Multimodal Prediction of Primary Open-Angle Glaucoma Using Polygenic Risk Scores and Clinical Features in a High-Risk African Ancestry Cohort

Yan Zhu[1], Aude Benigne Ikuzwe Sindikubwabo[2], Yuki Bradford[3], Lannawill Caruth[2], Rebecca Salowe[1], Kenneth Pham[1], Laxmi Moksha[1], Vrathasha Vrathasha[1], Roy Lee[1], Mina Halimitabrizi[1], Isabel Di Rosa[1], Leila Ghaffari[1], Jie He[1], Marylyn D. Ritchie[3], Joan M. O'Brien[1]*, Shefali Setia-Verma[2]*

[1]Center for Genetics of Complex Disease, Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA 19104, USA
[2]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[3]Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA
Present address: 25 N 38th Street, Philadelphia, PA 19104, USA
*Lead contact
*Correspondence: joan.o'brien@pennmedicine.upenn.edu, Shefali.SetiaVerma@pennmedicine.upenn.edu

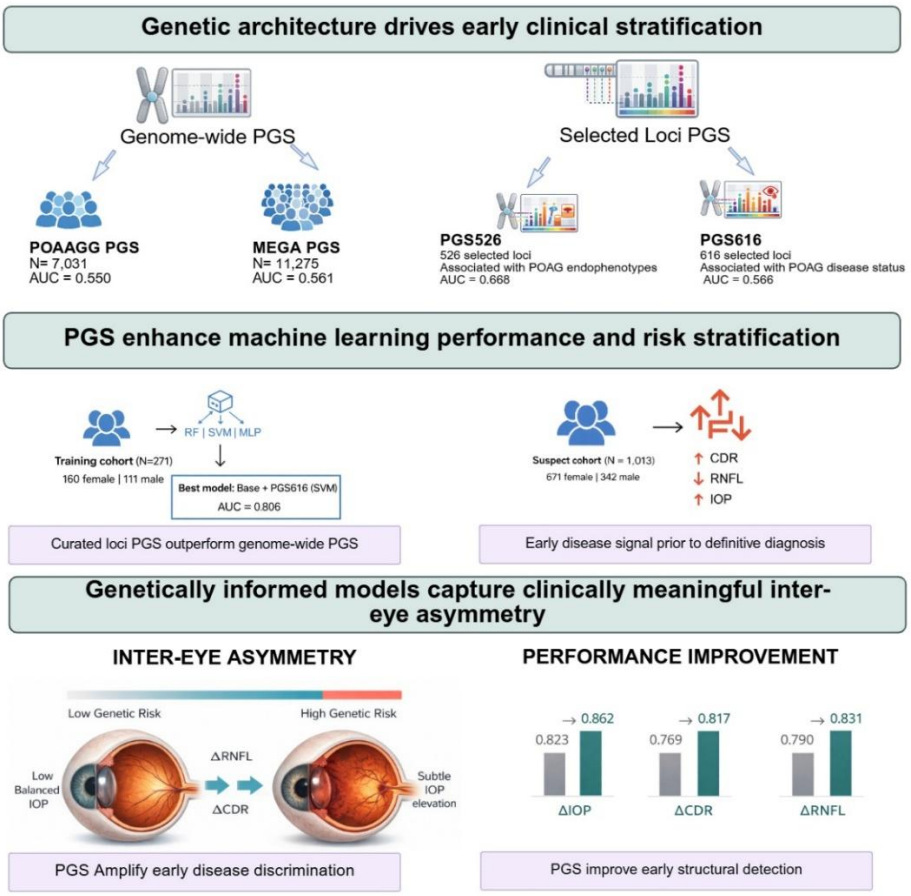
HIGHLIGHTS

- Ancestry-matched polygenic scores improve POAG risk prediction in African ancestry populations
- Curated loci-based PGS outperform genome-wide scores in machine learning models
- Genetic risk integrates with minimal demographics to achieve AUC up to 0.806
- Predicted risk aligns with optic nerve damage and inter-eye asymmetry in suspects

SUMMARY

We developed four polygenic risk scores (PGS) for primary open-angle glaucoma (POAG), which is the leading cause of irreversible blindness worldwide and remains undiagnosed in over half of patients. We constructed two genome-wide PGS using genome wide association study from African ancestry subjects: 1) the Primary Open-Angle African Ancestry Glaucoma Genetics (POAAGG) study (N = 7,031; POAAGG PGS) and an African ancestry GWAS (N = 11,275; MEGA PGS). We also derived two selected loci PGS from six multi-ancestry glaucoma GWAS and weighted these scores using African ancestry effect sizes (PGS616 and PGS526). In an independent training cohort (N = 271), the curated loci-based score PGS526 demonstrated the strongest standalone performance (mean AUC = 0.668), outperforming the genome-wide PGS constructed using PRS-CS. Integration with baseline demographic features (age and gender) further improved prediction, with the base + PGS616 model achieving a peak AUC of 0.806 with support vector machine model. Clinical enrichment in an independent suspect cohort (N = 1,013) showed that higher predicted genetic risk was significantly associated with elevated intraocular pressure, larger cup-to-disc ratio, and thinner retinal nerve fiber layer, which are all POAG diagnostic features. Leveraging inter-eye asymmetry, PGS further enhanced early disease discrimination, improving AUC from 0.823 to 0.862 for ΔIOP, from 0.769 to 0.817 for ΔCDR, and from 0.790 to 0.831 for ΔRNFL. These results demonstrate that PGS enhances prediction in phenotype-rich settings and enables accurate risk stratification in deep phenotype-limited cohorts, supporting earlier glaucoma detection.

GRAPHICAL ABSTRACT



KEYWORDS

primary open-angle glaucoma, machine learning, multimodal modeling, polygenic risk score, genetic risk score, phenotype integration, African ancestry, risk prediction, ocular phenotypes, precision medicine

INTRODUCTION

Glaucoma is the leading cause of irreversible blindness worldwide, affecting over 70 million individuals and projected to rise dramatically with the aging population.[1,2] Primary open-angle glaucoma (POAG), the most prevalent form of this disease, accounts for approximately 74% of all diagnosed glaucoma cases.[3] POAG is characterized by progressive degeneration of the optic nerve, often accompanied by elevated intraocular pressure (IOP), thinning of the retinal nerve fiber layer (RNFL), structural changes in the optic disc such as increased cup-to-disc ratio (CDR), and corresponding loss of visual field.[4-7] Importantly, individuals of African ancestry are disproportionately affected by POAG, experiencing earlier onset, more rapid progression, and up to five-fold higher risk of blindness compared to those of European ancestry.[8,9] Despite these known disparities, early detection and risk stratification in this high-risk population remain inadequate, with over 50% of cases remaining undiagnosed.[10,11] This underscores an urgent need for precision tools that can support earlier and more accurate diagnosis.[12]

In recent years, advances in genomics have enabled the development of polygenic risk scores (PGS), which aggregate the effects of multiple genetic variants to estimate an individual's inherited susceptibility to complex diseases.[13-15] These tools have shown promise in a range of common diseases, including type 2 diabetes,[16,17] coronary heart disease,[18,19] and breast cancer,[20-22] and are beginning to gain traction in ophthalmic diseases with a strong genetic component, such as POAG.[23,24] However, there are several major limitations to existing PGS for POAG. First, the majority of existing risk scores are trained in predominantly European ancestry cohorts and often lack transferability to other ancestries.[25,26] Therefore,

PGS built from carefully selected, biologically validated, and ancestry-matched variants may offer a more interpretable and potentially more reliable approach for glaucoma risk prediction.[22,27,28] Additionally, while existing PGS models reflect a person's inherited susceptibility to the disease, these risk scores do not consistently integrate current disease status or severity. In contrast, routinely measured clinical features, such as IOP, RNFL thickness, and CDR, are valuable diagnostic indicators that capture disease after structural or functional damage has already occurred.[29] To solve this, integrating these complementary data types could yield more accurate and equitable risk prediction models for POAG.[25]

Artificial intelligence (AI), particularly supervised machine learning (ML), offers a powerful framework to harness the complexity of multidimensional health data.[30,31] In contrast to traditional regression based risk models, ML algorithms can capture non-linear interactions among diverse inputs and automatically learn relevant feature representations.[32,33] In ophthalmology, ML techniques have been successfully employed to analyze fundus photographs, segment optic nerve head structures, and quantify RNFL thickness using OCT scans.[34-36] These models can detect structural changes that are often imperceptible to the human eye, providing a critical advantage in screening and disease monitoring.[37-39] However, relatively few studies have examined the utility of integrating genetic risk, particularly PGS, with ocular phenotypes in large, ancestrally diverse POAG cohorts.[40,41] Moreover, no prior work has systematically compared multiple ML models across different combinations of phenotypic and genetic features in the most overaffected African ancestry cohort. This gap highlights the need to evaluate whether ancestry-matched genetic scores, especially loci-focused PGS, can provide added predictive value beyond standard phenotypes in the context of ML based models.

In this study, we developed ML models incorporating PGS for POAG prediction in the Primary Open-Angle African Ancestry Glaucoma Genetics (POAAGG) cohort, which is a large and deeply phenotyped African ancestry dataset.[9,42,43] We constructed four complementary PGS, including two genome-wide scores (POAAGG PGS and MEGA PGS) that capture broad polygenic effects, and two selected list of loci-based scores (PGS526 and PGS616) derived from curated multi-ancestry glaucoma loci and weighted using African ancestry effect sizes. Using these PGS, we trained three supervised learning models in a POAAGG training cohort of 271 participants. These models included multilayer perceptron (MLP),[44] random forest (RF),[45-47] and support vector machine (SVM),[48-50] and were evaluated for performance and generalizability in an independent testing cohort of 1,013 participants. Each PGS was integrated into the ML models alongside demographic and clinical features to assess whether genetic information could improve predictive accuracy across the three models.

RESULTS

Study Design and Data Integration Framework

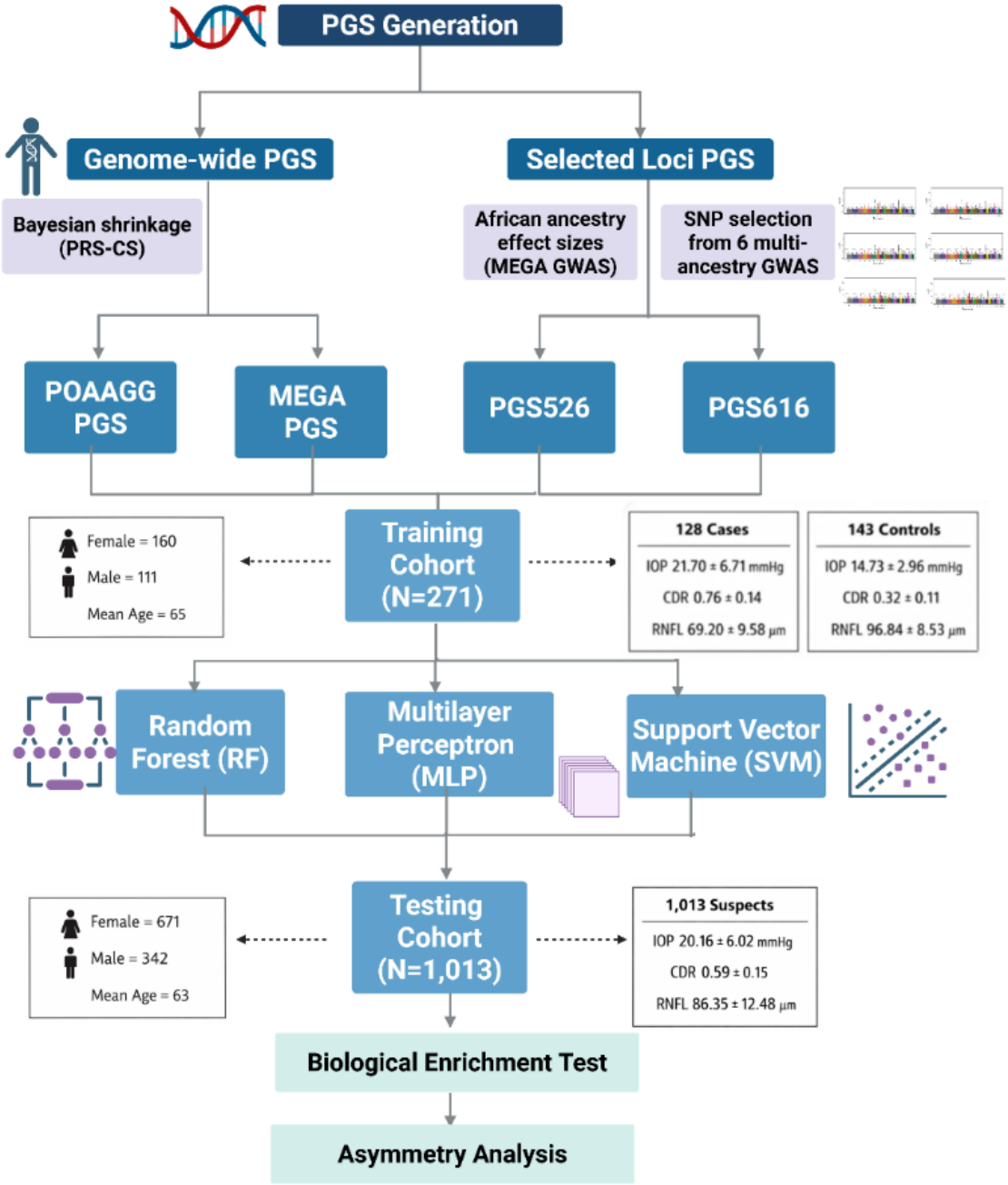
We developed a structured pipeline that integrates PGS generation, model training, and biological validation (Figure 1). Four PGS were constructed using two complementary strategies: genome-wide PGS derived from African ancestry GWAS via PRS-CS (POAAGG PGS and MEGA PGS) and loci-based PGS built from six multi-ancestry glaucoma GWAS[14,51-55] weighted by African ancestry-specific effect sizes (PGS526 and PGS616).

Model training was conducted in a training cohort of 271 subjects (128 cases, 143 controls; mean age = 65 years; 160 females, 111 males). The training set showed clear phenotypic separation between cases and controls. Cases exhibited significantly higher baseline IOP (21.70 ± 6.71 mmHg vs. 14.73 ± 2.96 mmHg; Welch's t-test $p = 5.44 \times 10^{-21}$), larger baseline CDR (0.76 ± 0.14 vs. 0.32 ± 0.11; $p = 4.86 \times 10^{-74}$), and thinner baseline RNFL (69.20 ± 9.58 μm vs. 96.84 ± 8.53 μm; $p = 4.01 \times 10^{-59}$).

To evaluate model generalizability, predictions were tested in an independent suspect cohort of 1,013 subjects (mean age = 63 years; 671 female, 342 male).[56] This group represents individuals with early or subclinical glaucoma features, exhibiting intermediate clinical profiles for IOP (20.16 ± 6.02 mmHg), CDR (0.59 ± 0.15), and RNFL (86.35 ± 12.48 μm). Because definitive case and control labels are unavailable in this suspect population, model performance was assessed using biological enrichment analyses, testing whether higher predicted genetic risk aligned with established ocular phenotypes (IOP, CDR, RNFL thickness). In addition, given that early POAG frequently presents with inter-eye asymmetry, we further evaluated whether PGS related models could capture clinically meaningful asymmetry patterns, providing an independent and biologically grounded validation of early disease risk. This design provided a

biologically grounded framework to link ancestry informed genetic risk with clinically measurable glaucoma phenotypes.

Figure 1. Overview of Study Design, Cohorts, and Phenotypic Characteristics



Two complementary strategies were used to generate PGS: genome-wide PGS from African ancestry GWAS using PRS-CS (POAAGG PGS and MEGA PGS), and selected loci-based PGS from six multi-ancestry glaucoma GWAS weighted by meta-analysis effect sizes (PGS526 and PGS616). These four PGS were applied to a training cohort (N = 271; 128 cases, 143 controls) for model development using RF, MLP, and SVM. Biological relevance was tested in an independent suspect cohort (N = 1,013) through enrichment analysis of IOP, CDR, and RNFL traits. Asymmetry analysis was also conducted on the training cohort and was tested on 1,013 suspects.

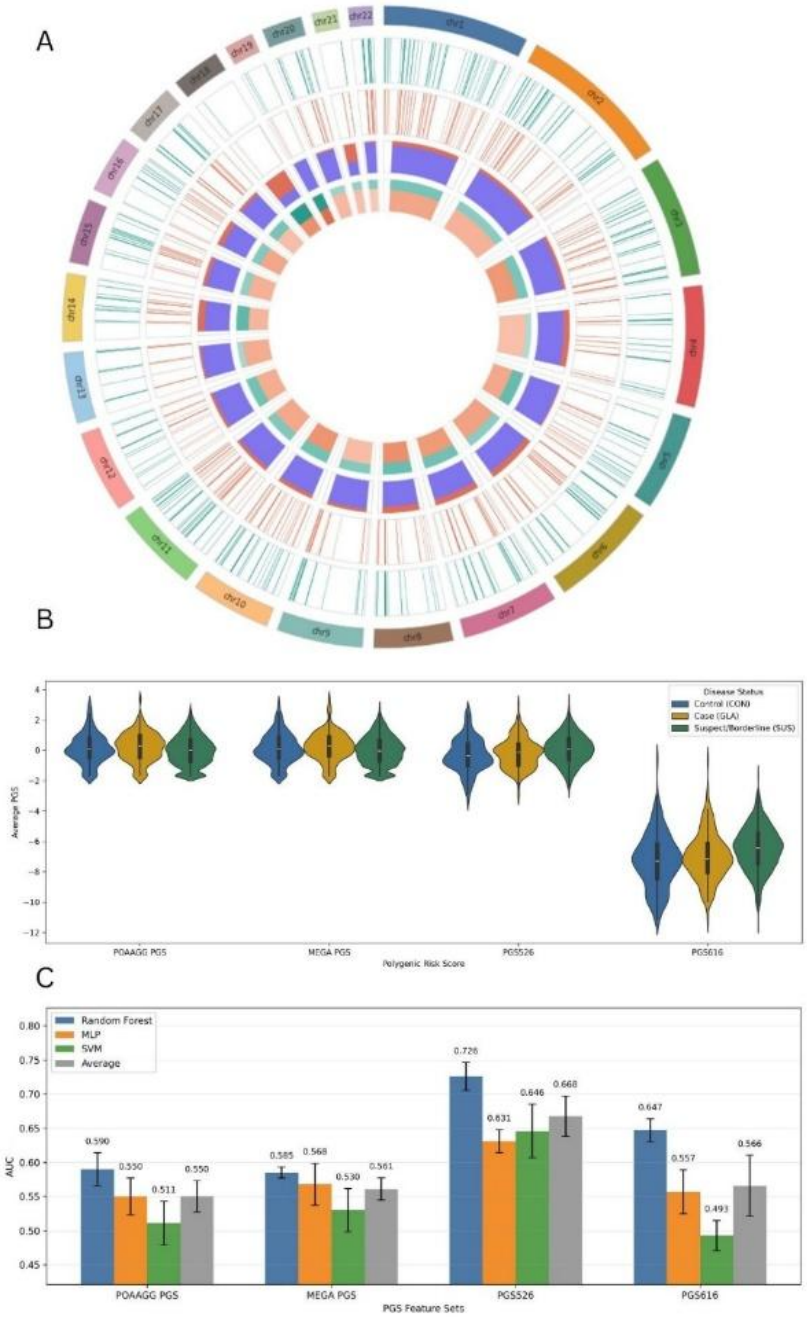
PGS-Based Prediction Models Reveal Limited but Interpretable Signals

To capture the inherited component of POAG risk, we constructed four PGS using two complementary strategies (See *Star Methods*). The first strategy employed a genome-wide approach that aggregates risk across millions of common variants. We generated two genome-wide scores: (1) The POAAGG PGS, based solely on GWAS summary statistics from the POAAGG cohort (N = 7,031),[9,57] and (2) The MEGA PGS, based on a larger African ancestry GWAS meta-analysis (MEGA, N = 11,275) combining three cohorts: POAAGG, the African Descent and Glaucoma Evaluation Study (ADAGES), and Genetics of Glaucoma in People of African Descent (GGLAD).[14,15] The second strategy focused on biologically curated, loci-based scores. We systematically extracted genome-wide significant SNPs from six large glaucoma GWAS across multiple ancestries (*Craig et al., 2019;*[51] *Han et al., 2023;*[54] *Gharahkhani et al., 2021;*[53] *Faro et al., 2024;*[52] *Chang-Wolf et al., 2024;*[55] *Verma et al., 2024*[14]) and weighted them using effect sizes derived from the African ancestry MEGA GWAS, resulting in two distinct curated scores: (3) PGS526, which emphasizes variants associated with IOP and optic nerve structure (CDR, RNFL), and (4) PGS616, which includes broader disease associated loci. SNP selection, weighting method, filtering, LD pruning, and normalization procedures for each score are detailed in the *STAR Methods*. A complete list of SNPs and weights for PGS526 and PGS616 is provided (Table S1).

To assess genomic coverage, we examined the chromosomal distribution of variants included in the two loci-based PGS, highlighting their overlap and complementarity (Figure 2A). The circos plot shows that PGS526 and PGS616 comprise partially overlapping but distinct sets of loci distributed across all autosomes, consistent with their design to capture both shared and complementary genetic contributions to glaucoma risk. We next evaluated how these variant-level differences translated to score-level behavior across cohorts. As shown in Figure 2B, standardized PGS distributions exhibited a clear gradient across disease groups, with higher scores in POAG cases, and lower scores in controls for all four PGS, supporting their biological relevance. Summary statistics for PGS distributions across disease groups are also included (Table S2).

When used as standalone features, all four PGS demonstrated moderate but interpretable predictive performance in the training cohort (Figure 2C). Average AUC values across the three models (RF, MLP, SVM) showed a consistent ranking among the PGS. PGS526 achieved the highest mean AUC (0.668 ± 0.067), followed by PGS616 (0.566 ± 0.092), MEGA PGS (0.561 ± 0.048), and POAAGG PGS (0.550 ± 0.063). Detailed performance metrics are listed (Table S3), including AUC, accuracy, F1 score, sensitivity, and specificity for all three models. Although the absolute values were moderate, the relative consistency across models indicates that each PGS captures reproducible genetic signal. While genome-wide PGS capture broad heritable risk, curated PGS such as PGS616 and PGS526 provide clearer separation between glaucoma cases and controls.

Figure 2. Performance and Characteristics of Four PGS in African Ancestry Glaucoma Cohorts



(A) Five-layer circos visualization of the chromosomal distribution and composition of variants included in the loci-based PGS. The outer ideogram represents chromosomes 1–22. The two outer tick rings mark the genomic positions of variants included in PGS526 (teal) and PGS616 (coral), illustrating per-chromosome variant distribution across the genome. The third ring summarizes per-chromosome loci counts, stacked by category: loci unique to PGS526 (teal), loci shared between PGS526 and PGS616 (violet), and loci unique to PGS616 (coral). The innermost split ring summarizes the fraction of variants with positive effect sizes ($\beta > 0$) per chromosome, shown separately for PGS526 (teal) and PGS616 (coral), and is included as a descriptive summary of effect direction rather than effect magnitude.

(B) Violin plots showing standardized distributions of all four PGS (POAAGG PGS, MEGA PGS, PGS526, PGS616) across training cases (orange), training controls (blue), and suspect individuals (green).
(C) Predictive performance of each PGS across three machine learning models (RF, SVM, MLP) in the training cohort (N = 271). Bars show mean AUC ± 95%CI, with gray bars representing the average across all three models.

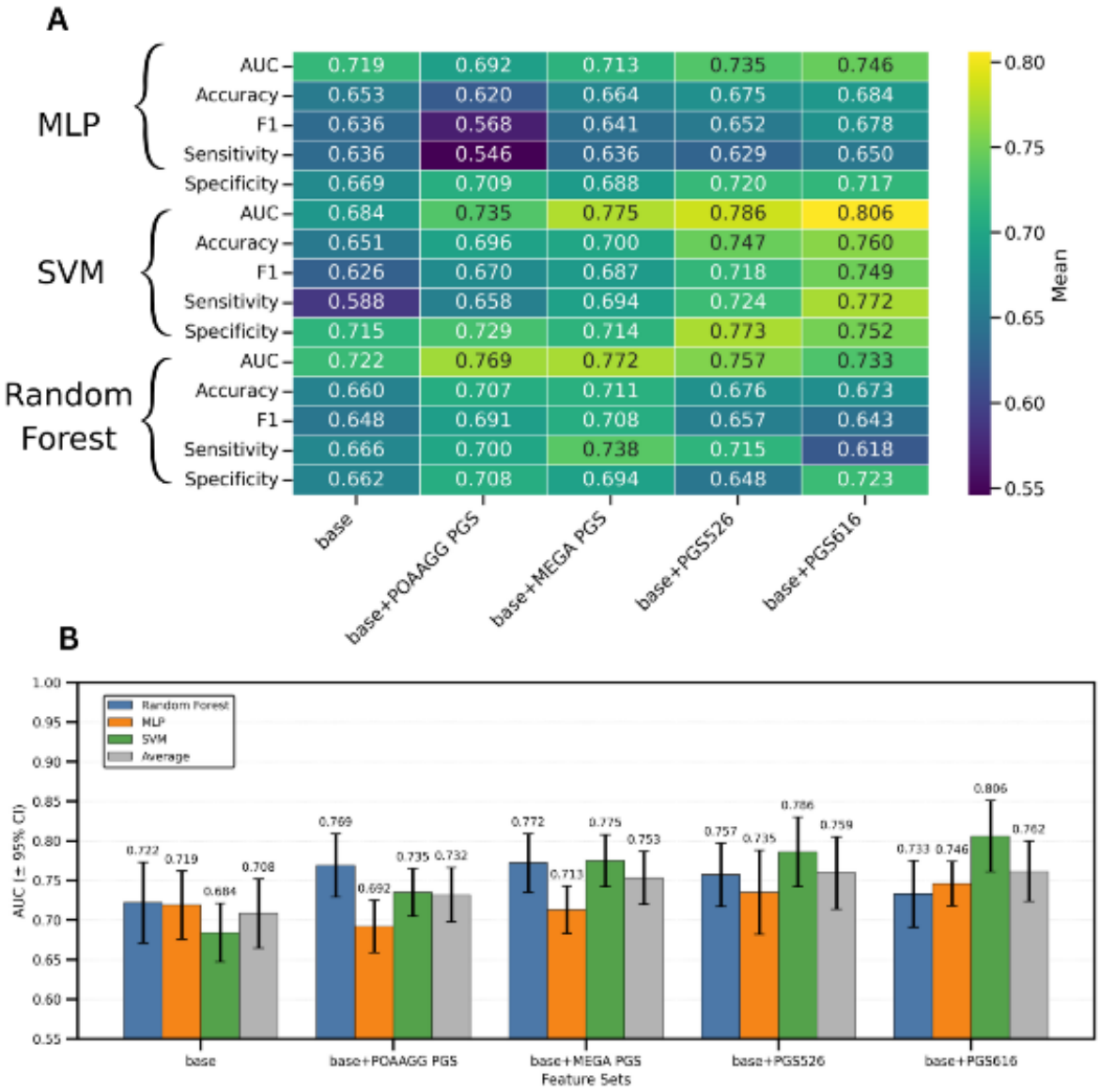
Training Cohort Analysis Reveals Consistent Performance Gains from Polygenic Risk Integration

In the training cohort (N = 271), we evaluated the predictive performance of three supervised models (RF, SVM, and MLP) with five feature sets: a baseline model with age and gender only, and four extended models that each incorporated one of the finalized PGS. Ancestry principal components (PCs) were not included in the baseline model, as the training cohort was genetically homogeneous and inclusion of PCs increased the risk of overfitting. However, we still evaluated models incorporating age, gender, and the first 20 ancestry PCs for completeness and comparison (Figure S1).

Across all three classifiers, integrating PGS consistently improved predictive performance over the baseline model, as reflected across multiple metrics (Figure 3A). For the SVM classifier, PGS616 produced the strongest overall performance, increasing AUC from 0.684 (baseline) to 0.806, accompanied by gains in accuracy (0.651 vs. 0.760), F1 score (0.626 vs. 0.749), sensitivity (0.588 vs. 0.772), and specificity (0.715 vs. 0.752). RF models showed more modest but consistent improvements, with AUC increasing from 0.722 at baseline to 0.733 with PGS616 and to 0.757 with PGS526, alongside improvements in accuracy and F1 score. In MLP models, PGS integration primarily improved discrimination and balance between sensitivity and specificity, with PGS616 yielding the highest AUC (0.746) and F1 score (0.678) relative to the baseline model. Across all three models, the selected loci-based PGS526 and PGS616 outperformed genome-wide PRS-CS scores (POAAGG PGS and MEGA PGS), demonstrating more consistent gains across complementary evaluation metrics. Model-specific performance metrics for all feature sets and models are also reported (Table S4).

To evaluate overall trends independent of classifier choice, we summarized performance by averaging AUC across the three models for each feature set (grey bar, Figure 3B). Relative to the baseline model (mean AUC = 0.708), all PGS-enhanced models showed improved performance. The largest average gain was observed for PGS616, which increased mean AUC to 0.762 (+0.054), followed closely by PGS526 (mean AUC = 0.759, +0.051). Genome-wide PGS also provided measurable but smaller improvements, with mean AUCs of 0.753 for MEGA PGS (+0.045) and 0.732 for POAAGG PGS (+0.024). These results demonstrate that incorporating ancestry-matched PGS consistently enhances predictive accuracy beyond demographic features alone, with curated, loci-based scores providing the greatest incremental benefit in this African ancestry training cohort.

Figure 3. Model performance across feature sets in the training cohort (N = 271).



(A) Heatmap summarizing performance metrics (AUC, accuracy, F1, sensitivity, and specificity) across three supervised machine learning models (MLP, SVM, and RF) under five feature sets: base model (age + gender) and versions incorporating each of the four PGS (POAAGG PGS, MEGA PGS, PGS526, and PGS616). Warmer colors indicate higher mean performance values. All models show improved metrics after adding any PGS, with the most consistent gains observed for the curated PGS616 feature set.

(B) Bar plots showing the mean area under the ROC curve (AUC ± 95%CI) for each classifier and the overall average across feature sets. Integrating genetic risk information increased performance for all models compared with the demographic-only baseline. The PGS616 produced the highest overall AUC (0.733–0.806 range), followed by PGS526, demonstrating that loci-based PGS substantially enhance glaucoma risk prediction in the training cohort.

Testing Cohort Analysis: Clinical Enrichment Supports Biological Relevance of Predicted Risk in an independent suspect cohort

To assess the clinical utility of our trained models, we applied the three classfiers (RF, MLP, SVM) to an independent test cohort of 1,013 POAAGG suspects. For primary analyses, we focused on the base + PGS616 feature set, which achieved the strongest and most consistent performance across models in the training cohort and therefore served as the primary risk model for downstream biological validation. Given the comparable performance of PGS526 in the training analyses, we additionally evaluated base + PGS526 models in the suspect cohort as a robustness analysis, and these results are presented in supplemental documents (Figure S2).

Using this feature set (base+PGS616), each classifier generated predicted glaucoma risk scores for the individuals in suspect cohort. Specifically, we tested whether higher predicted POAG risk scores were associated with established clinical indicators of glaucomatous damage, including elevated IOP, enlarged CDR, and thinning of the RNFL. Importantly, these ocular phenotypes (IOP, CDR, RNFL thickness) were not included as input features when applying the models to the suspect cohort and were used exclusively for post hoc biological validation. Because definitive case-control labels are unavailable in this population, traditional classification metrics (e.g., AUC) could not be computed; instead, we assessed Pearson correlations and linear regression analyses using continuous predicted risk scores at the individual level, while stratifying risk into quintiles (1–5) for visualization of enrichment patterns across increasing genetic risk.

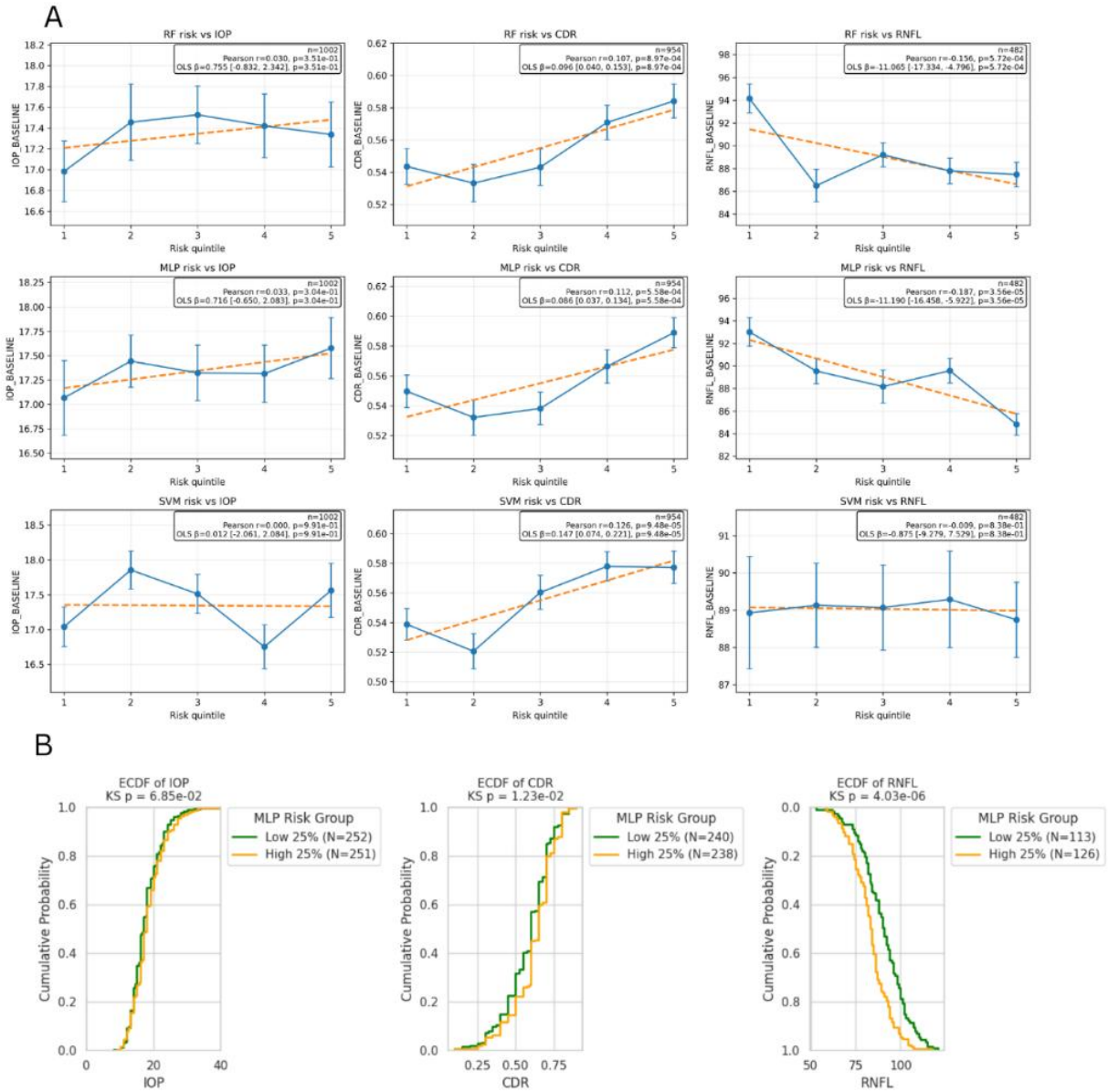
Across all three models, predicted glaucoma risk showed consistent enrichment for structural markers of disease severity, particularly CDR and RNFL thickness (Figure 4A). Using continuous individual-level risk scores, RF predictions were significantly positively correlated with CDR ($r = 0.107$, $p = 8.97\times10^{-4}$) and negatively correlated with RNFL thickness ($r = -0.156$, $p = 5.72\times10^{-4}$), while the association with baseline IOP was weak and not statistically significant ($r = 0.030$, $p = 0.351$).

The SVM model demonstrated a significant association with CDR ($r = 0.126$, $p = 9.48\times10^{-5}$), indicating enrichment of optic nerve structural damage with increasing predicted risk; however, associations with IOP ($r = 0.000$, $p = 0.991$) and RNFL thickness ($r = -0.009$, $p = 0.838$) were not statistically significant, suggesting reduced concordance with RNFL based structural severity in the suspect cohort.

Among the three classifiers, the MLP model demonstrated the most consistent and biologically coherent pattern across phenotypes. MLP predicted risk was significantly positively correlated with CDR ($r = 0.112$, $p = 5.58\times10^{-4}$) and showed the strongest negative association with RNFL thickness ($r = -0.187$, $p = 3.56\times10^{-5}$), consistent with progressive optic nerve damage at higher genetic risk levels. Although the association with IOP did not reach statistical significance ($r = 0.033$, $p = 0.304$), the monotonic enrichment across quintiles was directionally consistent. In all, these findings indicate that the MLP model provides the most stable stratification of structural glaucoma severity within the suspect cohort.

Since MLP gives strongest treadline, we employed this model and compared the distributions of key ocular phenotypes (IOP, CDR, and RNFL thickness) between individuals in the top and bottom 25% of predicted risk using empirical cumulative distribution functions (ECDFs) in Figure 4B. Among these, CDR and RNFL thickness showed the strongest and most clinically meaningful differences. Individuals in the high-risk group had a clear rightward shift in CDR values (Kolmogorov–Smirnov $p = 0.012$), indicating more pronounced optic nerve cupping, and a leftward shift in RNFL thickness ($p = 4.03\times10^{-6}$), consistent with earlier retinal nerve fiber thinning. To emphasize this pattern, RNFL ECDFs were visualized with the Y-axis inverted, making thinning in high-risk individuals easier to visualize. In contrast, IOP showed only a modest, non-significant trend ($p = 0.069$), which is biologically plausible given that glaucomatous structural damage, particularly in normal-tension glaucoma, may occur in the absence of sustained IOP elevation. This observation may also be attrubutable to the treatment of this cohort with IOP lowering drugs.

Figure 4. Predicted glaucoma risk scores from the PGS616 model correlate with clinical phenotypes in the suspect cohort (N=1,013).



(A) Associations between predicted glaucoma risk scores and baseline clinical indicators of disease severity in 1,013 suspect individuals. Panels show IOP, CDR, and RNFL thickness across three classifiers: RF (top row), MLP (middle row), and SVM (bottom row). Each model was trained using the base + PGS616 feature set. Predicted risk scores were stratified into quintiles (1–5) for visualization, and mean ± SEM of each phenotype within quintile are shown. Dashed lines represent linear regression fits. Pearson correlation coefficients (r) and corresponding p-values, as well as ordinary least squares (OLS) regression statistics, were calculated using continuous individual-level predicted risk scores. Across models, higher predicted risk was associated with larger CDR and thinner RNFL, consistent with structural glaucomatous damage, with the MLP model demonstrating the most consistent monotonic enrichment across quintiles.

(B) ECDF comparing key ocular traits between individuals in the top and bottom 25% of MLP-predicted glaucoma risk. High-risk individuals (orange) showed higher IOP, greater CDR, and thinner RNFL (for CDR and RNFL, Kolmogorov–Smirnov $p < 0.05$). These enrichment patterns confirm that PGS informed model predictions align with clinical markers of disease, validating their biological relevance in an unlabeled, real-world suspect population.

Phenotypic Asymmetry as a Clinically Meaningful Marker, Enhanced by Genetic Risk Scores

In glaucoma diagnosis and risk stratification, asymmetry between the two eyes is considered an early sign of possible disease.[58,59] In early-stage POAG, one eye may begin to show structural or functional changes while the other remains within normal limits.[59,60] This asymmetry can manifest in IOP, CDR, or RNFL thinning, often before the disease becomes symmetric or clinically definitive. Moreover, since absolute values of these phenotypes were used to define training cases and controls (see Star Method), to avoid circularity, we instead derived inter-eye differences (ΔIOP, ΔCDR, ΔRNFL) between two eyes. These asymmetry features reflect eye-specific deviation and are statistically independent of cohort selection criteria. As illustrated in Figure 5A, inter-eye differences can reveal early glaucomatous changes, even when absolute measurements fall within normal clinical ranges. These features carry strong clinical relevance and are especially valuable for detecting early or unilateral disease.

We evaluated the predictive value of asymmetry features alone and in combination with each of the four PGS (POAAGG PGS, MEGA PGS, PGS526, PGS616). Three machine learning classifiers (RF, SVM, MLP) were trained using bootstrap resampling, and model performance was evaluated using multiple metrics (Table S5). Across all feature sets, Random Forest showed the most stable and highest performance and is thus presented in Figure 5B; results from SVM and MLP are provided in Figures S3 and S4.

Asymmetry features alone were moderately predictive, with ΔIOP yielding the strongest standalone performance among the three phenotypes (Figure 5B). Importantly, adding PGS consistently improved model discrimination across all asymmetry features. For ΔIOP, adding PGS increased AUC from 0.823 for ΔIOP-only models to as high as 0.862 when combined with PGS616. Similar gains were observed for ΔCDR and ΔRNFL, with PGS models outperforming asymmetry models in both CDR and RNFL. Among the four scores, the loci-based PGS526 and PGS616 (in pink and purple) provided the largest and most consistent improvements, suggesting that loci-informed genetic risk meaningfully augments inter-eye asymmetry for early glaucoma risk prediction.

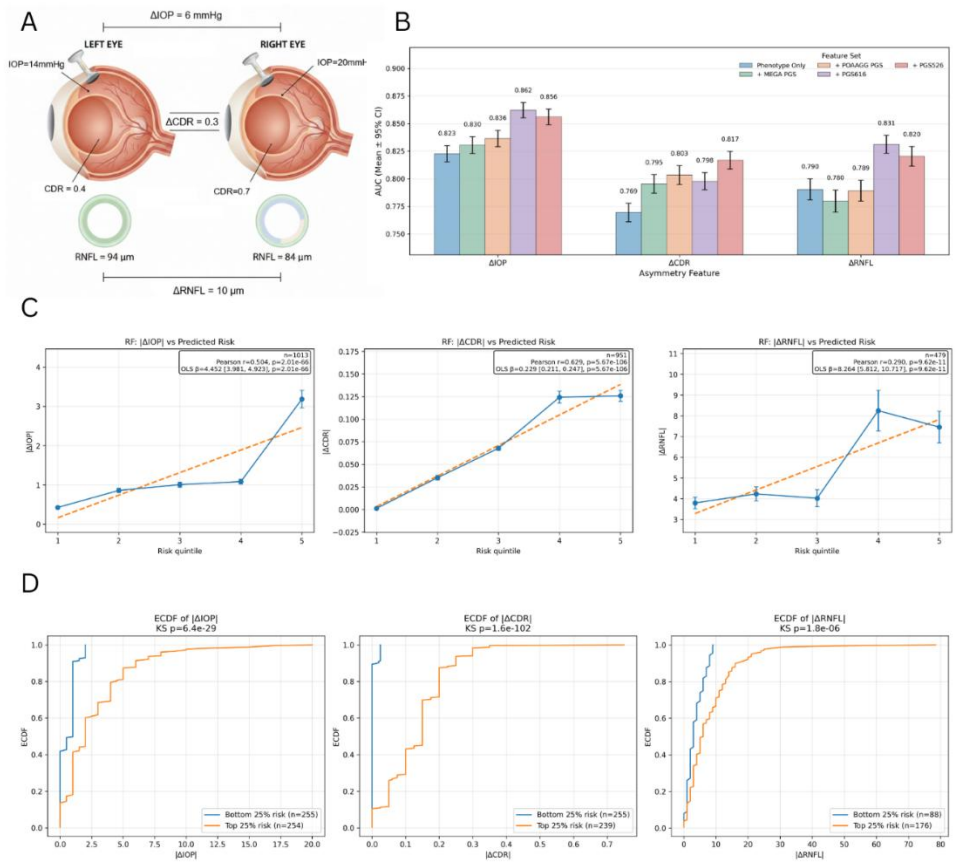
To evaluate asymmetry risk prediction in an independent population, we trained three machine learning models (RF, MLP, and SVM) in the labeled training cohort (N = 271) and applied them without retraining to a suspect cohort of 1,013 individuals. We performed biological validation by testing whether genetically informed risk predictions aligned with established asymmetry patterns in glaucoma, under the premise that inter-eye differences reflect early, unilateral structural damage and disease progression in POAG.

As shown in Figure 5C, predicted glaucoma risk in the suspect cohort demonstrated strong and statistically significant associations with the magnitude of inter-eye asymmetry across all three phenotypes (|ΔIOP|, |ΔCDR|, and |ΔRNFL|). Using continuous individual-level predicted risk scores, RF derived risk was positively correlated with |ΔIOP| (Pearson $r = 0.504$, $p = 2.01\times10^{-66}$) and |ΔCDR| ($r = 0.629$, $p = 5.67\times10^{-106}$), indicating progressively greater asymmetry in intraocular pressure and optic nerve cupping at higher predicted risk levels. A significant, though more moderate, association was also observed for |ΔRNFL| ($r = 0.290$, $p = 9.62\times10^{-11}$), consistent with increased inter-eye RNFL thinning among high-risk individuals. For visualization, predicted risk was stratified into quintiles, demonstrating monotonic enrichment of asymmetry across increasing risk groups. These results reflect models trained using PGS616, which showed the strongest and most consistent enrichment patterns. Comparable analyses using PGS526 (Figure S5) demonstrated directionally similar associations, supporting the robustness of loci-based genetic risk integration. Together, these findings indicate that genetically informed models capture clinically meaningful inter-eye asymmetry patterns indicative of early glaucomatous damage.

Complementary distributional analyses further supported this biological alignment. In Figure 5D, ECDFs comparing individuals in the top and bottom 25% of predicted risk revealed enrichment of inter-eye asymmetry among high-risk suspects. High-risk individuals (top 25%) showed significantly greater |ΔIOP| (Kolmogorov–Smirnov p value = 6.4×10^{-29}), |ΔCDR| ($p = 1.6\times10^{-102}$), and |ΔRNFL| ($p = 1.8\times10^{-6}$) relative to low-risk individuals (bottom 25%). The developed model (inter-eye phenotypes with PGS616) captures

clinically meaningful variation in an unlabeled population, providing strong biological validation of the model's risk stratification.

Figure 5. Phenotypic asymmetry as a clinically meaningful marker, enhanced by genetic risk scores



(A) *Conceptual illustration of clinically meaningful inter-eye asymmetry in glaucoma.*
Schematic depiction of left and right eyes highlighting inter-eye differences in three routinely assessed ocular phenotypes: ΔIOP, ΔCDR, and ΔRNFL. Example values (e.g., IOP = 14 vs. 20 mmHg; CDR = 0.4 vs. 0.7; RNFL = 94 vs. 84 μm) are shown for illustration only and do not correspond to specific study subjects. The schematic emphasizes that glaucoma often manifests asymmetrically, with early or unilateral changes detectable through inter-eye differences even when absolute measurements may remain within clinically normal ranges.

(B) *Random Forest model performance using inter-eye asymmetry features with and without genetic risk scores.*
AUC plotted for RF models trained on individual asymmetry features (ΔIOP, ΔCDR, ΔRNFL) alone or combined with four PGS (POAAGG PGS, MEGA PGS, PGS526, and PGS616) in the training cohort (N = 271). For all three asymmetry features, incorporation of PGS, particularly the curated PGS526 and PGS616, consistently improved predictive performance relative to phenotype models.

(C) *Consistent relationship between predicted risk in validation cohort (N=1,013) and inter-eye asymmetry severity.*
Mean inter-eye asymmetry (± SEM) for |ΔIOP|, |ΔCDR|, and |ΔRNFL| is shown across predicted risk quintiles (1–5) in the suspect cohort. For statistical inference, Pearson correlations and linear regression analyses were performed using continuous individual-level predicted risk scores. Predicted risk demonstrated significant positive associations with |ΔIOP| and |ΔCDR|, as well as a significant but more moderate association with |ΔRNFL|, indicating that higher model-predicted risk corresponds to greater inter-eye asymmetry. These findings support the biological relevance of genetically informed risk stratification in identifying early or asymmetric glaucomatous changes.

(D) *Biological validation of asymmetry-based risk prediction in an independent suspect cohort (N=1,013).*
ECDFs comparing the magnitude of inter-eye asymmetry (|ΔIOP|, |ΔCDR|, |ΔRNFL|) between individuals in the top and bottom quartiles of RF model. It predicted glaucoma risk in the suspect cohort (N = 1,013). High-risk individuals (top 25%) exhibited significantly greater inter-eye asymmetry across all three phenotypes (Kolmogorov–Smirnov test $p < 0.01$), consistent with early or unilateral glaucomatous damage.

DISCUSSION

In this study, we constructed and evaluated four PGS for POAG and examined their performance when integrated with basic clinical information. Two scores captured genome-wide polygenic burden in African ancestry populations, while two were biologically curated and loci based, drawing variants from six large multi-ancestry GWAS and applying effect size weights derived from African ancestry data. Across machine learning models, the selected loci-based PGS showed more consistent predictive value than genome-wide scores, with the strongest performance observed when PGS were combined with age and gender, reaching an AUC of 0.806 in the training cohort. Beyond case-control prediction, risk scores derived from models showed clinical relevance in an independent suspect cohort, where higher predicted risk aligned with established clinical markers and inter-eye asymmetry patterns characteristic of early POAG.

First, our findings highlight the importance of developing risk models that are generalizable across genetically diverse populations. PGS trained in European ancestry populations frequently show reduced portability, especially in African ancestry groups with greater genetic diversity and distinct linkage disequilibrium structure.[26,61] By constructing ancestry-matched genome-wide scores and curated loci-based scores weighted using African ancestry effect sizes, we demonstrate that population-aware PGS design improves predictive consistency.

Next, in the training cohort, PGS alone demonstrated measurable discriminatory ability, with selected loci-based scores outperforming genome-wide approaches. While genome-wide PGS achieved modest performance (AUC ~0.550–0.561), PGS526 showed stronger discrimination (AUC 0.668), highlighting the advantage of curated, biologically informed variants. Later, the baseline model achieved an average AUC of 0.708 alone, consistent with the well-established role of age as a major risk factor for POAG. Adding PGS provided complementary information beyond demographics, with the strongest overall gains observed for base plus PGS616 (average AUC 0.762), and the best performing model achieved by SVM with base plus PGS616 (AUC 0.806). These results indicate that genetic risk provides complementary information beyond demographics and that combining PGS with minimal clinical features enhances risk stratification in modest-sized cohorts. Meanwhile, our comparison of RF, SVM, and MLP models highlights trade-offs between predictive performance and interpretability. All three classifiers achieved comparable performance using limited clinical and genetic features, indicating that effective POAG risk prediction can be achieved.

Moreover, models trained in the discovery cohort retained clinical validity when applied to an independent suspect cohort (N = 1,013). Using RF, SVM, and MLP trained on base + PGS616, we generated individual risk scores and evaluated them against established clinical indicators (IOP, CDR, RNFL thickness) through enrichment analyses. Despite the absence of case/control labels, higher predicted risk was consistently associated with higher IOP, larger CDR, and thinner RNFL, indicating stratification along a clinically meaningful disease gradient. This demonstrates that our models capture meaningful subclinical variation and can stratify individuals along a biologically plausible disease continuum. These results have important clinical implications. In real-world settings, often in underserved or resource-limited environments, many patients fall into a diagnostic gray zone and remain untreated or not followed over time, despite showing early signs of damage. Our framework offers a scalable, genetics-informed risk stratification approach that could flag such individuals for closer monitoring, earlier intervention, or targeted genetic counseling. Importantly, this strategy does not rely on costly imaging or expert annotations, making it adaptable to a range of care settings.

From a practical perspective, our findings further identify inter-eye asymmetry as a clinically meaningful feature for early POAG risk stratification. Because absolute IOP, CDR, and RNFL thresholds were used to define training cases and controls, we derived inter-eye differences (ΔIOP, ΔCDR, ΔRNFL) to avoid circularity and capture eye-specific deviations. These asymmetry features showed predictive value across models, particularly for ΔIOP, and performance improved further when combined with PGS. In the independent suspect cohort, higher predicted risk was consistently associated with greater asymmetry across all three phenotypes, supporting the biological plausibility of asymmetry-based risk prediction for early or unilateral disease.

Notably, our findings provide a practical and flexible framework for POAG risk prediction across different clinical settings. In data-rich environments, combining selected loci-based PGS with clinical phenotypes like IOP, CDR, and RNFL thickness improved model performance, capturing both genetic risk and disease expression. For example, a patient with subtle optic nerve changes but a high PGS might benefit from earlier referral or more frequent monitoring. Importantly, this approach also works in data-limited settings, where clinical measurements may be unavailable. In such cases, a blood-based test to generate PGS offers a low-cost, ancestry-matched tool for estimating risk, even before symptoms appear. This is especially valuable for African ancestry populations, where underdiagnosis is common and access to advanced imaging may be limited. With further validation, integrating PGS into electronic health records and screening programs could support more equitable, scalable, and personalized glaucoma care. In conclusion, our study shows that combining PGS within ML models provides a powerful and equitable approach for early POAG risk prediction, particuarly in African ancestry populations. By rigorously evaluating performance across both curated training data and a real-world suspect cohort, we demonstrate the scientific strength and practical relevance of this approach. These findings support the development of population-specific, AI-driven screening tools that can improve early detection, reduce diagnostic delays, and advance health equity in glaucoma care.

Limitations of the study

Despite the strong performance of our models, several limitations should be acknowledged. First, the machine learning models were trained and evaluated using POAAGG participants recruited from a single city, which may limit portability to other recruitment settings and clinical workflows. Although we incorporated PGS derived from African ancestry GWAS, including a multi-cohort mega-analysis, prediction performance may still vary across African ancestry subgroups due to substantial differences in allele frequencies and linkage disequilibrium patterns among continental African, African American, and admixed populations. In addition, effect-size estimates used for PGS weighting may not fully capture the breadth of African genetic diversity represented across geographic regions and admixture profiles. Future work should externally validate these models in larger, multi-center cohorts and evaluate recalibration strategies to ensure robust and equitable performance across diverse populations.

Second, although we used defined cases and controls to train models and validated them in a large suspect cohort, the binary classification framework may oversimplify the continuum of POAG. Suspect individuals in enrichment analysis may represent transitional or early-stage disease. Future models should incorporate probabilistic or multi-class labeling to better reflect real-world diagnostic uncertainty and support early detection.

Third, our baseline models included only age and gender and did not incorporate ancestry PCs. This decision was intentional, as the training cohort was relatively homogeneous with respect to ancestry and the modest sample size increased the risk of overfitting when adding additional covariates. While sensitivity analyses including the first 20 PCs yielded similar qualitative trends (Figure S1), we cannot exclude the possibility that subtle population structure may influence risk prediction in larger or more heterogeneous cohorts. Future studies with expanded sample sizes will be better positioned to systematically evaluate the role of ancestry PCs alongside genetic risk scores in baseline models.

Fourth, while we integrated key clinical features such as age, IOP, CDR, and RNFL thickness, other glaucoma-relevant phenotypes (e.g., visual field loss, optic disc morphology, or corneal hysteresis) were excluded due to data sparsity or inconsistency. Their absence may limit the comprehensiveness of clinical modeling, especially for atypical presentations or advanced diseases. Additionally, gender was included as a covariate in all models, although stratified analyses were not performed due to limited sample size in the training cohort. We acknowledge that gender-specific genetic and clinical factors may influence glaucoma risk and these deserve further investigation in future studies.

Finally, the integration of complex ML models such as MLP and RF enhances predictive performance but introduces interpretability challenges that may hinder clinical translation. Machine learning models like MLP operate as black boxes, limiting their appeal in settings where decision transparency is critical.[62,63] Hybrid approaches that incorporate interpretable algorithms (e.g., logistic regression or decision trees) or techniques such as SHAP[64] or LIME[65] may help bridge this gap.[66] In addition, although we demonstrated consistent trends across full and training cohorts, external validation using truly independent, prospective

datasets remains necessary to assess robustness, generalizability, and potential confounders related to data collection protocols, healthcare access, or socioeconomic bias.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources, data, and code should be directed to and will be fulfilled by the lead contact, Joan M. O'Brien (joan.o'brien@pennmedicine.upenn.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The de-identified genotype data from the POAAGG study is available through dbGAP accession number phs001312.
- Due to POAAGG subject privacy concerns and IRB regulations, identifiable data cannot be deposited in a public repository. However, de-identified datasets used for model training and evaluation are available from the lead contact upon request and pending data use agreement.
- Summary statistics and curated SNP lists used to compute PGS are available in Table S1.
- All analysis scripts and notebooks are publicly available at: https://github.com/yzhu13/poag-pgs-ml (Release v0.1-submission corresponds to the version used in this manuscript).

ACKNOWLEDGMENTS

This study was supported by the National Eye Institute, National Institutes of Health, through a Research Project Grant (R01EY023557) and a Vision Research Core Grant (2P30EY001583-51). Additional funding was provided by the F.M. Kirby Foundation, Research to Prevent Blindness, the University of Pennsylvania Hospital Board of Women Visitors, and the Paul and Evanina Bell Mackall Foundation Trust. Institutional support was received from the Department of Ophthalmology at the Perelman School of Medicine and the VA Hospital in Philadelphia, Pennsylvania.

The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

AUTHOR CONTRIBUTIONS

Conceptualization: Yan Zhu, Shefali Setia-Verma, Joan M. O'Brien
Methodology: Yan Zhu, Aude Benigne Ikuzwe, Yuki Bradford, Lannawill Caruth
Software and Formal Analysis: Yan Zhu, Aude Benigne Ikuzwe, Yuki Bradford, Lannawill Caruth
Investigation: Yan Zhu, Rebecca Salowe, Laxmi Moksha, Vrathasha Vrathasha, Kenneth Pham, Roy Lee, Mina Halimitabrizi, Isabel Di Rosa, Leila Ghaffari, Jie He

Resources: Marylyn D. Ritchie, Shefali Setia-Verma, Joan M. O'Brien
Data Curation: Yan Zhu, Aude Benigne Ikuzwe
Writing – Original Draft: Yan Zhu, Rebecca Salowe
Writing – Review & Editing: Laxmi Moksha, Vrathasha Vrathasha, Kenneth Pham, Roy Lee, Mina Halimitabrizi, Isabel Di Rosa, Leila Ghaffari, Jie He, Joan O'Brien

Supervision: Marylyn D. Ritchie, Shefali Setia-Verma, Joan M. O'Brien
Funding Acquisition: Joan M. O'Brien

All authors have read and agreed to the published version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

SUPPLEMENTAL INFORMATION

Document S1. Figures S1–S5 and supplemental references

Table S1. SNP lists and African ancestry effect sizes for curated PGS526 and PGS616, related to Figure 2
Table S2. Summary statistics for PGS distributions across disease groups, related to Figure 2
Table S3. Standalone PGS performance metrics across machine learning models, related to Figure 2
Table S4. Model performance metrics across feature sets in the training cohort, related to Figure 3
Table S5. Asymmetry-based model performance metrics across classifiers, related to Figure 5

STAR★METHODS

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Genotype data for ADAGES participants	dbGaP	Accession phs001673
Genotype data for GGLAD participants	Hauser et al.[67]	https://s3-ap-southeast-1.amazonaws.com/gis-hg4-poag/ poag_africa_4_collection_ logistic.meta. sorted_N4.tsv.zip
Genotype data for POAAGG participants	dbGaP	Accession phs001312
Software and algorithms		
Python 3.11	Python Software Foundation	RRID:SCR_008394
R 4.2.2	https://www.r-project.org	https://cloud.r-project.org/bin/windows/ base
Scikit-learn v1.3	scikit-learn developers	RRID:SCR_002577
TensorFlow 2.13 (Keras API)	Google / Meta	RRID:SCR_016345 / RRID:SCR_018194
TOPMED Imputation Reference Panel	Das et al.[68]	https://imputation.biodatacatalyst.nhlbi.nih.gov/#!
EIGENSOFT package	Galinsky et al.[69,70]	http://www.hsph.harvard.edu/alkes-price/ software
SAIGE version 1.1 software	Mbatchou et al.[71]	https://saigegit.github.io/SAIGE-doc/
Matplotlib / Seaborn	Python	N/A
PRS-CS	https://github.com/getian107/PRScs	N/A
PLINK v1.9 / v2.0	Chang et al.[72,73]	https://www.cog-genomics.org/plink/
Other		
Google Colab environment	https://colab.research.google.com	N/A
GitHub repository	https://github.com/yzhu13/poag-pgs-ml	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human Study Participants

Two African ancestry cohorts were used to generate and validate the genome-wide PGS: the POAAGG study cohort and the MEGA cohort.

The POAAGG study cohort consists of self-identified Blacks (African Americans, African descent, or African Caribbean), 35 years or older, identified from the Philadelphia region. Each subject was classified as a case, control, or suspect by a glaucoma specialist or ophthalmologist, based on previously detailed criteria.[74] Subjects underwent a comprehensive onsite interview and examination, including the collection of demographic information and comprehensive ocular phenotypes. All subjects signed an informed consent form and provided a genomic DNA sample. [9,56,75] The study was approved by the Institutional Review Board at the University of Pennsylvania and was conducted in accordance with the tenets of the Declaration of Helsinki.

For the current analysis, a total of 8,315 POAAGG subjects (including 2,723 cases, 4,579 controls, and 1013 suspects) were included. All subjects had available genotype data, demographic information (age, gender), and at least one ocular phenotype (IOP, vertical CDR, and RNFL thickness). This comprised:

- **POAAGG PGS Generation** cohort (N=7,031, with 2,595 cases and 4,436 controls)**:** This cohort was used in the generation of the POAAGG PGS, described in the below section.
- **Training case/control cohort** (N=271, with 128 cases and 143 controls): This cohort consists of cases and controls <u>not</u> included in the generation of the PGS scores. To ensure that subjects correctly fit their classification of case or control, we carefully reviewed their phenotypes. 99 clean cases were required to meet the following phenotypes in both eyes: IOP ≥ 22.1 mmHg, CDR ≥ 0.7, and RNFL ≤ 80 μm. 97 clean controls were required to have normal phenotypes for both eyes, including: IOP between 10–16 mmHg, CDR ≤ 0.5, and RNFL ≥ 82 μm. The remaining 75 subjects were drawn from recently accrued, real-world clinical POAG cases and controls within the same healthcare system.
- **Suspect cohort** (N=1,013): This cohort was used to assess model generalizability and perform clinical enrichment analyses.[56] The individuals in suspect cohort were <u>not</u> included in PGS generation and were <u>not</u> used in model training.

Additionally, the MEGA dataset was used to generate the MEGA PGS has been described previously.[14] This cohort consists of 11,275 subjects (6,003 cases and 5,272 controls) from three African ancestry glaucoma cohorts: (1) ADAGES study (n=1,999), (2) GGALD study (n=2,952), and (3) POAAGG study (n=6,324).

METHOD DETAILS

1. Data Preprocessing and Cohort Selection

This study used two distinct categories of data sources with non-overlapping roles: (i) large-scale genetic datasets for PGS derivation and (ii) a deeply phenotyped clinical cohort for machine learning model development and validation.

PGS derivation datasets: PGS were derived using GWAS summary statistics and individual-level data from African ancestry subjects in the POAAGG study and the MEGA cohort, described above. All MEGA data were imputed to the TOPMed reference panel,[76] and association testing accounted for population structure and relatedness. These datasets were used exclusively for PGS construction and weighting and did not contribute to machine learning model training or testing.

Machine learning training and validation cohorts: All machine learning model training, internal evaluation, and external validation were performed exclusively within the POAAGG cohort. A curated case-control subset (N = 271) was used for model training and bootstrap-based performance evaluation, while an independent suspect cohort (N = 1,013) was used for biological validation and clinical enrichment analyses. No individuals used for PGS derivation were included in model training or testing, ensuring strict separation between genetic discovery and predictive modeling.

Demographic features were numerically encoded (male = 1, female = 0). Continuous clinical traits, including IOP, vertical CDR, and RNFL thickness, were standardized using z-score transformation. Missing clinical values were imputed using cohort-wise means. All preprocessing steps, including imputation and scaling, were implemented within scikit-learn pipelines to ensure consistent application across bootstrap resampling iterations and prevent data leakage.

2.PGS Construction

We constructed four PGS for POAG using two complementary strategies: genome-wide modeling to capture diffuse polygenic burden and curated, locus-based scores for interpretable biological relevance (Figure 2). All four scores were computed using African ancestry-specific effect sizes to ensure appropriate calibration for our target population.

1. POAAGG PGS: The POAAGG PGS is a genome-wide polygenic score derived from GWAS summary statistics generated within the POAAGG cohort (N = 7,031), which includes African ancestry individuals. We used the PRS-CS framework, a Bayesian method that applies continuous shrinkage priors to SNP effect sizes while accounting for linkage disequilibrium (LD).[77] Effect size modeling was performed using the African ancestry reference panel from the 1000 Genomes Project, with default parameters (φ = auto, 1,000 burn-in and 2,000 MCMC iterations).[14,26] Ambiguous SNPs (A/T or G/C) were removed prior to scoring.[78] Genome-wide scores were computed using the weighted sum of alleles in PLINK 2.0 and then standardized (z-score) within the POAAGG cohort.[72,73] This score reflects cohort-specific polygenic risk and serves as a comparator for broader African ancestry models.
2. MEGA PGS: The MEGA PGS was constructed using the same methodology as the POAAGG PGS but applied to the meta-analysis of African ancestry individuals (N = 11,275) drawn from three large glaucoma cohorts (ADAGES, GGLAD, and POAAGG). By aggregating individual-level data across multiple cohorts, this score captures a more comprehensive and generalizable estimate of POAG polygenic burden in African ancestry populations. As with the POAAGG PGS, we used PRS-CS with the African ancestry LD reference panel and computed weighted allele sums in PLINK 2.0,[72,73] followed by z-score normalization.
3. PGS526: PGS526 is a biologically curated, locus-based score composed of 526 SNPs selected for their strong associations with POAG endophenotypes, including intraocular pressure (IOP and IOP_max), optic nerve structure (CDR and RNFL thickness), and additional functional and anatomical markers such as visual acuity (VA), mean deviation (MD), pattern standard deviation (PSD), and central corneal thickness (CCT). SNPs were drawn from six large multi-ancestry glaucoma GWAS (Craig et al.,[51] Gharahkhani et al.,[53] Han et al.,[54] Faro et al.,[52] Chang-Wolf et al.,[55] Verma et al.[14]), harmonized to the hg38 genome build, and filtered to retain variants present in both the POAAGG genotype data and MEGA summary statistics. Importantly, all effect sizes (β) were derived exclusively from the MEGA African ancestry GWAS to ensure ancestry-matched weighting.[14] Final scores were computed in PLINK 2.0,[72,73] normalized using rank-based inverse normal transformation, and standardized prior to model integration.[72,73] PGS526 emphasizes mechanistic genetic risk relevant to measurable physiological traits underlying POAG.
4. PGS616: PGS616 is a complementary, selected locus-based PGS designed to capture broader POAG disease risk by integrating 616 SNPs derived from genome-wide or suggestive POAG-associated loci across the same six GWAS used for PGS526. Unlike PGS526, which focused on endophenotype-associated loci, PGS616 prioritized SNPs associated with POAG disease status, including those that were replicated in the MEGA dataset.[14] All variants were subjected to the same filtering, pruning, and weighting pipeline as PGS526, with African ancestry-specific effect sizes drawn from MEGA summary statistics. Like the other scores, PGS616 was computed using PLINK 2.0 and normalized before downstream use.[72,73] Together with PGS526, this score represents a curated, interpretable risk model that highlights genetic signals with strong disease-specific relevance in African ancestry populations.

These four scores represent complementary genetic risk models that balance breadth, biological specificity, and ancestry-matched calibration, forming the foundation for downstream machine learning and clinical enrichment analyses.

3. Machine Learning Models: RF, MLP, and SVM

We applied three ML models to each feature set: RF, MLP, and SVM, implemented using the scikit-learn library (v1.3.0).[79] Each classifier was trained to distinguish POAG cases from controls based on different combinations of demographic, clinical, and genetic features.

- RF[46,80] was configured with 100 decision trees, using the default max_depth=None and class_weight='balanced' to correct for label imbalance.
- MLP[44] employed a single hidden layer with 32 neurons, ReLU activation, and stochastic gradient descent with early stopping enabled (max_iter=300).
- SVM[50] was used with a radial basis function (RBF) kernel, with probability estimates enabled (probability = True) to support AUC calculation.

Model performance was assessed using repeated bootstrap resampling. [80,81] In each bootstrap iteration, the dataset was resampled with replacement at the individual level, followed by a stratified 80/20 train–test split to preserve the original case-control ratio.[80] The training pipeline included mean imputation for missing values (SimpleImputer), z-score standardization of continuous features (StandardScaler), and model fitting.[82,83]

Each PGS was incorporated as a continuous variable alongside demographic covariates (age and gender) in all machine learning analyses. We constructed and evaluated five feature sets for model training and evaluation:

- (1) base = age + gender
- (2) base + POAAGG PGS
- (3) base + MEGA PGS
- (4) base + PGS526
- (5) base + PGS616

This structured design enabled systematic assessment of the individual and combined contributions of demographic and genetic features to glaucoma risk prediction. Full feature definitions and modeling pipelines are publicly available at: https://github.com/yzhu13/poag-pgs-ml (release v0.1-submission).

Performance metrics, including AUC, accuracy, precision, sensitivity, specificity, and F1-score, were computed on the held-out test set for each bootstrap iteration.[84] Final reported values represent the mean and standard deviation across bootstrap iterations.[81] This resampling-based evaluation strategy was applied consistently across all models and feature sets to ensure fair model comparison and robust estimation.

4. Testing Cohort Analysis and Clinical Enrichment in Suspect Population

To evaluate the biological relevance and generalizability of PGS-informed models in the absence of definitive diagnostic labels, trained ML models were applied without retraining to an independent suspect cohort (N = 1,013). Models were trained exclusively in the labeled training cohort (N = 271) and then used to generate continuous, individual-level predicted glaucoma risk scores for all suspects.

For this analysis, we focused on models trained using the base + PGS616 feature set, which showed the most consistent and highest performance across models in the training cohort. Predicted risk scores were retained as continuous individual level outputs for statistical inference. Associations between predicted risk and quantitative ocular phenotypes (IOP, CDR, RNFL thickness) were evaluated using Pearson correlation and OLS linear regression across all 1,013 individuals in the suspect cohort.

For visualization of enrichment patterns, predicted risk scores were stratified into quintiles (1–5), and mean phenotype values within each quintile were plotted with corresponding standard errors (± SEM). Because the suspect cohort lacks definitive diagnostic labels, classification metrics such as AUC could not be computed; therefore, model evaluation focused on clinical enrichment analyses using continuous statistical testing complemented by stratified visualization.

654 In addition, ECDFs were used to compare phenotype distributions between individuals in the top and bottom
655 quartiles of predicted risk. Differences between distributions were evaluated using the Kolmogorov–
656 Smirnov test.

657 This enrichment-based validation framework enables assessment of biological plausibility and disease
658 relevance in unlabeled populations and is consistent with prior approaches for evaluating risk prediction
659 models in real-world clinical cohorts.

660 **5. Asymmetry Analysis and Biological Validation**

661 To evaluate inter-eye asymmetry as an independent and clinically meaningful disease signal while avoiding
662 circularity from cohort definition criteria, we derived inter-eye difference features for ΔIOP, ΔCDR, and
663 ΔRNFL.[58,85] Each asymmetry feature was calculated as the absolute difference between the right and left
664 eyes (|Right − Left|).

665 Asymmetry features were evaluated alone and in combination with PGS using all three models (RF, MLP,
666 SVM) trained in the training cohort (N = 271) under the same evaluation framework described above. Across
667 feature sets, RF models demonstrated the most stable and highest overall performance and were therefore
668 selected as the primary models for downstream validation (Figure 5); SVM and MLP results are reported in
669 Figure S3 and Figure S4.

670 For biological validation, a predefined RF model trained on combined asymmetry features (ΔIOP, ΔCDR,
671 ΔRNFL) and PGS616 was applied without retraining to an independent suspect cohort (N = 1,013), which
672 lacks definitive diagnostic labels. The model generated continuous, individual-level predicted glaucoma risk
673 scores for all 1,013 suspects. Validation focused on clinical enrichment analyses, assessing whether higher
674 predicted risk aligned with greater inter-eye asymmetry. Predicted risk scores were analyzed as continuous
675 individual-level variables for statistical inference. Associations between predicted risk and asymmetry
676 magnitude were evaluated using Pearson correlation and OLS regression across all individuals. For
677 visualization, predicted risk was stratified into quintiles (1–5), and mean inter-eye asymmetry values within
678 each quintile were plotted (± SEM). Distributional differences between high-risk and low-risk groups were
679 further assessed using ECDF analysis with Kolmogorov–Smirnov testing.

680 **6. Software and Reproducibility**

681 All data preprocessing, feature engineering, and ML modeling were performed in Python (version 3.11)[86]
682 using open-source libraries, including scikit-learn (version 1.3.0)[79] for model implementation, pandas and
683 numpy for data manipulation, matplotlib and seaborn for visualization, and SciPy for statistical analysis.
684 Genetic data processing, including risk score calculation and dataset merging, was conducted using PLINK
685 v2.0[73] and PRS-CS[77], with ancestry-specific LD panels used to infer SNP weights for PGS.

686 All scripts were developed and executed within Google Colaboratory (Colab)[87], facilitating a cloud-based
687 computational environment with consistent software dependencies across users. The project followed
688 reproducible research practices, with each model training and evaluation task encapsulated in modular
689 code blocks with clearly defined input and output interfaces. Outputs for each model–feature set
690 combination were stored as labeled CSV and Excel files, with naming conventions reflecting the model type
691 (e.g., RF, MLP, SVM), feature set, and performance metrics.

692 To promote transparency and facilitate replication, all Google Colab notebooks, including those used for
693 bootstrap-based model evaluation and visualization generation, have been deposited in a GitHub repository
694 and are available publicly, as described in the Data and Code Availability section.

695 All visualizations presented in the main text were generated directly from model outputs using standardized
696 plotting functions that support error bar overlays and stratified grouping. These functions are also included
697 in the repository and can be reused or adapted by other researchers. Together, this combination of modular
698 scripting, version controlled code, and cloud based execution supports full reproducibility of the AI modeling
699 pipeline and facilitates broader adoption of these tools for glaucoma risk prediction.

700 ## QUANTIFICATION AND STATISTICAL ANALYSIS

701 Model performance was quantified using standard classification metrics, including AUC, accuracy, F1-
702 score, sensitivity (recall for the positive class), and specificity (recall for the negative class).[84] Metrics were

computed on held-out test sets across repeated bootstrap resampling iterations, and final reported values represent the mean and standard deviation across iterations.[81]

Performance summaries for individual models and feature sets are presented in Figures 2–6 and Tables S1–S5. Given the nonparametric nature of the ML models and resampling-based evaluation framework, no assumptions of normality or homoscedasticity were required for statistical inference.

ADDITIONAL RESOURCES

This study is not a clinical trial; there is no clinical registry number. There are no external sites that have been generated to support discussion or use of the information/data/material created by the manuscript.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI, GPT-5.2) to assist with language refinement, clarity, and editorial revisions of the manuscript. All content generated with AI assistance was carefully reviewed, edited, and validated by the authors. The authors take full responsibility for the accuracy, integrity, and final content of the published article.

REFERENCES

1. Weinreb, R.N., and Khaw, P.T. (2004). Primary open-angle glaucoma. The lancet *363*, 1711-1720.
2. Zhang, N., Wang, J., Li, Y., and Jiang, B. (2021). Prevalence of primary open angle glaucoma in the last 20 years: a meta-analysis and systematic review. Scientific reports *11*, 13762.
3. Quigley, H.A., and Broman, A.T. (2006). The number of people with glaucoma worldwide in 2010 and 2020. British journal of ophthalmology *90*, 262-267.
4. Weinreb, R.N., Leung, C.K., Crowston, J.G., Medeiros, F.A., Friedman, D.S., Wiggs, J.L., and Martin, K.R. (2016). Primary open-angle glaucoma. Nature reviews Disease primers *2*, 1-19.
5. Sommer, A., Tielsch, J.M., Katz, J., Quigley, H.A., Gottsch, J.D., Javitt, J., and Singh, K. (1991). Relationship between intraocular pressure and primary open angle glaucoma among white and black Americans: the Baltimore Eye Survey. Archives of ophthalmology *109*, 1090-1095.
6. Hollands, H., Johnson, D., Hollands, S., Simel, D.L., Jinapriya, D., and Sharma, S. (2013). Do findings on routine examination identify patients at risk for primary open-angle glaucoma?: The rational clinical examination systematic review. Jama *309*, 2035-2042.
7. Zhu, Y., Moksha, L., Salowe, R., Vrathasha, V., Pham, K., Aibo, M.-A.I., Lee, R., Halimitabrizi, M., Di Rosa, I., and O'Brien, J.M. (2025). Integrating neuroprotection, antioxidative effects, and precision medicine in glaucoma management with bioactive compounds. Biomedicine & Pharmacotherapy *190*, 118319.
8. Tielsch, J.M., Sommer, A., Katz, J., Royall, R.M., Quigley, H.A., and Javitt, J. (1991). Racial variations in the prevalence of primary open-angle glaucoma: the Baltimore Eye Survey. Jama *266*, 369-374.
9. Khachatryan, N., Pistilli, M., Maguire, M.G., Salowe, R.J., Fertig, R.M., Moore, T., Gudiseva, H.V., Chavali, V.R., Collins, D.W., and Daniel, E. (2019). Primary open-angle African American glaucoma genetics (POAAGG) study: gender and risk of POAG in African Americans. PloS one *14*, e0218804.
10. Mitchell, P., Smith, W., Attebo, K., and Healey, P.R. (1996). Prevalence of open-angle glaucoma in Australia: the Blue Mountains Eye Study. Ophthalmology *103*, 1661-1669.
11. Hennis, A., Wu, S.-Y., Nemesure, B., Honkanen, R., Leske, M.C., and Group, B.E.S. (2007). Awareness of incident open-angle glaucoma in a population study: the Barbados Eye Studies. Ophthalmology *114*, 1816-1821.

12. Bettin, P., and Di Matteo, F. (2013). Glaucoma: present challenges and future trends. Ophthalmic research *50*, 197-208.
13. Lewis, C.M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. Genome medicine *12*, 44.
14. Verma, S.S., Gudiseva, H.V., Chavali, V.R., Salowe, R.J., Bradford, Y., Guare, L., Lucas, A., Collins, D.W., Vrathasha, V., and Nair, R.M. (2024). A multi-cohort genome-wide association study in African ancestry individuals reveals risk loci for primary open-angle glaucoma. Cell *187*, 464-480. e410.
15. Igo Jr, R.P., Kinzy, T.G., and Cooke Bailey, J.N. (2019). Genetic risk scores. Current protocols in human genetics *104*, e95.
16. Läll, K., Mägi, R., Morris, A., Metspalu, A., and Fischer, K. (2017). Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. Genetics in Medicine *19*, 322-329.
17. Chasman, D.I., Giulianini, F., Demler, O.V., and Udler, M.S. (2020). Pleiotropy-based decomposition of genetic risk scores: association and interaction analysis for type 2 diabetes and CAD. The American Journal of Human Genetics *106*, 646-658.
18. Ripatti, S., Tikkanen, E., Orho-Melander, M., Havulinna, A.S., Silander, K., Sharma, A., Guiducci, C., Perola, M., Jula, A., and Sinisalo, J. (2010). A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. The Lancet *376*, 1393-1400.
19. Morrison, A.C., Bare, L.A., Chambless, L.E., Ellis, S.G., Malloy, M., Kane, J.P., Pankow, J.S., Devlin, J.J., Willerson, J.T., and Boerwinkle, E. (2007). Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. American journal of epidemiology *166*, 28-35.
20. Shieh, Y., Hu, D., Ma, L., Huntsman, S., Gard, C.C., Leung, J.W., Tice, J.A., Vachon, C.M., Cummings, S.R., and Kerlikowske, K. (2016). Breast cancer risk prediction using a clinical risk model and polygenic risk score. Breast cancer research and treatment *159*, 513-525.
21. Roberts, E., Howell, S., and Evans, D.G. (2023). Polygenic risk scores and breast cancer risk prediction. The Breast *67*, 71-77.
22. Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J.P., Chen, T.-H., Wang, Q., and Bolla, M.K. (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. The American Journal of Human Genetics *104*, 21-34.
23. Qassim, A., Souzeau, E., Siggs, O.M., Hassall, M.M., Han, X., Griffiths, H.L., Frost, N.A., Vallabh, N.A., Kirwan, J.F., and Menon, G. (2020). An intraocular pressure polygenic risk score stratifies multiple primary open-angle glaucoma parameters including treatment intensity. Ophthalmology *127*, 901-907.
24. Gao, X.R., Huang, H., and Kim, H. (2019). Polygenic risk score is associated with intraocular pressure and improves glaucoma prediction in the UK Biobank cohort. Translational vision science & technology *8*, 10-10.
25. Han, X., Hewitt, A.W., and MacGregor, S. (2021). Predicting the future of genetic risk profiling of glaucoma: a narrative review. JAMA ophthalmology *139*, 224-231.
26. Wang, Y., Kanai, M., Tan, T., Kamariza, M., Tsuo, K., Yuan, K., Zhou, W., Okada, Y., Huang, H., and Turley, P. (2023). Polygenic prediction across populations is influenced by ancestry, genetic architecture, and methodology. Cell Genomics *3*.
27. Lambert, S.A., Abraham, G., and Inouye, M. (2019). Towards clinical utility of polygenic risk scores. Human molecular genetics *28*, R133-R142.
28. Shieh, Y., Roger, J., Yau, C., Wolf, D.M., Hirst, G.L., Swigart, L.B., Huntsman, S., Hu, D., Nierenberg, J.L., and Middha, P. (2023). Development and testing of a polygenic risk score for breast cancer aggressiveness. NPJ Precision Oncology *7*, 42.

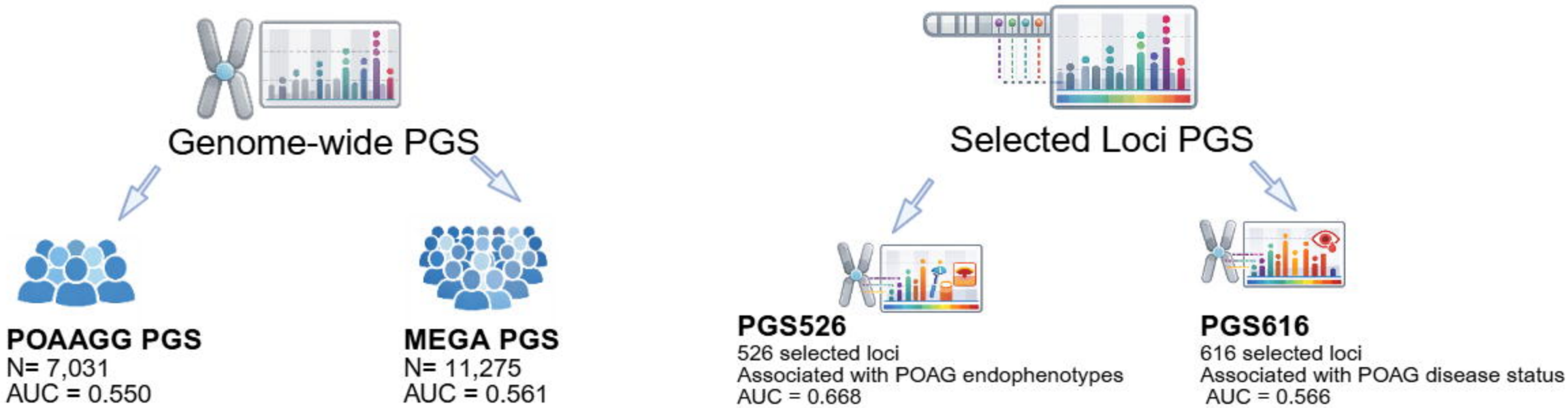
29. Sekimitsu, S., Xiang, D., Smith, S.L., Curran, K., Elze, T., Friedman, D.S., Foster, P.J., Luo, Y., Pasquale, L.R., and Peto, T. (2023). Deep ocular phenotyping across primary open-angle glaucoma genetic burden. JAMA ophthalmology *141*, 891-899.
30. Rajpurkar, P., Chen, E., Banerjee, O., and Topol, E.J. (2022). AI in health and medicine. Nature medicine *28*, 31-38.
31. Wang, F., and Preininger, A. (2019). AI in health: state of the art, challenges, and future directions. Yearbook of medical informatics *28*, 016-026.
32. Ngiam, K.Y., and Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. The Lancet Oncology *20*, e262-e273.
33. Beam, A.L., and Kohane, I.S. (2018). Big data and machine learning in health care. Jama *319*, 1317-1318.
34. Zhu, Y., Salowe, R., Chow, C., Li, S., Bastani, O., and O'Brien, J.M. (2024). Advancing glaucoma care: integrating artificial intelligence in diagnosis, management, and progression detection. Bioengineering *11*, 122.
35. Gurudath, N., Celenk, M., and Riley, H.B. (2014). Machine learning identification of diabetic retinopathy from fundus images. (IEEE), pp. 1-7.
36. An, G., Omodaka, K., Hashimoto, K., Tsuda, S., Shiga, Y., Takada, N., Kikawa, T., Yokota, H., Akiba, M., and Nakazawa, T. (2019). Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images. Journal of healthcare engineering *2019*, 4061313.
37. Alipanahi, B., Hormozdiari, F., Behsaz, B., Cosentino, J., McCaw, Z.R., Schorsch, E., Sculley, D., Dorfman, E.H., Foster, P.J., and Peng, L.H. (2021). Large-scale machine-learning-based phenotyping significantly improves genomic discovery for optic nerve head morphology. The American Journal of Human Genetics *108*, 1217-1230.
38. Han, X., Steven, K., Qassim, A., Marshall, H.N., Bean, C., Tremeer, M., An, J., Siggs, O.M., Gharahkhani, P., and Craig, J.E. (2021). 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. The American Journal of Human Genetics *108*, 1204-1216.
39. Zeppieri, M., Gardini, L., Culiersi, C., Fontana, L., Musa, M., D'Esposito, F., Surico, P.L., Gagliano, C., and Sorrentino, F.S. (2024). Novel Approaches for the Early Detection of Glaucoma Using Artificial Intelligence. Life *14*, 1386.
40. Truong, B., Hull, L.E., Ruan, Y., Huang, Q.Q., Hornsby, W., Martin, H., Van Heel, D.A., Wang, Y., Martin, A.R., and Lee, S.H. (2024). Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. Cell Genomics *4*.
41. Zhai, S., Mehrotra, D.V., and Shen, J. (2024). Applying polygenic risk score methods to pharmacogenomics GWAS: challenges and opportunities. Briefings in Bioinformatics *25*, bbad470.
42. Mamidipaka, A., Di Rosa, I., Lee, R., Zhu, Y., Chen, Y., Salowe, R., Addis, V., Sankar, P., Daniel, E., and Ying, G.-S. (2023). Factors associated with large cup-to-disc ratio and blindness in the primary open-angle African American glaucoma genetics (POAAGG) study. Genes *14*, 1809.
43. Gudiseva, H.V., Verma, S.S., Chavali, V.R., Salowe, R.J., Lucas, A., Collins, D.W., Rathi, S., He, J., Lee, R., and Merriam, S. (2020). Genome wide-association study identifies novel loci in the Primary Open-Angle African American Glaucoma Genetics (POAAGG) study. BioRxiv, 2020.2002. 2027.968156.
44. Popescu, M.-C., Balas, V.E., Perescu-Popescu, L., and Mastorakis, N. (2009). Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems *8*, 579-588.
45. Qi, Y. (2012). Random forest for bioinformatics. In Ensemble machine learning, (Springer), pp. 307-323.

46. Belgiu, M., and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing *114*, 24-31.
47. Breiman, L. (2001). Random forests. Machine learning *45*, 5-32.
48. Suthaharan, S. (2016). Support vector machine. In Machine learning models and algorithms for big data classification: thinking with examples for effective learning, (Springer), pp. 207-235.
49. Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., and Scholkopf, B. (1998). Support vector machines. IEEE Intelligent Systems and their applications *13*, 18-28.
50. Cortes, C., and Vapnik, V. (1995). Support-vector networks. Machine learning *20*, 273-297.
51. Craig, J.E., Han, X., Qassim, A., Hassall, M., Cooke Bailey, J.N., Kinzy, T.G., Khawaja, A.P., An, J., Marshall, H., and Gharahkhani, P. (2020). Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. Nature genetics *52*, 160-166.
52. Faro, V.L., Bhattacharya, A., Zhou, W., Zhou, D., Wang, Y., Läll, K., Kanai, M., Lopera-Maya, E., Straub, P., and Pawar, P. (2024). Novel ancestry-specific primary open-angle glaucoma loci and shared biology with vascular mechanisms and cell proliferation. Cell Reports Medicine *5*.
53. Gharahkhani, P., Jorgenson, E., Hysi, P., Khawaja, A.P., Pendergrass, S., Han, X., Ong, J.S., Hewitt, A.W., Segrè, A.V., and Rouhana, J.M. (2021). Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. Nature communications *12*, 1258.
54. Han, X., Gharahkhani, P., Hamel, A.R., Ong, J.S., Rentería, M.E., Mehta, P., Dong, X., Pasutto, F., Hammond, C., and Young, T.L. (2023). Large-scale multitrait genome-wide association analyses identify hundreds of glaucoma risk loci. Nature genetics *55*, 1116-1125.
55. Chang-Wolf, J.M., Kinzy, T.G., Driessen, S.J., Cruz, L.A., Iyengar, S.K., Peachey, N.S., Aung, T., Khor, C.C., Williams, S.E., and Ramsay, M. (2025). Performance of Polygenic Risk Scores for Primary Open-Angle Glaucoma in Populations of African Descent. JAMA ophthalmology *143*, 7-14.
56. Zhao, A.T., Di Rosa, I., Salowe, R., Jin, F., Lee, R., Aibo, M.-A.I., Halimitabrizi, M., Chow, C., Zhu, Y., and Addis, V. (2025). Predictors of Glaucoma Conversion in an African Ancestry Cohort: A Longitudinal Study. American Journal of Ophthalmology.
57. Mamidipaka, A., Shi, A., Lee, R., Zhu, Y., Chen, Y., Di Rosa, I., Salowe, R., Ying, G.-S., and O'Brien, J.M. (2025). Socioeconomic and environmental factors associated with glaucoma in an African Ancestry Population: findings from the Primary Open-Angle African American Glaucoma Genetics (POAAGG) study. Eye *39*, 1086-1092.
58. Hou, H., Moghimi, S., Zangwill, L.M., Shoji, T., Ghahari, E., Manalastas, P.I.C., Penteado, R.C., and Weinreb, R.N. (2018). Inter-eye asymmetry of optical coherence tomography angiography vessel density in bilateral glaucoma, glaucoma suspect, and healthy eyes. American journal of ophthalmology *190*, 69-77.
59. Maguire, M.G. (2020). Assessing intereye symmetry and its implications for study design. Investigative Ophthalmology & Visual Science *61*, 27-27.
60. Eppig, T., Spira-Eppig, C., Goebels, S., Seitz, B., El-Husseiny, M., Lenhart, M., Papavasileiou, K., Szentmáry, N., and Langenbucher, A. (2018). Asymmetry between left and right eyes in keratoconus patients increases with the severity of the worse eye. Current Eye Research *43*, 848-855.
61. Fatumo, S., and Inouye, M. (2023). African genomes hold the key to accurate genetic risk prediction. Nature Human Behaviour *7*, 295-296.
62. Doshi-Velez, F., and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
63. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence *1*, 206-215.

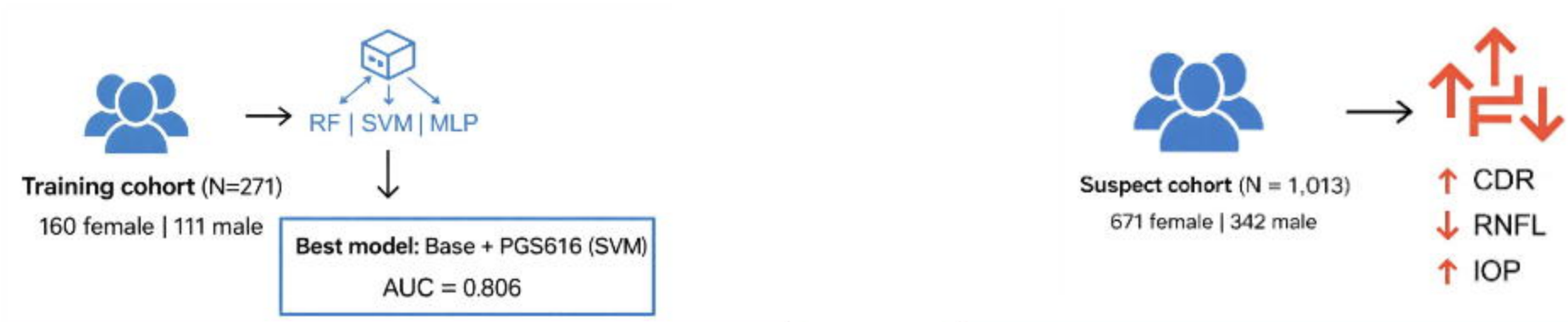
64. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems *30*.
65. Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). " Why should i trust you?" Explaining the predictions of any classifier. pp. 1135-1144.
66. Molnar, C., Casalicchio, G., and Bischl, B. (2020). Interpretable machine learning–a brief history, state-of-the-art and challenges. (Springer), pp. 417-431.
67. Consortium, G.o.G.i.P.o.A.D., Hauser, M.A., Allingham, R.R., Aung, T., Van Der Heide, C.J., Taylor, K.D., Rotter, J.I., Wang, S.-H.J., Bonnemaijer, P.W., and Williams, S.E. (2019). Association of genetic variants with primary open-angle glaucoma among individuals with African ancestry. Jama *322*, 1682-1691.
68. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., and McGue, M. (2016). Next-generation genotype imputation service and methods. Nature genetics *48*, 1284-1287.
69. Galinsky, K.J., Loh, P.-R., Mallick, S., Patterson, N.J., and Price, A.L. (2016). Population structure of UK Biobank and ancient Eurasians reveals adaptation at genes influencing blood pressure. The American Journal of Human Genetics *99*, 1130-1139.
70. Galinsky, K.J., Bhatia, G., Loh, P.-R., Georgiev, S., Mukherjee, S., Patterson, N.J., and Price, A.L. (2016). Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. The American Journal of Human Genetics *98*, 456-472.
71. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., and Boutkov, B. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. Nature genetics *53*, 1097-1103.
72. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, s13742-13015-10047-13748.
73. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics *81*, 559-575.
74. Charlson, E.S., Sankar, P.S., Miller-Ellis, E., Regina, M., Fertig, R., Salinas, J., Pistilli, M., Salowe, R.J., Rhodes, A.L., and Merritt III, W.T. (2015). The primary open-angle african american glaucoma genetics study: baseline demographics. Ophthalmology *122*, 711-720.
75. Salowe, R.J., Lee, R., Zenebe-Gete, S., Vaughn, M., Gudiseva, H.V., Pistilli, M., Kikut, A., Becker, E., Collins, D.W., and He, J. (2022). Recruitment strategies and lessons learned from a large genetic study of African Americans. PLOS global public health *2*, e0000416.
76. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., and Kang, H.M. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature *590*, 290-299.
77. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nature communications *10*, 1776.
78. Jin, J., Zhan, J., Zhang, J., Zhao, R., O'Connell, J., Jiang, Y., Aslibekyan, S., Auton, A., Babalola, E., and Bell, R.K. (2024). MUSSEL: Enhanced Bayesian polygenic risk prediction leveraging information across multiple ancestry groups. Cell genomics *4*.
79. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research *12*, 2825-2830.
80. Efron, B., and Tibshirani, R.J. (1994). An introduction to the bootstrap (Chapman and Hall/CRC).
81. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In 2. (Montreal, Canada), pp. 1137-1145.

82. Little, R.J., and Rubin, D.B. (2019). Statistical analysis with missing data (John Wiley & Sons).
83. Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning. Springer series in statistics New-York.
84. Powers, D.M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.
85. Cui, H., and Zhang, Z. (2025). Interocular symmetry and asymmetry in ocular health and disease. Frontiers in Medicine *12*, 1626329.
86. Van Rossum, G., and Drake Jr, F.L. (1995). Python tutorial (Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands).
87. Bisong, E. (2019). Google colaboratory. In Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners, (Springer), pp. 59-64.

Genetic architecture drives early clinical stratification



PGS enhance machine learning performance and risk stratification

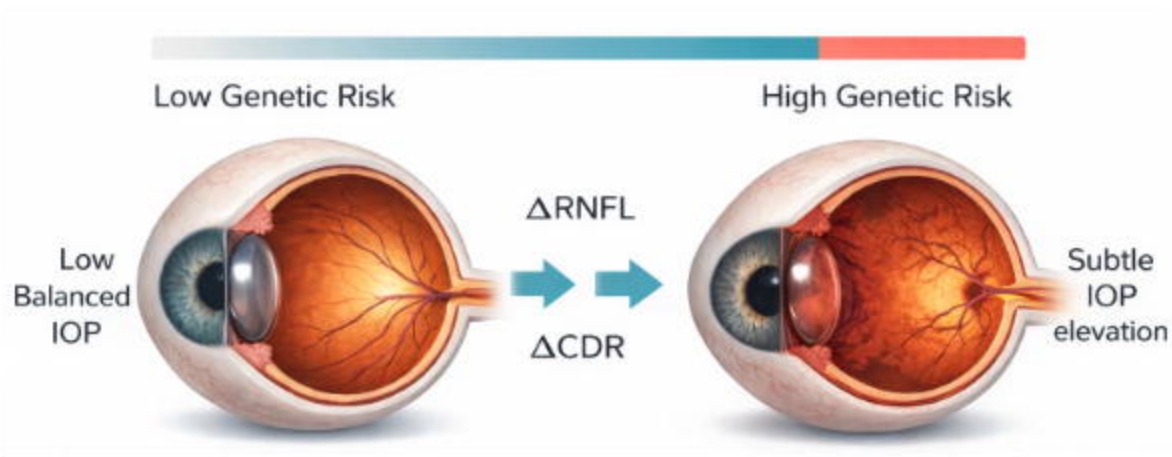


Curated loci PGS outperform genome-wide PGS

Early disease signal prior to definitive diagnosis

Genetically informed models capture clinically meaningful inter-eye asymmetry

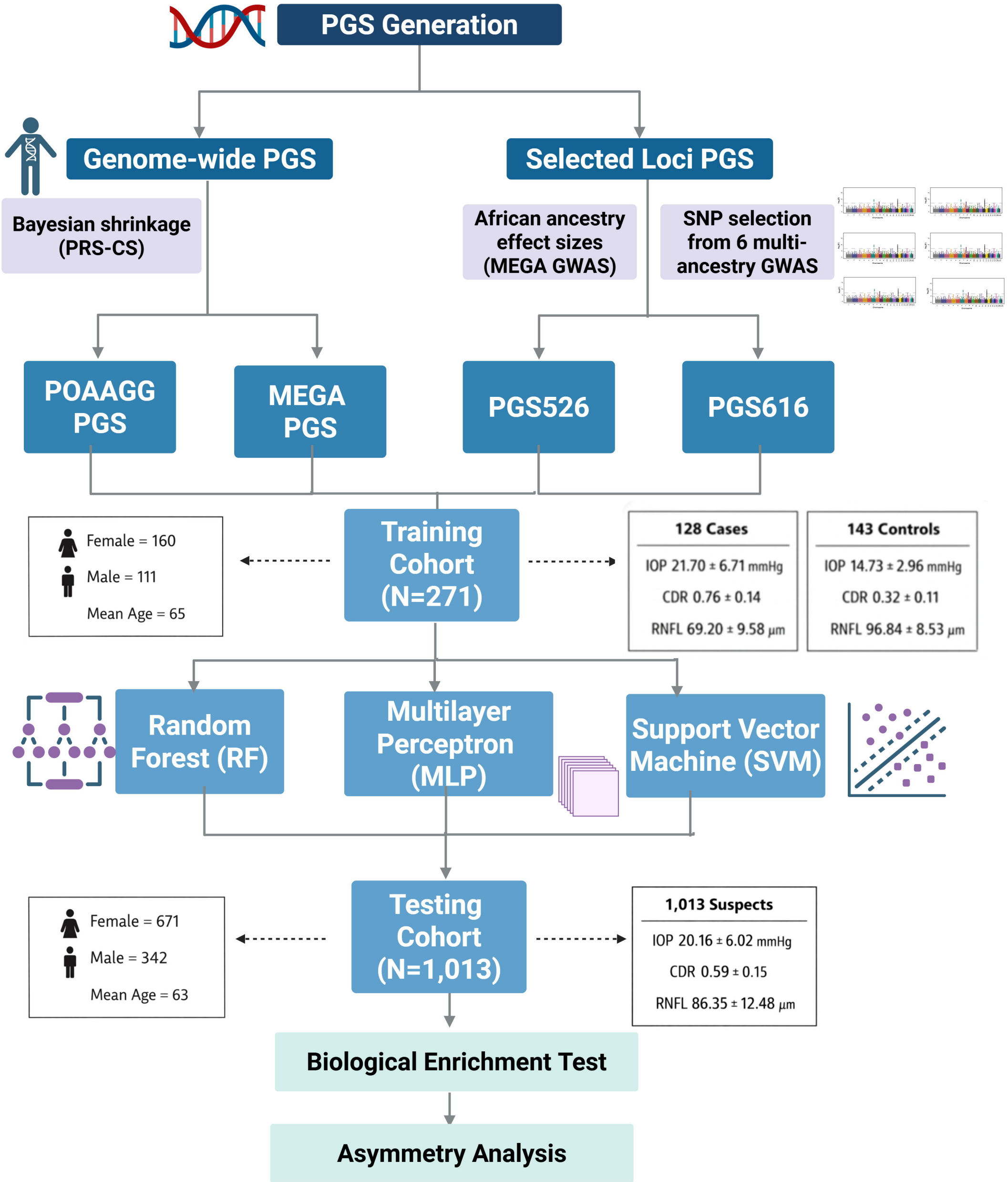
INTER-EYE ASYMMETRY



PGS Amplify early disease discrimination

PERFORMANCE IMPROVEMENT

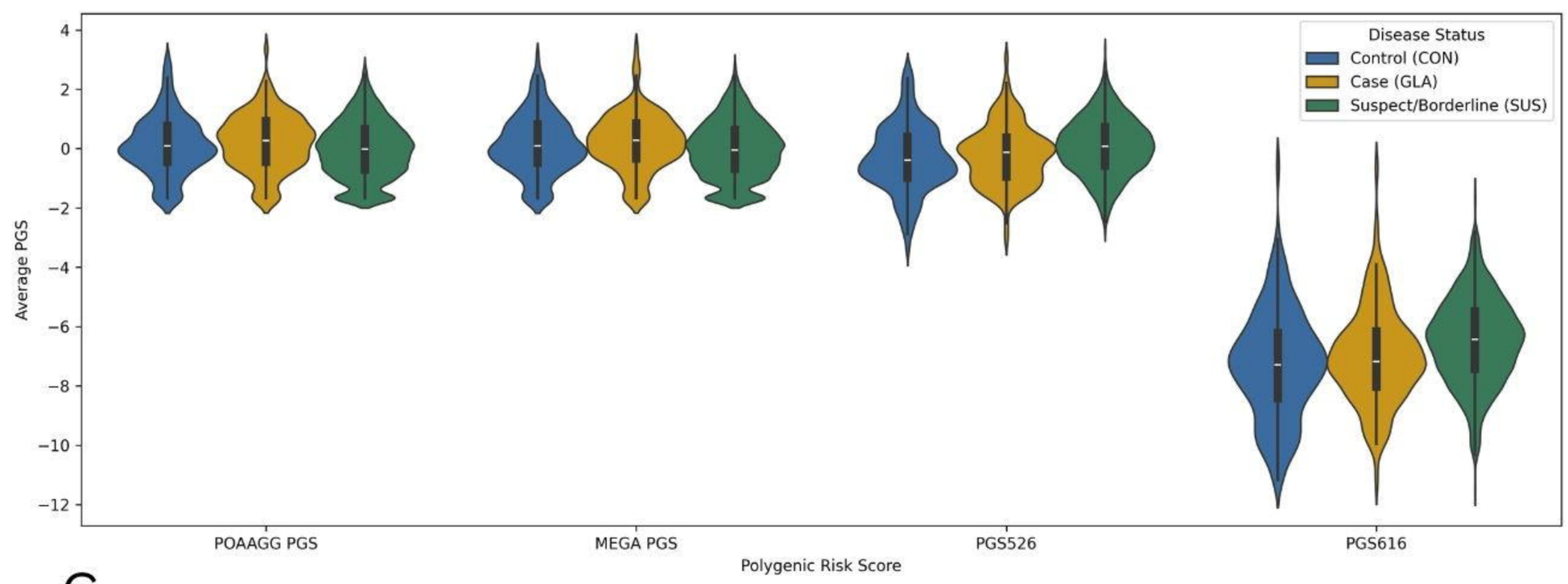
PGS improve early structural detection


PGS Generation
Genome-wide PGS
Selected Loci PGS
Bayesian shrinkage (PRS-CS)
African ancestry effect sizes (MEGA GWAS)
SNP selection from 6 multi-ancestry GWAS
POAAGG PGS
MEGA PGS
PGS526
PGS616
Training Cohort (N=271)
Female = 160
Male = 111
Mean Age = 65
128 Cases
IOP 21.70 ± 6.71 mmHg
CDR 0.76 ± 0.14
RNFL 69.20 ± 9.58 µm
143 Controls
IOP 14.73 ± 2.96 mmHg
CDR 0.32 ± 0.11
RNFL 96.84 ± 8.53 µm
Random Forest (RF)
Multilayer Perceptron (MLP)
Support Vector Machine (SVM)
Testing Cohort (N=1,013)
Female = 671
Male = 342
Mean Age = 63
1,013 Suspects
IOP 20.16 ± 6.02 mmHg
CDR 0.59 ± 0.15
RNFL 86.35 ± 12.48 µm
Biological Enrichment Test
Asymmetry Analysis

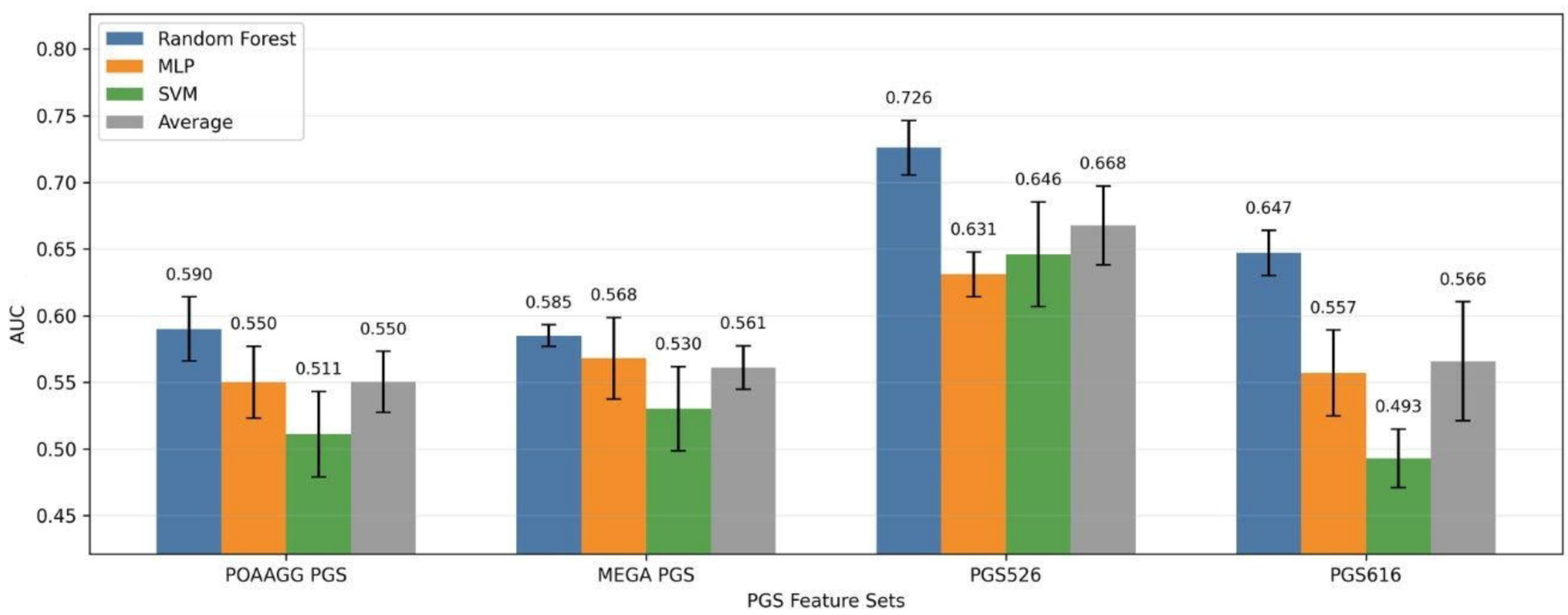
A

chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22

B


Disease Status
Control (CON)
Case (GLA)
Suspect/Borderline (SUS)
Average PGS
4
2
0
−2
−4
−6
−8
−10
−12
POAAGG PGS
MEGA PGS
PGS526
PGS616
Polygenic Risk Score

C

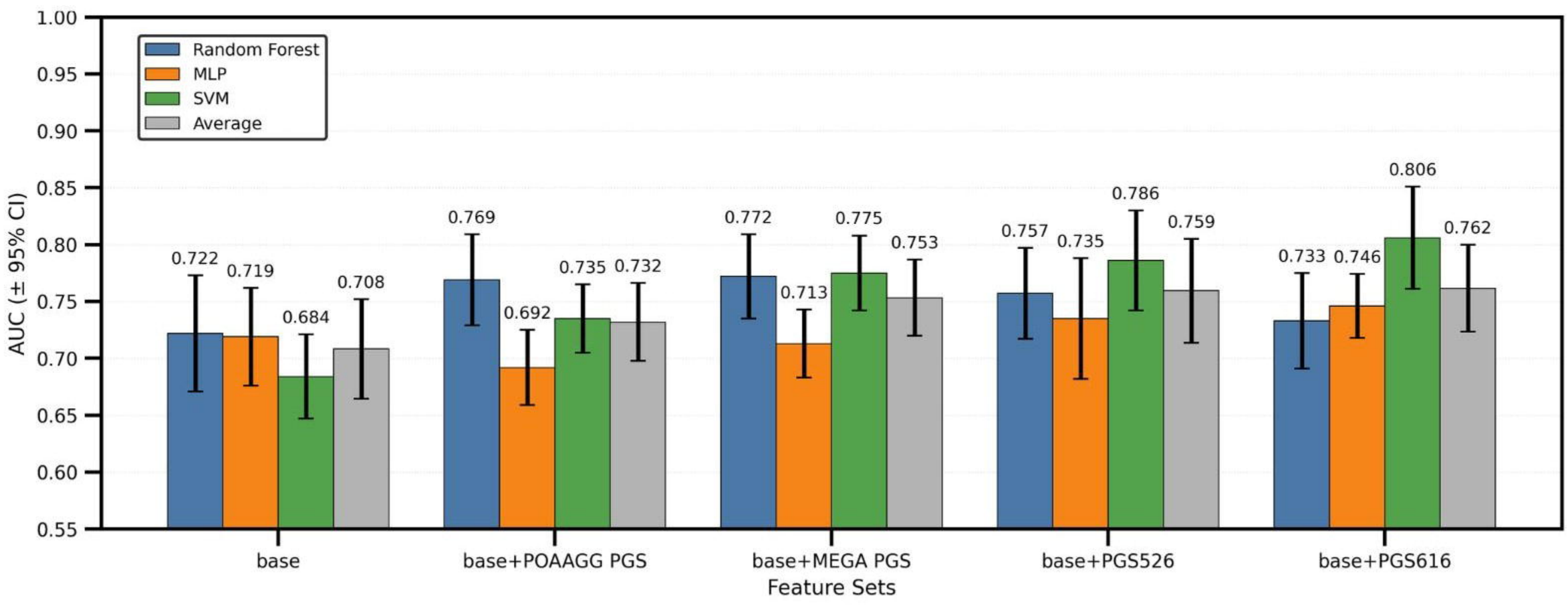

Random Forest
MLP
SVM
Average
AUC
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.590
0.550
0.511
0.550
0.585
0.568
0.530
0.561
0.726
0.631
0.646
0.668
0.647
0.557
0.493
0.566
POAAGG PGS
MEGA PGS
PGS526
PGS616
PGS Feature Sets

A

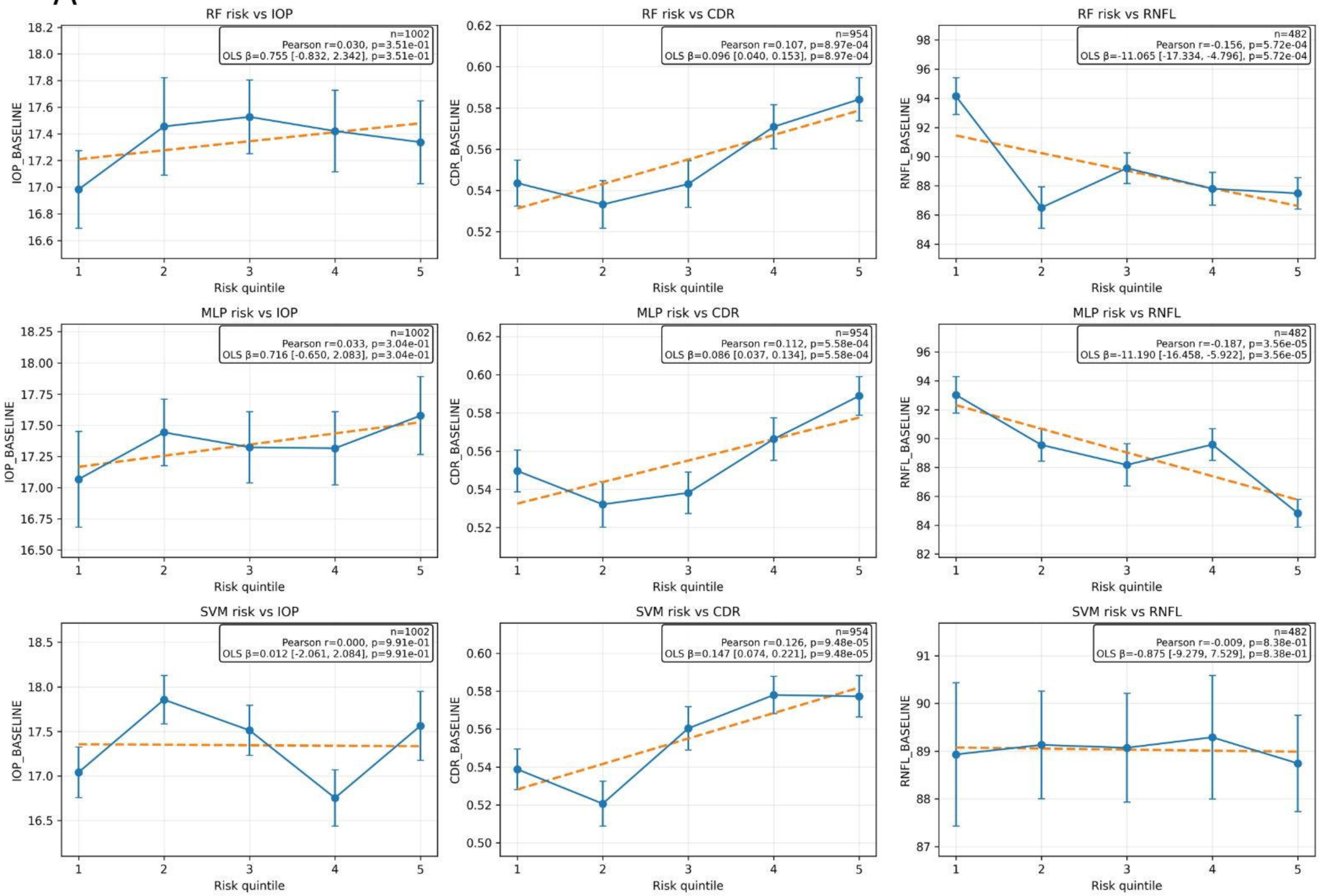
		base	base+POAAGG PGS	base+MEGA PGS	base+PGS526	base+PGS616
MLP	AUC	0.719	0.692	0.713	0.735	0.746
	Accuracy	0.653	0.620	0.664	0.675	0.684
	F1	0.636	0.568	0.641	0.652	0.678
	Sensitivity	0.636	0.546	0.636	0.629	0.650
	Specificity	0.669	0.709	0.688	0.720	0.717
SVM	AUC	0.684	0.735	0.775	0.786	0.806
	Accuracy	0.651	0.696	0.700	0.747	0.760
	F1	0.626	0.670	0.687	0.718	0.749
	Sensitivity	0.588	0.658	0.694	0.724	0.772
	Specificity	0.715	0.729	0.714	0.773	0.752
Random Forest	AUC	0.722	0.769	0.772	0.757	0.733
	Accuracy	0.660	0.707	0.711	0.676	0.673
	F1	0.648	0.691	0.708	0.657	0.643
	Sensitivity	0.666	0.700	0.738	0.715	0.618
	Specificity	0.662	0.708	0.694	0.648	0.723

Mean: 0.80, 0.75, 0.70, 0.65, 0.60, 0.55

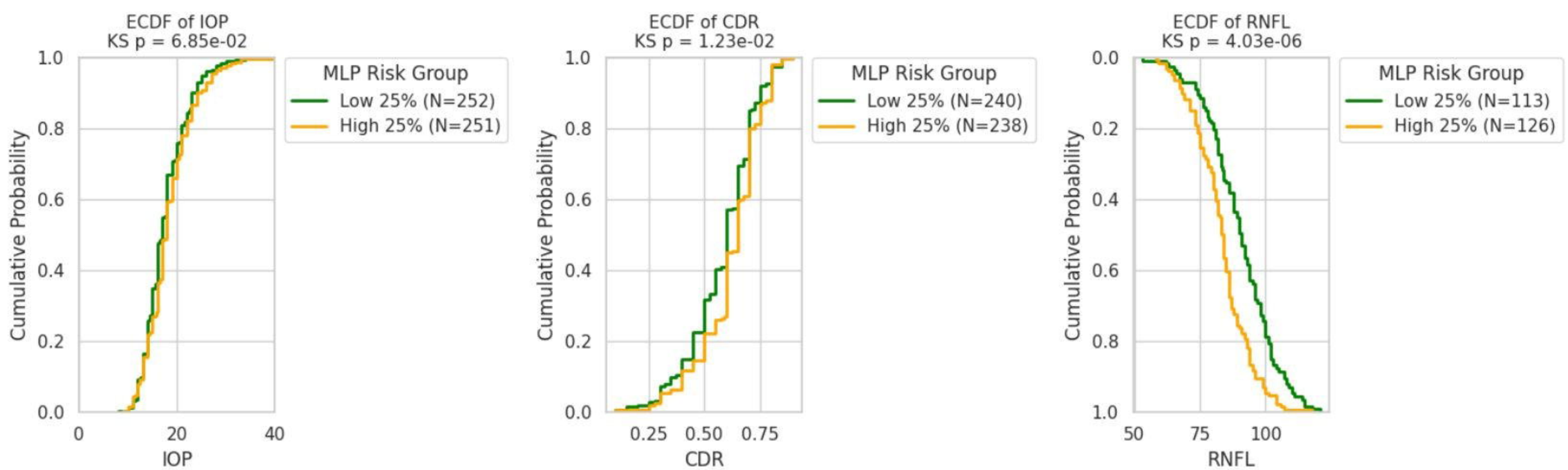
B



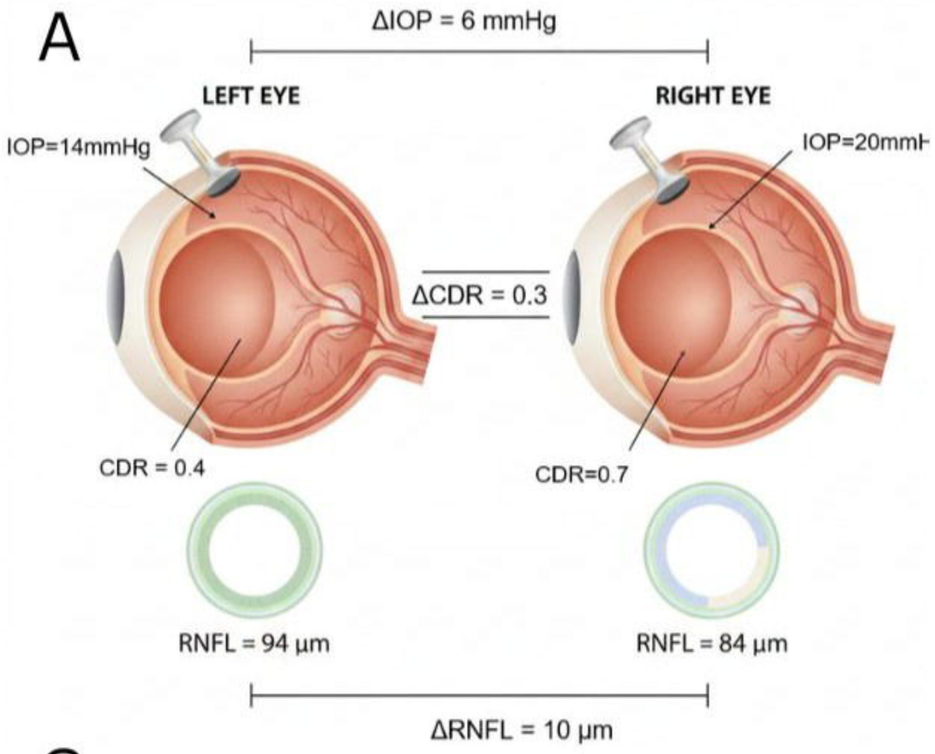
A


RF risk vs IOP
n=1002
Pearson r=0.030, p=3.51e-01
OLS β=0.755 [-0.832, 2.342], p=3.51e-01
IOP_BASELINE
Risk quintile
RF risk vs CDR
n=954
Pearson r=0.107, p=8.97e-04
OLS β=0.096 [0.040, 0.153], p=8.97e-04
CDR_BASELINE
Risk quintile
RF risk vs RNFL
n=482
Pearson r=-0.156, p=5.72e-04
OLS β=-11.065 [-17.334, -4.796], p=5.72e-04
RNFL_BASELINE
Risk quintile
MLP risk vs IOP
n=1002
Pearson r=0.033, p=3.04e-01
OLS β=0.716 [-0.650, 2.083], p=3.04e-01
IOP_BASELINE
Risk quintile
MLP risk vs CDR
n=954
Pearson r=0.112, p=5.58e-04
OLS β=0.086 [0.037, 0.134], p=5.58e-04
CDR_BASELINE
Risk quintile
MLP risk vs RNFL
n=482
Pearson r=-0.187, p=3.56e-05
OLS β=-11.190 [-16.458, -5.922], p=3.56e-05
RNFL_BASELINE
Risk quintile
SVM risk vs IOP
n=1002
Pearson r=0.000, p=9.91e-01
OLS β=0.012 [-2.061, 2.084], p=9.91e-01
IOP_BASELINE
Risk quintile
SVM risk vs CDR
n=954
Pearson r=0.126, p=9.48e-05
OLS β=0.147 [0.074, 0.221], p=9.48e-05
CDR_BASELINE
Risk quintile
SVM risk vs RNFL
n=482
Pearson r=-0.009, p=8.38e-01
OLS β=-0.875 [-9.279, 7.529], p=8.38e-01
RNFL_BASELINE
Risk quintile

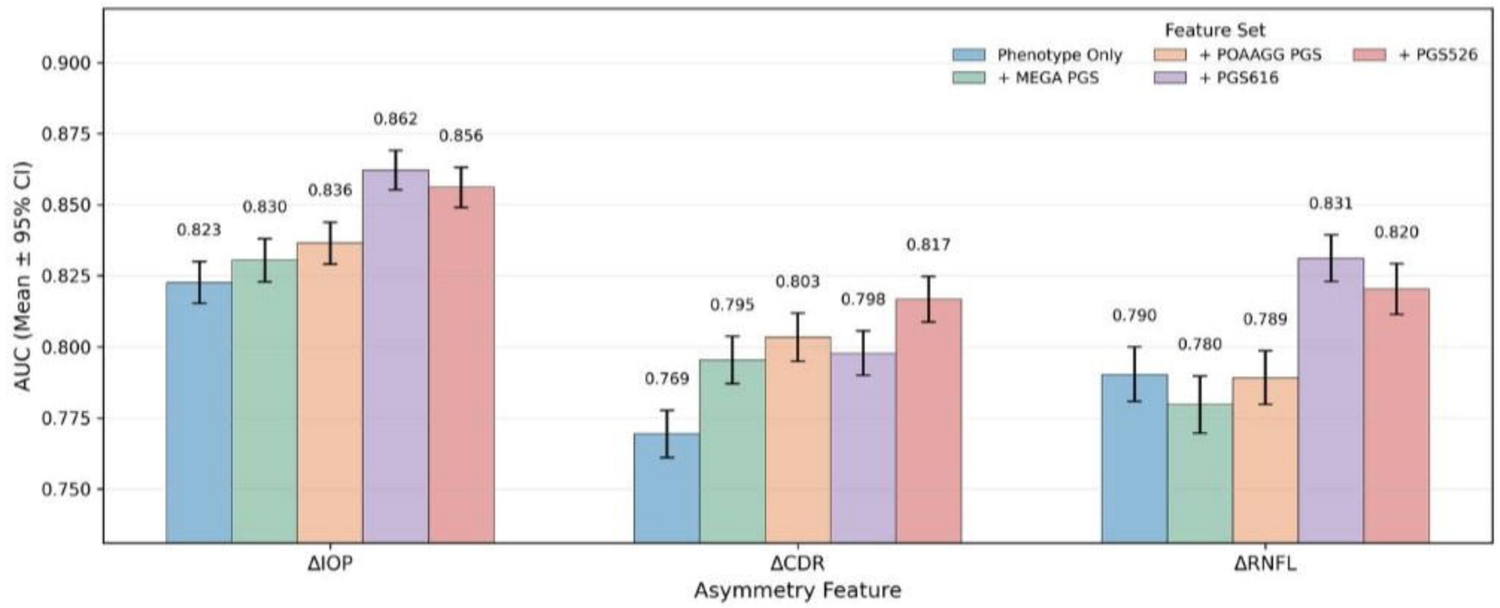
B


ECDF of IOP
KS p = 6.85e-02
MLP Risk Group
Low 25% (N=252)
High 25% (N=251)
Cumulative Probability
IOP
ECDF of CDR
KS p = 1.23e-02
MLP Risk Group
Low 25% (N=240)
High 25% (N=238)
Cumulative Probability
CDR
ECDF of RNFL
KS p = 4.03e-06
MLP Risk Group
Low 25% (N=113)
High 25% (N=126)
Cumulative Probability
RNFL

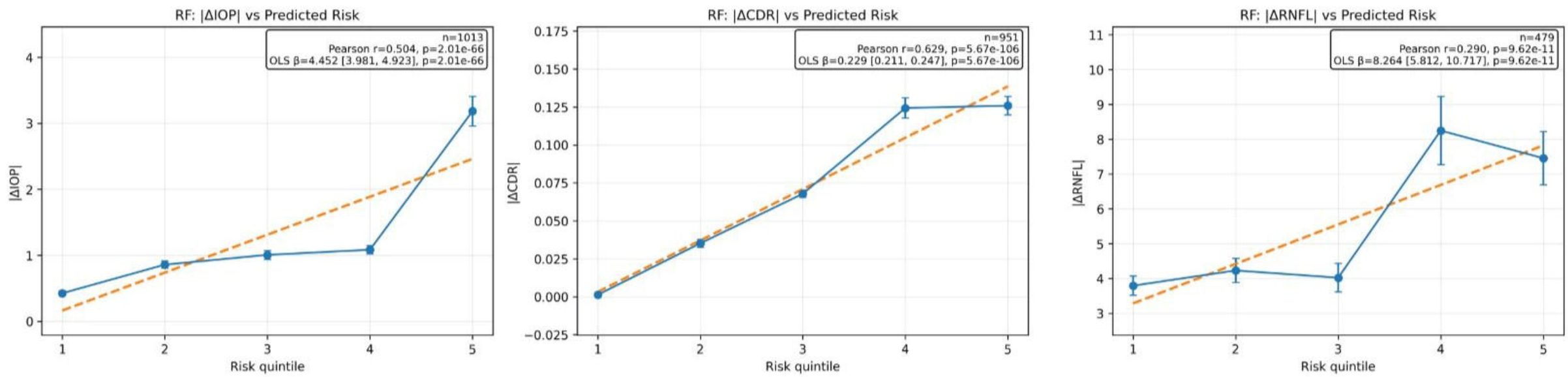
A


ΔIOP = 6 mmHg
LEFT EYE
RIGHT EYE
IOP=14mmHg
IOP=20mmH
ΔCDR = 0.3
CDR = 0.4
CDR=0.7
RNFL = 94 µm
RNFL = 84 µm
ΔRNFL = 10 µm

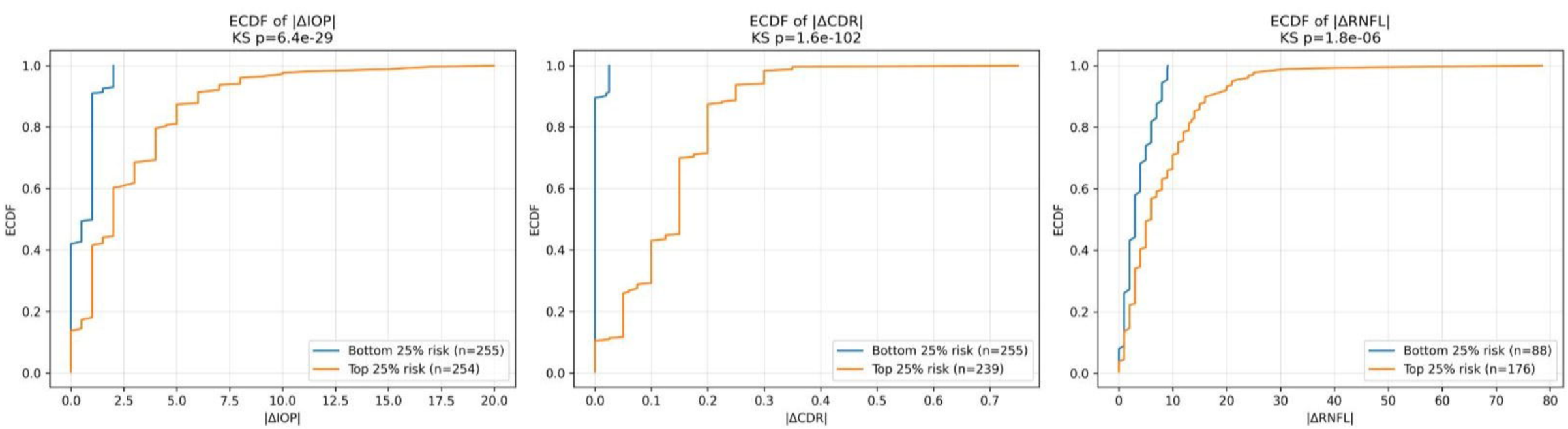
B


Feature Set
Phenotype Only
+ MEGA PGS
+ POAAGG PGS
+ PGS616
+ PGS526
AUC (Mean ± 95% CI)
0.823
0.830
0.836
0.862
0.856
0.769
0.795
0.803
0.798
0.817
0.790
0.780
0.789
0.831
0.820
ΔIOP
ΔCDR
ΔRNFL
Asymmetry Feature

C


RF: |ΔIOP| vs Predicted Risk
n=1013
Pearson r=0.504, p=2.01e-66
OLS β=4.452 [3.981, 4.923], p=2.01e-66
|ΔIOP|
Risk quintile
RF: |ΔCDR| vs Predicted Risk
n=951
Pearson r=0.629, p=5.67e-106
OLS β=0.229 [0.211, 0.247], p=5.67e-106
|ΔCDR|
Risk quintile
RF: |ΔRNFL| vs Predicted Risk
n=479
Pearson r=0.290, p=9.62e-11
OLS β=8.264 [5.812, 10.717], p=9.62e-11
|ΔRNFL|
Risk quintile

D


ECDF of |ΔIOP|
KS p=6.4e-29
Bottom 25% risk (n=255)
Top 25% risk (n=254)
ECDF
|ΔIOP|
ECDF of |ΔCDR|
KS p=1.6e-102
Bottom 25% risk (n=255)
Top 25% risk (n=239)
ECDF
|ΔCDR|
ECDF of |ΔRNFL|
KS p=1.8e-06
Bottom 25% risk (n=88)
Top 25% risk (n=176)
ECDF
|ΔRNFL|