Genome-Wide Association Studies of retinal vessel tortuosity identify 173 novel loci, capturing genes and pathways associated with disease and vascular tissue pathomechanics

Mattia Tomasoni[1,2], Michael Johannes Beyeler[1,2], Ninon Mounier[2,3], Eleonora Porcu[2,3,4], Sofia Ortin Vela[1,2], Alexander Luke Button[1,2], Tanguy Corre[1,2,3], Hana Abouzeid[5,6], Murielle Bochud[3], Daniel Krefl[1,2], Sven Bergmann[1,2,7]

1 Dept. of Computational Biology, University of Lausanne, Lausanne, Switzerland
2 Swiss Institute of Bioinformatics, Lausanne, Switzerland
3 Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland
4 Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland
5 Division of Ophthalmology, Geneva University Hospitals, Switzerland
6 Clinical Eye Research Center Memorial Adolphe de Rothschild, Geneva, Switzerland
7 Dept. of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa

Corresponding authors:
mattia.tomasoni@unil.ch
sven.bergmann@unil.ch

Abstract

Fundus images of the eye allow for non-invasive inspection of the microvasculature system of the retina, which is informative of systemic cardiovascular health. We set up a fully automated image processing pipeline enabling the massively parallelised annotation of such images in terms of vessel type (i.e., artery or vein) and quantitative morphological properties, such as tortuosity ("bendiness"). We tested seven different tortuosity measures by applying this pipeline to 175,821 fundus images from 62,751 genotyped subjects of the UK Biobank and the SKIPOGH cohort. We present the largest Genome-Wide Association Study (GWAS) of vessel tortuosity conducted to date, capturing 175 loci (173 of which are novel). Vessel-type specific GWA Studies revealed 114 and 63 loci, respectively. SNP-based heritability was highest when using the Distance Factor (curve over chord length) as tortuosity measure (h^2=0.25 (SE=0.025) when considering all vessels types, h^2=0.23 (SE=0.020) for arteries and h^2=0.15 (SE=0.021) for veins). We showed that a number of alleles associated with retinal vessel tortuosity point to a common genetic architecture of this trait with CVD and metabolic syndrome. Using *PascalX*, a new high-precision method to aggregate SNP-wise effects on genes and annotated gene-sets, we obtained new insights into key components and the molecular pathways modulating the structural properties of the retinal vasculature. Our results shed new light on the genetics of vascular pathomechanisms and systemic disease, and highlight how GWAS can be used to inform phenotype extraction from high-dimensional data, such as images.

Table of contents

Introduction 2

Results 2

Automated processing of retinal images defines vessel tortuosity phenotypes 2

GWAS as a tool for method development and phenotype selection 3

Retinal vessel tortuosity GWA Studies identify 173 novel loci 3

Trend of effect sizes replicates in the SKIPOGH cohort 3

Tortuosity loci serve pleiotropic functions as disease variants 4

Genetic overlap with metabolic syndrome and other CVD risk factors 4

Mendelian Randomisation reveals causal effect of LDL on tortuosity 4

Gene and pathway scoring reveals key players affecting tortuosity 5

Tortuosity genes are overexpressed in arteries and heart 5

Link with disease is supported by correlation with patient records 6

Discussion 6

Materials and Methods 9

Definition of tortuosity 9

UK Biobank phenotypes 9

Data extraction 10

Classification of vessels into arteries and veins 10

UK Biobank genotype data 11

The SKIPOGH study 11

Genome-wide association analysis 11

Shared genetic architecture with disease 11

Mendelian Randomisation analysis 12

Gene-based tests 12

Pathway and GO enrichment 12

Tissue-wide gene expression analysis 12

Figures 13

Tables 17

Author contributions 20

Acknowledgements 20

Funding 20

Additional information 20

References 21

Introduction

The fundus of the eye is covered by blood vessels which are essential for bringing oxygen and nutrients to the various tissues of the retina. Fundus photography allows for easy and non-invasive inspection of this retinal microvasculature, and it is well known that there are disease-related changes in the morphometric properties of retinal vessels. The fundus is of interest beyond the field of ophthalmology, since pathological changes in the retinal vessels often reflect those in the microvasculature of other organs. Indeed, based on the homology between the microvasculature in the retina and that found in other organs, retinal analysis has the potential of becoming a powerful screening tool for diseases elsewhere in the body, notably the brain[1,2,3,4,5], kidney[6,7] and ear[8]. The retinal microvasculature can therefore provide signs of systemic disease, including increased risk of diabetes[9,10,11,12], obesity[13] and cardiovascular disease (CVD)[14,15,16,17,18], specifically stroke[16,19,19,20,21], coronary heart disease[22], coronary artery disease[23], hypertension[11,24,25,20,26,27,28,29,30,31,32,33,34], atherosclerosis[19,20,35] and myocardial infarction[36,37].

In recent years, measuring retinal features in large genotyped cohorts has paved the way for studying the genetic underpinning of these phenotypes, with Genome Wide Association Studies (GWA Studies) having already identified a number of loci[38,39] associated with renal vessel size[40,41,42], optic disc morphology[43,44] and vessel tortuosity[23].

In the following we report the results from our GWA Studies on retinal vessel tortuosity. Our analysis is based on the largest collection of retinal images performed to date. We used images[45] and genotypes from 62,751 subjects of the UK Biobank[46] and from 436 subjects of the much smaller, yet independent, population-based cohort SKIPOGH[47,48]. Our fully automated image processing pipeline combined different existing tools to annotate the retinal vessels in each image and assign the vessel type as either artery or vein. Tortuosity was quantified using seven different measures, computed for all vessel segments in a given image. We then used the median value of the most heritable tortuosity measure (i.e. the Distance Factor) across arterial, venous and combined segments as our three principle traits to be studied by GWAS, which led to the confirmation of two known trait variants and the discovery of 173 new ones. Many of the variants we identified had previously been associated with other traits, specifically CVD, metabolic syndrome and some of their risk factors. Furthermore, our downstream analysis aggregating the SNP-wise association signals onto genes and gene-sets provides new insights into the molecular pathways modulating the structural properties of the retinal vasculature. Importantly, many of the underlying mechanisms and their response to disease risk factors are likely to be relevant for the vascular pathomechanisms in other organs.

Results

Automated processing of retinal images defines vessel tortuosity phenotypes

For our discovery study, we processed 175,821 images from 62,751 individuals available in the UK Biobank. We first applied the ARIA[49] software for automated vessel annotation of these images. To this end we modified this software to operate in batch mode. Using classical image processing tools ARIA delivers a set of identified blood vessel segments for each image, with each represented by a list of dense points along the vessel midline. Using these annotated vessel points, we measured the tortuosity for each vessel segment in terms of the so-called Distance Factor, i.e. the ratio between the path length $s(C)$ along the segment and the distance $chord(C)$ between its start and end point, as was first suggested in Smedby, O. et al.[50]. We also tested six alternative measures of tortuosity that integrate the local absolute value of curvature, or its square, along the segment curve, and optionally use devison by $s(C)$ or $chord(C)$ for normalisation (see Supplementary Material, Definitions of alternative tortuosity measures). We then used the median retinal vessel tortuosity over all annotated vessels (averaged over multiple images, if available) as our trait (see Methods).

Independently, we also applied the recently published retina analysis tool *Little W-Net,* which uses a deep neural network approach[51] exclusively to assign vessel type predictions to each image pixel. Using a simple majority vote for all points attributed to a vessel segment by ARIA, we classified vessel segments as being part of an artery or a vein. This allowed us to define tortuosity specific to arteries and veins (see Figure 1 for an overview of our retinal phenotype extraction and imaging GWAS pipeline).

GWAS as a tool for method development and phenotype selection

We carried out LD Score Regression[52] to determine the SNP-based heritability of our seven vessel tortuosity phenotypes. Heritability differed greatly across measures, providing us with a principled method to pick the one best reflecting genetic, and thereby, most likely to capture molecular contributions modulating vessel tortuosity. The Distance Factor obtained the highest heritability: $h^2_{SNP} = 0.25$ $(SE = 0.025)$ when considering all vessels, $h^2_{SNP} = 0.23$ $(SE = 0.020)$ for arteries, and $h^2_{SNP} = 0.15$ $(SE = 0.021)$ for veins. We did not observe any significant genomic inflation (see Table 6 for details). Traits defined by the six alternative tortuosity measures were less heritable ($0.11 < h^2_{SNP} < 0.13$, see Supplementary Material).

Retinal vessel tortuosity GWA Studies identify 173 novel loci

Applying linear regression of confounder-corrected Q-Q normalized median retinal vessel tortuosity on the genotypes of the matching subjects imputed to a panel of about 15M genetic variants, we identified 6,725 significantly associated SNPs when considering all vessels (both arteries and veins) in a combined-vessel-type GWAS (see Supplementary File 2A and Methods for details). Vessel-type-specific GWAS resulted in 6,332 and 2,842 significantly associated SNPs for arteries and veins, respectively (see Supplementary File 2B and 2C). We applied LD pruning with a threshold of r^2<0.1 within a window of 500k bases to define independence. This procedure resulted in the identification of 131 independent loci in the combined-vessel-type GWAS, 114 in the artery-specific GWAS, and 63 in the vein-specific GWAS. In total, we obtained a list of 175 independent lead SNPs. The top 10 are listed in Table 1, ordered by statistical significance (for complete listing see Supplementary File 1 and Supplementary Files 2A/B/C).

Among the significantly associated variants, rs1808382 and rs7991229 had previously been associated with retinal vessel tortuosity[23], while the remaining 173 independent lead SNPs represented novel loci associated with this trait (see Figure 2 for a Manhattan plot of our genome-wide signals). According to the GWAS Catalogue[53], there is a third known locus for retinal[23] vessel tortuosity, namely rs73157566. However, its association is controversial as it was only marginally significant in the discovery cohort and failed to be replicated in the independent cohort of the study that proposed it[23]. We did not replicate this finding either (for details about these three variants and the respective genomic regions, see Supplementary Material on known associations with retinal vessel tortuosity).

Finally, we performed an additional GWAS using the 90th percentile (D9) (as opposed to the median) of the retinal tortuosity measures from each image as the trait (considering all blood vessels, both arteries and veins): of the resulting 151 significant loci, 11 hits were specific to this "D9 tortuosity", while the remaining ones overlapped with those of median tortuosity. Despite this, D9 tortuosity was much less heritable: LD Score Regression $h^2_{SNP} = 0.07$ $(SE = 0.010)$.

Trend of effect sizes replicates in the SKIPOGH cohort

We attempted replication of the lead SNPs from our combined vessel-type UK Biobank[46] GWAS (considering both arteries and veins) using data from the SKIPOGH cohort[47,48] of 397 individuals (multiple images per eye, for a total of 1,352 images). 118 out of 131 lead SNPs were available for comparison, while the remaining 13 could not be tested in the replication, as they had minor allele frequencies which were too small. Given the limited sample size of the replication cohort (*N*=397), we lacked power to replicate any of the individual associations found in the discovery cohort, as none of them remained significant after Bonferroni correction

(p=0.05/118=4.2E-4). Nevertheless, the effect size estimates using the SKIPOGH data showed good concordance with those from the UK Biobank (see Supplementary File 3). First, 81 of 118 lead SNPs had the same sign of their effect size estimate in both studies (binomial test p=6.3E-5). Second, we observed a Pearson correlation of r=0.46 (p=5.51E-7) across these estimates (see Figure 2). Both results remain significant when removing outliers (see Supplementary Material, for replication of effect sizes without outliers).

Tortuosity loci serve pleiotropic functions as disease variants

A shared genetic basis of retinal tortuosity and Coronary Artery Disease had already been noted for locus rs1808382 (mapped to the ACTN4/CAPN12 genes), underlining the usefulness of retinal vascular traits as biomarkers for cardiovascular diseases[23]. We replicated this finding and asked to what extent it also applies to the large panel of new variants we associated with retinal vessel tortuosity. Querying the GWAS Catalogue[53] for our hits revealed 9 loci linked to genes that had been reported as genome-wide significant in associations with other diseases, including coronary heart disease, myocardial infarction, arterial hypertension, type 2 diabetes, chronic lymphocytic leukemia, Alzheimer's disease, diverticular disease, glaucoma, and myopia (see Table 2). Besides these 9 loci, we also uncovered 26 additional SNPs with pleiotropic effects on various diseases which could not be confidently mapped to a specific gene (see full list in Supplementary Material variants associated with disease outcome). We next expanded our query to include phenotypes known to confer a disease risk. We report a list of 12 loci linked to genes influencing both tortuosity and disease risk factors (see Table 3). Furthermore, we uncovered another 9 SNPs showing similar pleiotropic properties, which could not be confidently mapped to a specific gene (see Supplementary Material).

Genetic overlap with metabolic syndrome and other CVD risk factors

We expanded our analysis of overlap with known genetic signals beyond variants with the same rsID, considering SNPs belonging to the same LD block (r^2>0.8). Figure 4 shows how many of the variants identified by our retinal tortuosity GWAS had previously been reported as being associated with any of the complex traits in the GWAS Catalogue[53]. A number of traits related to metabolic syndrome[54] stand out: blood pressure (55 SNPs for SBP, 49 for DBP, 15 for pulse pressure), blood cholesterol levels (54 SNPs), BMI (54 SNPs), blood pressure in relation to alcohol intake and smoking (44 SNPs for SBP, 27 for BDP) and type2 diabetes (5 SNPs). Also other CVD risk factors share a high number of associated variants with tortuosity, such as protein levels[55] (27 SNPs) and type1 diabetes (9 SNPs). We note that we intentionally did not use any of these traits as covariates in our GWAS (nevertheless, correcting tortuosity for these traits leaves our association signals largely unchanged, see Supplementary Material). Furthermore, we observe a sizable number of tortuosity-associated variants overlapping with actual disease, namely CVD (54 SNPs). Besides various eye morphology traits (e.g., optic disc morphometry) (40 SNPs), we observed four additional phenotypes with a sizable overlap of trait associated variants for which the relationship to tortuosity is less obvious, namely: 25 SNPs for lung function, 21 for bone mineral density, 26 for educational attainment, and 21 for hair morphology traits (see Figure 4 for artery/vein-specific results and Supplementary File 4 for the full list of phenotypes and referenced publications).

Mendelian Randomisation reveals causal effect of LDL on tortuosity

We performed bi-directional Mendelian Randomisation[56,57] (MR) to search for evidence of causal effects between tortuosity and some of the traits within the metabolic syndrome. Specifically, we investigated phenotypes for which we had access to results from GWAS with large sample sizes but no overlap with UK Biobank, namely: Body Mass Index (BMI)[58], Coronary artery disease (CAD)[59], Systolic Blood Pressure (SBP)[60], and blood lipids[61] (see Methods for details). We used arterial venous and combined-vessel tortuosity both as outcome and exposure. Using inverse variance weighting MR, we observed that exposure to elevated levels of the blood lipid Low-density lipoprotein (LDL) had a significant negative effect on

arterial (but much less venous or combined) tortuosity (p=0.001). Conversely, venous (but not arterial or combined) tortuosity positively affects BMI (p=0.01) (see Supplementary Table 6).

Gene and pathway scoring reveals key players affecting tortuosity

We obtained gene-wise association scores using PascalX[62], a new high-precision tool to aggregate SNP-wise effects. This resulted in 275 significant genes: 152 from the artery-specific GWAS, 72 from the vein-specific GWAS, and 51 genes that were shared between them. Furthermore, 37 of the 265 genes from the GWAS combining vessel types were specific to this trait. The top genes are listed in Table 5 (for complete listing see Supplementary File 6A/B/C).

We identified a cluster of genes with very low p-values on chromosome 19, which includes ACTN4, a very high-scoring hit in both arteries and veins, related to actin filament bundling (and previously implicated in vessel tortuosity and disease[23]). Functionally related, TNS1 was another top top scoring gene (involved in crosslinking of actin filaments). The locus of ACTN4 (see Suppl. Fig. 1) includes several other genes with strongly associated SNPs that probably all tag the causal variant (or variants). Yet, the functional annotations of some of these genes appear to be relevant for vessel properties: CAPN12, a calcium-activated neutral protease, is involved in extracellular matrix degradation[63]. Moreover, this locus includes three genes coding for galectins, involved in adhesion to the connective tissue via modulation of cell-cell and cell-matrix interactions[64]: LGALS7, its paralog LGALS7B and LGALS4. Other highly significant genes, which are part of the same cluster, are less likely to be directly involved (i.e., EIF3K, MAP4K1, ECH1 and AC104534.3). An additional high-scoring gene was COL4A2 (type IV collagen, alpha 2). It was significant in both arteries and veins, though with a much higher score in arteries. COL4A1 also received a very low p-value in our artery-specific analysis (replicating findings in the literature[23]). Finally, among the artery-specific genes, we found FLT1, which plays a role in vessel formation[65].

Pathway scoring elucidates functional gene groups modulating tortuosity. We aggregated the GWAS summary statistics onto annotated gene-sets using PascalX[62]. Specifically, we first scored only a set of 2,868 strictly canonical pathways, consisting of BioCarta, KEGG, PID, Reactome, and WikiPathways[66,67]. This approach yielded 7 significant hits, 3 artery-specific, 3 veins-specific, and 1 shared. We then scored the full MSigDB[66] datasets sets, which includes GO, as well as sets for microRNA and transcription factor targets for a total of 31,120 gene-sets. This resulted in 53 significant pathways, 46 of which were artery-specific, 4 vein-specific, and 3 shared (shown in Figure 5). Our analysis identified a gene set from "human retinal fibroblasts" and two sets annotated as "vascular smooth muscle cells" in the kidney and the neuroepithelium. Support for tortuosity genes being involved in the epithelium (which protects the outer surface of blood vessels) came from enrichment of the GO category "epithelium development". We also identified four pathways related to endothelium (which lines the interior of blood vessels). Furthermore, we recovered a pathway containing Vascular Endothelial Growth Factors (VEGF), namely VEGFA-VEGFR2, which is the major signaling pathway to activate endothelial angiogenesis. Enrichment of three GO terms related to "response to growth factor" further supported an involvement of tortuosity genes in VEGF-mediated angiogenesis. The role of tortuosity genes in the development of blood vessels was confirmed by GO terms for "circulatory system development", "anatomical structure morphogenesis", and "tube development". Another significantly enriched gene set belonged to "kidney mesangial cells" (which contract to regulate capillary flow and initiate the formation of convoluted capillaries). Finally, three GO terms hinted at structural stability of vessels, providing information about localization ("cell substrate junction", "anchoring junction") and mechanisms ("response to endogenous stimuli"). Nine additional GO categories (related to "actin" and "actomyosin") revealed some of the molecular players involved.

Tortuosity genes are overexpressed in arteries and heart

In order to assess whether genes associated with tortuosity are predominantly expressed in certain tissues, we performed a tissue-specific gene expression analysis on 54 tissues from GTEx[68]. The analysis was again performed by applying PascalX[62] to the sets comprising all

significant genes from each of the three GWAS we performed (i.e., only arteries, only veins, or combining both).

The artery and combined-vessel type analyses were shown to be in close agreement with each other and revealed that genes associated with tortuosity are over-expressed in three types of arteries (i.e., aorta, tibial artery and coronary artery) and two tissues from the heart (i.e., ventricle and atrial appendage). We also found over-expression in fibroblasts and muscular tissues. In contrast to these findings, the vein enrichment analysis did not result in any significant tissue association (using a strict Bonferroni threshold of p=0.05/54=9.2E-4). Figure 6 reports the significance level of all tissues and the Supplementary Material (on tissue-wide gene expression analysis) includes hierarchically clustered heat maps of the expression values of significant gene-sets across tissues.

Link with disease is supported by correlation with patient records

In order to provide further support linking our genetic analysis of retinal vessel tortuosity with disease, we measured the direct association between tortuosity and the disease status reported in patient medical records (UK Biobank[46] medical phenotypes). Associations between retinal vessel tortuosity and disease have been reported in the literature in the case of arterial stiffness[69] and type 2 diabetes[70]. In support of the molecular pathophysiological analysis proposed so far, we report five additional novel associations with disease status: angina, myocardial infarction, stroke, deep vein thrombosis (DVT) and hypertension. P-values were obtained via Welch's t-test (comparing the distribution of tortuosity measurements between cases and controls), while effect sizes were estimated through a linear model evaluating disease status as a function of tortuosity. In the case of angina, z-scored median tortuosity of retinal arteries associated significantly (*beta*=0.09, p=6E-04) and was shown to be even more significant for veinal tortuosity (*beta*=0.18, p=2E-10). For myocardial infarction, z-scored median tortuosity of retinal arteries was not significantly associated with disease, but it was for veinal tortuosity (*beta*=0.12, p=2E-04). Similarly, for stroke and deep vein thrombosis (DVT), significant association of z-scored median tortuosity was only observed for veins (stroke *beta*=0.16, p=5E-05, DVT *beta*=0.11, p=5E-04). Overall, the clearest association was measured between hypertension and median tortuosity of both arteries (*beta*=0.19, p=3E-56) and veins (*beta*=0.25, p=7E-99), confirming a known association between tortuosity and systolic blood pressure[69]. Despite observing a significant association with disease outcome, we found that, after correcting for known risk factors, increased retinal vessel tortuosity hardly improved health risk prediction based on a logistic regression model (see Predictive power over disease status in Supplementary Material).

Discussion

Adapting the retina image processing tool ARIA[49] to run on a cluster facilitated the extraction of median retinal vessel tortuosity estimates for close to 63 thousand subjects of the UK Biobank, enabling Genome Wide Associations Studies for this trait with substantially increasing power compared to previous studies. This gain in power resulted in the identification of 173 novel loci and the replication of all associations known from previous studies, providing a substantially improved picture of the genetic architecture of this trait. We did not observe any significant genomic inflation: LD Score Regression[52] *Intercept* was 1.01 (0.01), indicating that most of the signal was due to polygenicity, with little influence of confounders.

We detected pleiotropic effects of 8 tortuosity loci, which we showed to be independently associated with several diseases, specifically: CVD (Coronary Artery disease, Coronary Heart disease, Myocardial infarction, Hypertension), systemic disorders (diabetes, chronic lymphocytic leukemia, Alzheimer's disease), eye conditions (myopia, glaucoma). All these traits received further support in our extended analysis of SNPs in high LD. An additional locus was shared with diverticular disease, a condition which is characterized by a hyperproliferation of the epithelium[71], a tissue which our gene and pathway analysis clearly pointed to. Our results link these diseases through common associated genetic variants to retinal vessel

tortuosity. Moreover, the effects obtained from our tortuosity GWA Studies were sufficiently strong to apply Mendelian Randomisation[56,57], allowing us to establish that elevated LDL exposure causally reduced arterial tortuosity. Indeed, it is well known that high LDL causes atherosclerosis[72], which has been clinically linked to arterial tortuosity[73,74]: atherosclerotic plaque buildup could make arterial walls less flexible and thereby reduce their tortuosity. The potential causal effect of venous tortuosity on BMI we observed is more enigmatic, yet its significance is only marginal and does not survive correction for multiple hypothesis testing.

Our gene analysis revealed multiple signals pointing towards vascular and connective tissue stability in the context of disease, in line with observational studies[75]. Complementing our analysis of pleiotropic tortuosity variants associated with disease risk (see Table 3), one of the strongest genome-wide associations belonged to an actin gene: ACTN4, coding for the alpha-actinin-4 protein, mediates actin filament bundling, thereby contributing to cell adhesion and motility and is involved in the tight junction assembly in epithelial cells. Additionally, ACTN4 has recently been associated with vasorelaxation[76] (the widening of vessels) which results in a lowering of blood pressure, a mechanism that, when functioning inappropriately, can lead to hypertension. The ACTN4 tortuosity lead SNP is also independently associated with Coronary Artery Disease[23]. Another top hit, TNS1, is functionally connected to ACTN4: its protein crosslinks actin filaments and localizes to focal adhesions (regions of the plasma membrane where the cell attaches to the extracellular matrix). Among the most significant genes, we also found COL4A1 and COL4A2. These paralogues form the hetero-trimer α1α1α2, a tertiary helical structure that serves as the most abundant fundamental unit of the collagen 4 lattice-like network. This network is a part of the basement membrane and forms an extracellular barrier between tissue compartments[77] that potentially influences blood vessel stability. A missense mutation in COL4A1 has been reported in a case study[78] as the cause of familial retinal tortuosity (a rare disorder with autosomal dominant inheritance[79]). COL4A2 figured among our variants with pleiotropic effects on disease risk (see Table 3). High expression levels of ACTN4 and COL4A2 in relevant tissues from the Genotype-Tissue Expression database[68] further corroborated evidence for their involvement in tortuosity (see Supplementary Material). Another particularly interesting hit was FLT1, coding for VEGFR1, which is transcribed in most tissues, including arteries and heart (although not as high as in some tissues like thyroid[68]). It is known to play a suppressive role during embryonic vascular development and promotes atherosclerosis during adulthood[65], but our strong association with retinal tortuosity suggests that it continues to play an important role in vessel formation beyond this stage (at least for the retina). The fact that VEGFR-1 is also expressed in macrophages, may be one of the drivers for the observed genetic overlap of the tortuosity-associated variants with those of colorectal cancer and inflammatory diseases.

Our GO enrichment results contribute further details to the overall picture of the genetic architecture. Tortuosity, as we measure it, can be seen as a simple elongation of vessels due to cardiovascular risk factors such as elevated blood pressure: in line with this, we retrieved the GO categories related to "morphogenesis of anatomical structures", "development of circulatory system" and "tube development" (i.e., a process whose outcome is the progression of a tube, such as the vascular system[80]). The GO category "cellular response to endogenous stimulus", in combination with several GO terms related to structural stability of cells and tissues, suggested that tortuosity might be brought about by a reduced ability of vessels to regain their original shape after they have been internally stretched. Specifically, we found enrichment of "cell substrate junction" and "anchoring junction", which are responsible for the mechanical attachment of a cell and its cytoskeleton to the extracellular matrix. A related GO category, "actin cytoskeleton", points to the actin filament molecule in particular: actin filaments are among the major components of the cytoskeleton and several enriched GO categories point to it ("actin binding", "actin filament bundle organization", "positive regulation of actin filament bundle assembly"). Part of the involvement of actin in the tortuosity phenotype is brought about in its combined form with myosin, as suggested by the enriched GO term "actomyosin structure organization": actomyosin is a constituent of muscle fibers and responsible for the contraction and relaxation of muscles.

Our pathway analysis retrieved gene-sets which had been generated by experiments conducted on the retina itself, or on organs for which microcirculation is known to develop

similarly to that of the retina (namely the kidneys, brain, and pancreas). For instance, we found enrichment in a gene set from transcriptomic analysis of human retinal fibroblasts[81]. Fibroblasts are cells that synthesize structural collagen and the extracellular matrix, which is known to remodel itself in response to signals from the embedded vasculature. This suggests that genetic variants specific to fibroblasts may modulate the response of vessel shape to external mechanical factors, such as those induced by elevated blood pressure.

Our analysis further illustrates how mechanical stimuli act on certain tissues inside and around blood vessels, leading to a change in their morphometric properties, which we measure as tortuosity. Specifically, we recovered two gene-sets belonging to "vascular smooth muscle cells", which compose the walls of blood vessels. A dysregulated response in the natural tendency of vascular smooth muscle cells to contract leads to high blood pressure, and their excessive proliferation[82] contributes to the progression of CVD: this matches our analysis of shared genetic signals between tortuosity and disease GWAS. The first of these two gene-sets referred to the epithelium, the protective layer of cells which lines the outer surfaces of blood vessels (separate support was offered by enrichment for a related GO category, "epithelium development"). The second of the two gene-sets referred to pericytes, which are mural cells that wrap around the endothelium, lining capillaries and venules. The endothelium itself was the source of four other gene-sets. This is particularly fitting as endothelial cells maintain blood vessel tone, a lack of which may result in the tissue bulging and deforming, increasing vessel tortuosity. Importantly, endothelial dysfunction is itself a hallmark of vascular diseases (i.e., coronary artery disease, diabetes mellitus and arterial hypertension). This fits both our analysis of direct correlation of tortuosity with patient records, and our analysis of shared genetic signals with disease GWAS. In the retina, the endothelium plays a direct role in angiogenesis, while two major eye diseases are caused by abnormal neovascularization disease, namely Age-related Macular Degeneration (AMD) and proliferative Diabetic Retinopathy (DR). Fittingly, we specifically recovered VEGFA-VEGFR2, the major signaling pathway to activate angiogenesis (with further support from two GO categories related to the regulation of response to growth factors). Intravitreal anti-VEGF injections is a widely used therapy for both neovascular DR and AMD. These findings thus highlight the potential of the present tortuosity GWAS to identify potential drug targets. Further directions of exploration are opened: we recovered a GO category related to the nuclear transforming growth factor-kappa B (NF-kappaB) which is known to interact with VEGFs and to promote proliferation of vascular smooth muscle cells in several vascular disorders[82]. Moreover, our gene set enrichment suggests a specific mechanism linking TF-mediated angiogenesis and tortuosity in the endothelial cells, namely, tissue stretching in response to elevate blood pressure. Specifically, we found enrichment in a gene set originating from mesangial cells (part of the so-called vascular pole of the kidneys): during development, local endothelial precursors produce a growth factor (PDGF-B) in response to capillary stretching which induces mesangial cells to initiate the development of convoluted capillaries[83]. Mesangial cells also play a role in circulation once they have reached maturity, contracting to regulate capillary blood flow.

The proposed pathways and GO enrichment confirms the results of our analysis of individual gene functions, which highlighted processes related to cell structure maintenance, adhesion, and smooth muscle function. Furthermore, in line with the results of pathways and GO enrichment, our tissue expression results indicate that arterial tortuosity genes are over-expressed in arteries, heart, fibroblasts and muscular tissue. Expression of tortuosity genes was generally elevated across most tissue types, indicating they might be required for the maintenance of basic cellular functions (see Supplementary Material on Tissue-wide gene expression analysis).

This study was subject to a number of limitations. First, the small size of our replication cohort prevented us from replicating any individual hits. Yet, the effects sizes in the replication study correlated strongly with those in the discovery cohort, providing independent evidence that they were not driven by any artifacts specific to the UK Biobank[46]. Further support was provided by a similar trend in the relationships between gene scores in discovery and replication cohorts. The significance of many of our newly discovered loci is so strong that it is very unlikely that they will turn out to be false positives, also because our gene and pathways analysis give further support for these signals making biological sense. Nevertheless, we

make our complete analysis pipeline available for researchers with access to other large collections of fundus images from genotyped subjects, which will enable further validation of our findings and potentially the discovery of additional trait loci.

A further potential limitation of our study is our focus on the Distance Factor as a tortuosity measure for our GWAS. This was motivated by the fact that we obtained a higher estimate of heritability for this measure than for the six other measures that integrate absolute curvature, or its squared, along the vessel. Indeed, while the overall signals are consistent, none of these measures gave rise to as many and as significant associations as the simple Distance Factor, even though various arguments exist in their favor[84]. An explanation could be that the key mechanisms modulating tortuosity through genetic variation affect structural vessel properties that are mainly expressed in terms of vessel elongation, the extent of which is indeed what is measured by the Distance Factor. Further work would be needed to investigate whether the number of inflection points (as captured by the tortuosity density measure[85]) may help to extract additional signals that can elucidate other mechanisms impacting vessel geometry. Finally, we did not adjust for spherical equivalent refractive error, which might have confounded our measurements to some degree. Investigating the impact of this and other corrections on the heritability estimates could shed more light on their usefulness. Indeed, highly powered GWAS may turn out to be a formidable tool to inform method development in defining the most biologically relevant phenotypes from high-dimensional data, such as images.

Materials and Methods

Definition of tortuosity

Our study assessed tortuosity using a measure originally proposed for the femoral artery, known as the Distance Factor (DF)[50]:

$$\tau_{DF} = \frac{s(C)}{chord(C)}$$

where the total vessel length, $s(C)$, is divided by the Euclidean distance between the vessel segment endpoints, $chord(C)$. Estimates for $s(C)$ and $chord(C)$ were provided by the ARIA software[49]. τ_{DF} is referred to in a recent review as the AOC measure (Arc Over Chord ratio)[84]. Some authors subtract 1 from AOC to obtain a tortuosity measure that vanishes for straight vessels ($s(C) = chord(C)$), but this additive constant is immaterial for the GWAS regression analysis. We performed GWAS on six additional tortuosity measures. For their definition, their correlation with τ_{DF} and an analysis of the impact on genetic associations, please refer to Supplementary Material.

UK Biobank phenotypes

Our data were collected as part of the UK Biobank[46] effort, a large-scale study that includes over half a million volunteers from the UK (502,505 participants, collection years 2006-2010). It includes a repeated assessment phase (20,000 participants, collection years 2012-2013). The age of participants ranges between 40 and 69 (median age 59), roughly balanced between sexes (229,122 males and 273,383 females).

175,821 fundus eye images (87,562 images of left eyes and 88,259 images of right eyes) were available at the time of data extraction. We processed all images, including those from the reassessment time point. Other phenotypes were used to correct for biases in the genetic associations, namel: age, sex, and 9 Principal Components (PCs) of genotypes (PCs: 1, 2, 5, 6, 7, 8, 16, 17, 18). Other phenotypes were used to study the correlation with disease and lifestyle. The following health statistics are of interest to interpret the medical implications of our analysis: among the participants for which at least one fundus image was available, 2,644 had been diagnosed with type 2 diabetes, 1,448 with angina, 1,077 with myocardial infarction, 1,072 with deep-vein thrombosis (i.e., blood clot in leg), 750 with stroke, 8,797 lived with stage 2 hypertension (i.e., automated reading of blood pressure >90 mmHg diastolic or >140

mmHg systolic).among the participants for which at least one fundus image was available, 2,644 had been diagnosed with type 2 diabetes, 1,448 with angina, 1,077 with myocardial infarction, 1,072 with deep-vein thrombosis (i.e., blood clot in leg), 750 with stroke, 8,797 lived with stage 2 hypertension (i.e., automated reading of blood pressure >90 mmHg diastolic or >140 mmHg systolic).

Data extraction

The tortuosity phenotype measure was extracted using the software ARIA by Peter Bunkehad[49], which was modified to run in batch. The ARIA[49] parameters used for vessel extraction were kept at their default values (originally fixed via images from the database REVIEW[86] as part of the tool's own system tests).

We now describe the phenotype extraction quality control procedure. The software ARIA[49] was used to perform segmentation of blood vessels and measurement of both their tortuosity and diameters. The software is designed to perform vessel diameter measurements at regular intervals along the centerline of each vessel so that the number of measured diameters could be used as a proxy for the total length of the vascular system depicted in one image. We only retained images in which the total number of equally spaced diameters fell between two thresholds: a minimum of 11,000 and a maximum of 20,000 diameters per image. Similarly, we set thresholds on the number of vessels contained in each eye: a minimum of 100 and a maximum of 250 vessels per image were required(see Supplementary Material). By visual inspection, we confirmed the effectiveness of these thresholds in discriminating low quality images that were too dark, too light, out of focus (lower-bound thresholds), or that contained spurious vessels, that is, artifacts of the picture that were being erroneously segmented as vessels (higher-bound thresholds). Roughly two out of three images passed this strict quality control, with a total of 116,639 images being sent forward in the pipeline. Postprocessing of the data consisted of averaging the values derived from the left and right eye of each participant (for the resulting distribution, refer to Supplementary Material).

The data extraction pipeline was written in python and bash and was run on a cluster using the SLURM job scheduler.

Classification of vessels into arteries and veins

We calculated pixel-wise artery and vein classifications by running raw fundus images through a publicly available pre-trained convolutional neural network, Little W-Net[51]. We defined a score for each vessel as the difference between ARIA's midline vessel points classified by Little W-Net to as arterial (based on the closest pixel) minus those classified being venous. We used this score to call the type of vessel. Testing our approach on a set 44 images manually annotated by an ophthalmologist we obtained an AUC of 0.93 and an accuracy of 0.88. We then performed vessel type classification for the entire set of retina images, allowing us to compute artery- and vein-specific tortuosity values for all vessel segments in each image, from which we calculated the median artery tortuosity and mean vein tortuosity phenotypes (averaged over multiple images, when available).

We estimated the independence of the signals arising from the above described classification of arteries and veins as follows: we modified the pipeline to perform random calling of arteries and veins (by shuffling the vector of artery and veins scores computed for each eye). We then compared the similarity in the signal between two random vessel-type GWAS: we clearly show that the effect sizes (which are significant in at least one of the two GWAS) are nearly identical (r=0.99, p=4E-82) when the vessel type calling is random. By comparison, the effect sizes are much less coupled (r=0.76, p=1E-20) in the artery- and vein-specific GWAS based on the vessel type calling procedure that has been described (for a plot, refer to Supplementary Material). This supports the effectiveness of our vessel type calling procedure.

UK Biobank genotype data

Around 488,000 participants were genotyped on Axiom arrays for a total of 805,426 markers. From this, about 96 million genotypes were imputed using a combined reference panel from the 1000 Genomes and UK10K projects[87]. The annotation used to report variant positions is the Genome Reference Consortium Human genome build 37 (known as GRChb37 or hg37). We subset the genotypes using the software BGENIX[88], shrinking the list of investigated variants to those that have been assigned an rsID (15,599,830 SNPs). An additional Quality Control was performed via a postprocessing step on the GWAS output. We filtered out SNPs with MAF < 5E-4 (which, given the sample size of 62,751 subjects, translates to an expected minimal number of ~60 individuals having at least one minor allele). We applied a filtering procedure[89] removing SNPs with imputation quality < 0.3.

The SKIPOGH study

We performed replication of the GWAS results in the *Swiss Kidney Project On Genes in Hypertension* (SKIPOGH) cohort[47,48]. This is a family- and population-based cohort including 1'042 participants (493 males and 549 females), aged between 18 and 96 years old, which have been extensively phenotyped at baseline and in a 3-year follow-up. Participants were recruited from three different locations in Switzerland, namely, Bern, Geneva and Lausanne. For 397 participants (186 males and 211 females) retina images from both eyes were available. The genotyping was performed with the Illumina Omni 2.5 chip, followed by an imputation based on HRC v1.1 panel using Minimach3. The annotation used to report variant positions is the Genome Reference Consortium Human genome build 37 (GRCh37/hg19).

Genome-wide association analysis

The raw tortuosity measures extracted from the image data were normalized using a rank-based inverse normal transformation, and genetic association studies were run using the software BGENIE[90]. In order to account for confounding effects[91], the following variables were provided as regression covariates: age, sex, Principal Components (PCs) of the genotypes (only 9 PCs, which showed a statistically significant correlation to tortuosity, were considered). For an overview of the confounder analysis that we performed, please refer to Supplementary Material.

Using a Bonferroni threshold of 5E-8 the (unpruned) sets of significant SNPs consist of 6,725 SNPs for the combined-vessel-type GWAS, 6,332 SNPs for the artery-specific GWAS, and 2,895 SNPs for the vein-specific GWAS (see Supplementary Files 2A, 2B and 2C).

All significant genetic associations were pulled and the list of independent SNPs was calculated by performing LD pruning using the LDpair function of the R package LDlinkR[92], selecting as a reference panel "GBR" (Great Britain) from the 1k genome project. Two SNPs were considered as independent if they had LD r^2<0.1 or were more than 500K bases apart (see Supplementary File 1). This resulted in 175 independent lead SNPs.

The association analysis at replication stage for the SKIPOGH cohort was performed using the Emmax function of the Epacts software[93] in order to account for family structure by using the kinship matrix in the model. Additionally, the recruitment center was included as a covariable.

Summary plots were generated using the R packages GWASTools[94].

Shared genetic architecture with disease

The SNP variants' overlap with disease phenotypes (matching rsid) was analysed using the GWAS Catalogue[53]. We report independent SNPs in the combined-vessel-type tortuosity GWAS that are part of the GWAS Catalogue[53], and therefore have been previously associated with disease (or with a disease-related phenotype) (see Figure 3, Table 2 and Table 3). We also list independent SNPs in the combined-vessel-type tortuosity GWAS who were in LD with SNPs in the GWAS Catalogue[53] (see Figure 4 and Supplementary Files 4 and 5).

Mendelian Randomisation analysis

We performed two sample bi-directional Mendelian Randomisation (MR) to search for evidence of causal effect between tortuosity and the following traits: BMI, CAD, SBP and lipid traits, namely high-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol (TC) and triglycerides (TG). For each trait, we used independent (r^2<0.01) significant (P<5E-08) SNPs as instrumental variables. All summary statistics (estimated univariate effect size and standard error) originate from the most recent meta-analyses (not including UK Biobank individuals) and were downloaded from the publicly available NIH Genome-wide Repository of Associations Between SNPs and Phenotypes[98]. We only used SNPs on autosomal chromosomes and available in the UK10K reference panel[97], which allowed us to estimate the linkage disequilibrium among these SNPs and prune them. Strand ambiguous SNPs were removed. Causal estimates were based on the inverse variance weighted (IVW) method[97,99] and were calculated using the Mendelian Randomisation R package[100].

Gene-based tests

For this study, we used PascalX[62], a novel high precision implementation of the Pascal[95] pathway scoring algorithm, to aggregate SNP-wise summary statistics into gene scores using a sum of χ^2 statistics: PascalX[62] takes into account LD by effectively transforming the sum of χ^2 from all SNPs within the gene window into a new basis of independent "eigen-SNPs" corresponding to a weighted sum of χ^2 statistic. Using multiple-precision arithmetic, PascalX[62] computes the corresponding null cumulative probability distribution to essentially arbitrary precision (while other tools only approximate the underlying distribution). Using this novel software implementation, we computed p-values up to a precision of 1E-100, allowing for accurate scoring of genes with contributions from extremely significant SNPs, which become increasingly frequent in highly powered GWAS such as this one.

The following configurations were used. Gene scores were computed from SNPs that lie within the extended transcript window of plus 50kb at both the start and end sites. The annotation on the gene positions was based on the Genome Reference Consortium Human genome build 37 (GRCh37/hg19), downloaded from the ensembl biomart[96]; only protein-coding and lincRNA genes were considered. The reference panel from the UK10K project[97] was used to estimate the SNP-SNP correlations (LD effects). PascalX[62] uncovered 265 genome-wide significant genes (after Bonferroni correction for 25,489 gene-based tests p<0.05/25,489=2.0E-6).

Pathway and GO enrichment

We also used PascalX[62] to compute pathway scores based on a gene ranking derived from the gene scores (since a large number of genes have inflated p-values, this ranking approach is more conservative). We first computed scores for 2,868 canonical pathways (BioCarta, KEGG, PID, Reactome, and WikiPathways[66,67]) and then extended our analysis to all of the 31,120 pathways in MSigDB (v7.2)[66]. To adjust for statistical dependence and co-expression, genes which are less than 100kb apart were "fused" (i.e. considered as single entities termed "fusion genes"). Scores for these fusion genes were computed using the same parameters as for the ordinary gene scoring.

Tissue-wide gene expression analysis

Tissue-wide gene expression analysis was performed using PascalX[62] on the whole GTEx[68] (v8) dataset (54 tissues). We defined gene-sets based on the significant genes from each of our three GWAS on median vessel tortuosity (artery, vein and combined-vessel-type). PascalX[62] performed an enrichment analysis to calculate whether these sets were over-expressed in any particular tissue. PascalX[62] corrected for co-expression of gene sub-clusters within each gene set by merging nearby genes to meta-genes. Gene expression values in Transcripts Per Kilobase Million (TPM) were computed for these meta-genes from the raw read counts. The TPM values were uniformized via ranking, transformed to χ^2-distributed random variables, summed, and tested against a χ^2 distribution with as many degrees of freedom as there were "fusion genes" in each set.

Figures

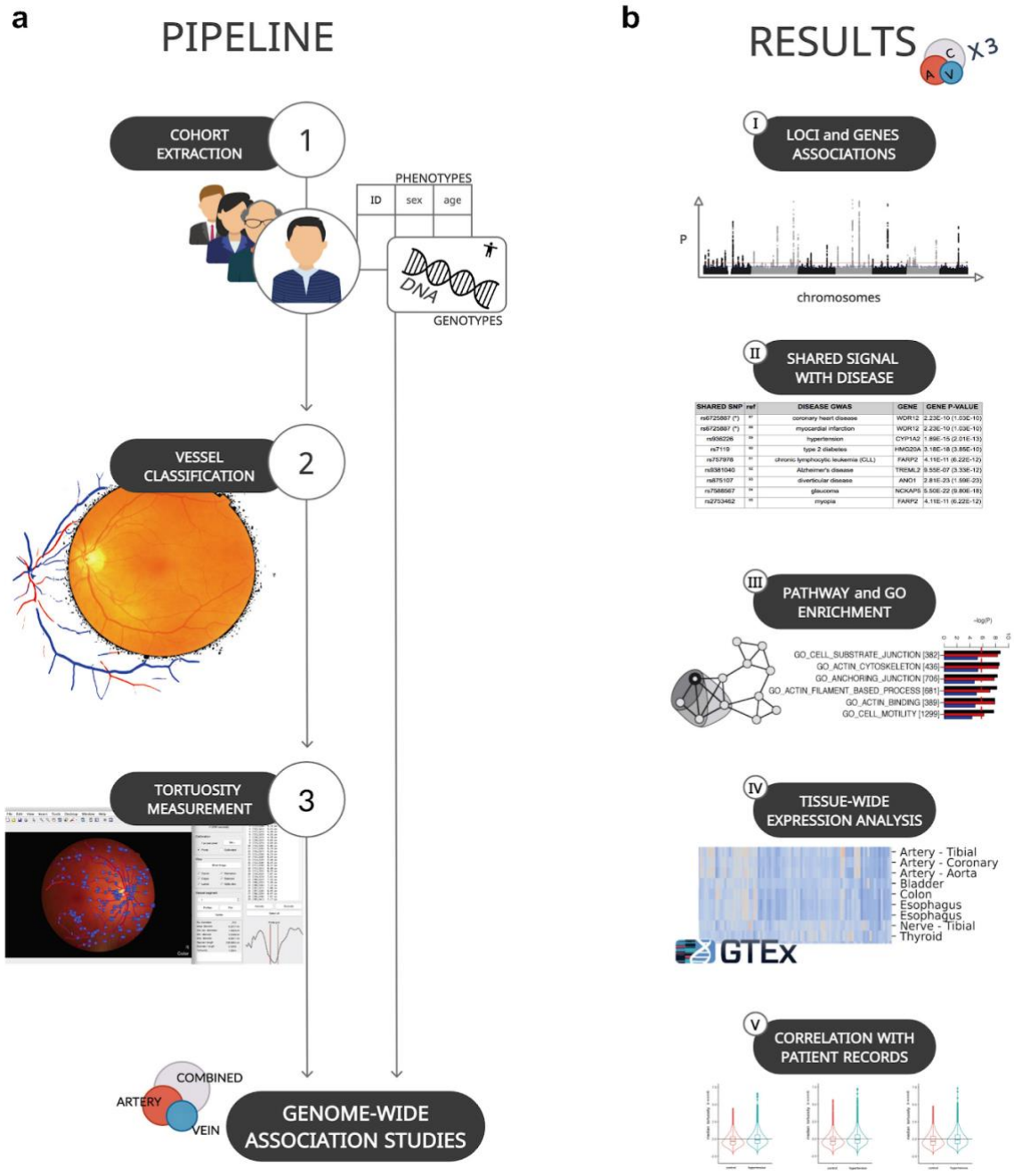

Figure 1 **| Pipeline and results. a,** The analysis pipeline of this study in three steps. First, data were extracted (from the UK Biobank and SKIPOGH cohorts). Relevant phenotypes (i.e., confounders) and genotypes were collected, as well as fundus eye images. Second, the images were processed to extract positions of blood vessel segments and classify them as arterial or venous. Third, a range of tortuosity measures were calculated, which provided the phenotypes for combined and vessel-type specific GWAS. **b,** The primary results were 173 novel genetic trait loci and 275 trait-associated genes. The association signals pointed to a shared genetic architecture of retinal tortuosity and disease (metabolic syndrome and CVD). Their aggregation on annotated gene-sets identified relevant pathways and GO terms. Tissue-wide expression analysis revealed the most relevant tissues for these pathways. Correlation analysis revealed the predictive value of retinal tortuosity with patient records on relevant diseases.

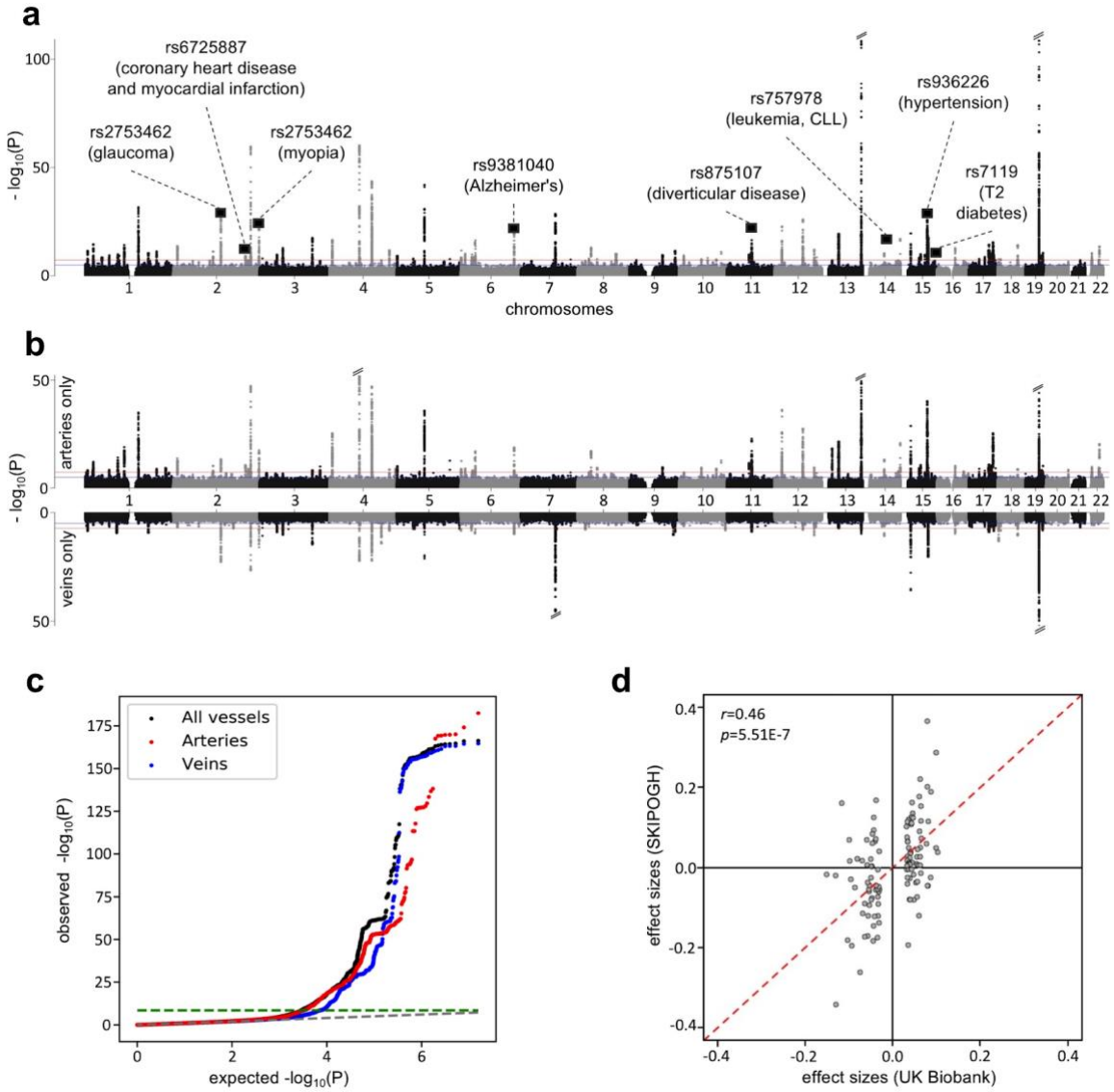

Figure 2 **| SNP p-values and effects. a**, Manhattan plot of Genome-Wide Association Study (GWAS) of median retinal vessel tortuosity, combining all vessel types (both arteries and veins). The red line indicates the genome-wide significance level after Bonferroni correction (*p*=5E-8). Oblique dashes on top of peaks mark extremely significant p-values that have been cropped. Squares mark the position of disease SNPs (see Table 2). The trait was corrected for phenotypic variables which showed a statistically significant association, i.e.: age, sex, and a subset of principal components (PCs) of genotypes. **b**, Manhattan plots of the vessels-specific GWAS (artery-specific on top, vein-specific at the bottom). Confounder correction, significance level and cropping of extremely significant p-values as in the (a). **c**, GWAS Q-Q plot: arteries in red, veins in blue, combined signal in black; the genome-wide significance level is represented as a green dashed line. **d**, Statistically significant correlation between the measured effect sizes in the discovery cohort (UK Biobank, *N*=62,751) and replication cohort (SKIPOGH, *N*=397). We considered all lead (independent) SNPs in the UK Biobank. We could find 118 with matching rsIDs in SKIPOGH, 81 of which had the same sign of their effect size estimate in the UK Biobank. The resulting correlation has Pearson *r*=0.46 and *p*=5.51E-7.

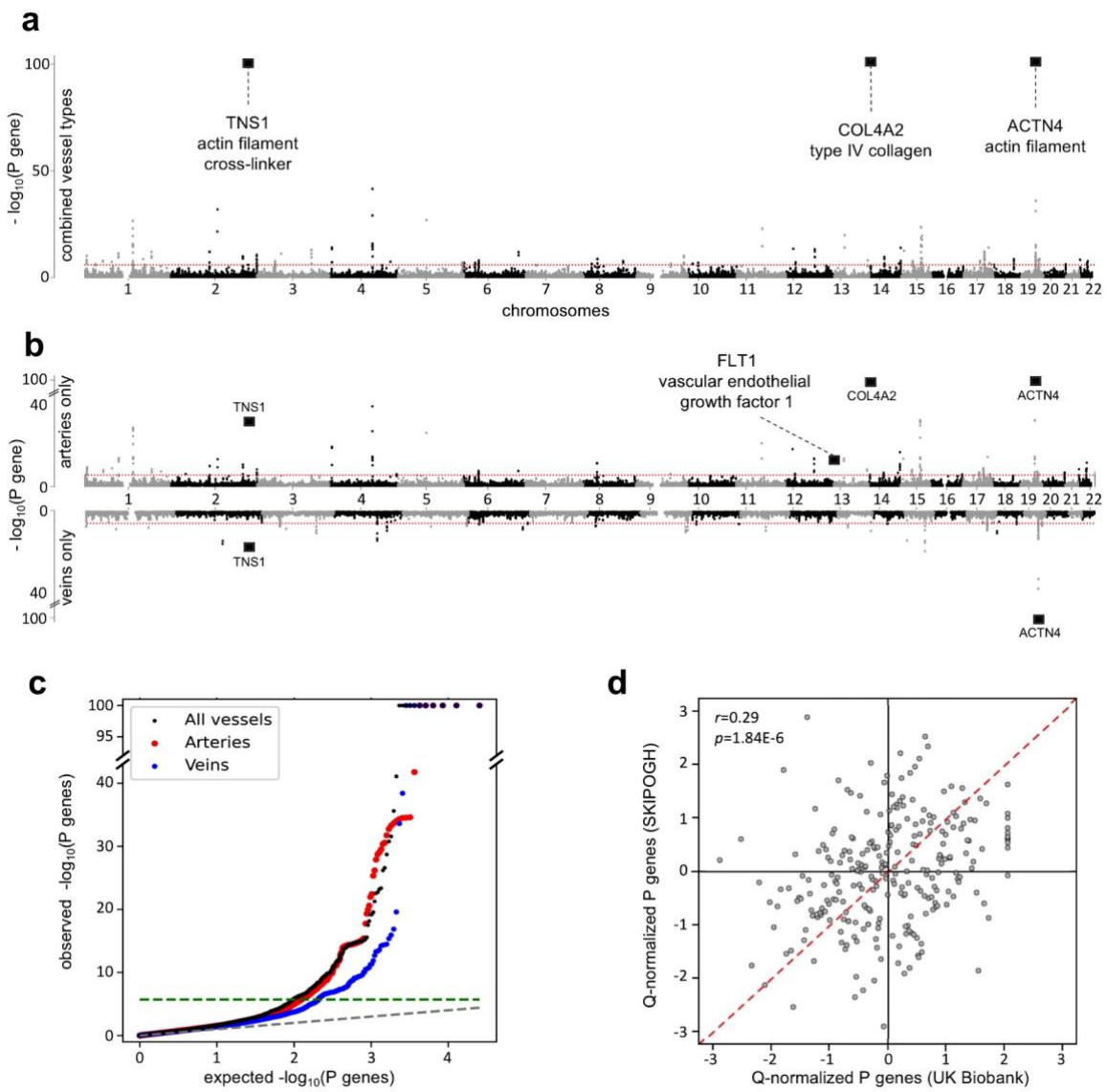

Figure 3 **| Gene p-values and replication scores. a,** Gene-based Manhattan plot of median retinal vessel tortuosity, combining all vessel types (both arteries and veins). Gene-based tests were computed by PascalX[62]. The red line indicates the genome-wide significance level after Bonferroni correction (p=0.05/25,489=2E-6). Squares mark the position of particularly relevant genes (see corresponding Results section). **b,** Gene-based Manhattan plots of the vessels-specific GWAS (artery-specific on top, vein-specific at the bottom). **c,** Q-Q plot of gene p-values: arteries in red, veins in blue, combined signal in black; the genome-wide significance level is represented as a green dashed line. **d,** Statistically significant correlation between Q-Q normalized genes' p-values in the discovery (UK Biobank) and in the replication cohort (SKIPOGH). SNP-wise effects were aggregated to genes using PascalX[62]. Only genes that were significant in the discovery cohort were considered. The resulting correlation has Pearson r=0.29 and p=1.84E-6.

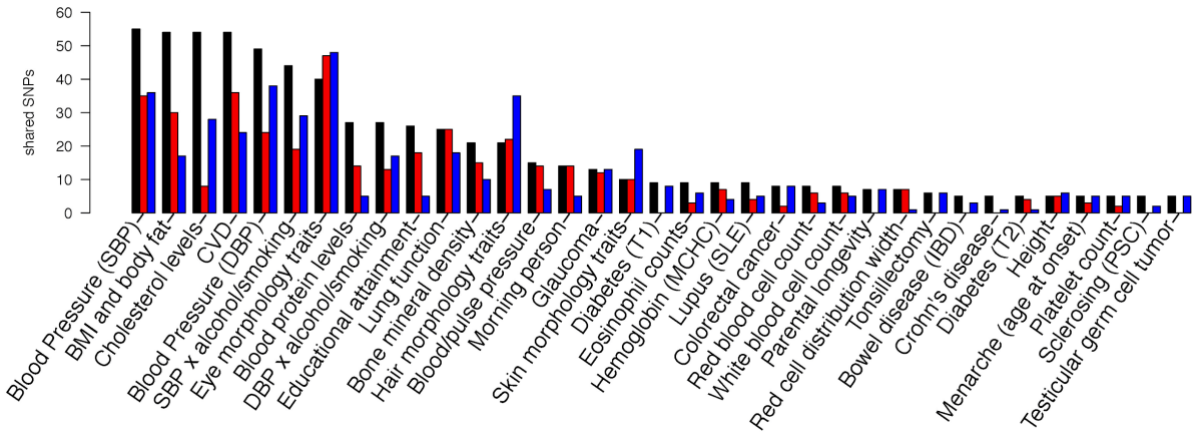

Figure 4 **| Overlap in genetic signals with diseases and other complex traits.** Number of variants shared with other traits reported in the GWAS Catalogue[53] (also considering SNPs in high LD with the lead SNP, r^2>0.8). Only traits with at least 5 shared associations are included (for a full list, including rsIDs, refer to the Supplementary File 5). The traits with the highest number of shared SNPs belong to metabolic syndrome (blood pressure, BMI, blood cholesterol levels) and CVD. This analysis was generated using FUMA[101].

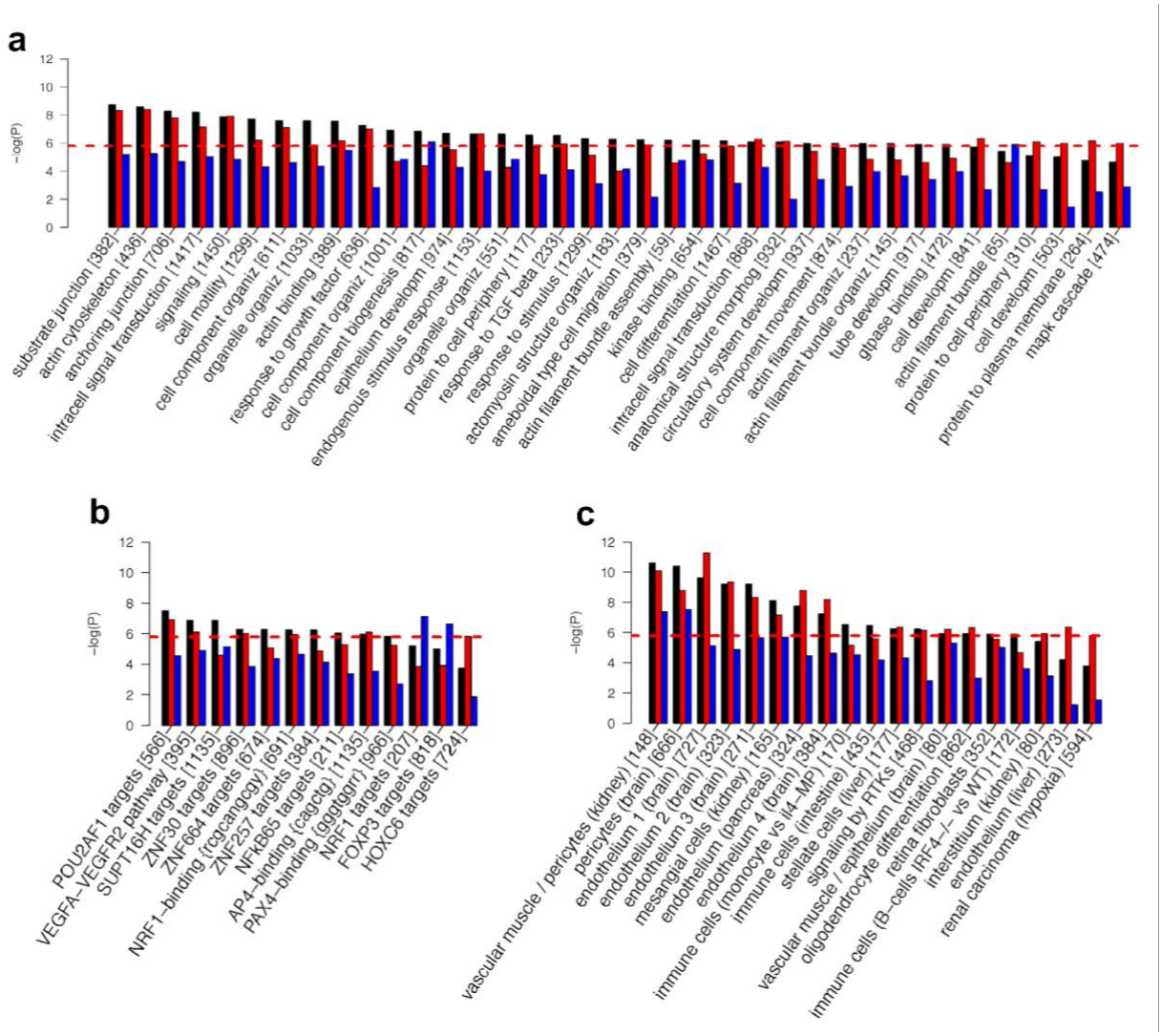

Figure 5 **| Enriched pathways and gene-sets.** Scores for 31,120 gene-sets in MSigDB (v7.2)[66] were calculated by PascalX[62]. Red bars correspond to enrichment results from the artery GWAS, blue correspond to veins, and black correspond to the combined-vessel-type GWAS. Only gene-sets for which significance was reached by at least one GWAS are shown. The red dashed line indicates Bonferroni-threshold (-log10 p=5.7). The number of

genes in each set is indicated in squared brackets. Gene-set names have been shortened and some GO categories are not shown (because too general or redundant). For details, refer to the extended plot in Supplementary Material. **a,** Enrichment in Gene Ontology (GO) categories. **b,** Enrichment in pathways referring to a particular molecule (typically a transcription factor) and/or binding motif. **c,** Enrichment in gene-set obtained from transcriptomic analysis of tissues of treated cell types.

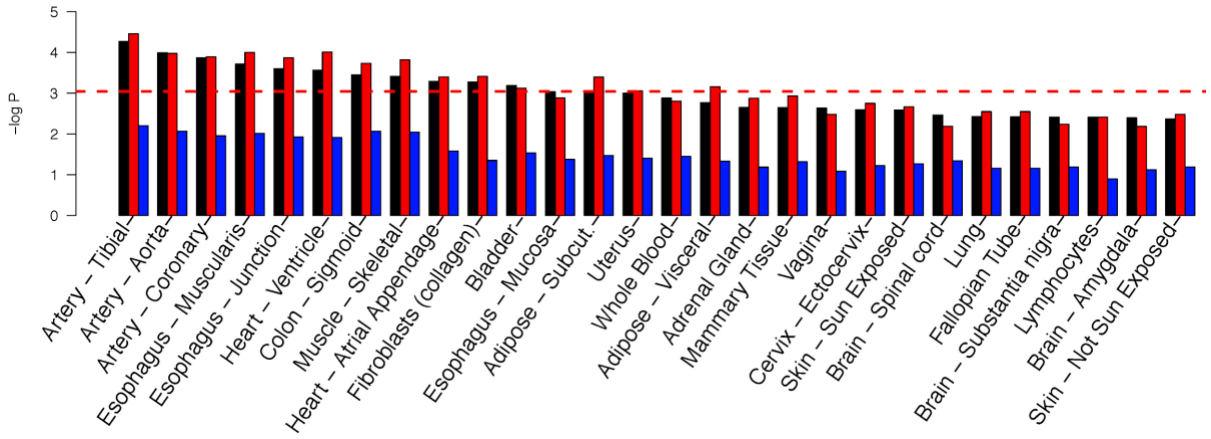

Figure 6 **| Tissue expression results.** Tissue-specific gene expression analysis of GTEx (v8)[68] performed using PascalX[62]. We defined sets based on the significant genes from each of the three GWAS we carried out and asked whether they were over-expressed in a particular tissue: red bars represent the estimated significance for artery genes, while significance of vein genes is in blue, and in black os the significance of genes from the combined-vessel-type GWAS.

Tables

CHR	SNP	BP	EA	NEA	FREQ	BETA	-log P	signal type
13	rs9559797	111085411	G	C	0.580	-0.162	182.4	artery and vein
19	rs16972767	39153044	G	A	0.473	-0.150	164.6	artery and vein
4	rs17008193	85078215	T	C	0.403	-0.089	55.0	artery and vein
7	rs187691758	96411753	A	G	0.005	0.627	53.2	vein
4	rs12506823	119974397	G	A	0.406	0.083	47.9	artery and vein
2	rs2571461	218670945	T	G	0.601	-0.082	47.1	artery and vein
15	rs12913832	28365618	A	G	0.744	0.080	43.6	artery and vein
12	rs11045245	20589390	A	G	0.375	0.073	37.3	artery
5	rs784420	77987524	A	G	0.281	0.078	37.0	artery and vein
4	rs11727963	120278100	G	A	0.166	0.092	35.1	artery

Table 1 **| Top retinal tortuosity SNPs.** The 10 most significant retinal tortuosity SNPs, ordered by p-value. For full results, refer to the list of 175 independent lead SNPs in Supplementary File 1. Abbreviations: CHR, Chromosome; BP, base pair position; EA, effect allele; NEA, non-effect allele; FREQ, allele frequency of effect allele; BETA, effect size estimate; Chromosomal positions are in GRCh37 coordinates.

SHARED SNP	ref	DISEASE GWAS	GENE	-log P
rs875107	[102]	diverticular disease	ANO1	22.5

rs7588567	[103]	glaucoma	NCKAP5	21.2
rs7119	[104]	type 2 diabetes	HMG20A	17.5
rs936226	[105]	hypertension	CYP1A2	14.7
rs757978	[106]	chronic lymphocytic leukemia (CLL)	FARP2	10.4
rs2753462	[107]	myopia	FARP2	10.4
rs6725887 (*)	[108]	coronary heart disease	WDR12	9.6
rs6725887 (*)	[109]	myocardial infarction	WDR12	9.6
rs9381040	[110]	Alzheimer's disease	TREML2	6.0

Table 2 **| Pleiotropic disease-variants.** List of variants identified by the UK Biobank tortuosity GWAS which were independently found to be associated with disease outcome in an independent study. We report only exact variants (same rsID in both tortuosity and disease GWAS) and variants which we could confidently map to a gene (in the combined-vessel-type analysis, which included both arteries and veins). Gene p-values were computed by PascalX[95]. Variants associated with more than one disease are marked by a star (*).

SHARED SNP	ref	RISK FACTOR (diseases)	GENE	-log P
rs11083475	[111]	heart rate (rhythm disorders)	ACTN4	> 100
rs9555695	[112]	waist-hip ratio (obesity)	COL4A2	> 100
rs2571445	[113]	lung function (pulmonary disease)	TNS1	> 100
rs3791979	[114]	intraocular pressure (open angle glaucoma)	TNS1	> 100
rs17263971	[115]	eGFR (Chronic Kidney Disease)	SYNPO2	28.7
rs35252676	[116]	pulse pressure (CVD)	LHFPL2	26.6
rs1378942 (*)	[117]	diastolic blood pressure (CVD)	CSK	23.4
rs1378942 (*)	[118]	mean arterial pressure (CVD)	CSK	23.4
rs17355629	[119]	pulse pressure (CVD)	LRCH1	19.6
rs7655064	[112]	waist-hip ratio (obesity)	MYOZ2	14.5
rs6495122	[120]	diastolic blood pressure (CVD)	CPLX3	14.5
rs12913832	[121]	intraocular pressure (open angle glaucoma)	HERC2	12.3
rs9303401	[122]	cognitive ability (mental disorders)	PPM1E	10.01

Table 3 **| Pleiotropic risk factor variants.** List of variants identified by the UK Biobank tortuosity GWAS which were independently found to be associated with disease risk factors in an independent study. We report only exact variants (same rsID in both tortuosity and disease GWAS). We report only variants which we could confidently map to a gene (in the combined-vessel-type analysis, which included both arteries and veins). Gene p-values were computed by PascalX[95] (precision cutoff: 1E-100). Variants associated with more than one disease are marked by a star (*).

GENE	CHR	BP	-log P COMBINED	-log P ARTERIES	-log P VEINS
ACTN4	19	39,138,289	> 100	> 100	> 100
CAPN12	19	39,220,827	> 100	> 100	> 100

EIF3K	19	39,109,735	> 100	> 100	> 100
LGALS7	19	39,261,611	> 100	> 100	> 100
LGALS7B	19	39,279,851	> 100	> 100	> 100
COL4A2	13	110,958,159	> 100	> 100	9.5
LGALS4	19	39,292,311	> 100	34.5	> 100
MAP4K1	19	39,078,281	> 100	34.3	> 100
TNS1	2	218,664,512	> 100	32.7	16.9
ECH1	19	39,306,062	> 100	15.3	> 100
AC104534.3	19	39,310,806	> 100	14.0	> 100

Table 5 | **Top retinal tortuosity genes.** The 15 most significant retinal tortuosity genes. P-values were computed by PascalX[95] (precision cutoff: 1E-100). For full results, refer to Supplementary Files 6A/B/C.

GWAS type	h^2_{SNP}	*Lambda GC*	*Mean Chi²*	*Intercept*	*Ratio*
combined-vessel-types	0.25 (0.025)	1.14	1.31	1.01 (0.01)	0.03 (0.03)
artery-only	0.23 (0.020)	1.12	1.27	1.00 (0.01)	< 0
vein-only	0.15 (0.021)	1.10	1.18	1.00 (0.01)	< 0

Table 6 **| SNP-based heritability.** h^2_{SNP} is the portion of phenotypic variance cumulatively explained by the SNPs. *Lambda GC* is the measure of inflation (it measures the effect of confounding and polygenicity acting on the trait). *Intercept* is the LD Score Regression intercept (values close to 1 indicates little influence of confounders, mostly of population stratification). *Ratio* is the ratio of the proportion of the inflation in the *Mean Chi²* that is not due to polygenicity (a ratio close to, or smaller than, 0 is desirable as it indicates low inflation from population stratification). SE are given in parentheses.

Author contributions

MT, MB and SB designed the study. MT and MJB performed QC on the raw images. MT extracted tortuosity measurements from the image data (UK Biobank and SKIPOGH cohorts) and post-processed them. MJB performed classification of arteries and veins. MT carried out the median tortuosity GWA studies (based on the Distance Factor measure), with the guidance of SB, DK, NM and EP. MJB carried out gene and pathway scoring. ALB performed LD Score Regression analysis. MT performed subsequent bioinformatics analyses. SOV evaluated the correlation between different tortuosity measures and their impact on genetic associations. TC and MT performed the replication analysis in SKIPOGH. HA provided ophthalmological expertise and manually annotated the raw image data. MT and SB drafted the manuscript, which was revised by MJB, HA, MB, DK and ALB.

Competing interests

The authors declare no competing interests.

Acknowledgements

This work was conducted using the UK Biobank (application ID 43805) and SKIPOGH. Thanks to Micha Hersch for inspiring this project, to the UK Biobank team for their support and responsiveness, and to all UK Biobank participants for sharing their personal data.

Funding

This work was supported by the Swiss National Science Foundation (#FN 310030_152724/1 to SB). The SKIPOGH study was also supported by the Swiss National Science Foundation (#FN 33CM30-124087 to MB).

Additional information

Please refer to the Supplementary material and files.

References

1. MacCormick, I. J. C., Czanner, G. & Faragher, B. Developing retinal biomarkers of neurological disease: an analytical perspective. *Biomark. Med.* **9**, 691–701 (2015).
2. Patton, N. *et al.* Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: a rationale based on homology between cerebral and retinal microvasculatures. *J. Anat.* **206**, 319–348 (2005).
3. Liao, H., Zhu, Z. & Peng, Y. Potential Utility of Retinal Imaging for Alzheimer's Disease: A Review. *Front. Aging Neurosci.* **10**, 188 (2018).
4. Dumitrascu, O. M. & Qureshi, T. A. Retinal Vascular Imaging in Vascular Cognitive Impairment: Current and Future Perspectives. *J. Exp. Neurosci.* **12**, 1179069518801291 (2018).
5. Baker, M. L., Hand, P. J., Wang, J. J. & Wong, T. Y. Retinal signs and stroke: revisiting the link between the eye and brain. *Stroke* **39**, 1371–1379 (2008).
6. Sabanayagam, C. *et al.* A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *The Lancet Digital Health* **2**, e295–e302 (2020).
7. Park, H. C. *et al.* Diabetic retinopathy is a prognostic factor for progression of chronic kidney disease in the patients with type 2 diabetes mellitus. *PLoS One* **14**, e0220506 (2019).
8. Nash, B., Carlson, M. L. & Van Gompel, J. J. Microvascular decompression for tinnitus: systematic review. *J. Neurosurg.* **126**, 1148–1157 (2017).
9. Weiler, D. L., Engelke, C. B., Moore, A. L. O. & Harrison, W. W. Arteriole tortuosity associated with diabetic retinopathy and cholesterol. *Optom. Vis. Sci.* **92**, 384–391 (2015).
10. Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **316**, 2402–2410 (2016).
11. Smith, W. *et al.* Retinal arteriolar narrowing is associated with 5-year incident severe hypertension: the Blue Mountains Eye Study. *Hypertension* **44**, 442–447 (2004).
12. Mookiah, M. R. K. *et al.* Application of different imaging modalities for diagnosis of Diabetic Macular Edema: A review. *Comput. Biol. Med.* **66**, 295–315 (2015).
13. Wang, J. J. *et al.* Retinal vessel diameters and obesity: a population-based study in older persons. *Obesity* **14**, 206–214 (2006).
14. Poplin, R. *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* **2**, 158–164 (2018).
15. Grunwald, J. E. *et al.* Retinopathy and the risk of cardiovascular disease in patients with chronic kidney disease (from the Chronic Renal Insufficiency Cohort study). *Am. J. Cardiol.* **116**, 1527–1533 (2015).
16. Ikram, M. K., Ong, Y. T., Cheung, C. Y. & Wong, T. Y. Retinal vascular caliber measurements: clinical significance, current knowledge and future perspectives. *Ophthalmologica* **229**, 125–136 (2013).
17. Flammer, J. *et al.* The eye and the heart. *Eur. Heart J.* **34**, 1270–1278 (2013).
18. Cuspidi, C., Sala, C. & Grassi, G. Updated classification of hypertensive retinopathy: which role for cardiovascular risk stratification? *Journal of hypertension* vol. 33 2204–2206 (2015).
19. Seidelmann, S. B. *et al.* Retinal Vessel Calibers in Predicting Long-Term Cardiovascular Outcomes: The Atherosclerosis Risk in Communities Study. *Circulation* **134**, 1328–1338 (2016).
20. Kawasaki, R. *et al.* Retinal vessel diameters and risk of hypertension: the Multiethnic Study of Atherosclerosis. *J. Hypertens.* **27**, 2386–2393 (2009).
21. Ikram, M. K. *et al.* Retinal vessel diameters and risk of stroke: the Rotterdam Study. *Neurology* **66**, 1339–1343 (2006).
22. Liew, G. *et al.* Fractal analysis of retinal microvasculature and coronary heart disease mortality. *Eur. Heart J.* **32**, 422–429 (2011).
23. Veluchamy, A. *et al.* Novel Genetic Locus Influencing Retinal Venular Tortuosity Is Also Associated With Risk of Coronary Artery Disease. *Arterioscler. Thromb. Vasc. Biol.* **39**, 2542–2552 (2019).
24. Wong, T. & Mitchell, P. The eye in hypertension. *Lancet* **369**, 425–435 (2007).
25. Cheung, C. Y.-L. *et al.* Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. *Ophthalmology* **118**, 812–818 (2011).
26. Grosso, A., Veglio, F., Porta, M., Grignolo, F. M. & Wong, T. Y. Hypertensive retinopathy revisited: some answers, more questions. *Br. J. Ophthalmol.* **89**, 1646–1654 (2005).
27. Wong, T. Y., Shankar, A., Klein, R., Klein, B. E. K. & Hubbard, L. D. Prospective cohort study of retinal vessel diameters and risk of hypertension. *BMJ* **329**, 79 (2004).
28. Wong, T. Y., Klein, R., Klein, B. E. K., Meuer, S. M. & Hubbard, L. D. Retinal vessel diameters and their associations with age and blood pressure. *Invest. Ophthalmol. Vis. Sci.* **44**, 4644–4650 (2003).
29. Dimmitt, S. B. *et al.* Usefulness of ophthalmoscopy in mild to moderate hypertension. *Lancet* **1**, 1103–1106 (1989).
30. Leung, H. *et al.* Impact of current and past blood pressure on retinal arteriolar diameter in an older population. *J. Hypertens.* **22**, 1543–1549 (2004).
31. Wang, J. J. *et al.* Hypertensive retinal vessel wall signs in a general older population: the Blue Mountains Eye Study. *Hypertension* **42**, 534–541 (2003).
32. Wong, T. Y. *et al.* Retinal arteriolar diameter and risk for hypertension. *Ann. Intern. Med.* **140**, 248–255 (2004).
33. Wong, T. Y. *et al.* Retinal microvascular abnormalities and blood pressure in older people: the Cardiovascular Health Study. *Br. J. Ophthalmol.* **86**, 1007–1013 (2002).
34. Ikram, M. K. *et al.* Retinal vessel diameters and risk of hypertension: the Rotterdam Study. *Hypertension* **47**, 189–194 (2006).
35. Sharrett, A. R. *et al.* Retinal arteriolar diameters and elevated blood pressure: the Atherosclerosis Risk in Communities Study. *Am. J. Epidemiol.* **150**, 263–270 (1999).

36. Woo, S. C. Y., Lip, G. Y. H. & Lip, P. L. Associations of retinal artery occlusion and retinal vein occlusion to mortality, stroke, and myocardial infarction: a systematic review. *Eye* **30**, 1031–1038 (2016).
37. Rim, T. H. *et al.* Retinal vein occlusion and the risk of acute myocardial infarction development: a 12-year nationwide cohort study. *Sci. Rep.* **6**, 22351 (2016).
38. Smith, W. *et al.* Retinal arteriolar narrowing is associated with 5-year incident severe hypertension: the Blue Mountains Eye Study. *Hypertension* **44**, 442–447 (2004).
39. Mookiah, M. R. K. *et al.* Application of different imaging modalities for diagnosis of Diabetic Macular Edema: A review. *Comput. Biol. Med.* **66**, 295–315 (2015).
40. Nannini, D. R. *et al.* A Genome-Wide Association Study of Vertical Cup-Disc Ratio in a Latino Population. *Invest. Ophthalmol. Vis. Sci.* **58**, 87–95 (2017).
41. Jensen, R. A. *et al.* Novel Genetic Loci Associated With Retinal Microvascular Diameter. *Circ. Cardiovasc. Genet.* **9**, 45–54 (2016).
42. Ikram, M. K. *et al.* Four novel Loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. *PLoS Genet.* **6**, e1001184 (2010).
43. Springelkamp, H. *et al.* Meta-analysis of Genome-Wide Association Studies Identifies Novel Loci Associated With Optic Disc Morphology. *Genet. Epidemiol.* **39**, 207–216 (2015).
44. Han, X. *et al.* Genome-wide association analysis of 95 549 individuals identifies novel loci and genes influencing optic disc morphology. *Hum. Mol. Genet.* **28**, 3680–3690 (2019).
45. MacGillivray, T. J. *et al.* Suitability of UK Biobank Retinal Images for Automatic Analysis of Morphometric Properties of the Vasculature. *PLoS One* **10**, e0127914 (2015).
46. Sudlow, C. *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* vol. 12 e1001779 (2015).
47. Pruijm, M. *et al.* Heritability, determinants and reference values of renal length: a family-based population study. *Eur. Radiol.* **23**, 2899–2905 (2013).
48. Ponte, B. *et al.* Reference values and factors associated with renal resistive index in a family-based population study. *Hypertension* **63**, 136–142 (2014).
49. Bankhead, P., Scholfield, C. N., McGeown, J. G. & Curtis, T. M. Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PLoS One* **7**, e32435 (2012).
50. Smedby, O. *et al.* Two-dimensional tortuosity of the superficial femoral artery in early atherosclerosis. *J. Vasc. Res.* **30**, 181–191 (1993).
51. Adrian Galdran, André Anjos, José Dolz, Hadi Chakor, Hervé Lombaert, Ismail Ben Ayed. The Little W-Net That Could: State-of-the-Art Retinal Vessel Segmentation with Minimalistic Models. *arXiv* (2020) doi:The Little W-Net That Could: State-of-the-Art Retinal Vessel Segmentation with Minimalistic Models.
52. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
53. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
54. Lindsay, R. S. & Howard, B. V. Cardiovascular risk associated with the metabolic syndrome. *Curr. Diab. Rep.* **4**, 63–68 (2004).
55. Halbesma, N. *et al.* High protein intake associates with cardiovascular events but not with loss of renal function. *J. Am. Soc. Nephrol.* **20**, 1797–1804 (2009).
56. Burgess, S., Small, D. S. & Thompson, S. G. A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26**, 2333–2355 (2017).
57. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
58. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
59. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
60. International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
61. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
62. Krefl, D. & Bergmann, S. *PascalX v0.0.1*. (2021). doi:10.5281/zenodo.4429922.
63. Fonović, M. & Turk, B. Cysteine cathepsins and extracellular matrix degradation. *Biochim. Biophys. Acta* **1840**, 2560–2570 (2014).
64. Kuwabara, I. *et al.* Galectin-7 (PIG1) exhibits pro-apoptotic function through JNK activation and mitochondrial cytochrome c release. *J. Biol. Chem.* **277**, 3487–3497 (2002).
65. Shibuya, M. Vascular endothelial growth factor receptor-1 (VEGFR-1/Flt-1): a dual regulator for angiogenesis. *Angiogenesis* **9**, 225–30; discussion 231 (2006).
66. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
67. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
68. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
69. Tapp, R. J. *et al.* Associations of Retinal Microvascular Diameters and Tortuosity With Blood Pressure and Arterial Stiffness: United Kingdom Biobank. *Hypertension* **74**, 1383–1390 (2019).
70. Sasongko, M. B. *et al.* Retinal vascular tortuosity in persons with diabetes and diabetic retinopathy. *Diabetologia* **54**, 2409–2416 (2011).
71. Tursi, A., Brandimarte, G., Elisei, W., Inchingolo, C. D. & Aiello, F. Epithelial cell proliferation of the colonic

mucosa in different degrees of colonic diverticular disease. *J. Clin. Gastroenterol.* **40**, 306–311 (2006).
72. Ference, B. A. *et al.* Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur. Heart J.* **38**, 2459–2472 (2017).
73. Han, H.-C. Twisted blood vessels: symptoms, etiology and biomechanical mechanisms. *J. Vasc. Res.* **49**, 185–197 (2012).
74. Kwa, V. I. H. *et al.* Retinal arterial changes correlate with cerebral small-vessel disease. *Neurology* **59**, 1536–1540 (2002).
75. Welby, J. P. *et al.* Carotid Artery Tortuosity Is Associated with Connective Tissue Diseases. *AJNR Am. J. Neuroradiol.* **40**, 1738–1743 (2019).
76. Won, K.-J. *et al.* Monoclonal antibody against α-actinin 4 from human umbilical vein endothelial cells inhibits endothelium-dependent vasorelaxation. *J. Vasc. Res.* **50**, 210–220 (2013).
77. J.M.B. Sand, F. Genovese, N.S. Gudmann, M.A. Karsdal. Type IV collagen. in *Biochemistry of Collagens, Laminins and Elastin. 2nd edition* (ed. Karsdal, M. A.) (2019).
78. Zenteno, J. C. *et al.* Next generation sequencing uncovers a missense mutation in COL4A1 as the cause of familial retinal arteriolar tortuosity. *Graefes Arch. Clin. Exp. Ophthalmol.* **252**, 1789–1794 (2014).
79. Sutter, F. K. P. & Helbig, H. Familial retinal arteriolar tortuosity: a review. *Surv. Ophthalmol.* **48**, 245–255 (2003).
80. Lubarsky, B. & Krasnow, M. A. Tube morphogenesis: making and shaping biological tubes. *Cell* **112**, 19–28 (2003).
81. Hu, Y. *et al.* Dissecting the transcriptome landscape of the human fetal neural retina and retinal pigment epithelium by single-cell RNA-seq analysis. *PLoS Biol.* **17**, e3000365 (2019).
82. Yang, D. *et al.* Proliferation of vascular smooth muscle cells under inflammation is regulated by NF-κB p65/microRNA-17/RB pathway activation. *Int. J. Mol. Med.* **41**, 43–50 (2018).
83. Schell, C., Wanner, N. & Huber, T. B. Glomerular development--shaping the multi-cellular filtration unit. *Semin. Cell Dev. Biol.* **36**, 39–49 (2014).
84. Abdalla, M., Hunter, A. & Al-Diri, B. Quantifying retinal blood vessels' tortuosity — Review. *2015 Science and Information Conference (SAI)* (2015) doi:10.1109/sai.2015.7237216.
85. Lisowska, A., Annunziata, R., Loh, G. K., Karl, D. & Trucco, E. An experimental assessment of five indices of retinal vessel tortuosity with the RET-TORT public dataset. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2014**, 5414–5417 (2014).
86. Al-Diri, B. *et al.* REVIEW - a reference data set for retinal vessel profiles. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2008**, 2262–2265 (2008).
87. Chou, W.-C. *et al.* A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples. *Sci. Rep.* **6**, 39313 (2016).
88. Band, G. & Marchini, J. BGEN: a binary file format for imputed genotype and haplotype data. doi:10.1101/308296.
89. Pistis, G. *et al.* Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. *Eur. J. Hum. Genet.* **23**, 975–983 (2015).
90. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 166298 (2017) doi:10.1101/166298.
91. Pain, O., Dudbridge, F. & Ronald, A. Are your covariates under control? How normalization can re-introduce covariate effects. *Eur. J. Hum. Genet.* **26**, 1194–1201 (2018).
92. Myers, T. A., Chanock, S. J. & Machiela, M. J. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Frontiers in Genetics* vol. 11 (2020).
93. Kang, H. M. *EPACTS: efficient and parallelizable association container toolbox*. (2016).
94. Gogarten, S. M. *et al.* GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics* **28**, 3329–3331 (2012).
95. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
96. Kinsella, R. J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* **2011**, bar030 (2011).
97. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 8111 (2015).
98. Genome-wide Repository of Associations Between SNPs and Phenotypes. *National Institutes of Health (NIH)* https://grasp.nhlbi.nih.gov/.
99. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
100. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
101. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
102. Maguire, L. H. *et al.* Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. *Nat. Genet.* **50**, 1359–1365 (2018).
103. Osman, W., Low, S.-K., Takahashi, A., Kubo, M. & Nakamura, Y. A genome-wide association study in the Japanese population confirms 9p21 and 14q23 as susceptibility loci for primary open angle glaucoma. *Hum. Mol. Genet.* **21**, 2836–2842 (2012).
104. Sim, X. *et al.* Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet.* **7**, e1001363 (2011).
105. German, C. A., Sinsheimer, J. S., Klimentidis, Y. C., Zhou, H. & Zhou, J. J. Ordered multinomial regression

for genetic association analysis of ordinal phenotypes at Biobank scale. *Genet. Epidemiol.* **44**, 248–260 (2020).
106. Slager, S. L. *et al.* Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. *Blood* **120**, 843–846 (2012).
107. Tedja, M. S. *et al.* Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat. Genet.* **50**, 834–848 (2018).
108. Mehta, N. N. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Circulation. Cardiovascular genetics* vol. 4 327–329 (2011).
109. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
110. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
111. den Hoed, M. *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat. Genet.* **45**, 621–631 (2013).
112. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
113. Wain, L. V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* **3**, 769–781 (2015).
114. Khawaja, A. P. *et al.* Genome-wide analyses identify 68 new loci associated with intraocular pressure and improve risk prediction for primary open-angle glaucoma. *Nature Genetics* vol. 50 778–782 (2018).
115. Stapleton, C. P. *et al.* The impact of donor and recipient common clinical and genetic variation on estimated glomerular filtration rate in a European renal transplant population. *Am. J. Transplant* **19**, 2262–2273 (2019).
116. Warren, H. R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* **49**, 403–415 (2017).
117. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).
118. Wain, L. V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* **43**, 1005–1011 (2011).
119. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **51**, 51–62 (2019).
120. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677–687 (2009).
121. Craig, J. E. *et al.* Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. *Nat. Genet.* **52**, 160–166 (2020).
122. Seshadri, S. *et al.* Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med. Genet.* **8 Suppl 1**, S15 (2007).