Multi-trait Polygenic Probability Risk Score Enhances Glaucoma Prediction Across Ancestries

Xiaoyi Raymond Gao[1,2]

[1]Department of Ophthalmology and Visual Sciences, The Ohio State University, Columbus, OH 43210, USA

[2]Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA

[*]Correspondence should be addressed to:
Xiaoyi Raymond Gao, PhD
Associate Professor
Department of Ophthalmology and Visual Sciences
Department of Biomedical Informatics
The Ohio State University
Columbus, OH 43212
USA
Phone: (614) 293-5287
Fax: (614) 293-5602
Email: raymond.gao@osumc.edu

Abstract:

Primary open-angle glaucoma (POAG) remains the leading cause of irreversible blindness worldwide, with early detection crucial for preventing vision loss. However, current risk assessment methods have limited predictive power. Here, we present a multi-trait polygenic probability risk score (PPRS) approach that integrates multiple glaucoma-related traits and leverages functional genomic annotations to enhance POAG prediction across diverse ancestries. We constructed PRSs for POAG, intraocular pressure (IOP), vertical cup-to-disc ratio (VCDR), and retinal nerve fiber layer (RNFL) thickness using extensive genomic coverage (>7 million variants) and 96 functional annotations through the SBayesRC method. Validation in the UK Biobank (n=324,713, European ancestry) and Mexican American Glaucoma Genetic Study (MAGGS, n=4,549, Latino ancestry) demonstrated significant improvements in predictive accuracy over conventional approaches. Our multi-trait PPRS achieved area under the curve (AUC) values of 0.814 in Europeans and 0.801 in Latinos, compared to AUC ≤0.79 for single-trait models. We identified ancestry-specific differences in genetic contributions, with IOP demonstrating the strongest association in Europeans (OR=1.63, $P = 5.37 \times 10^{-89}$), while VCDR was predominant in Latinos (OR=1.64, $P = 2.04 \times 10^{-11}$). The model achieved remarkable risk stratification, with the highest PPRS decile showing 80.2-fold and 51.1-fold increased POAG risk in Europeans and Latinos, respectively, compared to the lowest decile. Importantly, the top PPRS quintile captured 65.9% and 62.2% of POAG cases in Europeans and Latinos, substantially improving upon previous approaches. Our findings demonstrate that integrating multiple disease-relevant traits and functional annotations significantly enhances polygenic prediction of POAG across diverse populations, with significant implications for targeted screening, early intervention, and reduction of disease burden.

Keywords:

glaucoma, polygenic risk score, genetic prediction, multi-ancestry, precision medicine

Introduction

Glaucoma is the leading cause of irreversible blindness[1, 2], affecting about 80 million individuals worldwide[3, 4], with primary open-angle glaucoma (POAG) constituting approximately 90% of glaucoma cases in North America[4, 5]. Despite available treatments to slow disease progression, the irreversible nature of glaucoma-induced vision loss underscores the critical importance of early detection and intervention. While traditional risk assessment relies primarily on clinical measurements and family history, recent advances in genomics offer promising opportunities for enhanced risk stratification.

Polygenic risk scores (PRSs) have emerged as powerful tools for quantifying genetic predisposition to complex diseases[6-8]. However, their application to POAG has faced three key limitations. First, previous studies utilizing single-trait PRSs have achieved limited stratification ability, with about 50% of cases being identified in the top PRS quintile in a recent report[9]. Second, existing approaches have failed to capture the inherent complexity of POAG's pathophysiology, which involves multiple quantitative endophenotypes[1, 10], including intraocular pressure (IOP), vertical cup-to-disc ratio (VCDR), and retinal nerve fiber layer (RNFL) thickness. Though the multi-trait analysis of genome-wide association studies (MTAG) approach can analyze multi-traits jointly, previous studies indicated MTAG-derived PRS did not provide the optimal prediction accuracy for POAG[11]. For POAG PRS prediction per se, these endophenotypes may be best modeled separately in PRS prediction. Third, most POAG PRS prediction studies have employed limited genomic coverage (< 1M markers) and have lacked integration of functional annotations, potentially missing important biological signals. These limitations underscore the necessity for more advanced models that can capture the multifaceted genetic underpinnings of POAG.

Here, we present a multi-trait polygenic probability risk score (PPRS) approach, integrating PRSs derived from different modalities, including imaging-derived phenotypes, that addresses these limitations through three key innovations: 1) integrating multiple disease-relevant quantitative traits into PRS model-

ing; 2) incorporating extensive genomic coverage (over seven million variants) and 96 functional annotations; 3) validating across diverse ancestral populations. We hypothesized that this comprehensive approach would significantly improve POAG risk prediction across diverse populations compared to existing methods. To test this hypothesis, we evaluated our model in two independent cohorts: the UK Biobank (UKB) for European ancestry individuals and the Mexican American Glaucoma Genetic Study (MAGGS) for Latino participants.

Methods

Study Design and Populations

We conducted a two-stage study: first, deriving genome-wide association study (GWAS) summary statistics from previously published data and our own GWAS analyses (details provided below); second, validating PRSs in a subset of the UKB cohort (n = 324,713, POAG cases: 2,168) and the MAGGS cohort (n = 4,549, POAG cases: 275), both independent of the GWAS discovery samples. UKB and MAGGS received approval from respective institutional review boards[12-14]. All participants provided written informed consent. The study adhered to the tenets of the Declaration of Helsinki. We obtained fully de-identified data to ensure participant confidentiality.

UKB dataset

The UKB is an ongoing, large-scale prospective cohort study with over 500,000 adult participants aged 40 to 69 at enrollment (2006–2010), registered with the National Health Service in the United Kingdom[12, 13]. The study collected medical information, lifestyle data, and DNA samples, including ophthalmologic data for a subset of approximately 118,000 participants. Our access and use of the UKB data were approved under application #23424. Our analysis focused on participants of European ancestry. Genotyping was performed using the UK BiLEVE Axiom Array or the UKB Axiom Array, with imputation based on reference panels from the 1000 Genomes Project (1KGP), UK10K, and the Haplotype Reference Consortium (HRC)[15]. Variants with an imputation INFO score less than 0.3 and minor allele frequency less than

0.01 were excluded, resulting in approximately 10.3 million variants for downstream analysis. For individuals with IOP measurements and RNFL data, we conducted IOP analyses[16] and a GWAS of RNFL to derive summary statistics. For those without ophthalmic exams (n = 324,713, unrelated, independent of the IOP and RNFL GWAS samples)[17], we utilized their data for PRS construction and association testing with POAG. POAG (H40.1) cases were identified using the International Classification of Diseases Tenth Revision (ICD-10) codes. Controls were identified as those who did not have glaucoma.

MAGGS dataset

MAGGS samples were sourced from the Los Angeles Latino Eye Study (LALES)[18], a population-based epidemiologic study examining visual impairment and ocular diseases in 6,357 Latino individuals aged 40 or older in Los Angeles County, California. The presence of POAG was determined by agreement among three glaucoma specialists using clinical data, including the presence of open angles, characteristic visual field abnormalities, and optic disc damage in at least one eye. MAGGS and LALES genotyped 4,996 Latino individuals using either the Illumina OmniExpress BeadChip or the Illumina Hispanic/SOL BeadChip[14, 19]. Imputation was performed using 1KGP reference panel, excluding variants with a Mach[20] Rsq less than 0.3 and minor allele frequency less than 0.01, resulting in approximately 9.9 million imputed variants. Our analysis included 4,549 participants (4,108 unrelated and 441 related individuals) from this dataset.

GWAS summary statistics

We used previously published GWAS summary statistics for POAG, IOP, and vertical cup-to-disc ratio (VCDR), as well as our own GWAS results for RNFL. The details are as follows:

1) POAG: We used GWAS results from Gharahkhani et al. (2021), which performed a multi-ethnic meta-analysis using the International Glaucoma Genetics Consortium (IGGC) and UKB datasets, comprising 383,500 individuals of European, Asian, and African descent[21]. For building PRSs for UKB

participants, we used results from IGGC only, excluding UKB individuals, resulting in a cross-ancestry meta-analysis with 192,702 individuals suitable for UKB PRS construction.

2) IOP: We used GWAS results from Gao et al. (2018), including 115,486 UKB participants of European ancestry[16].

3) VCDR: We used GWAS summary statistics from Han et al. (2021), involving 111,724 individuals of European and Asian descent from UKB, the Canadian Longitudinal Study on Aging (CLSA), and IGGC[22]. VCDR measurements were obtained from human graders and deep learning models applied to retinal fundus images.

4) RNFL: We conducted a GWAS with 52,902 European ancestry UKB participants using the REGENIE[23] software and RNFL thickness values derived from optical coherence tomography images[24]. We excluded outliers and low-quality images (image quality score less than 45). We averaged RNFL values between the left and right eyes and applied rank-based inverse normalization. We adjusted for age, sex, diabetes, spherical equivalent, and the first 10 principal components of genetic ancestry. Supplementary Figure 1 presents a Manhattan plot of the genome-wide -log10(*P* values) for our RNFL GWAS analysis.

These summary statistics provided the foundation for constructing PRSs in our analysis, tailored to different phenotypes relevant to POAG risk.

PRS construction

We employed three different approaches to construct PRSs:

First, we implemented a traditional clumping and thresholding (C+T) based PRS[25] using PLINK[26, 27] to construct the POAG PRS, denoted as PRS-POAG-C+T. Independent single nucleotide polymorphisms (SNPs) were identified using linkage disequilibrium (LD)-based clumping ($r^2 < 0.3$) and selected based on a p-value threshold of $P < 5 \times 10^{-5}$.

Second, we constructed enhanced PRSs using the SBayesRC[28] method, which leverages functional genomic annotations and extensive SNP coverage to improve polygenic risk prediction for complex

traits or diseases. By integrating 96 functional annotations, including regulatory elements, chromatin states, and transcription factor binding sites, and analyzing over seven million SNPs, we sought to substantially improve PRS accuracy. Using SBayesRC, we constructed PRSs for POAG (PRS-POAG), IOP (PRS-IOP), VCDR (PRS-VCDR), and RNFL thickness (PRS-RNFL).

Third, we constructed a MTAG-derived PRS for POAG to enable direct comparison. We utilized summary statistics for 2,763 uncorrelated SNPs (identified through LD clumping at $r^2 = 0.1$ and a significance threshold of $P < 0.001$) from Craig et al.'s (2020) MTAG analysis[29]. The MTAG-derived PRS was calculated using the formula: PRS-POAG-MTAG(i) = $\sum \beta_j G_{ij}$, where β_j represents the regression coefficient for SNP_j, and G_{ij} indicates the risk allele dosage for SNP_j in individual i. Given that Craig et al. used UKB participants in their MTAG analysis, we avoided applying the MTAG-derived PRS to UKB participants to prevent potential data leakage. Consequently, this PRS was exclusively applied to MAGGS study participants.

All constructed PRSs were standardized to have a mean of zero and a variance of one.

Statistical analysis

To evaluate the association between PRSs and POAG, we performed logistic regression analyses adjusted for age and sex. We assessed the predictive ability of the PRSs for POAG using AUC. For unrelated samples, we employed leaving-one-out cross-validation to obtain predicted probabilities of POAG for each individual. For MAGGS related samples, predictive models were developed using MAGGS unrelated samples and subsequently applied to the related individuals to derive predicted probabilities. We termed this derived probability measure as polygenic probability risk score (PPRS), integrating multiple PRSs through statical modeling, to differentiate it from traditional single-trait PRSs and the MTAG-derived PRS. To check potential multicollinearity among PRSs, we used two complementary approaches. First, we computed the variance inflation factor (VIF) for each predictor, noting that VIF values less than 5 typically indicate low correlation[30]. Second, we employed XGBoost[31], a tree-based method known to be less sensitive to collinearity, and examined feature importances via SHapley Additive exPlanations (SHAP)

values[32], following our previous methodology[33]. For risk stratification, we divided the PPRS into deciles and compared the odds of POAG among study participants in higher PPRS deciles versus those in the lowest PPRS decile in both the UKB and MAGGS datasets. A statistical significance threshold of $P < 0.05$ was used for all analyses. All analyses and plots were performed using R version 4.4, SAS version 9.4, and Python version 3.12.

Results

Table 1 summarizes the characteristics of the study samples from the UK Biobank (UKB) and the Mexican American Glaucoma Genetic Study (MAGGS). The UKB cohort comprised 324,713 individuals with 2,168 glaucoma cases and 322,545 controls. The mean (standard deviation [SD]) age of glaucoma and non-glaucoma were 63.0 (5.7) and 57.1 (8.0) years, respectively. Females consisted of 48.4% and 54.3% in glaucoma and non-glaucoma participants, respectively. The MAGGS dataset included 4,549 participants with 275 glaucoma cases and 4,274 controls. The mean (SD) age of glaucoma and non-glaucoma were 66.4 (11.1) and 56.1 (10.2), respectively. Females constituted 52.4% and 59.2% of glaucoma and non-glaucoma participants, respectively.

Table 2 presents the results of the single PRS logistic regression analyses, adjusted for age and sex, among UKB and MAGGS participants. In the UKB cohort (Table 2a), the traditional clumping and thresholding (C+T) method PRS for POAG (PRS-POAG-C+T) yielded an odds ratio (OR) of 1.71 (95% confidence interval [CI]: 1.64–1.79; $P = 8.71 \times 10^{-141}$). The SBayesRC-derived PRS for POAG (PRS-POAG) showed a stronger association with an OR of 2.14 (95% CI: 2.05–2.24; $P = 8.54 \times 10^{-269}$). The SBayesRC-derived PRSs for IOP (PRS-IOP), VCDR (PRS-VCDR), and RNFL thickness (PRS-RNFL) also demonstrated significant associations with POAG. Both PRS-POAG and PRS-IOP gave similar AUC (0.79) for predicting glaucoma. Similarly, in the MAGGS cohort (Table 2b), the PRS-POAG-MTAG had a lower OR of 1.44 (95% CI: 1.24–1.60; $P = 1.17 \times 10^{-7}$) and a lower AUC of 0.75 (95% CI: 0.72–0.78). PRS-VCDR showed the highest OR of 1.90 (95% CI: 1.66-2.17; $P = 1.39 \times 10^{-20}$) and the highest AUC

of 0.78 (95% CI: 0.75-0.81). The PRS-POAG-C+T, PRS-IOP, PRS-POAG, and PRS-RNFL in MAGGS participants also exhibited significant associations with POAG.

Table 3 displays the results of the multiple PRS logistic regression analyses, assessing the combined association of PRSs with POAG and their relative importance in each cohort. In the UKB participants, PRS-POAG had an OR of 1.59 ($P = 1.23 \times 10^{-74}$), PRS-IOP showed the strongest association with an OR of 1.63 ($P = 5.37 \times 10^{-89}$), PRS-VCDR had an OR of 1.28 ($P = 1.99 \times 10^{-27}$), and PRS-RNFL exhibited an OR of 0.94 ($P = 0.002$), indicating a slight inverse association (larger PRS-RNFL indicated thicker RNFL and hence is protective). The OR for a simultaneous one-SD-unit-increase in all four standardized PRS variables was 3.11 (95% CI: 2.89-3.34). In the MAGGS cohort, PRS-POAG had an OR of 1.40 ($P = 7.63 \times 10^{-5}$), PRS-IOP an OR of 1.15 ($P = 0.07$), PRS-VCDR showed the strongest association with an OR of 1.64 ($P = 2.04 \times 10^{-11}$), and PRS-RNFL had an OR of 0.85 ($P = 0.016$), also suggesting an inverse relationship. The OR for a simultaneous one-SD-unit-increase in all four standardized PRS variables was 2.25 (95% CI: 1.79-2.82). These results highlight the significant contributions of the different PRSs in predicting POAG risk in both cohorts. Notably, PRS-IOP contributed most significantly to POAG risk prediction in the UKB dataset, whereas PRS-VCDR had the strongest contribution in the MAGGS dataset.

To evaluate potential collinearity among these PRSs (PRS-POAG, PRS-IOP, PRS-VCDR, and PRS-RNFL), we applied two complementary approaches: (1) VIF analysis and (2) XGBoost with SHAP-based feature importance. Supplementary Table 1 presents the VIF values for each predictor in the logistic regression models. These values are only slightly above 1, indicating minimal collinearity, and all remain well below the typical cutoff of 5. Supplementary Figure 2 displays the SHAP feature importance rankings from our XGBoost models. In UKB participants, PRS-IOP and PRS-POAG both rank highly among

the PRSs. In MAGGS, PRS-VCDR shows the highest importance among the PRSs, aligning with the logistic regression findings. Collectively, these results underscore the validity and robustness of our logistic regression analyses.

We examined the discriminatory ability of the PRSs (PRS-POAG, PRS-IOP, PRS-VCDR, and PRS-RNFL) for predicting POAG using leave-one-out cross-validation logistic regression models and calculated AUC. Figure 1 illustrates these results. In UKB participants (Figure 1A), the baseline model including only age and sex had an AUC of 0.711. Adding PRS-POAG-C+T increased the AUC to 0.761, while incorporating the SBayesRC-derived PRS-POAG improved it to 0.790, a 3% increase over the C+T approach ($P = 2.31 \times 10^{-29}$). Incorporating all four PRSs (PRS-POAG, PRS-IOP, PRS-VCDR, and PRS-RNFL) with adjustment for age and sex resulted in an AUC of 0.814 (95% CI: 0.805–0.822), representing a 10.3% improvement over the baseline model ($P = 4.37 \times 10^{-135}$). In MAGGS participants (Figure 1B), the baseline AUC was 0.746, which increased to 0.760 with PRS-POAG-MTAG, to 0.773 with PRS-POAG-C+T and to 0.782 with PRS-POAG. The full model with all four PRSs (PRS-POAG, PRS-IOP, PRS-VCDR, and PRS-RNFL) achieved an AUC of 0.801 (95% CI: 0.775–0.828), providing a 5.5% enhancement over the baseline model ($P = 3.05 \times 10^{-7}$). These findings indicate that incorporating multiple glaucoma-related PRSs significantly improves the prediction of POAG in both cohorts.

We further evaluated POAG risk stratification using PPRS from our leave-one-out logistic regression models, which aggregate the effects of age, sex, the PRSs of POAG, IOP, VCDR, and RNFL and predict on the leave-one-out sample. Analysis of PPRS revealed strong risk stratification capabilities (Figure 2). In the UKB cohort (Figure 2A), clear trend emerged with increasing ORs across higher PPRS deciles. Individuals in the highest decile had 80.24-fold increased odds of POAG (95% CI: 46.42–138.69; $P = 1.53 \times 10^{-55}$) compared to those in the lowest decile. Similarly, in the MAGGS cohort (Figure 2B), participants in the highest decile demonstrated 51.11-fold increased odds of POAG (95% CI: 16.11–162.11; $P = 2.43 \times 10^{-11}$) compared to the lowest decile. Most POAG cases (46.63% in UKB and 41.82% in MAGGS) are

in the highest decile category of PPRS (supplementary Table 2). Notably, when examining the top quintile of PPRS, we identified 65.95% and 62.2% of POAG cases for Europeans and Latinos, respectively, representing a substantial improvement over previous approaches. Take together, these findings underscore the strong relationship between PPRS deciles and POAG risk across UKB and MAGGS, highlighting the potential utility of the PPRS for risk stratification in clinical settings.

Discussion

Our study demonstrates that integrating multiple glaucoma-related traits with functional annotations substantially improves the polygenic prediction for POAG across diverse ancestries. Three key findings emerge from our analysis: First, our multi-modal PPRS approach achieved higher predictive accuracy, with AUC values of 0.814 in Europeans and 0.801 in Latinos, compared to our single-trait models (AUC $\leq$ 0.79). This improvement likely reflects the complex genetic architecture of POAG, where multiple biological pathways contribute to disease development. Second, we identified ancestry-specific differences in the relative contributions of component PRSs. In Europeans, IOP showed the strongest association, whereas VCDR was most strongly associated in Latinos. This finding underscores the importance of population-specific risk assessment and suggests distinct genetic mechanisms underlying POAG across ancestral groups. Third, the strong risk stratification achieved by our PPRS, demonstrating over 80-fold and 51-fold risk differences between extreme deciles in Europeans and Latinos, respectively, suggests significant clinical utility. This high level of stratification can effectively identify individuals at greatest risk who might benefit from enhanced screening and early intervention.

Direct comparison with previous glaucoma PRS studies is challenging due to differences in target datasets. However, several observations can be highlighted. First, the correlation between phenotypes is not equivalent to correlations among PRSs. Despite high phenotypic correlations among POAG, IOP, and VCDR[34], correlations among the corresponding PRSs are notably lower based on our VIF analysis. By incorporating quantitative traits like IOP, VCDR, and RNFL thickness into a unified PPRS framework,

we effectively address the heterogeneous nature of POAG. Each of these traits reflects distinct biological mechanisms and provides unique predictive information, enhancing the model's capability to capture overall glaucoma risk. Second, MTAG-derived PRS has not yielded optimal prediction accuracy for POAG, as demonstrated in recent studies[11]. Although several recent reports employed lassosum for PRS construction [9, 35-38], previous methodological evaluations indicate that lassosum does not achieve optimal predictive performance[11, 39]. We employed the SBayesRC approach, which integrates functional genomic annotations to prioritize biologically relevant variants, leveraging millions of SNPs. This method offers marked advantages over conventional PRS methodologies[28]. These improvements in polygenic risk estimation have substantial implications for effective sample stratification.

One of the most direct clinical applications of PRS is identifying high-risk individuals. MacGregor et al. (2018), using an IOP PRS, reported an increased glaucoma risk (OR = 5.6, 95% CI: 4.1–7.6) comparing the top and bottom deciles[40]. Craig et al. (2020), employing an MTAG approach for their POAG PRS, observed a higher OR of 14.9 (95% CI: 10.7–20.9) for comparing the top and bottom deciles[29]. Gao et al. (2019) found an OR of 6.3 for their top IOP PRS quintile versus the bottom quintile comparison and approximately 40.6% of POAG cases were found in the top quintile[17], while de Vries et al. (2025) identified about 50% in their top glaucoma PRS quintile[9]. Our study substantially improves upon these results, demonstrating ORs of 80.2 (95% CI: 46.4–138.7) in Europeans and 51.1 (95% CI: 16.1–162.1) in Latinos when comparing extreme deciles of our PPRS. Additionally, we identified 65.9% and 62.2% of glaucoma cases within the top quintile of our PPRS scores for Europeans and Latinos (supplementary Table 2), respectively, underscoring significant advancement in polygenic prediction performance relative to previous studies.

Our study has several strengths, notably the use of a comprehensive multi-trait PPRS approach that integrates multiple glaucoma-related traits, including imaging-derived ones, and leverages functional annotations (through SBayesRC) to improve prediction accuracy. By evaluating PRS performance in both

European and Latino populations, we address the crucial need for inclusive and representative genetic research, especially among traditionally understudied groups[41]. However, our study also has limitations. Due to resource constraints, we included only European and Latino populations in this study. Expanding future research to additional populations would enhance generalizability and ensure broader applicability of our findings. Furthermore, our PRS construction was limited to common variants. Incorporating rare variants, similar to previous reports on rare variants associated with IOP[42], warrants further investigation.

We utilized logistic regression for our statistical modeling, providing interpretable outcomes suitable for analyzing associations between PRSs and glaucoma. Nonetheless, prediction accuracy might be further improved through advanced machine learning techniques. For example, gradient boosting methods like XGBoost, which we have successfully applied in related studies[33], could potentially capture more complex non-linear interactions. However, the implementation of these machine learning techniques typically requires more complex hyperparameter tuning and computational resources. Future research could explore the potential of these advanced modeling techniques while balancing their complexity with interpretability and practical application.

In conclusion, our multi-trait PPRS approach, combining multiple glaucoma-related traits and incorporating functional annotations, represents a significant advancement in polygenic risk prediction for POAG. The robust performance across different ancestral groups and substantial risk stratification capabilities suggest considerable potential for enhancing early detection and preventative strategies. By facilitating the identification of individuals at high genetic risk, our approach has significant implications for timely clinical intervention and could ultimately reduce the public health burden of POAG.

Acknowledgements

We would like to thank the study participants from the UK Biobank, the Los Angeles Latino Eye Study, the Mexican American Glaucoma Genetic Study, and the staff who aided in data collection and processing. This work was supported in part by National Institutes of Health (NIH; Bethesda, MD, USA) grants P30EY032857 and R01EY022651. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Fan BJ, Wiggs JL. Glaucoma: genes, phenotypes, and new directions for therapy. J Clin Invest 2010;120(9):3064-72.
2. Blindness GBD, Vision Impairment C, Vision Loss Expert Group of the Global Burden of Disease S. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health 2021;9(2):e144-e60.
3. Weinreb RN, Leung CK, Crowston JG, et al. Primary open-angle glaucoma. Nat Rev Dis Primers 2016;2:16067.
4. Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology 2014;121(11):2081-90.
5. Lang GK. Ophthalmology: A Pocket Textbook Atlas, 2nd ed. Stuttgart, New Yok: Thieme 2007.
6. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460(7256):748-52.
7. Kachuri L, Chatterjee N, Hirbo J, et al. Principles and methods for transferring polygenic risk scores across global populations. Nat Rev Genet 2024;25(1):8-25.
8. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet 2018;19(9):581-90.
9. de Vries VA, Hanyuda A, Vergroesen JE, et al. The Clinical Usefulness of a Glaucoma Polygenic Risk Score in 4 Population-Based European Ancestry Cohorts. Ophthalmology 2025;132(2):228-37.
10. Ramdas WD, Amin N, van Koolwijk LM, et al. Genetic architecture of open angle glaucoma and related determinants. J Med Genet 2011;48(3):190-6.
11. Biradar M, Luben R, Stuart K, et al. Comparison of multiple polygenic risk score derivation approaches for population-level prediction of primary open-angle glaucoma. Investigative Ophthalmology & Visual Science 2023;64(8):492-.
12. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779.
13. Allen NE, Sudlow C, Peakman T, et al. UK biobank data: come and get it. Sci Transl Med 2014;6(224):224ed4.
14. Nannini DR, Torres M, Chen YI, et al. A Genome-Wide Association Study of Vertical Cup-Disc Ratio in a Latino Population. Invest Ophthalmol Vis Sci 2017;58(1):87-95.
15. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562(7726):203-9.
16. Gao XR, Huang H, Nannini DR, et al. Genome-wide association analyses identify new loci influencing intraocular pressure. Hum Mol Genet 2018;27(12):2205-13.
17. Gao XR, Huang H, Kim H. Polygenic Risk Score Is Associated With Intraocular Pressure and Improves Glaucoma Prediction in the UK Biobank Cohort. Transl Vis Sci Technol 2019;8(2):10.
18. Varma R, Paz SH, Azen SP, et al. The Los Angeles Latino Eye Study: design, methods, and baseline data. Ophthalmology 2004;111(6):1121-31.
19. Nannini D, Torres M, Chen YD, et al. African Ancestry Is Associated with Higher Intraocular Pressure in Latinos. Ophthalmology 2016;123(1):102-8.
20. Li Y, Abecasis GR. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. Am J Hum Genet 2006;S79.
21. Gharahkhani P, Jorgenson E, Hysi P, et al. Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. Nat Commun 2021;12(1):1258.
22. Han X, Steven K, Qassim A, 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 2021.
23. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet 2021;53(7):1097-103.

24. Ko F, Foster PJ, Strouthidis NG, et al. Associations with Retinal Pigment Epithelium Thickness Measures in a Large Cohort: Results from the UK Biobank. Ophthalmology 2017;124(1):105-17.
25. Prive F, Vilhjalmsson BJ, Aschard H, Blum MGB. Making the Most of Clumping and Thresholding for Polygenic Scores. Am J Hum Genet 2019;105(6):1213-21.
26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559-75.
27. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.
28. Zheng Z, Liu S, Sidorenko J, et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. Nat Genet 2024;56(5):767-77.
29. Craig JE, Han X, Qassim A, et al. Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. Nat Genet 2020;52(2):160-6.
30. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R, 1st ed. New York, NY: Springer, 2009; 426 p.
31. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: Association for Computing Machinery, 2016.
32. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2020;2(1):56-67.
33. Gao XR, Chiariglione M, Qin K, et al. Explainable machine learning aggregates polygenic risk scores and electronic health records for Alzheimer's disease prediction. Scientific Reports 2023;13(1):450.
34. He W, Lee SS, Diaz Torres S, et al. Predictive Power of Polygenic Risk Scores for Intraocular Pressure or Vertical Cup-Disc Ratio. JAMA Ophthalmol 2025;143(1):15-22.
35. Tran JH, Kang J, Han E, et al. Use of Diagnostic Codes for Primary Open-Angle Glaucoma Polygenic Risk Score Construction in Electronic Health Record-Linked Biobanks. Am J Ophthalmol 2024;267:204-12.
36. Sekimitsu S, Ghazal N, Aziz K, et al. Primary Open-Angle Glaucoma Polygenic Risk Score and Risk of Disease Onset: A Post Hoc Analysis of a Randomized Clinical Trial. JAMA Ophthalmol 2024;142(12):1132-9.
37. Singh RK, Zhao Y, Elze T, et al. Polygenic Risk Scores for Glaucoma Onset in the Ocular Hypertension Treatment Study. JAMA Ophthalmol 2024;142(4):356-63.
38. Zhao H, Pasquale LR, Zebardast N, et al. Screening by Glaucoma Polygenic Risk Score to Identify Primary Open-Angle Glaucoma in Two Biobanks: an Interim Report. Investigative Ophthalmology & Visual Science 2024;65(7):1010-.
39. Prive F, Arbel J, Vilhjalmsson BJ. LDpred2: better, faster, stronger. Bioinformatics 2021;36(22-23):5424-31.
40. MacGregor S, Ong JS, An J, et al. Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma. Nat Genet 2018;50(8):1067-71.
41. Martschenko DO, Wand H, Young JL, Wojcik GL. Including multiracial individuals is crucial for race, ethnicity and ancestry frameworks in genetics and genomics. Nat Genet 2023;55(6):895-900.
42. Gao XR, Chiariglione M, Arch AJ. Whole-exome sequencing study identifies rare variants and genes associated with intraocular pressure and glaucoma. Nature Communications 2022;13(1):7376.

Figure legends

Figure 1. Receiver Operating Characteristic Curves Predicting POAG Using Polygenic Risk Scores

Receiver operating characteristic curves illustrating the predictive performance of polygenic risk scores (PRSs) for primary open-angle glaucoma (POAG) in (A) the UK Biobank cohort and (B) the Mexican American Glaucoma Genetic Study (MAGGS) cohort. In both cohorts, models including age, sex, and PRSs significantly improved the area under the curve (AUC) compared to the baseline models with only age and sex.

Abbreviations: IOP, intraocular pressure; RNFL, retinal nerve fiber layer; VCDR, vertical cup-to-disc ratio.

Figure 2. Association Between Polygenic Probability Risk Score Deciles and POAG Risk

Odds ratios (ORs) for primary open-angle glaucoma (POAG) across deciles of the polygenic probability risk score (PPRS) in (A) the UK Biobank cohort and (B) the Mexican American Glaucoma Genetic Study (MAGGS) cohort. The lowest decile (Decile 1) serves as the reference group. Error bars represent 95% confidence intervals (CIs). There is a clear trend of increasing POAG risk with higher PPRS deciles, demonstrating strong risk stratification capabilities of PPRS in both cohorts.

Supplementary Figure 1. Manhattan Plot Displaying the –log10(*P* values) for the Association Between RNFL Thickness and Genome-wide Genetic Variants

Solid and dotted horizontal lines represent genome-wide significant associations ($P < 5 \times 10^{-8}$) and suggestive associations ($P < 5 \times 10^{-5}$), respectively. Genetic variants are plotted by genomic position.

Supplementary Figure 2. SHAP Feature Importance for XGBoost Models

Bar charts illustrating feature importance, as evaluated via SHapley Additive exPlanations (SHAP) values, for the (A) UK Biobank and (B) Mexican American Glaucoma Genetic Study (MAGGS) cohorts. The y-axis displays the ranked features, and the x-axis shows the average impact on the model's output. PRS-POAG, PRS-IOP, PRS-VCDR, and PRS-RNFL were SBayesRC-derived.

457 Abbreviations: IOP, intraocular pressure; RNFL, retinal nerve fiber layer; VCDR, vertical cup-to-disc ra-
458 tio.
459
460

461 **Table 1. Sample Characteristics**
462

	UK Biobank (n = 324,713)		**MAGGS** (n = 4,549)	
	Controls (322,545)	Cases (2,168)	Controls (4,274)	Cases (275)
Age (yrs, SD)	57.1 (8.0)	63.0 (5.7)	56.1 (10.2)	66.4 (11.1)
Female (%)	54.3	48.4	59.2	52.4

463 Abbreviation: MAGGS, Mexican American Glaucoma Genetic Study; SD: standard deviation.
464
465

Table 2. Single-PRS Logistic Regression Results

(a) UK Biobank participants

	OR (95% CI)	*P* value	AUC (95% CI)
PRS-POAG-C+T	1.71 (1.64, 1.79)	8.71×10^{-141}	0.76 (0.75, 0.77)
PRS-POAG	2.14 (2.05, 2.24)	8.54×10^{-269}	0.79 (0.78, 0.80)
PRS-IOP	2.08 (2.00, 2.17)	1.04×10^{-248}	0.79 (0.79, 0.80)
PRS-VCDR	1.61 (1.54, 1.68)	9.50×10^{-107}	0.75 (0.74, 0.76)
PRS-RNFL	0.92 (0.89, 0.96)	2.52×10^{-4}	0.72 (0.71, 0.73)

(b) MAGGS[†] participants

	OR (95% CI)	*P* value	AUC (95% CI)
PRS-POAG-C+T	1.64 (1.44, 1.88)	1.68×10^{-13}	0.76 (0.73, 0.79)
PRS-POAG-MTAG	1.44 (1.24, 1.60)	1.17×10^{-7}	0.75 (0.72, 0.78)
PRS-POAG	1.82 (1.59, 2.08)	2.79×10^{-18}	0.77 (0.74, 0.80)
PRS-IOP	1.45 (1.27, 1.65)	1.66×10^{-8}	0.75 (0.72, 0.78)
PRS-VCDR	1.90 (1.66, 2.17)	1.39×10^{-20}	0.78 (0.75, 0.81)
PRS-RNFL	0.85 (0.74, 0.96)	0.011	0.75 (0.72, 0.77)

Results of single-PRS logistic regression analyses, adjusted for age and sex. [†]Unrelated MAGGS participants (n = 4,108) were used. PRS-POAG, PRS-IOP, PRS-VCDR, and PRS-RNFL were SBayesRC-derived. To prevent potential data leakage, MTAG-derived PRS (based on Craig et al.'s summary statistics, which used UK Biobank data) was not applied to UK Biobank participants. PRS-POAG-MTAG was exclusively applied to MAGGS study participants.

Abbreviations: C+T, clumping and thresholding; IOP, intraocular pressure; MAGGS, Mexican American Glaucoma Genetic Study; OR, odds ratio; PRS, polygenic risk score; POAG, primary open-angle glaucoma; RNFL, retinal nerve fiber layer thickness; VCDR, vertical cup-to-disc ratio.

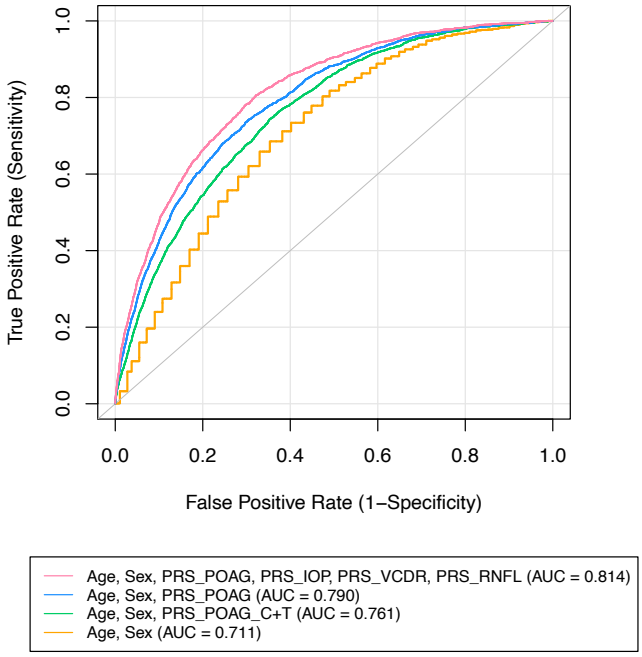
Table 3. Multiple Logistic Regression Results

	UK Biobank		MAGGS[†]	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
PRS-POAG	1.59 (1.51, 1.67)	1.23×10^{-74}	1.40 (1.18, 1.65)	7.63×10^{-5}
PRS-IOP	1.63 (1.55, 1.71)	5.37×10^{-89}	1.15 (0.99, 1.35)	0.07
PRS-VCDR	1.28 (1.23, 1.34)	1.99×10^{-27}	1.64 (1.42, 1.90)	2.04×10^{-11}
PRS-RNFL	0.94 (0.90, 0.98)	0.002	0.85 (0.74, 0.97)	0.016

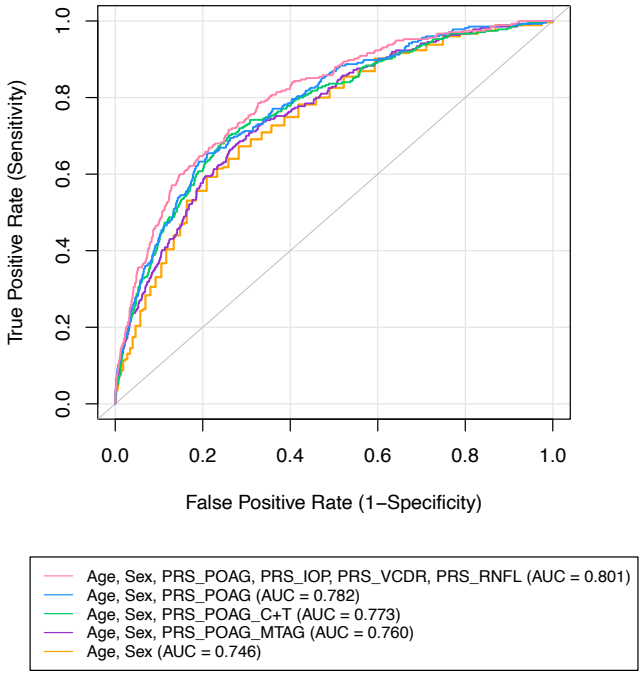
Results of multiple logistic regression analyses, adjusted for age and sex, assessing the combined association of polygenic risk scores (PRSs) with primary open-angle glaucoma. [†]Unrelated MAGGS participants (n = 4,108) were used. PRS-POAG, PRS-IOP, PRS-VCDR, and PRS-RNFL were SBayesRC-derived.

Abbreviations: IOP, intraocular pressure; MAGGS, Mexican American Glaucoma Genetic Study; OR, odds ratio; PRS, polygenic risk score; POAG, primary open-angle glaucoma; RNFL, retinal nerve fiber layer thickness; VCDR, vertical cup-to-disc ratio.

Figure 1. Receiver Operating Characteristic Curves Predicting POAG Using Polygenic Risk Scores

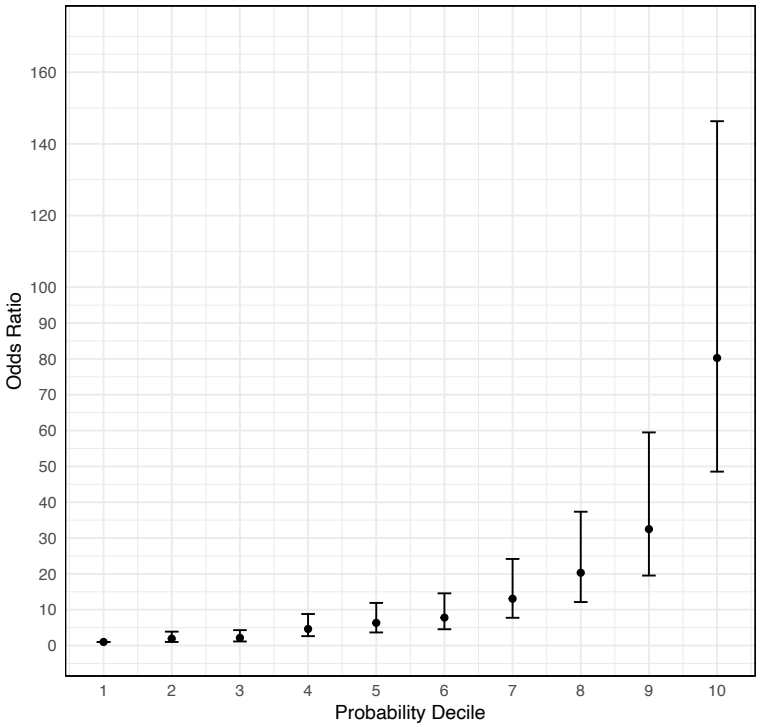


(a) UK Biobank participants

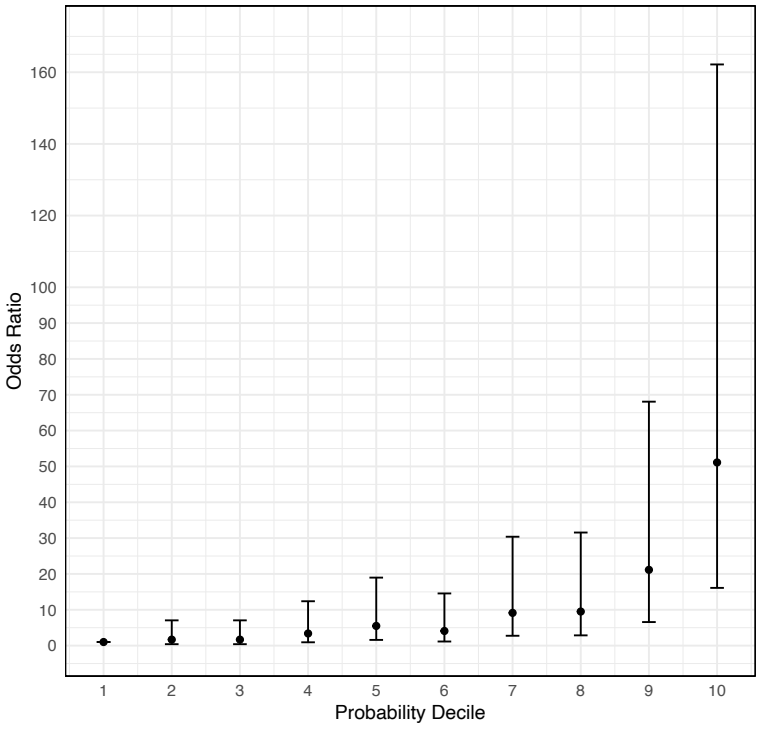


(b) MAGGS participants

Figure 2. Association Between Polygenic Probability Risk Score Deciles and POAG Risk



(a) UK Biobank participants



(b) MAGGS participants

Supplementary Table 1. Variance Inflation factors for polygenic risk scores in our logistic regression models

(a) UK Biobank

Age	Sex	PRS-POAG	PRS-IOP	PRS-VCDR	PRS-RNFL
1.002	1.002	1.347	1.240	1.112	1.003

(b) MAGGS[†]

Age	Sex	PRS-POAG	PRS-IOP	PRS-VCDR	PRS-RNFL
1.024	1.006	1.555	1.371	1.166	1.004

Displayed are the variance inflation factors for the polygenic risk scores (PRSs) used in our logistic regression analyses of a) UK Biobank; b) MAGGS. [†]Unrelated MAGGS participants were used. PRS-POAG, PRS-IOP, PRS-VCDR, and PRS-RNFL were SBayesRC-derived.

Abbreviations: IOP, intraocular pressure; MAGGS, Mexican American Glaucoma Genetic Study; PRS, polygenic risk score; POAG, primary open-angle glaucoma; RNFL, retinal nerve fiber layer thickness; VCDR, vertical cup-to-disc ratio.

Supplementary Table 2. Number of POAG Cases in Each Decile Category

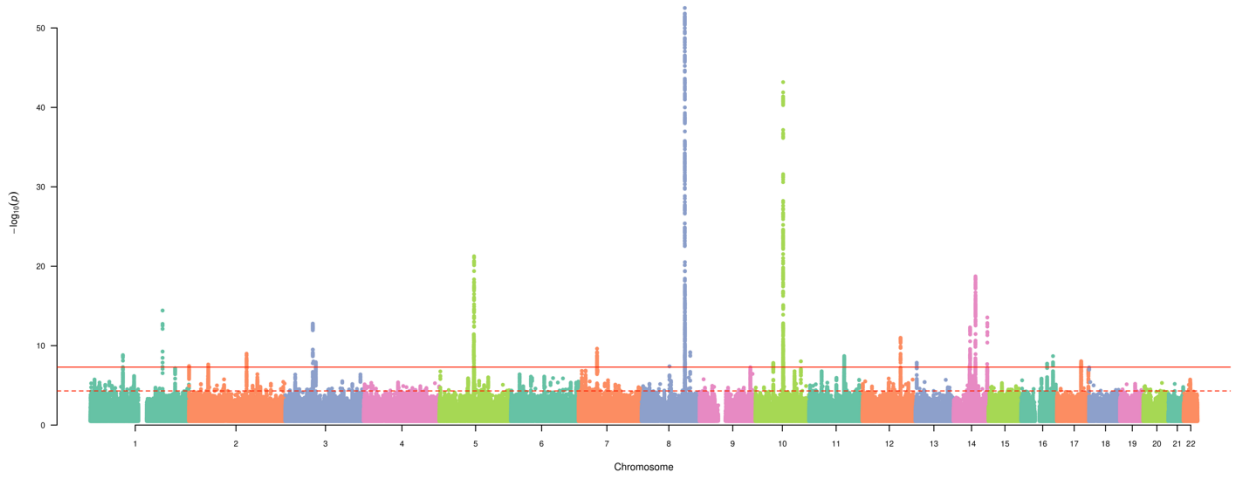
(a) UK Biobank

	Predicted Probability Risk Score Decile									
	1	2	3	4	5	6	7	8	9	10
POAG, n (%)	13 (0.60)	25 (1.15)	28 (1.29)	60 (2.77)	82 (3.78)	101 (4.66)	169 (7.80)	262 (12.08)	417 (19.23)	1011 (46.63)

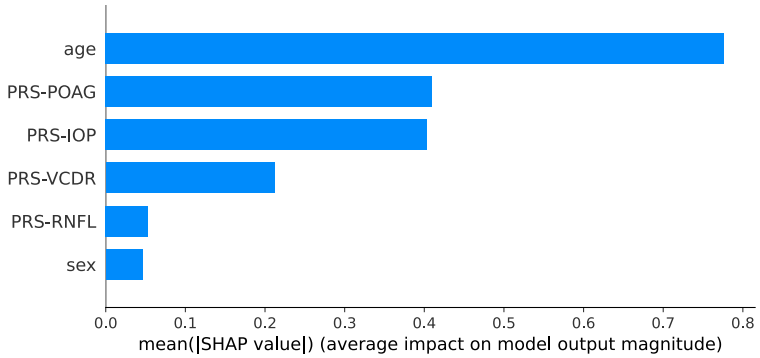
(b) MAGGS

	Predicted Probability Risk Score Decile									
	1	2	3	4	5	6	7	8	9	10
POAG, n (%)	3 (1.09)	5 (1.82)	5 (1.82)	10 (3.64)	16 (5.82)	12 (4.36)	26 (9.45)	27 (9.82)	56 (20.36)	115 (41.82)

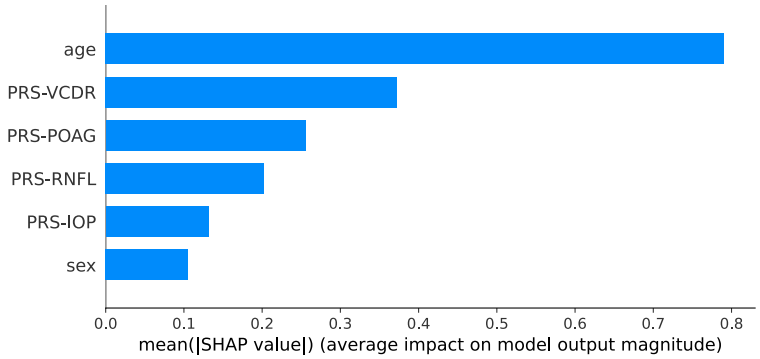
Supplementary Figure 1. Manhattan Plot Displaying the –log10(*P* values) for the Association Between RNFL Thickness and Genome-wide Genetic Variants



Supplementary Figure 2. SHAP Feature Importance for XGBoost Models



(a) UK Biobank participants



(b) MAGGS participants