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

MEGASTROKE Consortium; Franceschini, Nora; Giambartolomei, Claudia; de Vries, Paul S.; Finan, Chris; Bis, Joshua C.; Huntley, Rachael P.; Lovering, Ruth C.; Tajuddin, Salman M.; Winkler, Thomas W.

Published in:
Nature Communications

DOI:
10.1038/s41467-018-07340-5

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
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.
Atherosclerosis is characterized by an accumulation of lipid-rich and inflammatory deposits (plaques) in the subintimal 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 US. Because atherosclerosis has a long pre-clinical phase, early detection of atherosclerosis using non-invasive methods may help identify individuals at risk for atherothrombotic clinical events, 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 plaques.

Subclinical and clinical atherosclerosis has known genetic components. Genome-wide association studies (GWAS) of subclinical atherosclerosis have previously identified three loci significantly associated with cIMT at ZHX2, APOCI, and PINX1, and two loci associated with common carotid artery plaque at PIK3CG and EDNRA. An exome-wide-association study identified significant associations of the APOE e2 allele with cIMT and coronary artery calcification. The APOE single nucleotide polymorphism (SNP) rs7412 is in linkage disequilibrium (LD) with the APOCI variant, thus representing the same signal. Additional GWAS-identified associations were reported for carotid plaque at the 9p21 and SEXN2 loci, and for cIMT at the CFDP1-TMEM170A locus. 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 atherothrombotic 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) consortium and the University College London-Edinburgh-Bristol (UCLEB) consortium. 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 colocalization. These methods integrate for established and novel loci of cIMT and carotid plaque using non-invasive methods may help identify individuals at risk for atherothrombotic clinical events, 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 plaques.

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 × 10−8), of which eight were newly described and three have been previously reported (Table 1). The closest genes for the eight loci were: lq32 intergenic (rs201648240), ATP6AP1L1 (rs224904), AIG1 (rs6907215), PIK3CG (rs13225723), MCPH1 (rs2912063), SGK223 (rs11785239), VTN1 (rs1196033), and CFA2A73 (rs344396). For carotid plaque, three loci 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 plaque. The two

![Fig. 1](https://example.com/image1.png)

**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 PINXI 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.

**Table 1** Loci significantly associated with cIMT and plaque GWAS

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chrposition</th>
<th>Nearest coding gene</th>
<th>Alleles (effect/other)</th>
<th>Effect allele freq.</th>
<th>Beta (SE)</th>
<th>p</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs201648240</td>
<td>1:208953176-indel</td>
<td>LINC01717</td>
<td>/A</td>
<td>0.83</td>
<td>-0.0062 (0.0011)</td>
<td>( 4 \times 10^{-9} )</td>
<td>54,752</td>
</tr>
<tr>
<td>rs224904</td>
<td>5:81637916</td>
<td>ATP6AP1</td>
<td>C/G</td>
<td>0.95</td>
<td>-0.0088 (0.0016)</td>
<td>( 5 \times 10^{-8} )</td>
<td>68,962</td>
</tr>
<tr>
<td>rs6907215</td>
<td>6:143608988</td>
<td>AIG1</td>
<td>T/C</td>
<td>0.60</td>
<td>-0.0040 (0.0007)</td>
<td>( 5 \times 10^{-8} )</td>
<td>64,586</td>
</tr>
<tr>
<td>rs13225723</td>
<td>7:106416467</td>
<td>PIK3CG</td>
<td>A/G</td>
<td>0.22</td>
<td>0.0052 (0.0009)</td>
<td>( 3 \times 10^{-9} )</td>
<td>68,070</td>
</tr>
<tr>
<td>rs2912063</td>
<td>8:6486033</td>
<td>MCPH1</td>
<td>A/G</td>
<td>0.71</td>
<td>0.0045 (0.0008)</td>
<td>( 9 \times 10^{-9} )</td>
<td>67,401</td>
</tr>
<tr>
<td>rs1785239</td>
<td>8:8205010</td>
<td>SGK222</td>
<td>T/C</td>
<td>0.65</td>
<td>-0.0043 (0.0008)</td>
<td>( 9 \times 10^{-9} )</td>
<td>67,107</td>
</tr>
<tr>
<td>rs11196033</td>
<td>10:114410998</td>
<td>VTI1A</td>
<td>A/C</td>
<td>0.48</td>
<td>0.0042 (0.0008)</td>
<td>4 \times 10^{-8}</td>
<td>57,995</td>
</tr>
<tr>
<td>rs844396</td>
<td>16:8896667</td>
<td>CBFA2T3</td>
<td>T/C</td>
<td>0.30</td>
<td>-0.0051 (0.0009)</td>
<td>( 6 \times 10^{-9} )</td>
<td>50,377</td>
</tr>
<tr>
<td>rs200495339</td>
<td>19:11189298-indel</td>
<td>LDLR</td>
<td>/G</td>
<td>0.11</td>
<td>-0.0103 (0.0179)</td>
<td>1 \times 10^{-8}</td>
<td>36,569</td>
</tr>
</tbody>
</table>

\( p = p \)-values of association from linear regression analysis; N = total number in meta-analyses

\( ^* \)Published cIMT SNP in LD with our most significant SNP: rs178551 (\( p = 0.95 \) with rs148147734), rs6601530 (\( p = 0.0 \) with rs200482500), and rs445925 (\( p = 0.60 \) with rs7412)

\( ^* \)Published plaque SNP in LD with our most significant SNP: rs1878406 (\( p = 0.98 \) with rs14137744), rs17398575 (\( p = 0.8 \) with rs17477177), rs9644682 (\( p = 0.79 \) with rs9632884), and rs48868378 (\( p = 0.94 \) with rs113309775).

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 study \(^10\).

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 locus \(^10\). 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...
probability of no colocalization of the same SNP with tissue eQTLs. Red indicates a high posterior probability of colocalization and blue a high probability of no colocalization of the SNP with tissue eQTLs.

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

<table>
<thead>
<tr>
<th>SNP</th>
<th>eQTL(^a) (Gene, p) GTEx</th>
<th>eQTL(^b) (Gene, p) STARNET tissues</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs201648240</td>
<td>CAMK1G, 0.0094</td>
<td>CD34, 0.00532</td>
</tr>
<tr>
<td></td>
<td>AL031361, 0.0040</td>
<td>TRAF3IP3, 0.0097</td>
</tr>
<tr>
<td>rs6907215</td>
<td>AL023584.1, 0.005384704 (VEN)</td>
<td>ENSG00000217648, 0.00046</td>
</tr>
<tr>
<td>rs13225723</td>
<td>ENSG00000177820.5, 7.0 × 10(^{-5})</td>
<td>CCDC71L, 6 × 10(^{-3})</td>
</tr>
<tr>
<td></td>
<td>PRKAR2B, 4 × 10(^{-8})</td>
<td>PRKAR2B, 6 × 10(^{-8})</td>
</tr>
<tr>
<td></td>
<td>PIK3CG, 10 × 10(^{-3})</td>
<td>SYPL1, 0.0043</td>
</tr>
<tr>
<td>rs2912063</td>
<td>ENSG00000271743.1, 0.0093 (VEN)</td>
<td>MCHP1-AS1, 0.0020</td>
</tr>
<tr>
<td>rs11785239</td>
<td>ENSG00000267052.1, 6 × 10(^{-11})</td>
<td>ERII, 0.0069</td>
</tr>
<tr>
<td>rs844396</td>
<td>AC92284.2, 0.001</td>
<td>PPP1R3B, 0.0036</td>
</tr>
<tr>
<td>rs7412</td>
<td>ENSG00000267163.1, 0.007</td>
<td>TRAPPCL2L, 0.0040</td>
</tr>
<tr>
<td>rs1413744</td>
<td>PMT9, 0.004</td>
<td>ZNF276, 0.0070</td>
</tr>
<tr>
<td>rs17477177</td>
<td>CCDC71L, 6 × 10(^{-3})</td>
<td>TRAPPCL2L, 0.0091</td>
</tr>
<tr>
<td>rs9632884</td>
<td>BCAP29, 0.002</td>
<td>ZNF276, 0.0059</td>
</tr>
<tr>
<td>rs113309773</td>
<td>BCAR1, 6 × 10(^{-11})</td>
<td>TRAPPCL2L, 0.0059</td>
</tr>
<tr>
<td></td>
<td>ENSG0000026783.1, 2 × 10(^{-16})</td>
<td>ZNF276, 0.0059</td>
</tr>
<tr>
<td></td>
<td>GABARAPL2, 0.004</td>
<td>TRAPPCL2L, 0.0091</td>
</tr>
</tbody>
</table>

\(p\) = \(p\)-values of association from linear regression analysis

\(^a\)The 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 (\(p < 10^{-8}\)).

\(^b\)This includes heart atrial (ATR) and heart left ventricle (VEN).

\(^c\)This includes aorta (AOR).

\(^d\)This includes heart atrial (ATR) and heart left ventricle (VEN).

\(\text{SNP eQTL} \text{ GTEx}\)

EQUIL-GWAS dataset pair (Supplementary Table 6). A posterior probability of ≥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 (>98%) 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%).

---

**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.
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). Attributing 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, \( p = 0.0033 \)) and rs55917128, chr10:99017729-101017321, LOXL4, \( p = 0.0003 \).

### 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 ≥20%, Supplementary Data 1). In contrast, there was strong evidence (≥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).

<table>
<thead>
<tr>
<th>Region (chr:start-stop)</th>
<th>Trait</th>
<th>Gene</th>
<th>SNP with best joint probability</th>
<th>( p, \text{BETA (SE)} ), Tissue posterior probability (PPA)</th>
<th>Direction of effect GWAS/eQTL</th>
<th>( \text{PPA} ) of colocalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr3:63560280-65833136</td>
<td>cIMT</td>
<td>ADAMTS9</td>
<td>rs17676309 (T/C)</td>
<td>2 × 10^{-6}, -0.0035 (0.0007)</td>
<td>AOR eQTL</td>
<td>-0.61 (0.06)</td>
</tr>
<tr>
<td>chr10:99017729-101017321</td>
<td>Plaque</td>
<td>LOXL4</td>
<td>rs55917128 (T/C)</td>
<td>5 × 10^{-7}, 0.0037 (0.0007)</td>
<td>AOR eQTL</td>
<td>-0.69 (0.06)</td>
</tr>
<tr>
<td>chr7:105299372-107743409</td>
<td>cIMT</td>
<td>CCDC71L</td>
<td>rs12705390 (A/G)</td>
<td>5 × 10^{-9}, 0.0049 (0.0008)</td>
<td>AOR eQTL</td>
<td>-0.75 (0.06)</td>
</tr>
<tr>
<td>chr3:63560280-65833136</td>
<td>cIMT</td>
<td>ADAMTS9</td>
<td>rs17676309 (T/C)</td>
<td>4 × 10^{-8}, 0.12 (0.022)</td>
<td>AOR eQTL</td>
<td>-0.79 (0.06)</td>
</tr>
</tbody>
</table>

PPA posterior probability of sharing same SNP higher than 75%, cIMT common carotid artery intima-media thickness, AOR aorta, MAM mammary artery.

*This signal reaches genome-wide significance in cIMT/plaque, and reaches a high probability of being mediated by the genes in AOR and MAM.

### 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 druggable (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 approved clinical 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 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 data. 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.
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 subclinical 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 (*LOXL4*). *ADAMTS9* is a metalloproteinase involved in thrombosis 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. *LOXL4* 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 models.

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

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

\(^a\)CARDIoGRAMPlusC4D
\(^b\)MEGASTROKE 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 is 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 atherosclerosis. 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 outcomespecificity. 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 pathway.

Measures of cIMT and carotid plaque reflect vascular pathophysiologic 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 CHD. This gene encodes a protein involved in cell–cell junctions in endothelial cells, which was recently shown to be involved in pathologic angiogenic process in vitro and in vivo experimental models. 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 molcor and eCAVIAR 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 CHARGE 22 and the UCLeB consortia,23 for genetic studies of CHD 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 reported. 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/ miniMaCH. We 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–Weiner 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/G/C/T/D/I markers; duplicate markers with low call rate; monomorphic SNPs and those with low frequency 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 r2 ≥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 as p < 5.0 × 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 (colco)29. 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 STARNET study, including blood, MAM, AOR, subcutaneous fat (SF), visceral fat (VAF), skeletal muscle (SKLM), and liver (LIV) obtained from 600 patients during open heart surgery31. 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 to form 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 consortia39. 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 moloc24 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 Project20,31, 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 DNase Hypersensitivity sites (DHS) for nine cells lines. From FANTOM33, 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 IGVAS13 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 estimation to identify a set of annotations that significantly improve the model fit, and reverse selection of each annotation included in the model, as suggested in the IGVAS 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 and stroke subtypes: cardioembolic stroke (CE, large vessel disease, and SVD) with stroke traits was calculated using LD-score, which uses GWAS summary statistics and is not affected by sample overlap. This method relies on the fact that the χ2 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 heritabilities25. Genetic correlations between stroke subtypes (CE, IS,CE, large vessel disease, SVD) and carotid plaque were calculated using software available at http://github.com/bulik/dscl with GWAS summary statistics for our cIMT/carotid plaque GWAS, CARDIOGRAMplusC4D data, and stroke GWAS. We used the LD-scores35, which are based on the 1000 Genomes European population and estimated within 1-cM windows based on ten ten-times repeated (two subclinical traits and five outcomes), we estimated the significance threshold to p = 0.005.

**Methods for GO Slim 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 r2 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 passed the SNP selection 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: 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-therapeutics36. 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 significant sequence similarity to approved drug targets. Tier 3 contained 2370 genes encoding secreted or extracellular proteins, proteins with more distant similarity to known bioactive drug-like small molecule binding partners and those with significant sequence 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://gexportal.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.cardiogramusc4d.org/. For LD scores, see broadinstitute.org/~bulik/eur/ldevscore/. For UniProt ID, see uniprot.org/uploadlists/. For QuickGO, see www.ebi.ac.uk/QuickGO. For VLAD tool, see http://proto.informatics.jax.org/

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 accession number 365 (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

References

Acknowledgements
The work was supported by the following grants: National Institute of Health grants: R21HL1123677, R21-HL140385, DK104806-01A1, R01-MD012765-01A1 (NF), National Institutes of Health awards R01HG009120, R10HG006389, U01CA194935, T32NS048004 (CG), the American Heart Association Grant 17POST33350042 (PY), the British Heart Foundation (RG/135/30112) and the National Institute for Health Research University College London Hospitals Biomedical Research Centre (RCL and RPH), the British Heart Foundation FS/14/55/10806 (JCH), the German Federal Ministry of Education and Research (BMBF) in the context of the e:Ather-oSysMed, the DFG as part of the CRC 1123 (BS), and the FP7/2007-2013 European Union project CYgenes@target (grant number 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 acknowledgments are included in Supplementary Note 2.

Author contributions


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/Malaise. E.I. is a scientist 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.

MEGASTROKE Consortium

Yukinori Okada\textsuperscript{133,134,135}, Aniket Mishra\textsuperscript{136,137}, Loes Rutten-Jacobs\textsuperscript{138}, Anne-Katrin Giese\textsuperscript{139,140}, Sander W. van der Laan\textsuperscript{141}, Solveig Gretarsdottir\textsuperscript{142}, Christopher D. Anderson\textsuperscript{140,143,144}, Michael Chong\textsuperscript{145}, Hieab H.H. Adams\textsuperscript{8}, Tetsuro Ago\textsuperscript{146}, Peter Almgren\textsuperscript{147}, Philippe Amouyel\textsuperscript{148,149}, Hakan Ay\textsuperscript{140,150}, Traci M. Bartz\textsuperscript{5,71}, Oscar R. Benavente\textsuperscript{151}, Steve Bevan\textsuperscript{152}, Giorgio B. Boncoraglio\textsuperscript{153}, Robert D. Brown Jr.\textsuperscript{154}, Adam S. Butterworth\textsuperscript{155,156}, Caty Carrera\textsuperscript{157,158}, Cara L. Carty\textsuperscript{159,160}, Daniel I. Chasman\textsuperscript{132,161}, Wei-Min Chen\textsuperscript{162}, John W. Cole\textsuperscript{163}, Ioana Cotlaruc\textsuperscript{164}, Carlos Cruchaga\textsuperscript{165,166}, John Danesh\textsuperscript{155,156,167}, Paul I.W. de Bakker\textsuperscript{168,169}, Anita L. DeStefano\textsuperscript{27,170}, Marcel den Hoed\textsuperscript{171}, Ping Duan\textsuperscript{172}, Stefan T. Engelter\textsuperscript{173,174}, Guido J. Falcone\textsuperscript{144,175}, Rebecca F. Gottesman\textsuperscript{176}, Raji P. Grewal\textsuperscript{177}, Stefan Gustafsson\textsuperscript{160}, Jeffrey Haessler\textsuperscript{178}, Tamara B. Harris\textsuperscript{6}, Ahamad Hassan\textsuperscript{179}, Aki S. Havulinna\textsuperscript{180,181}, Elizabeth G. Holloway\textsuperscript{182,183}, George Howard\textsuperscript{184}, Fang-Chi Hsu\textsuperscript{15}, Hyacinth I. Hyacinth\textsuperscript{185}, M. Arfan Ikram\textsuperscript{8}, Marguerite R. Irvin\textsuperscript{186}, Xueqiu Jian\textsuperscript{187}, Jordi Jiménez-Conde\textsuperscript{188}, Julie A. Johnson\textsuperscript{189,190}, J. Wouter Jukema\textsuperscript{186}, Masahiro Kanai\textsuperscript{2}, Keith L. Keene\textsuperscript{191,192}, Brett M. Kissela\textsuperscript{193}, Dawn O. Kleindorfer\textsuperscript{193}, Charles Kooperberg\textsuperscript{178}, Michiaki Kubo\textsuperscript{194}, Leslie Lange\textsuperscript{195}, Carl D. Langefeld\textsuperscript{196}, Claudia Langenberg\textsuperscript{172}, Jin-Moo Lee\textsuperscript{197}, Robin Lemmens\textsuperscript{198,199}, Didier Leys\textsuperscript{200}, Cathryn M. Lewis\textsuperscript{201,202}, Wei-Yu Lin\textsuperscript{203,204}, Arne G. Lindgren\textsuperscript{205,206}, Erik Lorentzen\textsuperscript{207}, Patrik K. Magnusson\textsuperscript{103}, Jane Maguire\textsuperscript{208}, Ani Manichaikul\textsuperscript{162}, Patrick F. McArdele\textsuperscript{209}, James F. Meschia\textsuperscript{210}, Thomas H. Mosley\textsuperscript{211,212}, Toshiharu Ninomiya\textsuperscript{213}, Martin J. O’Donnell\textsuperscript{145,214}, Sara L. Pulin\textsuperscript{215}, Kristiina Rannikmäe\textsuperscript{216}, Alexander P. Reiner\textsuperscript{178,217}, Kathryn M. Rexrode\textsuperscript{218}, Stephen S. Rich\textsuperscript{162}, Paul M. Ridker\textsuperscript{132,161}, Natalia S. Rost\textsuperscript{139,140}, Wei-Yu Lin\textsuperscript{203,204}, Arne G. Lindgren\textsuperscript{205,206}, Erik Lorentzen\textsuperscript{207}, Patrik K. Magnusson\textsuperscript{103}, Jane Maguire\textsuperscript{208}, Ani Manichaikul\textsuperscript{162}, Patrick F. McArdele\textsuperscript{209}, James F. Meschia\textsuperscript{210}, Thomas H. Mosley\textsuperscript{211,212}, Toshiharu Ninomiya\textsuperscript{213}, Martin J. O’Donnell\textsuperscript{145,214}, Sara L. Pulin\textsuperscript{215}, Kristiina Rannikmäe\textsuperscript{216}, Alexander P. Reiner\textsuperscript{178,217}, Kathryn M. Rexrode\textsuperscript{218}, Stephen S. Rich\textsuperscript{162}, Paul M. Ridker\textsuperscript{132,161}, Natalia S. Rost\textsuperscript{139,140},