

Background Polygenic Risk Modulates the Association between Glaucoma and Cardiopulmonary Diseases and Measures: an Analysis from the UK Biobank.

Supplementary File

Supplement 1: International Classification of Disease, Ninth (ICD-9) and Tenth (ICD-10) codes used for the identification of cardiopulmonary diseases.

Type 2 Diabetes Mellitus:

- ICD-9: 250.00, 250.09, 250.10, 250.20, 250.3, 250.4, 250.5, 250.6, 250.7, 250.8, 250.9, 250.90
- ICD-10: E11.XX

Hypertension:

- ICD-9: 401.XX
- ICD-10: I10.XX

Dyslipidemia:

- ICD-9: 272.XX
- ICD-10: E78.XX

Cardiovascular Disease:

- ICD-9: 410.XX, 411.XX, 412.XX, 413.XX, 414.XX
- ICD-10: I20.XX, I21.XX, I22.XX, I24.XX, I25.XX

Stroke

- ICD-9: 434.XX
- ICD-10: I63.XX, I64.XX

Chronic Kidney Disease

- ICD-9: 585.XX
- ICD-10: N18.XX, N19.XX

Chronic Obstructive Pulmonary Disease

- ICD-9: 491.XX, 492.XX, 494.XX
- ICD-10: J41.XX, J42.XX, J43.XX, J44.XX

Supplement 2: Supplementary Methods Technical Details

Technical Details of Genotyping and Creation of the Primary Open Angle Glaucoma Polygenic Risk Score

Bycroft et al. has previously described the array genotype curation process implemented by the UK Biobank in detail.[1] Using PLINK 1.9 and best practice approaches from the GTEx consortium, we applied further quality control steps on directly genotyped variants of UKBB samples.[2] This process iteratively examines genotype efficiency, gender discrepancies, allele frequencies, Hardy-Weinberg equilibrium tests, and variants, including indel and single nucleotide polymorphisms [SNPs]. Sample duplicates, contamination, and cryptic relatedness were also evaluated.

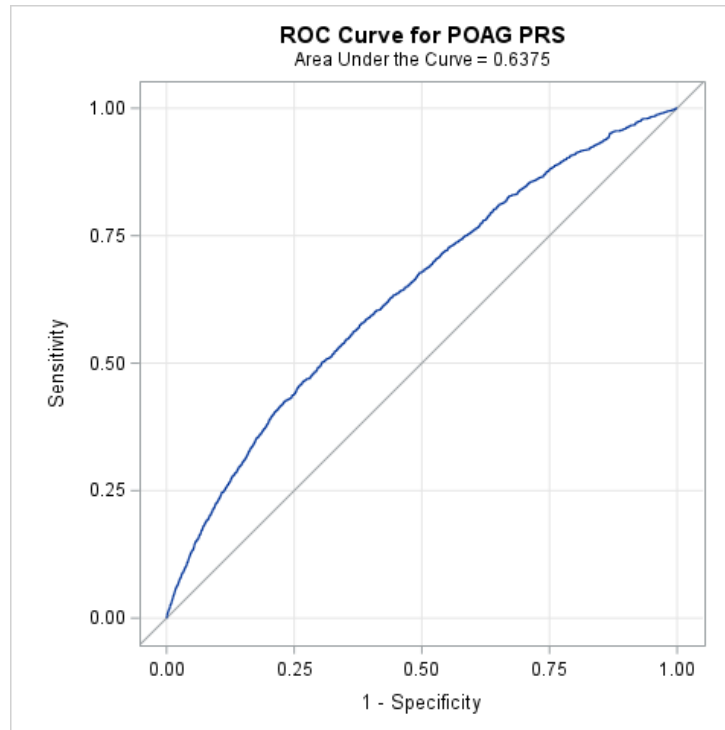
Via this process, we removed participants with unresolved differences between genotype-inferred and reported sex (n=449), genotyping call rate <97% (n=229), high cryptic relatedness defined as >0.1875 $\hat{\pi}$ (n=3,381), and outlying heterozygosity defined as 4 standard deviations from the mean heterozygosity rate after accounting for inferred ancestry (n=98).

To predict the ancestral background of participants using ancestral labels from the 1000 Genomes Project Phase 3 reference panel, Principal Component Analysis (PCA) to linkage disequilibrium (LD)-pruned ($r^2 < 0.1$ in 200kb windows) genetic markers with minor allele frequency (MAF) >1% and the k-nearest neighbors algorithm were used. This method demonstrated good correlation between self-reported and inferred ancestry. We used the inferred ancestry for quality control (QC) and downstream analyses for those with mismatched ancestry. The present study only includes participants with inferred European ancestry. Variants with call rate < 97%, MAF < 0.01 and Hardy-Weinberg equilibrium test $p < 1e-5$ were removed. As a result, 533,176 variants were used in the creation of the POAG PRS.

Data from genome-wide associate study (GWAS) summary statistics from the Caucasian subset of a large cross-ancestry meta-analysis[3] was used to create a POAG PRS. The POAG PRS was created using *LDpred2*[4] in the R package *bigsnpr*[5] The summary statistics used are publicly available (<https://segrelab.meei.harvard.edu/data/>). Data from the UKBB cohort was excluded in the creation of the PRS, and the PRS was subsequently applied to 442,097 UKBB participants' data. The posterior mean effect sizes from GWAS

summary statistics were estimated using a point-normal mixture prior for the variant effects and were adjusted for linkage disequilibrium. Model hyperparameters included SNP heritability (h^2) and the fraction of casual variants (p). The posterior mean effect sizes from GWAS summary statistics were estimated using a point-normal mixture prior for the variant effects and were adjusted for linkage disequilibrium. The hyperparameters in the model included SNP heritability (h^2) and the fraction of casual variants (p). We ran :Dpred2-grid to tune the hyperparameters with p from a sequence of 21 values from 10^{-5} to 1 on log-scale, and $h^2 = (0.7, 1, 1.4) * h^2_{LDSC}$ where h^2_{LDSC} is the heritability estimated from the LD score regression. Hyperparameters to construct the final PRS were chosen based on the best prediction performance measured by the area under the curve (AUC). Two definitions of glaucoma cases were used to create the PRS: 1) ICD9/10 diagnosis code (747 cases, 75624 controls, 1,588 missing) and 2) combination of ICD 9/10 diagnosis code and glaucoma self-report (2001 cases and 75624 controls). The model using ICD 9/10 diagnosis codes only demonstrated better performance and was thus used in the following analysis.

A receiver operating curve with area under the curve was calculated for the POAG PRS as a predictor of glaucoma, as shown below. In a logistic regression model of POAG PRS as a predictor of glaucoma yielded a c statistic of 0.637. A c statistic in this range was expected, given that glaucoma is a complex polygenic disease with both genetic and non-genetic risk factors. Similar c statistics have been reported for PRSs predicting other common polygenic diseases, such as type 2 diabetes mellitus, depression, and asthma.[6-8]



1. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*. 2017:166298.
2. Aguet F, Brown AA, Castel SE, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213.
3. Gharahkhani P, Jorgenson E, Hysi P, et al. Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nat Commun* 2021;12:1258. doi:10.1038/s41467-020-20851-4
4. Privé F, Arbel J, Vilhjálmsón BJ. LDpred2: better, faster, stronger. *Bioinformatics*. Published online December 16, 2020. doi:10.1093/bioinformatics/btaa1029
5. Privé F, Aschard H, Ziyatdinov A, Blum MGB. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics*. 2018;34(16):2781-2787. doi:10.1093/bioinformatics/bty185
6. Liu W, Zhuang Z, Wang W, Huang T, Liu Z. An Improved Genome-Wide Polygenic Score Model for Predicting the Risk of Type 2 Diabetes. *Front Genet*. 2021 Feb 11;12:632385. doi: 10.3389/fgene.2021.632385.
7. Sordillo JE, Lutz SM, Jorgenson E, Iribarren C, McGeachie M, Dahlin A, Tantisira K, Kelly R, Lasky-Su J, Sakornsakolpat P, Moll M, Cho MH, Wu AC. A polygenic risk score for asthma in a large racially diverse population. *Clin Exp Allergy*. 2021 Nov;51(11):1410-1420. doi: 10.1111/cea.14007. Epub 2021 Sep 5.
8. Halldorsdottir T, Piechaczek C, Soares de Matos AP, Czamara D, Pehl V, Wagenbuechler P, Feldmann L, Quickenstedt-Reinhardt P, Allgaier AK,

Freisleder FJ, Greimel E, Kvist T, Lahti J, Räikkönen K, Rex-Haffner M, Arnarson EÖ, Craighead WE, Schulte-Körne G, Binder EB. Polygenic Risk: Predicting Depression Outcomes in Clinical and Epidemiological Cohorts of Youths. *Am J Psychiatry*. 2019 Aug 1;176(8):615-625. doi: 10.1176/appi.ajp.2019.18091014. Epub 2019 Apr 5.

Technical Details of Ganglion Cell Complex Thickness Assessment

Spectral domain OCT scans of the macula were obtained on a subset of 67,321 participants using Topcon 3D OCT 1000 Mk2 (Topcon, Inc, Japan).[35] All OCT images were stored in .fda image files without prior analysis of macular thickness. Topcon Advanced Boundary Segmentation (TABS) algorithm was used to automatically segment all scans, using dual-scale gradient information to allow for automated segmentation of the inner and outer retinal boundaries and retinal sublayers.[9] The software provides an image quality score and segmentation indicators which was used for quality control. Segmentation indicators included the Inner Limiting Membrane (ILM) Indicator, a measure of the minimum localized edge strength around the ILM boundary across the scan, can be used to identify blinks, scans that contain regions of signal fading, and errors in segmentation. We excluded all images with image quality less than 40 and images representing the poorest 10% of ILM indicator.[10] We also excluded any image with a layer thickness greater than 2.5 standard deviations away from the mean. Ganglion cell complex (GCC) thickness was defined as the total thickness of the retinal nerve fiber layer, retinal ganglion cell layer, and inner plexiform layer. GCC thickness was measured for both

eyes of each participant, and the thinner GCC thickness value was used in our analyses.

9. Keane PA, Grossi CM, Foster PJ, *et al.* Optical Coherence Tomography in the UK Biobank Study - Rapid Automated Analysis of Retinal Thickness for Large Population-Based Studies. *PLoS One* 2016;**11**:e0164095.
doi:10.1371/journal.pone.0164095
10. Curren H, Hysi P, Fitzgerald TW, *et al.* Genetic variation affects morphological retinal phenotypes extracted from UK Biobank optical coherence tomography images. *PLOS Genetics* 2021;17:e1009497. doi:10.1371/journal.pgen.1009497

Supplement 3: Cardiopulmonary Characteristics by Glaucoma Status and Primary Open Angle Glaucoma Polygenic Risk Score Decile

| Characteristic | Decile 1 | | Decile 2 | | Decile 3 | | Decile 4 | | Decile 5 | |
|---------------------------------------|----------------|---------------------|----------------|---------------------|----------------|---------------------|----------------|---------------------|----------------|----------------------|
| | Cases n=417 | Controls n=44458 | Cases n=525 | Controls n=44286 | Cases n=648 | Controls n=44162 | Cases n=742 | Controls n=44013 | Cases n=801 | Controls n=43,926 |
| Medical History | Prevalence | | | | | | | | | |
| Diabetes | 17.5% | 6.5% | 14.1% | 6.6% | 12.7% | 6.6% | 12.7% | 6.7% | 14.7% | 6.4% |
| Hypertension | 50.1% | 33.4% | 48.4% | 33.5% | 46.1% | 33.2% | 49.3% | 32.9% | 49.6% | 32.5% |
| Dyslipidemia | 31.2% | 18.3% | 30.5% | 18.2% | 31.3% | 18.1% | 31.4% | 17.9% | 28.8% | 17.5% |
| Cardiovascular Disease | 18.9% | 9.2% | 13.0% | 9.4% | 15.3% | 9.3% | 16.0% | 9.2% | 17.2% | 9.0% |
| Stroke | 1.9% | 1.3% | 3.8% | 1.3% | 2.9% | 1.3% | 2.6% | 1.3% | 2.5% | 1.3% |
| Chronic Kidney Disease | 6.7% | 2.0% | 4.4% | 2.0% | 5.7% | 2.1% | 4.3% | 2.0% | 4.9% | 2.1% |
| Chronic Obstructive Pulmonary Disease | 9.4% | 4.1% | 8.4% | 3.9% | 7.4% | 3.9% | 8.0% | 3.9% | 5.4% | 3.8% |
| Examination Measure | Mean | | | | | | | | | |
| BMI (kg/m ²) | 27.2 | 26.9 | 27.4 | 26.9 | 27.2 | 26.9 | 27.0 | 26.9 | 27.0 | 26.9 |
| Systolic BP (mmHg) | 140.9 | 136.9 | 140.9 | 136.8 | 140.5 | 136.6 | 140.3 | 136.7 | 140.2 | 136.6 |
| Diastolic BP (mmHg) | 81.5 | 81.8 | 82.1 | 81.8 | 82.0 | 81.8 | 81.9 | 81.7 | 82.1 | 81.7 |
| Pulse Rate | 69.1 | 68.3 | 69.3 | 68.4 | 68.7 | 68.3 | 68.0 | 68.4 | 68.6 | 68.4 |
| Waist to Hip Ratio | 0.89 | 0.87 | 0.89 | 0.87 | 0.88 | 0.87 | 0.88 | 0.87 | 0.89 | 0.87 |
| FEV1/FVC Ratio | 0.40 | 0.35 | 0.33 | 0.36 | 0.33 | 0.35 | 0.31 | 0.35 | 0.36 | 0.36 |

| Laboratory Measure | Mean | | | | | | | | | |
|----------------------|---------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | Hemoglobin A1C (mmol/mol) | 38.51 | 35.92 | 38.28 | 35.94 | 37.86 | 35.92 | 37.93 | 35.93 | 38.24 |
| Cholesterol (mmol/L) | 20.43 | 21.18 | 20.59 | 21.31 | 20.61 | 21.36 | 20.95 | 21.39 | 20.53 | 21.47 |
| HDL (mmol/L) | 1.42 | 1.45 | 1.45 | 1.45 | 1.44 | 1.45 | 1.44 | 1.45 | 1.41 | 1.45 |
| LDL (mmol/L) | 3.54 | 3.58 | 3.43 | 3.58 | 3.55 | 3.57 | 3.43 | 3.56 | 3.47 | 3.57 |
| Vitamin D (nmol/L) | 47.75 | 49.45 | 48.28 | 49.70 | 49.72 | 49.57 | 50.87 | 49.51 | 48.24 | 49.44 |
| Creatinine (umol/L) | 74.16 | 72.29 | 73.93 | 72.26 | 73.63 | 72.37 | 73.88 | 72.25 | 73.39 | 72.32 |

| Characteristic | Decile 6 | | Decile 7 | | Decile 8 | | Decile 9 | | Decile 10 | |
|------------------------|----------------|---------------------|-----------------|---------------------|-----------------|---------------------|-----------------|---------------------|-----------------|---------------------|
| | Cases n=900 | Controls n=43807 | Cases n=1007 | Controls n=43705 | Cases n=1155 | Controls n=43504 | Cases n=1449 | Controls n=43144 | Cases n=2135 | Controls n=42413 |
| Medical History | Prevalence | | | | | | | | | |
| Diabetes | 12.9% | 6.6% | 12.0% | 6.5% | 11.7% | 6.6% | 12.1% | 6.4% | 9.9% | 6.4% |
| Hypertension | 51.1% | 32.6% | 47.4% | 32.8% | 47.0% | 32.4% | 45.5% | 32.2% | 43.8% | 32.4% |
| Dyslipidemia | 29.9% | 18.0% | 28.0% | 18.0% | 27.4% | 18.0% | 27.1% | 17.4% | 27.7% | 17.3% |
| Cardiovascular Disease | 16.3% | 9.1% | 14.7% | 9.1% | 13.3% | 9.3% | 16.0% | 9.0% | 15.1% | 9.2% |
| Stroke | 2.6% | 1.3% | 2.5% | 1.3% | 1.8% | 1.3% | 3.0% | 1.3% | 2.1% | 1.2% |
| Chronic Kidney Disease | 4.3% | 2.0% | 3.9% | 2.0% | 4.5% | 1.9% | 3.8% | 1.9% | 4.0% | 2.1% |

| | | | | | | | | | | |
|---------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Chronic Obstructive Pulmonary Disease | 7.1% | 4.0% | 7.8% | 4.0% | 6.4% | 4.0% | 7.5% | 3.8% | 5.4% | 4.0% |
| Examination Measure | Mean | | | | | | | | | |
| BMI | 27.2 | 26.8 | 26.9 | 26.8 | 27.1 | 26.8 | 27.0 | 26.8 | 27.0 | 26.8 |
| Systolic BP | 141.0 | 136.7 | 141.8 | 136.7 | 140.5 | 136.7 | 140.3 | 136.6 | 140.1 | 136.6 |
| Diastolic BP | 82.1 | 81.6 | 82.5 | 81.7 | 82.4 | 81.7 | 81.9 | 81.7 | 81.9 | 81.7 |
| Pulse Rate | 67.9 | 68.4 | 68.4 | 68.4 | 68.5 | 68.5 | 68.4 | 68.3 | 68.0 | 68.4 |
| Waist to Hip Ratio | 0.88 | 0.87 | 0.88 | 0.87 | 0.89 | 0.87 | 0.88 | 0.87 | 0.88 | 0.87 |
| FEV1/FVC Ratio | 0.38 | 0.35 | 0.42 | 0.36 | 0.38 | 0.35 | 0.36 | 0.35 | 0.40 | 0.36 |
| Laboratory Measure | Mean | | | | | | | | | |
| Hemoglobin A1C | 37.99 | 35.95 | 37.38 | 35.94 | 37.25 | 35.95 | 37.26 | 35.86 | 36.88 | 35.93 |
| Cholesterol | 20.55 | 21.41 | 21.12 | 21.42 | 21.33 | 21.42 | 21.09 | 21.51 | 21.11 | 21.55 |
| HDL | 1.45 | 1.46 | 1.45 | 1.45 | 1.42 | 1.45 | 1.43 | 1.45 | 1.44 | 1.46 |
| LDL | 3.51 | 3.56 | 3.56 | 3.57 | 3.50 | 3.57 | 3.54 | 3.56 | 3.57 | 3.56 |
| Vitamin D | 49.80 | 49.39 | 48.44 | 49.46 | 49.92 | 49.36 | 49.55 | 49.38 | 49.94 | 49.32 |
| Creatinine | 73.21 | 72.20 | 72.58 | 72.19 | 75.24 | 72.08 | 73.67 | q | 73.33 | 72.09 |

Abbreviations: BMI: body mass index; BP: blood pressure, HDL: high density lipoprotein; LDL: low density lipoprotein. Decile 1 indicates lowest genetic risk for primary open angle glaucoma, and decile 10 indicates highest risk.

Supplement 4: Results of the sensitivity analysis assessing associations of Ganglion Cell Complex Thickness (Assessed by Ocular Coherence Tomography) with Cardiopulmonary Diseases and Factors that were Associated with Glaucoma in the Main Analysis.

Among the subset of individuals with available GCC thickness data, glaucoma cases had thinner GCC (n=739; mean: 96.7 μm ; SD: 10.0 μm) compared to controls (n= 38,730; mean: 102.4 μm ; SD: 8.2 μm). Mean (SD) spherical equivalents of refractive error was -0.30 (2.7) in glaucoma cases and -0.29 (2.3) in controls. Each 0.01 unit increase in POAG PRS was associated with a 25 μm unit decrease in GCC thickness (p=0.0003). In logistic regression models adjusted for age, gender, and refractive error (spherical equivalents), lower GCC thickness was not associated with higher odds of diabetes (OR: 1.016; 95% CI: 0.998, 1.033; p=0.08), dyslipidemia (OR: 1.007; 95% CI: 0.996, 1.018; p=0.20), CKD (OR: 1.004; 95% CI: 0.974, 1.034; p=0.79), or COPD (OR: 0.991; 95% CI: 0.969, 1.015; p=0.46) within decile 1. Within decile 10, smaller GCC thickness (per 1 μm unit decrease) was associated with higher adjusted odds of diabetes (OR: 1.031; 95% CI: 1.015, 1.048; p=0.0002) and dyslipidemia (OR 1.018; 95% CI: 1.008, 1.029; p=0.002), but not CKD (OR: 1.027; 95% CI: 0.997, 1.057; p=0.032) or COPD (OR: 1.012; 95% CI: 0.990, 1.034; p=0.030). Within decile 1, smaller GCC thickness (per 1 μm unit decrease) was associated with lower cholesterol (Beta: -0.032; 95% CI: -0.053, -0.010; p=0.0042), but was not associated with HbA1c (Beta: 0.004 mmol/mol; 95% CI: -0.020, 0.029; p=0.72). Within decile 10, lower GCC thickness (per 1 μm unit decrease) was associated with both lower cholesterol (Beta: -0.033 mmol/L, 95% CI: -0.012, -0.055; p=0.0024; p=0.012) and higher HbA1c (Beta: 0.026 mmol/mol; 95% CI: 0.009, 0.057; p=0.0063).