is reported alongside the primary list and represents a more conservative interpretation of novelty. We additionally compared our genome-wide significant loci with the 299 lead variants reported by Alipanahi et al.[28] for machine-learning-derived optic nerve head vCDR in UKBB and CLSA. Nine of our genome-wide significant loci overlapped Alipanahi lead variants at $r^2 \geq 0.1$, the majority of which represent shared previously reported retinal-layer-thickness loci. Among our 21 novel-to-glaucoma loci, only one (10:69971810_TAG_T) overlapped an Alipanahi locus at $r^2 \geq 0.1$, and none did at $r^2 \geq 0.8$. We interpret this limited overlap as evidence that our disease-trained macular-OCT representation captures structural variation that is largely distinct from vCDR-based ML phenotyping of the optic disc, consistent with the two approaches being complementary rather than redundant. Locus-level details are reported in SI File 13.

To provide a more direct, disease-level evaluation of the 21 novel loci, we tested their aggregate association with POAG case-control status in an independent cohort, the Mass General Brigham Biobank (MGBB), adjusting for age, age^2, sex, genotyping platform, and ten genotype PCs. Because MGBB and our discovery data draw on overlapping health systems, we first removed the 298 MGBB participants who were also part of the Mass Eye and Ear endophenotype-GWAS cohort (146 of them in the analysis sample, all cases), leaving a genuinely external sample of 2731 cases and 53,479 controls. As the loci were discovered across multiple endophenotypes and may act on disease risk in heterogeneous directions, we used a direction-robust variant-set test (SKAT, with its optimal extension SKAT-O) as the primary analysis[36]. The novel-locus set was significantly associated with POAG status. For the 15 directly genotyped loci, SKAT gave $P = 2.6 \times 10^{-4}$, and SKAT-O gave $P = 4.8 \times 10^{-4}$, and adding two LD-proxy-rescued loci (17 loci) gave SKAT $P = 7.1 \times 10^{-5}$ and SKAT-O 1.4×10^{-4}. The signal was stronger in the European-ancestry subset ($n = 47{,}889$, SKAT $P = 1.3 \times 10^{-5}$), was supported by concordant joint-model evidence (likelihood-ratio

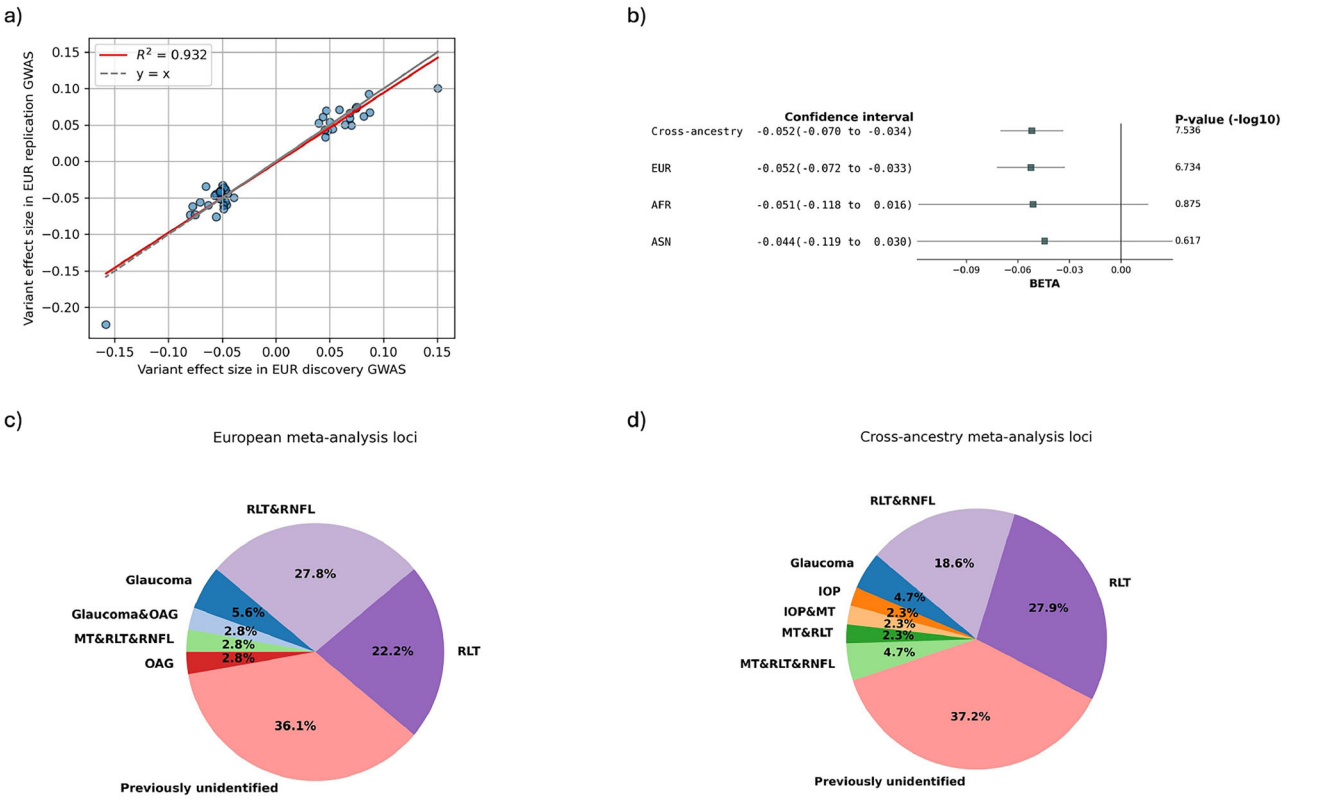


Fig. 3 | ML-OCT-derived endophenotypes replicated and expanded glaucoma genetic associations with consistent effect sizes across cohort subsets and ancestries. a The effect sizes of significant loci in the EUR discovery GWAS stayed highly concordant in the replication GWAS (slope = 0.97, R^2 = 0.93, OLS regression of replication β on discovery β over the 45 EUR-significant trait × locus points). **b** Representative forest plot to show that the effect sizes of genome-wide significant loci were relatively consistent across meta-analyses with different ancestries. The plotted variant is rs35045925 (chr14:59,270,382, hg38, effect allele A, non-effect allele G, $\beta = -0.052$), shown for endophenotype auto_GCC_q10. The large standard errors in AFR and AMR were due to the small sample sizes. **c**, **d** Pie plots showing the proportions of EUR and cross-ancestry meta-analysis genome-wide significant loci that were previously reported in glaucoma, open-angle glaucoma (OAG), intraocular pressure (IOP), retinal nerve fiber layer thickness (RNFL), retinal layer thickness (RLT), and macular thickness (MT) (an upset-plot version is provided as Fig. S6 to more clearly show the overlap structure across categories).

$\chi^2 = 49.2$, $P = 5.5 \times 10^{-5}$), was carried by multiple variants rather than any single locus, and was robust to leave-one-out removal. The signal was marginally stronger after excluding the overlapping participants than in the full sample (17-locus SKAT $P = 1.0 \times 10^{-4}$), confirming that it is not an artifact of discovery-cohort overlap. The SKAT-O data-adaptive parameter selected $\rho = 0$ (a pure variance-component test), indicating that an unsigned, equal-weight burden would be mis-specified for this mixed-direction locus set. Given the modest case count and that four of the 21 loci lacked a usable proxy, we interpret this as an exploratory, disease-level anchor for the novel loci rather than formal external replication, and not as a replication of the endophenotype GWAS itself, since MGBB has no OCT.

Gene-based and gene-set enrichment analyses using MAGMA

We performed gene-based association analysis for each ML-OCT-derived endophenotype using MAGMA and identified 28 and 27 genes that passed the gene-based Bonferroni-corrected threshold ($P < 0.05/18{,}685$ genes tested) for both the EUR and cross-ancestry meta-analyses, respectively. Effect sizes from the MAGMA gene-based test (the gene-level z-score) for these significant genes ranged from 4.56 to 7.40 in the EUR meta-analysis and 4.56 to 7.52 in the cross-ancestry meta-analysis, with corresponding Bonferroni-adjusted P-values reported in SI File 6. Among them, 68%[19] of the genes significant in EUR were also significant in the cross-ancestry meta-analyses. Full gene-based test results are in SI File 6.

Gene-set enrichment highlighted vascular and oxidative-stress processes relevant to glaucoma. Within each endophenotype, we corrected enrichment P-values across all 17,023 tested gene sets using the Benjamini-Hochberg FDR. Although no set reached FDR < 0.05, the sets emerging at the suggestive FDR < 0.2 level centered on biologically relevant processes. Examples include VEGF-receptor/vascular signaling (regulation and negative regulation of VEGFR signaling; $\beta = 0.62$ and 1.06; $q = 0.16$), oxidative-radical handling (regulation of superoxide removal; $\beta = 1.38$; $q = 0.11$), and smooth-muscle tissue development ($\beta = 0.70$; $q = 0.16$). These results are consistent with the vascular and neurodegenerative components of glaucomatous optic-nerve injury. Because FDR < 0.2 is a permissive threshold, we present these enrichments as hypothesis-generating (SI File 10).

Colocalization analysis of GWAS significant loci with *cis* eQTLs and *cis* sQTLs

To further propose putative causal genes that may underlie genome-wide significant loci for these endophenotypes through regulatory effects, we used e/sQTLs in brain, nerve, artery, and fibroblast tissues from the Genotype-Tissue Expression (GTEx) project[37], as well as peripheral retina eQTLs[38]. We applied eCAVIAR, a Bayesian colocalization method, to the genome-wide significant loci from each of the endophenotypes' meta-analyses that had at least one genome-wide significant hit. e/sQTLs from the selected tissues that overlapped each GWAS locus LD interval were included for colocalization.

We found that 72% (26/36) of EUR GWAS loci and 40% (17/43) of cross-ancestry GWAS loci identified in our study showed colocalization evidence with at least one eQTL and/or sQTL (colocalization posterior probability [CLPP] > 0.01). The retinal eQTLs that colocalized at CLPP > 0.01 largely overlapped with colocalizing e/sQTLs identified in GTEx tissues. In total, 25 genes from the EUR meta-analysis and 24 genes from the cross-ancestry meta-analysis showed evidence of colocalization with the

https://doi.org/10.1038/s41525-026-00612-w

Table 1 | Novel loci from EUR and cross-ancestry ML-OCT endophenotype GWAS meta-analyses

Population	Locus	Chromosome	Position (hg38)	Effect allele	Beta	Standard error	*P*	Direction[a]	Subsets tested (*N*)	Concordant (*N*)	Sign-test *P*	Nearest gene
Cross-ancestry	1:161386627_AT_A	1	161416837	A	−0.2308	0.0408	1.60E-08	------	6	6	0.016	ENSG00000283360
Cross-ancestry	rs3767216	1	19813847	A	−0.0437	0.0076	8.66E-09	-+---+	6	4	0.34	*RNF186*
EUR	rs3767216	1	19813847	G	-0.0449	7.90E-03	1.48E-08	--	2	2	0.25	*RNF186*
EUR	2:145315140_GGTGT_G	2	144557573	G	−0.0453	8.00E-03	1.86E-08	--	2	2	0.25	*lnc-ZEB2-3/ LINC01412*
Cross-ancestry	2:145315140_GGTGT_G	2	144557573	G	−0.0426	0.0077	3.56E-08	-+---+	6	4	0.34	ENSG00000235435
Cross-ancestry	rs62172787	2	173960997	A	−0.1375	0.0251	4.59E-08	??--??	2	2	0.25	*SP3*
EUR	rs62172787	2	173960997	G	−0.1375	2.51E-02	4.59E-08	--	2	2	0.25	*SP3*
Cross-ancestry	rs61416912	3	194092883	T	−0.075	0.0127	3.06E-09	----+-	6	5	0.11	NA
Cross-ancestry	rs10933712	3	194093395	A	−0.0712	0.0127	1.92E-08	----+-	6	5	0.11	ENSG00000229964
Cross-ancestry	rs11728835	4	4361027	A	−0.091	0.0162	2.02E-08	------	6	6	0.016	*NSG1*
Cross-ancestry	rs56834952	4	147970941	A	0.0679	0.0119	1.14E-08	++++++	6	6	0.016	*ARHGAP1*
Cross-ancestry	8:109194601_CA_C	8	108182372	CA	−0.0498	0.0085	4.75E-09	------	6	6	0.016	NA
EUR	8:109194601_CA_C	8	108182372	CA	−0.0483	8.80E-03	4.70E-08	--	2	2	0.25	*EIF3E*
EUR	rs4750249	10	12706817	C	0.0507	9.00E-03	1.93E-08	++	2	2	0.25	*CAMK1D*
EUR	rs112115087	14	60341147	G	−0.0512	7.80E-03	4.76E-11	--	2	2	0.25	NA
Cross-ancestry	rs141927261	14	74148741	G	0.0661	0.0112	3.32E-09	++++++	6	6	0.016	NA
EUR	rs72766456	16	8622271	G	−0.0653	1.18E-02	3.50E-08	--	2	2	0.25	*METTL22*
EUR	16:80430159_CG_C	16	80396262	C	0.1751	3.05E-02	9.48E-09	++	2	2	0.25	*MAF*
EUR	rs72855624	17	81562724	A	−0.0781	0.0116	1.61E-11	--	2	2	0.25	*NPLOC4*
Cross-ancestry	rs72855624	17	81562724	A	−0.0732	0.011	2.32E-11	-----+	6	5	0.11	*NPLOC4*
Cross-ancestry	17:79614932_TTAAC_T	17	81647906	T	0.0722	0.0076	2.37E-21	-+++++	6	5	0.11	*NPLOC4*
EUR	17:79614932_TTAAC_T	17	81647906	TTAAC	−0.0621	7.80E-03	1.71E-15	--	2	2	0.25	*NPLOC4*

The table lists the 21 novel, LD-independent loci that reached genome-wide significance ($P \leq 5 \times 10^{-8}$) and have not been previously associated with glaucoma or its related traits. For each lead variant, the table details the population in which it was identified, its chromosomal position (genome build hg38), the effect allele, the effect size (beta) and standard error, the association *P*-value from the corresponding meta-analysis, and the nearest mapped gene.

[a]The direction of effect signs corresponds to the discovery and replication GWAS in the EUR meta-analysis, and discovery and replication for each of the populations in the following order: AFR, EUR, and ASN. The three rightmost columns summarize directional concordance and replace reliance on the per-subset effect-direction string. "Subsets tested (*N*)" is the number of GWAS subsets with a non-missing effect direction (two for EUR loci, namely discovery and replication, and up to six for cross-ancestry loci), "Concordant (*N*)" is the number agreeing with the majority direction, and "Sign-test *P*" is the one-sided exact binomial probability of observing at least that many concordant directions under a null of random sign, so that smaller values indicate stronger concordance. The per-locus test is floor-limited by the number of subsets (minimum attainable $P = 0.25$ for EUR and 0.016 for cross-ancestry loci), and the better-powered aggregate directional concordance is 100% (EUR) and 90.3% (cross-ancestry).

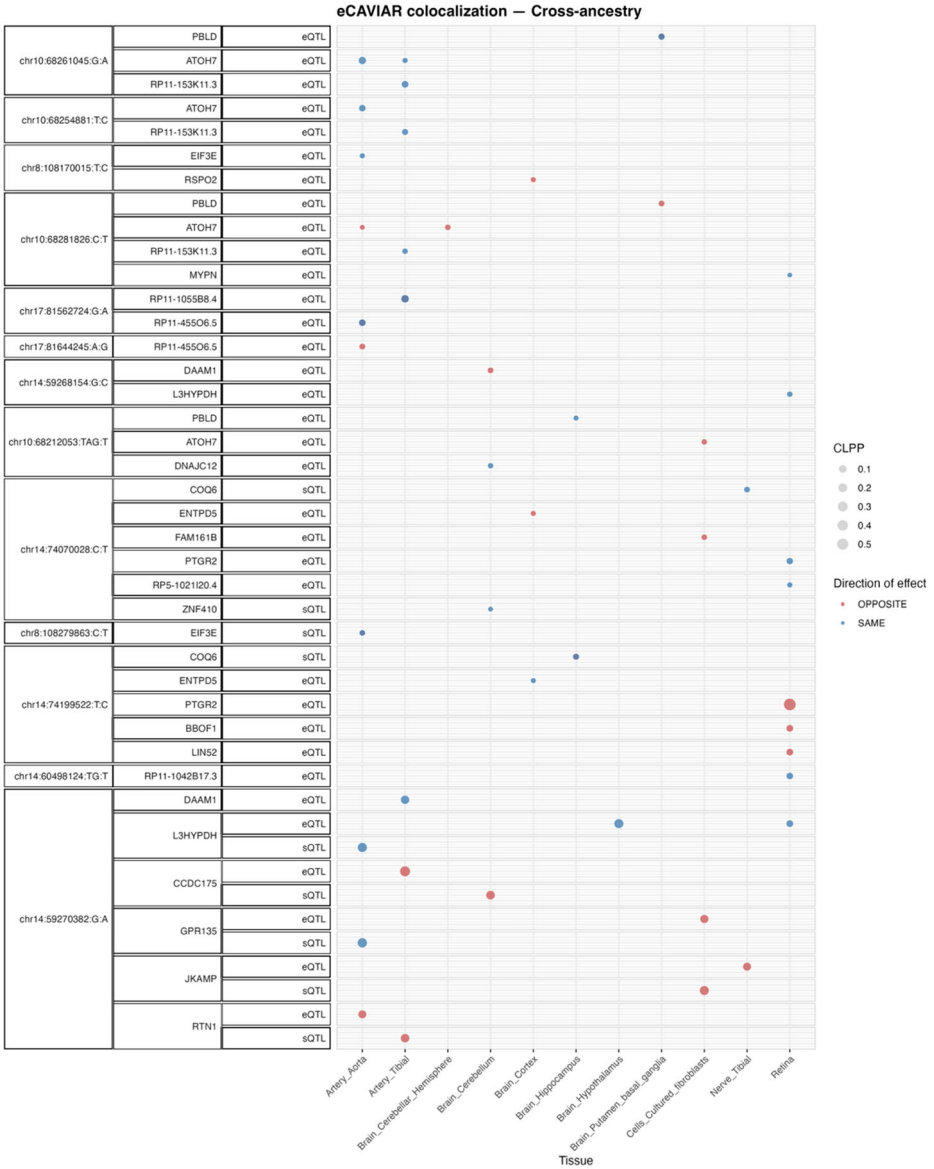


Fig. 4 | Colocalizing e/sQTLs in GTEx tissues and retina with genome-wide significant loci in the cross-ancestry meta-analyses. Shown are genes with at least one colocalization (CLPP > 0.01) between e/sQTLs tested across selected GTEx tissues and peripheral retina and genome-wide significant loci identified in the cross-ancestry meta-analysis. Bubble size represents the maximum colocalization posterior probability among all e/sVariants tested for each gene, QTL type, and tissue combination. Points are color-coded according to the relative direction of effect between the QTL and the GWAS meta-analysis.

identified genome-wide significant loci (SI File 7). Of these, 95% (20/25) and 79% (19/24) of the colocalizing genes in EUR and cross-ancestry meta-analyses were protein-coding, respectively. We note that CLPP > 0.01 is the conventional eCAVIAR screening threshold but corresponds to only a 1% posterior probability of a shared causal variant and should be interpreted as a sensitive but permissive criterion rather than strong evidence of colocalization. To highlight high-confidence cases, 7 EUR lead loci showed substantially stronger evidence at CLPP > 0.5 (six lead variants in the 17q25.3 region all colocalizing with TSPAN10 in Brain Cerebellum at CLPP ≈ 1.0, plus SLC25A16 colocalizing with a peripheral-retina eQTL at CLPP = 0.62), and 1 cross-ancestry lead locus showed CLPP > 0.5 (PTGR2 at rs45569432 with a peripheral-retina eQTL, CLPP = 0.60).

Figures 4 and S1 summarize the e/sGenes that colocalized (CLPP > 0.1) with at least one of the genome-wide significant loci from the cross-ancestry and EUR meta-analysis, respectively. Several e/sGenes, including *EIF3E, PCNX4, PBLD, TSPAN10*, and *DNAJC12*, colocalized with the novel loci in this study. In addition, 36%[10] and 30%[8] of the significant genes from the EUR and cross-ancestry gene-based tests overlapped with the significant e/sGenes identified through the colocalization analysis, respectively.

Cell-type enrichment analysis

To connect the prioritized genes to specific cell types, we applied ECLIPSER[39,40] to test whether the colocalizing e/sGenes at our loci are enriched for cell-type-specific expression in the eye tissues most relevant to glaucoma, namely the aqueous outflow pathways, retina, and optic nerve head (ONH). For the RNFL endophenotypes, mapped genes were enriched in fibroblasts of both the unconventional (ciliary body and iris) and conventional (trabecular meshwork/Schlemm's canal) outflow pathways and in ONH fibroblasts and astrocytes (Table 2, SI File 8, and Figs. 5a, S2 and S3), consistent with the established roles of these cells in glaucomatous tissue remodeling. These outflow-pathway and ONH signals reached only the more permissive tissue-wide FDR < 0.2, whereas the GCC-endophenotype enrichments described below reached tissue-wide FDR < 0.05. Because the genes tested were drawn from the colocalization analysis, these enrichments add cell-type resolution to genes that were already nominated rather than constituting independent discoveries, and we therefore interpret them as supportive evidence.

The recurrent fibroblast enrichments across anterior- and posterior-segment cell types in both meta-analyses were driven by RSPO2 (Fig. 5a, b), which recurred throughout our functional analyses. For the GCC endophenotypes, the strongest enrichment (tissue-wide FDR < 0.05) was in the

Table 2 | Summary of the ocular cell types showing significant enrichment of expression of e/sQTL-mapped genes at endophenotype loci

Tissue	Cross-ancestry, GCC	Cross-ancestry, RNFL	EUR, GCC	EUR, RNFL
Anterior segment (trabecular meshwork/ Schlemm's canal, iris, central cornea, corneoscleral wedge, lens)	Iris_sphincter*, CB_NPCE, goblet, lens_fiber, Uveal_Melanocyte	goblet, CB_NPCE, iris_fibro, ciliary_fibro, K_fibro, iris_RPE, TM_fibro, K_Endo	CB_NPCE, Lens_Fiber, Uveal_Melanocyte, CB_PCE, Iris_PPE, Conj_Melanocyte, Iris_APE, Lens_EarlyFiber, Iris_Sphincter, Pericyte†	CB_NPCE, Iris_Fibro, Ciliary_Fibro, Iris_PPE, TM_Fibro, Goblet
Retina	RPE*, Both_AC, ON_MGC, OFF_MGC	Astrocyte	RPE*, Both_AC, Other_RGC, Cone	
Optic nerve head, optic nerve, and surrounding posterior tissues	Fibroblasts, melanocytes, immune cells, 6_Astro, 8_Astro	Fibroblasts, astrocytes, 6_Astro, 8_Astro	3-Melanocyte*, 14-Melanocyte*	17-Fibro, 28-Fibro, 5-Fibro, 10-Fibro, 22-Fibro, 22-Astro
Macula	RPE*,	H2	RPE*, H2	

An asterisk (*) denotes cell types that passed a tissue-wide threshold of BH FDR < 0.1, while all other listed cell types passed nominal significance ($P < 0.05$). Cell types are listed in order of significance in GCC and RNFL endophenotypes. Cell type descriptions are available in SI File 8.

retinal ganglion cell (RGC) cluster comprising all RGCs other than ON- and OFF-midget cells, driven by DYNC1I2, NPLOC4, and SLC25A12 (Fig. 5c, d). Further GCC signals were seen in ONH melanocytes, retinal and macular RPE, and iris sphincter muscle (FDR 0.004 to 0.042, Fig. 5c). The pigmented-cell enrichments were driven by OCA2, NPLOC4, and TSPAN10. Because OCA2 and NPLOC4 have been associated with myopia[41], a known POAG risk factor, this signal may reflect genetic architecture shared between refractive error and glaucomatous structural change, a hypothesis that will require direct functional follow-up. Together, these analyses localize the prioritized genes to fibroblast populations of the outflow pathway and ONH and to RGC and pigmented/RPE populations of the retina, the cell types most plausibly tied to the mechanical and neurodegenerative components of POAG.

Mendelian randomization of putative causal genes

To assess whether genetically predicted cis-regulated expression of these prioritized genes is associated with our endophenotypes, and thereby provide evidence consistent with a regulatory mechanism at the locus, we applied two-sample Generalized Summary-data-based Mendelian Randomization (GSMR) to the 12 genes that were significant in at least two of three functional analyses applied to the endophenotype GWAS, namely colocalization, gene-based association testing, and cell-type enrichment analysis (Table 3). The exposure was cis-eQTL summary statistics from GTEx and retina (EyeGEx) tissues (independent of the UKBB GWAS used as the outcome), and the outcome was the corresponding endophenotype meta-analysis. The selection of endophenotype for each gene followed the colocalization-best-supported pairing reported in Table 3, and in cases where a gene colocalized with multiple endophenotypes, the most significant pairing was tested. After removing variants flagged by HEIDI as likely horizontally pleiotropic, 11 of 12 genes (92%) showed an association between genetically predicted expression and the corresponding endophenotype at a Bonferroni-corrected threshold ($P < 0.05/12 = 4.17 \times 10^{-3}$). Complete MR results are reported in SI File 9.

Discussion

Our study shows that disease-focused, deep-learning phenotypes extracted from macular OCT can expand the genetic understanding of POAG. Our European and cross-ancestry meta-analyses uncovered 36 and 43 genome-wide significant loci, respectively. Among these loci, roughly 65% were previously reported to be associated with OAG or at least one well-established glaucoma-related trait, such as IOP and retinal layer thickness. The concordance of our GWAS hits with established glaucoma loci suggests that our ML-OCT-derived endophenotypes are disease-specific and able to capture contributions of glaucomatous structural damage. More importantly, however, we identified 21 novel genome-wide significant loci, where several of the mapped genes are highly expressed in relevant eye tissues, such as ONH and retina. The biological relevance of the identified loci is further supported by multiple lines of evidence.

Using systematic functional analyses, we advance from association to higher-resolution gene and cell-type inference, providing a basis for mechanistic hypotheses that can be followed up for POAG. Bayesian colocalization with multiple GTEx tissues that contain relevant cell types plus retinal eQTLs linked 72% of European association signals to likely regulatory gene/s in tissues with eye-relevant cell types. MAGMA gene-based tests uncovered 28 significant genes from the European meta-analysis, 68% of which overlapped with those from the cross-ancestry analysis. Importantly, 36% of MAGMA hits in the EUR cohort overlapped with colocalizing e/sGenes, providing further statistical support. Cell-type specific enrichment analysis connected these genes to the cellular architecture of the eye, where pigmented cells in the retina and ONH, ONH and outflow pathway fibroblasts, and ONH astrocytes emerged as the most over-represented populations, echoing the cell-types produced by Hamel et al.[42]. The enrichment in outflow pathway cells suggested that elevated IOP also plays a role in affecting glaucomatous damage in the retina. Not previously observed with POAG GWAS loci was enrichment in non-midget

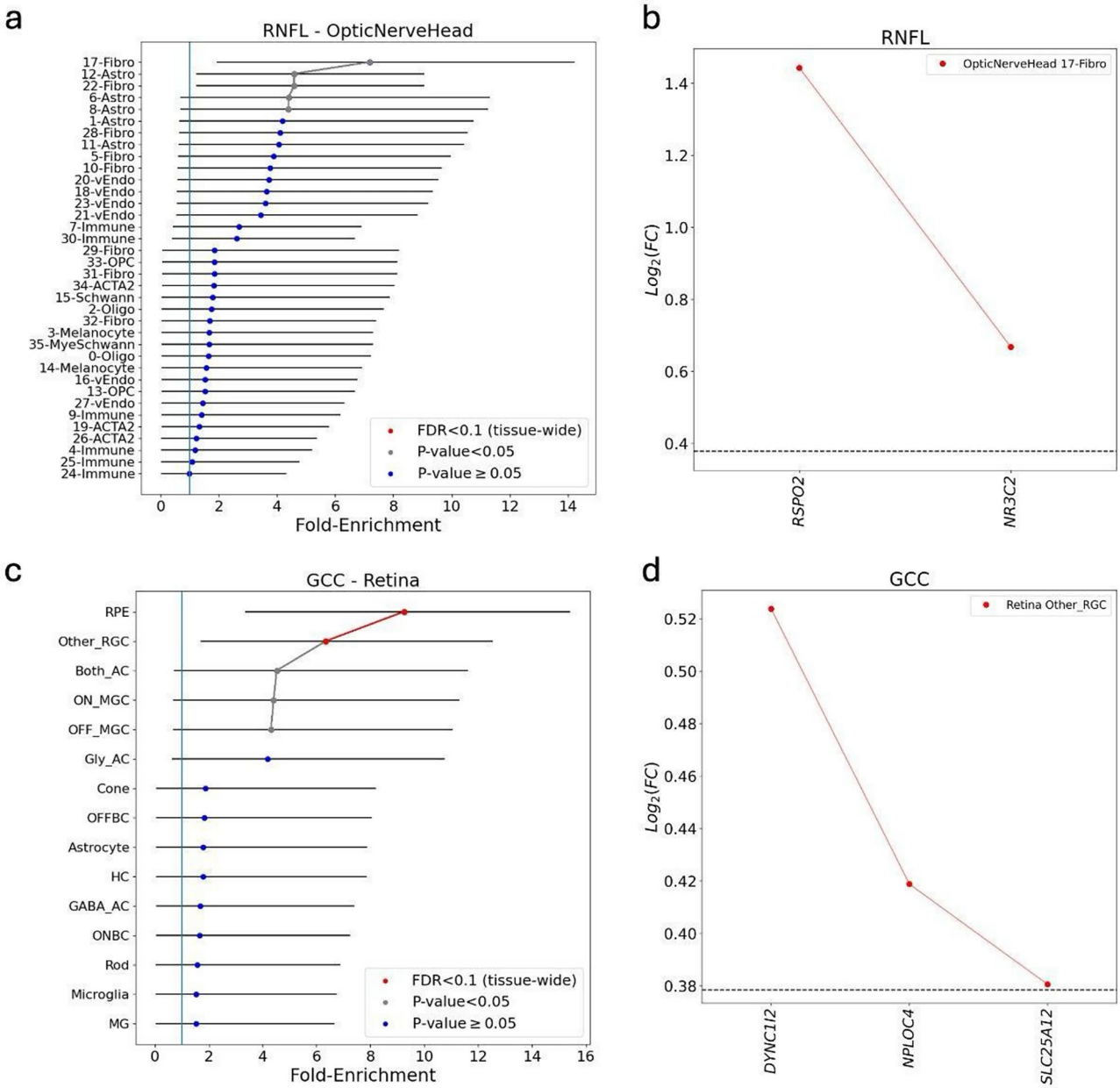


Fig. 5 | Cell type enrichment analysis of the cross-ancestry meta-analysis in the optic nerve head and retina. a, **c** The fold-enrichment of cell type-specific expression for genes mapped from the cross-ancestry GWAS meta-analysis loci in the optic nerve head for RNFL-derived endophenotypes and retina for GCC-derived endophenotypes, respectively. Cell types are ranked by their enrichment score (*x*-axis). Points are colored by statistical significance, with red indicating a tissue-wide significance level of FDR < 0.2 and gray indicating nominal significance ($P < 0.05$). **b**, **d** The specific genes driving the most significant enrichment signals from the corresponding plots on the left. **b** shows that RSPO2 is the primary driver of the enrichment in optic nerve head fibroblasts. **d** shows that DYNC1I2, NPLOC4, and SLC25A12 collectively drive the enrichment in RGC clusters. The horizontal dashed line represents log2(Fold-change) of 0.375 (FC = 1.3) and FDR < 0.2 used as the cell type-specificity enrichment cutoff. FC, fold-change of gene expression in the given cell type relative to all other cells in the given tissue.

ganglion cell RGCs driven by *NPLOC4, SLC25A12*, and *DYNC1I2*. Two-sample Mendelian randomization using GTEx and retina cis-eQTLs provided evidence consistent with a regulatory effect of these genes on the corresponding endophenotypes. This layered approach prioritized 11 putative causal genes (*EIF3E, LIN52, NPLOC4, PBLD, RSPO2, SLC25A16, TSPAN10, DYNC1I2, SLC25A12, ATOH7*, and *C14orf39*). Each of these genes was supported by more than one line of analysis, having been significant in at least two of the three functional tests (gene-based association, colocalization, and cell-type enrichment) before we applied Mendelian randomization. The strongest colocalizations fall in disease-relevant tissue, with *SLC25A16* and the cross-ancestry locus *PTGR2* colocalizing with peripheral-retina eQTLs (CLPP 0.62 and 0.60), and *TSPAN10* has previously been linked to ocular structural traits. The cell-type enrichments recurrently implicate the same outflow-pathway, RPE, and retinal ganglion cell populations across endophenotypes, adding cellular resolution to genes that were already implicated by colocalization. Among these cell-type signals, only the GCC pigment/RPE and RGC enrichments reached tissue-wide FDR < 0.05, whereas the outflow-pathway and ONH fibroblast/astrocyte enrichments were suggestive (FDR < 0.2).

Among the prioritized genes, some are known to be associated with POAG-related traits, while others are novel. Specifically, *ATOH7* and *C14orf39* are genes that have been previously proposed to potentially

Table 3 | Summary of functional evidence for prioritized genes

Gene	Locus	Gene-based test	Colocalization	Cell-type enrichment analysis	MR
ATOH7	chr10:68281826:C:T	+	+	−	+
EIF3E	chr8:108279863:C:T	+	+	−	+
LIN52	chr14:74199522:T:C	+	+	−	+
MYPN	chr10:68281826:C:T	+	+	−	−
NPLOC4	chr14:74199522:T:C	+	+	+	+
PBLD	chr10:68281826:C:T	+	+	−	+
RSPO2	chr8:108170015:T:C	+	+	+	+
SLC25A16	chr10:68263296:A:G	+	+	+	+
TSPAN10	chr17:81644245:A:G	+	+	+	+
C14orf39	chr14:60490453:G:A	+	−	+	+
SLC25A12	chr2:171739103:T:TG	−	+	+	+
DYNC1I2	chr2:171739103:T:TG	+	−	+	+

The table highlights genes that were statistically significant in at least two of the primary functional analyses performed. A "+" indicates that a gene passed the significance threshold for the corresponding analysis. A "−" indicates the gene did not meet the significance threshold for that analysis.
MR Mendelian randomization.

contribute to POAG risk[5,12]. *SLC25A16* and *PBLD* are genes near *ATOH7* and have been reported to be associated with VDCR. *TSPAN10* and *NPLOC4* have been reported to be associated with ocular traits like macular thickness and optic disc parameters. Importantly, five genes (*RSPO2, EIF3E, LIN52, DYNC1I2, SLC25A12*) that showed strong supportive causal evidence in our studies have not been reported previously. For example, *EIF3E*, a gene from one of our novel loci, was highly significant in the gene-based association test ($P = 4.52 \times 10^{-13}$), showed a colocalization signal with multiple brain and artery tissues, and had elevated expression across several outflow pathway, ONH, and retinal cell types (Fig. S4). As another example, *RSPO2* was significant in gene-based association tests ($P = 2.13 \times 10^{-9}$) and colocalized with eQTLs in brain cortex. Single-cell expression in ocular cell types showed elevated *RSPO2* expression in the fibroblasts of ONH and multiple cell types in the anterior segment, with the highest expression levels in fibroblasts in the trabecular meshwork, iris, and ciliary body. Cell-type enrichment analysis further supported *RSPO2* as the main driver of fibroblast enrichment in these regions.

The genes and ocular cell types identified in this study point to two connected biologic mechanisms of POAG risk. RSPO2 is a potentiator of canonical Wnt signaling[43]. Its high expression in anterior-segment and ONH fibroblasts suggests a role in extracellular matrix remodeling and aqueous outflow regulation that potentially leads to elevated IOP. On the neuronal side, EIF3E, an initiator for protein synthesis and DNA damage repair[44], shows strong gene-based and colocalization support with broad retinal expression, implicating a potential role in RGC resilience. *DYNC1I2* underlies enrichment in an RGC subtype cluster and encodes a dynein-1 intermediate chain, and dynein-dependent axonal transport is known to fail early in experimental glaucoma[45], potentially linking this locus to RGC transport vulnerability. *SLC25A12*, a mitochondrial aspartate-glutamate carrier, contributes to the same RGC enrichment and supports neuronal energy metabolism. These data support a model where Wnt-tuned fibroblast programs in the ONH and outflow pathway modulate IOP and biomechanical stress, which then interface with RGC-intrinsic capacities for protein synthesis/repair, axonal transport, and mitochondrial bioenergetics to determine susceptibility to glaucomatous injury.

While these novel genes do not have established therapeutic agents, publicly available data suggest they represent candidate targets that merit further evaluation for future drug discovery programs[46]. For instance, *RSPO2, EIF3E,* and *DYNC1I2* possess known protein-ligand structures that could inform small molecule development. Specifically, the existence of a co-crystal structure of RSPO2 with a small molecule ligand offers a structural starting point for hypothesis-driven structure-based drug design but does not by itself validate RSPO2 as a therapeutic target. Furthermore, as RSPO2 is a secreted potentiator of canonical Wnt signaling highly expressed in anterior-segment fibroblasts, a neutralizing antibody may represent a plausible therapeutic strategy worth exploring. However, demonstrating feasibility would require substantial target validation, safety, tissue-delivery, and efficacy studies that are outside the scope of the present work. These results demonstrated that our genetic findings contribute to the understanding of POAG biology and propose hypothesis-generating starting points for functional validations with patient-derived cells and animal models.

This work has several limitations. Our ML models were trained on patients identified by POAG-related ICD codes, and although the size of the training set should largely offset the resulting label noise, chart-confirmed cases might yield more specific endophenotypes. Because no external cohort with both macular OCT and genotype data of comparable scale was available, replication relied on a 75:25 split of the UK Biobank, in which the discovery and replication subsets share genotyping platform, recruitment and selection biases, imaging protocol, and population structure. The high effect-size concordance ($R^2 = 0.93$) is therefore expected under this design and does not substitute for replication in a separately ascertained cohort. The aggregate association of the 21 novel loci with POAG status in MGBB provides an orthogonal, disease-level anchor, but it is not a replication of the endophenotype GWAS, as the MGBB lacks OCT. Re-deriving the ML-OCT endophenotypes and their genetic associations in an independent cohort with paired OCT and genotypes remains an essential next step before the prioritized loci and genes can be considered firmly established. Residual population substructure may persist despite adjustment for ten genotype PCs, and the European-centric UK Biobank limits generalizability, so the smaller but directionally consistent effects in our African and Asian groups will need confirmation in larger non-European OCT cohorts. Several interpretive caveats also apply. Our Mendelian randomization tests the effect of cis-regulated expression on the endophenotype rather than on POAG itself and thus supports a regulatory effect at the locus rather than a causal role in disease. The cell-type enrichments are largely suggestive, with most reaching only FDR < 0.2, and remain hypothesis-generating until confirmed in independent single-cell data. Finally, the druggability discussion is intended to generate hypotheses rather than to establish any prioritized gene as a validated therapeutic target.

In conclusion, by integrating disease-trained ML-OCT-derived endophenotypes with systematic functional genomics, we have successfully expanded the number of glaucoma-associated loci, linked them to plausible causal genes, and mapped them to specific eye cell types. Beyond

advancing our understanding of glaucoma's genetic mechanisms, this study demonstrates the effectiveness of enhancing disease phenotyping by transfer learning from a clinical dataset to a population cohort, a methodology applicable to the investigation of other complex diseases.

Methods

MEE cohort definition

This study was conducted in accordance with the tenets of the Declaration of Helsinki. The use of OCT scans acquired at MEE was approved by the Harvard Longwood Campus (HLC) IRB (protocol 2024P003507), with a waiver of informed consent granted owing to the retrospective use of de-identified data.

OCT scans from glaucoma patients seen in MEE between October 1, 2020 and June 1, 2023 were used for model training. Glaucoma Individuals were defined based on the International Classification of Diseases (ICD) code in the electronic health record (i.e., ICD9: 365.x, ICD10: H40.x). Specifically, 99% (8207) of these patients contain at least one of the OAG ICD codes. The other 1% are patients with other types of glaucoma, including angle-closure glaucoma and secondary glaucoma.

Quality control of UKBB genotype data

Access to UKBB OCT and genetic data was obtained with approval from the UKBB Committee. All participants in UKBB provided informed consent at recruitment.

Genotype array quality control steps were applied with PLINK v1.9 and PLINK v2.0 (https://www.cog-genomics.org/plink/) on directly genotyped and imputed variants of 488,395 UKBB samples. The hard call sample quality control (QC) was conducted to remove withdrawn samples ($N = 107$), to exclude samples that are not in impute data ($N = 969$), to exclude samples with no sample QC data ($N = 112$), to remove ambiguous sex ($N = 1930$), to remove heterozygosity outliers by using 5 standard deviations from the mean heterozygosity rate within each inferred ancestry subpopulation ($N = 978$) and to remove related individuals with high cryptic relatedness (>0.1875 Pi-hat) ($N = 35{,}113$). To infer ancestry, Principal Component Analysis (PCA) was first applied to linkage disequilibrium (LD)-pruned ($r^2 < 0.2$ in 200 kb windows) genetic markers with MAF > 5%. The SHARDS algorithm selected the first 5 important PCs and then implemented k-nearest neighbors classification to predict the ancestral background of participants using labels from the 1000 Genomes Project Phase 3 reference panel. Participants were then categorized based on primary genetic ancestry. In terms of the relative removal algorithm, graph theory was also applied to keep the maximal independent set. The application of the maximal independent set contributed to keeping as many samples as possible. Concurrently, another relative removal algorithm was also conducted to keep the highest ratio of the number of cases versus the number of controls to improve the power.

Regarding the imputed variant QC, the pipeline iteratively examines variants, INFO scores, allele frequencies, and tests of Hardy-Weinberg equilibrium. More specifically, the missingness rate per sample using posterior probability > 0.9 for genotype calling was inspected. A small missingness rate of autosomes was observed, and no significant difference was observed between males and females. Monomorphic sites were removed if only one allele occurs at a site or locus in the population. We further filtered out the variants with indel length > 50 ($N = 2312$), and subsequently, the variants with INFO score < 0.6 ($N = 45{,}927{,}754$) were excluded. The variants with allele frequency < 0.1% for each ancestry subpopulation ($N = 35{,}850{,}305$ in EUR, $N = 28{,}792{,}905$ in AMR, $N = 24{,}954{,}341$ in AFR, $N = 35{,}437{,}737$ in SAS, and $N = 37{,}236{,}001$ in EAS) were further excluded. Then the Hardy-Weinberg equilibrium was calculated and inspected for each ancestry subpopulation. The mid p-value was utilized in exact tests for Hardy-Weinberg equilibrium analyses. Therefore, a total of 227,173 variants in EUR, 98 variants in AMR, 32,317 variants in AFR, 38,173 variants in SAS, and 3711 variants in EAS of autosomes failed Hardy-Weinberg and were removed. To be conservative, a variant was kept if it passed the Hardy–Weinberg test for all ancestry subpopulations. Another MAF filter was applied to keep the common variants with MAF > 1% for further analyses. After QC, the number of variants in EUR, AFR, and ASN are 9,329,555, 8,477,280, and 7,963,945, respectively.

GWAS and meta-analysis

Macular GCC and RNFL endophenotypes were derived from 18,985 OCT scans of 8323 glaucoma patients treated at MEE. These endophenotypes were mapped to OCT scans from the UK Biobank, including 35,739 participants of European ancestry, 1730 of African ancestry, and 1677 of Asian ancestry. For patients with scans from both eyes and multiple visits, we prioritized the right eye. If no scans were available for the right eye, the left eye was used instead. When multiple scans existed for the same eye, the most recent scan was used as the endophenotype.

GWAS were conducted on a total of 21 endophenotypes, including 7 AUTO GCC phenotypes, 2 MOCO GCC phenotypes, 7 AUTO RNFL phenotypes, and 5 MOCO RNFL phenotypes. Throughout this work, each endophenotype is named by the convention {model}_{layer}_q{index} (e.g., auto_GCC_q10, moco_RNFL_q3), where the model is AUTO (unsupervised autoencoder) or MOCO (momentum contrastive learning), the layer is GCC or RNFL, and q{index} denotes the latent dimension or spatial sector index returned by the trained model. When the MEE-trained AUTO and MOCO models were applied to the UKBB macular OCT scans, 21 latent dimensions (14 from the AUTO model and 7 from the MOCO model) took non-zero values across UKBB participants and were retained as endophenotypes. The remaining model dimensions were zero for every UKBB participant, corresponding to MEE-derived cluster patterns not represented in the UKBB population, and were therefore uninformative and excluded from GWAS. No further selection or filtering was applied among the 21 retained dimensions. Because the raw endophenotype values are not normally distributed, each endophenotype was transformed by a rank-based inverse normal transformation prior to GWAS. Observations were ranked, the ranks were converted to quantiles using the empirical cumulative distribution function, and the quantiles were mapped to the standard normal distribution using the inverse normal (probit) function, which reduces the influence of outliers and improves conformity to the normality assumptions commonly used in GWAS[47]. GWAS was performed separately for each ancestry group (AFR, Asian, and EUR). Within each ancestry, samples were randomly divided into discovery (75%) and replication (25%) sets, with matched age distributions and female-to-male proportions across the two subsets. Related individuals (PI_HAT > 0.1875) were already excluded during genotype QC, so the discovery and replication split is over the unrelated set. Analyses were performed using PLINK v2.0, adjusting for age, sex, the first 10 PCs, and refractive error. PCs were calculated separately within each ancestry group. For meta-analysis, we used METAL (version 2011-03-25), applying a fixed-effects model based on inverse-variance weighting, under the assumption that all samples came from the same biobank. Heterogeneity statistics were computed, and beta coefficients and standard errors were used as inputs. Meta-analyses were first conducted within each ancestry, followed by a trans-ancestry meta-analysis.

LD clumping was performed with PLINK v2.0 to identify independent loci in each meta-analysis. LD calculations were based on population-specific samples in the UKBB. For genome-wide significant signals, the LD threshold r^2 was set to 0.2, and the LD radius was set to be 500 kb for LD clumping.

None of the GWAS or meta-analyses showed any evidence of population structure inflation, with the genomic lambda values around 1. Using Cochran's Q test as the primary measure of heterogeneity, no EUR locus and no cross-ancestry locus showed significant heterogeneity after Benjamini-Hochberg or Bonferroni correction, with only one cross-ancestry locus (rs7816434) reaching nominal Cochran's Q $P = 0.036$; median I^2 was 0% in both meta-analyses. Because I^2 is upwardly biased and unstable when the number of contributing studies is small, we report Cochran's Q-based tests as the primary heterogeneity statistic and treat I^2 descriptively. We note, however, that absence of evidence is not evidence of absence, and therefore

report observed I^2 values alongside heterogeneity p-values for every locus in SI File 2.

Comparison with previously reported associations

We compared our identified independent genome-wide significant loci to genome-wide significant variants previously reported to be associated with glaucoma and glaucoma-related traits, which are retinal layer thickness, IOP, and CDR. The set of variants used for comparison was downloaded from GWAS Catalog (MONDO_0005041, EFO_0004190, OBA_2050110, EFO_0004695) and from three POAG GWAS studies[5,12,15]. To ensure that our classification of novel vs previously reported loci is LD-aware, we included LD proxies ($r^2 > 0.1$, UKBB as reference panel) of the identified loci in this comparison. We adopted $r^2 > 0.1$ to maximize sensitivity in detecting overlap with previously reported associations and to be conservative in claiming novelty. We additionally performed sensitivity analyses at $r^2 \geq 0.5$ and $r^2 \geq 0.8$ (SI File 13).

Colocalization analysis with expression and splicing QTL datasets

We conducted a Bayesian colocalization analysis using eCAVIAR (https://github.com/fhormoz/caviar) to pinpoint candidate genes and regulatory elements (e/sQTLs) potentially involved in how glaucoma-linked genetic variations exert their effects. This method estimates the probability that overlapping GWAS and e/sQTL signals share the same causal variant or haplotype. This method considers local LD and the presence of multiple causal variants (allelic heterogeneity).

We used cis-expression quantitative trait loci (cis-eQTLs) and cis-splicing quantitative trait loci (cis-sQTLs) data from multiple tissues in the GTEx project (v8 release)[48] and peripheral retina eQTLs from the EyeGEx dataset[38]. The GTEx data were based on GENCODE v26 annotations, while EyeGEx data used GENCODE v25. Significant e/sQTLs (FDR < 0.05), gene expression levels, and intron excision ratios (LeafCutter values) were downloaded from the GTEx portal (https://gtexportal.org/home/datasets) for GTEx and EyeGEx. These summary statistics for all tested variant-gene pairs were then used for colocalization analysis.

Colocalization was performed on the genome-wide significant LD-independent loci for each GWAS meta-analysis. Every GWAS locus was assessed for colocalization with all eQTLs and sQTLs containing at least five e/sVariants (FDR < 0.05) located within the corresponding GWAS LD interval. For each locus, LD windows were defined by selecting variants in LD ($r^2 > 0.1$) with the lead GWAS variant, extending 50 kb upstream and downstream. For loci derived from European populations, LD calculations were based on European samples from the 1000 Genomes Project[49], while for the cross-ancestry loci, LD was computed using the European, African, and Asian samples in 1000 Genomes, using PLINK v1.9. When the lead variant was missing from the 1000 Genomes dataset, a high LD proxy ($r^2 > 0.8$) or the nearest variant was used. Z-scores for GWAS and e/sQTL summary statistics were computed as the effect size divided by its standard error for each variant. For retina-specific eQTLs, Z-scores were derived from association p-values, assuming a chi-square distribution with 1 degree of freedom. We tested all variants with MAF > 1% that were common to the GWAS and e/sQTL datasets. Variants in GWAS datasets were aligned to the alternative allele (ALT) reported in GTEx and EyeGEx datasets. Only eQTLs were analyzed for retina-specific data, while both eQTLs and sQTLs were analyzed for GTEx tissues. A locus was considered to show colocalization evidence when the colocalization posterior probability (CLPP) exceeded 0.01, following eCAVIAR's recommendations. We note that CLPP > 0.01 is a sensitive but permissive screening threshold (a 1% posterior probability of a shared causal variant) rather than strong evidence of colocalization. To reduce false-positive findings, we excluded any variant–gene–tissue–trait combinations in which the colocalized e/sVariant had a GWAS p-value > 0.05, an e/sQTL p-value > 1×10^{-4}, or failed to meet the FDR < 0.05 criterion (i.e., labeled "FALSE" in the column "QC_name_QTL_FDR05_P1E04_GWAS_P5E02" of SI File 7).

Gene-based tests and gene set enrichment analyses

We conducted gene-based association tests and gene set enrichment analyses with MAGMA (v1.10, https://ctg.cncr.nl/software/magma). In this gene-based analysis, genetic variants were mapped to 18,685 genes based on distance (±100 kb) to compute a gene-level z-score, which was then tested for association with the ML-derived endophenotypes, adjusting for confounding factors. The derived gene-based P-values were adjusted using the Bonferroni method to account for multiple testing ($P < 0.05/18{,}685 = 2.68 \times 10^{-6}$). In the pathway analysis, a list of 17,023 curated gene sets from the MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/) human database was used, and the P-values of pathway analysis were adjusted by applying a Benjamini-Hochberg FDR across all sets, with FDR < 0.05 as significant and FDR 0.05–0.2 as suggestive signals.

Mendelian randomization

For each gene identified by our colocalization or our gene-based analysis, we performed summary-data Mendelian randomization using GSMR (https://yanglab.westlake.edu.cn/software/gsmr/) to test for a putative causal effect of cis-regulated expression on the endophenotypes. The causal effect of interest in this MR analysis is the effect of genetically predicted cis-regulated gene expression on the corresponding ML-derived endophenotype, and we do not interpret these estimates as the causal effect of expression on POAG clinical status. The selection of endophenotype tested for each gene followed the colocalization-best-supported pairing (Table 3 and SI File 7), and where a gene colocalized with multiple endophenotypes, the most significant pairing was tested. The exposure (eQTL) summary statistics were derived from GTEx v8 and the EyeGEx peripheral retina dataset, which are independent of the UKBB GWAS used as the outcome, mitigating the bias that would arise from sharing samples between exposure and outcome. For each target gene, we first extracted independent cis-eQTL instruments (defined as SNPs within ±1 Mb of the transcription start site with $P < 5 \times 10^{-8}$) from the GTEx and retina eQTL summary statistics. Instruments were LD-pruned at $r^2 < 0.05$ in 1 Mb windows using a European-ancestry reference panel ($n = 5000$ unrelated UK Biobank participants). To guard against horizontal pleiotropy, we applied the HEIDI-outlier test implemented in GSMR with a PHEIDI threshold of 0.01, removing SNPs inconsistent with a single causal pathway. GSMR then estimated the causal effect of gene expression on the endophenotype with a generalized least-squares framework that accounts for sampling variance and residual LD among instruments. Effect estimates and their standard errors were computed under the default GSMR settings. Multiple-testing correction used the Bonferroni method based on the 12 genes carried into MR ($P < 0.05/12 = 4.17 \times 10^{-3}$). HEIDI only partially addresses horizontal pleiotropy, and undetected pleiotropy remains a residual risk that we acknowledge as a limitation.

External aggregate validation of the novel loci. To assess whether the 21 novel ML-OCT loci are associated with clinical glaucoma, we tested their aggregate association with POAG case-control status in MGBB (imputed GRCh38 hard-call genotypes), adjusting for age, age^2, sex, genotyping platform, and ten genotype PCs. Because MGBB and the discovery cohort draw on overlapping health systems, we identified the 298 MGBB participants who were also part of the Mass Eye and Ear endophenotype-GWAS cohort (via the Biobank-to-MEE crosswalk) and removed them for the primary, genuinely external analysis (2731 cases and 53,479 controls, of which 146 of the removed participants were in the analysis sample and all were cases). The full sample (2877 cases and 53,479 controls) was retained as a sensitivity analysis. For each locus, a single sentinel was selected (the cross-ancestry lead where available, otherwise the EUR lead) and matched to MGBB by position and allele set on the forward GRCh38 strand. Fifteen sentinels were directly present in MGBB. Of the six absent after imputation quality control (Rsq ≥ 0.6, MAF ≥ 0.01), two were rescued using 1000 Genomes European-founder LD proxies ($r^2 \geq 0.8$ with documented allele phasing) and four had no usable proxy. Because the loci act in mixed directions, the primary test was a direction-robust sequence-kernel association test (SKAT, equal weights) with its optimal extension SKAT-O, complemented by a joint logistic likelihood-ratio test against the covariate-

only model and by leave-one-out and European-ancestry-only sensitivity analyses. The SKAT-O data-adaptive parameter selected $\rho = 0$ (a pure variance-component signal), confirming that an unsigned, equal-weight burden test would be mis-specified for this locus set. Analyses used PLINK v2.0 and R (SKAT v2.2.5).

Ocular single-cell expression

Publicly available single-cell RNA-seq datasets from the posterior segment, anterior segment, and retina were used to plot the single-cell expression patterns of our prioritized genes across a variety of ocular cell types[50–52].

Cell type-specific enrichment

To identify ocular cell types that are enriched for cell type-specific expression of genes that mapped to GWAS loci based on colocalizing e/sQTLs, we used a previously described pipeline, ECLIPSER. Briefly, ECLIPSER operates by determining if genes mapped to GWAS loci for a specific complex disease or trait based on e/sQTLs or other functional genomic data show greater cell type-specific expression compared to genes linked to a background set of GWAS loci from thousands of unrelated traits. The fundamental principle of ECLIPSER is that in a relevant tissue, multiple disease-associated genes will exhibit elevated expression in pathogenic cell types relative to non-pathogenic cells, to a greater extent than genes associated with unrelated traits. Enrichment was assessed by tissue-wide Benjamini-Hochberg FDR levels computed across the cell types within each tissue. Cell types with a fold-change > 1.3 (log2 fold-change > 0.375) were evaluated. We report tissue-wide FDR < 0.05 as significant and FDR 0.05–0.2 as suggestive.

The cell type-specific expression data we used were previously described in Hamel et al.[42]. It is single-nucleus RNA-seq of 13 tissues dissected from non-diseased human eyes, including central cornea, corneoscleral wedge (CSW), trabecular meshwork (TM) including Schlemm's canal, iris, ciliary body (CB), lens (all from anterior segment[50]), peripheral and macular retina[50,53], the optic nerve head (ONH), optic nerve (ON), peripapillary sclera (PPS), peripheral sclera, and choroid (all from posterior segment[51]). GWAS-specific genome-wide significant loci from the EUR and cross-ancestry meta-analysis were used as input for enrichment.

Data availability

The GWAS summary statistics generated in this study are available in Figshare (https://doi.org/10.6084/m9.figshare.30227764). UK Biobank data are available by request through the UK Biobank Access Management System https://www.ukbiobank.ac.uk/. The 1000 Genomes Phase 3 data are available at https://www.internationalgenome.org/. The GTEx protected data are available through the database of Genotypes and Phenotypes (dbGaP) (accession no. phs000424.v8). The processed GTEx eQTL and sQTL summary statistics are available on the GTEx portal (https://gtexportal.org/home/datasets). The EyeGEx retina eQTL summary statistics is available at https://visiongenomics.org. The snRNA-seq data for the anterior segment and macula are available in Gene Expression Omnibus (GEO) accession number GSE199013, for the optic nerve head and posterior tissues in GSE236566, and for the retina in GSE226108. All the results from our analyses (e.g., GWAS, MAGMA, colocalization, Mendelian randomization, cell-type enrichment) can be found in the Supplementary Data.

Received: 8 December 2025; Accepted: 18 August 2026;
Published online: 19 September 2026

References

1. Davuluru, S. S. et al. Identifying, understanding, and addressing disparities in glaucoma care in the United States. *Transl. Vis. Sci. Technol.* **12**, 18 (2023).
2. Tham, Y.-C. et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040. *Ophthalmology* **121**, 2081–2090 (2014).
3. Bailey, J. N. C. et al. Genome-wide association analysis identifies TXNRD2, ATXN2 and FOXC1 as susceptibility loci for primary open-angle glaucoma. *Nat. Genet.* **48**, 189–194 (2016).
4. Gharahkhani, P. et al. Common variants near ABCA1, AFAP1 and GMDS confer risk of primary open-angle glaucoma. *Nat. Genet.* **46**, 1120–1125 (2014).
5. Gharahkhani, P. et al. Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nat. Commun.* **12**, 1258 (2021).
6. Han, X. et al. Automated AI labeling of optic nerve head enables insights into cross-ancestry glaucoma risk and genetic discovery in >280,000 images from UKB and CLSA. *Am. J. Hum. Genet.* **108**, 1204–1216 (2021).
7. Hysi, P. G. et al. Genome-wide analysis of multi-ancestry cohorts identifies new loci influencing intraocular pressure and susceptibility to glaucoma. *Nat. Genet.* **46**, 1126–1130 (2014).
8. Pasquale, L. R. et al. CDKN2B-AS1 genotype-glaucoma feature correlations in primary open-angle glaucoma patients from the United States. *Am. J. Ophthalmol.* **155**, 342–353.e5 (2013).
9. Springelkamp, H. et al. New insights into the genetics of primary open-angle glaucoma based on meta-analyses of intraocular pressure and optic disc characteristics. *Hum. Mol. Genet.* **26**, 438–453 (2017).
10. Wiggs, J. L. et al.GENEVA Consortium Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma in Caucasians from the USA. *Hum. Mol. Genet.* **20**, 4707–4713 (2011).
11. Wiggs, J. L. et al. Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma. *PLoS Genet.* **8**, e1002654 (2012).
12. Han, X. et al.23andMe Research Team, International Glaucoma Genetics Consortium Large-scale multitrait genome-wide association analyses identify hundreds of glaucoma risk loci. *Nat. Genet.* **55**, 1116–1125 (2023).
13. Johnson, K. B. et al. Precision medicine, AI, and the future of personalized health care. *Clin. Transl. Sci.* **14**, 86–93 (2021).
14. Shickel, B., Tighe, P. J., Bihorac, A. & Rashidi, P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J. Biomed. Health Inform.* **22**, 1589–1604 (2018).
15. Khawaja, A. P. et al. Genome-wide analyses identify 68 new loci associated with intraocular pressure and improve risk prediction for primary open-angle glaucoma. *Nat. Genet.* **50**, 778–782 (2018).
16. Hoffmann, E. M., Zangwill, L. M., Crowston, J. G. & Weinreb, R. N. Optic disk size and glaucoma. *Surv. Ophthalmol.* **52**, 32–49 (2007).
17. Weinreb, R. N., Aung, T. & Medeiros, F. A. The pathophysiology and treatment of glaucoma: a review. *JAMA* **311**, 1901–1911 (2014).
18. Scuderi, G., Fragiotta, S., Scuderi, L., Iodice, C. M. & Perdicchi, A. Ganglion cell complex analysis in glaucoma patients: What can it tell us? *Eye Brain* **12**, 33–44 (2020).
19. Currant, H. et al. Genetic variation affects morphological retinal phenotypes extracted from UK Biobank optical coherence tomography images. *PLoS Genet.* **17**, e1009497 (2021).
20. 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 (2019).
21. Kirchler, M. et al. transferGWAS: GWAS of images using deep transfer learning. *Bioinformatics* **38**, 3621–3628 (2022).
22. Xie, Z. et al. iGWAS: Image-based genome-wide association of self-supervised deep phenotyping of retina fundus images. *PLoS Genet.* **20**, e1011273 (2024).
23. Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018).
24. Wu, J. et al. DeepMoIC: multi-omics data integration via deep graph convolutional networks for cancer subtype classification. *BMC Genom.* **25**, 1209 (2024).

25. Chen, L. Can classical statistics and deep learning converge on explainable, causally driven target discovery? *DNA Res.* **32**, dsaf024 (2025).
26. Zekavat, S. M. et al. Deep learning of the retina enables phenome- and genome-wide analyses of the microvasculature. *Circulation* **145**, 134–150 (2022).
27. Tomasoni, M. et al. Genome-wide association studies of retinal vessel tortuosity identify numerous novel loci revealing genes and pathways associated with ocular and cardiometabolic diseases. *Ophthalmol. Sci.* **3**, 100288 (2023).
28. Alipanahi, B. et al. Large-scale machine-learning-based phenotyping significantly improves genomic discovery for optic nerve head morphology. *Am. J. Hum. Genet.* **108**, 1217–1230 (2021).
29. Jackson, V. E. et al. Multi-omic spatial effects on high-resolution AI-derived retinal thickness. *Nat. Commun.* **16**, 1317 (2025).
30. Sergouniotis, P. I. et al. Autoencoder-based phenotyping of ophthalmic images highlights genetic loci influencing retinal morphology and provides informative biomarkers. *Bioinformatics* **41**, btae732 (2024).
31. Yun, T. et al. Unsupervised representation learning on high-dimensional clinical data improves genomic discovery and prediction. *Nat. Genet.* **56**, 1604–1613 (2024).
32. Kazeminasab, S. et al. An artificial intelligence method for phenotyping of OCT-derived thickness maps using unsupervised and self-supervised deep learning. *J. Imaging Inform. Med.* https://doi.org/10.1007/s10278-025-01539-x (2025).
33. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).
34. von Hippel, P. T. The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Med. Res. Methodol.* **15**, 35 (2015).
35. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
36. Lee, S. et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* **91**, 224–237 (2012).
37. GTEx Consortium The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
38. Ratnapriya, R. et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).
39. Eraslan, G. et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science* **376**, eabl4290 (2022).
40. Rouhana, J. M. et al. ECLIPSER: identifying causal cell types and genes for complex traits through single cell enrichment of e/sQTL-mapped genes in GWAS loci. https://doi.org/10.1101/2021.11.24.469720 (2021).
41. Shah, R. L. & Guggenheim, J. A. & UK Biobank Eye and Vision Consortium. 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 (2018).
42. Hamel, A. R. et al. Integrating genetic regulation and single-cell expression with GWAS prioritizes causal genes and cell types for glaucoma. *Nat. Commun.* **15**, 396 (2024).
43. Srivastava, A., Rikhari, D. & Srivastava, S. RSPO2 as Wnt signaling enabler: Important roles in cancer development and therapeutic opportunities. *Genes Dis.* **11**, 788–806 (2024).
44. Masutani, M., Sonenberg, N., Yokoyama, S. & Imataka, H. Reconstitution reveals the functional core of mammalian eIF3. *EMBO J.* **26**, 3373–3383 (2007).
45. Martin, K. R. G., Quigley, H. A., Valenta, D., Kielczewski, J. & Pease, M. E. Optic nerve dynein motor protein distribution changes with intraocular pressure elevation in a rat model of glaucoma. *Exp. Eye Res.* **83**, 255–262 (2006).
46. Buniello, A. et al. Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res.* **53**, D1467–D1475 (2025).
47. McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. & Lin, X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* **76**, 1262–1272 (2020).
48. GTEx Consortium The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
49. Fairley, S., Lowy-Gallego, E., Perry, E. & Flicek, P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* **48**, D941–D947 (2020).
50. van Zyl, T. et al. Cell atlas of the human ocular anterior segment: Tissue-specific and shared cell types. *Proc. Natl. Acad. Sci. USA* **119**, e2200914119 (2022).
51. Monavarfeshani, A. et al. Transcriptomic analysis of the ocular posterior segment completes a cell atlas of the human eye. *Proc. Natl. Acad. Sci. USA* **120**, e2306153120 (2023).
52. Hahn, J. et al. Evolution of neuronal cell classes and types in the vertebrate retina. *Nature* **624**, 415–424 (2023).
53. Liang, Q. et al. A multi-omics atlas of the human retina at single-cell resolution. *Cell Genom.* **3**, 100298 (2023).

Acknowledgements

The authors are grateful to the patients of the MEE glaucoma clinic and the participants of the UK Biobank whose contributions were essential for this research. The authors would also like to acknowledge Yuyang Luo, Puja Mehta, and Susan Xu, who provided technical assistance with some of the computational tools. The authors also acknowledge these funding sources: National Institutes of Health grants R01 EY036518 (N.Z.), R01 EY036222 (M.W.), R01 EY030575 (T.E.), R01EY031424 (A.V.S.); National Eye Institute: P30EY003790 (T.E.), P30EY014104 (A.V.S.), Research to Prevent Blindness (N.Z., T.E.).

Author contributions

N.Z., A.V.S., J.W., L.C., S.K.H., T.E., M.W., M.E., and Y.Z. contributed to the conceptualization of the study. N.Z., A.V.S., J.W., L.C., S.K.H., T.E., M.W., and M.E. developed the methodology. N.Z., A.V.S., J.W., L.C., S.K.H., T.E., M.W., M.E., and Y.Z. carried out the investigation. L.C., N.Z., and A.V.S. prepared the visualizations. N.Z., A.V.S., J.W., T.E., M.W., and M.E. secured funding. N.Z., A.V.S., J.W., L.C., S.K.H., T.E., M.W., M.E., and Y.Z. oversaw project administration. N.Z., A.V.S., J.W., T.E., M.W., and M.E. provided supervision. L.C. wrote the original manuscript draft. All authors reviewed the manuscript.

Competing interests

N.Z. serves as a consultant for Sanofi/Genzyme, which is unrelated to this work. All other authors declare no competing financial or non-financial interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-026-00612-w.

Correspondence and requests for materials should be addressed to Nazlee Zebardast.

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

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.