Genetic Insights from Automated Lumen Diameter Measurements in Carotid Ultrasounds of the UK Biobank

Authors | Sofía Ortín Vela[1,2,&], Dennis Bontempi[1,2,&], Bianca Mazini[3], Leah Böttger[1,2], Olga Trofimova[1,2], Sven Bergmann[1,2,4*].

Affiliations | [1] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [2] Swiss Institute of Bioinformatics, Lausanne, Switzerland. [3] Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland. [4] Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa.

&: joint first authors (sofia.ortinvela@unil.ch and dennis.bontempi@unil.ch)

Sofía Ortín Vela

Dennis Bontempi

Bianca Mazini

Leah Böttger

Olga Trofimova

Sven Bergmann

***: Corresponding author |** Sven Bergmann, Ph.D., Department of Computational Biology; Genopode, office 2025.1 - CH1015 Lausanne, Switzerland.

E-mail: sven.bergmann@unil.ch

Abstract

Carotid ultrasound is routinely used in clinical practice for non-invasive vascular anatomical and functional assessment, such as measuring carotid Intima-Media Thickness (cIMT), an important marker for quantifying atherosclerotic burden in the common carotid arteries (CCAs). Recent research suggests that several risk factors associated with higher cIMT, such as high blood pressure, can induce a compensatory increase in the carotid Lumen Diameter (cLD) of the CCAs. However, the genetic architecture of cLD and its association with other cardiovascular traits are still poorly understood. To investigate these questions, we trained a Deep Learning model to segment the carotid artery from ultrasound images and developed an algorithm to measure the cLD. We compared multiple measures of cLD corresponding to lateral and central views of the left and right carotid arteries. By applying genome-wide association studies (GWAS), we investigated the genetic architecture of cLD and the relationship between cLD and cIMT in a cohort of 18 804 individuals imaged with carotid ultrasound from the UK Biobank. We found that cLD has an estimated heritability of 31%, substantially higher than that of cIMT (23%). Furthermore, these phenotypes only have a mild phenotypic (37%) but much higher genotypic (58%) correlation.

Introduction

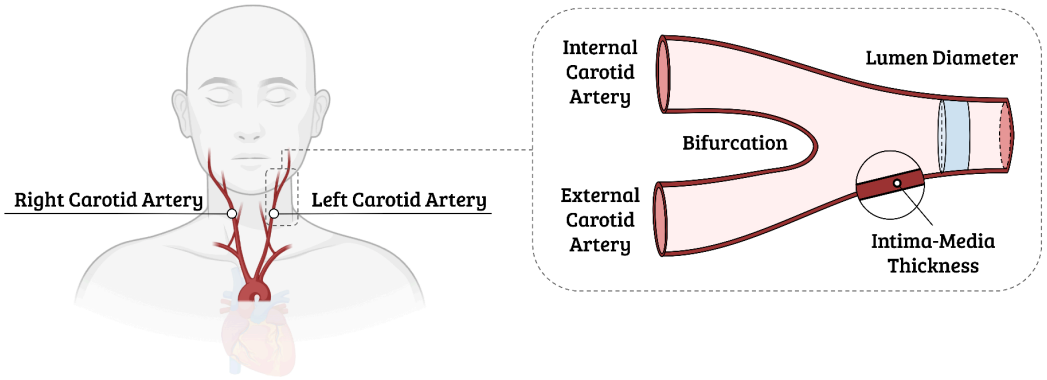
The right and left common carotid arteries (CCAs) supply the head, face, and neck with oxygenated blood (**Fig. 1a**). Atherosclerosis causes plaque to form within the carotid artery walls, narrowing the carotid lumen diameter (cLD), a condition called stenosis, which presents a serious risk for ischemic stroke [1]. According to Glagov's hypothesis, the CCAs adapt to atherosclerotic plaque buildup by expanding their luminal diameter. This serves as a compensatory mechanism to maintain blood flow despite structural changes and increased stiffness within the artery walls [2,3].

Carotid ultrasound imaging is a standard tool for evaluating cardiovascular health, offering a non-invasive approach to characterise carotid anatomy and function [4]. One of the primary measurements in clinical settings is the carotid intima-media thickness (cIMT), widely recognised as a significant marker for atherosclerotic burden [5,6]. In contrast, the assessment of cLD has not been implemented as a standard of care, also because its diagnostic utility has not been fully demonstrated, despite its known implication in compensatory vascular remodelling. Indeed, a recent study suggested that cLD is associated with all-cause mortality in the general population [7], and other studies [8,9] have explored the clinical significance of cLD for cardiovascular events. Two publications studied the genetic architecture of cLD: using genome-wide linkage analysis with data from 3 300 American Indian participants in the Strong Heart Family Study, Bella et al. [10] identified a locus influencing cLD on chromosome (Chr) 7 and suggested that cLD has a distinct genetic makeup from cIMT. Another study from Proust et al. [11], performed a GWAS involving a sample of 3 681 participants, on the right carotid diameter, finding a significant association between cLD (and a trend for the external diameter) and the single nucleotide polymorphisms (SNP) rs2903692 mapping to the *CLEC16A* gene on 16p13.

To investigate the genetic architecture of cLD and to evaluate its relationship with cIMT, we developed a fully automated analysis pipeline enabling high-throughput assessment of this feature. Specifically, we leveraged a deep convolutional neural network (CNN) trained on a dataset annotated by a clinical expert to segment a section of the carotid artery from ultrasound images and devised a method to measure the median cLD on such segmentations. Our tool allowed us to measure cLDs phenotypes of 18 808 participants of the UK Biobank (UKB), establishing the largest dataset to date. This enabled us to investigate the correlations between cLD measures from the left and right carotid, as well as their lateral and central view. We then explored the genetic architecture of these measures through genome-wide association analyses (GWAS) and subsequent post-processing analysis, pointing to overall mean cLD as the most robust measure. Comparing the latter

with mean cIMT revealed that these two ultrasound-derived phenotypes have a higher genetic than phenotypic correlation, with some distinct genetic associations, supporting the notion of complementary genetic mechanisms modulating these phenotypes. Building on the complementary nature of these two phenotypes, we propose the ratio cIMT/cLD as an additional, normalised phenotype to capture distinct aspects of the CCAs characteristics that may not be fully represented by the two phenotypes individually.

a



b

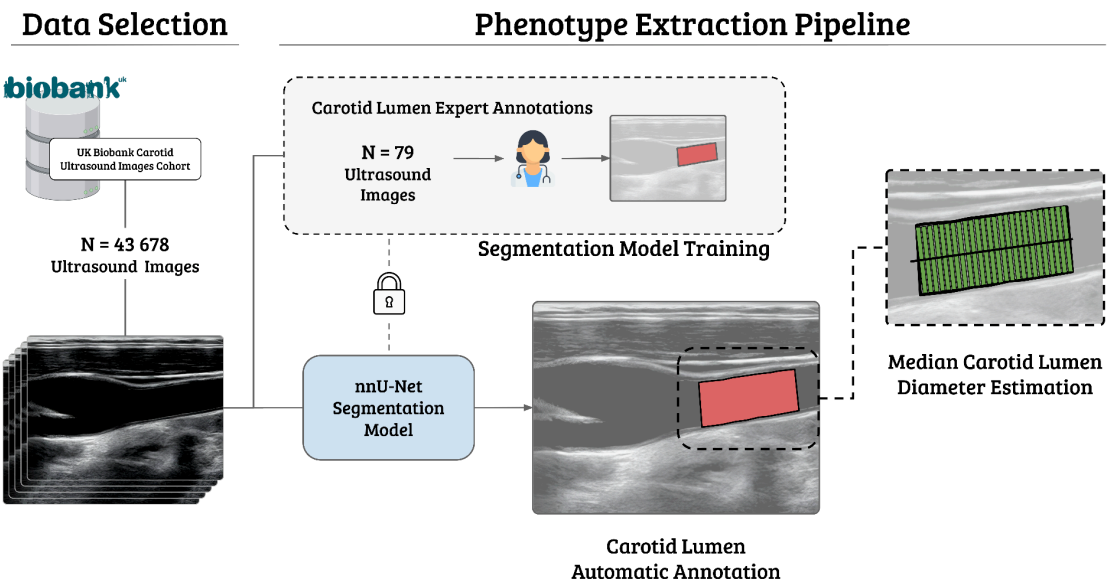


Figure 1 | Study Overview. a) Anatomy of the carotid artery. The left and right carotid arteries bifurcate to internal and external carotid arteries. The carotid intima-media thickness (cIMT) and the carotid lumen diameter (cLD) are measured before this bifurcation. **b)** We first trained a convolutional neural network (CNN) to segment the carotid from ultrasound still images automatically, using ground truth provided by a human annotator (N = 79 manual segmentations, shown in red). We then used this CNN to process 43 678 ultrasound images of the left and right carotid, including central and lateral views, from the UK Biobank. Each segmentation was post-processed for quality control (QC) and then underwent classical image processing to estimate the cLD for each image by computing the median of multiple diameter measurements (green) perpendicular to the central axis of the segmentation mask.

Results

The UKB dataset includes 21 838 left and 21 840 right carotid ultrasound DICOM image series, derived from 20 031 and 20 033 subjects, respectively. Each series includes several still images captured from ultrasound movies during diastole. These images include views of the central and lateral regions of the right and left CCAs from which we extracted the *right lateral*, *right central*, *left lateral*, and *left central cLD*. From these four primary cLD measurements, we also computed five additional derived cLD phenotypes, namely the *left*, *right*, *central*, *lateral*, and (overall) *mean cLD* by averaging across the relevant primary cLD measurements. Furthermore, we computed the *mean cIMT* and the *mean cIMT over mean cLD*. An overview of our study design is presented in **Fig. 1b** and detailed in the Methods section.

Correlations and Heritabilities of different Carotid Lumen Diameter measures

All nine cLD measures were first adjusted by regressing out the effects of common covariates (see Methods). We then computed pairwise phenotypic correlations between the corrected phenotypes (**Fig. 2a**, lower triangle). We observed high phenotypic correlations across the different cLD phenotypes ($r \in [0.69, 0.98]$), with central and lateral cLD phenotypes showing the strongest correlations. In contrast, weaker correlations were noted between the left and right cLDs ($r \in [0.69, 0.75]$), particularly when comparing lateral with central views ($r \in [0.70, 0.90]$) (Suppl. Data 1).

To investigate the genetic basis of cLD, we performed GWAS for each measure and analysed the resulting summary statistics using Linkage Disequilibrium Score Regression (LDSR) [12,13] to estimate cross-phenotype genetic correlations (**Fig. 2a**, upper triangle) and heritabilities (h^2; **Fig. 2b**). Genetic correlations followed similar patterns to their phenotypic counterparts but were generally higher ($r \in [0.92,1]$) (Suppl. Data 1). Manhattan plots summarising significant genetic loci are shown in **Suppl. Fig. 4**.

Heritability estimates were similar for the nine cLD measures (**Fig. 2b**), with mean cLD and lateral cLD showing the highest values ($h^2 = 0.31 \pm 0.06$). The smallest h^2 was observed for the left central cLD ($h^2 = 0.22 \pm 0.04$) (Suppl. Data 2).

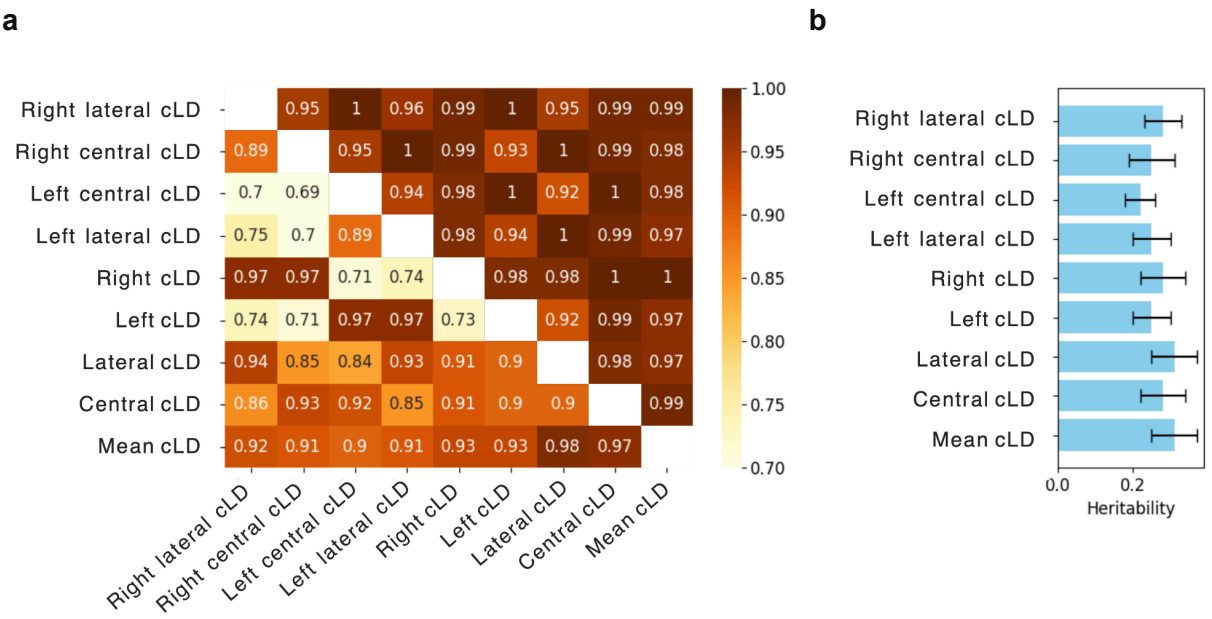


Figure 2 | Correlation and heritability analysis. a) Phenotypic (lower-left triangle) and genetic (upper-right triangle) correlations of cLDs. Phenotypes were corrected for covariates and z-scored before phenotypic correlation, while for the GWAS phenotypes were rank-normalised and regressed on genotypes and covariates (see Methods). **b)** Heritability (h^2) estimates. The corresponding phenotype h^2 were estimated using LDSR. The sample sizes are as follows: Right lateral cLD (N = 17 951), right central cLD (N = 17 839), left central cLD (N = 17 758), left lateral cLD (N = 17 680), right cLD (N = 18 634), left cLD (N = 18 584), lateral cLD (N = 18 686), central cLD (N = 18 689), and mean cLD (N = 18 808).

Genes and Pathways influencing different Carotid Lumen Diameter measures

To identify genes associated with each phenotype, we used our *PascalX* analysis tool [14,15]. The number of genes associated with cLD varied across phenotypes, ranging from 6 to 47 (**Fig. 3a** diagonal and Suppl. Data 3). Notably, the left central cLD showed the fewest associated genes, consistent with its lower h^2 estimate (**Fig. 2b**). Similarly, left lateral cLD and left cLD had fewer associated genes, consistent with their smaller h^2 values. While lateral cLD and mean cLD obtained both the highest h^2 values, lateral cLD had slightly more associated genes.

Five genes were significantly associated with all cLD phenotypes, *TAGLN, SIDT2, PAFAH1B2, PCSK7*, and *SIK3* (**Fig. 3b**). Other genes, such as *C8orf12*, and *RP11-10A14.5* were associated with all the cLD phenotypes except the left cLD and left central cLD (Suppl. Data 3).

Using PascalX, we identified annotated gene sets (i.e., “pathways”) enriched with high-scoring genes (**Fig. 3c, d**). Three phenotypes were associated with more than one pathway: Left (4), lateral (2), and mean (2) cLD. Notably, the gene set *‘chr8p23’* was associated with all phenotypes except the left central cLD (Suppl. Data 4).

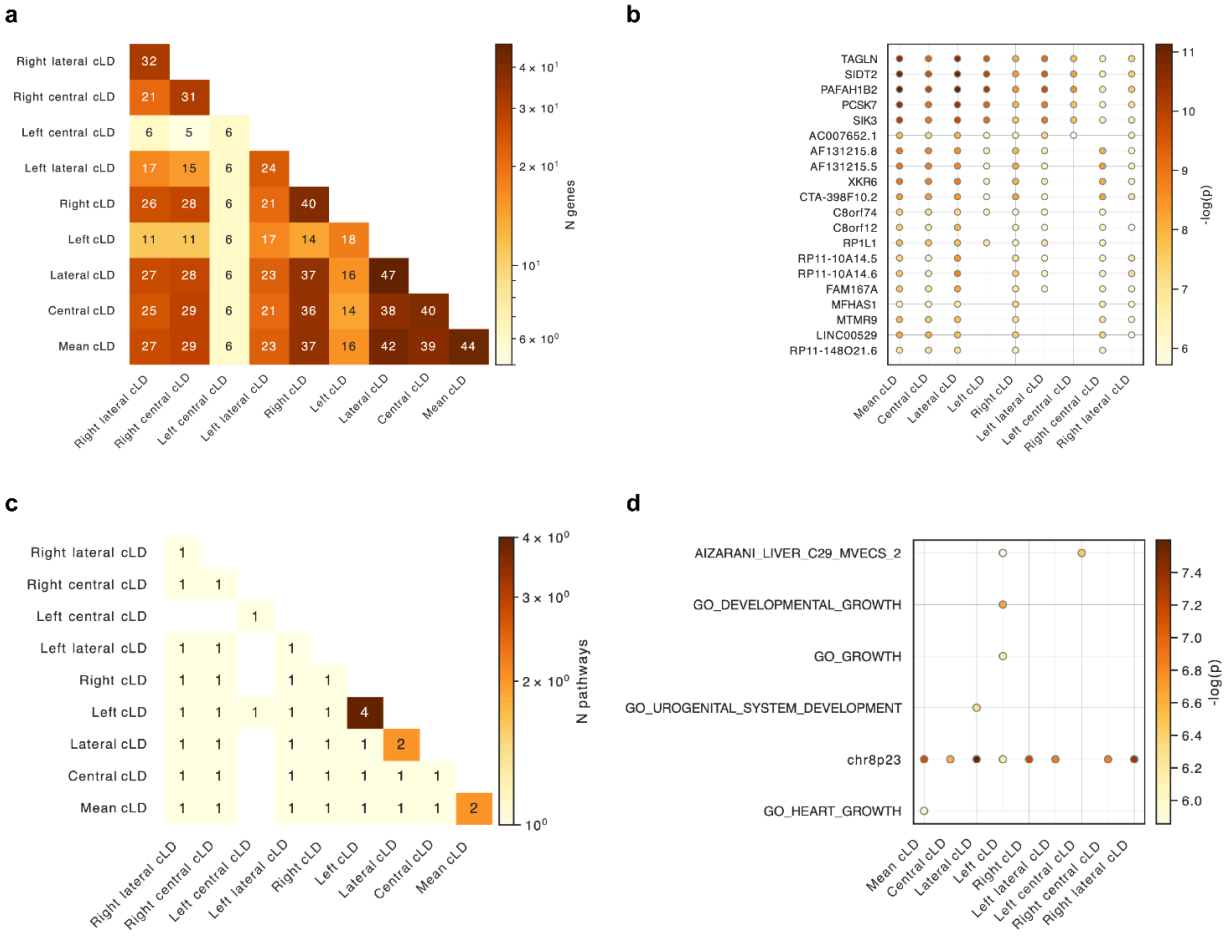


Figure 3 | Gene and pathways analyses. a) Gene-scoring intersection showing genes in common between cLD phenotypes. **b)** Name of the most frequent genes across the cLD phenotypes. **c)** Pathway-scoring intersection showing gene sets in common between cLD phenotypes. **d)** Name of the most frequent gene sets across the cLD phenotypes.

Comparison of Carotid Lumen Diameter and Carotid Intima-Media Thickness

To compare cLD phenotypes with cIMT, we focused on mean cLD as the representative for all cLD phenotypes due to its comprehensive nature, being the average of all cLD measurements, and its high heritability. Additionally, we also considered the ratio of mean cIMT over mean cLD as a composite phenotype presenting a "normalised" cIMT.

The covariate-adjusted phenotypic correlation between mean cLD and mean cIMT was moderate (0.37) (**Fig. 4a**). In contrast, the genetic correlation was substantially higher (0.58 ± 0.10). For the ratio cIMT/cLD, the phenotypic correlations were -0.23 with mean cLD and 0.81 with mean cIMT. The corresponding genetic correlations were -0.11 ± 0.11 and 0.74 ± 0.05, respectively (Suppl. Data 1).

Manhattan plots revealed heterogeneity in genetic signals among the phenotypes (**Fig. 4b**). Despite identical sample sizes, mean cLD exhibited stronger association signals than mean cIMT or cIMT/cLD. For mean cLD, the most significant SNPs were located at the start of Chr 7 (rs343029; p = 2.73E-26), with additional associations on other chromosomes, including Chr 8 (rs7838131; p = 3.20E-08, among others) and 11 (rs111677878; p = 8.62E-12). Mean cIMT showed its strongest associations at the start of Chr 7 (rs342988; p = 4.25E-10) and at the end of Chr 19 (rs1065853; p = 8.99E-11), while cIMT/cLD had key signals on Chr 7 (rs7792074; p = 5.21E-09, mid-region), 15 (rs625034; p = 8.16E-11) and 19 (rs111688353; p = 3.11E-08, rs1065853; p = 2.10E-10), among others. For more details, see **Suppl. Table 2**.

Heritability estimates confirmed mean cLD ($h^2 = 0.31 \pm 0.06$) as the most heritable, followed by mean cIMT (0.23 ± 0.04) and cIMT/cLD (0.14 ± 0.03) (**Table 1a**; Suppl. Data 2). Mean cLD was also associated with the highest number of genes (44), compared to mean cIMT (14) and their ratio (2). 11 genes were significantly associated with both mean cLD and mean cIMT, including *XKR6*, *LINC00529*, *MTMR9*, *FAM167A*, *C8orf12*, *LINC00208*, *BLK*, and *MFHAS1* (Suppl. Data 3). The *PascalX* cross-GWAS coherence test [15] resulted in more coherent than anti-coherent signals between the three phenotypes (**Table 1b**; Suppl. Data 3). In particular, cLD and cIMT shared 47 coherent genes, which included the 11 genes from the intersection of their individual gene scores and 36 other genes, such as *RP1L1*, *GMDS*, *ERI1*, *SGK223*, *MSRA*, *SOX7*, *GATA4*, and *TMEM170A.* Of note, 43 of the 47 coherent genes are located on 8p23.1. Anti-coherent signals were *ELN* and *RP11-731K22.1.* The ratio between cIMT and cLD shared coherent signals with cIMT (*LINC00670*) and cLD (8 genes, including *LINC00670*, *TMEM170A*, and *CBFA2T3*). The six anti-coherent signals included *ELN* and *PAFAH1B2*.

For cIMT, we identified a single gene set, *'chr8p23'*, while two gene sets were found for mean cLD, namely *'chr8p*23' and *'GO HEART GROWTH'* (**Table 1a**; Suppl. Data 4).

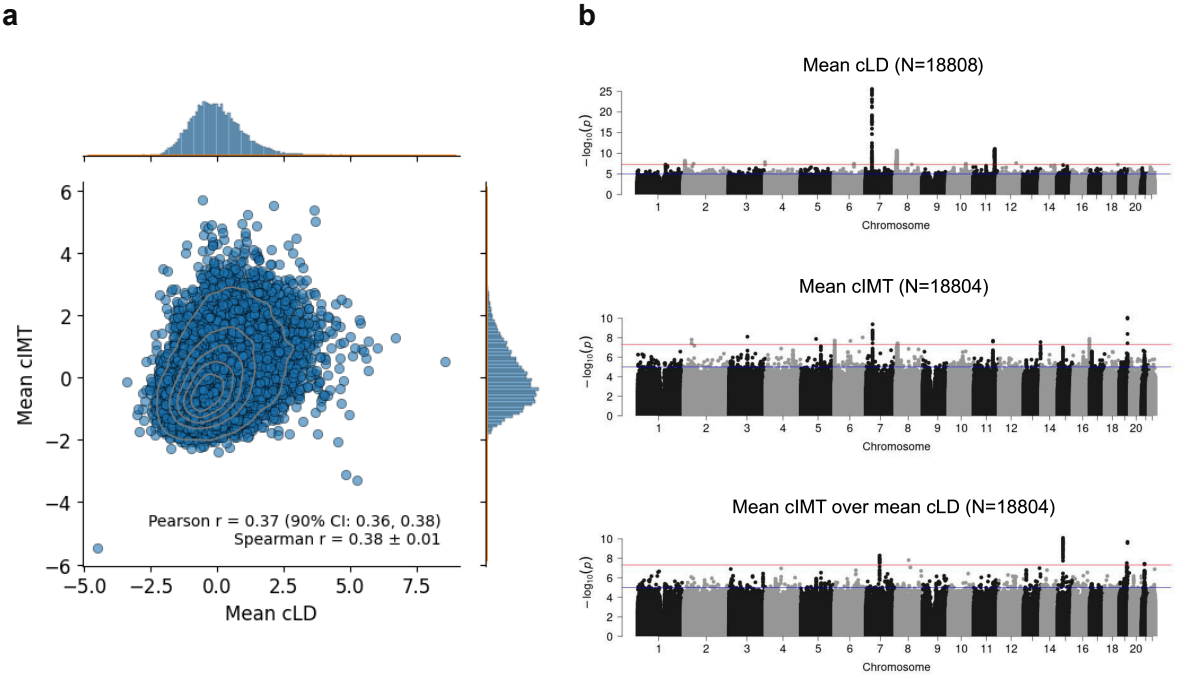


Figure 4 | Mean cLD, mean cIMT, and ratio comparison. **a)** Scatter-plot of *mean cIMT* against *mean cLD* after regressing out covariates effects and z-scoring. **b)** Manhattan plots of *mean cLD*, *mean cIMT* and *mean cIMT over mean cIMT*. The sample sizes for each analysis are reported in the title of each figure and are identical for the phenotypes shown in panel **a** (after filtering and subtracting covariates) and the GWAS summary statistics used for the genetic correlation analysis.

a

Phenotype	h^2 (std)	Number of genes	Number of pathways
Mean cLD	0.31 (0.06)	44	2
Mean cIMT	0.23 (0.04)	14	1
Mean cIMT over mean cLD	0.14 (0.03)	2	0

b

	Mean cLD	Mean cIMT	Mean cIMT over mean cLD
Mean cLD		47	8
Mean cIMT	2		1
Mean cIMT over mean cLD	2	4	

Table 1 | Genetic *mean cLD*, *mean cIMT*, and *ratio* comparison. a) The heritability estimate (h^2) was obtained using LDSR, and the number of genes and pathways was obtained using PascalX. **b)** Number of genes showing coherent (top right) or anti-coherent (bottom left) signals between pairs of phenotypes, obtained using PascalX cross-scoring.

Discussion

This study aimed to measure the carotid lumen diameter (cLD) in a large population with the following three goals: (i) elucidate the relationships between the left and right cLD, as well as the potential impact of the ultrasound orientation, (ii) compare cLD to the well-established carotid intima-media thickness (cIMT) phenotype, and (iii) unravel their common and distinct genetic architectures. By employing a convolutional neural network to segment carotid ultrasound images and incorporating image processing techniques, we automated the extraction of the cLD phenotypes, overcoming limitations in prior studies such as small sample sizes and labour-intensive methodologies. The resulting dataset, the largest of its kind, enabled a robust investigation of the genetic and phenotypic relationships of the cLD, revealing that the cLD provides complementary information to cIMT, with distinct genetic architectures and phenotypic patterns.

The results underline the distinct yet interconnected roles of the cLD and cIMT in vascular health, with a relatively high phenotypic correlation (37%). The even higher genetic correlation we observed (58%) suggests substantial but not entirely overlapping genetic influences. Furthermore, the lower heritability and weaker genetic signals of the ratio cIMT/cLD in comparison to the two original phenotypes may reflect a larger contribution of environmental factors to its variability.

This study significantly advances our understanding of cLD genetics, which, to the best of our knowledge, has only been explored in two previous studies. The linkage study by Bella et al. [10] found a significant SNP influencing cLD on Chr 7 at 120 centimorgans (cM), as well as a suggestive linkage on Chr 12 at 153 cM and 9 at 154 cM. Notably, while we found significant signals for rs343029 and rs11108966 on Chr 7 and 12, respectively, as plausible candidates for the respective linkage regions, we did not observe any significant associations on Chr 9. The GWAS by Proust et al. [11] found a marginally significant association (p = 4E-7) for the right internal carotid diameter (rs2903692, Chr 16) mapping to the gene *CLEC16A*, which we did not replicate for our right or mean cLD.

In our study, with a significantly larger sample size, we uncovered several novel loci for cLD, among those our strongest association, rs343029 (Chr 7). This SNP is close to the long non-coding RNA *AC007652.1* (i.e., ENSG00000235464) located between the protein-coding genes *TBX20,* associated with heart-related diseases [16–20], and *HERPUD2*, which has been linked with *TBX20* to electrocardiogram signals [21]. These genes were not detected using *PascalX*, because they are outside of its default 50kb window around rs343029. Nevertheless, a regulatory effect by *AC007652.1* seems plausible. Furthermore, this SNP has previously been associated with cIMT [22]. Other new loci uncovered are rs7838131

(Chr 8), which has been associated with body mass index [23], systolic blood pressure [23], hypertension [23], as well as coronary artery disease [24], and rs111677878 (Chr 11).

In contrast to cLD, the genetic architecture of cIMT has been extensively studied with high-powered GWAS [22,25]. Consistent with prior cIMT research [22,26], SNPs rs342988 (Chr 7) and rs1065853 (Chr 19) were also found to be significant in our analysis. These SNPs have further been associated with lipoprotein levels [27,28] and myocardial infarction [29], emphasising their importance for cardiovascular disease risk. Additionally, rs974819 (Chr 11), which was significantly associated with mean cIMT in this study, has previously been linked to coronary heart disease and was found to exhibit a sex-dependent effect [30]. To our knowledge, however, several loci identified in this study have not been reported previously in relation to cIMT. Importantly, one of the lead SNPs for cIMT (rs342988) on Chr 7 is in strong linkage disequilibrium with the nearby lead SNP for cLD (rs343029) with $R^2 = 0.77$ (see Suppl. Section Genetic Variants) [31].

For the cIMT/cLD), this is the first GWAS to explore its genetic basis, motivated by the potential biological significance of the ratio in reflecting the relationship between arterial wall thickening and cLD, which may inform mechanisms such as arterial distensibility and remodelling. SNP rs1065853 (Chr 19) was the only one significantly associated with the cIMT/cLD and mean cIMT, while all other significant hits were distinct. Among those were rs625034 (Chr 15,) which has previously been associated with thoracic aortic aneurysm [32]. These findings suggest that the cIMT/cLD captures aspects of vascular health that are not fully represented by either the cLD or cIMT alone.

Comparative analyses of Manhattan plots revealed notable differences, with mean cLD showing a stronger genetic signal compared to mean cIMT or cIMT/cLD. This highlights the distinct genetic architecture of the cLD compared to cIMT and emphasises its substantial signal strength, despite both phenotypes not being highly polygenic.

Our gene-wise analysis highlighted several genes associated with cLD, such as *PCSK7*, associated with blood lipids [33], *SIK3*, associated with high-density lipoprotein [34], *TAGLN*, associated with triglycerides [35], and *APOA1* associated with blood pressure [36] and lipoproteins [37]. Furthermore, genes in the *'chr8p23'* gene set, implicated in embryonic development [38–40], metabolism [41–43], and inflammation [44–46], shared coherent signals with both cLD and cIMT. These associations emphasise the metabolic and inflammatory pathways underlying vascular health. Interestingly, *ELN*, a gene involved in tissue elasticity and associated with heart diseases [47–49], exhibited an anti-coherent association between cLD and cIMT. This aligns with studies showing that reduced elastin

levels result in a narrowed arterial lumen and increased arterial stiffness in both humans and mice [47,48].

Despite its promise, the study has practical limitations that warrant consideration. First, while the cLD and cIMT were evaluated in a large, population-based cohort, the UKB is not fully representative of the general population, which may limit the generalisability of these findings. Additionally, the prognostic utility of the cLD needs further validation in diverse demographic and clinical populations to establish its role in routine cardiovascular screening.

In conclusion, this study explored the genetic architecture of the cLD and showed that this is distinct from the more commonly assessed phenotype of cIMT. By automating the extraction of the cLD from carotid ultrasound images and investigating its genetic determinants, we have established a foundation for incorporating the cLD as a routine measure in cardiovascular screenings. Our findings underline the importance of both cIMT and the cLD as complementary markers.

Methods

UK Biobank and Carotid Ultrasound Images

The UKB is a large-scale biomedical database and research resource containing anonymised genetic, lifestyle, and health information from half a million UK participants. The UKB's database, which includes blood samples, heart and brain scans, and genetic data of the volunteer participants, is globally accessible to approved researchers who are undertaking health-related research that is in public interest. UKB recruited 500 000 people between the ages of 40-69 years in 2006-2010 from across the UK. With their consent, they provided detailed information about their lifestyle, and physical measures and had blood, urine, and saliva samples collected and stored for future analysis. It includes multi-organ imaging for many participants, such as magnetic resonance image scans of the brain, heart, and liver, carotid ultrasounds, and retinal colour fundus images [50].

Carotid ultrasound data, available for around 20 000 participants, was collected to measure cIMT, a marker for subclinical atherosclerosis and cardiovascular disease risk. Images were acquired from both left and right carotid arteries using standardised protocols across all assessment centres. Images were taken at four angles (120°, 150°, 210°, and 240°) below and near to the carotid bifurcation. The angle of acquisition for each still image did not always align exactly with the target reference angles, often resulting in multiple images attempting to capture the same reference angle. For such cases, in our study, we averaged cLD to yield a single cLD value per reference angle per subject during each medical visit. Only some of the images, among all the ones captured for each subject, show the CCA and are correctly locked on the diastole, featuring a small bounding box marking the cIMT. For these images, maximum, mean, and minimum cIMT values are available [Carotid Ultrasound Documentation] as part of the UKB. To ensure uniformity in the measurements, we used these images for our analyses. It is important to note that cIMT values were available for more subjects than the carotid images themselves.

Lumen Diameter Segmentation

To automate the carotid artery segmentation from ultrasound images, we trained a Deep Learning CNN using N = 79 randomly sampled images labelled by an expert radiologist with more than six years of experience using itk-snap [51]. During the manual segmentation process, we only selected images showing clear interfaces for blood/intima and media/adventitia. We carefully segmented the anechoic portion of the CCAs, avoiding the inclusion of the intima, to ensure accurate delineation of the lumen. Additionally, to prevent overestimation of cLD, we systematically excluded the carotid bulb. Then, we cropped the

labelled ultrasound stills to the image area and converted them to grayscale. For the training, we used the PyTorch-based nnU-Net deep framework [52], which implements a heuristic that enables data-driven hyperparameters search. The nnU-Net framework has been shown in numerous studies to be very data-efficient as it only needs small training sample sizes to show substantial generalisability. Given the nature of our images, we limited the training to a 2D patch-based model only. We first trained five different 2D models using nnU-Net's default 5-fold cross-validation to assess the models' performance on previously unseen data. After observing a high enough Dice Coefficient on all the folds (Dice Coefficient > 0.95), we trained a model using all the human-labelled data. All the training runs minimised a composite loss of Dice Coefficient and Cross-Entropy and consisted of up to 1000 epochs. Only the best model checkpoint was saved.

Using this model, we segmented all the carotid images in the UKB dataset with an associated cIMT measurement (i.e., correctly locked on diastole, as explained in the previous section). We used an out-of-the-box optical character recognition model (PyTesseract) to identify such images.

Lumen Median Diameter Measurement Per Image

The QC began with evaluating the number of segmented objects in each image. If more than one object was identified, their areas were compared, and objects with an area smaller than half the maximum object area were excluded. If no object met this criterion, the image was discarded (**Suppl. Fig. 1a**). Next, the shape and regularity of the remaining segmented object were assessed. Segmented objects were required to resemble rectangles or squares and to have fewer than a defined number of contour points. This threshold was set at 215 based on an initial analysis of a subset of our dataset. While a perfect rectangle or square would have four points, minor pixel-level irregularities justified a higher threshold (**Suppl. Fig. 1b**).

After QC, the median cLD for each image in pixels was measured using the segmented cLD (described in the previous section), for that purpose, the centre of gravity of the segmentation was calculated using the central moments (**Equation 1**) along with the major axis angle (Θ) (**Equation 2**). To ensure consistent alignment of the major axis with the cLD, the angle was adjusted when $|\Theta|$ exceeded $\pi/4$ by applying the transformation $\Theta = |\Theta| - \pi/2$.

Transverse lines were then plotted across the length of the major axis at intervals of two pixels. For each line, the distance between the upper and lower segmented boundaries was measured (**Suppl. Fig. 2**; green lines in **Fig. 1b**). To mitigate irregularities, lines with

distances less than half the median line length were excluded. The median of the remaining line distances was calculated to represent the median cLD for the image (ϕ).

This approach assumes that defining the cLD linearly is a reasonable approximation for this dataset. If this assumption does not hold for other datasets, alternative methods incorporating higher-order approximations may be required.

$\{\bar{x}, \bar{y}\} = \left\{\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}\right\}$, with the moments defined as $M_{ij} = \sum_{x}\sum_{y} x^{i} y^{j} I(x, y)$ **(Eq. 1)**

and $I(x, y)$ being the image pixel intensities.

$\Theta = \frac{1}{2} arctan\left(\frac{2\mu'_{11}}{\mu'_{20} - \mu'_{02}}\right)$, with the central moments defined as

(Eq. 2)

$$\mu_{pq} = \sum_{m}^{p}\sum_{n}^{q} C_{m}^{p} C_{n}^{q} (-\bar{x})^{(p-m)} (-\bar{y})^{(q-n)} M_{mn}$$

Lumen Diameter Phenotypes Measurement Per Subject

After obtaining the median cLD (ϕ) for each image, cLD phenotypes were calculated for each participant. At least four carotid images were measured per participant during the same session, corresponding to four reference angles (α = 120º, 150º, 210º, and 240º). The first two represented the right carotid, and the latter two represented the left carotid (**Fig. 1a**).

In some cases, multiple images were taken for the same reference angle for a single participant. Analysis revealed that measurements for images targeting the same angle were nearly identical, with only minimal variations. Therefore, for participants with multiple images per angle, the median cLD values were averaged to generate a single value per angle.

Sometimes there were multiple instances per subject; however, we only included the first instance, since not many subjects would have been added by using different instances (**Suppl. Table 3**; **Suppl. Fig. 5**).

Lumen Diameter Phenotypes Definition

For each participant, a maximum of one median cLD (ϕ) measurement per reference angle was retained. From these, four initial cLD phenotypes were defined: median(ϕ) for α=120º, 150º, 210º, and 240º named as right lateral cLD, right central cLD, left central cLD, and left lateral cLD.

To incorporate combinations of individual angles, we computed additional phenotypes:

- Right and left cLD:
  - mean(median(ϕ for α=120º), median(ϕ for α=150º)),
  - mean(median(ϕ for α=210º), median(ϕ for α=240º)).
- Lateral and central cLD:
  - mean(median(ϕ for α=120º), median(ϕ for α=240º)),
  - mean(median(ϕ for α=150º), median(ϕ for α=210º)).
- Mean cLD: mean(median(ϕ for α=120º), median(ϕ for α=150º), median(ϕ for α=210º), median(ϕ for α=240º)).

We computed the mean for all available angles, ensuring that if one or more angle-specific cLD measurements were missing, the mean was calculated from the remaining available values. This approach allowed for the inclusion of as much data as possible while maintaining consistency across participants.

Additionally, carotid ultrasound images displayed data on the mean, minimum, and maximum cIMT. Using an optical character recognition model, we extracted this information from each image and applied the same preprocessing pipeline applied to the cLD measurements, we calculated the mean cIMT across the four reference angles. This resulted in a single measure of mean cIMT for each participant.

To examine the relationship between cIMT and cLD, we derived a normalised phenotype as the ratio of mean cIMT over mean cLD. The distributions of all phenotypes can be found in the **Suppl. Fig. 3**, and baseline data can be found in **Suppl. Table 1**.

The resulting dataset comprised nine cLD phenotypes per participant, along with the mean cIMT and the normalised phenotype.

Genome-wide association analysis (GWAS)

The GWAS for all phenotypes was performed using regenie [53]. Prior to the analysis, the genotype data underwent QC as recommended for UKB genotype data [54] using PLINK2

[55] (MAF = 0.01, MAC = 100, SNP genotype missingness = 0.1, individual genotype missingness = 0.1, HWE = 1E-15). Phenotypes were rank-inverse normal transformed prior to analysis. The covariates included in the GWAS were sex, age, age-squared, assessment centre, standing height, and the first 20 genetic principal components (PCs).

Genetic Correlations and SNP-Heritabilities

Summary statistics were used as input to compute the genetic correlations and h^2, which were computed using LDSR [12,13].

Genes and Pathways

Gene and pathway scores were computed using PascalX [14,15]. Both protein-coding genes and lincRNAs were scored using the approximate “saddle" method, taking into account all SNPs with a minor allele frequency > 0.05 within a 50 kb window around each gene. All pathways available in MSigDB v7.2 were scored using PascalX’s ranking mode, fusing and rescoring any co-occurring genes less than 100kb apart. PascalX requires linkage disequilibrium structure to accurately compute gene scores, which in our analyses was provided with the UK10K (hg19) reference panel. 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.

Data and Code Availability

GWAS summary statistics will be available on Zenodo after the peer-reviewed publication. Image-derived phenotypic data is under restricted access and will only be available through the UKB cohort platform (https://www.ukbiobank.ac.uk/) after peer-reviewed publication. The raw UKB data are protected and not open access; however, they can be obtained upon project creation and acceptance. The code will be available on GitHub after the peer-reviewed publication of the manuscript.

Acknowledgements

The authors are grateful to the study participants and the staff from the UKB.

Funding

This research was supported by the Swiss National Science Foundation grant no. CRSII5 209510 for the “VascX” Sinergia project.

Author contributions

SOV and SB conceptualised and designed the study. DB performed image preprocessing and trained the image segmentation pipeline. SOV performed phenotype extraction and quality control. SOV and LB performed GWAS analyses collaboratively. SOV estimated heritability, cross-phenotype correlations, and conducted gene and pathway analyses. OT analysed cross-phenotype gene associations using PascalX’s coherence scores. BM and SOV worked on segmenting the ground truth images. The manuscript was written by SOV, DB, and SB, with contributions from all other authors.

Competing Interests

All the authors declare no competing interests, including both financial and non-financial interests.

References

1. Howard DPJ, Gaziano L, Rothwell PM, Oxford Vascular Study. Risk of stroke in relation to degree of asymptomatic carotid stenosis: a population-based cohort study, systematic review, and meta-analysis. Lancet Neurol. 2021;20: 193–202.

2. Korshunov VA, Schwartz SM, Berk BC. Vascular Remodeling. Arteriosclerosis, Thrombosis, and Vascular Biology. 2007 [cited 16 Dec 2024]. doi:10.1161/ATVBAHA.106.129254

3. Watase H, Sun J, Hippe DS, Balu N, Li F, Zhao X, et al. Carotid artery remodeling is segment specific: An in vivo study by vessel wall magnetic resonance imaging: An in vivo study by vessel wall magnetic resonance imaging. Arterioscler Thromb Vasc Biol. 2018;38: 927–934.

4. Pignoli P, Tremoli E, Poli A, Oreste P, Paoletti R. Intimal plus medial thickness of the arterial wall: a direct measurement with ultrasound imaging. Circulation. 1986;74: 1399–1406.

5. Baldassarre D, Amato M, Bondioli A, Sirtori CR, Tremoli E. Carotid artery intima-media thickness measured by ultrasonography in normal clinical practice correlates well with atherosclerosis risk factors. Stroke. 2000;31: 2426–2430.

6. Simova’] [’iana. Intima-media thickness: appropriate evaluation and proper measurement. [cited 17 Dec 2024]. Available: https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-13/Intima-media-thickness-Appropriate-evaluation-and-proper-measurement-described

7. Fritze F, Groß S, Ittermann T, Völzke H, Felix SB, Schminke U, et al. Carotid Lumen Diameter Is Associated With All-Cause Mortality in the General Population. J Am Heart Assoc. 2020;9: e015630.

8. Eigenbrodt ML, Sukhija R, Rose KM, Tracy RE, Couper DJ, Evans GW, et al. Common carotid artery wall thickness and external diameter as predictors of prevalent and incident cardiac events in a large population study. Cardiovasc Ultrasound. 2007;5: 11.

9. Sedaghat S, van Sloten TT, Laurent S, London GM, Pannier B, Kavousi M, et al. Common carotid artery diameter and risk of cardiovascular events and mortality: Pooled analyses of four cohort studies: Pooled analyses of four cohort studies. Hypertension. 2018;72: 85–92.

10. Bella JN, Cole SA, Laston S, Almasy L, Comuzzie A, Lee ET, et al. Genome-wide linkage analysis of carotid artery lumen diameter: the strong heart family study. Int J Cardiol. 2013;168: 3902–3908.

11. Proust C, Empana J-P, Boutouyrie P, Alivon M, Challande P, Danchin N, et al. Contribution of Rare and Common Genetic Variants to Plasma Lipid Levels and Carotid Stiffness and Geometry: A Substudy of the Paris Prospective Study 3. Circ Cardiovasc Genet. 2015;8: 628–636.

12. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47: 291–295.

13. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47: 1236–1241.

14. 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. 2016;12: e1004714.

15. Krefl D, Bergmann S. Cross-GWAS coherence test at the gene and pathway level. PLoS Comput Biol. 2022;18: e1010517.

16. Luyckx I, Kumar AA, Reyniers E, Dekeyser E, Vanderstraeten K, Vandeweyer G, et al. Copy number variation analysis in bicuspid aortic valve-related aortopathy identifies TBX20 as a contributing gene. Eur J Hum Genet. 2019;27: 1033–1043.

17. Yang XF, Zhang YF, Zhao CF, Liu MM, Si JP, Fang YF, et al. Relationship between TBX20 gene polymorphism and congenital heart disease. Genet Mol Res. 2016;15. doi:10.4238/gmr.15027374

18. Zhou Y-M, Dai X-Y, Huang R-T, Xue S, Xu Y-J, Qiu X-B, et al. A novel TBX20 loss-of-function mutation contributes to adult-onset dilated cardiomyopathy or congenital atrial septal defect. Mol Med Rep. 2016;14: 3307–3314.

19. Amor-Salamanca A, Santana Rodríguez A, Rasoul H, Rodríguez-Palomares JF, Moldovan O, Hey TM, et al. Role of Truncating Variants in Dilated Cardiomyopathy and Left Ventricular Noncompaction. Circ Genom Precis Med. 2024;17: e004404.

20. Zhao C-M, Bing-Sun, Song H-M, Wang J, Xu W-J, Jiang J-F, et al. TBX20 loss-of-function mutation associated with familial dilated cardiomyopathy. Clin Chem Lab Med. 2016;54: 325–332.

21. Scholman KT, Meijborg VMF, Gálvez-Montón C, Lodder EM, Boukens BJ. From Genome-Wide Association Studies to Cardiac Electrophysiology: Through the Maze of Biological Complexity. Front Physiol. 2020;11: 557.

22. Yeung MW, Wang S, van de Vegte YJ, Borisov O, van Setten J, Snieder H, et al. Twenty-five novel loci for carotid intima-media thickness: A genome-wide association study in >45 000 individuals and meta-analysis of >100 000 individuals. Arterioscler Thromb Vasc Biol. 2022;42: 484–501.

23. Christakoudi S, Evangelou E, Riboli E, Tsilidis KK. GWAS of allometric body-shape indices in UK Biobank identifies loci suggesting associations with morphogenesis, organogenesis, adrenal cell renewal and cancer. Sci Rep. 2021;11: 10688.

24. Gill D, Georgakis MK, Zuber V, Karhunen V, Burgess S, Malik R, et al. Genetically predicted midlife blood pressure and coronary artery disease risk: Mendelian randomization analysis. J Am Heart Assoc. 2020;9: e016773.

25. Nikolajevic Starcevic J, Petrovic D. Carotid Intima Media-thickness and Genes Involved in Lipid Metabolism in Diabetic Patients using Statins – a Pathway Toward Personalized Medicine. 2013 [cited 24 Dec 2024]. Available: https://www.ingentaconnect.com/content/ben/chamc/2013/00000011/00000001/art00003

26. Strawbridge RJ, Ward J, Bailey MES, Cullen B, Ferguson A, Graham N, et al. Carotid intima-media thickness: Novel loci, sex-specific effects, and genetic correlations with obesity and glucometabolic traits in UK Biobank. Arterioscler Thromb Vasc Biol.

2020;40: 446–461.

27. Surakka I, Horikoshi M, Mägi R, Sarin A-P, Mahajan A, Lagou V, et al. The impact of low-frequency and rare variants on lipid levels. Nat Genet. 2015;47: 589–597.

28. Richardson TG, Leyden GM, Wang Q, Bell JA, Elsworth B, Davey Smith G, et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. PLoS Biol. 2022;20: e3001547.

29. Hartiala JA, Han Y, Jia Q, Hilser JR, Huang P, Gukasyan J, et al. Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. Eur Heart J. 2021;42: 919–933.

30. Zhou J, Huang Y, Huang RS, Wang F, Xu L, Le Y, et al. A case-control study provides evidence of association for a common SNP rs974819 in PDGFD to coronary heart disease and suggests a sex-dependent effect. Thromb Res. 2012;130: 602–606.

31. NCI, CBIIT, DCEG, Machiela. LDlink. [cited 22 Dec 2024]. Available: https://ldlink.nih.gov/?var1=rs343029&var2=rs7792074&pop=GBR&genome_build=grch37&tab=ldpair

32. Ashvetiya T, Fan SX, Chen Y-J, Williams CH, O'Connell JR, Perry JA, et al. Analysis of UK Biobank Cohort Reveals Novel Insights for Thoracic and Abdominal Aortic Aneurysms. Genomics. bioRxiv; 2021. Available: https://www.biorxiv.org/content/10.1101/2021.02.05.429911v1.full.pdf

33. GeneCards Human Gene Database. PCSK7 Gene - GeneCards. [cited 23 Dec 2024]. Available: https://www.genecards.org/cgi-bin/carddisp.pl?gene=PCSK7&keywords=PCSK7

34. GeneCards Human Gene Database. SIK3 Gene - GeneCards. [cited 23 Dec 2024]. Available: https://www.genecards.org/cgi-bin/carddisp.pl?gene=SIK3&keywords=SIK3

35. GeneCards Human Gene Database. TAGLN Gene - GeneCards. [cited 23 Dec 2024]. Available: https://www.genecards.org/cgi-bin/carddisp.pl?gene=TAGLN&keywords=TAGLN

36. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok P-Y, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nat Genet. 2017;49: 54–64.

37. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. Nat Genet. 2021;53: 185–194.

38. Simpson NH, Ceroni F, Reader RH, Covill LE, Knight JC, SLI Consortium, et al. Genome-wide analysis identifies a role for common copy number variants in specific language impairment. Eur J Hum Genet. 2015;23: 1370–1377.

39. SOX7 expression is critically required in FLK1-expressing cells for vasculogenesis and angiogenesis during mouse embryonic development. Mechanisms of Development. 2017;146: 31–41.

40. Afouda BA. Towards Understanding the Gene-Specific Roles of GATA Factors in Heart Development: Does GATA4 Lead the Way? International Journal of Molecular Sciences. 2022;23: 5255.

41. Manning AK, Goustin AS, Kleinbrink EL, Thepsuwan P, Cai J, Ju D, et al. A long

non-coding RNA, LOC157273, is an effector transcript at the chromosome 8p23.1-PPP1R3B metabolic traits and type 2 diabetes risk locus. Front Genet. 2020;11: 615.

42. The MTMR9 rs2293855 polymorphism is associated with glucose tolerance, insulin secretion, insulin sensitivity and increased risk of prediabetes. Gene. 2014;546: 150–155.

43. Secolin R, Gonsales MC, Rocha CS, Naslavsky M, De Marco L, Bicalho MAC, et al. Exploring a Region on Chromosome 8p23.1 Displaying Positive Selection Signals in Brazilian Admixed Populations: Additional Insights Into Predisposition to Obesity and Related Disorders. Front Genet. 2021;12: 636542.

44. Guo Y, Liu Q, Zheng Z, Qing M, Yao T, Wang B, et al. Genetic association of inflammatory marker GlycA with lung function and respiratory diseases. Nat Commun. 2024;15: 3751.

45. Zhong J, Shi Q-Q, Zhu M-M, Shen J, Wang H-H, Ma D, et al. MFHAS1 Is Associated with Sepsis and Stimulates TLR2/NF-κB Signaling Pathway Following Negative Regulation. PLOS ONE. 2015;10: e0143662.

46. Website. doi:10.1182/blood-2011-11-394072

47. Cocciolone AJ, Hawes JZ, Staiculescu MC, Johnson EO, Murshed M, Wagenseil JE. Elastin, arterial mechanics, and cardiovascular disease. American Journal of Physiology-Heart and Circulatory Physiology. 2018 [cited 30 Dec 2024]. doi:10.1152/ajpheart.00087.2018

48. Wahart A, Hocine T, Albrecht C, Henry A, Sarazin T, Martiny L, et al. Role of elastin peptides and elastin receptor complex in metabolic and cardiovascular diseases. The FEBS Journal. 2019;286: 2980–2993.

49. Francis CM, Futschik ME, Huang J, Bai W, Sargurupremraj M, Teumer A, et al. Genome-wide associations of aortic distensibility suggest causality for aortic aneurysms and brain white matter hyperintensities. Nature Communications. 2022;13: 1–18.

50. Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nat Commun. 2020;11: 2624.

51. 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability.

52. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18: 203–211.

53. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet. 2021;53: 1097–1103.

54. UKBB Analysis - regenie. [cited 23 Dec 2024]. Available: https://rgcgithub.github.io/regenie/recommendations/

55. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4: 7.