Genome-wide association study and multi-ancestry meta-analysis identify common variants associated with carotid artery intima-media thickness

Devendra Meena[1], Jian Huang[1], Marjan Zare[2], Natalie R. Hasbani[3], BOUA Palwendé Romuald[4], Rima Mustafa[1], Sander W van der Laan[5], Huichun Xu[6], James G. Terry[7], Joshua C. Bis[8], Deepti Jain[9], Nicholette D. Palmer[10], Nancy Heard-Costa[11,12], Yuan-I Min[13], Xiuqing Guo[14], Jie Yao[14], Kent D. Taylor[14], Jingyi Tan[14], Juan Peralta[15,16], Alexandre C. Pereira[17,18], Alyna Khan[9], Ananyo Choudhury[19], Anne B. Newman[20], Anny H. Xiang[21], Aroon Hingorani[22], Barry I. Freedman[23], Christopher J. O'Donnell[24], Claudia Giambartolomei [25], David M. Herrington[26], David R. Jacobs Jr[27], Derek Klarin[28,29], Fei Fei Wang[9], Gerardo Heiss[30], HarshaVardhan Doddapaneni[31], Howard N. Hodis[32], Jai Broome[9], James G. Wilson[33], Jean-Tristan Brandenburg[19], John Blangero[15,16], Jose E Krieger[18], Josh D. Smith[34,35], Karine A. Viaud-Martinez[36], Kathleen A. Ryan[6], Leslie A. Lange[37], May E. Montasser[6], Michael C. Mahaney[15,16], Michal Mokry[39], Myriam Fornage[40], Patricia Munroe[41,42], Richard A. Gibbs[31,43], Russell P. Tracy[44], Ryan W. Kim[45], Scott M. Damrauer[46], Stephen S. Rich[47], Willa A. Hsueh[48], Yii-Der Ida Chen[14], NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, The Million Veteran Program (MVP), TOPMed Atherosclerosis Working Group, Alanna C. Morrison[3], Braxton D. Mitchell[6,49], John Jeffrey Carr[7], Bruce M. Psaty[8,50,51], Donald W. Bowden[10], Ramachandran S. Vasan[12,52,53], Adolfo Correa[13,54], Wendy S. Post[55], Mark O. Goodarzi[56], Leslie J. Raffel[57], Joanne E. Curran[15,16], Michele Ramsay[19], Jerome I. Rotter[14], Paul Elliott[58,59,60], Nora Franceschini[61], Paul S. de Vries[3], Ioanna Tzoulaki[1,62,63], Abbas Dehghan[1]

[1]Department of Epidemiology and Biostatistics, Imperial College London School of Public Health, London, UK

[2]Maternal-fetal medicine Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

[3]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

[4]Clinical Research Unit of Nanoro, Institut de Recherche en Sciences de la Santé, CNRST, Burkina Faso

[5]Central Diagnostics Laboratory, Division Laboratories, Pharmacy, and Biomedical genetics, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

[6]Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

[7]Department of Radiology, Vanderbilt Translational and Clinical Cardiovascular Research Center, Vanderbilt University Medical Center, Nashville, TN, USA

[8]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA

[9]Genetic Analysis Center, Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, USA

[10]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA

[11]Boston University School of Medicine, Boston, MA, USA

[12]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA

[13]Jackson Heart Study, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA

[14]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA

[15]Department of Human Genetics, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA

[16]South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA

[17]Department of Genetics, Harvard Medical School, Boston, MA, USA

[18]Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of S√£o Paulo, S√£o Paulo, Brazil

[19]Division of Human Genetics at the National Health Laboratory Service, University of the Witwatersrand, Johannesburg, South Africa

[20]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[21]Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA USA

[22]Institute of Cardiovascular Science, Faculty of Population Health, University College London, London, United Kingdom

[23]Section on Nephrology, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[24]VA Boston Healthcare System, West Roxbury VA Medical Center, Cardiology Section, Boston, MA, USA

[25]Department of Pathology and Laboratory Medicine, University of California (UCLA), Los Angeles, Los Angeles, CA, USA

[26]Department of Internal Medicine, Section of Cardiovascular Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[27]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN, USA

[28]VA Palo Alto Healthcare System, Palo Alto, CA

[29]Department of Surgery, Stanford University Medical Center, Stanford, CA

[30]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

[31]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA

[32]Department of Preventive Medicine, Department of Medicine, Atherosclerosis Research Unit, Keck School of Medicine, University of Southern California, Los Angeles, CA USA

[33]Division of Cardiology, Beth Israel Deaconess Medical Center, Boston, MA, USA

[34]Department of Genome Sciences, University of Washington, Seattle, WA, USA

[35]Northwest Genomics Center, University of Washington, Seattle, WA, USA

[36]Illumina Laboratory Services, Illumina Inc., San Diego, CA, USA

[37]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA

[38]Central Diagnostics Laboratory, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands

[39]Laboratory of Experimental Cardiology, Department of Cardiology, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands

[40]Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA

[41]Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, Charterhouse Square, London, UK

[42]NIHR Barts Biomedical Research Centre, Queen Mary University of London, Charterhouse Square, London, UK

[43]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

[44]Department of Pathology and Laboratory Medicine, Robert Larner, M.D. College of Medicine, University of Vermont, Burlington, VT, USA

[45]Psomagen Inc., Rockville, MD, USA

[46]Division of Vascular Surgery, Hospital of the University of Pennsylvania, Philadelphia, PA 19104

[47]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA

[48]Department of Internal Medicine, The Ohio State University Wexner Medical Center, Columbus, OH USA

[49]Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA

[50]Department of Epidemiology, University of Washington, Seattle, WA, USA

[51]Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA

[52]Department of Medicine, Boston University School of Medicine, Boston, MA, USA

[53]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

[54]Department of Population Health Science, University of Mississippi Medical Center, Jackson, MS, USA

4

[55]Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[56]Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA USA

[57]Division of Genetic and Genomic Medicine, Department of Pediatrics, University of California, Irvine, CA USA

[58]MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, UK

[59]UK Dementia Research Institute, Imperial College London, London UK

[60]National Institute for Health Research Imperial College Biomedical Research Centre, Imperial College London, London, UK

[61]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

[62]British Heart Foundation Centre of Excellence, Imperial College London, London, UK

[63]Department of Hygiene and Epidemiology, University of Ioannina, Ioannina, Greece

Abstract

Carotid artery intima-media thickness (cIMT) is a measurement of subclinical atherosclerosis that predicts future cardiovascular events, including stroke and myocardial infarction. Genome-wide association studies (GWAS) have identified only a fraction of the genetic variants associated with cIMT. We performed the largest GWAS for cIMT involving up to 131,000 individuals. For the first time, we meta-analysed a diverse range of ancestries including populations with African, Asian (Chinese), Brazilian, European, and Hispanic ancestries. Our study identified 59 independent loci (53 loci from the multi-ancestry single variant analysis of which 19 are novel, $P<5\times10^{-8}$; 6 novel in gene-based analysis from single variant analysis, $P=2.6\times10^{-6}$, 2 novel in meta-regression) associated with cIMT. Gene-based, tissue-expression and gene-set enrichment analyses revealed novel genes of potential interest and highlighted significant relationships between vascular tissues (aorta, coronary and tibial arteries) and genetic associations. We found that circulatory levels of seven proteins, including ACAN, BCAM, DUT, ERI1, APOE, FN1, and GLRX were potentially causally associated with cIMT levels. We found a strong genome-wide correlation between cIMT and various cardiometabolic, smoking phenotypes, and lipid traits. Using Mendelian randomisation, our analyses provide robust evidence for causal associations between cIMT and several clinically relevant traits, including lipids, blood pressure, and waist circumference. Our study extends our genetic knowledge of atherosclerosis and highlights potential causal relations between risk factors, atherosclerosis and clinical diagnoses.

Introduction

Carotid artery intima-media thickness (cIMT), a measurement of subclinical atherosclerosis, is correlated with change in coronary artery disease assessed by quantitative coronary angiography[1] and is a strong predictor of future clinical cardiovascular events[2] that can be measured non-invasively by B-mode ultrasound imaging of the carotid arteries[3]. cIMT is significantly heritable with study-specific heritability estimates ranging from 35-60%, suggesting a substantial genetic component underlying individual differences in cIMT measurements[4-7].

To date, genome-wide association studies (GWAS) have identified at least 35 loci robustly associated with cIMT ($P<5\times10^{-8}$)[8-14]. The largest GWAS meta-analysis of cIMT to date (comprising up to 100,253 individuals of European ancestry) came from a meta-analysis combining data from the UK Biobank (UKBB), the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), and the University College London-Edinburgh-Bristol (UCLEB) consortia. However, most previous GWAS of cIMT focused on individuals of European ancestry only.

Here, we conduct the first multi-ancestry GWAS meta-analysis of cIMT that includes individuals of European, African, Hispanic, Asian (Chinese), and Brazilian populations from the UKBB, CHARGE-UCLEB consortium[9]; Trans-Omics for Precision Medicine (TOPMed) Program; the Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen), the Mexican-American Coronary Artery Disease (MACAD); the Hypertension-Insulin Resistance Family Study (HTN-IR); and the Baependi Heart Study (BHS).

Results

A European- and multi-ancestry GWAS meta-analysis of cIMT was conducted, aggregating data from seven consortia/studies (**Supplementary Fig. 1)** that identified 51 loci, applying

7

stringent quality control filters before meta-analysis. The genomic inflation factor showed little evidence for population stratification (λ_{GC} range 0.95–1.09) (**ST 1–2**).

European- and multi-ancestry GWAS meta-analysis

A European-only meta-analysis was first performed, combining UKBB (IMT mean maximum) with CHARGE-UCLEB cIMT GWAS summary statistics and identified 39 loci at $P<5\times10^{-8}$ (**ST 3**). A subsequent fixed-effects inverse variance-weighted multi-ancestry meta-analysis included non-overlapping samples from UKBB and CHARGE-UCLEB, TOPMed, AWI-Gen, MACAD, HTN-IR, and BHS (**Supplementary Figs. 2-3**), identifying an additional twelve loci, taking the total to 51, of which 17 were novel (**Fig. 1, ST 4**). A meta-regression was applied using MR-MEGA[15] to account for the heterogeneity in allelic effects between the ancestral groups, identifying two additional loci on chromosome 7 near *CPVL* (rs1522919, $P_{MR-MEGA}=2.5\times10^{-8}$) and chromosome 15 near *TMC3/STARD5 (*rs75789774, $P_{MR-MEGA}=3\times10^{-8}$*)*. We replicated all 35 known signals from published GWAS at a p-value of 0.05 (**ST4a**).

In a conditional analysis at the 51 loci, we found further independent associations at four loci including *SVIL* (rs2369339, rs9337951), *MYOCD* (rs6502187, rs7504018), *COL4A2* (rs9515203, rs4773141), and *APOE* (rs7412, rs429358) (**ST4**). We have generated regional plots (**Supplementary Figure 4**), forest plots for the most significant SNPs at each locus (**Supplementary Figure 5**) and Circos plots to show eQTL and chromatin interaction in the identified loci (**Supplementary Figure 6**).

Sex-specific GWAS in the UKBB

A sex-specific cIMT GWAS was conducted in the UKBB to detect sex-related genetic differences. The analysis identified 17 loci (11 for women and 6 for men; **Supplementary Fig 7, ST 5-6**), of which three were not detected in the main GWAS (**Supplementary Fig. 7, ST 7-8**). Two loci were significant only in women including rs4953207 (β=0.021, $P=4\times10^{-8}$, near *COX7A2L*, p for interaction = 0.03) and rs746610273 (β=0.027, $P=3.7\times10^{-8}$ *near CDK13*, p

8

for interaction = 0.01). The effect estimates and p-values from the UKBB sex-specific GWAS showed a moderate correlation (r=.78) (**Supplementary Fig 8**).

Functional annotation and gene mapping of cIMT loci

Functional annotation of genome-wide significant single nucleotide polymorphisms (SNPs) using FUMA showed that the vast majority (~84%) of these lies within intronic or intergenic regions (**ST 9**). The lead SNPs at *APOE* (rs7412) and *LTBP4* (rs34093919*)* on chromosome 19 were exonic, non-synonymous and had the highest observed probability of a deleterious protein effect (Combined Annotation Dependent Depletion (CADD) score >25), indicating a probable deleterious impact of these variants on gene function. The remaining 10 exonic non-synonymous SNPs at *FN1*, *MCPH1*, *PPP1CC*, *SERPINA1*, *CBFA2T3* (rs12930980 and rs12924185), *APOE* (rs28399653, rs28399654, rs429358) and *MKL1* were in high linkage disequilibrium (LD) with either a lead SNP or with one of the independent significant SNPs (r^2>0.6). Out of 246 genes that were mapped to independent significant SNPs or their proxies (r^2 >0.6), gene mapping prioritised 106 genes based on *genomic position*, 47 based on expression quantitative loci (*eQTLs*), and 176 based on *3D chromatin interaction* mapping (**Fig. 2d**, **ST 10**).

Statistical fine-mapping (Credible-set analysis)

Bayesian fine-mapping was performed using summary statistics from the European-only GWAS and multi-ancestry meta-analysis to identify potential causal variants at the identified loci. At 3 loci (*COL4A2*, *LTBP4, APOE*), a single variant was highlighted as the likely single causal SNP (**ST 11**). For 22 out of 39 loci, the multi-ancestry meta-analysis results yielded smaller 95% credible sets compared to the European-only (**ST 11a**, **Supplementary Fig. 9**), with the highest reduction in the number of variants at *ZBTB10*, *ZHX2*, and *CFDP1* loci.

Genome-wide gene-based association analysis

A genome-wide gene-based association study (GWGAS) was conducted on the summary statistics from a multi-ancestry meta-analysis using MAGMA in FUMA. This identified 66 protein-coding genes significantly associated with cIMT (Bonferroni-corrected P-value < 2.6×10^{-6} (0.05/19,220 genes)), sixty of which were also included in the 246 prioritised genes mapped to 51 loci in the single variant analysis (**Fig. 1c**, **ST12**). The most significant association was with *TOMM40* (P=6.1×10^{-16}) followed by *APOE* (P=5.5×10^{-15}) and *CFDP1* (P=3.1×10^{-14}).

Rare variant analysis for in TOPMed

Gene-based rare variant analyses on whole genome sequence data from 24,953 participants across nine TOPMed studies (five ancestry groups) using Efficient Variant-Set Mixed Model Association Test (SMMAT) revealed 56 significant genes (Supplementary document) with 10 genes showing nominal significance (P<0.05), and MSRA consistently meeting significance across all aggregation strategies (Supplementary document, **ST 12a**).

Gene-set enrichment and Tissue expression (gene-property) analysis

Gene-set enrichment and tissue expression analyses were conducted on 246 prioritised genes in sets of genes defined by MsigDB v7.0, WikiPathways, and the NHGRI-EBI GWAS catalog to link multi-ancestry GWAS findings to biological functions. The strongest enrichment was found for NHGRI-EBI GWAS catalog-reported genes for a general factor of neuroticism (P=5.1×10^{-18}) followed by cIMT (P=3.4×10^{-13}) (**Fig. 2c**, **ST13**).

Differentially expressed gene (DEG) analysis across 30 tissue types showed a significant upregulation of these genes in blood vessel tissues, including coronary and aorta arteries (P<0.03) (**Fig. 2b**, **ST13a**). Tissue expression analysis found that the genes identified by gene-based analysis were significantly enriched for expression in blood vessel tissues: tibial artery (P=1.3×10^{-3}); coronary artery (P=1.4×10^{-3}); and aorta artery (P=7.9×10^{-5}) (**Fig. 2a**, **ST13b**).

Genetic correlation and heritability estimate of cIMT-associated variants

The SNP-based heritability (h^2) of the European-only cIMT-associated variants was 0.1095 (se=0.001). A genome-wide genetic correlation (r_g) was estimated between the European-only IMT GWAS and 53 other complex trait/disease categories using LD score regression (http://ldsc.broadinstitute.org/ldhub/). After adjusting for multiple testing (P=0.00094 (0.05/53)), a significant genetic correlation was found with ten traits/diseases (**ST14**) six of which had rg values stronger than 0.2, including type 2 diabetes (T2D; r_g=0.29), waist circumference (WC; r_g=0.24), overweight (r_g=0.23), waist-to-hip ratio (r_g=0.22), obesity (r_g=0.22) and body mass index (BMI; r_g=0.21).

Association with clinical outcomes

Lookups of all independent SNPs (r^2<0.6, n=399) in publicly available GWAS summary statistics of various diseases and cardiovascular risk factors relevant to cIMT were performed. We found that 41 loci were related to at least one of the 15 traits and disease categories, including general factors of neuroticism, coronary artery disease (CAD), immune-, blood- and lipid-related traits (**Fig. 3a**, **ST15**). Blood pressure (BP; systolic blood pressure (SBP) and diastolic blood pressure (DBP)) had the largest number of overlapping genes with 28 genes associated with both BP and cIMT. We also identified 10 loci for CAD, 8 loci for carotid plaque, 4 for lipid levels and peripheral artery disease (PAD), 2 for ischemic stroke (AIS), 2 for any stroke (AS), 1 for arterial stiffness index (ASI), and 1 for Alzheimer's disease (AD) (**Figs. 3b-3c**).

Mendelian randomisation

A bi-directional MR analysis was conducted on cIMT and various clinically relevant traits and diseases (**ST16**). We found supporting evidence for a positive effect of low-density lipoprotein cholesterol (LDL; β_{IVW} =0.013, P=7.4×10^{-10}), BMI (β_{IVW}=0.018, P=8.2×10^{-6}), waist circumference (β_{IVW} =0.023, P=6.2×10^{-7}), SBP (β_{IVW} =0.002, P=9.1×10^{-41}), DBP (β_{IVW} =0.001, P=2.1×10^{-13}), pulse pressure (PP; β_{IVW} =0.0046, P=2.0×10^{-74}), and life-time smoking (β_{IVW}=0.026, P=4.1×10^{-4}) and an inverse effect of high-density lipoprotein cholesterol (HDL;

β_{IVW} =-0.007, P=0.001) on cIMT (**Fig. 4a**, **ST17**). MR-Egger did not suggest a directional pleiotropic effect.

Conversely, cIMT was significantly associated with SBP (β_{IVW} =20.4, P=3.2×10^{-22}), PP (β_{IVW} =20.8, P=5.4×10^{-21}), risk of any stroke (OR$_{IVW}$=7.9, P=2.3×10^{-4}), ischemic stroke (OR$_{IVW}$ =9.6, P=2.0×10^{-4}), and large artery stroke (LAS; OR$_{IVW}$=46.1, P=0.003) (**Fig. 4b**, **ST18**) with support from at least one robust method to rule out pleiotropic effects.

Given the bidirectional association observed between SBP and cIMT, a sensitivity analysis was performed excluding shared genetic instruments for cIMT and SBP, showing a consistent association between SBP and cIMT supported by WM, and MR-Egger (β_{IVW}=0.0020, P=1.1×10^{-40}) (**ST19**). The reverse direction showed a slight decrease in magnitude (β_{IVW}=15.1, P=1.9×10^{-14}), without supporting evidence by MR-Egger and potential evidence of horizontal pleiotropy (P=0.042) (**ST19**). The MR-Steiger estimates suggested a bidirectional relationship between cIMT and SBP with a stronger causal effect of SBP on cIMT (p=3.7×10^{-271}) compared to the effect of cIMT on SBP (p=1.5×10^{-138}) (**ST19a**).

We used data from the UK Biobank in a framework of Mendelian randomisation (MR) analysis to examine whether the protein products of the identified genes are causally related to cIMT. Of the 246 genes prioritised by FUMA, 39 genes were found to be coding for a protein that was measured by Olink assay and were carried forward for performing cis-MR. Our cis-MR revealed that genetically predicted ACAN, BCAM, DUT, and ERI1 levels showed a positive association while APOE, FN1, and GLRX showed a negative association with cIMT levels (**ST20; Fig. 4c**)

cIMT loci associate with cell-type specific plaque expression and morphology

We assessed the cell-type specific expression of the 54 nearest protein-coding genes at each of 51 loci in atherosclerotic plaques using the single-cell RNA sequencing data from the Athero-Express Biobank study[16] that included 35 patients undergoing carotid

endarterectomy[17]. Of note was the cell-type-specific expression of the arterial wall: *FN1*, *LRP1*, *COL4A2*, *FBN1*, *CDH13*, *LTBP4* in ACTA2+ smooth muscle cells, and *SOX7*, *JCAD*, *PDGFD*, *CDH13*, *LTBP4*, *COL4A2* in endothelial cells. We also observed inflammatory cell-specific expression of *TBD52*, *CBFA2T3*, and *GOSR2* in CD79A+ B-cells, *WDR1*, *PPP1CC*, *ISG20* in T-cells, *SERPINA1*, *APOE*, *LRP1* in macrophages, *SLC24A3* in mast-cells, and *PREX1* in natural killer cells (**Figure 5**). These data are also available through www.plaqview.com together with other plaque-derived scRNAseq datasets[18].

We tested the 51 lead variants for association with plaque histological markers in 1500 patients from the Athero-Express Biobank. Of note was the significant association of the *APOE* locus with plaque fat content, and positive correlation with collagen content, calcification, and intraplaque vessel density at the *CFDP1*, *FGGY*, and *MCPH1* loci, respectively (**ST21**).

Discussion

We performed a large-scale multi-ancestry GWAS meta-analysis of cIMT, aggregating data from seven studies or consortia, comprising more than ~130,000 participants. The meta-analysis identified 53 cIMT-associated loci for maximum mean cIMT, of which 19 were novel. An additional two loci were identified when we accounted for heterogeneity in allelic effects between ancestral groups. A sex-specific GWAS using the UK Biobank, identified 17 loci, 11 for women and 6 for men, with 14 overlapping with sex-combined GWAS. We highlighted 4 potentially causal genes that are highly expressed in the arterial wall of smooth muscle, endothelial cells, and inflammatory cells, and show evidence for locus-specific effects on plaque morphology. We found that circulatory levels of protein products of seven identified genes, including ACAN, BCAM, DUT, ERI1, APOE, FN1, and GLRX were potentially causally associated with cIMT levels. We found a considerable genetic correlation between anthropometric measures, glycaemic homeostasis, and CAD, however less with lipids

13

(except for HDL-cholesterol). Mendelian randomisation analyses showed significant associations with pulse pressure, risk of any stroke, ischemic stroke, and large artery stroke as well as a strong and potentially bidirectional causal relationship between SBP and cIMT.

Our study is a large multi-ancestry GWAS meta-analysis for cIMT performed in individuals from five diverse populations including, African, Asian (Chinese), Brazilian, European, and Hispanic. Our findings further support the efficiency of conducting multi-ancestry GWAS. First, by using a Bayesian fine-mapping approach in European-only and multi-ancestry GWAS, we showed that multi-ancestry meta-analysis resulted in considerable refinement of causal variants (due to less extensive linkage disequilibrium) at 22 loci associated with cIMT. Second, the identification of 12 additional loci associated with cIMT from the multi-ancestry meta-analysis emphasises the importance of aggregating GWAS results from diverse populations to increase sample size and statistical power which is in line with previous research[19].

Observational research has linked modifiable cardiovascular risk factors with cIMT but whether the links are causal is unclear. Here for the first time, using a Mendelian randomisation framework, we present evidence for the causal role of anthropometric measures, blood pressure, smoking and dyslipidaemia in subclinical atherosclerosis. We provide strong evidence of a causative effect of higher waist circumference on higher cIMT suggesting that central fat distribution (abdominal obesity) is a causal risk factor for cIMT. Waist circumference is associated with subclinical atherosclerosis[20], and it may influence changes in cIMT, possibly via its effect on insulin resistance and glucose levels[21,22]. Our MR results are consistent with the results from epidemiological studies which showed that HDL-C is inversely associated with cIMT[23], while increased levels of total cholesterol and LDL-C are positively associated with cIMT.

In bidirectional two-sample MR analyses, we found causal associations between SBP and cIMT, and vice versa. These associations remained robust in sensitivity analyses as well as

14

after excluding genetic variants associated with both cIMT and SBP. Studies have suggested that elevated blood pressure results in increased cIMT possibly by the vascular damage caused by local shearing force in this condition[24]. On the other hand, increased cIMT may result in increased blood pressure by a possible mechanism leading to vascular endothelial dysfunction[24] and arterial stiffness.

Several studies have linked cardiovascular diseases with AD, possibly due to shared risk factors between these two diseases, especially APOE which is a known gene for cIMT and AD. Moreover, there is growing evidence to suggest that vascular dysfunction plays a central role in the development of AD. Thus, we did a lookup of the cIMT-associated variants in AD GWAS and investigated the potential causal role of cIMT in AD using Mendelian randomisation analysis. Although previous research has repeatedly implicated vascular factors induced by cardiovascular diseases in AD pathology[25], our MR analysis did not find any causal association between cIMT and AD, indicating that factors such as atherosclerosis might not be materially involved in the pathophysiology of AD. *APOE* is a pleiotropic genetic locus and could be linked to cardiovascular diseases and AD through different pathways.

Cross-sectional studies have shown higher cIMT in men compared to women; however, the contribution of genetic factors to these differences is largely unknown. We ran a sex-stratified analysis and identified two novel loci (*COX7A2L* and *CDK13*) that were genome-wide significant in women but not in men. *COX7A2L*, an estrogen-responsive gene, regulates energy metabolism, its deficiency has been shown to lower blood glucose levels[26], and it has also been associated with thrombosis (P =2 x10^{-7})[27]. *De novo* mutations in *CDK13* have been associated with congenital heart defects[28]. Our sex-specific GWAS had a fairly balanced gender breakdown (48% men versus 52% women), similar to that of previous UKBB cIMT GWAS, so it is less likely that the lack of association in men is due to a lack of statistical power. Previous studies have suggested that the sex difference in cIMT could be explained by differences in cardiovascular risk factors, including obesity and blood pressure[29].

However, we found two loci that were only significant in women, and we showed that the genetic effects at the *APOE* locus are stronger in men (β= 0.030) compared to women (β= 0.018). This might indicate that the sex differences in cIMT levels are not fully driven by risk factor distribution but also have roots in genes.

The strength of our study is the large sample size and multi-ancestry meta-analysis, which has enabled us to significantly enhance the number of known loci. We further searched for the role of rare variants in the genetic architecture of cIMT using whole-genome sequence data from TOPMed and applied gene-based methods. We applied various methods to study the functional characteristics of the identified genetic loci. Finally, we used various methods including genetic correlation, locus lookups, colocalization, and Mendelian randomisation to study the link between cIMT and cardiovascular risk factors and clinical outcomes. Several limitations should be mentioned. First, this study has collected all available studies in the discovery panel to maximise the statistical power and therefore findings are not replicated in independent samples. Second, the cIMT measurement protocols were slightly different across participating studies, however, any systematic differences in calibration should have been taken care of after variable transformations. Moreover, the findings are not driven by a single study. Third, the majority of the samples analysed in the current study were of European ancestry, which highlights the need for further genome-wide scans in populations other than those of European ancestry.

In conclusion, we report the largest multi-ancestry GWAS meta-analysis of cIMT. These findings extend our understanding of the genetic architecture of cIMT. We have comprehensively studied the link between cardiovascular risk factors, cIMT, and complex disorders. We report that systolic blood pressure and cIMT could have a bidirectional synergic effect on each other.

Methods

16

Study population

UKBB sample: The UK Biobank (UKBB), a large-scale prospective epidemiological cohort, recruited ~500,000 individuals aged 40-69 years from across the United Kingdom between 2006 and 2010. In 2014, UKBB began a pilot phase for imaging modalities relevant to cardiovascular research, including ultrasound of the carotid arteries in 100,000 participants[30]. Carotid intima-media thickness (cIMT) phenotyping was performed at one of the UKBB imaging centres. A detailed description of carotid ultrasound protocol (UKBB Resource 511, https://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=511) is provided elsewhere[31]. After performing quality control on phenotypic data, 42,449 cIMT measurements were available for further analysis. We use natural log-transformed values of the mean of the maximum (IMTmean-max) and mean of the mean (IMT mean) cIMT for normality. A detailed description of genotyping measures, imputation and central QC performed by the UKBB can be found elsewhere[32]. The Genome-Wide Association Study (GWAS) of cIMT was performed using a linear mixed non-infinitesimal model as implemented in BOLT-LMM[33], adjusted for age, sex, genotyping centre and first four principal components. Strict quality control filters were applied on genotypic as well as genome-wide imputed data (**Supplementary Tables 1-2**).

CHARGE-UCLEB: We obtained summary statistics from a previous GWAS meta-analysis of cIMT reported by the CHARGE-UCLEB consortia[9]. This GWAS meta-analysis analysed ~9.5 million non-sex-linked SNPs in up to 71,128 individuals of European ancestry from 31 studies [8,9]. Each participating study excluded SNPs and samples based on standard quality measures (**ST 1-2**). cIMT was defined by the mean of the maximum of several common carotid artery measurements, measured at the far wall or the near wall.

TOPMed: Whole genome sequencing (WGS) data from the Trans-Omics for Precision Medicine (TOPMed) program was used, involving 24,953 individuals from 9 different studies and 4 different ancestry groups. cIMT was measured using carotid ultrasound measurements with two primary measures of IMT were presented: 1) IMT_{mean}, and 2) $IMT_{mean-max}$. Phenotype

17

harmonization was performed by the TOPMed Data Coordinating Centre. A GWAS of cIMT was performed using the software package 'genesis'[34] adjusting for age, sex, study and 11 principal components (PCs) (**ST 1-2**).

AWI-Gen: AWI-Gen, a population-based cross-sectional study, included populations from six sub-Saharan African sites in four countries (Burkina Faso, Ghana, Kenya and South Africa)[35]. Over 12,000 sub-Saharan Africans, aged from 40 to 60 years, were enrolled from rural and urban settings from 2012 to 2016 [35,36]. Ultrasound scans were performed to assess the right and left cIMT (two measurements: $IMT_{mean-max}$ and IMT_{mean}) using dual B-mode ultrasound images of the left and right common carotid arteries. Samples were genotyped using the H3Africa genotyping array on ~2.3 million SNPs. A GWAS of cIMT was performed using fastGWA[37] software, adjusting for age, sex, study site and first four PCs (**ST 1-2**).

MACAD and HTN-INR: The MACAD and HTN-IR studies were designed to examine the genetic basis of coronary artery disease and insulin resistance in Mexican-Americans using a family-based design. cIMT was measured using standardized procedures and technology developed specifically for longitudinal measurement of atherosclerosis[38] (Patents 2005, 2006, 2011). The coefficient of variation of repeated cIMT measurements is typically <3% and often approaches 1%. Samples with call rates >0.98, SNPs with call rates >0.99, MAF >0.001, and HWE ($P<1\times10^{-6}$) passed laboratory quality control [39]. Pedigrees were examined for consistency of stated family structure. Software package 'Rvtests'[40] was used for cIMT genetic association analysis adjusting for age, sex and admixture proportions (**ST 1-2**).

BHS: The Baependi Heart Study is an epidemiological study in Baependi, a rural city in Minas Gerais State, Brazil. We used a cross-sectional analysis of data collected at the second evaluation visit (from 2010 to 2015) on subjects that underwent carotid ultrasonography. Individuals with angina, infarction, cardiac insufficiency, and revascularization were removed from our analysis. Imputation was performed using the Haplotype Reference Consortium Michigan Imputation Server using the TOPMed reference haplotype panel as a reference.

After imputation markers were kept if $R^2 > 0.3$, MAF>0.01 and HWE p-value $> 1 \times 10^{-20}$. The cIMT GWAS was performed using FaST-LMM [41] adjusting for age, sex and the first four PCs (**ST 1-2**).

Pre-meta-analysis quality control (QC) and Statistical analyses: Pre-meta-analysis QC on UKBB and CHARGE-UCLEB, TOPMed, AWI-Gen, HTN-IR, MACAD, and BHS summary-level data was performed using the R software package 'EasyQC'[42]. We used Haplotype Reference Consortium (HRC)[43] for UKBB and the 1000 Genomes phase 1 version 3 for CHARGE-UCLEB. We excluded markers based on missing alleles, allele frequency, effect size estimates and their associated standard errors. After performing central QC on all participating cohorts, ~16M autosomal markers were available for meta-analysis. Meta-analysis of summary statistics was performed using the inverse-variance fixed-effects method as implemented in the software METAL[44]. To account for heterogeneous loci, we also implemented a meta-regression approach using MR-MEGA v0.2[15] (https://genomics.ut.ee/en/tools). Principal components generated using MR-MEGA are shown in Supplementary Figure 10.

Functional annotation of cIMT-associated variants: We used the web-based tool Functional Mapping and Annotation (FUMA) v1.3.6[45] for annotating the cIMT-associated variants identified in the multi-ancestry meta-analysis. Independent significant SNPs with P $<5 \times 10^{-8}$ in the meta-analysis, and in moderate linkage disequilibrium (LD) ($r^2 < 0.6$) within a 500 kb window were used to define a locus. These 'independent significant SNPs' were further narrowed down to 'lead SNPs' having $r^2 < 0.1$ with each other [45]. The most likely causal genes associated with cIMT were identified using positional mapping, eQTL mapping, and chromatin interaction mapping, prioritising 246 genes. Tissue expression analysis employed the results originating from the gene-based genome-wide association analysis using MAGMA.

Gene-based analysis: The gene-based genome-wide association analysis was performed using MAGMA v1.07[46] as implemented in FUMA. This analysis used GWAS-identified SNPs and mapped them to protein-coding genes. A principal components regression model computed gene-level p-value (F-test)[46] and multiple testing was addressed using Bonferroni correction ($\alpha = 2.6 \times 10^{-6} = 0.05/19220$).

Gene-set and tissue expression analysis**:** A gene-set analysis was performed using MAGMA's competitive testing as implemented in FUMA to identify key biological pathways. This analysis tests whether genes in a gene-set are more strongly associated with a phenotype of interest than other genes. A gene-property analysis was also applied, testing for the relationship between highly expressed genes in specific tissue types and genetic associations.

Conditional analysis using COJO:

A stepwise model selection procedure was performed using the COndition and JOint analysis tool implemented in the Genome-wide Complex Trait Analysis software[47] (COJO-GCTA; cojo-slct) to select independently associated SNPs. For each locus where we observed multiple independent signals, the lead SNP was considered the first independent signal and a conditional analysis was performed by selecting SNPs within 500 kb of the lead SNP, using the 1000 genomes data as reference panel (N=503). Conditional analysis included SNPs with MAF > 0.01 and a p-value $<= 5 \times 10^{-8}$.

Statistical fine-mapping: We performed Bayesian fine-mapping by constructing 95% credible sets for each locus identified in the European-only GWAS and compared it with the multi-ethnic meta-analysis. Index SNPs were selected for all 39 loci identified in the European-only GWAS and defined regions for fine-mapping by extracting all SNPs within a 50 kb upstream/downstream window around each index SNP. For each SNP within a given region, Bayes factor (BF) was calculated based on GWAS p-values and the posterior probability that the SNP is causal (SNP's BF / summation of the BFs in the region)[48]. We

20

generated 95% credible sets by summing the posterior probabilities until the cumulative value exceeded 95% of the total cumulative posterior probability for all SNPs in the region.

Estimating heritability and genetic correlation with other traits: Genome-wide genetic correlations between cIMT GWAS (European-only) summary statistics and various cardiometabolic, anthropometric, and glycemic traits and diseases were evaluated to investigate the shared genetic architecture. Genetic correlation analysis was performed through the web-platform LDHub[49]. Statistical significance for the genetic correlation analysis was defined at P=0.00094 (0.05/53) after correction for the number of evaluated associations. We used linkage disequilibrium score regression to estimate the SNP-based heritability. The summary level data from CHARGE-UCLEB and UK Biobank was used for this analysis, restricted to HapMap3 SNPs as these are well-imputed in most studies.

Two-sample univariable Mendelian randomisation: We performed two-sample Mendelian randomisation (MR) using 'TwoSampleMR' package to investigate both the causal role of various risk factors on cIMT and the causal role of cIMT on various outcomes. For MR analysis, we used the genetic association for cIMT from the European-only meta-analysis in this study. The details of risk factors and outcomes are available in Supplementary Table 21, including sources of genetic association and sample size of the corresponding GWAS. For each MR analysis, we selected SNPs associated with the exposure of interest at P-value $<5\times10^{-8}$, with an F-statistic >10, and MAF > 5%. The genetic associations of the same SNPs with the outcome of interest were obtained from the corresponding outcome GWAS (**ST18**). The independent ($r^2<0.001$) harmonised SNPs were used as genetic instruments for the MR analysis.

SNP-specific effects for an exposure-outcome pair were estimated using the Wald ratio. We used the multiplicative random-effects inverse-variance weighted (IVW) method to pool the SNP-specific estimates for the MR effect estimates. We also used two sensitivity methods, namely weighted median, and MR-Egger regression[50]. In all analyses, we excluded potential

outlier SNPs identified by MR-PRESSO[51]. Specifically for the bidirectional analysis for cIMT and SBP, we performed a sensitivity analysis excluding common SNPs associated with both traits. We excluded 43,514 common SNPs and their proxies ($r^2 \geq 0.001$). We also used MR-Steiger[52] to investigate the direction of causality between cIMT and SBP.

We accounted for multiple comparisons using Bonferroni correction (α = 0.05/13 (N risk factors and outcomes) =0.004). For cis-MR on circulatory proteins, we used independent (r2=0.001) protein quantitative trait loci (pQTL) (p-value < 5×10^{-8}) from Sun et al 2023[53] within +/-500kb of the coding gene.

Single-cell expression in carotid plaque. The Athero-Express Biobank Study (www.atheroexpress.nl) in Utrecht, the Netherlands collected atherosclerotic plaques during surgery blood and plaques from patients undergoing carotid endarterectomy and stored at -80°C [16,54]. The plaques were analysed to quantitatively score the number of macrophages (CD68) and smooth muscle cells (SMCs, α-actin), as a percentage of the microscope field area by computerized analysis. Vessel density (CD34) was assessed as the average number per 3 hotspots. Intraplaque haemorrhage (IPH), and intraplaque fat, were scored using hematoxylin and eosin staining (HE). Calcification (using HE), and collagen (picrosirius red) were scored binary [55].

DNA was isolated for genetic analysis of these plaque characteristics [56]. The study was conducted in three separate experiments: Athero-Express Genomics Study 1 (AEGS1) including 891 patients, the Athero-Express Genomics Study 2 (AEGS2) including 954 patients, and the Athero-Express Genomics Study 3 (AEGS3) including 658 patients. All experiments were carried out according to OECD standards and we adhered to community standard quality control and assurance (QCA) procedures of the genotype data as described before[56,57]. After QCA these comprise 890 samples and 407,712 SNPs in AEGS1, 869 samples and 534,508 SNPs in AEGS2, and 649,954 samples and 534,508 SNPs in AEGS3 remained. Missing genotypes were imputed with 1000G phase 3, version 5 and HRC release

1.1 as a reference using the Michigan Imputation Server[58]. These results were further integrated using QCTOOL v2, where HRC imputed variants are given precedence over 1000G phase 3 imputed variants.

The association of 51 independent loci with plaque characteristics was analysed adjusted for age, sex and principal components. Continuous variables were inverse-rank normal transformed. Codes and raw results from these analyses are found here: https://github.com/CirculatoryHealth/CHARGE_CIMT.

For single-cell RNA sequencing, we collected carotid plaques from 35 individuals, isolated and processed RNA as described before[59], while employing the CEL-seq2 method[60].

Contributors

D.M., J.G.T., J.C.B., N.D.P., N.H.C., Y.I.M., G.H., D.R.J.J., M.F., A.B.N., J.B., A.K., F.F.W., B.I.F., L.A.L., J.G.W., D.M.H., S.S.R., W.A.H., A.H.X., H.N.H., M.C.M., J.B., M.R., B.P.R., A.C.M., B.D.M., J.J.C., B.M.P., D.W.B., R.S.V., A.C., J.I.R., W.S.P., M.O.G., L.J.R., A.C.P., J.E.C., P.S.V., A.D., I.T. contributed to the phenotype acquisition or harmonization including study design or funding of contributing studies. D.M., D.J., N.H.C., K.D.T., J.G.W., S.S.R., Y.D.I.C., M.C.M., J.B., J.D.S., R.A.G., H.V.D., K.A.V.M., M.R., B.P.R., B.D.M., A.C., J.I.R., A.C.P., J.E.C., A.D., I.T. contributed to acquisition of genotyping data or quality control. D.M., J.H., N.R.H., D.J., X.G., J.Y., J.T., M.R., B.P.R., A.C.P., P.S.V., A.D., I.T., S.W.L. contributed to statistical analysis of the data. D.M., J.H., D.J., M.R., B.P.R., P.S.V., A.D., I.T., S.W.L. contributed to the drafting of the manuscript. All authors contributed to the critical revision of the manuscript.

Declaration of interest: C.J.O. is a full-time employee at Novartis Institute of Biomedical Research. S.W.L. has received Roche funding for unrelated work. R.W.K. is an employee at

Correspondence

Dr Abbas Dehghan, MD, PhD, Department of Epidemiology and Biostatistics, Imperial College London School of Public Health, London, UK. Email: a.dehghan@imperial.ac.uk

Data availability

The summary statistics of the cIMT European-only and multi-ancestry GWAS will be made available for download upon publication via the database of Genotypes and Phenotypes (dbGaP) and/or GWAS catalog (https://www.ebi.ac.uk/gwas/).

Code availability

The following software packages were used for data analysis: R (https://www.r-project.org) version 3.6, EasyQC version 9.2, Plink1.9 (https://www.cog-genomics.org/plink/1.9); Plink2 (https://www.cog-genomics.org/plink/2.0); BOLT-LMM version v2.3.6, METAL (version

2011-03-25), FUMA version v1.3.6a (https://fuma.ctglab.nl), MAGMA version v1.08 (https://ctg.cncr.nl/software/magma), and LDHub (http://ldsc.broadinstitute.org/ldhub/), MR-MEGA v0.215 (https://genomics.ut.ee/en/tools).

Acknowledgements

References

1. Mack, W.J., LaBree, L., Liu, C., Selzer, R.H. & Hodis, H.N. Correlations between measures of atherosclerosis change using carotid ultrasonography and coronary angiography. *Atherosclerosis* **150**, 371-9 (2000).
2. Hodis, H.N. *et al.* The role of carotid arterial intima-media thickness in predicting clinical coronary events. *Ann Intern Med* **128**, 262-9 (1998).
3. Blankenhorn, D.H. & Hodis, H.N. George Lyman Duff Memorial Lecture. Arterial imaging and atherosclerosis reversal. *Arterioscler Thromb* **14**, 177-92 (1994).
4. Fox, C.S. *et al.* Genetic and environmental contributions to atherosclerosis phenotypes in men and women: heritability of carotid intima-media thickness in the Framingham Heart Study. *Stroke* **34**, 397-401 (2003).
5. Manolio, T.A., Boerwinkle, E., O'Donnell, C.J. & Wilson, A.F. Genetics of ultrasonographic carotid atherosclerosis. *Arterioscler Thromb Vasc Biol* **24**, 1567-77 (2004).
6. Sacco, R.L. *et al.* Heritability and linkage analysis for carotid intima-media thickness: the family study of stroke risk and carotid atherosclerosis. *Stroke* **40**, 2307-12 (2009).
7. Wang, D. *et al.* A genome-wide scan for carotid artery intima-media thickness: the Mexican-American Coronary Artery Disease family study. *Stroke* **36**, 540-5 (2005).
8. Bis, J.C. *et al.* Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet* **43**, 940-7 (2011).
9. Franceschini, N. *et al.* GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat Commun* **9**, 5141 (2018).
10. Strawbridge, R.J. *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* **40**, 446-461 (2020).
11. Natarajan, P. *et al.* Multiethnic Exome-Wide Association Study of Subclinical Atherosclerosis. *Circ Cardiovasc Genet* **9**, 511-520 (2016).
12. Gertow, K. *et al.* Identification of the BCAR1-CFDP1-TMEM170A locus as a determinant of carotid intima-media thickness and coronary artery disease risk. *Circ Cardiovasc Genet* **5**, 656-65 (2012).
13. Wai Yeung, M. *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*, ATVBAHA121317007 (2021).
14. Boua, P.R. *et al.* Genetic associations with carotid intima-media thickness link to atherosclerosis with sex-specific effects in sub-Saharan Africans. *Nat Commun* **13**, 855 (2022).
15. Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639-3650 (2017).
16. Verhoeven, B.A. *et al.* Athero-express: differential atherosclerotic plaque expression of mRNA and protein in relation to cardiovascular events and patient characteristics. Rationale and design. *Eur J Epidemiol* **19**, 1127-33 (2004).
17. Slenders, L. *et al.* Intersecting single-cell transcriptomics and genome-wide association studies identifies crucial cell populations and candidate genes for atherosclerosis. *Eur Heart J Open* **2**, oeab043 (2022).

18. Ma, W.F. *et al.* Enhanced single-cell RNA-seq workflow reveals coronary artery disease cellular cross-talk and candidate drug targets. *Atherosclerosis* **340**, 12-22 (2022).

19. Peterson, R.E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589-603 (2019).

20. Al-Domi, H. & Al-Shorman, A. Increased waist circumference is associated with subclinical atherosclerosis in schoolchildren. *Diabetes Metab Syndr* **13**, 264-269 (2019).

21. Xiang, A.H. *et al.* Effect of thiazolidinedione treatment on progression of subclinical atherosclerosis in premenopausal women at high risk for type 2 diabetes. *J Clin Endocrinol Metab* **90**, 1986-91 (2005).

22. Xiang, A.H. *et al.* Effect of pioglitazone on progression of subclinical atherosclerosis in non-diabetic premenopausal Hispanic women with prior gestational diabetes. *Atherosclerosis* **199**, 207-14 (2008).

23. Colombo, G.I. *et al.* The Association between HDL-C and Subclinical Atherosclerosis Depends on CETP Plasma Concentration: Insights from the IMPROVE Study. *Biomedicines* **9**(2021).

24. Zhang, L. *et al.* The association between carotid intima-media thickness and new-onset hypertension in a Chinese community-based population. *BMC Cardiovasc Disord* **19**, 269 (2019).

25. de Bruijn, R.F. & Ikram, M.A. Cardiovascular risk factors and future risk of Alzheimer's disease. *BMC Med* **12**, 130 (2014).

26. Shiba, S. *et al.* Deficiency of COX7RP, a mitochondrial supercomplex assembly promoting factor, lowers blood glucose level in mice. *Sci Rep* **7**, 7606 (2017).

27. Hinds, D.A. *et al.* Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. *Hum Mol Genet* **25**, 1867-74 (2016).

28. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet* **48**, 1060-5 (2016).

29. Juonala, M. *et al.* Effect of age and sex on carotid intima-media thickness, elasticity and brachial endothelial function in healthy adults: the cardiovascular risk in Young Finns Study. *Eur Heart J* **29**, 1198-206 (2008).

30. Littlejohns, T.J., Sudlow, C., Allen, N.E. & Collins, R. UK Biobank: opportunities for cardiovascular research. *Eur Heart J* **40**, 1158-1166 (2019).

31. Coffey, S. *et al.* Protocol and quality assurance for carotid imaging in 100,000 participants of UK Biobank: development and assessment. *Eur J Prev Cardiol* **24**, 1799-1806 (2017).

32. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

33. Loh, P.R., Kichaev, G., Gazal, S., Schoech, A.P. & Price, A.L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906-908 (2018).

34. Gogarten, S.M. *et al.* Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* **35**, 5346-5348 (2019).

35. Ramsay, M. *et al.* H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob Health Epidemiol Genom* **1**, e20 (2016).

36. Ali, S.A. *et al.* Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Glob Health Action* **11**, 1507133 (2018).

37.   Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet* **51**, 1749-1755 (2019).

38.   Hodis, H.N. *et al.* Vascular Effects of Early versus Late Postmenopausal Treatment with Estradiol. *N Engl J Med* **374**, 1221-31 (2016).

39.   Palmer, N.D. *et al.* Genetic Variants Associated With Quantitative Glucose Homeostasis Traits Translate to Type 2 Diabetes in Mexican Americans: The GUARDIAN (Genetics Underlying Diabetes in Hispanics) Consortium. *Diabetes* **64**, 1853-66 (2015).

40.   Zhan, X., Hu, Y., Li, B., Abecasis, G.R. & Liu, D.J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423-6 (2016).

41.   Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat Methods* **8**, 833-5 (2011).

42.   Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).

43.   McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).

44.   Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).

45.   Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).

46.   de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).

47.   Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).

48.   van Rooij, F.J.A. *et al.* Genome-wide Trans-ethnic Meta-analysis Identifies Seven Genetic Loci Influencing Erythrocyte Traits and a Role for RBPMS in Erythropoiesis. *Am J Hum Genet* **100**, 51-63 (2017).

49.   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).

50.   Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-25 (2015).

51.   Verbanck, M., Chen, C.Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* **50**, 693-698 (2018).

52.   Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13**, e1007081 (2017).

53.   Sun, B.B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329-338 (2023).

54.   van Lammeren, G.W. *et al.* Atherosclerotic plaque vulnerability as an explanation for the increased risk of stroke in elderly undergoing carotid artery stenting. *Stroke* **42**, 2550-5 (2011).

55.   Hellings, W.E. *et al.* Intraobserver and interobserver variability and spatial differences in histologic examination of carotid endarterectomy specimens. *J Vasc Surg* **46**, 1147-54 (2007).

56.   van der Laan, S.W. *et al.* Genetic Susceptibility Loci for Cardiovascular Disease and Their Impact on Atherosclerotic Plaques. *Circ Genom Precis Med* **11**, e002115 (2018).
57.   Laurie, C.C. *et al.* Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* **34**, 591-602 (2010).
58.   Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
59.   Depuydt, M.A.C. *et al.* Microanatomy of the Human Atherosclerotic Plaque by Single-Cell Transcriptomics. *Circ Res* **127**, 1437-1455 (2020).
60.   Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* **17**, 77 (2016).

Figure legends

Figure 1. **Manhattan and Fuji plots of common variants associated with cIMT (a)** Fuji plot showing independent loci associated with cIMT. The colour-coded circles (from inside to outside) represent GWAS of 1) IMTmean-max GWAS in women in UKBB; 2) IMTmean-max GWAS in men in UKBB; 3) IMTmean in UKBB; 4) IMT mean-max in UKBB and 5) IMTmean-max multi-ancestry GWAS meta-analysis. Every dot represents a locus associated with cIMT; large dots represent cross-trait loci, while small dots represent trait-specific loci. Genes with an asterisk denote novel cIMT loci; **(b)** Manhattan plot for the multi-ancestry GWAS on cIMT showing the negative $\log_{10}$ transformed P value for each SNP on the y axis and the base-pair position of the SNPs on each chromosome on the x-axis. The genome-wide-significance threshold ($P<5\times10^{-8}$) is represented as the horizontal red line. The 51 cIMT loci are annotated; red and black dots refer to novel and known loci, respectively; and **(c)** Gene-based analysis: Manhattan plot showing the genes associated with cIMT in the gene-based analysis. The y-axis shows the negative $\log_{10}$ transformed P value of the gene-based test computed in MAGMA, and the x-axis shows the genomic position on each chromosome. The red line indicates the Bonferroni-corrected threshold for genome-wide significance ($P=2.6x10^{-6}$ (0.05/19,220 genes). The novel signals from the gene-based test are highlighted in red.

Figure 2. Functional annotation of cIMT-associated variants a) MAGMA tissue expression analysis using gene expression per tissue based on GTEx version 8 data for 53 specific tissue types. The horizontal dotted line indicates the Bonferroni-corrected significance threshold (P = 0.05/53; $-\log_{10}$ transformed). Significant tissues are shown in red. **(b)** Differential expression gene analysis of prioritized genes across GTEx version 8 30 general tissue types. Genes with P-value ≤ 0.05 after Bonferroni correction and absolute log fold change ≥ 0.58 were defined as differentially expressed genes in a given tissue compared to others. Significantly enriched DEG (Bonferroni corrected P-value ≤ 0.05) are highlighted in red. **(c)** Gene-set and pathway

31

enrichment analysis; and **(d)** Venn diagram showing overlap of genes implicated by positional mapping, chromatin interactions, GWGAS, and eQTL mapping. GWGAS, genome-wide gene-based association study; eQTL, Expression quantitative trait loci.

Figure 3. Association of the identified cIMT SNPs with clinical traits and diseases (a) Look up in GWAS catalog **(b) Look up in published GWAS on cardiovascular** diseases, and **(c) Look up in published GWAS on cardiovascular** traits. PAD, peripheral arterial disease; AD, Alzheimer's disease; CAD, coronary artery disease; Plaque, carotid plaque; ASI, arterial stiffness index; BP, blood pressure.

Figure 4. **Forest plot of the IVW, WM and MR-Egger estimates. (a)** Effect of cardiovascular risk factors on cIMT, **(b)** Effect of cIMT on cardiovascular traits and diseases and c) Forest plot for the effect estimates of selected plasma proteins on cIMT. *Represents proteins with Wald ratio (causal estimate obtained for a single genetic variant). MR, Mendelian randomization; IVW, inverse-variance weighted; WM, weighted median.

Figure 5. Cell-type specific expression in carotid plaques of candidate genes. The y-axis shows cell types as previously described [16]. The size of the dot shows the percentage of cells expressing the respective genes, and the colour indicates the average total expression in that cell-types.
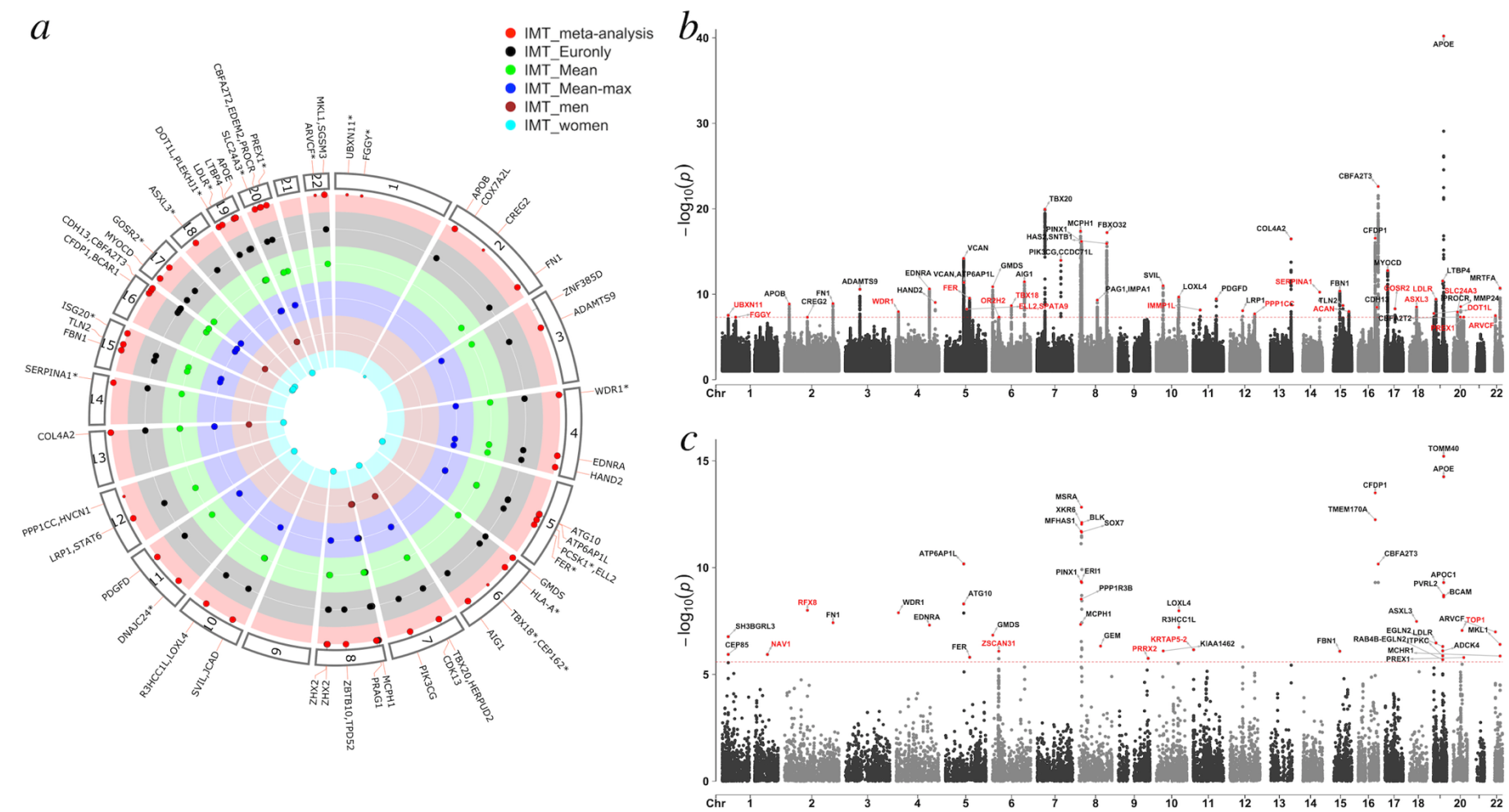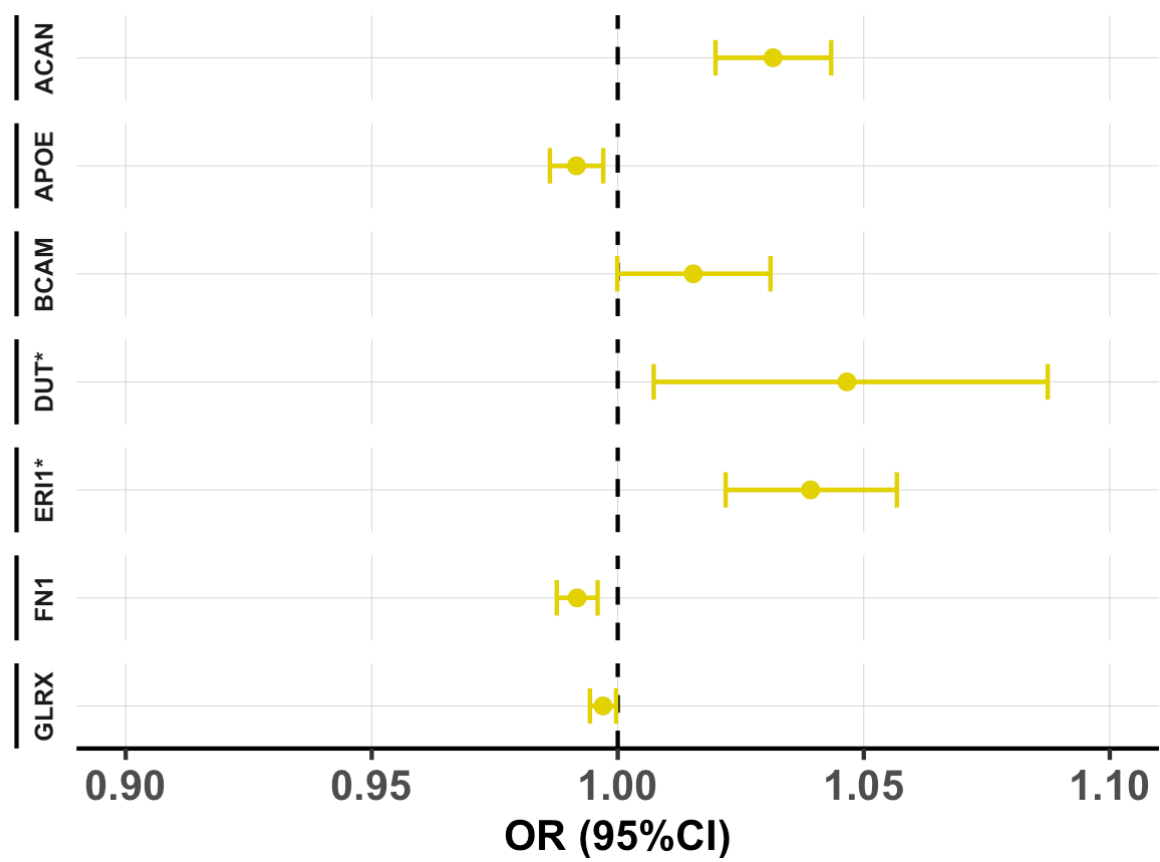
Figure 1

Figure 2

Figure 3

Figure 4

Figure 4

c

Figure 5