medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Phenotypic and Genetic Characteristics of Retinal Vascular Parameters and their Association with Diseases

Sofía Ortín Vela*†[1,2], Michael J. Beyeler*†[1,2], Olga Trofimova[1,2], Ilaria Iuliani[1,2], Jose D. Vargas Quiros[3,4], Victor A. de Vries[3,4], Ilenia Meloni[5], Adham Elwakil[5], Florence Hoogewoud[5], Bart Liefers[3,4], David Presby[1,2], Wishal D. Ramdas[3], Mattia Tomasoni[5], Reinier Schlingemann[5,6], Caroline C.W. Klaver[3,4,7,8], and Sven Bergmann†[1,2,9]

[1]Dept. of Computational Biology, University of Lausanne, Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Department of Ophthalmology, Erasmus MC University Medical Center, Rotterdam, The Netherlands
[4]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands
[5]Jules-Gonin Eye Hospital, FAA, University of Lausanne, Lausanne, Switzerland
[6]Department of Ophthalmology, Amsterdam University Medical Centres, Amsterdam, The Netherlands
[7]Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands
[8]Institute of Molecular and Clinical Ophthalmology, University of Basel, Basel, Switzerland
[9]Dept. of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa

Abstract

Fundus images allow for non-invasive assessment of the retinal vasculature whose features provide important information on health. Using a fully automated image processing pipeline, we extracted 17 different morphological vascular phenotypes, including median vessels diameter, diameter variability, main temporal angles, vascular density, central retinal equivalents, the number of bifurcations, and tortuosity, from over 130k fundus images of close to 72k UK Biobank subjects. We performed Genome-Wide Association Studies of these phenotypes. From this, we estimated their heritabilities, ranging between 5 and 25%, and genetic cross-phenotype correlations, which mostly mirrored the corresponding phenotypic correlations, but tended to be slightly larger. Projecting our genetic association signals onto genes and pathways revealed remarkably low overlap suggesting largely decoupled mechanisms modulating the different phenotypes. Our disease phenotype associations confirmed some previously known findings and revealed many novel connections. Notably, diameter variability, especially for the veins, seems to have new and interesting associations with diseases, including heart attack, pulmonary embolism, and age of death. Mendelian Randomization analysis suggests a causal influence of blood pressure and body mass index on retinal vessel morphology, among other results. We validated key findings in two independent smaller

*Equal first authors in random order
†E-mails: sofia.ortinvela@unil.ch, michael.beyeler@unil.ch, sven.bergmann@unil.ch

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

cohorts. Our analyses provide evidence that large-scale analysis of image-derived vascular phenotypes has sufficient power for obtaining functional, as well as some initial causal insights into the processes modulating the retinal vasculature.

Keywords: GWAS, retina, CFI, IDPs, vasculature, genes, vascular diseases, heritability, MR.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Nonstandard Abbreviations and Acronyms

BMI: Body mass index

CFI: Color fundus image

CNN: Convolutional neural network

DBP: Diastolic blood pressure

DL: Deep learning

DVT: Deep vein thrombosis

ED: Eye diseases

FDR: False discovery rate

HDL: High-density lipoprotein

HbA1c: Glycated hemoglobin

IDP: Image-derived phenotype

IVW: Inverse variance weighted

LDL: Low-density lipoprotein

LDSR: Linkage disequilibrium score regression

LOO: Leave-one-out

LWNET: Little W-Net

MR: Mendelian Randomisation

OD: Optic disc

PAD: Peripheral artery disease

PC: Principal component

PE: Pulmonary embolism

PR: Pulse rate

PWASI: Pulse wave arterial stiffness index

QC: Quality control

RS: Rotterdam Study

SBP: Systolic blood pressure

SNP: Single nucleotide polymorphism

UKBB: UK Biobank

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Introduction

The retina provides a unique opportunity for imaging human vasculature. In particular, retinal color fundus images (CFIs) allow for noninvasive in-vivo assessment of the vascular system of the superficial inner layer of the retina. Such images have been acquired in several cohorts and there is a large body of research on automatic extraction of vascular properties and their associations with medically relevant information.

It is well established that vascular properties obtained from retinal imaging not only enable the monitoring of ocular diseases such as diabetic retinopathy, macular degeneration, and glaucoma, but can also serve as a powerful screening tool for early detection of systemic diseases, including stroke [1–4], coronary heart disease [5, 6], peripheral artery diseases [7], hypertension [3, 4, 8–18], atherosclerosis [2, 3, 19], and myocardial infarction [20, 21]. Retinal abnormalities have also been associated with common comorbidities such as diabetes [8, 22–24] and obesity [25].

The processing of CFIs typically can be divided into three steps. First, the image is processed at the level of pixels to identify which of them represents blood vessels (possibly distinguishing between arteries and veins), or other structures, like the optic disc (OD). In the second step this pixel-wise information is used to annotate the retina in terms of objects, such as vessel segments represented as a list of points along their midline, as well as the vessel widths along these points. Finally, the information from these objects is used to measure different vascular properties such as vessel diameter or tortuosity, number of bifurcations, as well as certain angles between major vessels. Some simple vascular phenotypes, such as vascular density or fractal dimension, can also be computed directly from the pixel-wise information.

While there have been several studies analysing retinal vascular phenotypes [26–31], most of them focused on measuring just one or few retinal phenotypes, often in small image sets, and some required expert input rather than being fully automated [32–34]. Furthermore, the software used for vascular phenotyping is usually not openly accessible, with one very recent exception [35]. Together this precludes the establishment of a comprehensive and reproducible characterisation of large retinal image collections.

Recently, Deep Learning (DL) approaches have gained popularity in retinal image analysis. The first contribution is at the level of pixel-wise annotation, where state-of-the-art segmentation can be achieved with Convolutional Neural Network (CNN) architectures. For example, the little W-Net (LWNET) annotates pixels as being part of an artery or vein, outperforming classical segmentation approaches [36]. Such networks can also be used to annotate pixels belonging to the OD [37], vessel bifurcations [38], or other structures. The second contribution of deep CNNs is to learn latent variables providing efficient low-dimensional representations of retinal images [39, 40]. Such self-supervised, image-based phenotyping can generate novel phenotypes complementing explicit retinal features. Finally, DL approaches have been used to directly predict health-relevant phenotypes from retinal images [41, 42].

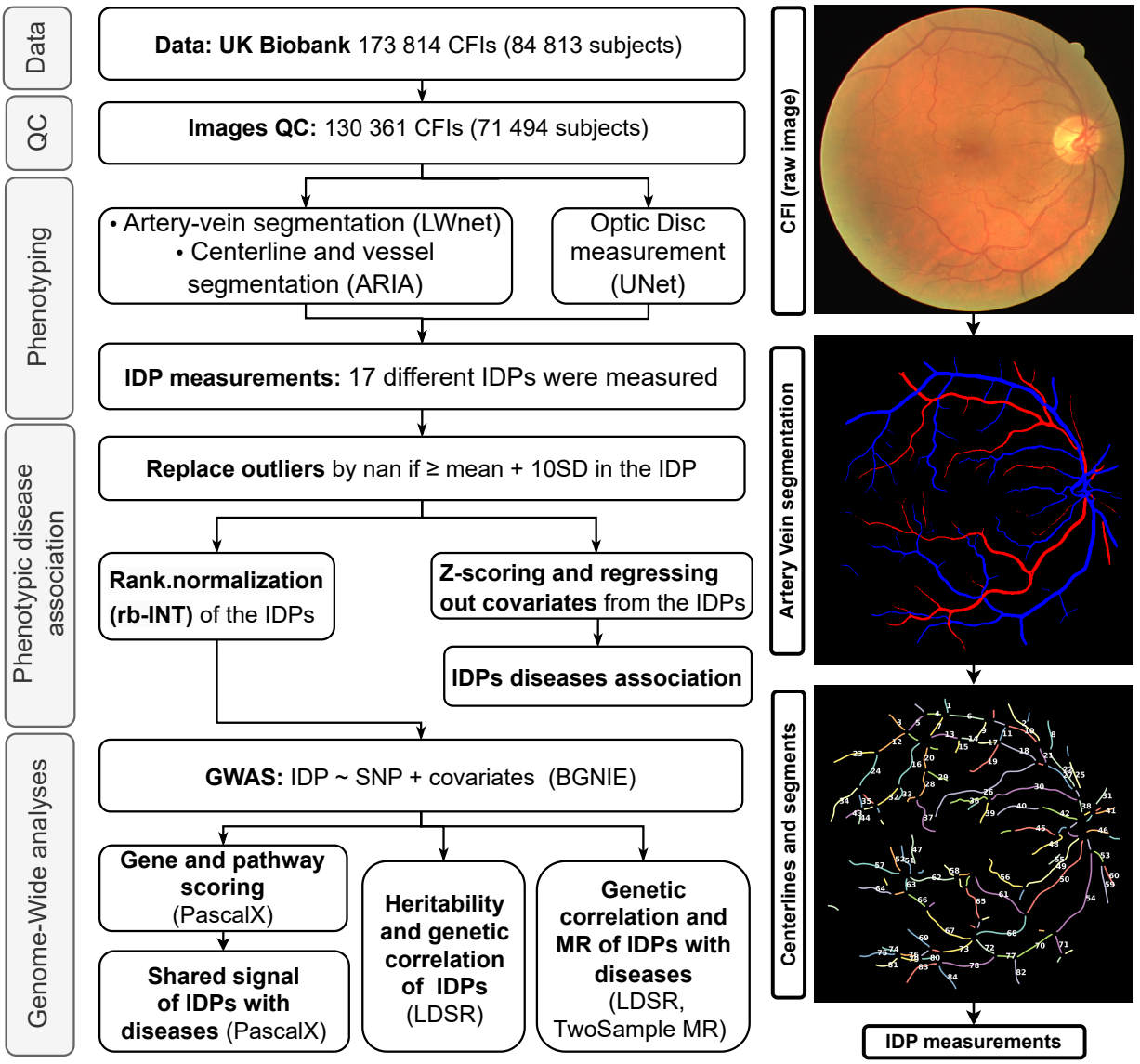
Beyond their value for assessing ocular or systemic health, retinal phenotypes have also been used in Genome-Wide Association Studies (GWAS) to identify genetic variants modulating these phenotypes. How-

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

ever, previous GWAS have focused on a limited set of vascular properties, such as vessel diameter [43–46], tortuosity [47, 48], vascular density, fractal dimension [49], and certain deep latent variables [39] (see Suppl. Text "Comparison with previous GWAS").

In this study, we present the first joint analysis of 17 retinal vascular phenotypes, including novel features like temporal angles, number of bifurcations, or diameter variability, which have not been analyzed previously at a large scale. Leveraging data from the UK Biobank (UKBB), and employing our open-source fully automated analysis platform, we provide retinal vascular phenotyping for close to 72k subjects, after quality control (QC). For all phenotypes, we performed GWAS, heritability estimates, gene and pathway analyses, as well as associations with a broad set of systemic and ocular diseases. This allowed us to compare cross-phenotype correlations, both at the phenotypic and genotypic levels, study in detail which genes affect individual or multiple phenotypes, and identify potential causal relationships with diseases (see Fig. 1 depicting the overall methodology of our discovery study). We reproduced the phenotypic correlation structure in two independent smaller cohorts: the Rotterdam Study (RS, N=8.1k) and OphtalmoLaus (N=2.2k) participants. Analysis of RS data also provided consistent estimates for phenotypes heritabilities and genetic correlations, as well as replication of a large number of our genome-wide significant hits for the vast majority of phenotypes.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .



(a) Discovery pipeline (b) CFI processing

Figure 1: **a)** Overview of discovery pipeline. Subjects' basic and medical information, genotypes, and CFIs were collected from the UKBB. Applying the Image QC method of [49] removed $\sim 25\%$ of all CFIs. Pixel-wise vessel segmentation and classification were performed using LWNET [36]. ARIA [50] was used to identify vessel segment objects. A DL network was used to measure the position of the OD [37]. Based on this primary information our bespoke algorithms measured vascular image-derived phenotypes (IDPs). IDPs (z-scored and corrected for covariates) were associated with diseases through linear and logistic regressions, and Cox models. GWAS was performed on all IDPs after rank-based inverse normal transformation (rb-INT) and correction for covariates, and the resulting summary statistics were used to estimate heritabilities and genetic correlations, to identify relevant genes and pathways, and to study the genetic association and potential causal relationships between the IDPs and some of the disease phenotypes. **b)** Overview of IDP measuring process. Top: original CFI from DRIVE dataset. Middle: Pixel-wise segmented vasculature with artery-vein classification using LWNET [36]. Bottom: Vessel segment objects in terms of centerlines and diameters were identified using ARIA [50], providing the starting point for measuring vascular IDPs (see Suppl. Text "Methods for phenotype extraction" for details).

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Results

Automated pipeline for phenotyping of the retinal vasculature

To analyze the retinal vasculature, we developed an automated pipeline building on our previous work on retinal vessel tortuosity [47]. This pipeline enabled us to segment and annotate the retinal vasculature and OD from 130 361 color fundus images (CFIs) of 71 494 subjects after Quality Control (QC). We used a previously established method for computing a QC score [49], and removed images in the lowest quartile of this score. Applying a range of quality control thresholds, we observed that phenotype heritabilities tended to be higher, while disease incident rates were lower when applying more stringent thresholds, possibly because some diseases impact the vascular density which is a good proxy for image quality (see Suppl. Text "Quality control threshold effect on results" for details).

We selected 17 representative image-derived phenotypes (IDPs) from a broader set of 36 phenotypes to characterize each image, including vascular densities, median tortuosities, central retinal equivalents, median diameters, diameter variabilities, and main temporal angles for arteries and veins. We also calculated the ratios between artery and vein values for the first four phenotypes, and we estimated the total number of vessel bifurcations. These IDPs were chosen based on their associations with diseases and the reliability of measurement. The distributions of these main IDPs are presented in Suppl. Text "Distribution of retinal vascular IDPs", and their specific sample sizes in Suppl. Table "Sample size per IDP". The broader set of 36 IDPs can be found in Suppl. Text "Methods for phenotype extraction".

Correlation structure and heritabilities of retinal vascular IDPs

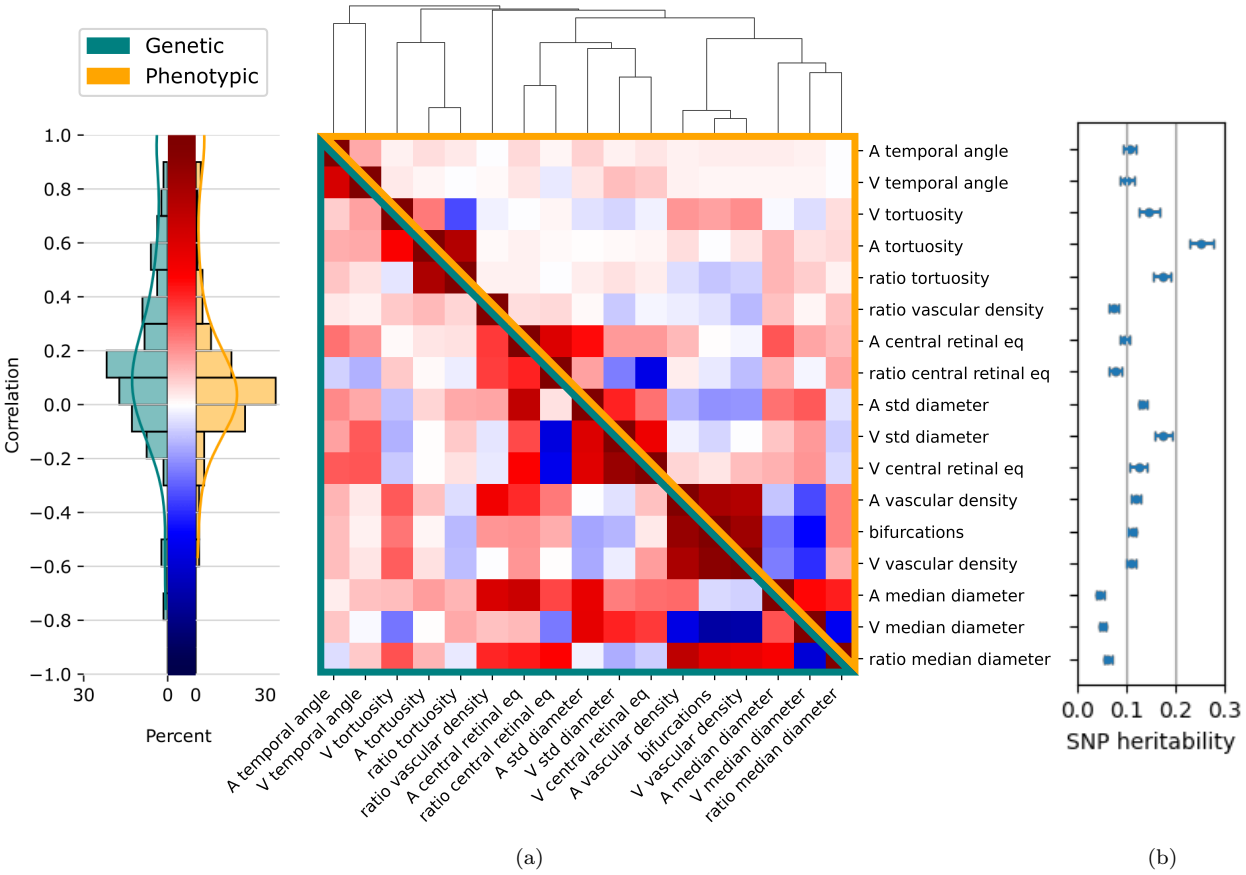
To explore the relationships between our main IDPs, we calculated pair-wise phenotypic Pearson correlations, $r_{ij}^{(p)}$ (upper right triangle in Fig. 2a). We observed that the same IDPs measured for arteries and veins tended to cluster together, in particular for the temporal angles, tortuosities, and vascular densities. Vascular densities were highly correlated with each other and the number of bifurcations, the strongest correlation observed. They also correlated with venous, but not arterial, tortuosity and anti-correlated with median diameters. The temporal angles exhibited low correlations to the other phenotypes. Finally, most artery-vein ratios were highly correlated with their corresponding arterial, and highly anti-correlated with their corresponding venous measures.

To assess the genetic relationships between the IDPs, we estimated pair-wise genetic correlations, $r_{ij}^{(g)}$, using cross-trait Linkage Disequilibrium Score Regression (LDSR) [51] (lower left triangle in Fig. 2a). On average, genetic correlations were slightly higher than phenotypic correlations (standardized mean difference across the 136 pairs (ij) is $d = 0.34$, t(135)=6.53, $p = 6.2 \times 10^{-10}$). The largest difference occurred between the temporal angles. Overall there was a strong correspondence between genetic and phenotypic correlations, $corr(r^{(g)}, r^{(p)}) = 0.86$, permutation $p < 1 \times 10^{-4}$ (see Suppl. Text "Correlation structure and heritabilities of extended list of retinal vascular IDPs" for the corresponding analysis for our broad set of 36 IDPs and

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Suppl. Data "Phenotypic and genetic correlations, discovery and replications" for the phenotypic and genetic correlation values of the main IDPs).

To gain deeper insights into the genetic control of retinal vascular morphology, we estimated the single nucleotide polymorphism (SNP) heritability, h^2, for each IDP (Fig. 2b), using LDSR [52]. Our analysis revealed varying levels of heritability across our IDPs. Arterial tortuosity exhibited the highest heritability, h^2=0.25±0.02, followed by the tortuosity ratio, h^2=0.20±0.02, and venous diameter variability, h^2=0.18±0.02. In contrast, the median diameter measures, particularly arterial median diameter, showed the lowest heritability, h^2=0.05±0.01 (see Suppl. Text "Manhattan and Quantile-Quantile plots" for Manhattan and Quantile-Quantile plots from the GWAS of each IDP).



(a) (b)

Figure 2: **a)** Phenotypic (upper-right orange triangle) and genetic (lower-left green triangle) correlations between retinal vascular phenotypes, clustered by absolute phenotypic correlation distance, $1 - |corr|$. The 17 phenotypes are (A: artery, V: vein): main temporal angles ('A/V temporal angle'), median tortuosities and their ratio ('A/V tortuosity' and 'ratio tortuosity'), central retinal equivalents and their ratio ('A/V central retinal eq' and 'ratio central retinal eq'), diameter variabilities ('A/V std diameter'), vascular densities and their ratio ('A/V vascular density' and 'ratio vascular density'), median diameters and their ratio ('A/V median diameter' and 'ratio median diameter'), and the number of bifurcations ('bifurcations'). Phenotypes were corrected for age, sex, eye geometry, batch effects, and ethnicity before phenotypic clustering and before GWAS (see Methods). **b)** Corresponding phenotype SNP heritabilities, h^2, and their standard error, estimated using LDSR [52].

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Gene and pathway level analysis of retinal vascular IDPs

To identify the genes modulating specific vascular IDPs, we employed our *PascalX* analysis tool [53, 54]. This tool aggregates SNP-wise association signals within gene windows and generates gene scores.

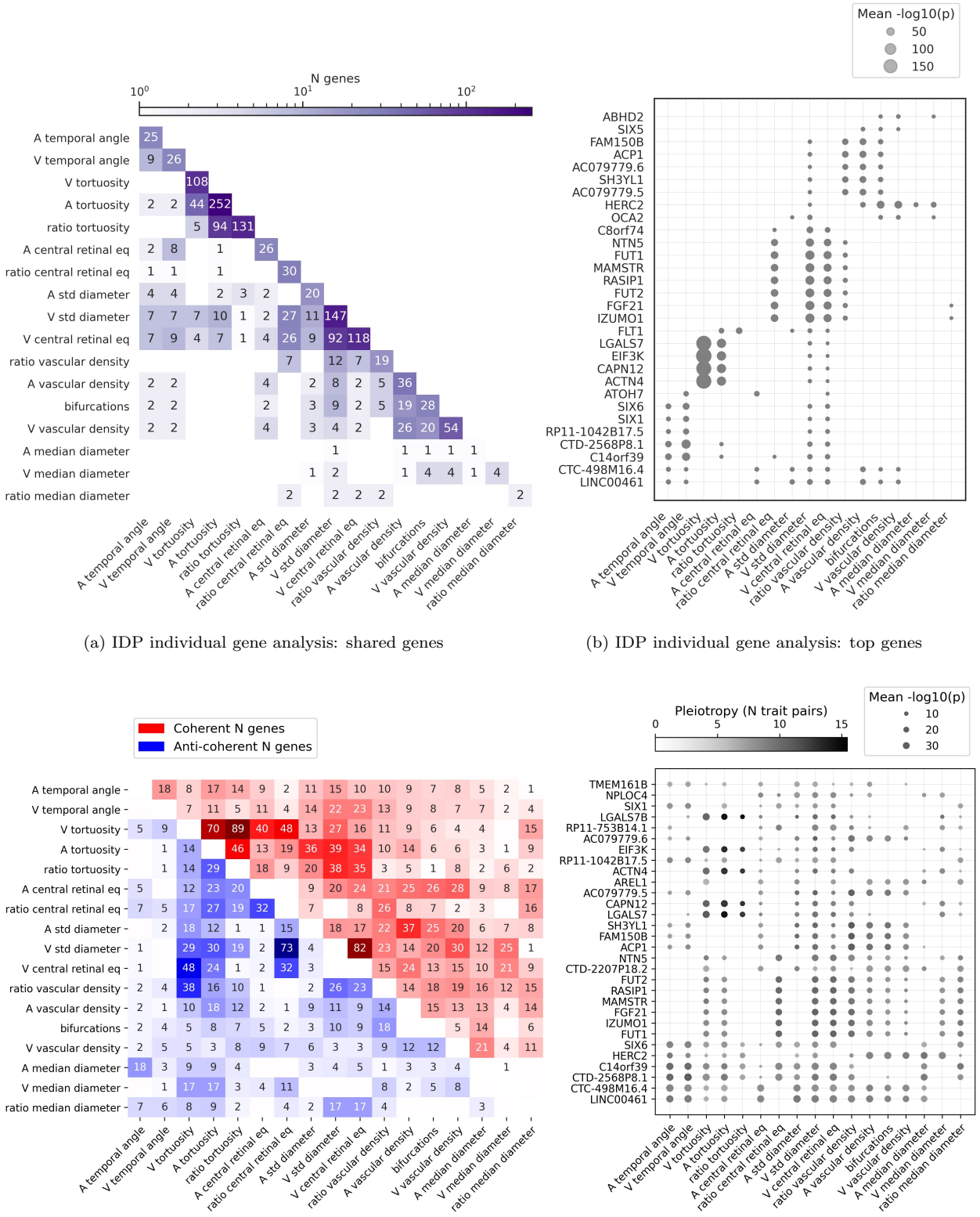
The number of genes associated with IDPs (diagonal of Fig. 3a) ranged from 1 gene for the arterial median diameter to 252 genes for arterial tortuosity. We observed the highest number of genes for tortuosities, venous diameter variability, and venous central retinal equivalent. The off-diagonal elements in Fig. 3a show the number of common genes for each pair of IDPs. Generally, more correlated IDPs tended to share more genes. However, even among highly similar IDPs, a significant portion of genes were phenotype-specific, and some IDPs shared only a few or no genes.

While no single gene was shared among all IDPs, two genes, *LINC00461* and *CTC-498M16.4*, were associated with 9 of the 17 IDPs (Fig. 3b). Other genes associated with multiple IDPs included *SIX6*, *FLT1*, *FUT1*, *HERC2*, and *PDE6G*. The complete list of significant genes associated with each IDP is available as Suppl. Data "Significant genes per IDP", and "IDP Gene scores (UKBB)".

Furthermore, we conducted pathway analysis using *PascalX* to identify gene sets, or "pathways", that exhibited higher association signals for each IDP GWAS than expected by chance. Although no single pathway was shared among all IDPs, some of the most frequent pathways included 'Fetal retina fibroblast', and 'Abnormal retinal morphology' (Suppl. Text "Pathway analyses"). The complete list of significant pathways per IDP is available as Suppl. Data "Significant pathways per IDP".

To further investigate pleiotropic genes, we employed *PascalX* cross-GWAS analysis [54], a method that examines coherent effects across the SNPs associated with two phenotypes within a gene window. This approach has more power than just intersecting the gene sets of two individual phenotypes and also allows for distinguishing between coherent and anti-coherent effects. However, even in this more sensitive analysis, no single gene was shared between all pairs of phenotypes. The most pleiotropic genes largely overlapped with those identified by simple gene-scoring, yet they tended to be shared among more IDPs (Fig. 3c and Fig. 3d). See Suppl. Data "PascalX IDP-IDP cross gene scores" for a comprehensive list of genes. Generally, IDP pairs with positive LDSR genetic correlation shared more coherent gene signals, while those with negative LDSR genetic correlation had predominantly anti-coherent signals (see Suppl. Text "LDSR genetic correlation against PascalX").

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .



(a) IDP individual gene analysis: shared genes

(b) IDP individual gene analysis: top genes

(c) IDP pairs gene analysis: shared genes

(d) IDP pairs gene analysis: top genes

Figure 3: **a)** Number of genes associated with the different IDPs, obtained using *PascalX* gene-scoring [54]. The diagonal shows the number of genes significantly associated with each IDP. The lower triangle shows the number of genes in the intersection between pairs of IDPs. **b)** 30 genes most frequently associated with the IDPs. Dot sizes are inversely proportional to p-values. **c)** Number of genes showing coherent (top right) or anti-coherent (bottom left) signal between pairs of IDPs, obtained using *PascalX* cross-scoring [54]. **d)** 30 genes most frequently found in the cross-phenotype analysis. Dot color represents pleiotropy, i.e. the number of phenotype pairs showing (anti-)coherent signal for a given gene. Dot sizes are inversely proportional to p-values.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Phenotypic association with diseases and risk factors

To evaluate the clinical relevance of retinal vascular morphology, we examined the phenotypic associations between the retinal vascular IDPs and various eye-related diseases, vascular diseases, and their associated risk factors. We employed linear regression for continuous diseases variables (that correspond to risk factors) (Fig. 4a), logistic regression for binary disease states (Fig. 4b), and Cox models for diagnoses with age-of-onset (Fig. 4c). All IDPs were standardized (z-scored) and adjusted for potential confounders.

Amongst the risk factors, diastolic blood pressure (DBP) and systolic blood pressure (SBP) displayed similar associations with most IDPs. Smoking pack-years exhibited the strongest positive association with venous diameter variability, $\beta = 0.12$, while the strongest negative correlations were observed between blood pressure (BP) and arterial-related IDPs, particularly the central retinal equivalent, vascular density, and median diameter $\beta \in$ [-0.18, -0.13] (Fig. 4a).

For binary disease states, hypertension was positively associated with venous tortuosity and less with arterial tortuosity, while negatively associated with central retinal artery equivalent, and the ratios of central retinal equivalents and vascular densities, among others. Eye diseases such as amblyopia were negatively associated with the number of bifurcations and venous vascular density, $\beta \approx$ -0.16. Others like presbyopia, hypermetropia, and myopia showed smaller effects, $\beta \in$ [-0.08, 0.06], but were still statistically significant. Retinal diabetes was negatively associated with the vascular densities and the number of bifurcations, $\beta \in$ [-0.30, -0.25], while atherosclerosis was positively associated with venous diameter variability, $\beta = 0.34$ (Fig. 4b).

The Cox model analysis for age-of-onset phenotypes included severe eye- and cardiovascular diseases, diabetes, and age at death. Both vascular densities displayed consistent associations with the age of diagnosis for all eye-related diseases. Diabetes age-at-diagnosis shared similar associations with our vascular IDPs as retinal diabetes, but was also associated with venous tortuosity and central retinal arterial equivalent, among others. Heart attack was associated with larger median arterial diameter and central retinal arterial equivalent, as well as diameter variability in both vessel types. Importantly, earlier deaths were most strongly associated with increased venous diameter variability and less strongly with increased central venous retinal equivalent, increased venous temporal angles, decreased vascular density, and fewer bifurcations. In general, venous diameter variability was associated with almost all diseases, including the only associations with pulmonary embolism and stroke (Fig. 4c). For a complete table of the standardized effect sizes, hazard ratios and p-values, see Suppl. Data "IDPs phenotypic association with diseases".

While the vascular densities and the number of bifurcations were highly inter-correlated, their associations with diseases sometimes differed. For instance, SBP and hypertension were associated with arterial vascular density and the number of bifurcations but not with venous vascular density. An equivalent analysis of 17 leading principal components (PCs) of our IDPs revealed no novel or stronger disease associations compared to the raw IDPs (see Suppl. Text "Principal Component Analysis").

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

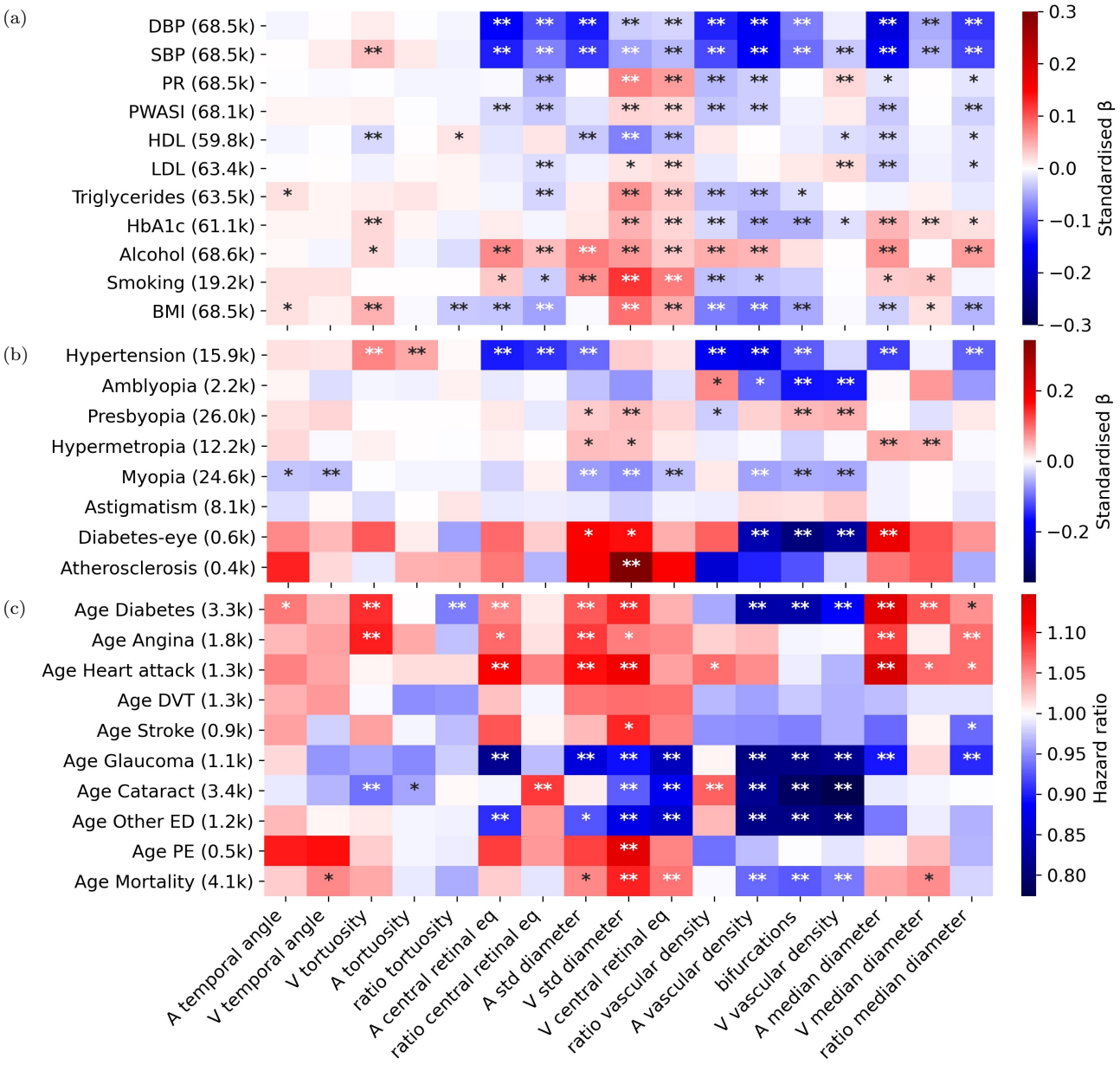


Figure 4: Phenotypic association of IDPs with risk factors and diseases. The x-axis shows IDPs and the y-axis shows risk factors and diseases. The numbers in parentheses correspond to the number of subjects with this information for which we were able to measure at least one of the 17 IDPs, for continuous diseases. For binary disease states, it represents the number of subjects who were cases and had data for at least one of the 17 IDPs. Linear (**a**) and logistic (**b**) regressions were used for continuous and binary disease states, respectively. For age-of-death and other severe diseases with the age-of-onset information, Cox proportional hazards regression was performed (**c**). In all models, phenotypes were corrected for age, sex, eye geometry, batch effects, and ethnicity. The color indicates standardized effect sizes for linear and logistic regressions or hazard ratios for Cox models. Asterisks indicate the level of statistical significance ($* : p < 0.05/N_{tests}$, $** : p < 0.001/N_{tests}$, where $N_{tests} = N_{IDPs} \times N_{traits}$, and N_{traits} is the number of diseases or risk traits considered in each panel). Labels: 'PR': Pulse rate, 'PWASI': Pulse wave arterial stiffness index, 'HDL': High-density lipoprotein, 'LDL': Low-density lipoprotein, 'HbA1c': Glycated hemoglobin, 'Alcohol': Alcohol intake frequency, 'Smoking': pack-years, 'BMI': Body mass index, 'Diabetes-eye': Diabetes related to the eye, 'DVT': Deep vein thrombosis, 'Other ED': all types of severe eye diseases not included explicitly, 'PE': Pulmonary embolism.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Genetic associations with diseases and causality analysis

To investigate the extent of common genetic architectures between the vascular IDPs and risk factors, we first used LDSR to estimate their genetic cross-correlations. Notably, we observed the strongest negative correlations (≈ -0.30) between BP measures and the ratios of the central retinal equivalents, and vascular densities. Body mass index (BMI) was correlated positively with median vein diameter (≈ 0.15), and its variation (≈ 0.18), and negatively with arterial vascular density and the number of bifurcations (Fig. 5a). HbA1c was also negatively correlated with the number of bifurcations, while HDL was negatively correlated with venous diameter variability, but only with marginal significance. For binary disease states, we found weaker associations; however, hypertension exhibited associations similar to those of blood pressure. For further details on binary disease states, see Suppl. Text "Genetics associations between vascular IDPs and binary diseases".

To compare the phenotypic and genetic associations between vascular IDPs and risk factors, we plotted the effect sizes of the phenotypic linear regressions against genetic correlations for each IDP/risk factor pair (Fig. 5b). The slopes of the best-fitting lines through the corresponding points were positive for all risk factors, except for LDL cholesterol.

A strong genetic association between two phenotypes does not imply the existence of a causal link. To systematically assess potential causal relationships between vascular IDPs and risk factors, we performed bidirectional two-sample Mendelian Randomisation (MR) analyses [55, 56]. Causal estimates were derived from the inverse variance weighted method using the TwoSampleMR R package [57, 58]. We used false discovery rate (FDR) to correct for multiple testing and we also performed sensitivity analyses that confirmed the robustness of our results (see Methods and Suppl. Data "MR risk factors (IVW)" for complete results).

Using risk factors as exposures, we observed evidence for causal effects on most of the IDPs, even after adjusting for multiple testing (see Fig. 5c). Notably, we found strong evidence for negative effects of DBP and SBP on many different vascular IDPs, including arterial (and ratio) central retinal equivalent, arterial (and ratio) median diameter, vascular density and bifurcations. Also, SBP had a positive effect on arterial and venous tortuosity. BMI had a positive effect on venous (and arterial) diameter variability and venous central retinal equivalent, while it showed a negative effect on ratio central equivalent, ratio vascular density and bifurcations. Furthermore, alcohol intake had a positive effect on arterial (but not venous) diameter variability, while HbA1c levels had a weak negative effect on arterial vascular density, bifurcations and arterial diameter variability. We also observed other potential causal effects, but only with marginal significance ($p_{uncorrected} < 0.05$). For example, PR had a positive effect on venous diameter variability and a negative effect on arterial temporal angle, while LDL levels had a positive effect on arterial tortuosity and ratio central equivalent.

In contrast, using IDPs as exposures, only the causal effect of ratio central equivalent on BP (SBP and DPB) survived FDR correction. We also observed potential causal effects on risk factors for many other vascular

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

traits, but only with marginal significance ($p_{uncorrected} < 0.05$) (see Fig. 5d). For example, we identified arterial tortuosity and venous diameter variability as positive causal factors for BP, while the ratio of central equivalent had a negative effect on BP. Venous diameter variability (and venous median diameter) also had a negative effect on HDL levels, while arterial median diameter had a positive effect on PR. Significant causal effects were also found between vascular IDPs and binary disease states; see Suppl. Text "Mendelian Randomization analysis between vascular IDPs and binary diseases".

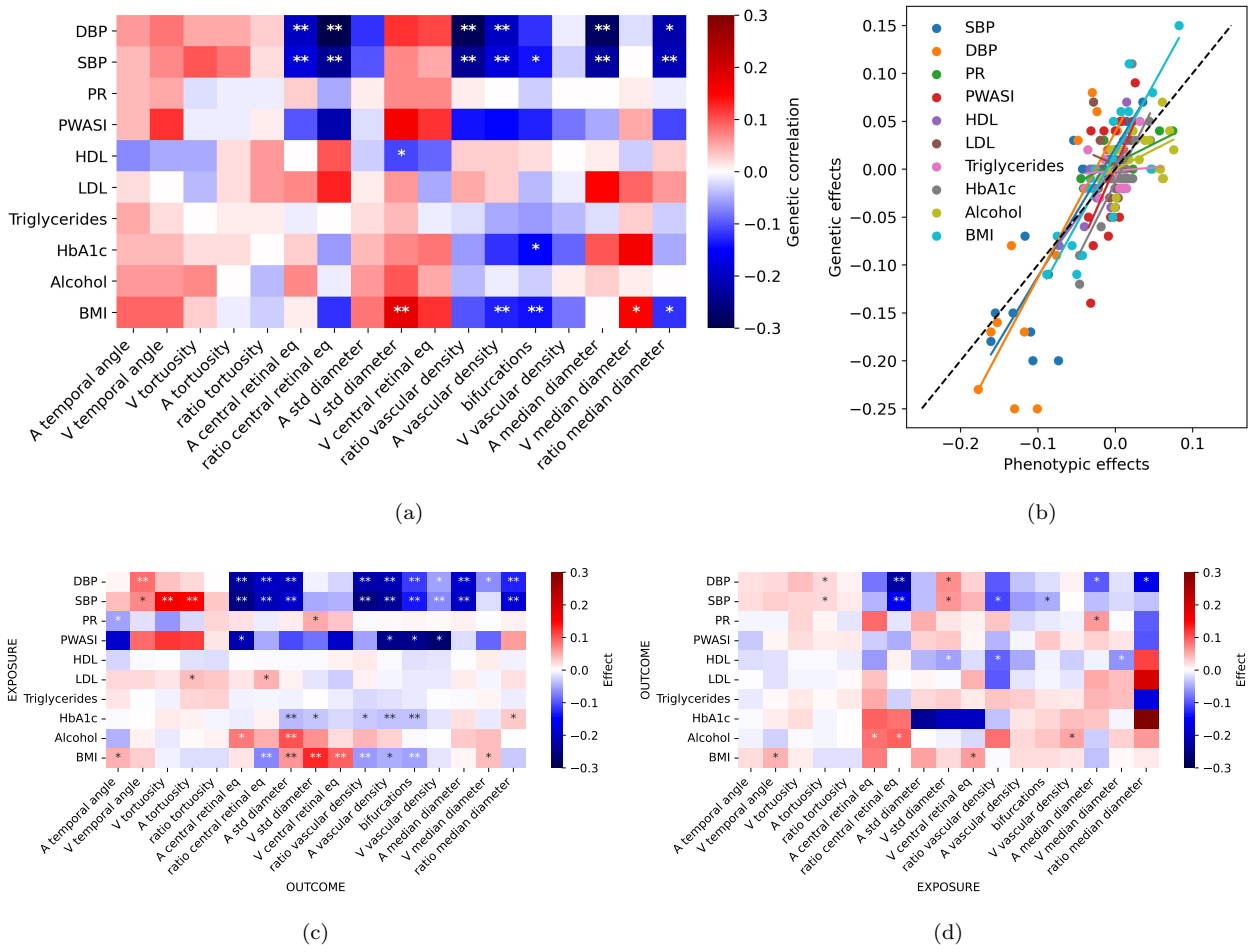


Figure 5: **a)** Genetic correlation between IDPs and risk factors, computed using LDSR [52]. The color indicates the genetic correlation coefficient and the asterisks indicate the level of statistical significance ($* : p < 0.05/N_{test}$, $** : p < 0.001/N_{test}$, being $N_{test} = N_{IDPs} \times N_{LinearDiseases}$). **b)** Correlation between phenotypic and genetic correlations of IDPs with risk factors. **c)** Causal effect estimates with risk factors as exposures and IDPs as outcomes. **d)** Causal effect estimates with IDPs as exposures and risk factors as outcomes. The color indicates the causal effect estimates based on the inverse variance-weighted MR method. The level of statistical significance is indicated with a single asterisk for nominal significance without correction for multiple testing ($* : p_{uncorrected} < 0.05$) and two asterisks for a FDR ($** : p_{FDR} < 0.05$).

In order to identify genes that were jointly associated with IDPs and risk factors, we first computed simple intersections between the corresponding gene sets of such pairs (Fig. 6a). This revealed some sizable overlaps, in particular for vein diameter variability and central equivalent with BP measures, triglycerides, HbA1c,

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

and BMI. Applying *PascalX* cross-GWAS analysis [54], as previously for the pairs of two vascular IDPs, we were able to identify many new candidates for pleiotropic genes (Fig. 6b and 6c). We observed that the sets of coherent genes (Fig. 6b) tended to be largest for phenotype pairs whose sets of associated genes already had a sizable overlap, while we found large sets of anti-coherent genes (Fig. 6c) for some phenotype pairs whose associated genes had no or only little overlap, notably for DBP and the ratio of the central retinal equivalents, the ratio of the vascular densities and the median arterial diameter and its variability. Moreover, the ratio of central retinal equivalents shared multiple anti-coherent genes with BP and BMI (see Suppl. Data "PascalX IDP-disease cross gene scores" for a comprehensive list of genes). For a similar analysis for binary disease states, we found weaker associations; however, hypertension exhibited associations similar to those of blood pressure. For further details on binary disease states, see Suppl. Text "Genetics associations between vascular IDPs and binary diseases".

Shared pathways between the IDPs and diseases can be found in the Suppl. Text "Pathway analyses between vascular IDPs and diseases".

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

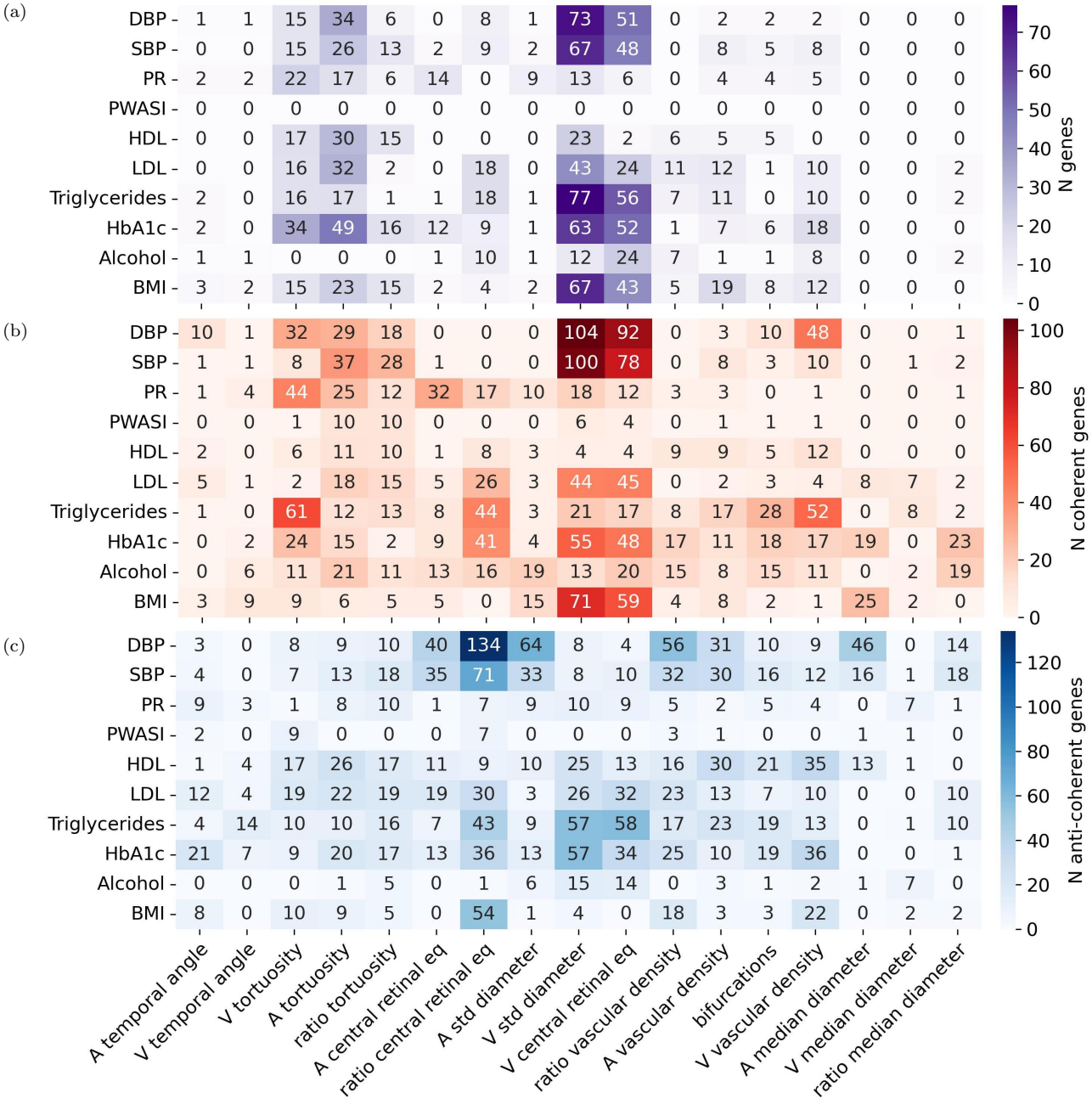


Figure 6: **a)** Gene-scoring plain intersection showing genes in common between IDPs and risk factors. Each cell shows the number of intersected genes in phenotype pairs. **b)** and **c)** Cross-phenotype coherence analysis showing the number of coherent **b)** and anti-coherent **c)** genes between phenotype pairs. Summary statistics for risk factors were obtained from `http://www.nealelab.is/uk-biobank`, and gene-level analyses were computed using *PascalX* [54].

Replication analysis

We had access to CFIs from two independent smaller cohorts, namely the Rotterdam Study (RS, N=8 142 participants) and OphtalmoLaus (N=2 276 participants), from which we computed our 17 IDPs. While we used an identical analysis pipeline for OphtalmoLaus, a novel annotation software was developed for the RS

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

images, adapted to their specifics, notably including a dedicated and internally validated vessel segmentation tool ("VascX", manuscript in preparation). Phenotypic correlations $r_{ij}^{(p)}$ between IDPs obtained in the two replication cohorts were globally concordant with those from the UKBB data (ρ=0.86, p=7.5×10^{-42} for OphtalmoLaus and ρ=0.69, p=1.8×10^{-20} for RS, see Fig. 7a). Notably, we replicated the high phenotypic correlation between vascular densities and the number of bifurcations and the low correlation between the temporal angles and the other IDPs. The main difference was that for RS data we observed a positive correlation between the variability in arterial diameter and vascular densities, while it was negative for UKBB and OphtalmoLaus (see Suppl. Text "IDPs phenotypic replication" for details), which may be due to the different vessel segmentation tool used for RS images. We also observed many consistent associations between IDPs and disease traits in the RS (when available, see Suppl. Text "IDPs phenotypic replication" for details).

Because of its larger sample size we used exclusively RS data to attempt replication of our genetic associations results in the UKBB. Since RS consists of four sub-studies with different sample sizes (RS-I: 2 391, RS-II: 877 RS-III: 2 811, and RS-IV: 2 063 participants) GWAS were performed independently for each of them, and then meta-analysed (see Suppl. Text "IDPs genetic replication" for the corresponding Manhattan and QQ-plots). Applying LDSR to estimate IDP SNP heritabilities and genetic cross-trait correlations, we observed overall consistency with the estimates from the UKBB data (ρ=0.74, p=0.001 for the former and ρ=0.45, p=5.63×10^{-12} for the latter, see Fig. 7b and c). Notably, in RS arterial tortuosity (h^2=0.22±0.07), the tortuosity ratio (h^2=0.20±0.06) and variability in venous diameter (h^2=0.18±0.06) also had the highest heritability estimates. Plotting the effect sizes for significant SNPs in the UKBB against those of the RS (Fig. 7d) revealed highly significant correlations and concordance in direction for the vast majority of IDPs. We then sought to replicate individual genetic associations. To this end, we applied the well-established Benjamini-Hochberg procedure [59]. With a fixed FDR of 0.05, we replicated 86 SNPs out of 195 across the four IDPs shown in Fig. 7e, and 232 SNPs out of 566 were replicated across the 17 main IDPs (see Suppl. Text "IDPs genetic replication"). The complete figure with all the 17 IDPs and the RS SNP heritabilities can be found in Suppl. Text "IDPs genetic replication".

Finally, to identify the genes modulating specific vascular IDPs in the replication cohort (RS), we employed our *PascalX* analysis tool [53, 54]. We applied the Benjamini-Hochberg procedure [59] for the genes. With a fixed FDR of 0.05, we replicated 93 genes out of 310 across the four IDPs shown in Fig. 7f. The complete set of RS gene scores can be found in the Suppl. Data "IDP Gene scores (RS)" and a table with genes associated with each RS IDP can be found in Suppl. Text "IDPs genetic replication". For a more detailed explanation of the SNPs and genes replication using the Bonferroni threshold as discovery and the Benjamini-Hochberg procedure as replication, see Suppl. Data "SNPs replication: UKBB replicated in RS, and RS replicated in UKBB", and "Genes replication: UKBB replicated in RS, and RS replicated in UKBB".

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

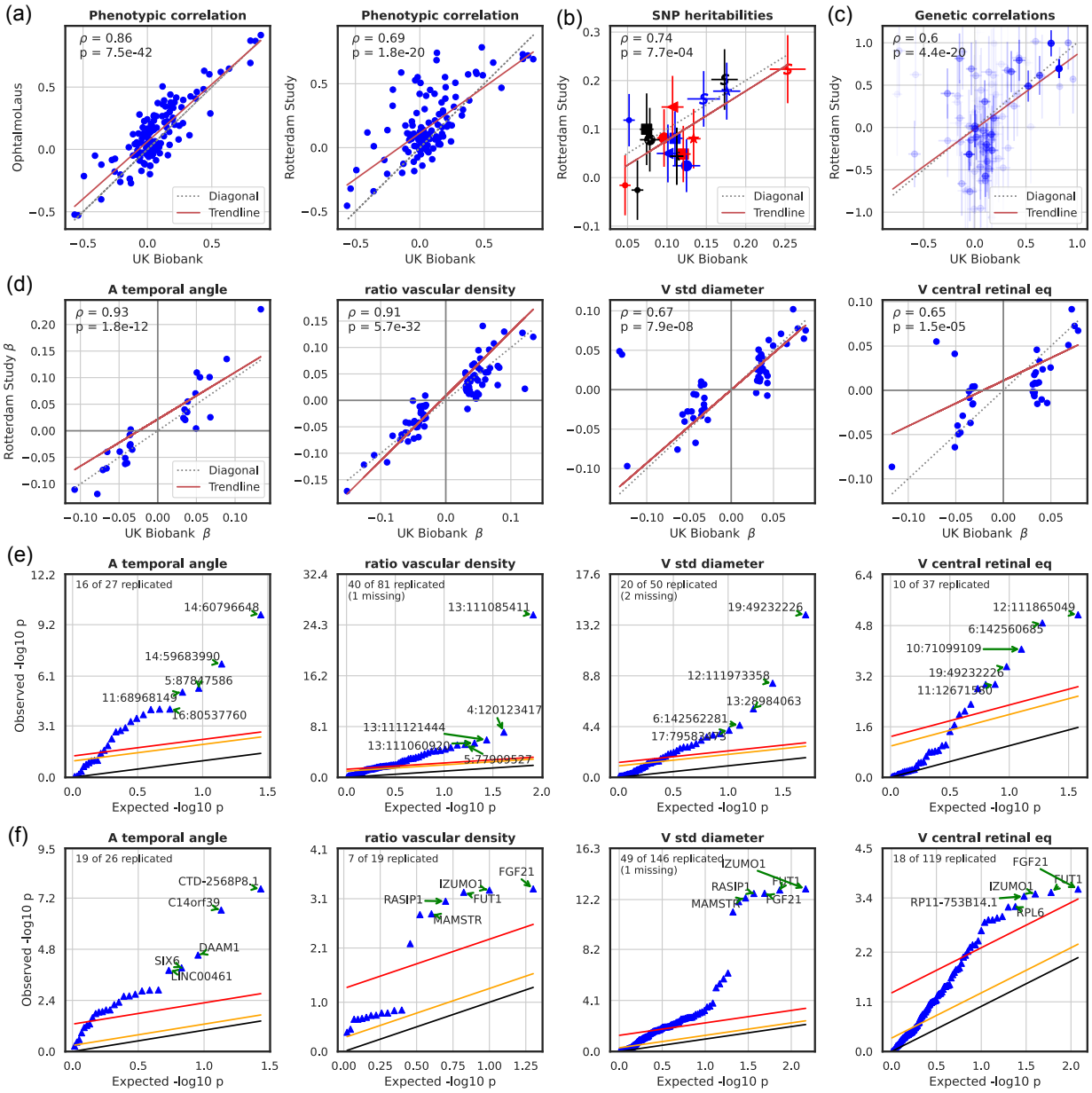


Figure 7: **a)** Scatter plot of the phenotypic correlations between our 17 IDPs in the UKBB and in the replication cohorts, OphtalmoLaus (left) and RS (right). IDPs were corrected for age, sex, eye geometry, and ethnicity (see Methods). **b)** Correlation of SNP heritabilities, using LDSR, between our 17 IDPs in the discovery (UKBB) and the replication cohort (RS). **c)** Scatter plot of the genetic correlations, using LDSR, between our 17 IDPs in the discovery (UKBB) and the replication cohort (RS). Weighted-least square regression was used to determine trendline and the significance of the association. To distinguish between the different IDPs, the following color and shape legend was utilized: 'S' denoted tortuosity, '*' for standard deviations, '◁' for temporal angles, the '≺' for bifurcations, and '□' for vascular density. While the red color is used for arteries, blue for veins, and black for no specific vessel type. **d)** Correlation of effect sizes at the SNP level in the discovery (UKBB) and the replication cohort (RS). **e)** Benjamini-Hochberg procedure on discovery lead SNPs from the UKBB using the RS. FDR = 0.05 in red, FDR = 0.5 in orange, and observed = expected line in black. The label "missing" indicates that these SNPs were not available in the replication cohort. **f)** Benjamini-Hochberg procedure on genes discovered in the UKBB using the RS. The color code is the same as in the previous subfigure. The complete figures can be found in Suppl. Text "IDPs genetic replication".

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Discussion

In this study, we established an automated analysis pipeline to extract 17 retinal vascular phenotypes from CFIs and applied it to over 130k CFIs of close to 72k UKBB subjects. While some of these phenotypes had previously been studied individually, our work is the first to provide a common reference. Our phenotyping procedure, automated and open access, enabled us to study jointly a large panel of retinal vascular phenotypes, some of which (temporal angles, central equivalents, number of bifurcations) were assessed for the first time in a large cohort. We provided a comparison of the phenotypic with genotypic correlation structures of these IDPs. We estimated their heritabilities, and elucidated associated genes and pathways, allowing us to identify common and disjoint genetic architectures. We studied associations of our IDPs with a spectrum of diseases and risk factors providing evidence of their complementarity for indicating specific disease risks. For validation, we reproduced the phenotypic correlation structure in two independent cohorts, the RS and OpthalmoLaus, with 8.1k and 2.2k participants respectively, and validated numerous GWAS results in the RS. Importantly, the RS analysis pipeline was coded independently (to adapt to the specifics of their CFIs), such that these successful replication results provide strong evidence that our phenotyping is robust and not driven by cohort specific effects.

The phenotypic correlations revealed distinct clusters of vascular IDPs. One cluster includes vascular arterial and venous density, previously linked to fractal dimension [49], and the number of bifurcations. Our results, both the discovery and replications, suggest that the number of bifurcations, which is challenging to identify, can be reliably estimated using vascular densities. Interestingly, diameter variabilities showed stronger correlations with central retinal equivalents (of the same vessel type) than with the median diameters, suggesting that larger vessels dominate diameter variability. Indeed, median diameters (especially for veins) are confounded by vascular density, which is probably due to CFIs with lower vascular density exhibiting less blood vessels of small caliber. The latter may be due to degeneration of the vasculature, but also to poorer image quality, in particular blurriness, which is impacted by corneal opacity. More work is needed to disentangle the different distributions to establish robust diameter measures for blood vessels not proximal to the OD. Finally, the temporal angles exhibited little correlation with other vascular IDPs, both in the discovery and replication cohorts, indicating that they may be influenced by non-vascular factors, such as eye anatomy.

The genetic correlation between IDPs largely mirrored their phenotypic correlation, supporting Cheverud's conjecture which states that phenotypic correlation can be used as a proxy for genetic correlation [60]. However, in our study genetic correlations were slightly larger on average, indicating potential differential effects of environmental factors on correlated IDPs.

We observed significant variation in heritability estimates among different vascular IDPs. Tortuosity and vascular density showed comparatively high heritability, consistent with some previous findings [47, 49]. A recent study [48] estimated heritability of retinal arteriolar tortuosity at 0.51 using UKBB data, substan-

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

tially higher than our 0.25 estimate. Our point estimate in the RS is 0.22±0.07, which is consistent with our estimate from UKBB data. We note that here we used median tortuosity across all vessel segments independent of their caliber and length, even though tortuosity may vary as a function of blood vessel length and diameter, calling for more refined or stratified measures.

Notably, venous diameter variability also exhibited high heritability, which may be partially attributed to vascular beading, severe forms of which are known to be inherited [61]. Interestingly, the heritability of temporal angles was relatively low, likely reflecting sensitivity to environmental factors, as well as measurement noise and methodological variations, such as variability in the position of the OD across images, and variation due to refraction that may be insufficiently corrected by spherical and cylindrical power as our covariates. Future work could use heritability as a guide for developing more robust measures of these angles both in terms of the methodology for extracting them (see Suppl. Text "Methods for phenotype extraction" for our procedure) and implementing proper corrections for potential confounders. Finally, median vessel diameters obtained the smallest heritabilities, consistent with their known dependence on the environment, such as vasoconstriction induced by stress, and diseases, as well as the aforementioned confounding by image quality and age-dependent reduction in detectable small blood vessels.

Gene analysis of individual IDPs replicated some previously identified gene associations with explicit vascular phenotypes, namely tortuosity [46–48], fractal dimension, vascular density [49], and vessel width [48], as well as DL vascular phenotypes representing the latent space of an autoencoder [39] (see Suppl. Text "Comparison with previous GWAS" for complete list). Diameter variability and central retinal equivalent, particularly for veins, received numerous gene associations, providing evidence for the genetic complexity of these phenotypes. Finally, the ratios of vascular measures, combining arterial and venous components, showed associations with genes not found in individual measures, indicating sensitivity to vessel-type specific effects. Generally, phenotypes with higher heritability tended to have more associated genes, with some exceptions like arterial diameter variability.

No single gene was significantly associated with all the vascular IDPs. The most frequently observed gene was *LINC00461*, also known as visual cortex-expressed gene (VISC), an evolutionarily conserved long non-coding RNA that produces several alternatively spliced transcripts [62]. Other frequently observed genes are related to eye diseases (*PDE6G* has been linked to Retinitis Pigmentosa, and *SIX6* to glaucoma, myopia, and retinal degeneration) or vascular processes (*FLT1* linked to hypertension and heart disease). Also, *HERC2* and *OCA2*, two neighboring genes related to pigmentation, were linked to multiple IDPs. *OCA2* is involved in the production of melanosomes, and therefore directly contributes to the formation of pigments, while mutations in *HERC2* have been found to modulate *OCA2* expression [63]. We speculate that less eye pigmentation leads to reduced protection from damaging light, which could cause global changes on the retinal surface, including vascular morphology.

Our study of phenotypic associations between vascular IDPs and disease-relevant phenotypes confirmed some previously known associations, such as the link between vessel diameter and hypertension, likely due

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

to alterations in vascular resistance and blood flow [3, 4, 64]. Interestingly our MR analysis provides support for a low ratio of the central retinal equivalent being causal for high BP. This is consistent with a positive (but not significant) phenotypic and genetic correlation between central retinal venular equivalent and hypertension/BP. We also observed significant associations between IDPs capturing arterial vessel caliber and heart attack, consistent with previous findings [18, 43–45, 65, 66], and the previously observed link between tortuosity and cardiovascular risk factors [10, 67]. For a more complete discussion refer to Suppl. Text "Replication of previously identified associations of retinal vascular phenotypes with diseases".

Additionally, we found diameter variability in both vessel types to be positively associated with heart attack, possibly due to the formation of plaques that enlarge the vessel while reducing lumen diameter and negatively impacting blood flow. While atherosclerosis has a similar association profile across the vascular phenotypes, interestingly the association with venous diameter variability was the strongest and the only one passing statistical significance. This signal may be driven by artery-vein crossings and calls for further investigation.

Type-2 diabetes is known to affect the microvascular circulation in the retina resulting in a range of structural changes unique to this tissue, such as neovascularization [68], which can affect vascular density [49] and arteriolar tortuosity [69], among others. We note that the association profile of our IDPs with HbA1c, which is used to diagnose type-2 diabetes [70], is similar, but typically a bit weaker in terms of effect size and significance than the latter, suggesting that our Cox model taking into account the age of onset of the disease endpoint has more power. Given that diabetes is a known risk factor for cardiovascular disease, including angina, this could explain why many associations between type-2 diabetes and IDPs related to vessel caliber are also observed for angina and heart attack.

Age at death is the only trait, besides myopia, that is associated with the venous temporal angle. Moreover, age at death also shares some of the associations of specific diseases with our IDPs, such as a negative association with the number of bifurcations (in line with the results for hypertension and type-2 diabetes reported above), and positive associations with several vein-related phenotypes besides the temporal angle, including the diameter variability, the median diameter (in line with the results for stroke, type-2 diabetes, heart attack and atherosclerosis reported above). It seems plausible that the reduced lifespan for these common diseases explains the observed associations.

MR analysis allowed assessment of potential causal directions for the links observed in the correlation analysis. Overall the effects of CVD risk factors on our vascular IDPs tend to be stronger and more significant than the reverse, underlining the usefulness of IDPs for early diagnosis of CVD. Consistent with previous findings, we found that individuals with genetically elevated BP tend to have lower retinal vascular density [49]. In addition, we found several other IDPs being affected, suggesting that vascular remodeling can be caused by elevated BP. Also, elevated BMI may cause higher variability in venous and arterial diameter, consistent with the finding that obesity may decrease venous return of blood from the lower extremities thereby increasing the risk of chronic venous insufficiency [71]. Amongst the strongest potential reverse causal effects is the aforementioned decrease in BP due to a higher ratio of the central retinal equivalent. Additionally, we

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

found that higher arterial tortuosity tends to increase BP, confirming previous findings [48]. Interestingly, we also found that vein diameter variability may also cause higher BP, which could be mediated by anatomical alterations such as venous beading. However, those findings were not significant after correcting for multiple testing. The recent paper by Jiang et al. [48] also reported a causal effect of arterial tortuosity on Coronary Heart Disease (CHD). We were able to confirm this causal link of arterial tortuosity on CHD, while no other retinal trait was found to be causal (see Suppl. Text "Mendelian Randomization analysis between vascular IDPs and binary diseases").

In the gene analysis, we observed that the pairs of diseases and IDPs with the highest positive correlations tended to have a greater number of shared genes, both in the intersection and coherence analyses (see Suppl. Text "LDSR genetic correlation against PascalX"). It is worth noting that this was not limited to pairs with high-significance values. Besides, we again observed that the patterns differed in the analysis of coherent and anti-coherent genes. For example, vein diameter variability and BMI had multiple genes associated coherently, implying that the genes modulating them acted in the same direction. Conversely, between BMI and the central equivalent ratio, the majority of genetic effects acted in the opposite direction. This is consistent with the previous finding that vein diameter variability and ratio central equivalent shared many anti-coherent genes.

While our study pushes the boundaries of analysing retinal vascular phenotypes, it has several important limitations: First, limiting our study to data from the UKBB, RS, and OphtalmoLaus makes our findings specific to a population of mostly European ancestry. Second, there are some other potentially relevant vascular IDPs that we did not analyze, including branching angles, artery-vein crossings, and neovascularization. Third, our GWAS did not include the analysis of sex chromosomes or rare variants. The exclusion of these factors may have limited the scope of the study and prevented the identification of potential associations between genetic variations and the development of certain diseases or phenotypes. Finally, summary statistics used for the genetic analysis of binary disease states were not obtained using GWAS with logistic regression, but linear regression, which can have an effect a sensitivity effect on their results (see methods: Genetic association with diseases).

In summary, this study establishes a common framework for studying multiple vascular phenotypes of the retina. The explicit characterisation of retinal vasculatures will be useful both for clinical research further exploring their usefulness as biomarkers for systemic diseases, and fundamental research, where we provide an important alternative reference to implicit characterizations of the retina, such as the recent "Foundation model" for retinal images [40]. Our analysis of genes and pathways unveiled a strikingly limited intersection, indicating that the mechanisms governing these phenotypes are largely independent. Our findings regarding the association between disease phenotypes affirmed some established knowledge while uncovering numerous novel connections. Specifically, we observed a plethora of intriguing new links between diameter variability, particularly in veins, and various disease phenotypes such as age of mortality, pulmonary embolism, and myocardial infarction. While more work is needed to further validate and extend our findings, our analyses

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

provide evidence that we start to have sufficient power for obtaining functional, as well as some initial causal insights into the genetic and disease-related processes modulating the retinal vasculature.

Methods

UK Biobank data

The UKBB is a population-based cohort of approximately 488k subjects with rich, longitudinal phenotypic data including a complete medical history, and a median 10-year follow-up [72, 73]. Standard retinal 45° CFIs were captured using a Topcon Triton 3D OCT 1000. 173 814 images from 84 813 individuals, were analyzed. Genotyping was performed on Axiom arrays for a total of 805 426 markers, from which approximately 96 million genotypes were imputed. We used the subset of 15 599 830 SNPs that had been assigned a rsID. Baseline and disease information about the subjects whose images were analyzed can be found in Suppl. Text "Baseline characteristics".

Image segmentation and quality control

Raw CFIs were segmented into pixel-wise segmentations of arteries and veins using the DL model LWNET [36], which had been trained on the publicly available DRIVE dataset. To detect the OD, a random set of 100 CFIs of varying quality from the UKBB were first annotated, and the resulting ground truths were then used to retrain a standard U-Net previously trained to detect the OD in various public datasets [37]. Last, branch points of vessels and vessel-segment-wise and centerlines were extracted using skeletonization, and diameters were extracted using the distance transform, both provided in ARIA [50].

A published QC method [49] to assess image quality in the UKBB was used, and the 75% highest-quality CFIs according to this method were retained for further analysis. In this method, the image quality of 1 000 CFIs was quantified by professional graders, and a CNN was then trained to imitate the graders' quality assessment. A significant negative correlation between image quality and age was observed, r=-0.21, but was not corrected. For more information refer to Suppl. Text "Image segmentation". Additionally, to see how the threshold on the QC can affect the results refer to Suppl. Text "Quality control threshold effect on results".

Phenotyping

Retinal vessel morphology was broadly phenotyped, drawing on a set of known relevant ophthalmological phenotypes, including a few that were previously undescribed, such as diameter variability. Due to the lack of consensus definitions and methods for their measurements, their implementation can vary. Therefore, we implemented different definitions and methods for most phenotypes, see Suppl. Text "Methods for phenotype extraction" and Suppl. Text "Correlation structure and heritability of extended list of retinal vascular IDPs". In this study, we focused on 17 representative phenotypes with significant heritability and relevant disease

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

associations. These phenotypes were selected for independence, removing highly redundant phenotypes with similar definitions, but keeping some highly correlated phenotypes when they resulted from significantly different definitions. The mean between left and right eye measurements was taken whenever measurements from both eyes passed quality control, otherwise, the single eye measurement was used. Measurements from only the first measured time point were used whenever measurements from multiple time points were present. Cohen's d, $\frac{\overline{r^{(g)}}-\overline{r^{(p)}}}{\sqrt{\frac{SD^{(g)2}+SD^{(p)2}}{2}}}$, was used to quantify the mean difference between genetic and phenotypic correlations.

We used two validation procedures: (1) visual inspection of random images in the UKBB to identify and refine potential confounders affecting certain phenotypes, and (2) the use of the DRIVE dataset as a proxy to assess performance in other datasets with similar characteristics, using ground truth on it for the number of bifurcations and temporal angles (Suppl. Text "Validation of IDPs").

Correction for covariates

IDPs were corrected for sex, age, age-squared, sex-by-age, sex-by-age-squared, spherical power, spherical power-squared, cylindrical power, cylindrical power-squared, instance, assessment center, genotype measurement batch, and genomic PCs 1-20. Their associations with each phenotype have been visualized in Suppl. Text "Covariate effects". For GWAS, raw phenotypes were transformed with the rank-based inverse normal transformation (rb-INT) before correction.

Disease association

The list of diseases analyzed includes vascular and eye-related diseases, risk factors, mortality, and other conditions previously found to be associated with the retina vascular system. The disease data were collected from the UKBB, and the official datafield identifier corresponding to each disease can be found in the Suppl. Text "Diseases list UKBB", and in Suppl. Data "Diseases information".

Different regression models were employed based on the nature of the disease traits. For risk factors, ordinary least squares (OLS) linear regression was used to estimate standardized effects using the 'statsmodels.formula.api' library in Python 3.8.13. For binary and categorical disease phenotypes, logistic regression was applied, using the logit function from the 'statsmodels.formula.api' library.

Prior to conducting the regression analyses, a pre-processing step was performed to address potential confounding effects. Covariates were regressed out of retinal IDPs and the obtained residuals were then used as regressors in the linear/logistic regression analyses. For Cox models, the covariates were added again to the models to adjust for potential non-linear effects.

The regression models were fitted using the adjusted independent variables, the estimates of regression coefficients (betas), and their corresponding standard deviation (std), or odds ratios were obtained. To determine the significance of regression coefficients, p-values were computed and compared to predefined

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

alpha thresholds (0.05 and 0.001), divided by the total number of tests conducted (i.e., the number of independent variables multiplied by the number of diseases analyzed).

In cases where time-to-event data was available and accurate diagnosis was feasible, multivariate Cox proportional hazards regression was utilized. This approach allowed us to estimate hazard ratios for diseases with time-to-event outcomes and reliable diagnoses. 'coxph' from the R packages 'survival' and 'survminer' were used with default parameters.

Genome-wide analyses

GWAS for UKBB data was performed using BGENIE [73]. SNP-wise heritabilities and genetic correlations between IDPs were derived using LDSR [52]. Gene and pathway scores were computed using *PascalX* [54, 74]. Both protein-coding genes and lincRNAs were scored using the novel, approximate "saddle" method, taking into account all SNPs within a 50kb window around each gene. All pathways available in MSigDB v7.2 were scored using *PascalX*' ranking mode, fusing and rescoring any co-occurring genes less than 100kb apart. *PascalX* requires LD structure to accurately compute gene scores, which in our analyses was provided with the UK10K (hg19) reference panel. Gene-level cross-GWAS coherence test between IDP pairs and between IDPs and diseases or risk factors was computed using the *PascalX* cross-scoring *zsum* method, testing for both coherence and anti-coherence of GWAS signals. Variants with a minor allele frequency of at least 0.001 were considered. Correction for bias due to sample overlap was done using the intercept from pairwise LDSR genetic correlation. The significance threshold was set at 0.05 divided by the number of tested genes. Top hits GWAS summary statistics can be found in Suppl. Data "GWAS top hits (UKBB) before pruning".

Genetic association with diseases

For the genetic correlation between IDPs and diseases, the summary statistics of the diseases in LDSR format were obtained from the Neale lab (nealelab-ldsc-sumstat-files). LDSR was computed for the diseases and the IDPs. We limited ourselves to diseases that were categorized as 'High confidence' by the Neale lab (column 'Neale_ldsr' in Suppl. Data "Diseases information").

Regarding the genes shared between IDPs and diseases, we used the GWAS summary statistics for diseases (nealelab-sumstat-files). However, in this case, we applied some additional filters deleting rsid values that did not start with 'rs', p-values that were missing, and low confidence variants (Suppl. Data "Significant genes per diseases").

Binary disease states were included in the genetic analysis, however, their results require additional caution, since the GWAS summary statistics from the Neale lab used linear regression for all the disease phenotypes, which is not ideal for non-continuous response variables. It should be noted that the covariates used for the GWAS analysis of our IDPs and those used for the risk factors/diseases are almost, but not exactly, the same (age, sex, age-squared, age-by-sex, age-by-sex-squared, and first 20 PCs for risk factors/diseases).

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Mendelian Randomisation

To perform the bidirectional two-sample MR analyses we used the TwoSamplesMR package in R [57]. We first considered independent SNPs significantly ($p < 5 \times 10^{-8}$) associated with the exposure as genetic instruments, pruning SNPs with $r^2 > 0.001$ to a lead SNP according to LD estimates from the UK10K reference panel [75]. In cases where the number of instruments was below 10, we relaxed the selection threshold for instruments to $p < 10^{-6}$. In general, the number of instruments tended to be lower when using IDPs, rather than disease traits as exposures, which may be due, at least partially, to lower sample sizes. For this reason, when using IDPs as exposures we relaxed the selection threshold for instruments to $p < 10^{-6}$.

To assess instrument strength, we computed the F-statistic from the regression of the exposure on the instruments [76] defined as $F = \frac{R^2 \times (N-k-1)}{(1-R^2) \times k}$, where R^2 is the explained variance of the exposure by the instruments, N is the sample size of the GWAS for the exposure, and k is the number of instruments. For all the forward and reverse MR pairs, the F-statistics were > 50 suggesting that the selected SNPs were suitable instruments.

Causal estimates were based on the inverse variance-weighted method (IVW) [58]. In particular, we used a fixed-effect model when having three or less instruments, and otherwise a random-effects model. To complement and enhance the reliability of the results, we applied additional methods, namely MR-Egger, weighted median, and weighted mode [77, 78] MR. For all exposure-outcome pairs, the estimated causal effects were consistent in the direction across the four methods whenever significant. Differences in significance levels are likely because the power of these additional methods is smaller than that of the IVW method [79] (Suppl. Data "MR risk factors (all methods)").

Notably, MR-Egger intercepts of most of the associations were not significantly different from zero, suggesting that no significant pleiotropy was detected (Suppl. Data "MR risk factors (Egger intercept)"). Lastly, leave-one-out analyses showed that the estimates were not biased by any single SNP (Suppl. Data "MR risk factors (LOO)"). Overall, the sensitivity analyses confirmed the reliability of most of our putative causal effects in both directions.

Since the number of UKBB subjects for which we extracted IDPs was much smaller than the number of UKBB subjects used to study disease phenotypes, our analysis can still be considered a two-sample MR setup, and potential bias due to sample overlap is expected to be small and in direction of the null [80, 81].

Replication

The RS is a prospective population-based cohort study of people living in Ommoord, a district of the city of Rotterdam [82]. The RS consists of four cohorts, all of which were used in this replication. Each cohort was followed for multiple rounds of follow-up examinations every 4 to 5 years. Most of the patient visits in the RS involved the capture of CFIs on both eyes. Due to the multi-decade span of the RS, multiple devices, capture conditions and fields (macula and disc centered) are present in the dataset. In the RS,

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

DNA extraction was performed using whole blood samples following standardized and previously described protocols [82]. Genotyping was performed using both the Infinium II HumanHap550(-Duo) (RS-I & RS-II) and 610-Quad Genotyping BeadChip (RS-I & RS-III; Illumina, San Diego, CA, USA). Imputation of markers was performed using the Haplotype Reference Consortium version 1.1 as the reference panel [83]. The CoLaus study, initiated in 2003 in Lausanne, Switzerland, involves over 6 700 volunteers aged 35 to 75. OphtalmoLaus, a segment of CoLaus, delves into ocular health. OphtalmoLaus CFIs were acquired with Topcon 2000 or Topcon triton. The number of initial images was 6 503, corresponding to 2 276 participants.

Phenotyping and cross-correlations

In this study we made use of RS imaging from recent rounds (RS-I-4, RS-II-4, RS-III-1, RS-IV-1) due to the generally higher quality and number of these images, even though in some cases this meant less participants. We quality-controlled the images by automatically filtering out images where the OD was near or out of the bounds of the CFI. Both disc- and macula-centered images were included. After QC, the number of participants with usable images per RS cohort was 2 710 (RS-I), 1 169 (RS-II), 3 350 (RS-III), and 2 658 (RS-IV). Most visits in the RS captured multiple CFIs per participant. The average number of usable CFIs per subject varied between 2.96 (RS-I) and 7.75 (RS-IV), including both eyes. A participant's features were computed as the mean over all their images' features.

To accommodate the use of disc-centered images and the high variability in imaging conditions present in the RS, we made use of segmentation and feature extraction methods trained and tested on RS data. The 17 phenotypes computed for the main study were implemented in a fundus analysis software used in-house following the original implementation.

To compute the RS phenotypic cross-correlations in Fig. 7, IDPs were corrected for age, sex, eye geometry, imaging device (if multiple were used within the same cohort, dummy coded), and genomic PCs 1-10. Eye geometry was included as a combination of the spherical and cylindrical powers into one variable as spherical equivalent (spherical power + cylindrical power/2).

For OphtalmoLaus, in the case of multiple images for the same participant, we kept the one with the highest QC score. The number of participants after removing the images that failed the segmentation or IDPs computation, and after QC, was N=1 715 participants. If both eyes survived this screening, we averaged out the phenotypes of the two eyes, while if only one eye survived then we considered the phenotypes of that eye. After that, we cross-referenced with the sample file and the final number of subjects was 1 435. We corrected for the following covariates: age, sex, age-by-sex, age-squared, cylindrical and spherical powers, spherical-squared, cylindrical-squared, and genomic PCs 1-10.

GWAS

As imputation and QC of the four RS-cohorts were done separately, we also performed the GWAS analyses separately and then meta-analyzed the results. For the initial GWAS analyses, we performed linear

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

regressions using Plink 2.0 [84]. The included covariates were the same as for the phenotypic correlation, i.e. age, sex, eye geometry, imaging device, and genomic PCs 1-10. We performed an inverse variance weighted fixed-effect meta-analysis, using METAL software [85]. P-values for the association results were calculated by using the z-statistic. The meta-analyzed significant hits were pruned using Plink 2.0. Variants were considered independent if they were at least 500 kb apart and had an $R^2 < 0.1$. Top hits GWAS summary statistics in the RS can be found in Suppl. Data "GWAS top hits (RS) before pruning".

Data and code availability

The genetic and phenotypic UKBB data are available upon application to the UKBB, https://www.ukbiobank.ac.uk. Similarly for the replication data, RS and OphtalmoLaus, data can be made available upon request to researchers through a data transfer agreement. GWAS summary statistics will be available on the GWAS Catalog following publication. Code will be made available in our public GitHub repository following publication: https://github.com/BergmannLab/retina-phenotypes.

Author contributions

SOV, MJB, and SB designed this study. MJB performed QC and image preprocessing. SOV extracted the number of bifurcations, temporal angles, central retinal equivalents, and diameter variability. MJB extracted OD positions, vascular density, and fractal dimension. MT extracted median diameter and tortuosity. MJB performed postprocessing (covariate correction and normalization). GWAS was performed by MJB and SOV for UKBB data. MJB estimated heritabilities and cross-phenotype correlations. SOV and MJB performed gene and pathway analysis. Gene-level cross-phenotype analysis was done by OT. Linear and logistic regressions were done by SOV and OT, and Cox model analyses by MJB. Cross-phenotype LDSR and gene analysis between retina phenotypes and diseases were done by SOV and OT. FH helped SOV with the validation of the angles and bifurcation phenotypes. II performed MR analysis. JDVQ and VAV performed RS replication under the supervision of WR, BL, and CCWK. IM and AE performed OphtalmoLaus replication under the supervision of MT and RS. SB supervised all analyses. SOV, MJB, OT, II, MT, JDVQ, IM, and SB wrote the manuscript. All other authors contributed to the writing.

Acknowledgements

The authors thank Leah Böttger and Sacha Bors for their valuable comments on the manuscript. The authors thank the entire EyeNED team for their various contributions during the development of the algorithms used in the RS replication. The authors are grateful to the study participants and the staff from the UKBB, RS, and OphtalmoLaus.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Funding

Supported by the Swiss National Science Foundation grant no. CRSII5_209510 for the "VascX" Sinergia project. Supported by Oogfonds, Stichting voor Ooglijders, Stichting voor Blindenhulp, Henkes stichting, and Landelijke Stichting voor Blinden en Slechtzienden (LSBS). Additional support was given by Erasmus Medical Center, Erasmus University, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

Conflict of Interest statement: None declared.

References

[1] M Kamran Ikram, Yi Ting Ong, Carol Y Cheung, and Tien Y Wong. Retinal vascular caliber measurements: clinical significance, current knowledge and future perspectives. *Ophthalmologica*, 229(3): 125–136, 2013.

[2] Sara B Seidelmann, Brian Claggett, Paco E Bravo, Ankur Gupta, Hoshang Farhad, Barbara E Klein, Ronald Klein, Marcelo Di Carli, and Scott D Solomon. Retinal vessel calibers in predicting long-term cardiovascular outcomes: the atherosclerosis risk in communities study. *Circulation*, 134(18):1328–1338, 2016.

[3] Ryo Kawasaki, Ning Cheung, Jie Jin Wang, Ronald Klein, Barbara EK Klein, Mary Frances Cotch, A Richey Sharrett, Steven Shea, FM Amirul Islam, and Tien Y Wong. Retinal vessel diameters and risk of hypertension: the multiethnic study of atherosclerosis. *Journal of hypertension*, 27(12):2386, 2009.

[4] M Kamran Ikram, Frank Jan De Jong, Ewoud J Van Dijk, Niels D Prins, Albert Hofman, Monique MB Breteler, and Paulus TVM De Jong. Retinal vessel diameters and cerebral small vessel disease: the rotterdam scan study. *Brain*, 129(1):182–188, 2006.

[5] Raviv Allon, Michael Aronov, Michael Belkin, Elad Maor, Michael Shechter, and Ido Didi Fabian. Retinal microvascular signs as screening and prognostic factors for cardiac disease: a systematic review of current evidence. *The American Journal of Medicine*, 134(1):36–47, 2021.

[6] Gerald Liew, Paul Mitchell, Elena Rochtchina, Tien Yin Wong, Wynne Hsu, Mong Li Lee, Alan Wainwright, and Jie Jin Wang. Fractal analysis of retinal microvasculature and coronary heart disease mortality. *European heart journal*, 32(4):422–429, 2011.

[7] Maximilian WM Wintergerst, Peyman Falahat, Frank G Holz, Christian Schaefer, Nadjib Schahab, and Robert Finger. Retinal vasculature assessed by octa in peripheral arterial disease. *Investigative Ophthalmology & Visual Science*, 61(7):3203–3203, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[8] Wayne Smith, Jie Jin Wang, Tien Yin Wong, Elena Rochtchina, Ronald Klein, Stephen R Leeder, and Paul Mitchell. Retinal arteriolar narrowing is associated with 5-year incident severe hypertension: the blue mountains eye study. *Hypertension*, 44(4):442–447, 2004.

[9] Tien Wong and Paul Mitchell. The eye in hypertension. *The Lancet*, 369(9559):425–435, 2007.

[10] Carol Yim-lui Cheung, Yingfeng Zheng, Wynne Hsu, Mong Li Lee, Qiangfeng Peter Lau, Paul Mitchell, Jie Jin Wang, Ronald Klein, and Tien Yin Wong. Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. *Ophthalmology*, 118(5):812–818, 2011.

[11] Andrea Grosso, Franco Veglio, Massimo Porta, FM Grignolo, and TY Wong. Hypertensive retinopathy revisited: some answers, more questions. *British Journal of Ophthalmology*, 89(12):1646–1654, 2005.

[12] Tien Yin Wong, Anoop Shankar, Ronald Klein, Barbara EK Klein, and Larry D Hubbard. Prospective cohort study of retinal vessel diameters and risk of hypertension. *bmj*, 329(7457):79, 2004.

[13] Tien Yin Wong, Ronald Klein, Barbara EK Klein, Stacy M Meuer, and Larry D Hubbard. Retinal vessel diameters and their associations with age and blood pressure. *Investigative ophthalmology & visual science*, 44(11):4644–4650, 2003.

[14] SB Dimmitt, SM Eames, P Gosling, JNW West, JM Gibson, and WA Littler. Usefulness of ophthalmoscopy in mild to moderate hypertension. *The Lancet*, 333(8647):1103–1106, 1989.

[15] Harry Leung, Jie Jin Wang, Elena Rochtchina, Tien Y Wong, Ronald Klein, and Paul Mitchell. Impact of current and past blood pressure on retinal arteriolar diameter in an older population. *Journal of hypertension*, 22(8):1543–1549, 2004.

[16] Jie Jin Wang, Paul Mitchell, Harry Leung, Elena Rochtchina, Tien Yin Wong, and Ronald Klein. Hypertensive retinal vessel wall signs in a general older population: the blue mountains eye study. *Hypertension*, 42(4):534–541, 2003.

[17] Tien Yin Wong, Ronald Klein, A Richey Sharrett, Bruce B Duncan, David J Couper, Barbara EK Klein, Larry D Hubbard, and F Javier Nieto. Retinal arteriolar diameter and risk for hypertension. *Annals of internal medicine*, 140(4):248–255, 2004.

[18] Tien Yin Wong, Ronald Klein, A Richey Sharrett, Bruce B Duncan, David J Couper, James M Tielsch, Barbara EK Klein, and Larry D Hubbard. Retinal arteriolar narrowing and risk of coronary heart disease in men and women: the atherosclerosis risk in communities study. *Jama*, 287(9):1153–1159, 2002.

[19] A Richey Sharrett, Larry D Hubbard, Lawton S Cooper, Paul D Sorlie, Rosemary J Brothers, F Javier Nieto, Joan L Pinsky, and Ronald Klein. Retinal arteriolar diameters and elevated blood pressure: the atherosclerosis risk in communities study. *American journal of epidemiology*, 150(3):263–270, 1999.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[20] SCY Woo, GYH Lip, and PL Lip. Associations of retinal artery occlusion and retinal vein occlusion to mortality, stroke, and myocardial infarction: a systematic review. *Eye*, 30(8):1031–1038, 2016.

[21] Tyler Hyungtaek Rim, John Seungsoo Han, Jaewon Oh, Dong Wook Kim, Seok-Min Kang, and Eun Jee Chung. Retinal vein occlusion and the risk of acute myocardial infarction development: a 12-year nationwide cohort study. *Scientific reports*, 6(1):1–9, 2016.

[22] Danielle L Weiler, Carla B Engelke, Anna LO Moore, and Wendy W Harrison. Arteriole tortuosity associated with diabetic retinopathy and cholesterol. *Optometry and Vision Science*, 92(3):384–391, 2015.

[23] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[24] Muthu Rama Krishnan Mookiah, U Rajendra Acharya, Hamido Fujita, Jen Hong Tan, Chua Kuang Chua, Sulatha V Bhandary, Augustinus Laude, and Louis Tong. Application of different imaging modalities for diagnosis of diabetic macular edema: a review. *Computers in biology and medicine*, 66: 295–315, 2015.

[25] Jie J Wang, Bronwen Taylor, Tien Y Wong, Brian Chua, Elena Rochtchina, Ronald Klein, and Paul Mitchell. Retinal vessel diameters and obesity: a population-based study in older persons. *Obesity*, 14 (2):206–214, 2006.

[26] Dora Elisa Alvarado-Carrillo, Iván Cruz-Aceves, Martha Alicia Hernández-González, and Luis Miguel López-Montero. Robust detection and modeling of the major temporal arcade in retinal fundus images. *Mathematics*, 10(8):1334, 2022.

[27] Faraz Oloumi, Rangaraj M Rangayyan, and Anna L Ells. A graphical user interface for measurement of temporal arcade angles in fundus images of the retina. In *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4. IEEE, 2012.

[28] Álvaro S Hervella, José Rouco, Jorge Novo, Manuel G Penedo, and Marcos Ortega. Deep multi-instance heatmap regression for the detection of retinal vessel crossings and bifurcations in eye fundus images. *Computer Methods and Programs in Biomedicine*, 186:105201, 2020.

[29] Harry Pratt, Bryan M Williams, Jae Yee Ku, Charles Vas, Emma McCann, Baidaa Al-Bander, Yitian Zhao, Frans Coenen, and Yalin Zheng. Automatic detection and distinction of retinal vessel bifurcations and crossings in colour fundus photography. *Journal of Imaging*, 4(1):4, 2018.

[30] Xiayu Xu, Joseph M Reinhardt, Qiao Hu, Benjamin Bakall, Paul S Tlucek, Geir Bertelsen, and Michael D Abramoff. Retinal vessel width measurement at branchings using an improved electric field theory-based graph approach. *PloS one*, 7(11):e49668, 2012.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[31] Alauddin Bhuiyan, Baikunth Nath, and Kotagiri Ramamohanarao. Detection and classification of bifurcation and branch points on retinal vascular network. In *2012 International Conference on Digital Image Computing Techniques and Applications (DICTA)*, pages 1–8. IEEE, 2012.

[32] M Elena Martinez-Perez, Alun D Hughes, Alice V Stanton, Simon A Thom, Neil Chapman, Anil A Bharath, and Kim H Parker. Geometrical and morphological analysis of vascular branches from fundus retinal images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000: Third International Conference, Pittsburgh, PA, USA, October 11-14, 2000. Proceedings 3*, pages 756–765. Springer, 2000.

[33] Adria Perez-Rovira, T MacGillivray, Emanuele Trucco, KS Chin, K Zutis, C Lupascu, Domenico Tegolo, Andrea Giachetti, Peter J Wilson, A Doney, et al. Vampire: vessel assessment and measurement platform for images of the retina. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3391–3394. IEEE, 2011.

[34] Behdad Dashtbozorg, Ana Maria Mendonca, Susana Penas, and Aurelio Campilho. Retinacad, a system for the assessment of retinal vascular changes. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6328–6331. IEEE, 2014.

[35] Jonathan Fhima, Jan Van Eijgen, Ingeborg Stalmans, Yevgeniy Men, Moti Freiman, and Joachim A Behar. Pvbm: A python vasculature biomarker toolbox based on retinal blood vessel segmentation. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*, pages 296–312. Springer, 2023.

[36] Adrian Galdran, André Anjos, José Dolz, Hadi Chakor, Hervé Lombaert, and Ismail Ben Ayed. State-of-the-art retinal vessel segmentation with minimalistic models. *Scientific Reports*, 12(1):1–13, 2022.

[37] Artem Sevastopolsky. Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. *Pattern Recognition and Image Analysis*, 27(3):618–624, 2017.

[38] Harry Pratt, Bryan M Williams, Jae Yee Ku, Charles Vas, Emma McCann, Baidaa Al-Bander, Yitian Zhao, Frans Coenen, and Yalin Zheng. Automatic detection and distinction of retinal vessel bifurcations and crossings in colour fundus photography. *Journal of Imaging*, 4(1):4, 2017.

[39] Ziqian Xie, Tao Zhang, Sangbae Kim, Jiaxiong Lu, Wanheng Zhang, Cheng-Hui Lin, Man-Ru Wu, Alexander Davis, Roomasa Channa, Luca Giancarlo, et al. igwas: image-based genome-wide association of self-supervised deep phenotyping of human medical images. *medRxiv*, pages 2022–05, 2022.

[40] Yukun Zhou, Mark A Chia, Siegfried K Wagner, Murat S Ayhan, Dominic J Williamson, Robbert R Struyven, Timing Liu, Moucheng Xu, Mateo G Lozano, Peter Woodward-Court, et al. A foundation model for generalizable disease detection from retinal images. *Nature*, pages 1–8, 2023.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[41] Ryan Poplin, Avinash V Varadarajan, Katy Blumer, Yun Liu, Michael V McConnell, Greg S Corrado, Lily Peng, and Dale R Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3):158–164, 2018.

[42] Tao Li, Wang Bo, Chunyu Hu, Hong Kang, Hanruo Liu, Kai Wang, and Huazhu Fu. Applications of deep learning in fundus images: A review. *Medical Image Analysis*, 69:101971, 2021.

[43] Xueling Sim, Richard A Jensen, M Kamran Ikram, Mary Frances Cotch, Xiaohui Li, Stuart MacGregor, Jing Xie, Albert Vernon Smith, Eric Boerwinkle, Paul Mitchell, et al. Genetic loci for retinal arteriolar microcirculation. *PloS one*, 8(6):e65804, 2013.

[44] Richard A Jensen, Xueling Sim, Albert Vernon Smith, Xiaohui Li, Jóhanna Jakobsdóttir, Ching-Yu Cheng, Jennifer A Brody, Mary Frances Cotch, Barbara Mcknight, Ronald Klein, et al. Novel genetic loci associated with retinal microvascular diameter. *Circulation: Cardiovascular Genetics*, 9(1):45–54, 2016.

[45] M Kamran Ikram, Sim Xueling, Richard A Jensen, Mary Frances Cotch, Alex W Hewitt, M Arfan Ikram, Jie Jin Wang, Ronald Klein, Barbara EK Klein, Monique MB Breteler, et al. Four novel loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. *PLoS genetics*, 6(10):e1001184, 2010.

[46] Abirami Veluchamy, Lucia Ballerini, Veronique Vitart, Katharina E Schraut, Mirna Kirin, Harry Campbell, Peter K Joshi, Devanjali Relan, Sarah Harris, Ellie Brown, et al. Novel genetic locus influencing retinal venular tortuosity is also associated with risk of coronary artery disease. *Arteriosclerosis, thrombosis, and vascular biology*, 39(12):2542–2552, 2019.

[47] Mattia Tomasoni, Michael Johannes Beyeler, Sofia Ortin Vela, Ninon Mounier, Eleonora Porcu, Tanguy Corre, Daniel Krefl, Alexander Luke Button, Hana Abouzeid, Konstantinidis Lazaros, et al. Genome-wide association studies of retinal vessel tortuosity identify numerous novel loci revealing genes and pathways associated with ocular and cardiometabolic diseases. *Ophthalmology Science*, page 100288, 2023.

[48] Xiaofan Jiang, Pirro G Hysi, Anthony P Khawaja, Omar A Mahroo, Zihe Xu, Christopher J Hammond, Paul J Foster, Roshan A Welikala, Sarah A Barman, Peter H Whincup, et al. Gwas on retinal vasculometry phenotypes. *Plos Genetics*, 19(2):e1010583, 2023.

[49] Seyedeh Maryam Zekavat, Vineet K Raghu, Mark Trinder, Yixuan Ye, Satoshi Koyama, Michael C Honigberg, Zhi Yu, Akhil Pampana, Sarah Urbut, Sara Haidermota, Declan P O'Regan, Hongyu Zhao, Patrick T Ellinor, Ayellet V Segrè, Tobias Elze, Janey L Wiggs, James Martone, Ron A Adelman, Nazlee Zebardast, Lucian Del Priore, Jay C Wang, and Pradeep Natarajan. Deep learning of the retina enables phenome- and genome-wide analyses of the microvasculature. *Circulation*, 11 2021. ISSN 1524-4539. doi: 10.1161/CIRCULATIONAHA.121.057709. URL http://www.ncbi.nlm.nih.gov/pubmed/34743558.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[50] Peter Bankhead, C Norman Scholfield, J Graham McGeown, and Tim M Curtis. Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PloS one*, 7(3):e32435, 2012.

[51] Brendan Bulik-Sullivan, Hilary K Finucane, Verneri Anttila, Alexander Gusev, Felix R Day, Po-Ru Loh, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Laramie Duncan, et al. An atlas of genetic correlations across human diseases and traits. *Nature genetics*, 47(11):1236–1241, 2015.

[52] Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, and Benjamin M Neale. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, 47(3):291–295, 2015.

[53] David Lamparter, Daniel Marbach, Rico Rueedi, Zoltán Kutalik, and Sven Bergmann. Fast and rigorous computation of gene and pathway scores from snp-based summary statistics. 2016. doi: 10.1371/journal.pcbi.1004714. URL http://www2.unil.ch/cbg/index.php?title=PascalTestDatahttp://www2.unil.ch/cbg/index.php?title=Pascal.

[54] Daniel Krefl and Sven Bergmann. Cross-gwas coherence test at the gene and pathway level. *PLoS Computational Biology*, 18(9):e1010517, 2022.

[55] Stephen Burgess, Dylan S Small, and Simon G Thompson. A review of instrumental variable estimators for mendelian randomization. *Statistical methods in medical research*, 26(5):2333–2355, 2017.

[56] George Davey Smith and Shah Ebrahim. 'mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology*, 32(1):1–22, 2003.

[57] Danielle Rasooly and Chirag J Patel. Conducting a reproducible mendelian randomization analysis using the r analytic statistical environment. *Current protocols in human genetics*, 101(1):e82, 2019.

[58] Stephen Burgess, Adam Butterworth, and Simon G Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology*, 37(7):658–665, 2013.

[59] Yoav Benjamini and Yosef Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of educational and Behavioral Statistics*, 25(1):60–83, 2000.

[60] Sebastian M Sodini, Kathryn E Kemper, Naomi R Wray, and Maciej Trzaskowski. Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. *Genetics*, 209(3):941–948, 2018.

[61] Travis A Meredith. Inherited retinal venous beading. *Archives of Ophthalmology*, 105(7):949–953, 1987.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[62] Peter L Oliver, Rebecca A Chodroff, Amrit Gosal, Benjamin Edwards, Amanda FP Cheung, Julio Gomez-Rodriguez, Gene Elliot, Lisa J Garrett, Tom Lickiss, Francis Szele, et al. Disruption of visc-2, a brain-expressed conserved long noncoding rna, does not elicit an overt anatomical or behavioral phenotype. *Cerebral Cortex*, 25(10):3572–3585, 2015.

[63] Manfred Kayser, Fan Liu, A. Cecile J.W. Janssens, Fernando Rivadeneira, Oscar Lao, Kate van Duijn, Mark Vermeulen, Pascal Arp, Mila M. Jhamai, Wilfred F.J. van IJcken, Johan T. den Dunnen, Simon Heath, Diana Zelenika, Dominiek D.G. Despriet, Caroline C.W. Klaver, Johannes R. Vingerling, Paulus T.V.M. de Jong, Albert Hofman, Yurii S. Aulchenko, Andre G. Uitterlinden, Ben A. Oostra, and Cornelia M. van Duijn. Three genome-wide association studies and a linkage analysis identify herc2 as a human iris color gene. *American Journal of Human Genetics*, 82:411, 2 2008. ISSN 00029297. doi: 10.1016/J.AJHG.2007.10.003. URL /pmc/articles/PMC2427174//pmc/articles/PMC2427174/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2427174/.

[64] M Kamran Ikram, Frank Jan de Jong, Johannes R Vingerling, Jacqueline CM Witteman, Albert Hofman, Monique MB Breteler, and Paulus TVM de Jong. Are retinal arteriolar or venular diameters associated with markers for cardiovascular disorders? the rotterdam study. *Investigative ophthalmology & visual science*, 45(7):2129–2134, 2004.

[65] Thomas J MacGillivray, James R Cameron, Qiuli Zhang, Ahmed El-Medany, Carl Mulholland, Ziyan Sheng, Bal Dhillon, Fergus N Doubal, Paul J Foster, Emmanuel Trucco, et al. Suitability of uk biobank retinal images for automatic analysis of morphometric properties of the vasculature. *PLoS One*, 10(5): e0127914, 2015.

[66] Josef Flammer, Katarzyna Konieczka, Rosa M Bruno, Agostino Virdis, Andreas J Flammer, and Stefano Taddei. The eye and the heart. *European heart journal*, 34(17):1270–1278, 2013.

[67] WT Longstreth Jr, Emily K Marino Larsen, Ronald Klein, Tien Yin Wong, A Richey Sharrett, David Lefkowitz, and Teri A Manolio. Associations between findings on cranial magnetic resonance imaging and retinal photography in the elderly: the cardiovascular health study. *American journal of epidemiology*, 165(1):78–84, 2007.

[68] Eui Seok Shin, Christine M Sorenson, and Nader Sheibani. Diabetes and retinal vascular dysfunction. *Journal of ophthalmic & vision research*, 9(3):362, 2014.

[69] MB Sasongko, TY Wong, TT Nguyen, CY Cheung, JE Shaw, and JJ Wang. Retinal vascular tortuosity in persons with diabetes and diabetic retinopathy. *Diabetologia*, 54:2409–2416, 2011.

[70] Shariq I Sherwani, Haseeb A Khan, Aishah Ekhzaimy, Afshan Masood, and Meena K Sakharkar. Significance of hba1c test in diagnosis and prognosis of diabetic patients. *Biomarker insights*, 11:BMI–S38440, 2016.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[71] Torsten Willenberg, Anette Schumacher, Beatrice Amann-Vesti, Vincenzo Jacomella, Christoph Thalhammer, Nicolas Diehm, Iris Baumgartner, and Marc Husmann. Impact of obesity on venous hemodynamics of the lower limbs. *Journal of vascular surgery*, 52(3):664–668, 2010.

[72] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 2015.

[73] Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, et al. The uk biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209, 2018.

[74] Daniel Krefl, Alessandro Brandulas Cammarata, and Sven Bergmann. PascalX: a Python library for GWAS gene and pathway enrichment tests. *Bioinformatics*, 39(5):btad296, 05 2023. ISSN 1367-4811. doi: 10.1093/bioinformatics/btad296. URL https://doi.org/10.1093/bioinformatics/btad296.

[75] Jie Huang, Bryan Howie, Shane McCarthy, Yasin Memari, Klaudia Walter, Josine L Min, Petr Danecek, Giovanni Malerba, Elisabetta Trabetti, Hou-Feng Zheng, et al. Improved imputation of low-frequency and rare variants using the uk10k haplotype reference panel. *Nature communications*, 6(1):8111, 2015.

[76] Stephen Burgess, Simon G Thompson, and Crp Chd Genetics Collaboration. Avoiding bias from weak instruments in mendelian randomization studies. *International journal of epidemiology*, 40(3):755–764, 2011.

[77] Jack Bowden, George Davey Smith, and Stephen Burgess. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, 44(2):512–525, 06 2015. ISSN 0300-5771. doi: 10.1093/ije/dyv080. URL https://doi.org/10.1093/ije/dyv080.

[78] Jack Bowden, George Davey Smith, Philip C. Haycock, and Stephen Burgess. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology*, 40(4):304–314, 2016. doi: https://doi.org/10.1002/gepi.21965. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/gepi.21965.

[79] Fernando Hartwig, George Smith, and Jack Bowden. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *International journal of epidemiology*, 46, 07 2017. doi: 10.1093/ije/dyx102.

[80] Stephen Burgess, Neil M Davies, and Simon G Thompson. Bias due to participant overlap in two-sample mendelian randomization. *Genetic epidemiology*, 40(7):597–608, 2016.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292368; this version posted April 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

[81] Bernd Taschler, Stephen M Smith, and Thomas E Nichols. Causal inference on neuroimaging data with mendelian randomisation. *NeuroImage*, 258:119385, 2022.

[82] M Arfan Ikram, Guy Brusselle, Mohsen Ghanbari, André Goedegebure, M Kamran Ikram, Maryam Kavousi, Brenda C T Kieboom, Caroline C W Klaver, Robert J de Knegt, Annemarie I Luik, Tamar E C Nijsten, Robin P Peeters, Frank J A van Rooij, Bruno H Stricker, André G Uitterlinden, Meike W Vernooij, and Trudy Voortman. Objectives, design and main findings until 2020 from the rotterdam study. *Eur. J. Epidemiol.*, 35(5):483–517, May 2020.

[83] Shane McCarthy, Sayantan Das, Warren Kretzschmar, Olivier Delaneau, Andrew R Wood, Alexander Teumer, Hyun Min Kang, Christian Fuchsberger, Petr Danecek, Kevin Sharp, Yang Luo, Carlo Sidore, Alan Kwong, Nicholas Timpson, Seppo Koskinen, Scott Vrieze, Laura J Scott, He Zhang, Anubha Mahajan, Jan Veldink, Ulrike Peters, Carlos Pato, Cornelia M van Duijn, Christopher E Gillies, Ilaria Gandin, Massimo Mezzavilla, Arthur Gilly, Massimiliano Cocca, Michela Traglia, Andrea Angius, Jeffrey C Barrett, Dorrett Boomsma, Kari Branham, Gerome Breen, Chad M Brummett, Fabio Busonero, Harry Campbell, Andrew Chan, Sai Chen, Emily Chew, Francis S Collins, Laura J Corbin, George Davey Smith, George Dedoussis, Marcus Dorr, Aliki-Eleni Farmaki, Luigi Ferrucci, Lukas Forer, Ross M Fraser, Stacey Gabriel, Shawn Levy, Leif Groop, Tabitha Harrison, Andrew Hattersley, Oddgeir L Holmen, Kristian Hveem, Matthias Kretzler, James C Lee, Matt McGue, Thomas Meitinger, David Melzer, Josine L Min, Karen L Mohlke, John B Vincent, Matthias Nauck, Deborah Nickerson, Aarno Palotie, Michele Pato, Nicola Pirastu, Melvin McInnis, J Brent Richards, Cinzia Sala, Veikko Salomaa, David Schlessinger, Sebastian Schoenherr, P Eline Slagboom, Kerrin Small, Timothy Spector, Dwight Stambolian, Marcus Tuke, Jaakko Tuomilehto, Leonard H Van den Berg, Wouter Van Rheenen, Uwe Volker, Cisca Wijmenga, Daniela Toniolo, Eleftheria Zeggini, Paolo Gasparini, Matthew G Sampson, James F Wilson, Timothy Frayling, Paul I W de Bakker, Morris A Swertz, Steven McCarroll, Charles Kooperberg, Annelot Dekker, David Altshuler, Cristen Willer, William Iacono, Samuli Ripatti, Nicole Soranzo, Klaudia Walter, Anand Swaroop, Francesco Cucca, Carl A Anderson, Richard M Myers, Michael Boehnke, Mark I McCarthy, Richard Durbin, and Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, 48(10):1279–1283, October 2016.

[84] Christopher C Chang, Carson C Chow, Laurent Cam Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1):7, February 2015.

[85] Cristen J Willer, Yun Li, and Gonçalo R Abecasis. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17):2190–2191, September 2010.