nature genetics

Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms

Joanna M M Howson[1], Wei Zhao[2,64], Daniel R Barnes[1,64], Weang-Kee Ho[1,3], Robin Young[1,4], Dirk S Paul[1], Lindsay L Waite[5], Daniel F Freitag[1], Eric B Fauman[6], Elias L Salfati[7,8], Benjamin B Sun[1], John D Eicher[9,10], Andrew D Johnson[9,10], Wayne H H Sheu[11–13], Sune F Nielsen[14], Wei-Yu Lin[1,15], Praveen Surendran[1], Anders Malarstig[16], Jemma B Wilk[17], Anne Tybjærg-Hansen[18,19], Katrine L Rasmussen[14], Pia R Kamstrup[14], Panos Deloukas[20,21], Jeanette Erdmann[22–24], Sekar Kathiresan[25,26], Nilesh J Samani[27,28], Heribert Schunkert[29,30], Hugh Watkins[31,32], CARDIoGRAMplusC4D[33], Ron Do[34], Daniel J Rader[35], Julie A Johnson[36], Stanley L Hazen[37], Arshed A Quyyumi[38], John A Spertus[39,40], Carl J Pepine[41], Nora Franceschini[42], Anne Justice[42], Alex P Reiner[43], Steven Buyske[44], Lucia A Hindorff[45], Cara L Carty[46], Kari E North[42,47], Charles Kooperberg[46], Eric Boerwinkle[48,49], Kristin Young[42], Mariaelisa Graff[42], Ulrike Peters[46], Devin Absher[5], Chao A Hsiung[50], Wen-Jane Lee[51], Kent D Taylor[52], Ying-Hsiang Chen[50], I-Te Lee[53], Xiuqing Guo[52], Ren-Hua Chung[50], Yi-Jen Hung[13,54], Jerome I Rotter[55], Jyh-Ming J Juang[56,57], Thomas Quertermous[7,8], Tzung-Dau Wang[56,57], Asif Rasheed[58], Philippe Frossard[58], Dewan S Alam[59], Abdulla al Shafi Majumder[60], Emanuele Di Angelantonio[1,61], Rajiv Chowdhury[1], EPIC-CVD[33], Yii-Der Ida Chen[52], Børge G Nordestgaard[14,19], Themistocles L Assimes[7,8,64], John Danesh[1,61–64], Adam S Butterworth[1,61,64] & Danish Saleheen[1,2,58,64]

Coronary artery disease (CAD) is a leading cause of morbidity and mortality worldwide[1,2]. Although 58 genomic regions have been associated with CAD thus far[3–9], most of the heritability is unexplained[9], indicating that additional susceptibility loci await identification. An efficient discovery strategy may be larger-scale evaluation of promising associations suggested by genome-wide association studies (GWAS). Hence, we genotyped 56,309 participants using a targeted gene array derived from earlier GWAS results and performed meta-analysis of results with 194,427 participants previously genotyped, totaling 88,192 CAD cases and 162,544 controls. We identified 25 new SNP–CAD associations ($P < 5 \times 10^{-8}$, in fixed-effects meta-analysis) from 15 genomic regions, including SNPs in or near genes involved in cellular adhesion, leukocyte migration and atherosclerosis (*PECAM1*, rs1867624), coagulation and inflammation (*PROCR*, rs867186 (p.Ser219Gly)) and vascular smooth muscle cell differentiation (*LMOD1*, rs2820315). Correlation of these regions with cell-type-specific gene expression and plasma protein levels sheds light on potential disease mechanisms.

The CardioMetabochip is a genotyping array that contains 196,725 variants of confirmed or suspected relevance to cardiometabolic traits derived from earlier GWAS[10]. A previous meta-analysis by the CARDIoGRAMplusC4D consortium of 79,138 SNPs common to the CardioMetabochip and GWAS arrays identified 15 new loci associated with CAD[3]. Using the CardioMetabochip, we genotyped 56,309 additional samples of European (EUR; ~52%), South Asian (SAS; ~23%), East Asian (EAS; ~17%) and African-American (AA; ~8%) ancestry (**Supplementary Fig. 1**, **Supplementary Tables 1–3** and **Supplementary Note**). The results from our association analyses of these additional samples were subjected to meta-analysis with those reported by CARDIoGRAMplusC4D at 79,070 SNPs in two fixed-effects meta-analyses, one in European participants and a second across all four ancestry groups (**Figs. 1** and **2**). (Overlapping samples were removed before meta-analysis; Online Methods.) A genome-wide significance threshold ($P \leq 5 \times 10^{-8}$ in the fixed-effects meta-analysis) was adopted to minimize false positive findings, but there is still a small chance of a false positive result. The European fixed-effects meta-analysis identified 15 SNPs associated with CAD at genome-wide significance ($P < 5 \times 10^{-8}$) from nine distinct genomic regions that are not established CAD-associated loci (**Table 1**, **Supplementary Fig. 2** and **Supplementary Table 4**). Six additional distinct new CAD-associated regions were identified in the all-ancestry fixed-effects meta-analysis (**Fig. 2**, **Table 1** and **Supplementary Table 4**). In total, 15 new CAD-associated genomic regions (25 SNPs) were identified (**Supplementary Figs. 3** and **4**). The lead SNPs had at least nominal evidence of association ($P < 0.05$) in either a fixed-effects meta-analysis of the European studies with *de novo* genotyping or in a fixed-effects meta-analysis of all the studies with *de novo*

A full list of affiliations appears at the end of the paper.

Received 28 November 2016; accepted 26 April 2017; published online 22 May 2017; doi:10.1038/ng.3874

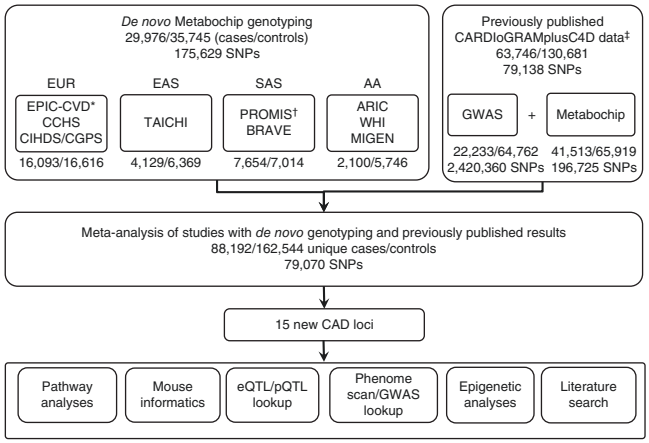

Figure 1 Schematic of the study design. The sample size information is provided as number of cases/number of controls. Note that samples with *de novo* genotyping that were also in the CARDIoGRAMplusC4D study were removed before meta-analysis. *, 1,826 CAD cases and 449 controls from EPIC-CVD with *de novo* genotyping were also included in CARDIoGRAMplusC4D and were therefore excluded from the larger meta-analysis. The actual number of European individuals contributed to the meta-analysis of our studies with *de novo* genotyping and the CARDIoGRAMplusC4D was 14,267 CAD cases and 16,167 controls. †, 3,704 CAD cases and 3,433 controls from PROMIS with *de novo* genotyping were also included in CARDIoGRAMplusC4D and were therefore excluded from the larger meta-analysis. The actual number of South Asian samples contributed to the meta-analysis of our studies with *de novo* genotyping and the CARDIoGRAMplusC4D was 3,950 CAD cases and 3,581 controls. CAD, coronary artery disease; EUR, European ancestry; EAS, East Asian ancestry; SAS, South Asian ancestry; AA, African-American ancestry.

genotyping (**Supplementary Fig. 5** and **Supplementary Table 5**). Within the CARDIoGRAMplusC4D results for these SNPs, there was no evidence of heterogeneity of effects ($P_{het} \geq 0.10$) and allele frequencies were consistent with our European studies (**Supplementary Table 5**). Tests for enrichment of CAD associations within sets of genes[11] and Ingenuity Pathway Analysis confirmed known CAD pathways (**Supplementary Tables 6–8** and **Supplementary Note**).

To prioritize candidate causal genes at the new loci, we defined regions encompassing the new CAD-associated SNPs on the basis of recombination rates (**Supplementary Table 9**) and cross-referenced them with expression quantitative trait locus (eQTL) databases including GTEx[12], MuTHER[13] and STARNET[14] (Online Methods). Twelve of the 15 new CAD-associated SNPs were identified as potential eQTLs in at least one tissue ($P < 5 \times 10^{-8}$; **Table 2** and **Supplementary Table 10**). HaploReg analysis[15] (Online Methods) showed that CAD-associated SNPs were enriched for H3K27ac enhancer marks ($P < 5.1 \times 10^{-4}$) in multiple heart-related tissues (left ventricle, right atrium, aorta) in the European results and in one heart-related tissue (right atrium) and liver in the all-ancestry analyses (**Supplementary Table 11**). We next tested for protein quantitative trait loci (pQTLs) in plasma on the aptamer-based Somalogic platform (Online Methods). Twenty-four proteins from the newly identified CAD regions were assayed and passed quality control. Of our 15 new CAD-associated SNPs, 2 associated with plasma protein abundance in *trans*: rs867186 (NP_006395.2:p.Ser219Gly), a missense variant in *PROCR*, was a *trans*-pQTL for protein C ($P = 1 \times 10^{-10}$) and rs1050362 (NP_054722.2:p.Arg140 =), a synonymous variant in *DHX38*, was a *trans*-pQTL for apolipoprotein L1 ($P = 5.37 \times 10^{-29}$; Online Methods), which is suggested to interact with HPR in the *DHX38* region (String database).

To further help prioritize candidate genes, we also queried the Mouse Genome Informatics database to discover phenotypes resulting from mutations in the orthologous genes for all genes in our 15 CAD-associated regions (**Table 2**). To understand the pathways by which our new loci might be related to CAD risk, we examined the associations of the 15 new CAD regions with a wide range of risk factors, molecular traits and clinical disorders, using PhenoScanner[16] (which encompasses the NHGRI-EBI GWAS catalog and other genotype–phenotype databases).

Six of our loci have previously been associated with known CAD risk factors, such as major lipids (*PCNX3* (ref. 17), *C12orf43–HNF1A*, *SCARB1*, *DHX38*) (ref. 18)) and blood pressure (*GOSR2* (ref. 19), *PROCR*[20]). The sentinel variants for the CAD and risk factor associations at *PCNX3*, *GOSR2* and *PROCR* were the same, implicating them in known biological pathways. Two correlated SNPs ($r^2 = 0.93$, $D' = 1.0$ in 1000 Genomes Project European samples) rs11057830 and rs11057841 tag the CAD association in the *SCARB1* region (**Table 1** and **Supplementary Table 4**), a region reported previously to be associated with HDL (rs838876, $\beta = -0.049$, $P = 7.33 \times 10^{-33}$)[18]. A rare nonsynonymous variant, rs74830677 (NP_005496.4:p.Pro376Leu) in *SCARB1* was also associated with high levels of HDL cholesterol (HDL-C)[21]. Conditional analyses showed that the CAD association was independent of the common variant HDL association (**Supplementary Fig. 6** and **Supplementary Note**). We found that the CAD SNPs and the common HDL-C SNP rs838880 overlap enhancers active in primary liver tissue (**Supplementary Fig. 7**). *SCARB1* is highly expressed in liver and adrenal gland tissues (GTEx; **Supplementary Fig. 7**)[12]. These findings suggest that the discovered genetic variants most likely have a role in regulation of the liver-restricted expression of *SCARB1*.

The *DHX38* region has previously been associated with increased total and LDL cholesterol (LDL-C)[18]. Both CAD-associated SNPs in *DHX38*, rs1050362 (NP_054722.2:p.Arg140 =) and rs2072142 (synonymous and intronic, respectively; **Table 1** and **Supplementary Table 4**) are in linkage disequilibrium (LD) but are not strongly correlated with the previously reported cholesterol-increasing SNP, intronic in *HPR*, rs2000999 ($r^2 = 0.41$, $D' = 1$ in 1000 Genomes Project European samples). Deletions in the *HP* gene have recently been shown to drive the reported cholesterol association in this region[22]. The CAD-associated SNPs are in strong LD with SNPs that increase haptoglobin levels[23] (rs6499560, $P = 2.92 \times 10^{-13}$, $r^2 = 0.97$), and haptoglobin has been reported to be associated with increased CAD risk[24]. *HP* encodes an α2 glycoprotein that is synthesized in the liver. It binds free hemoglobin and protects tissues from oxidative damage. Mouse models indicate the role of *Hp* in the development of atherosclerosis[25], where the underlying mechanism is disruption of the protective nature of the HP protein against hemoglobin-induced injury of atherosclerotic plaque. While the CAD-associated SNPs are eQTLs (or in LD with eQTLs) for multiple genes in the region (for example, *DHODH* in aorta artery[12] (rs1050362[A] allele, $\beta = 0.41$, $P = 1.4 \times 10^{-9}$), *DHX38* in peripheral blood[26] and atherosclerotic aortic root[14] ($P < 8 \times 10^{-26}$; **Table 2** and **Supplementary Table 10**)), the A allele at rs1050362 is also associated with increased expression of *HP* in heart left ventricle ($\beta = 0.535$, $P = 8.71 \times 10^{-10}$)[12] and decreased expression of *HP* in whole blood ($\beta = -0.27$, $P = 1.22 \times 10^{-10}$)[12]. While there could be multiple causal genes in the region, together these findings suggest that *HP* is a promising candidate gene.

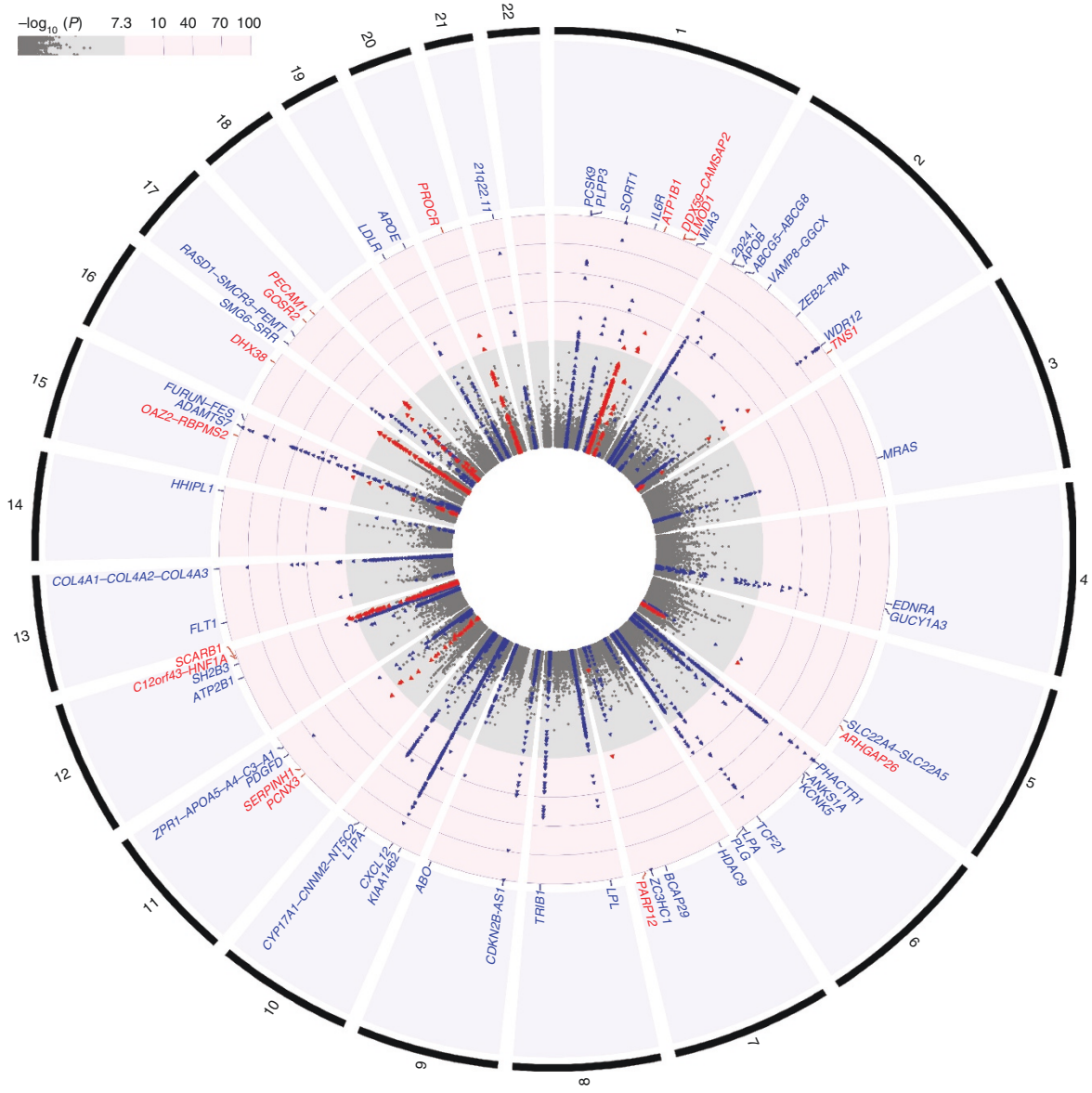

Figure 2 Plot showing the association of ~79,000 variants with CAD ($-\log_{10}P$) in up to 88,192 cases and 162,544 controls from the all-ancestry fixed-effects meta-analysis. The outer track represents the chromosomal number. SNPs are ordered by physical position. Blue dots represent known loci, and red dots are the new loci identified in the current study. Each association peak is labeled with the name of the closest gene(s) to the sentinel SNP. GWAS significance was set at $-\log_{10}P$ ~7.3. No adjustments to *P* values to account for multiple testing have been made.

PROCR encodes the endothelial protein C receptor (EPCR). We found that the G allele at rs867186 (which encodes the glycine residue at p.Ser219Gly) in *PROCR* confers protection from CAD (odds ratio (OR) = 0.93, 95% confidence interval (CI) = 0.91–0.96; **Table 1** and **Supplementary Fig. 8**). The same variant is also associated with increased circulating levels of soluble EPCR (which does not enhance protein C activation)[27], increased levels of protein C[28], increased factor VII levels[29] and increased risk of venous thrombosis[27]. Consistent with these associations, the variant has also been demonstrated to render EPCR more susceptible to proteolytic cleavage, resulting in increased shedding of membrane-bound EPCR from the endothelial surface[30] and causing elevated protein C levels in the circulation[31]. We found evidence of a second, independent CAD association at rs6088590 (r^2 = 0, D' = 0.01 with rs867186 in 1000 Genomes Project European samples; **Supplementary Fig. 8**), an intronic SNP in *NCOA6*, with the T allele conferring increased risk of CAD (conditional on rs867186, conditional $P = 1.14 \times 10^{-5}$, OR = 0.97, 95% CI = 0.95–0.98). No additional SNPs were associated with CAD after conditioning on rs867186 and rs6088590 ($P > 0.01$).

Five of the new CAD-associated regions identified in the current analysis include genes that encode proteins expressed in smooth muscle cells (*LMOD1*, *SERPINH1*, *DDX59–CAMSAP2*, *TNS1*, *PECAM1*)[32,33]. The CAD risk allele (T) of rs2820315, which is intronic in *LMOD1*, is associated with increased expression of *LMOD1* in omental and subcutaneous adipose tissues[13,34] (MuTHER, $\beta = 0.11$, $P = 1.43 \times 10^{-11}$). The protein is found in smooth muscle cells (SMCs)[32,33]. *In vitro* and transgenic mouse studies have demonstrated that CArG elements are essential for the expression of LMOD1 through both serum response factor (SRF) and myocardin (MYOCD)[35]. Myocardin has emerged as

Table 1 Newly identified CAD-associated genomic regions

			European				All ancestry				
Closest gene(s)	Variant and alleles	Chr:position (EA, AF)	OR	95% CI	*P*	*n*	OR	95% CI	*P*	$\log_{10}$ (BF)	*n*
ATP1B1	rs1892094C>T	1:169094459 (T, 0.50)	**0.96**	**0.94–0.97**	**3.99×10^{-8}**	**217,782**	**0.96**	**0.94–0.97**	**2.25×10^{-8}**	**6.33**	**243,623**
DDX59–CAMSAP2	rs6700559C>T	1:200646073 (T, 0.47)	**0.96**	**0.94–0.97**	**2.50×10^{-8}**	**221,073**	**0.96**	**0.95–0.97**	**1.13×10^{-8}**	**6.68**	**246,913**
LMOD1	rs2820315C>T	1:201872264 (T, 0.30)	**1.05**	**1.03–1.07**	**4.14×10^{-9}**	**214,844**	**1.05**	**1.03–1.07**	**7.70×10^{-10}**	**7.72**	**240,685**
TNS1[a]	rs2571445G>A	2:218683154 (A, 0.39)	1.04	1.02–1.06	3.58×10^{-6}	194,254	**1.05**	**1.03–1.06**	**4.55×10^{-10}**	**8.41**	**220,047**
ARHGAP26	rs246600C>T	5:142516897 (T, 0.48)	**1.05**	**1.03–1.06**	**1.29×10^{-8}**	**210,380**	**1.04**	**1.03–1.06**	**1.51×10^{-8}**	**6.39**	**236,223**
PARP12	rs10237377G>T	7:139757136 (T, 0.35)	0.95	0.93–0.97	1.70×10^{-7}	181,559	**0.95**	**0.93–0.97**	**1.75×10^{-8}**	**6.32**	**207,399**
PCNX3	rs12801636G>A	11:65391317 (A, 0.23)	0.95	0.93–0.97	1.00×10^{-7}	211,152	**0.95**	**0.94–0.97**	**9.71×10^{-9}**	**6.64**	**236,985**
SERPINH1	rs590121G>T	11:75274150 (T, 0.30)	**1.05**	**1.03–1.07**	**1.54×10^{-8}**	**207,426**	1.04	1.03–1.06	9.32×10^{-8}	5.80	233,249
C12orf43–HNF1A	rs2258287C>A	12:121454313 (A, 0.34)	**1.05**	**1.03–1.06**	**6.00×10^{-9}**	**221,068**	**1.04**	**1.03–1.06**	**2.18×10^{-8}**	**6.40**	**246,901**
SCARB1	rs11057830G>A	12:125307053 (A, 0.16)	**1.07**	**1.05–1.10**	**5.65×10^{-9}**	**177,550**	**1.06**	**1.04–1.09**	**1.34×10^{-8}**	**6.49**	**203,394**
OAZ2, RBPMS2	rs6494488A>G	15:65024204 (G, 0.18)	0.95	0.93–0.97	1.43×10^{-6}	205,410	**0.95**	**0.93–0.97**	**2.09×10^{-8}**	**6.41**	**228,578**
DHX38	rs1050362C>A	16:72130815 (A, 0.38)	1.04	1.03–1.06	2.32×10^{-7}	216,025	**1.04**	**1.03–1.06**	**3.52×10^{-8}**	**6.16**	**241,858**
GOSR2	rs17608766T>C	17:45013271 (C, 0.14)	**1.07**	**1.04–1.09**	**4.14×10^{-8}**	**215,857**	1.06	1.04–1.09	2.10×10^{-7}	5.30	231,213
PECAM1	rs1867624T>C	17:62387091 (C, 0.39)	0.96	0.94–0.97	1.14×10^{-7}	220,831	**0.96**	**0.95–0.97**	**3.98×10^{-8}**	**6.03**	**246,674**
PROCR[a]	rs867186A>G	20:33764554 (G, 0.11)	**0.93**	**0.91–0.96**	**1.26×10^{-8}**	**213,505**	**0.93**	**0.91–0.96**	**2.70×10^{-9}**	**7.11**	**239,340**

CAD association results for the lead SNPs from the European and all-ancestry meta-analyses are reported. SNP allele frequencies for each ancestry group are provided in **Supplementary Figure 3** and **Supplementary Table 5** for each of the studies with *de novo* genotyping. EA, effect allele; AF, effect allele frequency in Europeans; *n*, number of individuals in the analysis; OR, odds ratio; CI, confidence interval; $\log_{10}$ (BF), log 10 of the Bayes factor obtained from MANTRA analyses ($\log_{10}$ (BF) >6 is considered significant). There was no convincing evidence of heterogeneity at the new CAD-associated SNPs, $P_{het} \geq 0.01$. *P* values for heterogeneity across the data sets subjected to meta-analysis are provided in **Supplementary Table 4**, and I^2 statistics are provided in **Supplementary Figure 3**. Genome-wide significant results are shown in bold.
[a]Nonsynonymous SNP.

an important molecular switch for the programs of SMC and cardiac myocyte differentiation[36,37]. The CAD-associated SNP (or tag) is an eQTL for *IPO9* in peripheral blood mononuclear cells[38]; however, given the previous biological evidence, *LMOD1* would make the most plausible candidate gene in this region.

rs1867624 is upstream of *PECAM1*, which encodes platelet/endothelial cell adhesion molecule 1, a protein found on platelet, monocyte and neutrophil cell surfaces. The C allele is associated with reduced CAD risk (**Table 1**), increased expression of *PECAM1* in peripheral blood mononuclear cells[38] ($\beta = 0.1199$, $P = 1.38 \times 10^{-107}$) and is in LD with rs2070784 and rs6504218 ($D' = 1.0$, $r^2 > 0.8$ in 1000 Genomes Project European samples), which are eQTLs for *PECAM1* in aortic endothelial cells ($P = 4.35 \times 10^{-13}$) and stimulated $CD14^+$ monocytes ($P < 1.7 \times 10^{-24}$), respectively (**Supplementary Table 10**)[39]. PECAM-1 has been implicated in the maintenance of integrity of the vascular barrier, the breach of which is a sign of inflammatory response. Failure to restore barrier function contributes to the development of chronic inflammatory diseases such as atherosclerosis. PECAM-1-expressing endothelial cell monolayers have been shown to exhibit increased steady-state barrier function, as well as more rapid restoration of barrier integrity following thrombin-induced perturbation, in comparison to PECAM-1-deficient cells[40]. Expression of PECAM-1 has been shown to be correlated with increased plaque burden in athero-susceptible regions of the aorta in mice[41] and also with decreased atherosclerotic area in the aorta overall[42]. Together, these findings prioritize *PECAM1* as a candidate causal gene for this CAD-associated region in humans.

Of the 58 previously established CAD loci[3–9], 47 were included on the CardioMetabochip. Forty-five regions were directionally concordant with the previous reports (2 were neutral), and 34 of these 45 (42 SNPs) had at least nominal evidence of association in a fixed-effects meta-analysis ($P < 0.05$) in either our European or all-ancestry studies with *de novo* genotyping (**Supplementary Table 12**). Twenty-three of these formally replicated at a Bonferroni significance level of $P = 0.05/47 = 0.001$. *PHACTR1*, *CXCL12* and *COL4A1–COL4A2* had more statistical support of association (smaller *P* values despite fewer samples) in South Asians in comparison with the other ancestry groups. The *PHACTR1* SNP, rs9349379, is ancestrally informative, as frequency of the A allele ranges between 0.29 in the Taiwanese and 0.91 in African Americans (**Supplementary Table 12**). In contrast, the *COL4A1–COL4A2* SNP, rs4773144, had similar allele frequencies across ancestry groups (effect allele frequency (EAF) = 0.56–0.62). The stronger effect size in South Asians (OR = 0.91, 95% CI = 0.86–0.95 versus OR = 0.98, 95% CI = 0.95–1.02 in Europeans; $P_{het} = 0.0042$) could suggest gene–environment or gene–gene interactions at this locus.

We have reported 15 new CAD associations, which, together with previous efforts, bring the total number of CAD-associated regions to 73. In addition to implicating atherosclerosis and traditional risk factors as mechanisms in the pathobiology of CAD, our discoveries highlight the potential importance of biological processes active in the arterial wall involving endothelial, smooth muscle and white blood cells and promote coronary atherogenesis.

URLs. CARDIoGRAMplusC4D data on coronary artery disease and myocardial infarction, http://www.cardiogramplusc4d.org/; String database, http://string-db.org/; GTEx expression data, http://www.gtexportal.org/; Mouse Genome Informatics database, http://www.informatics.jax.org/; Protein Atlas, http://www.proteinatlas.org/; PhenoScanner, http://www.phenoscanner.medschl.cam.ac.uk/; R, http://www.R-project.org/; linkage disequilibrium information, http://www.1000genomes.org/ and http://snipa.helmholtz-muenchen.de/; gene information for *PECAM1*, http://www.ncbi.nlm.nih.gov/gene/5175.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

Table 2 Summary of functional data implicating candidate causal genes in newly identified CAD regions

SNP	Genes in region	Phenotype in mouse model	*Cis*-eQTL with SNP (or proxy with $r^2 > 0.9$)	Proteins expressed in SMC, heart, liver, blood	Candidate causal gene(s)
rs1892094C>T	*ATP1B1*, *BLZF1*, *CCDC181*, *F5*, *NME7*, *SELP*, *SLC19A2*	*ATP1B1* (cardiovascular, homeostasis, mortality/aging, muscle), *F5* (blood coagulation), *SELP* (cardiovascular, coagulation, inflammatory response)	*NME7**, *ATP1B1**	ATP1B1, NME7, SELP	*ATP1B1*, *NME7*
rs6700559C>T	*CAMSAP2*, *DDX59*, *KIF14*		*CAMSAP2**, *DDX59**	CAMSAP2, DDX59, KIF14	*CAMSAP2*, *DDX59*
rs2820315C>T	*IPO9*, *LMOD1*, *NAV1*, *SHISA4*, *TIMM17A*		*LMOD1*, *IPO9**	LMOD1	*LMOD1*
rs2571445G>A	*CXCR2*, *RUFY4*, *TNS1*	*CXCR2* (increased IL-6, abnormal interleukin levels)	*TNS1**	TNS1, RUFY4	*TNS1*
rs246600C>T	*ARHGAP26*, *FGF1*		None		
rs10237377G>T	*PARP12*, *TBXAS1*	*TBXAS1* (increased bleeding, decreased platelet aggregation)	*TBXAS1**		*TBXAS1*
rs12801636G>A	*PCNX3*, *POLA2*, *RELA*, *RNASEH2C*, *SAC3D1*, *SCYL1*, *SIPA1*, *SLC22A20*, *SLC25A45*, *SNX15*, *SNX32*, *SPDYC*, *SSSCA1*, *SYVN1*, *TIGD3*, *TM7SF2*, *TMEM262*, *VPS51*, *ZFPL1*, *ZNHIT2*	*CAPN1* (cardiovascular system), *CDCA5* (decreased mean corpuscular volume), *CFL1* (cardiovascular system), *EFEMP2* (cardiovascular), *MUS81* (cardiovascular system), *RELA* (CVD other), *SCYL1* (small myocardial fiber),	*SIPA1**	SIPA1	
rs590121G>T	*GDPD5*, *KLHL35*, *SERPINH1*	*SERPINH1* (hemorrhage)	*SERPINH1**	SERPINH1	*SERPINH1*
rs2258287C>A	*SPPL3*, *HNF1A-AS1*, *HNF1A*, *C12orf43*, *OASL*, *P2RX7*, *P2RX4*	*HNF1A* (increased cholesterol, decreased liver function), *P2RX4* (abnormal vascular endothelial cell physiology, abnormal vasodilation, abnormal common carotid artery morphology)		C12orf43, SPPL3, P2RX7, P2RX4	
rs11057830G>A	*SCARB1*, *UBC*	*SCARB1* (increased susceptibility to atherosclerosis, reduced heart rate, abnormal lipoprotein metabolism, abnormal vascular wound healing)	None	UBC	*SCARB1*
rs6494488A>G	*ANKDD1A*, *CSNK1G1*, *DAPK2*, *FAM96A*, *KIAA0101*, *OAZ2*, *PIF1*, *PLEKHO2*, *PPIB*, *RBPMS2*, *SNX1*, *SNX22*, *TRIP4*, *ZNF609*	*PIF1* (abnormal telomere length)	*ANKDD1A**, *RBPMS2**, *TRIP4**	TRIP4	*TRIP4*
rs1050362C>A	*AP1G1*, *ATXN1L*, *CALB2*, *CHST4*, *DHODH*, *DHX38*, *HP*, *HPR*	*HP* (renal, development of atherosclerosis[25])	*DHODH**, *HP**, *DHX38**	HP, DHX38, DHODH	*HP*
rs17608766T>C	*ARL17A*, *CDC27*, *GOSR2*, *MYL4*, *WNT9B*, *WNT3*		*GOSR2**	GOSR2	
rs1867624T>C	*DDX5*, *MILR1*, *PECAM1*, *POLG2*, *TEX2*	*DDX5* (abnormal vascular development), *PECAM1* (cardiovascular system, liver inflammation)	*PECAM1**	PECAM1, TEX2	*PECAM1*
rs867186A>G	*RALY*, *EIF2S2*, *ASIP*, *AHCY*, *ITCH*, *DYNLRB1*, *MAP1LC3A*,*PIGU*, *HMGB3P1*, *GGT7*, *ACSS2*, *NCOA6*, *GSS*, *MYH7B*, *TRPC4AP*, *EDEM2*, *PROCR*, *MMP24*, *EIF6*	*ASIP* (cardiovascular system), *NCOA6* (cardiovascular system), *PROCR* (abnormal circulating C-reactive protein and fibrinogen levels, thrombosis/blood coagulation)	*PROCR**, *EIF6**, *ITGB4BP**	EIF6, ITGB4BP	*PROCR*

"Genes in region" provides the genes in the LD block containing the CAD-associated SNP. "Phenotype in mouse model" lists the phenotype as provided in the Mouse Genome Informatics database; genes are listed if the phenotype affects the cardiovascular system, inflammation or liver function. eQTLs are listed where the SNP or a proxy with $r^2 > 0.9$ is an eQTL for the listed gene in one of refs. 12–14,26,38,43–49 (refer to **Supplementary Table 10** for an extended listing where $r^2 > 0.8$ between the CAD-associated SNP and the lead eQTL). Candidate genes are based on the most likely gene given the information ascertained on mouse phenotype, eQTLs, protein expression and any literature information described in the main text. Loci are further discussed in the **Supplementary Note**. An asterisk indicates that the eQTL was identified in one of blood (including peripheral blood mononuclear cells), heart, aorta/coronary artery or liver. Note that the *PCNX3* region also encompasses *AP5B1*, *ARL2*, *CAPN1*, *CDC42EP2*, *CDCA5*, *CFL1*, *CTSW*, *DPF2*, *EFEMP2*, *EHBP1L1*, *FAM89B*, *FAU*, *FRMD8*, *KAT5*, *KCNK7*, *LTBP3*, *MAP3K11*, *MRPL49*, *MUS81*, *NAALADL1* and *OVOL1*. The *DHX38* region also encompasses *IST1*, *MARVELD3*, *PHLPP2*, *PKD1L3*, *PMFBP1*, *TAT*, *TXNL4B*, *ZFHX3*, *ZNF19*, *ZNF23* and *ZNF821*. The *PROCR* region also includes *FAM83C*, *UQCC1*, *GDF5*, *SPAG4*, *CEP250*, *C20orf173*, *ERGIC3*, *FER1L4*, *CPNE1*, *RBM12*, *NFS1*, *ROMO1*, *RBM39*, *SCAND1*, *CNBD2*, *EPB41L1*, *LINC00657*, *AAR2* and *DLGAP4*.

ACKNOWLEDGMENTS

J.D. is a British Heart Foundation Professor, European Research Council Senior Investigator and NIHR Senior Investigator. J.D.E. and A.D.J. were supported by NHLBI Intramural Research Program funds. N.F. is supported by R21HL123677-01 and R56 DK104806-01A1. N.S. is supported by the British Heart Foundation and is an NIHR Senior Investigator. T.L.A. is supported by NIH career development award K23DK088942. This work was funded by the UK Medical Research Council (G0800270), the British Heart Foundation (SP/09/002), the UK National Institute for Health Research Cambridge Biomedical Research Centre, the European Research Council (268834), European Commission Framework Programme 7 (HEALTH-F2-2012-279233) and Pfizer. The eQTL database construction was supported by NHLBI intramural funds. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the US Department of Health and Human Services.

A full list of acknowledgments for the studies contributing to this work is provided in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Central analysis group: J.M.M.H., W.Z., D.R.B., T.L.A., A.S.B., D.S. Writing group: J.M.M.H., W.Z., D.R.B., D.S.P., T.L.A., A.S.B., J.D. Study analysts: J.M.M.H., W.-K.H., R.Y., L.L.W., E.L.S., S.F.N., W.-Y.L., R.D., N.F., A.J., A.P.R., C.L.C., K.Y., M.G., D.A., C.A.H., Y.-H.C., X.G., T.L.A. Study PIs and co-PIs: W.H.-H.S., P.D., J.E., S.K., N.J.S., H.S., H.W., D.J.R., J.A.J., S.L.H., A.A.Q., J.S., C.J.P., K.E.N., C.K., U.P., C.A.H., W.-J.L., I.-T.L., R.-H.C., Y.-J.H., J.I.R., J.-M.J.J., T.Q., T.-D.W., D.S.A., A.a.S.M., E.D.A., R.C., Y.-D.I.C., B.G.N., T.L.A., J.D., A.S.B., D.S., A.R., P.F. Bioinformatics, eQTL, pQTL and pathway analyses: D.S.P., W.Z., D.R.B., D.F.F., T.L.A., E.B.F., A.M., J.B.W., E.L.S., B.B.S., A.S.B., J.D.E., A.D.J., P.S., T.L.A., J.M.M.H. Genotyping: S.B., L.A.H., C.K., E.B., U.P., D.A., K.D.T., T.Q., T.L.A. Phenotyping: W.H.H.S., A.T.-H., K.L.R., P.R.K., K.E.N., C.K., C.A.H., W.-J.L., I.-T.L., R.-H.C., Y.-J.H., J.-M.J.J., T.Q., Y.-D.I.C.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Roth, G.A. *et al.* Demographic and epidemiologic drivers of global cardiovascular mortality. *N. Engl. J. Med.* **372**, 1333–1341 (2015).
2. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **385**, 117–171 (2015).
3. CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25–33 (2013).
4. Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
5. IBC 50K CAD Consortium. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet.* **7**, e1002260 (2011).
6. Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
7. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
8. Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.* **41**, 280–282 (2009).
9. CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
10. Voight, B.F. *et al.* The Metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
11. Segrè, A.V., Wei, N., Altshuler, D. & Florez, J.C. Pathways targeted by antidiabetes drugs are enriched for multiple genes associated with type 2 diabetes risk. *Diabetes* **64**, 1470–1483 (2015).
12. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
13. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
14. Franzén, O. *et al.* Cardiometabolic risk loci share downstream *cis*- and *trans*-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
15. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
16. Staley, J.R. *et al.* PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).
17. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
18. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
19. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
20. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).
21. Zanoni, P. *et al.* Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. *Science* **351**, 1166–1171 (2016).
22. Boettger, L.M. *et al.* Recurring exon deletions in the *HP* (haptoglobin) gene contribute to lower blood cholesterol levels. *Nat. Genet.* **48**, 359–366 (2016).
23. Johansson, Å. *et al.* Identification of genetic variants influencing the human plasma proteome. *Proc. Natl. Acad. Sci. USA* **110**, 4673–4678 (2013).
24. Holme, I., Aastveit, A.H., Hammar, N., Jungner, I. & Walldius, G. Haptoglobin and risk of myocardial infarction, stroke, and congestive heart failure in 342,125 men and women in the Apolipoprotein MOrtality RISk study (AMORIS). *Ann. Med.* **41**, 522–532 (2009).
25. Levy, A.P. *et al.* Haptoglobin genotype is a determinant of iron, lipid peroxidation, and macrophage accumulation in the atherosclerotic plaque. *Arterioscler. Thromb. Vasc. Biol.* **27**, 134–140 (2007).
26. Westra, H.J. *et al.* Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
27. Dennis, J. *et al.* The endothelial protein C receptor (PROCR) Ser219Gly variant and risk of common thrombotic disorders: a HuGE review and meta-analysis of evidence from observational studies. *Blood* **119**, 2392–2400 (2012).
28. Tang, W. *et al.* Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood* **116**, 5032–5036 (2010).
29. Smith, N.L. *et al.* Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: the CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation* **121**, 1382–1392 (2010).
30. Qu, D., Wang, Y., Song, Y., Esmon, N.L. & Esmon, C.T. The Ser219-->Gly dimorphism of the endothelial protein C receptor contributes to the higher soluble protein levels observed in individuals with the A3 haplotype. *J. Thromb. Haemost.* **4**, 229–235 (2006).
31. Reiner, A.P. *et al. PROC*, *PROCR* and *PROS1* polymorphisms, plasma anticoagulant phenotypes, and risk of cardiovascular disease and mortality in older adults: the Cardiovascular Health Study. *J. Thromb. Haemost.* **6**, 1625–1632 (2008).
32. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
33. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
34. Greenawalt, D.M. *et al.* A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res.* **21**, 1008–1016 (2011).
35. Nanda, V. & Miano, J.M. Leiomodin 1, a new serum response factor–dependent target gene expressed preferentially in differentiated smooth muscle cells. *J. Biol. Chem.* **287**, 2459–2467 (2012).
36. Chen, J., Kitchen, C.M., Streb, J.W. & Miano, J.M. Myocardin: a component of a molecular switch for smooth muscle differentiation. *J. Mol. Cell. Cardiol.* **34**, 1345–1356 (2002).
37. Wang, Z., Wang, D.Z., Pipes, G.C. & Olson, E.N. Myocardin is a master regulator of smooth muscle gene expression. *Proc. Natl. Acad. Sci. USA* **100**, 7129–7134 (2003).
38. Kirsten, H. *et al.* Dissecting the genetics of the human transcriptome identifies novel trait-related *trans*-eQTLs and corroborates the regulatory relevance of non-protein coding loci. *Hum. Mol. Genet.* **24**, 4746–4763 (2015).
39. Fairfax, B.P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
40. Privratsky, J.R. *et al.* Relative contribution of PECAM-1 adhesion and signaling to the maintenance of vascular integrity. *J. Cell Sci.* **124**, 1477–1485 (2011).
41. Harry, B.L. *et al.* Endothelial cell PECAM-1 promotes atherosclerotic lesions in areas of disturbed flow in ApoE-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **28**, 2003–2008 (2008).
42. Goel, R. *et al.* Site-specific effects of PECAM-1 on atherosclerosis in LDL receptor–deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **28**, 1996–2002 (2008).
43. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
44. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
45. Schröder, A. *et al.* Genomics of *ADME* gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. *Pharmacogenomics J.* **13**, 12–20 (2013).
46. Schadt, E.E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
47. Lin, H. *et al.* Gene expression and genetic variation in human atria. *Heart Rhythm* **11**, 266–271 (2014).
48. Narahara, M. *et al.* Large-scale East-Asian eQTL mapping reveals novel candidate genes for LD mapping and the genomic landscape of transcriptional effects of sequence variants. *PLoS One* **9**, e100924 (2014).
49. Innocenti, F. *et al.* Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* **7**, e1002078 (2011).

[1]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [2]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [3]Department of Applied Mathematics, University of Nottingham Malaysia Campus, Semenyih, Malaysia. [4]Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK. [5]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. [6]Pfizer Worldwide Research and Development, Cambridge, Massachusetts, USA. [7]Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, California, USA. [8]Stanford Cardiovascular Institute, Stanford University, Stanford, California, USA. [9]National Heart, Lung, and Blood Institute, Population Sciences Branch, Bethesda, Maryland, USA. [10]NHLBI and Boston University's The Framingham Heart Study, Framingham, Massachusetts, USA. [11]Division of Endocrine and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. [12]School of Medicine, National Yang-Ming University, Taipei, Taiwan. [13]College of Medicine, National Defense Medical Center, Taipei, Taiwan. [14]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark. [15]Northern Institute for Cancer Research, Newcastle University, Newcastle-upon-Tyne, UK. [16]Pfizer Worldwide Research and Development, Stockholm, Sweden. [17]Pfizer Worldwide Research and Development, Human Genetics, Cambridge, Massachusetts, USA. [18]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. [19]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [20]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [21]Centre for Genomic Health, Queen Mary University of London, London, UK. [22]Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany. [23]DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, Lübeck, Germany. [24]University Heart Center Lübeck, Lübeck, Germany. [25]Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [26]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [27]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. [28]NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK. [29]Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. [30]DZHK (German Center for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany. [31]Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [32]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [33]A list of members and affiliations appears in the **Supplementary Note**. [34]Charles Bronfman Institute for Personalized Medicine, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [35]Departments of Genetics, Medicine, and Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [36]University of Florida College of Pharmacy, Gainesville, Florida, USA. [37]Department of Cellular and Molecular Medicine, Lerner Research Institute, Cleveland, Ohio, USA. [38]Division of Cardiology, Emory University School of Medicine, Atlanta, Georgia, USA. [39]Saint Luke's Mid America Heart Institute, Kansas City, Missouri, USA. [40]Department of Biomedical and Health Informatics, University of Missouri–Kansas City, Kansas City, Missouri, USA. [41]College of Medicine, University of Florida, Gainesville, Florida, USA. [42]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA. [43]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [44]Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey, USA. [45]Division of Genomic Medicine, National Human Genome Research Institute, US National Institutes of Health, Bethesda,Maryland, USA. [46]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [47]Carolina Center for Genome Sciences, Chapel Hill, North Carolina, USA. [48]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA. [49]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. [50]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan. [51]Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan. [52]Institute for Translational Genomics and Population Sciences, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, California, USA. [53]School of Medicine, Chung Shan Medical University, Taichung, Taiwan. [54]Division of Endocrinology and Metabolism, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan. [55]Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, LABioMed at Harbor-UCLA Medical Center, Torrance, California, USA. [56]Cardiovascular Center and Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. [57]National Taiwan University College of Medicine, Taipei, Taiwan. [58]Centre for Non-Communicable Disease, Karachi, Pakistan. [59]School of Kinesiology and Health Science, York University, Toronto, Ontario, Canada. [60]National Institute of Cardiovascular Diseases, Sher-e-Bangla Nagar, Bangladesh. [61]National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK. [62]Wellcome Trust Sanger Institute, Hinxton, UK. [63]British Heart Foundation Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, UK. [64]These authors contributed equally to this work. Correspondence should be addressed to J.M.M.H. (jmmh2@medschl.cam.ac.uk).

ONLINE METHODS

Study participants. A full description of the component studies with *de novo* genotyping is given in the **Supplementary Note** and **Supplementary Table 1**. In brief, the European studies comprised 16,093 CAD cases and 16,616 controls from EPIC-CVD (a case cohort study embedded in the pan-European EPIC prospective study), the Copenhagen City Heart Study (CCHS), the Copenhagen Ischemic Heart Disease Study (CIHDS) and the Copenhagen General Population Study (CGPS), all recruited within Copenhagen, Denmark. The South Asian studies comprised up to 7,654 CAD cases and 7,014 controls from the Pakistan Risk of Myocardial Infarction Study (PROMIS), a case–control study that recruited samples from nine sites in Pakistan, and the Bangladesh Risk of Acute Vascular Events (BRAVE) study based in Dhaka, Bangladesh. The East Asian studies comprised 4,129 CAD cases and 6,369 controls recruited from seven studies across Taiwan that collectively comprise the TAIwan metaboCHIp (TAICHI) Consortium. The African-American studies comprised 2,100 CAD cases and 5,746 controls from the Atherosclerosis Risk in Communities Study (ARIC), Women's Health Initiative (WHI) and six studies from the Myocardial Infarction Genetics Consortium (MIGen).

Ethical approval was obtained from the appropriate ethics committees, and informed consent was obtained from all participants.

Genotyping and quality control in studies with *de novo* genotyping. Samples from EPIC-CVD, CCHS, CIHDS, CGPS, BRAVE and PROMIS were genotyped on a customized version of the Illumina CardioMetabochip (referred to as 'Metabochip+', Illumina), in two Illumina-certified laboratories located in Cambridge, UK, and Copenhagen, Denmark, by technicians masked to the phenotypic status of samples. The remaining studies were genotyped using the standard CardioMetabochip[10] in HudsonAlpha and Cedars Sinai (TAICHI[50], WHI and ARIC[51]) and the Broad Institute (MIGen).

Each collection was genotyped and underwent quality control separately (**Supplementary Tables 1** and **2**). In brief, studies genotyped on the Metabochip+ had genotypes assigned using Illumina GenCall software in GenomeStudio. Samples were removed if they had a call rate <0.97, if they had average heterozygosity >±3 s.d. from the overall mean heterozygosity or if their genotypic sex did not match their reported sex. One of each pair of duplicate samples and first-degree relatives (identified by a kinship coefficient >0.2) was removed.

Across all studies, SNP exclusions were based on MAF <0.01, $P < 1 \times 10^{-6}$ for Hardy–Weinberg equilibrium or call rate (CR) <0.97 (full details are given in **Supplementary Table 2**). These exclusions were also applied centrally to studies genotyped on the CardioMetabochip, namely the ARIC, WHI, MIGen and TAICHI studies. Principal-component analysis (PCA) was applied to identify and remove ancestral outliers. More stringent thresholds were adopted for SNPs used in the PCA for TAICHI and those studies genotyped on Metabochip+, namely CR <0.99, Hardy–Weinberg equilibrium $P < 1 \times 10^{-4}$ and MAF <0.05. In addition, one of each pair of SNPs in LD ($r^2 > 0.2$) was removed, as were variants in regions known to be associated with CAD.

SNP association analyses and meta-analyses. Statistical analyses were performed in R or PLINK[52] unless otherwise stated.

We collected sufficient samples to ensure the study was well powered to detect effect sizes in the range of OR = 1.05–1.10, which have typically been reported for CAD. With 88,000 cases, the study would have 88% power to detect OR = 1.05 for a SNP with MAF = 0.2 at $\alpha = 5 \times 10^{-8}$, assuming a multiplicative model on the OR scale. For a lower MAF of 0.1, the study would have 93% power to detect OR = 1.07 at $\alpha = 5 \times 10^{-8}$, assuming a multiplicative model. Power calculations were performed using Quanto.

Association with CAD was assessed in studies with *de novo* genotyping from European, South Asian and East Asian individuals, using the Genome-wide Efficient mixed-model analysis (GEMMA) approach[53]. This model includes both fixed effects and random effects of genetic inheritance. CAD (coded as 0/1) was the outcome variable; up to five principal components and the SNP of interest, coded additively, were included as fixed effects. *P* values from the score test are reported. The African-American studies were analyzed using a logistic model in PLINK, with CAD as the outcome variable and SNPs coded additively as the predictor. The covariates used by each study, including the number of principal components, are reported in the **Supplementary Note**. Genomic inflation was at most 5% for any given study (**Supplementary Fig. 1** and **Supplementary Note**). A subset of individuals from the PROMIS study and the EPIC-CVD consortium contributed to the CARDIoGRAMplusC4D 2013 report. To avoid any overlap of individuals in our studies with those in CARDioGRAMplusC4D, two analyses of these two studies were performed. One analysis included all the samples. A second analysis of the PROMIS and EPIC-CVD studies was performed after excluding all samples that had contributed to the CARDIoGRAMplusC4D study and before meta-analysis of our results with the results from the CARDIoGRAMplusC4D consortium. The CARDIoGRAMplusC4D SNP association results were converted onto the plus strand of GRh37, checked for heterogeneity and checked to ensure that allele frequencies were consistent with those for European populations.

Fixed-effects inverse-variance-weighted meta-analysis was used to combine results across studies in METAL[54]. Heterogeneity *P* values and I^2 values were calculated, and any SNP with $P < 0.0001$ for heterogeneity was removed. We performed two meta-analyses: the first involved just the European studies with *de novo* genotyping and the CARDIoGRAMplusC4D results to minimize ancestral diversity. The second involved all studies with *de novo* genotyping and the CARDIoGRAMplusC4D results to maximize sample size and statistical power. Given the ancestral diversity of the component studies with *de novo* genotyping, we also implemented meta-analyses with MANTRA[55], a meta-analysis approach designed to handle trans-ancestry study designs. However, for our studies, the data were broadly consistent with the results from METAL (**Table 1** and **Supplementary Table 4**), and we therefore primarily report the fixed-effect meta-analysis results.

Conditional association analyses. Analyses to test for secondary association signals across seven regions with potential for independent signals were performed using GCTA[56]. GCTA implements a method for conducting conditional analyses using summary-level statistics (effect size, standard error, *P* value, effective sample size) and LD information (r^2) between SNPs estimated from a reference panel[56]. Conditional analyses were performed separately in CARDIoGRAMplusC4D, European, South Asian, and East Asian samples, and the results were combined using an inverse-variance-weighted fixed-effects meta-analysis approach. The conditional analyses were not performed in African-American samples because the SNP-level case–control counts were not made available for ARIC, MIGen and WHI. The 1000 Genomes Project Phase 3 v5 ancestry-specific reference panel was used to provide LD information (r^2) for the conditioned SNPs and other SNPs in the test regions for each of the three ancestry groups considered in the analyses. As approximately 9% of CARDIoGRAMplusC4D samples were South Asian and the remainder were European, to calculate LD for this data set, we sampled with replacement the genotypes of 50 individuals from the 1000 Genomes Project South Asian reference panel and combined them with the genotypes of the 503 European individuals available in the 1000 Genomes Project. To identify SNPs that were associated with CAD independently of the lead SNP in the test region, the association of each SNP in the region was tested conditioning on the most significant SNP in the overall meta-analysis of European, South Asian, East Asian and CARIoGRAMplusC4D samples. SNPs were identified as independent signals for a specific region if the conditional *P* value was $\leq 1 \times 10^{-4}$. In each region, we performed several rounds of conditional analyses until the conditional *P* value was $>1 \times 10^{-4}$ for all SNPs in the region.

eQTL and epigenetic analyses. The MuTHER data set contains gene expression data from 850 UK twins for 23,596 probes and 2,029,988 (HapMap 2–imputed) SNPs. All *cis*-associated SNPs with FDR < 1% within each of the 15 newly identified CAD regions (IMPUTE info score > 0.8) were extracted from the MuTHER project data set for each of the tissues LCLs ($n = 777$), adipose ($n = 776$) and skin ($n = 667$).

The GTEx Project provides expression data from up to 449 individuals for 52,576 genes annotated in GENCODE v12 (including pseudogenes) and 6,820,472 genotyped SNPs (using the Human Omni5-Quad array).

In addition to the publicly available MuTHER and GTEx databases imputed to the HapMap and 1000 Genomes projects, respectively, we used a curated database of over 100 distinct eQTL data sets to determine whether our lead CAD-associated SNPs or SNPs in high LD with them ($r^2 > 0.8$ in Europeans from the HapMap or 1000 Genomes projects) were associated with the expres-

doi:10.1038/ng.3874

sion of one or more nearby genes in *cis*[57]. All our collated eQTL data sets meet criteria for statistical thresholds for SNP–gene transcript associations as described in the original studies[12,13,57]. In total, more than 30 different cells or tissues were queried, including circulating white blood cells of various types, liver, adipose, skin, brain, breast, heart and lung tissues. Complete details of the data sets and tissues queried in the current work can be found in the **Supplementary Note** and **Supplementary Table 10**, and a general overview of a subset of over 50 eQTL studies has been published[57]. We first identified all sets of eQTLs in perfect LD ($r^2 = 1$ among Europeans in the HapMap or 1000 Genomes project) with each other for each unique combination of study, tissue and transcript. We then determined whether any of these sets of eQTLs were either in perfect ($r^2 = 1$) or high ($1 > r^2 > 0.8$) LD with our lead CAD SNP (**Supplementary Table 10**).

We required that any eQTL had $P < 5 \times 10^{-8}$ for association with expression levels to be included in the eQTL tables.

We examined chromatin state maps of 23 relevant primary cell types and tissues. Chromatin states are defined as spatially coherent and biologically meaningful combinations of specific chromatin marks. These are computed by exploiting the correlation of such marks, including DNA methylation, chromatin accessibility and several histone modifications[58,59].

pQTL analyses. We conducted plasma protein assays in 3,301 healthy blood donors from the INTERVAL study[60] who had all been genotyped on the Affymetrix Axiom UK Biobank genotyping array and imputed to a combined 1000 Genomes Project + UK10K haplotype reference panel[61]. Proteins were assayed using the SomaLogic SomaScan platform, which uses high-specificity aptamer binding to provide relative protein abundances. Proteins passing stringent quality control (for example, coefficient of variation <20%) were log transformed, and age, sex, the duration between venepuncture and sample processing, and the first three principal components of genetic ancestry were regressed out. Residuals were then rank inverse normalized before genome-wide association testing using an additive model accounting for imputation uncertainty.

Enrichment analyses. *Ingenuity Pathway Analyses.* We used the Core Analysis function in Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems) to identify canonical pathways enriched for one or more SNPs with a low *P* value in the all-ancestry meta-analysis.

Modified MAGENTA. Given that the Metabochip comprises a select set of SNPs and lacks complete genomic coverage[10], MAGENTA, which assumes random sampling of variants from across the genome, could not be directly implemented. Therefore, a modified version of MAGENTA involving a hypergeometric test to account for the chip design was used to test for pathways that were enriched with CAD-associated variants[11]. This approach requires defining two sets of variants: a null set of variants that are not associated with CAD and a set of variants that are associated with CAD, referred to as the 'associated set'. Multiple variants can map to the same gene and still be included in the test. SNPs in LD were pruned out of the association results such that $r^2 < 0.2$ for all pairs of SNPs (based on 1000 Genomes Project data[62]; **Supplementary Table 6**) before implementation of modified MAGENTA. The null set was defined as the 1,000 remaining QT interval–associated SNPs with the largest *P* values (least evidence) for association with CAD. The associated set was defined as variants (after LD pruning) that showed evidence of association $P < 1 \times 10^{-6}$. This approach was adopted to select the null and associated sets so as to limit the number of variants included in the hypergeometric cumulative mass function, as a large number of variants results in an intractable calculation for the binomial coefficients. The observed *P* value from the hypergeometric test is compared to the *P* values obtained from 10,000 random sets to compute an empirical enrichment *P* value.

HaploReg: H3K27ac-based tissue enrichment analysis. The associated set as defined for MAGENTA was used for HaploReg analyses and compared to a background set of 12,000 SNPs previously associated with any trait at $P < 1 \times 10^{-5}$ (taken from sources such as the NHGRI-EBI GWAS catalog). Using data from HaploReg[15], we counted the number of SNPs with an H3K27ac annotation or in high LD ($r^2 > 0.8$ from the SNiPA[63] European 1000 Genomes Project maps) with a SNP with an H3K27ac annotation. The significance of the enrichment in H3K27ac marks from a particular tissue was determined by comparing the fraction of associated SNPs with that mark to the fraction of background SNPs with that same mark. A hypergeometric test was used to assign a *P* value to the enrichment.

Data availability. The full data from the trans-ancestry meta-analysis and the European meta-analysis from this report are available through http://www.phenoscanner.medschl.cam.ac.uk/.

50. Assimes, T.L. *et al.* Genetics of coronary artery disease in Taiwan: a cardiometabochip study by the Taichi Consortium. *PLoS One* **11**, e0138014 (2016).
51. Franceschini, N. *et al.* Prospective associations of coronary heart disease loci in African Americans using the MetaboChip: the PAGE study. *PLoS One* **9**, e113203 (2014).
52. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
53. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
54. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
55. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
56. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
57. Zhang, X. *et al.* Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. *BMC Genomics* **15**, 532 (2014).
58. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).
59. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
60. Moore, C. *et al.* The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. *Trials* **15**, 363 (2014).
61. Astle, W.J. *et al.* The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
62. Abecasis, G.R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
63. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmüller, G. SNiPA: an interactive, genetic variant–centered annotation browser. *Bioinformatics* **31**, 1334–1336 (2015).