

UCLA

UCLA Previously Published Works

Title

GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes

Permalink

<https://escholarship.org/uc/item/7sw5z41s>

Journal

Nature Communications, 9(1)

ISSN

2041-1723

Authors

Franceschini, Nora
Giambartolomei, Claudia
de Vries, Paul S
et al.

Publication Date

2018

DOI

10.1038/s41467-018-07340-5

Peer reviewed

ARTICLE

<https://doi.org/10.1038/s41467-018-07340-5>

OPEN

GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes

Nora Franceschini, Claudia Giambartolomei et al.[#]

Carotid artery intima media thickness (cIMT) and carotid plaque are measures of subclinical atherosclerosis associated with ischemic stroke and coronary heart disease (CHD). Here, we undertake meta-analyses of genome-wide association studies (GWAS) in 71,128 individuals for cIMT, and 48,434 individuals for carotid plaque traits. We identify eight novel susceptibility loci for cIMT, one independent association at the previously-identified *PINX1* locus, and one novel locus for carotid plaque. Colocalization analysis with nearby vascular expression quantitative loci (cis-eQTLs) derived from arterial wall and metabolic tissues obtained from patients with CHD identifies candidate genes at two potentially additional loci, *ADAMTS9* and *LOXL4*. LD score regression reveals significant genetic correlations between cIMT and plaque traits, and both cIMT and plaque with CHD, any stroke subtype and ischemic stroke. Our study provides insights into genes and tissue-specific regulatory mechanisms linking atherosclerosis both to its functional genomic origins and its clinical consequences in humans.

Correspondence and requests for materials should be addressed to J.L.M.B. (email: johan.bjorkegren@mssm.edu) or to C.J.O. (email: Christopher.ODonnell@va.gov). [#]A full list of authors and their affiliations appears at the end of the paper.

Atherosclerosis is characterized by an accumulation of lipid-rich and inflammatory deposits (plaques) in the sub-intimal space of medium and large arteries. Plaque enlargement leads to blood flow limitation, organ ischemia, and/or tissue necrosis. Plaque rupture can lead to abrupt vascular occlusion, which underlies clinical cardiovascular events, including myocardial infarction and ischemic stroke. Coronary heart disease (CHD) accounts for one in seven deaths, and stroke accounts for one in 20 deaths in the US¹. Because atherosclerosis has a long pre-clinical phase, early detection of atherosclerosis using non-invasive methods may help identify individuals at risk for atherosclerotic clinical events², and provides an opportunity for prevention. Subclinical atherosclerosis can be detected by B-mode ultrasound measurement of common carotid artery intima-media thickness (cIMT) or carotid plaques¹.

Subclinical and clinical atherosclerosis has known genetic components³. Genome-wide association studies (GWAS) of subclinical atherosclerosis have previously identified three loci significantly associated with cIMT at *ZHX2*, *APOC1*, and *PINX1*, and two loci associated with common carotid artery plaque at *PIK3CG* and *EDNRA*⁴. An exome-wide-association study identified significant associations of the *APOE* ϵ 2 allele with cIMT and coronary artery calcification⁵. The *APOE* single nucleotide polymorphism (SNP) rs7412 is in linkage disequilibrium (LD) with the *APOC1* variant, thus representing the same signal. Additional GWAS-identified associations were reported for carotid plaque at the 9p21 and *SEFN2* loci⁶, and for cIMT at the *CFDP1-TMEM170A* locus⁷. However, these prior studies were of limited sample size and genomic coverage, and failed to investigate the etiological role that subclinical atherosclerosis may have on atherosclerotic clinical events.

Herein, we perform a large meta-analysis of GWAS of subclinical atherosclerosis by analyzing 1000 Genomes imputed genotype data obtained from collaborations between the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium⁸ and the University College London-Edinburgh-Bristol (UCLEB) consortium⁹. One of the greatest challenges in the translation of GWAS findings to biological understanding is related to the limited access to RNA expression data from disease-relevant tissues. Consequently, we sought to reliably identify the tissue-specific gene regulatory functions responsible for the GWAS signals by prioritizing candidate genes for established and novel loci of cIMT and carotid plaque using statistical methods for colocalization¹⁰. These methods integrate identified loci with expression quantitative loci (eQTLs) inferred

from cardiovascular disease-relevant genetics of RNA expression, the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) study, where arterial wall and metabolic-related RNA samples were collected from up to 600 patients with CHD¹¹. We also evaluate the relationships of cIMT and carotid plaque with clinically apparent CHD and stroke using summary data from two large consortia. In summary, our study sequentially assesses the genetic epidemiology and tissue-specific patterns of gene regulation involved in the formation of sub-clinical atherosclerosis traits across cardiovascular disease-related tissues.

Results

Study description. The study design is shown in Fig. 1. We undertook meta-analysis of GWAS in individuals of European ancestry for cIMT (up to 71,128 participants from 31 studies) and carotid plaque (up to 48,434 participants from 17 studies; 21,540 with defined carotid plaque) (Supplementary Table 1). cIMT and plaque were evaluated using high-resolution B-mode ultrasonography and reading protocols as previously reported⁴. Carotid plaque was defined by atherosclerotic thickening of the common carotid artery wall or the proxy measure of luminal stenosis greater than 25% (Supplementary Table 2). Each cohort performed association analyses using standardized protocols (Methods) for variants imputed based on the 1000 Genomes Project (1000G) phase 1 v3 reference. Extensive quality control (QC) was applied to data, and there was little evidence for population stratification in any of the studies for either trait (Supplementary Table 3). The study-specific results were combined using fixed-effect meta-analyses, given the low heterogeneity across studies (0% heterogeneity)¹².

GWAS meta-analyses of cIMT and carotid plaque. For cIMT, 11 loci had at least one SNP association that reached the genome-wide association threshold ($p < 5 \times 10^{-8}$), of which eight were newly described and three have been previously reported (Table 1). The closest genes for the eight loci were: 1q32.2 intergenic (rs201648240), *ATP6AP1L* (rs224904), *AIG1* (rs6907215), *PIK3CG* (rs13225723), *MCPH1* (rs2912063), *SGK223* (rs11785239), *VTII* (rs1196033), and *CBFA2T3* (rs844396). For three loci previously reported, the closest genes were *ZHX2* (rs148147734), *PINX1* (rs200482500), and *APOE* (rs7412).

The *PIK3CG* is a newly described locus for cIMT, but has been previously reported in a GWAS of carotid plaque⁴. The two

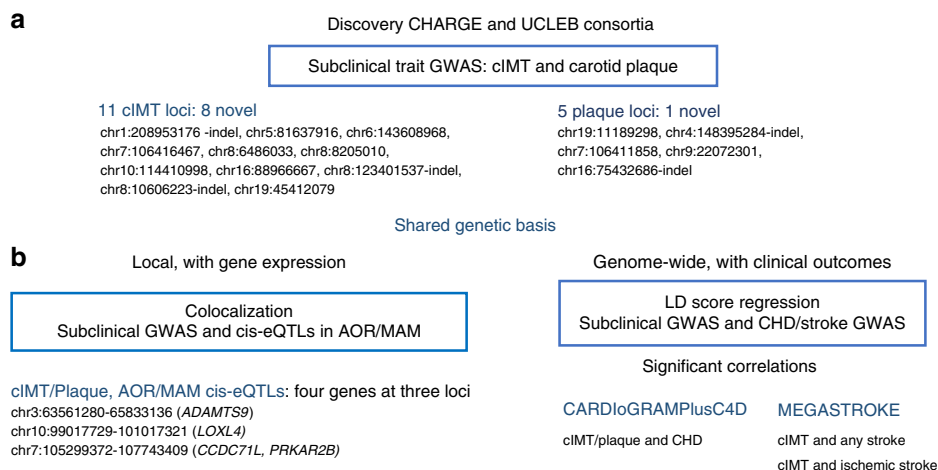


Fig. 1 Overall study design. **a** GWAS meta-analyses of cIMT and carotid plaque for gene discovery. **b** Local and genome-wide shared genetic basis using gene expression and clinical outcomes GWAS data

signals on chromosome 8 near *MCPH1* (rs2912063) and *SGK223* (rs11785239) were confirmed to be independent through conditional analysis (Supplementary Table 4). At the *PINX1* locus, the lowest association *p*-value variant (rs200482500) was not in LD with the previously reported associated variant in the region (rs6601530, $r^2 = 0.0$, Table 1), thus representing an independent signal at this locus. Two additional loci for cIMT had an SNP that reached suggestive evidence for association ($p < 1.0 \times 10^{-7}$) including an SNP nearby *APOB* (rs515135) and an intronic low frequency variant at *ATG4B* (rs139302128, minor allele frequency [MAF] = 0.03) (Supplementary Table 5).

The GWAS meta-analysis for carotid plaque identified five loci, of which one has not been previously described (nearby gene *LDLR*) (Table 1). At four known loci associated with carotid plaque (nearby genes *EDNRA*, *PIK3CG*, *CFDP1-TMEM170A*, and at the 9p21 region), the most significantly associated variants were in LD with the previously reported SNPs (Table 1)^{4,6,7}, indicating that these SNPs mark the same association at each locus. Two suggestive loci ($p < 10^{-7}$) were also identified nearby the genes *TMCO5B* and *STEAP2-AS1* (Supplementary Table 5). Conditional analyses confirmed the presence of a single independent signal at each locus. Manhattan and QQ plots from the meta-analysis of cIMT and carotid plaque are shown in Supplementary Figure 1 and regional plots in Supplementary Figure 2. Forest Plots for all loci are shown in Supplementary Figure 3.

Regulatory annotations of GWAS SNPs for cIMT/carotid plaque. To better define potentially causal variants within the identified genetic risk loci, we jointly analyzed the GWAS data with functional genomic information such as annotations on active transcription sites or open chromatin regions (i.e., performed a fine-mapping functional genome-wide association analysis using fGWAS¹³). Only variants in the *PINX1* region were

found to have a high probability that its association with cIMT is driven by SNPs that fall within transcription sites in adipose-derived mesenchymal stem cells at a DNaseI-hypersensitive site (Supplementary Figure 4), a finding that provides a down-stream mechanistic explanation for the cIMT signal in the *PINX1* locus.

To further explore the regulatory functions of variants in the identified loci for cIMT and carotid plaque, we investigated whether the identified lead SNPs were also eQTLs using vascular RNAseq data from GTEx (aorta, coronary and tibial arteries, heart atrial appendage, and heart left ventricle) and from the coronary artery disease cohort of STARNET (i.e., from the atherosclerotic-lesion-free internal mammary artery [MAM] and atherosclerotic aortic root [AOR]). Lead SNP associated with cIMT and carotid plaque (rs13225723) in the *PIK3CG* locus was found to be vascular-specific eQTLs for *CCDC71L* and *PRKAR2B* in GTEx aorta as well as in STARNET AOR and MAM tissues (Table 2, Fig. 2), suggesting that the genetic regulation of these two genes are responsible for risk variation in cIMT and carotid plaque development in this locus.

Colocalization analysis of GWAS data and STARNET eQTLs.

To identify further candidate genes in tissues affected by atherosclerosis that had strong evidence of sharing the same variant for cIMT and carotid plaque as found in our GWAS, we conducted pairwise colocalization analysis of these genetic variants with *cis*-eQTLs in the STARNET study¹⁰.

The pairwise colocalization analysis is based on coloc, a Bayesian statistical methodology that tests pairwise colocalization of SNPs in GWAS with eQTLs and, in this fashion, generates posterior probabilities for each locus weighting the evidence for competing hypothesis of either no colocalization or sharing of a distinct SNP at each locus¹⁰. We used summary statistics from all SNPs within a 200-kb window around each gene covered by the eQTL datasets ($N = 18,705$, see Methods), and analyzed each

Table 1 Loci significantly associated with cIMT and plaque GWAS

SNP	Chr:position	Nearest coding gene	Alleles (effect/other)	Effect allele freq.	Beta (SE)	<i>p</i>	<i>N</i>
Newly identified loci for cIMT							
rs201648240	1:208953176-indel	<i>LINC01717</i>	–/AA	0.83	–0.0062 (0.0011)	4×10^{-9}	54,752
rs224904	5:81637916	<i>ATP6AP1L</i>	C/G	0.95	–0.0088 (0.0016)	5×10^{-8}	68,962
rs6907215	6:143608968	<i>AIG1</i>	T/C	0.60	–0.0040 (0.0007)	5×10^{-8}	64,586
rs13225723	7:106416467	<i>PIK3CG</i>	A/G	0.22	0.0052 (0.0009)	3×10^{-9}	68,070
rs2912063	8:6486033	<i>MCPH1</i>	A/G	0.71	0.0045 (0.0008)	9×10^{-9}	67,401
rs11785239	8:8205010	<i>SGK223</i>	T/C	0.65	–0.0043 (0.0008)	9×10^{-9}	67,107
rs11196033	10:114410998	<i>VTI1A</i>	A/C	0.48	0.0042 (0.0008)	4×10^{-8}	57,995
rs844396	16:88966667	<i>CBFA2T3</i>	T/C	0.30	–0.0051 (0.0009)	6×10^{-9}	50,377
Newly identified loci for plaque							
rs200495339	19:11189298-indel	<i>LDLR</i>	–/G	0.11	–0.1023 (0.0179)	1×10^{-8}	36,569
Known loci for cIMT							
rs148147734 ^a	8:123401537-indel	<i>ZHX2</i>	–/G	0.54	0.0050 (0.0007)	3×10^{-11}	58,141
rs200482500 ^a	8:10606223-indel	<i>PINX1</i>	–/GTACC	0.52	0.0056 (0.0008)	7×10^{-12}	58,141
rs7412 ^a	19:45412079	<i>APOE</i>	T/C	0.08	–0.0119 (0.0015)	1×10^{-14}	44,607
Known loci for plaque							
rs11413744 ^b	4:148395284-indel	<i>EDNRA</i>	–/T	0.86	–0.1586 (0.0253)	4×10^{-10}	39,577
rs17477177 ^b	7:106411858	<i>PIK3CG</i>	T/C	0.79	–0.1305 (0.0197)	4×10^{-11}	47,863
rs9632884 ^b	9:22072301	9p21	C/G	0.48	0.1127 (0.0163)	5×10^{-12}	45,943
rs113309773 ^b	16:75432686-indel	<i>CFDP1-TMEM170A</i>	–/C	0.46	–0.1259 (0.0194)	9×10^{-11}	37,104

p = *p*-values of association from linear regression analysis, *N* = total number in meta-analyses

^aPublished cIMT SNP in LD with our most significant SNP: rs11781551 ($r^2 = 0.95$ with rs148147734), rs6601530 ($r^2 = 0$ with rs200482500), and rs445925 ($r^2 = 0.60$ with rs7412)

^bPublished plaque SNP in LD with our most significant SNP: rs1878406 ($r^2 = 0.98$ with rs11413744), rs17398575 ($r^2 = 0.8$ with rs17477177), rs9644862 ($r^2 = 0.79$ with rs9632884), and rs4888378 ($r^2 = 0.94$ with rs113309773)

Table 2 Gene expression results for significant SNPs in GTEx and STARNET tissues

SNP	eQTL ^a (Gene, p) GTEx		eQTL ^a (Gene, p) STARNET tissues	
	AOR ^b	HEART (ATR/VEN) ^c	AOR	MAM
rs201648240	CAMK1G, 0.0094 AL031316.1, 0.0040			CD34, 0.00532 TRAF3IP3, 0.0097
rs6907215		AL023584.1, 0.005384704 (VEN)	ENSG00000217648, 0.00046	ENSG00000217648, 0.8×10^{-5}
rs13225723	AC005050.1, 1×10^{-10} ENSG00000177820.5, 7.0×10^{-5} CCDC71L, 5×10^{-6} PRKAR2B, 4×10^{-8} PIK3CG, 10×10^{-3}		CCDC71L, 6×10^{-36} PRKAR2B, 7×10^{-7} SYPL1, 0.0043	CCDC71L, 3×10^{-33} PRKAR2B, 6×10^{-8} NAMPT, 6×10^{-6}
rs2912063	MCPH1, 0.0041	ENSG00000271743.1, 0.0093 (VEN)	MCPH1-AS1, 0.0020	
rs11785239		AC022784.1, 0.0078 (VEN)	ERI1, 0.0069	PPP1R3B, 0.0036
rs844396	ENSG00000141012.8, 0.003 AC092384.2, 0.001 CBFA2T3, 1×10^{-7}	ZNF469, 0.004 (ATR) AC092384.3, 5×10^{-6} (ATR) AC092384.1, 0.002 (ATR) CBFA2T3, 0.0004 (ATR) ZNF469, 0.002 (VEN) AC138028.4, 0.001 (VEN) ENSG00000224888.3, 0.009 (VEN) PIEZO1, 0.0004 (VEN) GALNS, 0.004 (VEN)	RPL13, 0.0024 ZNF276, 0.0070 TRAPPC2L, 0.0091	TRAPPC2L, 0.0040 ZNF276, 0.0059
rs200495339		ENSG00000267105.1, 0.0005 (VEN)		
rs148147734	DERL1, 0.0082			
rs200482500	AF131215.6, 0.005 AF131215.5, 0.001	AF131215.5, 0.002 (ATR) AF131215.6, 0.003 (VEN) AF131215.5, 0.004 (VEN)		
rs7412	ENSG00000267163.1, 0.007			
rs11413744	PRMT9, 0.004			
rs17477177	ENSG00000267052.1, 6×10^{-11} ENSG00000177820.5, 5×10^{-6} CCDC71L, 4×10^{-7} PRKAR2B, 2×10^{-8}	BCAP29, 0.002 (ATR)	CCDC71L, 2×10^{-37} PRKAR2B, 6×10^{-7} SYPL1, 0.0091	CCDC71L, 1×10^{-33} PRKAR2B, 2×10^{-8} NAMPT, 1×10^{-5}
rs9632884		DMRTA1, 0.007 (ATR)	CDKN2B, 2×10^{-3}	CDKN2B, 2×10^{-3}
rs113309773	BCAR1, 6×10^{-11} ENSG00000261783.1, 2×10^{-16} GABARAPL2, 0.004	ENSG00000261783.1, 1×10^{-5} (ATR) ENSG00000166822.8, 0.005 (ATR) ENSG00000261783.1, 0.0003 (VEN)	ZFP1, 4×10^{-4} AC009078.2, 0.002 BCAR1, 3×10^{-12} CFDP1, 0.002 TMEM170A, 0.009	

p = p-values of association from linear regression analysis

^aThe lead SNP from GWAS is considered an eQTL if the cis-association has a nominal p-value of association <0.01. Multiple but not all lead SNPs reach genome-wide significance ($p < 10^{-4}$).

^bThis includes aorta (AOR)

^cThis includes heart atrial (ATR) and heart left ventricle (VEN)

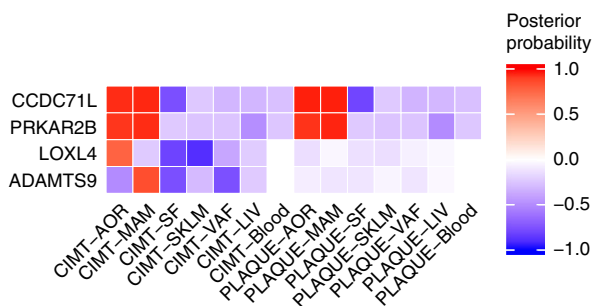


Fig. 2 Pairwise colocalization results for genes identified for cIMT and carotid plaque GWAS meta-analysis with STARNET expression datasets. Red indicates a high posterior probability of colocalization and blue a high probability of no colocalization of the same SNP with tissue eQTLs

eQTL-GWAS dataset pair (Supplementary Table 6). A posterior probability of $\geq 75\%$ was considered strong evidence of the tissue-specific eQTL-GWAS pair influencing both the expression and GWAS trait at a particular region. Results for this analysis are shown in Table 3 and Supplementary Figure 5. The strongest evidence for an effect on gene expression within the regions identified in our standard GWAS meta-analysis was for the *CCDC71L* and *PRKAR2B* genes at the previously described chromosome 7 cIMT locus (*PIK3CG* in Table 2, Fig. 2). These genes showed evidence of colocalization for both cIMT and carotid plaque in AOR and MAM tissues (Table 3, Fig. 3). *CCDC71L* had the highest probability (>95%) for colocalization for cIMT, and MAM and AOR tissue eQTLs, and for carotid plaque, and MAM and AOR tissue eQTLs. We found a low probability of colocalization of the SNP with the *PIK3CG* gene expression (<1%).

Table 3 Colocalization of cIMT and plaque with eQTLs in tissues from patients with CHD in STARNET tissues for genes/tissues combinations that have more than 75% probability to share the same associated variant

Region (chr:start-stop)	Trait	Gene	SNP with best joint probability	p , BETA (SE), Tissue posterior probability (PPA) ^a			Direction of effect GWAS/eQTL
				cIMT /plaque GWAS	AOR eQTL	MAM eQTL	
chr3:63561280-65833136	cIMT	ADAMTS9	rs17676309 (T/C)	2×10^{-6} , -0.0035 (0.0007)	2×10^{-25} , -0.65 (0.06) PPA=0.93	1×10^{-23} , -0.61 (0.06) PPA=0.89	-/-
chr10:99017729-101017321	cIMT	LOXL4	rs55917128 (T/C)	5×10^{-7} , 0.0037 (0.0007)	6×10^{-8} , 0.33 (0.06) PPA=0.79		+/+
chr7:105299372-107743409	cIMT	CCDC71L PRKAR2B	rs12705390 (A/G)	5×10^{-9} , 0.0049 (0.0008)	2×10^{-37} , 0.81 (0.06) PPA=0.97	1×10^{-33} , 0.755 (0.06) PPA=0.97	+/+ +/-
					6×10^{-7} , 0.34 (0.07) PPA=0.93	2×10^{-8} , 0.368 (0.06) PPA=0.96	
	Plaque	CCDC71L PRKAR2B	rs12705390 (A/G)	4×10^{-8} , 0.12 (0.022)	2×10^{-37} , 0.80 (0.06) PPA=0.97	1×10^{-33} , 0.75 (0.06) PPA=0.97	+/+ +/-
					6×10^{-7} , 0.33 (0.07) PPA=0.93	2×10^{-8} , 0.37 (0.06) PPA=0.96	

PPA posterior probability of sharing same SNP higher than 75%, cIMT common carotid artery intima-media thickness, AOR aorta, MAM mammary artery
^aThis signal reaches genome-wide significance in cIMT/plaque, and reaches a high probability of being mediated by the genes in AOR and MAM

The eQTL associations at two additional loci (*ADAMTS9*, *LOXL4*) in MAM or AOR showed evidence of colocalization with cIMT or carotid plaque, although GWAS association p -values at these loci did not meet the genome-wide significance threshold (Table 3, Supplementary Figure 5). Albeit with weaker magnitudes, the expression of these two genes were also associated with the top colocalizing SNPs as detected in RNAseq data in GTEx aorta (rs17676309, chr3:64730121, *ADAMTS9*, $p = 0.0003$ and rs55917128, chr10:100023359, *LOXL4*, $p = 0.0005$).

Colocalization of CHD and stroke GWAS and STARNET eQTLs. We next assessed if the four genes (*CCDC71L*, *PRKAR2B*, *ADAMTS9*, *LOXL4*) identified through colocalization of cIMT/carotid plaque with tissue-specific eQTLs also showed evidence for colocalization with CHD and stroke traits (Supplementary Data 1 and Supplementary Figure 6). We used GWAS summary data for CHD (CARDIoGRAMPlusC4D), and stroke subtypes (MEGASTROKE) and AOR and MAM STARNET tissue eQTLs for these analyses. *CCDC71L* and *PRKAR2B* had suggestive evidence of sharing the same variant with large vessel disease stroke in both AOR and MAM tissues (probability of colocalization $\geq 20\%$, Supplementary Data 1). In contrast, there was strong evidence ($\geq 75\%$) to reject a shared variant for CHD and eQTLs at this locus, thus suggesting there is atherosclerotic outcome specificity at vascular level for this locus (Supplementary Figure 5). Three of these genes, *CCDC71L*, *PRKAR2B*, and *ADAMTS9*, showed evidence for shared genetic influences of cIMT or carotid plaque on CHD/stroke outcomes when testing the joint association using moloc, a multiple-trait extension of coloc¹⁴ (Supplementary Table 7). We also highlight the expression of *KIAA1462* gene in MAM, carotid plaque/cIMT, and CHD, which were positively correlated (Supplementary Figure 7). This gene has suggestive evidence of pairwise colocalization with carotid plaque (67% of probability of shared variant between carotid plaque and eQTL in MAM), as well as a high probability of shared variant between MAM eQTL expression of this gene, GWAS carotid plaque or cIMT, and CHD traits (Supplementary Table 7). We note, however, that the GWAS signal for outcomes across the datasets did not reach genome-wide significance and larger sample sizes may be needed to strengthen the evidence for involvement in disease outcomes.

Genetic correlations of cIMT/carotid plaque and clinical outcomes. To provide etiological insights into the role of measures of

subclinical atherosclerosis and major atherosclerotic disease outcomes such as CHD and ischemic stroke, we quantified the genetic correlation using cross-trait LD score regression, a method that estimates genetic correlation across different traits using summary level data¹⁵. We used summary statistics between cIMT/carotid plaque with CHD and stroke meta-analysis of GWAS. Both cIMT and carotid plaque had positive significant genetic correlations with CHD (all $p < 0.05$ after adjusting for multiple testing), though the magnitude of the correlation was twice as strong for carotid plaque (0.52) as for cIMT (0.20) (Table 4). There was also evidence for genetic correlations between cIMT with any stroke and ischemic stroke subtype.

Pathway analysis and druggability. Gene Ontology (GO) analyses of genes identified in the loci for cIMT and carotid plaque according to our meta-analysis of GWAS (Table 1 and Supplementary Table 5) and in the colocalization analyses (Table 3, Supplementary Table 7) showed that cIMT genes are enriched in lipoprotein-related terms and cholesterol efflux, whereas carotid plaque genes are enriched in terms associated with fibroblast apoptosis (Supplementary Figure 8). Analysis of the cIMT genes using a GO Slim additionally identified several of the genes that were associated with terms describing cardiovascular development, cell adhesion, and immune processes, processes already considered relevant to atherosclerosis. Specifically, there is corroborating evidence from GO that *CCDC71L*, *PRKAR2B*, and *TWIST1* are associated with cIMT/carotid plaque as they are involved in lipid metabolism, with similar support that *ADAMTS9*, *CDH13*, and *KIAA1462* are associated with cIMT or carotid plaque risk as they are all involved in cell adhesion and, together with *TWIST1*, in cardiovascular system development (Supplementary Data 2).

From the loci associated with cIMT and carotid plaque, we identified seven genes (*ATG4B*, *ALPL*, *LDLR*, *APOB*, *EDNRA*, *APOE*, and *ADAMTS9*) whose encoded proteins are targets at various stages of the drug development process (Supplementary Tables 8 and 9). *ADAMTS9* gene encodes a protein likely to be druggable¹⁶. *ATG4B*, *ALPL*, and *LDLR* are proteins being targeted by compounds in pre-clinical phase (tier 2), while *APOB* and *EDNRA* are proteins targeted by drugs in clinical phase or licensed (tier 1). *APOB* is the target of an approved FDA drug for treatment of familial hypercholesterolemia. *EDNRA* gene encodes for endothelin A receptor, against which several antagonists have been developed for the treatment of pulmonary arterial

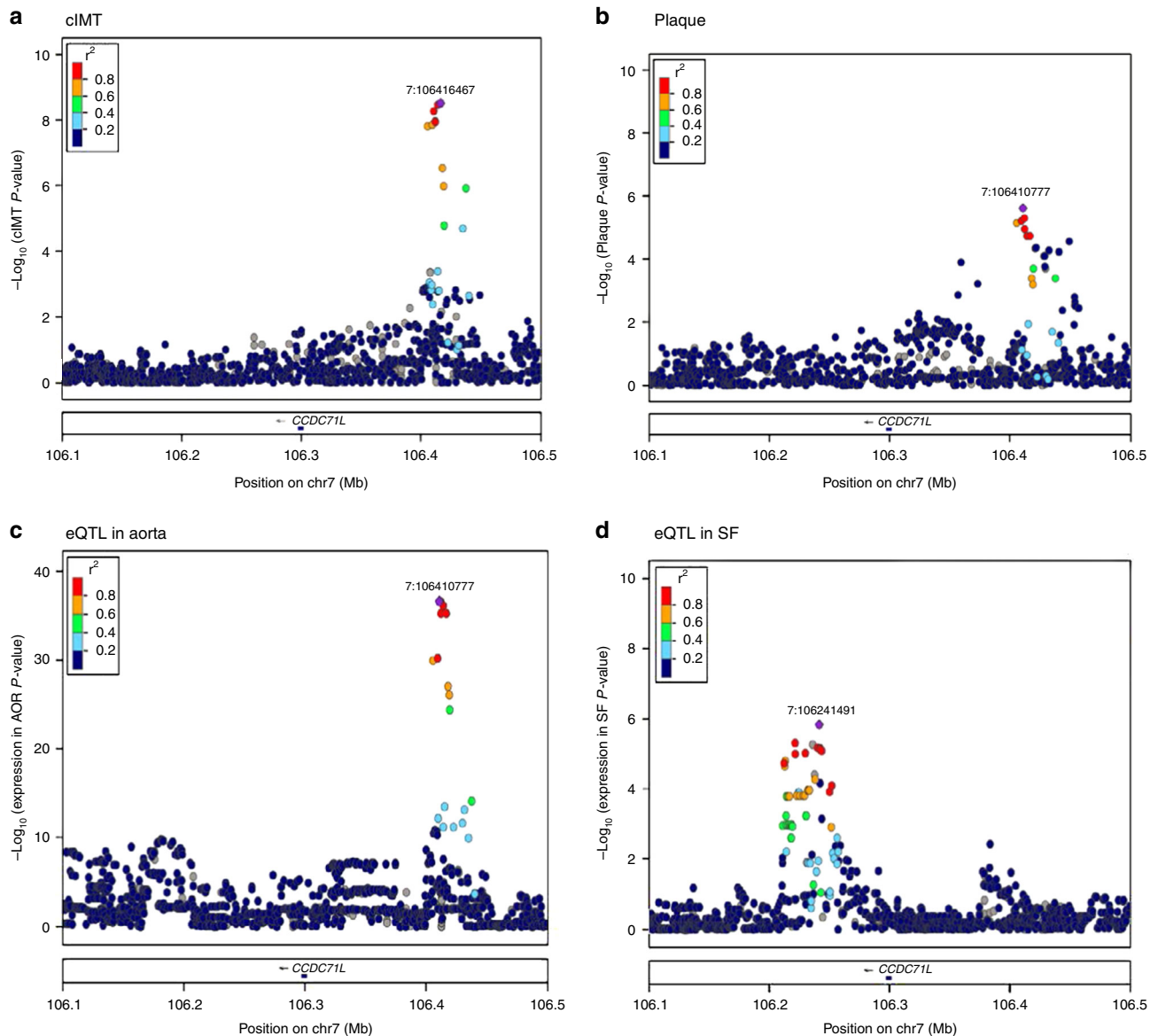


Fig. 3 Association results at the *CCDC71L* locus (chromosome 7), showing a high posterior probability of a shared variant for cIMT and carotid plaque in AOR and MAM eQTLs. $-\log_{10}(p)$ SNP association p -values for cIMT (plot A) and carotid plaque (plot B), and eQTL in AOR (plot C) and eQTL in SF (plot D). Association results in SF tissue have a low probability of a shared signal with cIMT and carotid plaque, possibly indicating a different mechanism in this tissue. eQTLs in MAM are identical to AOR and not shown. The p -values were calculated by fitting a linear regression model with cIMT or plaque as dependent variable and imputed SNPs as independent variables. Each dot is an SNP and the color indicates linkage disequilibrium (r^2) with the best hit (in purple)

hypertension or which are in advanced clinical phase development for non-small cell lung cancer and diabetic nephropathy.

Discussion

We provide results of a large meta-analysis of GWAS of sub-clinical atherosclerosis and we integrate our results with tissue-specific gene expression data using eQTLs from both the early (MAM) and late advanced (AOR) atherosclerotic arterial wall from the STARNET study to enable reliable discovery of genes with biological evidence of an increased probability for conferring inherited risk of atherosclerosis development. Our discovery approach using GWAS meta-analyses identified 16 loci significantly associated with either cIMT or carotid plaque, of which nine are novel.

The integration of GWAS and tissue-specific *cis*-eQTLs for the joint analyses of tissue-specific eQTLs from CHD patients identified two potentially additional loci colocalizing with cIMT or carotid plaque: chr3:63561280-65833136 (*ADAMTS9*), chr10:99017729-101017321 (*LOXLA*). *ADAMTS9* is a metalloproteinase involved in thrombolysis and angiogenesis and has been associated with cardiometabolic traits (waist-to-hip ratio, waist circumference, and type 2 diabetes) in GWAS, and with coronary artery calcification in a gene-by-smoking interaction GWAS^{17,18}. *LOXLA* encodes a lysyl oxidase involved in crosslinks of collagen and elastin in the extracellular matrix. This family of proteins are involved in the development of elastic vessels and mechanical strength of the vessel wall, and their inhibition was associated with the development of abdominal aortic aneurysms and more severe atherosclerosis in experimental models¹⁹.

Table 4 Genetic correlation between CHD and stroke traits with cIMT and plaque, and cIMT with plaque using LD score and meta-GWAS

Cardiovascular disease trait	Subclinical atherosclerosis trait	Genetic correlation	SE	z	p
CHD ^a	cIMT	0.20	0.05	4.1114	4×10^{-5}
Any stroke	cIMT	0.30	0.07	4.2301	2.3×10^{-5}
Ischemic stroke ^b	cIMT	0.31	0.07	4.646	3.4×10^{-6}
Cardio-embolic stroke ^b	cIMT	0.10	0.09	1.0729	0.28
Small vessel disease stroke ^b	cIMT	0.33	0.18	1.8728	0.06
CHD ^a	Carotid plaque	0.52	0.08	6.4263	1.3×10^{-10}
Any stroke ^b	Carotid plaque	0.28	0.10	2.7097	0.007
Ischemic stroke ^b	Carotid plaque	0.27	0.10	2.6578	0.008
Cardio-embolic stroke ^b	Carotid plaque	0.06	0.14	0.4684	0.64
Small vessel disease stroke ^b	Carotid plaque	-0.03	0.24	-0.1344	0.89
Plaque	cIMT	0.40	0.10	3.9667	7.3×10^{-5}

^aCARDIoGRAMPlusC4D^bMEGASTROKE consortium. Unable to estimate the genetic correlations with large vessel disease

Some loci identified in our meta-analysis of GWAS include genes in known pathways for atherosclerosis, including *LDLR*, which is related to lipid pathways and CHD, and identified for associations with carotid plaque in our study. For most of the loci, however, the underlying gene implicated in signals are unknown. Our colocalization approach found both *CCDC71L* and *PRKAR2B* as the most likely genes at the chromosome 7 locus, where *PIK3CG* was previously the suggested gene. This finding is in agreement with a targeted sequencing study of subclinical atherosclerosis¹⁵. An additional SNP (rs342286) at this locus has been associated with platelets volume and reactivity, and cardiovascular traits. However, rs342286 is not in LD with our most significant SNP and it is not associated with cIMT or carotid plaque in our studies ($p = 0.49$ and 0.01 , respectively). Of interest, the variant we identified in this study showed evidence for colocalization with cIMT/carotid plaque and large vessel disease stroke but not CHD, therefore showing tissue and outcome-specificity. *CCDC71L* has unknown function. *PRKAR2B* codes for one of the several regulatory subunits of cAMP-dependent protein kinase and its expression is ubiquitous. In vitro studies have shown that adenosine-induced apoptosis of arterial smooth muscle cells involves a cAMP-dependent pathway²⁰.

Measures of cIMT and carotid plaque reflect vascular pathophysiological and atherosclerosis processes, respectively, with carotid plaque more strongly reflecting atherosclerotic clinical events. An important contribution of this study is the supporting evidence for overall genetic correlations of CHD and stroke (any cause and ischemic stroke) with subclinical atherosclerosis traits, estimated using LD score methods. Further highlighting the potential biological relevance of our findings, the genetic correlations estimates for CHD were stronger for carotid plaque than for cIMT. However, cIMT and carotid plaque GWAS were correlated, and the genetic correlations estimates with stroke were similar for cIMT or carotid plaque, and not significant for carotid plaque. The colocalization analyses provided additional insights in the relationships between subclinical atherosclerosis, clinical outcomes, and tissue-specific regulation at specific genomic regions. For example, our suggestive top gene association in multi-trait colocalization for *KIAA1462* included MAM eQTLs, carotid plaque, and CHD, supporting the shared genetic effects at this locus of atherosclerosis in carotid and coronary arteries. *KIAA1462* has been previously reported in the same locus identified by GWAS for CHD²¹. This gene encodes a protein involved in cell-cell junctions in endothelial cells²², which was recently shown to be involved in pathological angiogenic process in vitro and in vivo experimental models²³. These findings suggest that

there may be important differences in vascular bed regulation at distinctive regions for atherosclerotic cardiovascular and stroke outcomes that may help to identify genes and specific targets for CHD or stroke prevention and treatment.

Additional studies in diverse and large samples across the multiple datasets are needed to explore these results further. As more summary statistics become available for other clinical endpoints beyond stroke and CHD (both in terms of larger sample size and richer genome coverage), and as further refinements in clinical phenotypes emerge (e.g. from CHD to acute coronary syndrome sub-components), strategies to integrate this knowledge using methods such as *moloc*¹⁰ and *eCAVIAR*²⁴ will continue to be essential for harnessing genome-wide findings in the drug-discovery process.

In summary, our study is a large GWAS meta-analysis of cIMT and carotid plaque. Through a sequential approach of discovery and colocalization studies, we provide deeper insights into disease causal genes of subclinical cIMT and carotid plaque formation. We confirmed three loci and identified nine novel loci in the meta-analyses of cIMT and carotid plaque. Additionally, we provide strong evidence for the role of three novel genes from our integrative analysis of GWAS and eQTL data. Moreover, the identified correlations with CHD and stroke highlight novel biological pathways that merit further assessments as novel targets for drug development.

Methods

Ethics statement. All human research was approved by the relevant institutional review boards for each study, and conducted according to the Declaration of Helsinki. All participants provided written informed consent.

Populations and phenotypes. The discovery GWAS in this study consists of a collaboration between the CHARGE⁸ and the UCLEB consortia⁹, for genetic studies of cIMT and carotid plaque among individuals of European ancestry (Supplementary Note 1). All studies followed standardized protocols for phenotype ascertainment and statistical analyses. The descriptive characteristics of participating studies are shown in Supplementary Table 1.

cIMT and carotid plaque measures were evaluated using high-resolution B-mode ultrasonography and reading protocols as previously reported⁴. We used data from the baseline examination or the first examination in which carotid ultrasonography was obtained. cIMT was defined by the mean of the maximum of several common carotid artery measurements, measured at the far wall or the near wall. For most studies, this was an average of multiple measurements from both the left and right arteries. We also examined a carotid plaque phenotype, defined by atherosclerotic thickening of the carotid artery wall or the proxy measure of luminal stenosis greater than 25% (Supplementary Table 2).

Genotyping, imputation, and study-level quality control. Genotyping arrays and QC pre-imputation are shown in Supplementary Table 3. Each GWAS study

conducted genome-wide imputation using a Phase 1 integrated (March 2012 release) reference panel from the 1000G Consortium using IMPUTE2²⁵ or MaCH/Minimac²⁶, and used Human Reference Genome Build 37. Sample QC was performed with exclusions based on call rates, extreme heterozygosity, sex discordance, cryptic relatedness, and outlying ethnicity. SNP QC excluded variants based on call rates across samples and extreme deviation from Hardy–Weinberg equilibrium (Supplementary Table 3). Non-autosomal SNPs were excluded from imputation and association analysis.

Pre-meta-analysis GWAS study-level QC was performed using EasyQC software²⁷. This QC excluded markers absent in the 1000G reference panel; non A/C/G/T/D/I markers; duplicate markers with low call rate; monomorphic SNPs and those with missing values in alleles, allele frequency, and beta estimates; SNPs with large effect estimates or standard error (SE) ≥ 10 ; and SNPs with allele frequency difference >0.3 compared to 1000G reference panel. There was a total of 9,574,088 SNPs for the cIMT meta-analysis and 8,578,107 SNPs for the carotid plaque meta-analysis.

Statistical analyses. Within each study, we used linear and logistic regression to model cIMT and carotid plaque, respectively, and an additive genetic model (SNP dosage) adjusted for age, sex, and up to 10 principal components. We combined summary estimates from each study and each trait using an inverse variance weighted meta-analysis. Additional filters were applied during meta-analyses including imputation quality (MACH $r^2 < 0.3$ and IMPUTE info < 0.4), a minor allele frequency (MAF) < 0.01 , and SNPs that were not present in at least four studies. The genome-wide significance threshold was considered at $p < 5.0 \times 10^{-8}$.

To assess the evidence for independent associations at each locus attaining genome-wide significance, we performed conditional analysis in a 1-Mb genomic interval flanking the lead SNP using GCTA²⁸. This approach uses summary meta-analysis statistics and a LD matrix from an ancestry-matched sample to perform approximate conditional SNP association analysis. The estimated LD matrix was based on 9713 unrelated individuals of European ancestry from the ARIC study, which was genotyped using an Affymetrix 6.0 array and imputed to the 1000G panel using IMPUTE2²⁵.

Gene expression analysis using GTEx. GTEx Analysis V6 (dbGaP Accession phs000424.v6.p1) eQTL results were downloaded from GTEx portal for 44 tissues, and then mapped to SNPs listed in Table 1. We used a false discovery rate (FDR) of ≤ 0.05 .

Colocalization analyses using eQTLs. We integrated our GWAS results with cis-eQTL data using a Bayesian method (coloc)¹⁰. This method evaluates whether the GWAS and eQTL associations best fit a model in which the associations are due to a single shared variant (summarized by the posterior probability). We used gene expression datasets from multiple tissues from patients with CHD of the STAR-NET study, including blood, MAM, AOR, subcutaneous fat (SF), visceral fat (VAF), skeletal muscle (SKLM), and liver (LIV) obtained from 600 patients during open heart surgery¹¹. Pairwise colocalization was tested between these expression disease tissue datasets and GWAS results from our cIMT/carotid plaque GWAS meta-analysis. We used GWAS and eQTL summary statistics of SNPs within a 200-kb window around each gene covered by the eQTL datasets. A posterior probability of colocalization ≥ 0.75 was considered a strong evidence for a causal gene. Next, we reported the gene(s) in the STARNET datasets that had the strongest evidence of sharing the same variant with cIMT or carotid plaque genome-wide. In an alternative analysis, we also tested loci with an SNP that reached a threshold of significant or suggestive genome-wide significance for cIMT or carotid plaque (reported in Table 1, Supplementary Table 5). For each region 200kb around the SNP with the lowest association p -value, we report the gene with the highest probability of being responsible for the GWAS signal (Supplementary Table 6).

Pairwise colocalization for these genes was also tested for publicly available GWAS for CHD case-controls (CARDIOGRAMPlusC4D) and stroke case-controls (MEGASTROKE consortium). The MEGASTROKE dataset uses genotypes imputed to the 1000G phase I haplotype panel. The European ancestry sample used to generate these results consisted of 40,585 stroke cases and 406,111 controls from 15 cohorts and two consortia: the METASTROKE and CHARGE consortia²⁹. The phenotypes used in this analysis were any stroke ($n = 39,067$ cases, total $n = 442,142$), ischemic stroke (IS, $n = 32,686$ cases, total $n = 423,266$), and etiologic stroke subtypes: cardioembolic stroke (CE, $n = 6,820$ cases, total $n = 314,368$), large vessel disease ($n = 4,113$, total $n = 202,263$), and small vessel disease (SVD, $n = 4,975$, total $n = 242,250$). To explore multi-trait colocalizations, we used moloc¹⁴ with prior probabilities of 10^{-4} for GWAS/GWAS/eQTL, 10^{-6} for GWAS+eQTL/GWAS or GWAS+GWAS/eQTL, and 10^{-7} for colocalization of all three association signals.

Functional annotation and epigenetic enrichment analyses. From the Epigenome Roadmap Project^{30,31}, we obtained regulatory information using broad classes of chromatin states ($n = 127$ tissues) capturing promoter-associated, transcription-associated, active intergenic, and large-scale repressed and repeat-associated states.

From ENCODE³², we obtained chromatin states, uniformly processed transcription factor (TF) Chip assays and DNaseI Hypersensitivity sites (DHS) for nine cell lines. From FANTOM5³³, we used information from expression of enhancers in each tissue ($n = 112$), and enhancers that are positively differentially expressed against any other tissue ($n = 110$).

We used fGWAS¹³ to identify genomic annotations that are enriched within the cIMT results and to select the variants with support for a functional role based on the most informative annotations. We only considered cIMT for these analyses because of the small number of identified loci for carotid plaque. We first estimated the enrichment parameters for each annotation individually and identified the set of annotations with significant marginal associations. We then applied 10-fold cross-validation likelihood and forward selection to identify the set of annotations that significantly improve the model fit, and reverse selection of each annotation included in the model, as suggested in the fGWAS workflow. We reported the model with the highest cross-validation likelihood and SNPs that have regional posterior probability of association (PPA) >0.9 and directly overlap the genomic annotations considered.

Overall genetic correlation analysis. Genetic correlation between cIMT/carotid plaque, CHD, and stroke traits were calculated using LD score regression approach LD-score, which uses GWAS summary statistics and is not affected by sample overlap. This method relies on the fact that the χ^2 association statistic for a given SNP includes the effects of all SNPs that are in LD with it and it calculates genetic correlation by partitioning the SNP heritabilities¹⁵. Genetic correlations between stroke traits (IS, CE, large vessel disease, and SVD) and cIMT and carotid plaque were calculated using software available at <http://github.com/bulik/ldsc> with GWAS summary statistics for our cIMT/carotid plaque GWAS, CARDIOGRAMPlusC4D data, and stroke GWAS. We used the LD-scores¹⁵, which are based on the 1000 Genomes European population and estimated within 1-cM windows. Based on ten tests performed (two subclinical traits and five outcomes), we set the significance threshold to $p = 0.005$.

PATHWAY ANALYSES. Methods for GO Slim: The Ensembl identifiers of all protein-coding genes identified as in LD with the 12 variants for cIMT and 15 variants for carotid plaque (including variants from main and suggestive signals, Table 1 and Supplementary Table 5), and five genes for which there is strong evidence of colocalization (Table 3), were mapped to UniProt accession numbers, using the UniProt ID mapping service (<http://www.uniprot.org/uploadlists/>). A GO Slim analysis was performed on this list using QuickGO (www.ebi.ac.uk/QuickGO) and the Generic GO Slim. The GO terms used in the final slim analysis were further refined by adding/removing GO terms to provide more detailed information about the processes covered.

Methods for GO term enrichment analysis: The VLAD gene list analysis and visualization tool (<http://proto.informatics.jax.org/prototypes/vlad/>) was used to perform a GO term enrichment analysis on the same UniProt accessions as listed for the GO Slim. The background annotation set was obtained from the goa_human.gaf file (dated 21 November 2017, downloaded from <ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/HUMAN/>) and the ontology data was obtained from the go-basic.obo file provided in the VLAD tool (analysis run 28 November 2017).

The LD block around top SNPs associated with cIMT and carotid plaque was constructed using LD information from the 1000 Genomes panel, as previously outlined in Finan et al.¹⁶. Briefly, the boundaries of the LD region were defined as the positions of the variants furthest upstream and downstream of a GWAS SNP with an r^2 value of ≥ 0.5 and within a 1-Mbp flank on either side of the GWAS variant. Associated variants that were not present in the 1000 Genomes panel that were not in LD with any other variants were given a nominal flank of 2.5 kbp on either side of the association. Gene annotations using Ensembl version 79 were then overlapped to the LD region.

Druggable genes. We examined the druggability status for the nearest coding genes identified in our GWAS analysis on cIMT and carotid plaque, including significant (novel and replicated) and suggestive ones, as well as genes identified through colocalization analysis. The druggable gene set was calculated using the previously described criteria: novel targets of first-in-class drugs licensed since 2005; the targets of drugs currently in late phase clinical development; pre-clinical phase small molecules with protein binding measurements reported in the ChEMBL database; and genes encoding secreted or plasma membrane proteins that are targets of monoclonal antibodies and other bio-therapeutics¹⁶. We defined three tiers of druggable gene sets based on their drug development. In Tier 1, 1427 genes were targets of approved small molecules and biotherapeutic drugs and clinical-phase drug candidates. Tier 2 comprised 682 genes encoding targets with known bioactive drug-like small molecule binding partners and those with significant sequence similarity to approved drug targets. Tier 3 contained 2370 genes encoding secreted or extracellular proteins, proteins with more distant similarity to approved drug targets, and druggable genes not included in Tier 1 or 2 such as GPCRs, nuclear hormone receptors, ion channels, kinases, and phosphodiesterases.

URLs. For GTEx, see <http://gtexportal.org/>. For Coloc, see <https://cran.r-project.org/web/packages/coloc/coloc.pdf>. For Moloc, see <https://github.com/clagiamba/>

[moloc/blob/master/man/moloc-package.Rd](#). For CARDIoGRAMPlusC4D, see www.cardiogramplusc4d.org/. For LD scores, www.broadinstitute.org/~bulik/eur_ldscores/. For UniProt ID, www.uniprot.org/uploadlists/. For QuickGO, www.ebi.ac.uk/QuickGO. For VLAD tool, see <http://proto.informatics.jax.org/prototypes/vlad/>.

Data availability

All relevant summary statistics data that support the findings of this study have been deposited in the database of Genotypes and Phenotypes (dbGaP) under the CHARGE acquisition number (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v6.p1; accession phs000930.v6.p1). GWAS data for most US studies are already available in dbGaP.

Received: 19 December 2017 Accepted: 24 September 2018

Published online: 03 December 2018

References

- Mozaffarian, D. et al. Heart Disease and Stroke Statistics-2016 Update: a report from the American Heart Association. *Circulation* **133**, e38–e360 (2016).
- Frieden, T. R. & Berwick, D. M. The Million Hearts initiative--preventing heart attacks and strokes. *N. Engl. J. Med.* **365**, e27 (2011).
- O'Donnell, C. J. & Nabel, E. G. Genomics of cardiovascular disease. *N. Engl. J. Med.* **365**, 2098–2109 (2011).
- 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**, 940–947 (2011).
- Natarajan, P. et al. Multiethnic Exome-Wide Association Study of Subclinical Atherosclerosis. *Circ. Cardiovasc. Genet.* **9**, 511–520 (2016).
- Pott, J. et al. Genome-wide meta-analysis identifies novel loci of plaque burden in carotid artery. *Atherosclerosis* **259**, 32–40 (2017).
- 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**, 656–665 (2012).
- Psaty, B. M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* **2**, 73–80 (2009).
- Shah, S. et al. Causal relevance of blood lipid fractions in the development of carotid atherosclerosis: Mendelian randomization analysis. *Circ. Cardiovasc. Genet.* **6**, 63–72 (2013).
- Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
- Franzen, O. et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
- Pfeiffer, R. M., Mitchell, H. G. & Pee, D. On combining data from genome-wide association studies to discover disease-associated SNPs. *Stat. Sci.* **24**, 547–560 (2009).
- Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
- Giambartolomei, C. et al. A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018).
- Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, eaag1166 (2017).
- Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).
- Polfus, L. M. et al. Genome-wide association study of gene by smoking interactions in coronary artery calcification. *PLoS ONE* **8**, e74642 (2013).
- Remus, E. W. et al. The role of lysyl oxidase family members in the stabilization of abdominal aortic aneurysms. *Am. J. Physiol. Heart Circ. Physiol.* **303**, H1067–H1075 (2012).
- Peyot, M. L. et al. Extracellular adenosine induces apoptosis of human arterial smooth muscle cells via A(2b)-purinoceptor. *Circ. Res.* **86**, 76–85 (2000).
- Erdmann, J. et al. Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur. Heart J.* **32**, 158–168 (2011).
- Akashi, M., Higashi, T., Masuda, S., Komori, T. & Furuse, M. A coronary artery disease-associated gene product, JCAD/KIAA1462, is a novel component of endothelial cell-cell junctions. *Biochem. Biophys. Res. Commun.* **413**, 224–229 (2011).
- Hara, T. et al. Targeted disruption of JCAD (Junctional Protein Associated With Coronary Artery Disease)/KIAA1462, a coronary artery disease-associated gene product, inhibits angiogenic processes in vitro and in vivo. *Arterioscler. Thromb. Vasc. Biol.* **37**, 1667–1673 (2017).
- Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014).
- Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
- Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
- Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Zhu, L. J. Integrative analysis of ChIP-chip and ChIP-seq dataset. *Methods Mol. Biol.* **1067**, 105–124 (2013).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Anderson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

Acknowledgements

The work was supported by the following grants: National Institute of Health grants: R21HL123677, R21-HL140385, DK104806-01A1, R01-MD012765-01A1 (NF), National Institutes of Health awards R01HG009120, R01HG006399, U01CA194393, T32NS048004 (CG), the American Heart Association Grant #17POST33350042 (PV), the British Heart Foundation (RG/13/5/30112) and the National Institute for Health Research University College London Hospitals Biomedical Research Centre (RCL and RPH), the British Heart Foundation FS/14/55/30806 (JCH), the German Federal Ministry of Education and Research (BMBF) in the context of the e:Med program (e:AtheroSysMed), the DFG as part of the CRC 1123 (B3), and the FP7/2007-2103 European Union project CVgenes@target (grant agreement number Health-F2-2013-601456). We thank Li-Ming Gan for assistance with the STARNET study and Jon White for assistance with UCLEB analyses. Additional acknowledgements are included in Supplementary Note 2.

Author contributions

N.F., C.G., J.C.B., M.K., C.D., M.S., S.K.M., U.S., W.P., A.B.Z., A.H., A.T., A.G.U., A.B.N., B.W.P., C.D.L., E.T., F.R., H.V., I.J.D., L.J.L., M.D., O.R., O.H.F., R.S., R.M., T.B.H., T.L., U.F., W.P., A.D., A.S., A.H., C.M.D., D.A.L., D.O.M.-K., D.W.B., H.S., J.F.W., J.G.W., J.I.R., J.M.W., M.L., M.K.E., S.E.H., U.V., V.G., A.D.H., J.P.C., and C.J.O. contributed to study concept and design. C.D., M.S., S.K.M., U.S., W.P., A.I., A.C., A.B.Z., A.T., A.G.U., A.W., A.J.S., A.B., A.R., B.H., B.W.P., C.F., D.B., D.H.O., D.T., D.K., E.T., E.M., E.G., E.B., E.E.S., E.I., F.R., F.B., G.H., H.C., H.V., H.S.M., I.J.D., J.W.J., J.G., J.P., J.T., J.E., K.D.T., L.K., L.L., L.J.L., L.H., M.D., M.S., M.K., M.K., M.A.N., O.M., O.R., P.H.W., P.G., L.A., R.R., R.B., R.H., R.S., R.M., R.W.M., S.G.W., S.M.L., S.T., S.K., S.R.H., S.R., T.B.H., T.L., T.G., T.S., U.F., V.P., W.R., W.P., X.Z., A.S., A.H., B.M.P., C.M.D., D.A.L., D.O.M.-K., D.W.B., H.S., J.F.W., J.G.W., J.I.R., J.C.H., J.M.W., J.D., J.H., M.K.E., S.E.H., U.V., V.G., A.D.H., J.L.M.B., J.P.C., and C.J.O. contributed to acquisition of genotyping or phenotypic data. N.F., C.G., C.F., J.C.B., R.P.H., R.C.L., S.M.T., T.W.W., M.G., M.K., C.D., A.V.S., E.H., E.M.L., I.M.N., L.L., M.S., M.S., N.P., O.F., P.K.J., R.N., R.E.M., S.-J.H., S.K.M., U.S., A.I., A.T., K.R., A.J.C., B.S., C.D.L., C.W., F.V., G.C., H.S., J.P., J.L., K.G., L.M.R., M.T., M.A.N., O.M., P.R., P.A., Q.W., R.J.S., S.H., S.S., S.M.L., T.S., X.Z., X.Z., X.G., Y.S., and L.-P.L. contributed to statistical analysis and interpretation of the data. N.F., C.G., P.S.V., J.C.B., M.K., S.K.M., A.D.H., and J.P.C. contributed to drafting of the manuscript. All authors contributed to the critical revision of the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-07340-5>.

Competing interests: C.F. received a fee for speaking at a course by Springer Healthcare/Malesci. E.I. is a scientific advisor for Precision Wellness, Cellink and Olink Proteomics

for work unrelated to the present project. M.A.N.'s participation in this project was supported by a consulting contract between Data Tecnica International and the National Institute on Aging, NIH, Bethesda, MD, USA. M.A.N. also consults for Illumina Inc., the Michael J. Fox Foundation, and University of California Healthcare. B.M.P. serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. D.A.L. has received support from Roche Diagnostics and Medtronic for biomarker research unrelated to this paper. J.P.C. has received funding from GSK regarding methodological work around electronic health records, and -omics for drug discovery. All remaining authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Nora Franceschini¹, Claudia Giambartolomei², Paul S. de Vries³, Chris Finan⁴, Joshua C. Bis⁵, Rachael P. Huntley⁴, Ruth C. Lovering⁴, Salman M. Tajuddin⁶, Thomas W. Winkler⁷, Misa Graff¹, Maryam Kavousi⁸, Caroline Dale⁹, Albert V. Smith^{10,11}, Edith Hofer^{12,13}, Elisabeth M. van Leeuwen⁸, Ilya M. Nolte¹⁴, Lingyi Lu¹⁵, Markus Scholz^{16,17}, Muralidharan Sargurupremraj¹⁸, Niina Pitkänen¹⁹, Oscar Franzén^{20,21}, Peter K. Joshi²², Raymond Noordam²³, Riccardo E. Marioni^{24,25}, Shih-Jen Hwang^{26,27}, Solomon K. Musani²⁸, Ulf Schminke²⁹, Walter Palmas³⁰, Aaron Isaacs^{8,31}, Adolfo Correa²⁸, Alan B. Zonderman⁶, Albert Hofman^{8,32}, Alexander Teumer^{33,34}, Amanda J. Cox^{35,36}, André G. Uitterlinden^{8,37}, Andrew Wong³⁸, Andries J. Smit³⁹, Anne B. Newman⁴⁰, Annie Britton⁴¹, Arno Ruusalepp^{21,42,43}, Bengt Sennblad^{44,45}, Bo Hedblad⁴⁶, Bogdan Pasaniuc^{2,47}, Brenda W. Penninx⁴⁸, Carl D. Langefeld¹⁵, Christina L. Wassel⁴⁹, Christophe Tzourio¹⁸, Cristiano Fava^{46,50}, Damiano Baldassarre^{51,52}, Daniel H. O'Leary⁵³, Daniel Teupser^{17,54}, Diana Kuh³⁸, Elena Tremoli^{52,55}, Elmo Mannarino⁵⁶, Enzo Grossi⁵⁷, Eric Boerwinkle^{3,58}, Eric E. Schadt^{20,21}, Erik Ingelsson^{59,60,61}, Fabrizio Veglia⁵², Fernando Rivadeneira^{8,37}, Frank Beutner⁶², Ganesh Chauhan^{18,63}, Gerardo Heiss¹, Harold Snieder¹⁴, Harry Campbell²², Henry Völzke^{33,34}, Hugh S. Markus⁶⁴, Ian J. Deary^{24,65}, J. Wouter Jukema⁶⁶, Jacqueline de Graaf⁶⁷, Jacqueline Price²², Janne Pott^{16,17}, Jemma C. Hopewell⁶⁸, Jingjing Liang⁶⁹, Joachim Thiery^{17,70}, Jorgen Engmann⁴, Karl Gertow⁴⁴, Kenneth Rice⁷¹, Kent D. Taylor⁷², Klodian Dhana⁷³, Lambertus A.L.M. Kiemeneij⁷⁴, Lars Lind⁷⁵, Laura M. Raffield⁷⁶, Lenore J. Launer⁶, Lesca M. Holdt^{17,54}, Marcus Dörr^{34,77}, Martin Dichgans^{78,79}, Matthew Traylor⁶⁴, Matthias Sitzer⁸⁰, Meena Kumari^{41,81}, Mika Kivimaki⁴¹, Mike A. Nalls^{82,83}, Olle Melander⁴⁶, Olli Raitakari^{19,84}, Oscar H. Franco^{8,85}, Oscar L. Rueda-Ochoa^{8,86}, Panos Roussos^{20,87,88}, Peter H. Whincup⁸⁹, Philippe Amouyel^{90,91,92}, Philippe Giral⁹³, Pramod Anugu²⁸, Quenna Wong⁹⁴, Rainer Malik⁷⁸, Rainer Rauramaa^{95,96}, Ralph Burkhardt^{17,97,98}, Rebecca Hardy³⁸, Reinhold Schmidt¹², Renée de Mutsert⁹⁹, Richard W. Morris¹⁰⁰, Rona J. Strawbridge^{44,101}, S. Goya Wannamethee¹⁰², Sara Hägg¹⁰³, Sonia Shah⁴, Stela McLachlan²², Stella Trompet^{23,66}, Sudha Seshadri¹⁰⁴, Sudhir Kurl¹⁰⁵, Susan R. Heckbert^{5,106}, Susan Ring^{107,108}, Tamara B. Harris⁶, Terho Lehtimäki^{109,110}, Tessel E. Galesloot⁷⁴, Tina Shah⁴, Ulf de Faire^{111,112}, Vincent Plagnol¹¹³, Wayne D. Rosamond¹, Wendy Post¹¹⁴, Xiaofeng Zhu⁶⁹, Xiaoling Zhang^{27,115}, Xiuqing Guo^{72,116}, Yasaman Saba¹¹⁷, MEGASTROKE Consortium, Abbas Dehghan^{8,118}, Adrie Seldenrijk¹¹⁹, Alanna C. Morrison³, Anders Hamsten⁴⁴, Bruce M. Psaty^{106,120}, Cornelia M. van Duijn^{8,68}, Deborah A. Lawlor^{107,108}, Dennis O. Mook-Kanamori^{99,121}, Donald W. Bowden¹²², Helena Schmidt¹¹⁷, James F. Wilson^{22,123}, James G. Wilson¹²⁴, Jerome I. Rotter^{72,116}, Joanna M. Wardlaw^{24,125}, John Deanfield⁴, Julian Halcox¹²⁶, Leo-Pekka Lyytikäinen^{109,110}

Markus Loeffler^{16,17}, Michele K. Evans⁶, Stéphanie Debette¹⁸, Steve E. Humphries¹²⁷, Uwe Völker^{34,128},
Vilmundur Gudnason^{10,11}, Aroon D. Hingorani⁴, Johan L.M. Björkegren^{20,21,42,129},
Juan P. Casas⁹ & Christopher J. O'Donnell^{130,131,132}

¹Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27516, USA. ²Department of Pathology and Laboratory Medicine, University of California (UCLA), Los Angeles, Los Angeles, CA 90095, USA. ³Human 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 77030, USA. ⁴Institute of Cardiovascular Science, University College London, London WC1 6BT, UK. ⁵Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA. ⁶Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD 20892, USA. ⁷Department of Genetic Epidemiology, University of Regensburg, Regensburg 93053, Germany. ⁸Department of Epidemiology, Erasmus Medical Center, Rotterdam 3015, The Netherlands. ⁹Institute of Health Informatics, University College London, London WC1E 6BT, UK. ¹⁰Icelandic Heart Association, Kopavogur IS-201, Iceland. ¹¹University of Iceland, Reykjavik 101, Iceland. ¹²Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz 8036, Austria. ¹³Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz 8036, Austria. ¹⁴Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen 3015, The Netherlands. ¹⁵Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA. ¹⁶Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig 04107, Germany. ¹⁷LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig 04107, Germany. ¹⁸Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, CHU Bordeaux, F-33000 Bordeaux, France. ¹⁹Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland. ²⁰Department of Genetics and Genomic Sciences, The Icahn Institute for Genomics and Multiscale Biology Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ²¹Clinical Gene Networks AB, Stockholm 104 62, Sweden. ²²Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, UK. ²³Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. ²⁴Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK. ²⁵Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. ²⁶Population Sciences Branch, Division of Intramural Research, NHLBI, NIH, Framingham, MA 01702-5827, USA. ²⁷National Heart, Lung and Blood Institute's Intramural Research Program, Framingham Heart Study, Framingham, MA 01702-5827, USA. ²⁸Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA. ²⁹Department of Neurology, University Medicine Greifswald, Greifswald 17475, Germany. ³⁰Department of Medicine, Columbia University, New York, NY 10032, USA. ³¹Department of Biochemistry, Maastricht Centre for Systems Biology (MaCSBio), CARIM School for Cardiovascular Diseases, Maastricht University, Maastricht 6229, The Netherlands. ³²Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. ³³Institute for Community Medicine, University Medicine Greifswald, Greifswald 17475, Germany. ³⁴DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald 17475, Germany. ³⁵Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA. ³⁶Menzies Health Institute Queensland, Griffith University, Southport, QLD 4222, Australia. ³⁷Department of Internal Medicine, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam 3015, The Netherlands. ³⁸MRC Unit for Lifelong Health and Ageing at UCL, London WC1E 6BT, UK. ³⁹Department of Medicine, University of Groningen, University Medical Center Groningen, Groningen 2300, The Netherlands. ⁴⁰Department of Epidemiology, and School of Medicine, Division of Geriatric Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA. ⁴¹Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK. ⁴²Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, University of Tartu, Biomeedikum, Tartu 51010, Estonia. ⁴³Department of Cardiac Surgery, Tartu University Hospital, Tartu 51010, Estonia. ⁴⁴Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm 17177, Sweden. ⁴⁵Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala 75108, Sweden. ⁴⁶Department of Clinical Sciences in Malmö, Lund University, Malmö SE-205 02, Sweden. ⁴⁷Department of Human Genetics, University of California (UCLA), Los Angeles, CA 90095, USA. ⁴⁸Department of Psychiatry, EMGO Institute for Health and Care Research and Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam 1081 HL, The Netherlands. ⁴⁹Applied Sciences, Premier, Inc., Charlotte, NC 28277, USA. ⁵⁰Department of Medicine, University of Verona, Verona 37134, Italy. ⁵¹Department of Medical Biotechnology and Translational Medicine, Università di Milano, Milan 20133, Italy. ⁵²Centro Cardiologico Monzino, IRCCS, Milan 20138, Italy. ⁵³St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, MA 02135, USA. ⁵⁴Institute of Laboratory Medicine, University Hospital Munich, LMU Munich 80539, Germany. ⁵⁵Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano, Milan 20133, Italy. ⁵⁶Department of Clinical and Experimental Medicine, Internal Medicine, Angiology and Arteriosclerosis Diseases, University of Perugia, Perugia 06123, Italy. ⁵⁷Centro Diagnostico Italiano, Milan 20147, Italy. ⁵⁸Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030-3411, USA. ⁵⁹Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94309, USA. ⁶⁰Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala 75185, Sweden. ⁶¹Stanford Cardiovascular Institute, Stanford University, Stanford, CA G1120, USA. ⁶²Heart Center Leipzig, Leipzig 04103, Germany. ⁶³Centre for Brain Research, Indian Institute of Science, Bangalore 560012, India. ⁶⁴Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0QQ, UK. ⁶⁵Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK. ⁶⁶Department of Cardiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. ⁶⁷Department of Internal Medicine, Radboud University Medical Center, Nijmegen 6525 GA, The Netherlands. ⁶⁸Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK. ⁶⁹Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA. ⁷⁰Institute for Laboratory Medicine, University of Leipzig, Leipzig 04109, Germany. ⁷¹Department of Biostatistics, University of Washington, Seattle, WA 98105, USA. ⁷²Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA. ⁷³Department of Internal Medicine, Rush University Medical Center, Chicago, IL 60612, USA. ⁷⁴Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, GA 6525, The Netherlands. ⁷⁵Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala 751 05, Sweden. ⁷⁶Department of Genetics, University of North Carolina, Chapel Hill, NC 27516, USA. ⁷⁷Department of Internal Medicine B, University Medicine Greifswald, Greifswald 17475, Germany. ⁷⁸Institute for Stroke and Dementia Research (ISD), University Hospital, Ludwig-Maximilians-University (LMU), Munich 80539, Germany. ⁷⁹Munich Cluster for Systems Neurology (SyNergy), Munich 81377, Germany. ⁸⁰Department of Neurology, Center for Neurology and Neurosurgery, Johann Wolfgang Goethe-University, Frankfurt am Main 60323, Germany. ⁸¹Institute for Social and Economic Research, Essex University, Colchester CO4 3SQ, UK. ⁸²Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA. ⁸³Data Tecnica International, Glen Echo, MD 20812, USA. ⁸⁴Department of Clinical Physiology and Nuclear

Medicine, Turku University Hospital, Turku 20521, Finland. ⁸⁵Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern 3012, Switzerland. ⁸⁶Electrocardiography Research Group, School of Medicine, Universidad Industrial de Santander, Bucaramanga, Santander 680003, Colombia. ⁸⁷Department of Psychiatry and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ⁸⁸Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, New York, NY 10468, USA. ⁸⁹Population Health Research Institute, St George's, University of London, London SW17 0RE, UK. ⁹⁰Inserm U1167, F-59000 Lille, France. ⁹¹Institut Pasteur de Lille, U1167, F-59000 Lille, France. ⁹²Université de Lille, U1167 - RID-AGE & Centre Hospitalier Universitaire de Lille, U1167, F-59000 Lille, France. ⁹³Sorbonne Université, Cardiovascular Prevention Unit, Pitié Salpêtrière Hospital, Paris 75013, France. ⁹⁴Collaborative Health Studies Coordinating Center, Department of Biostatistics, University of Washington, Seattle, WA 98195, USA. ⁹⁵Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio 70100, Finland. ⁹⁶Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio 70210, Finland. ⁹⁷Institute of Laboratory Medicine, University of Leipzig, Leipzig 04109, Germany. ⁹⁸Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg 93053, Germany. ⁹⁹Department of Clinical Epidemiology, Leiden University Medical Center, Leiden 2333, The Netherlands. ¹⁰⁰Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 1QU, UK. ¹⁰¹Mental Health and Wellbeing, Institute of Health and Wellbeing, University of Glasgow, Glasgow G12 0XH, UK. ¹⁰²Department of Primary Care & Population Health, University College London, London WC1E 6BT, UK. ¹⁰³Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden. ¹⁰⁴Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA. ¹⁰⁵Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio Campus, Kuopio FI-70210, Finland. ¹⁰⁶Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA. ¹⁰⁷Population Health Science, Bristol Medical School, University of Bristol, Bristol BS8 1QU, UK. ¹⁰⁸MRC Integrative Epidemiology Unit at the University of Bristol, Bristol BS8 1TH, UK. ¹⁰⁹Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33014, Finland. ¹¹⁰Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere 33014, Finland. ¹¹¹Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm S-171 77, Sweden. ¹¹²Department of Cardiology, Karolinska University Hospital, Stockholm S-171 77, Sweden. ¹¹³Genetics Institute, University College London, London WC1E 6BT, UK. ¹¹⁴Departments of Medicine and Epidemiology, Johns Hopkins University, Baltimore, MD 21205, USA. ¹¹⁵Section of Biomedical Genetics, School of Medicine, Boston University, Boston, MA 02215, USA. ¹¹⁶Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA. ¹¹⁷Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz 8010, Austria. ¹¹⁸Department of Epidemiology & Biostatistics, Imperial College London, London SW7 2AZ, UK. ¹¹⁹GGZ inGeest and Amsterdam Public Health Research Institute, Department of Psychiatry, Amsterdam University Medical Center, Amsterdam 1081 HV, The Netherlands. ¹²⁰Cardiovascular Health Research Unit and Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA 98195, USA. ¹²¹Department of Public Health and Primary Care, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands. ¹²²Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA. ¹²³MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. ¹²⁴Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA. ¹²⁵Centre for Clinical Brain Sciences, and UK Dementia Research Institute at the University of Edinburgh, Edinburgh EH16 4SB, UK. ¹²⁶Swansea University Medical School, Swansea SA2 8PP, UK. ¹²⁷Centre for Cardiovascular Genetics, Institute Cardiovascular Science, University College London, London WC1E 6BT, UK. ¹²⁸Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald 17475, Germany. ¹²⁹Integrated Cardio Metabolic Centre, Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge SE-141 57, Sweden. ¹³⁰Intramural Administration Management Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA. ¹³¹Cardiology Section, Boston Veteran's Administration Healthcare, Boston, MA 02130, USA. ¹³²Harvard Medical School, Boston, MA 02115, USA. These authors contributed equally: Nora Franceschini, Claudia Giambartolomei. A full list of consortium members can be found at the end of the article.

MEGASTROKE Consortium

Yukinori Okada^{133,134,135}, Aniket Mishra^{136,137}, Loes Rutten-Jacobs¹³⁸, Anne-Katrin Giese^{139,140}, Sander W. van der Laan¹⁴¹, Solveig Gretarsdottir¹⁴², Christopher D. Anderson^{140,143,144}, Michael Chong¹⁴⁵, Hieab H.H. Adams⁸, Tetsuro Ago¹⁴⁶, Peter Almgren¹⁴⁷, Philippe Amouyel^{148,149}, Hakan Ay^{140,150}, Traci M. Bartz^{5,71}, Oscar R. Benavente¹⁵¹, Steve Bevan¹⁵², Giorgio B. Boncoraglio¹⁵³, Robert D. Brown Jr.¹⁵⁴, Adam S. Butterworth^{155,156}, Caty Carrera^{157,158}, Cara L. Carty^{159,160}, Daniel I. Chasman^{132,161}, Wei-Min Chen¹⁶², John W. Cole¹⁶³, Ioana Cotlarciuc¹⁶⁴, Carlos Cruchaga^{165,166}, John Danesh^{155,156,167}, Paul I.W. de Bakker^{168,169}, Anita L. DeStefano^{27,170}, Marcel den Hoed¹⁷¹, Qing Duan¹⁷², Stefan T. Engelter^{173,174}, Guido J. Falcone^{144,175}, Rebecca F. Gottesman¹⁷⁶, Raji P. Grewal¹⁷⁷, Stefan Gustafsson⁶⁰, Jeffrey Haessler¹⁷⁸, Tamara B. Harris⁶, Ahamad Hassan¹⁷⁹, Aki S. Havulinna^{180,181}, Elizabeth G. Holliday^{182,183}, George Howard¹⁸⁴, Fang-Chi Hsu¹⁵, Hyacinth I. Hyacinth¹⁸⁵, M. Arfan Ikram⁸, Marguerite R. Irvin¹⁸⁶, Xueqiu Jian¹⁸⁷, Jordi Jiménez-Conde¹⁸⁸, Julie A. Johnson^{189,190}, J. Wouter Jukema⁶⁶, Masahiro Kanai², Keith L. Keene^{191,192}, Brett M. Kissela¹⁹³, Dawn O. Kleindorfer¹⁹³, Charles Kooperberg¹⁷⁸, Michiaki Kubo¹⁹⁴, Leslie Lange¹⁹⁵, Carl D. Langefeld¹⁹⁶, Claudia Langenberg¹⁷², Jin-Moo Lee¹⁹⁷, Robin Lemmens^{198,199}, Didier Leys²⁰⁰, Cathryn M. Lewis^{201,202}, Wei-Yu Lin^{203,204}, Arne G. Lindgren^{205,206}, Erik Lorentzen²⁰⁷, Patrik K. Magnusson¹⁰³, Jane Maguire²⁰⁸, Ani Manichaikul¹⁶², Patrick F. McArdle²⁰⁹, James F. Meschia²¹⁰, Thomas H. Mosley^{211,212}, Toshiharu Ninomiya²¹³, Martin J. O'Donnell^{145,214}, Sara L. Pulit²¹⁵, Kristiina Rannikmäe²¹⁶, Alexander P. Reiner^{178,217}, Kathryn M. Rexrode²¹⁸, Stephen S. Rich¹⁶², Paul M. Ridker^{132,161}, Natalia S. Rost^{139,140}

Peter M. Rothwell²¹⁹, Tatjana Rundek²²⁰, Ralph L. Sacco²²⁰, Saori Sakaue^{3,221}, Michele M. Sale¹⁶², Veikko Salomaa¹⁸⁰, Bishwa R. Sapkota²²², Reinhold Schmidt²²³, Carsten O. Schmidt²²⁴, Ulf Schminke²¹⁷, Pankaj Sharma¹⁶⁴, Agnieszka Slowik²²⁵, Cathie L.M. Sudlow²²⁶, Christian Tanislav²²⁷, Turgut Tatlisumak^{228,229}, Vincent N.S. Thijs^{230,231}, Gudmar Thorleifsson¹⁴², Unnur Thorsteinsdottir¹⁴², Steffen Tiedt⁷⁸, Stella Trompet²³, Matthew Walters²³², Nicholas J. Wareham¹⁷², Sylvia Wassertheil-Smoller²³³, Kerri L. Wiggins⁵, Qiong Yang¹⁷⁰, Salim Yusuf¹⁴⁵, Tomi Pastinen²³⁴, Arno Ruusalepp^{21,42,43}, Eric E. Schadt²⁰, Simon Koplev²⁰, Veronica Codoni^{235,236}, Mete Civelek^{162,237}, Nick Smith^{5,106,238}, David A. Tréguët^{235,236}, Ingrid E. Christophersen^{144,239,240}, Carolina Roselli¹⁴⁴, Steven A. Lubitz^{144,239}, Patrick T. Ellinor^{144,239}, E. Shyong Tai²⁴¹, Jaspal S. Kooner²⁴², Norihiro Kato²⁴³, Jiang He²⁴⁴, Pim van der Harst²⁴⁵, Paul Elliott²⁴⁶, John C. Chambers^{118,247}, Fumihiko Takeuchi²⁴³, Andrew D. Johnson²⁷, Dharambir K. Sanghera^{222,248,249}, Olle Melander⁴⁶, Christina Jern²⁵⁰, Daniel Strbian^{251,252}, Israel Fernandez-Cadenas^{157,158}, W.T. Longstreth Jr^{5,253}, Arndt Rolfs¹⁴⁷, Jun Hata²¹³, Daniel Woo¹⁹³, Jonathan Rosand^{140,143,144}, Guillaume Pare¹⁴⁵, Danish Saleheen²⁵⁴, Kari Stefansson^{142,255}, Bradford B. Worrall²⁵⁶, Steven J. Kittner¹⁶³, Joanna M.M. Howson¹⁵⁵ & Yoichiro Kamatani^{133,257}

¹³³Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan. ¹³⁴Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan. ¹³⁵Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan. ¹³⁶INSERM U1219 Bordeaux Population Health Research Center, Bordeaux F-33000, France. ¹³⁷University of Bordeaux, Bordeaux F-33000, France. ¹³⁸Stroke Research Group, Division of Clinical Neurosciences, University of Cambridge, Cambridge CB2 1TN, UK. ¹³⁹Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. ¹⁴⁰J. Philip Kistler Stroke Research Center, Department of Neurology, MGH, Boston, MA 02215, USA. ¹⁴¹Laboratory of Experimental Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht 3584 CX, Netherlands. ¹⁴²deCODE genetics/AMGEN inc, Reykjavik 101, Iceland. ¹⁴³Center for Genomic Medicine, Massachusetts General Hospital (MGH), Boston, MA 02114, USA. ¹⁴⁴Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA. ¹⁴⁵Population Health Research Institute, McMaster University, Hamilton L8L 2X2, Canada. ¹⁴⁶Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka 819-0935, Japan. ¹⁴⁷Albrecht Kossel Institute, University Clinic of Rostock, Rostock 18147, Germany. ¹⁴⁸INSERM U1167, Institut Pasteur de Lille, Lille F-59000, France. ¹⁴⁹Department of Public Health, Lille University Hospital, Lille F-59000, France. ¹⁵⁰Department of Radiology, Massachusetts General Hospital, Harvard Medical School, AA Martinos Center for Biomedical Imaging, Boston, MA 02129, USA. ¹⁵¹Division of Neurology, Faculty of Medicine, Brain Research Center, University of British Columbia, Vancouver 170-637, Canada. ¹⁵²School of Life Science, University of Lincoln, Lincoln LN6 7TS, UK. ¹⁵³Department of Cerebrovascular Diseases, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano 20133, Italy. ¹⁵⁴Department of Neurology, Mayo Clinic Rochester, Rochester, MN 55905, USA. ¹⁵⁵MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 1TN, UK. ¹⁵⁶The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge CB2 1TN, UK. ¹⁵⁷Neurovascular Research Laboratory, Vall d'Hebron Institut of Research, Neurology and Medicine Departments-Universitat Autònoma de Barcelona, Vall d'Hebrón Hospital, Barcelona 08193, Spain. ¹⁵⁸Stroke Pharmacogenomics and Genetics, Fundació Docència i Recerca MutuaTerrassa, Terrassa 08222, Spain. ¹⁵⁹Children's Research Institute, Children's National Medical Center, Washington, DC 20052, USA. ¹⁶⁰Center for Translational Science, George Washington University, Washington, DC 20052, USA. ¹⁶¹Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA. ¹⁶²Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA 22904-4259, USA. ¹⁶³Department of Neurology, University of Maryland School of Medicine and Baltimore VAMC, Baltimore, MD 21201, USA. ¹⁶⁴Institute of Cardiovascular Research, Royal Holloway University of London, Egham TW20 OEX, UK. ¹⁶⁵Department of Psychiatry, The Hope Center Program on Protein Aggregation and Neurodegeneration (HPAN), Washington University, School of Medicine, St. Louis, MO 98195, USA. ¹⁶⁶Department of Developmental Biology, Washington University School of Medicine, St. Louis, MO 98195, USA. ¹⁶⁷Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. ¹⁶⁸Department of Medical Genetics, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands. ¹⁶⁹Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands. ¹⁷⁰Boston University School of Public Health, Boston, MA 02118, USA. ¹⁷¹Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, Uppsala 751 05, Sweden. ¹⁷²MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0SL, UK. ¹⁷³Department of Neurology and Stroke Center, Basel University Hospital, Basel 4031, Switzerland. ¹⁷⁴Neurorehabilitation Unit, University and University Center for Medicine of Aging and Rehabilitation Basel, Felix Platter Hospital, Basel 4055, Switzerland. ¹⁷⁵Department of Neurology, Yale University School of Medicine, New Haven, CT 06510, USA. ¹⁷⁶Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. ¹⁷⁷Neuroscience Institute, SF Medical Center, Trenton, NJ 08629, USA. ¹⁷⁸Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024, USA. ¹⁷⁹Department of Neurology, Leeds General Infirmary, Leeds Teaching Hospitals NHS Trust, Leeds LS1 3EX, UK. ¹⁸⁰National Institute for Health and Welfare, Helsinki FI-00271, Finland. ¹⁸¹FIMM - Institute for Molecular Medicine Finland, Helsinki FI-00271, Finland. ¹⁸²Public Health Stream, Hunter Medical Research Institute, New Lambton NSW 2305, Australia. ¹⁸³Faculty of Health and Medicine, University of Newcastle, Newcastle 2308, Australia. ¹⁸⁴School of Public Health, University of Alabama at Birmingham, Birmingham, AL 35487, USA. ¹⁸⁵Aflac Cancer and Blood Disorder Center, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA. ¹⁸⁶Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham 35487, USA. ¹⁸⁷Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA. ¹⁸⁸Neurovascular Research Group (NEUVAS), Neurology Department, Institut Hospital del Mar d'Investigació Mèdica, Universitat Autònoma de Barcelona, Barcelona 08193, Spain. ¹⁸⁹Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL 32611, USA. ¹⁹⁰Division of Cardiovascular Medicine, College of

Medicine, University of Florida, Gainesville, FL 32611, USA. ¹⁹¹Department of Biology, East Carolina University, Greenville, NC 27858, USA. ¹⁹²Center for Health Disparities, East Carolina University, Greenville, NC 27858, USA. ¹⁹³University of Cincinnati College of Medicine, Cincinnati, OH 45220, USA. ¹⁹⁴RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan. ¹⁹⁵University of Colorado, Denver, CO 80203, USA. ¹⁹⁶Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA. ¹⁹⁷Department of Neurology, Radiology, and Biomedical Engineering, Washington University School of Medicine, St. Louis, MO 98195, USA. ¹⁹⁸Department of Neurosciences, Experimental Neurology, KU Leuven - University of Leuven, Leuven 3000, Belgium. ¹⁹⁹VIB Center for Brain & Disease Research, University Hospitals Leuven, Department of Neurology, Leuven 3000, Belgium. ²⁰⁰University of Lille, INSERM U1171, CHU Lille, Lille F-59000, France. ²⁰¹Department of Medical and Molecular Genetics, King's College London, London WC2R 2LS, UK. ²⁰²SGDP Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London WC2R 2LS, UK. ²⁰³Cardiovascular Epidemiology Unit, Department Public Health & Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. ²⁰⁴Northern Institute for Cancer Research, Paul O'Gorman Building, Newcastle University, Newcastle NE2 4AD, UK. ²⁰⁵Department of Clinical Sciences Lund, Neurology, Lund University, Lund 221 00, Sweden. ²⁰⁶Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund 222 29, Sweden. ²⁰⁷Bioinformatics Core Facility, University of Gothenburg, Gothenburg 405 30, Sweden. ²⁰⁸University of Technology Sydney, Faculty of Health, Ultimo NSW 2007, Australia. ²⁰⁹Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA. ²¹⁰Department of Neurology, Mayo Clinic, Jacksonville, FL 32224, USA. ²¹¹Division of Geriatrics, School of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA. ²¹²Memory Impairment and Neurodegenerative Dementia Center, University of Mississippi Medical Center, Jackson, FL 39216, USA. ²¹³Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University, Fukuoka 819-0395, Japan. ²¹⁴Clinical Research Facility, Department of Medicine, NUI Galway, Galway H91 TK33, Ireland. ²¹⁵Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht 3584, The Netherlands. ²¹⁶Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH4 2XU, UK. ²¹⁷Department of Neurology, University Medicine Greifswald, Greifswald 17489, Germany. ²¹⁸Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA. ²¹⁹Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford OX3 9DU, UK. ²²⁰Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL 33136, USA. ²²¹Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo 13-8654, Japan. ²²²Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA. ²²³Department of Neurology, Medical University of Graz, Graz 8036, Austria. ²²⁴University Medicine Greifswald, Institute for Community Medicine, SHIP-KEF, Greifswald 17489, Germany. ²²⁵Department of Neurology, Jagiellonian University, Krakow 31-007, Poland. ²²⁶University of Edinburgh, Edinburgh EH8 9JZ, UK. ²²⁷Department of Neurology, Justus Liebig University, Giessen 35390, Germany. ²²⁸Department of Clinical Neurosciences/Neurology, Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg SE-405, Sweden. ²²⁹Sahlgrenska University Hospital, Gothenburg SE-405, Sweden. ²³⁰Stroke Division, Florey Institute of Neuroscience and Mental Health, Heidelberg VIC 3084, Australia. ²³¹Austin Health, Department of Neurology, Heidelberg, Victoria 3084, Australia. ²³²School of Medicine, Dentistry and Nursing at the University of Glasgow, Glasgow G12 8QQ, UK. ²³³Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA. ²³⁴Department of Human Genetics, McGill University, Montreal H3A 0G4, Canada. ²³⁵Sorbonne Universités, UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Paris 75006, France. ²³⁶ICAN, Institute for Cardiometabolism and Nutrition, Paris 75013, France. ²³⁷Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22904-4259, USA. ²³⁸Seattle Epidemiologic Research and Information Center, VA Office of Research and Development, Seattle, WA 98108, USA. ²³⁹Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA. ²⁴⁰Department of Medical Research, Bærum Hospital, Vestre Viken Hospital Trust, Rud 3004, Norway. ²⁴¹Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore 119077, Singapore. ²⁴²National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK. ²⁴³Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo 162-8655, Japan. ²⁴⁴Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA 70112, USA. ²⁴⁵Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen 9700 RB, Netherlands. ²⁴⁶Department of Epidemiology and Biostatistics, Imperial College London, MRC-PHE Centre for Environment and Health, School of Public Health, London W2 1PG, UK. ²⁴⁷Department of Cardiology, Ealing Hospital NHS Trust, Southall HA1 3UJ, UK. ²⁴⁸Department of Pharmaceutical Sciences, College of Pharmacy, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA. ²⁴⁹Oklahoma Center for Neuroscience, Oklahoma City, OK 73104, USA. ²⁵⁰Department of Pathology and Genetics, Institute of Biomedicine, The Sahlgrenska Academy at University of Gothenburg, Gothenburg SE-405, Sweden. ²⁵¹Department of Neurology, Helsinki University Hospital, Helsinki FI-00029, Finland. ²⁵²Clinical Neurosciences, Neurology, University of Helsinki, Helsinki FI-00029, Finland. ²⁵³Department of Neurology, University of Washington, Seattle, WA 98195, USA. ²⁵⁴Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. ²⁵⁵Faculty of Medicine, University of Iceland, Reykjavik 201, Iceland. ²⁵⁶Departments of Neurology and Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA 22908, USA. ²⁵⁷Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan