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

Joanna M.M. Howson[1], Wei Zhao[2,65], Daniel R. Barnes[1,65], 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[11,12,13], 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[22,23,24], Sekar Kathiresan[25,26], Nilesh J. Samani[27,28], Heribert Schunkert[29,30], Hugh Watkins[31,32], CARDIoGRAMplusC4D, Ron Do[33], Daniel J. Rader[34], Julie A. Johnson[35], Stanley L. Hazen[36], Arshed Quyyumi[37], John A. Spertus[38,39], Carl J. Pepine[40], Nora Franceschini[41], Anne Justice[41], Alex P. Reiner[42], Steven Buyske[43], Lucia A. Hindorff[44], Cara L. Carty[45], Kari E. North[46,47], Charles Kooperberg[45], Eric Boerwinkle[48,49], Kristin Young[46], Mariaelisa Graff[46], Ulrike Peters[45], Devin Absher[5], Chao A. Hsiung[50], Wen-Jane Lee[51], Kent D. Taylor[52], Ying-Hsiang Chen[50], I-Te Lee[53,54,55], Xiuqing Guo[52], Ren-Hua Chung[50], Yi-Jen Hung[13,56], Jerome I. Rotter[57], Jyh-Ming J. Juang[58,59], Thomas Quertermous[7,8], Tzung-Dau Wang[58,59], Asif Rasheed[60], Philippe Frossard[60], Dewan S. Alam[61], Abdulla al Shafi Majumder[62], Emanuele Di Angelantonio[1,63], Rajiv Chowdhury[1], EPIC-CVD, Yii-Der Ida. Chen[52], Børge G. Nordestgaard[14,19], Themistocles L. Assimes[7,8,66], John Danesh[1,63,64,66], Adam S. Butterworth[1,63,66], Danish Saleheen[1,2,60,66]

1 . MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, UK.

2 . Department of Biostatistics and Epidemiology, University of Pennsylvania, USA.

3 . Department of Applied Mathematics, University of Nottingham Malaysia Campus, Malaysia.

4 . Robertson Centre for Biostatistics, University of Glasgow, UK.

5 . HudsonAlpha Institute for Biotechnology, USA.

6 . Pfizer Worldwide Research and Development, USA.

7 . Department of Medicine, Division of Cardiovascular Medicine, Stanford University, USA.

8 . Stanford Cardiovascular Institute, Stanford University, USA.

9 . National Heart, Lung and Blood Institute, Population Sciences Branch, USA.

10 . NHLBI and Boston University's The Framingham Heart Study, USA.

11 . Division of Endocrine and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taiwan ROC.
12 . School of Medicine, National Yang-Ming University, Taiwan ROC.
13 . College of Medicine, National Defense Medical Center, Taiwan ROC.
14 . Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Denmark.
15 . Northern Institute for Cancer Research, Paul O'Gorman Building, Newcastle University, UK.
16 . Pfizer Worldwide Research and Development, Sweden.
17 . Pfizer Worldwide Research and Development, Human Genetics, USA.
18 . Dept. of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Denmark.
19 . Faculty of Health and Medical Sciences, University of Copenhagen, Denmark.
20 . William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, UK.
21 . Centre for Genomic Health, Queen Mary University of London, UK.
22 . Institute for Cardiogenetics, University of Lübeck, Germany.
23 . DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, Germany.
24 . University Heart Center Luebeck, Germany.
25 . Center for Human Genetic Research | Massachusetts General Hospital, USA.
26 . Harvard Medical School, USA.
27 . Department of Cardiovascular Sciences, University of Leicester, UK.
28 . NIHR Leicester Biomedical Research Centre, Glenfield Hospital, UK.
29 . Deutsches Herzzentrum München, Technische Universität München, Germany.
30 . DZHK (German Center for Cardiovascular Research), partner site Munich Heart Alliance, Germany.
31 . Radcliffe Department of Medicine, University of Oxford, UK.
32 . Wellcome Trust Centre for Human Genetics, University of Oxford, UK.
33 . The Charles Bronfman Institute for Personalized Medicine, Department of Genetics and Genomic

Sciences, Icahn School of Medicine at Mount Sinai, USA.

34 . Departments of Genetics, Medicine, and Pediatrics, Perelman School of Medicine, University of Pennsylvania, USA.

35 . University of Florida College of Pharmacy, USA.

36 . Department of Cellular and Molecular Medicine, Lerner Research Institute, USA.

37 . Division of Cardiology, Emory University School of Medicine, USA.

38 . Saint Luke's Mid America Heart Institute, USA.

39 . University of Missouri-Kansas City, USA.

40 . College of Medicine, University of Florida, USA.

41 . Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, USA.

42 . Department of Epidemiology, University of Washington, Seattle, WA, USA.

43 . Department of Statistics & Biostatistics, Rutgers University, Piscataway, NJ, USA.

44 . Division of Genomic Medicine, National Human Genome Research Institute, NIH, USA.

45 . Public Health Sciences Division, Fred Hutchinson Cancer Research Center, USA.

46 . Gillings School of Global Public Health, University of North Carolina, USA.

47 . Carolina Center for Genome Sciences, USA.

48 . Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, USA.

49 . Human Genome Sequencing Center, Baylor College of Medicine, USA.

50 . Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Taiwan ROC.

51 . Department of Medical Research, Taichung Veterans General Hospital, Taiwan ROC.

52 . Institute for Translational Genomics and Population Sciences, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, USA.

53 . Division of Endocrine and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taiwan.

54 . School of Medicine, National Yang-Ming University, Taiwan.

55 . School of Medicine, Chung Shan Medical University, Taiwan.

56 . Division of Endocrinology and Metabolism, Tri-Service General Hospital, National Defense Medical Center, Taiwan ROC.

57 . Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, LABioMed at Harbor-UCLA Medical Center, USA.

58 . Cardiovascular Center and Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Taiwan ROC.

59 . National Taiwan University College of Medicine, Taiwan ROC.

60 . Centre for Non-Communicable Disease, Pakistan.

61 . School of Kinesiology and Health Science, York University, Canada.

62 . National Institute of Cardiovascular Diseases, Sher-e-Bangla Nagar, Bangladesh.

63 . The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, UK.

64 . Wellcome Trust Sanger Institute, UK.

65. Authors contributed equally

66. Authors contributed equally

Corresponding author: Joanna M M Howson, jmmh2@medschl.cam.ac.uk

Summary paragraph

Coronary artery disease (CAD) is a leading cause of morbidity and mortality worldwide[1,2]. Although 58 genomic regions have been associated with CAD to date[3-9], most of the heritability is unexplained[9], indicating 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 meta-analysed results with 194,427 participants previously genotyped to give a total of 88,192 CAD cases and 162,544 controls. We identified 25 new SNP-CAD-associations ($P < 5x10^{-8}$, in fixed effects meta-analysis) from 15 genomic regions, including SNPs in or near genes involved in cellular adhesion, leucocyte 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 shed light on potential novel disease mechanisms.

MAIN TEXT

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%) ancestries (Supplementary Information; Supplementary Tables 1, 2, 3; Supplementary Fig. 1). The results from our association analyses of these additional samples were meta-analysed with those reported by CARDIoGRAMplusC4D at 79,070 SNPs in two fixed effects meta-analyses, one in EUR participants and a second across all four ancestries (Figure 1 and 2). (Over-lapping samples were removed prior to meta-analysis [Methods]). A genome-wide significance threshold ($P \leq 5 \times 10^{-8}$ in the fixed effects meta-analysis) was adopted to minimise false positive findings. However, even at this strict *P*-value threshold, there is still a small chance of a false-positive result. The EUR 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 Table 4; Supplementary Fig. 2). An additional six distinct novel CAD-associated regions were identified in the all ancestries fixed effects meta-analysis (Table 1; Figure 2; Supplementary Table 4). In total, 15 novel CAD-associated genomic regions (25 SNPs) were identified (Supplementary Fig. 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 EUR studies with *de novo* genotyping, or in a fixed effects meta-analysis of all the studies with de novo genotyping (Supplementary Table 5, Supplementary Fig. 5). Within the CARDIoGRAMplusC4D results for these SNPs, there was no evidence of heterogeneity of effects ($P \geq 0.10$) and allele frequencies were consistent with our EUR studies (Supplementary Table 5). Tests for enrichment of CAD-associations within sets of genes[11] and Ingenuity Pathway Analysis confirmed known CAD pathways (Supplementary Information; Supplementary Tables 6, 7, 8).

To prioritize candidate causal genes at the new loci, we defined regions encompassing the novel CAD-associated SNPs based on recombination rates (Supplementary Table 9) and cross referenced them with expression quantitative trait loci (eQTL) databases including GTEx[12], MuTHER[13] and STARNET[14] (Methods). Twelve of the 15 novel CAD-associated SNPs were identified as potential eQTLs in at least one tissue ($P<5x10^{-8}$; Table 2, Supplementary Table 10). Haploreg analysis[15] (Methods) showed CAD-associated SNPs were enriched for H3K27ac enhancer marks ($P < 5.1x10^{-4}$) in multiple heart related tissues (left ventricle, right atrium, aorta) in the EUR 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 (pQTL) in plasma on the aptamer-based Somalogic platform (Methods). Twenty-four proteins from the newly identified CAD regions were assayed and passed QC. Of our 15 novel CAD-associated SNPs, two 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 ($P=10^{-10}$, discussed below) and rs1050362 (NP_054722.2:p.Arg140=) a synonymous variant in *DHX38* was a trans-pQTL for the apolipoprotein L1 ($P=5.37x10^{-29}$; 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 novel loci might be related to CAD risk, we examined the associations of the 15 novel CAD regions with a wide range of risk factors, molecular traits, and clinical disorders, using PhenoScanner[16] (which encompasses the NHGRI-EBI GWAS catalogue and other genotype-phenotype databases).

Six of our loci have previously been associated with known CAD risk factors, such as major lipids (*PCNX3*,[17] *C12orf43/HNF1A*, *SCARB1*, *DHX38*)[18] and blood pressure (*GOSR2*,[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) rs11057830 and rs11057841 tag the CAD-association in the *SCARB1* region (Table 1; Supplementary Table 4), a region reported previously to be associated with HDL (rs838876, β=-0.049, P=7.33x10^{-33})[18]. A *rare* nonsynonymous variant rs74830677 (NP_005496.4:p.Pro376Leu) in *SCARB1* also associated with high levels of high-density lipoprotein cholesterol (HDL-C)[21]. Conditional analyses showed that the CAD-association was independent of the common variant HDL association (Supplementary Information, Supplementary Fig. 6). We found 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 play a role in regulation of liver-restricted expression of *SCARB1*.

The *DHX38* region has previously been associated with increased total and LDL cholesterol[18]. Both CAD-associated SNPs in *DHX38*, rs1050362 (NP_054722.2:p.Arg140=) and rs2072142 (synonymous and intronic respectively; Table 1, Supplementary Table 4) are in LD but not strongly correlated with the previously reported cholesterol increasing SNP, intronic in *HPR*, rs2000999, (r^2=0.41, D'=1 in 1000 Genomes EUR). Deletions in the HP gene have recently been shown to drive the reported cholesterol association in this region[22]. The CAD SNPs are in strong LD with SNPs that increase haptoglobin levels[23] (rs6499560, P=2.92x10^{-13}, r^2=0.97), and haptoglobin has been reported to be associated with increased CAD risk[24]. HP encodes an alpha-2-glycoprotein which is synthesised in the liver. It binds free haemoglobin and protects tissues from oxidative damage. Mouse models indicate the role of *Hp* with 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 e.g. *DHODH* in aorta artery[12] (rs1050362 A allele, β=0.41, P=1.4x10^{-9}), *DHX38* in peripheral blood[26], atherosclerotic aortic root[14] (P<8x10^{-26}; Table 2, Supplementary Table 10), the A allele at rs1050362 is also associated with increased expression of *HP* in left ventricle heart (β=0.535, P=8.71x10^{-10})[12] and decreased expression of *HP* in whole blood (β=-0.27, P=1.22x10^{-10})[12]. While

199 there could be multiple causal genes in the region, together these findings suggest *HP* is a promising
200 candidate gene.

201

202 *PROCR* encodes the endothelial protein C receptor (EPCR). We found the G allele at rs867186
203 (which codes for the glycine residue at p.Ser219Gly) in *PROCR* confers protection from CAD
204 (OR[95%CI]=0.93[0.91-0.96]; Table 1, Supplementary Fig. 8). The same variant is also associated
205 with increased circulating levels of soluble EPCR (which does not enhance protein C activation)[27],
206 increased levels of protein C[28], increased factor VII levels[29], and increased risk of venous
207 thrombosis[27]. Consistent with these associations, the variant has also been demonstrated to render the
208 EPCR more susceptible to proteolytic cleavage, resulting in increased shedding of membrane-bound
209 EPCR from the endothelial surface[30] causing elevated protein C levels in the circulation[31]. We found
210 evidence of a second, independent CAD-association at rs6088590 (r^2=0, D'=0.01 with rs867186 in
211 1000G EUR samples; Supplementary Fig. 8), an intronic SNP in *NCOA6* with the T allele conferring
212 increased risk of CAD (conditional on rs867186, conditional P=1.14x10^{-5}, OR[95% CI]=0.97[0.95-
213 0.98]). No additional SNPs were associated with CAD after conditioning on rs867186 and rs6088590
214 (P>0.01).

215

216 Five of the novel CAD regions identified in the current analysis include genes that encode proteins
217 expressed in smooth muscle cells (*LMOD1*, *SERPINH1*, *DDX59*/*CAMSAP2*, *TNS1, PECAM1*)[32,33].
218 The CAD risk allele (T) of rs2820315, which is intronic in *LMOD1*, is associated with increased
219 expression of *LMOD1* in omental and subcutaneous adipose tissues[13,34] (MuTHER, β=0.11,
220 P=1.43x10^{-11}). The protein is found in smooth muscle cells (SMC)[32,33]. *In vitro* and transgenic mouse
221 studies demonstrate an essential requirement for CArG elements in the expression of LMOD1 through
222 both serum response factor (SRF) and myocardin (MYOCD)[35]. Myocardin has emerged as an
223 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 prior biological evidence *LMOD1* would make the most plausible candidate gene.

rs1867624 is upstream of *PECAM1*, which encodes platelet/endothelial cell adhesion molecule 1, a protein found on platelet, monocyte and neutrophil surfaces. The C-allele is associated with reduced CAD risk (Table 1), increased expression of *PECAM1* in peripheral blood mononuclear cells[38] (β=0.1199, P=1.38x10^{-107}) and is in LD with rs2070784 and rs6504218 (D'=1.0, r^2>0.8 in 1000G EUR samples), which are eQTL for *PECAM1* in aortic endothelial cells (P=4.35x10^{-13}) and stimulated CD14+ monocytes[39] respectively (P<1.7x10^{-24}; Supplementary Table 10)[39]. PECAM-1 has been implicated in the maintenance of vascular barrier integrity, 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 compared 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 prioritise *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 (two were neutral) and thirty-four of these 45 (42 SNPs) had at least nominal evidence of association in a fixed effects meta-analysis (P<0.05) in either our EUR or all ancestry studies with *de novo* genotyping (Supplementary Table 12). Twenty-three of these formally replicated at a Bonferroni significance level P=0.05/47=0.001). *PHACTR1*, *CXCL12* and *COL4A1-COL4A2* had more statistical support of association (smaller *P*-values despite fewer samples) in SAS compared with the other ancestries. The *PHACTR1* SNP,

rs9349379, is ancestrally informative, as the A allele frequency 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 ancestries (EAF=0.56-0.62). The stronger effect size in SAS (OR[95%CI]=0.91[0.86-0.95] versus 0.98[0.95-1.02] in EUR, heterogeneity *P*=0.0042) could suggest gene-environment or gene-gene interactions at this locus.

We have reported 15 novel CAD-associations, which, together with previous efforts, brings 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

Data on coronary artery disease / myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.cardiogramplusc4d.org; String database: http://string-db.org; GTEx expression data were obtained from: www.gtexportal.org; the mouse genome informatics database: http://www.informatics.jax.org; protein atlas: http://www.proteinatlas.org/; phenoscanner: www.phenoscanner.medschl.cam.ac.uk; R: www.R-project.org; linkage disequilibrium information: www.1000genomes.org, http://snipa.helmholtz-muenchen.de/; Gene information: http://www.ncbi.nlm.nih.gov/gene/5175

ACKNOWLEDGEMENTS

J Danesh is a British Heart Foundation Professor, European Research Council Senior Investigator, and NIHR Senior Investigator. J.D. Eicher and A.D. Johnson were supported by NHLBI Intramural Research Program funds. N Franceschini is supported by R21HL123677-01 and R56 DK104806-01A1. N Samani is supported by the British Heart Foundation and is a NIHR Senior Investigator. T.L. Assimes is supported by an NIH career development award K23DK088942. This work was funded by the UK Medical Research Council (G0800270), British Heart Foundation (SP/09/002), UK National Institute for Health Research Cambridge Biomedical Research Centre, 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 U.S. Department of Health and Human Services.

A full list of acknowledgements for the studies contributing to this work are provided in the Supplementary Information.

AUTHOR CONTRIBUTIONS

Central analysis group: JMMH, WZ, DRB, TLA, ASB, DS. Writing group: JMMH, WZ, DRB, DSP, TLA, ASB, JD. Study analysts: JMMH, W-KH, RY, LLW, EBF, ELS, SFN, W-YL, RD, NF, AJ, APR, CLC, KY, MG, DA, CAH, Y-HC, XG, TLA. Study PIs and co-PIs: WH-HS; PD, JE, SK, NJS, HS, HJW, DJR, JJ, SH, AQ, JS, CJP, KEN, CK, UP, CAH, W-JL, I-TL, R-HC, Y-JH, JIR, J-MJJ, TQ, T-DW, DSA, AalSM, EDA, RC, Y-DIC, BGN, TLA, JD, ASB, DS, AR, PF. Bioinformatics, eQTL, pQTL and pathway analyses: DSP, WZ, DRB, DFF, TLA, EBF, ELS, BBS, ASB, JDE, ADJ, PS, TLA, JMMH. Genotyping: SB, LAH, CK, EB, UP, DA, KDT, TQ, TLA. Phenotyping: WH-HS, AT-H, KLR, PRK, KEN, CK, CAH, W-JL, I-TL, R-HC, Y-JH, J-MJJ, TQ, Y-DIC

COMPETING FINANCIAL INTERESTS

AM, EBF and JBW are full time employees of Pfizer. DFF is now a full time employee of Bayer AG, Germany. JD reports personal fees and non-financial support from Merck Sharp & Dohme UK Atherosclerosis, Novartis Cardiovascular & Metabolic Advisory Board, Pfizer Population Research Advisory Panel, Sanofi Advisory Board.

REFERENCES

1. Roth, G.A. *et al.* Demographic and epidemiologic drivers of global cardiovascular mortality. *N Engl J Med* **372**, 1333-41 (2015).
2. G. B. D. Mortality & 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**, 117-71 (2015).
3. CARDIoGRAMplusC4D Consortium *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25-33 (2013).
4. Myocardial Infarction Genetics Consortium *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-41 (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**, 443-53 (2007).
7. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* **43**, 333-8 (2011).
8. Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet* **41**, 280-2 (2009).
9. CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-30 (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. Segre, A.V. *et al.* Pathways targeted by antidiabetes drugs are enriched for multiple genes associated with type 2 diabetes risk. *Diabetes* **64**, 1470-83 (2015).
12. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
13. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-9 (2012).
14. Franzen, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827-30 (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**, D930-4 (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 *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).
18. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-13 (2010).
19. International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
20. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* (2016).
21. Zanoni, P. *et al.* Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. *Science* **351**, 1166-71 (2016).
22. Boettger, L.M. *et al.* Recurring exon deletions in the HP (haptoglobin) gene contribute to lower blood cholesterol levels. *Nat Genet* **48**, 359-66 (2016).

23. Johansson, A. *et al.* Identification of genetic variants influencing the human plasma proteome. *Proc Natl Acad Sci U S A* **110**, 4673-8 (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**, 522-32 (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**, 134-40 (2007).
26. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (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**, 2392-400 (2012).
28. Tang, W. *et al.* Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood* **116**, 5032-6 (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**, 1382-92 (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**, 229-35 (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**, 1625-32 (2008).
32. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **28**, 1248-50 (2010).
33. Uhlen, M. *et al.* Proteomics. 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**, 1008-16 (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**, 2459-67 (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**, 1345-56 (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 U S A* **100**, 7129-34 (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 locidagger. *Hum Mol Genet* **24**, 4746-63 (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**, 1477-85 (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**, 2003-8 (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**, 506-11 (2013).

44. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
45. Schroder, 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**, 266-71 (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).

Figure Legends

Figure 1 Schematic of the study design. The sample-size information is provided as number of cases/number of controls. Note, samples with *de novo* genotyping that were also in the CARDIoGRAMplusC4D study were removed prior to 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 EUR individuals contributed to the meta-analysis of our studies with *de novo* genotyping and 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 SAS samples contributed to the meta-analysis of our studies with *de novo* genotyping and CARDIoGRAMplusC4D was 3,950 CAD cases and 3,581 controls.

Figure 2 Plot showing the association of ~79,000 variants with CAD ($-\log_{10}P$-value) in up to 88,192 cases and 162,544 controls from the all ancestry fixed effects meta-analysis. SNPs are ordered in physical position. No adjustments to *P*-values to account for multiple testing have been made. The outer track represents the chromosomal number. 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 ($-\log_{10}(P) \sim 7.3$).

441 Table 1 Newly identified CAD-associated genomic regions CAD-association results for the lead SNPs from the European and the all ancestry meta-analyses are reported.
442 Note, SNP allele frequencies for each ancestry are provided in, Supplementary Table 5 and in Supplementary Fig. 3 for each of the studies with *de novo* genotyping.

Closest gene(s)	Variant/alleles	Chr:Position (EA AF)	European				All Ancestries				
			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]	$3.99x10^{-8}$	217,782	0.96	[0.94-0.97]	$2.25x10^{-8}$	6.33	243,623
DDX59/*CAMSAP2*	rs6700559C>T	1:200646073 (T 0.47)	0.96	[0.94-0.97]	$2.50x10^{-8}$	221,073	0.96	[0.95-0.97]	$1.13x10^{-8}$	6.68	246,913
LMOD1	rs2820315C>T	1:201872264 (T 0.30)	1.05	[1.03-1.07]	$4.14x10^{-9}$	214,844	1.05	[1.03-1.07]	$7.70x10^{-10}$	7.72	240,685
TNS1[a]	rs2571445G>A	2:218683154 (A 0.39)	1.04	[1.02-1.06]	$3.58x10^{-6}$	194,254	1.05	[1.03-1.06]	$4.55x10^{-10}$	8.41	220,047
ARHGAP26	rs246600C>T	5:142516897 (T 0.48)	1.05	[1.03-1.06]	$1.29x10^{-8}$	210,380	1.04	[1.03-1.06]	$1.51x10^{-8}$	6.39	236,223
PARP12	rs10237377G>T	7:139757136 (T 0.35)	0.95	[0.93-0.97]	$1.70x10^{-7}$	181,559	0.95	[0.93-0.97]	$1.75x10^{-8}$	6.32	207,399
PCNX3	rs12801636G>A	11:65391317 (A 0.23)	0.95	[0.93-0.97]	$1.00x10^{-7}$	211,152	0.95	[0.94-0.97]	$9.71x10^{-9}$	6.64	236,985
SERPINH1	rs590121G>T	11:75274150 (T 0.30)	1.05	[1.03-1.07]	$1.54x10^{-8}$	207,426	1.04	[1.03-1.06]	$9.32x10^{-8}$	5.80	233,249
C12orf43/HNF1A	rs2258287C>A	12:121454313 (A 0.34)	1.05	[1.03-1.06]	$6.00x10^{-9}$	221,068	1.04	[1.03-1.06]	$2.18x10^{-8}$	6.40	246,901
SCARB1	rs11057830G>A	12:125307053 (A 0.16)	1.07	[1.05-1.10]	$5.65x10^{-9}$	177,550	1.06	[1.04-1.09]	$1.34x10^{-8}$	6.49	203,394
OAZ2, *RBPMS2*	rs6494488A>G	15:65024204 (G 0.18)	0.95	[0.93-0.97]	$1.43x10^{-6}$	205,410	0.95	[0.93-0.97]	$2.09x10^{-8}$	6.41	228,578
DHX38	rs1050362C>A	16:72130815 (A 0.38)	1.04	[1.03-1.06]	$2.32x10^{-7}$	216,025	1.04	[1.03-1.06]	$3.52x10^{-8}$	6.16	241,858
GOSR2	rs17608766T>C	17:45013271 (C 0.14)	1.07	[1.04-1.09]	$4.14x10^{-8}$	215,857	1.06	[1.04-1.09]	$2.10x10^{-7}$	5.30	231,213
PECAM1	rs1867624T>C	17:62387091 (C 0.39)	0.96	[0.94-0.97]	$1.14x10^{-7}$	220,831	0.96	[0.95-0.97]	$3.98x10^{-8}$	6.03	246,674
PROCR[a]	rs867186A>G	20:33764554 (G 0.11)	0.93	[0.91-0.96]	$1.26x10^{-8}$	213,505	0.93	[0.91-0.96]	$2.70x10^{-9}$	7.11	239,340

443 [a]These are nonsynonymous SNPs.

444 EA, Effect allele. AF, Effect allele frequency in Europeans. N, Number of individuals in the analysis. $Log_{10}BF$, log base 10 of the Bayes factor obtained from the MANTRA
445 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*-value for heterogeneity
446 across meta-analysed datasets are provided in Supplementary Table 4 and I^2 statistics in Supplementary Fig. 3.

447

448

449 Table 2 Summary of functional data implicating candidate causal genes in newly identified CAD regions. Genes in region, provides genes in the LD block containing
450 the CAD-associated SNP. Phenotype in murine model, lists the phenotype as provided in the mouse genome informatics database, genes are listed if the phenotype affects
451 the cardiovascular system, inflammation or liver function. eQTLs are listed where the SNP or a proxy with r^2> 0.9 are an eQTL for the listed gene in one of the following refs:
452 [12], [13], [26], [43], [44], [45], [46],[38],[47],[48],[14,49] (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
453 based on the most likely given the information ascertained on murine phenotype, eQTL, protein expression and any literature information described in the main text. Loci are
454 further discussed in the Supplementary Information.

SNP	Genes in region	Phenotype in murine model	Cis-eQTLs with SNP (or proxy 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 IL6, abnormal interleukin level)	*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 others), *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,*	*PIF1* (abnormal telomere length)	*ANKDD1A**, *RBPMS2**, *TRIP4**	TRIP4	*TRIP4*

	RBPMS2, SNX1, SNX22, TRIP4, ZNF609				
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 circulatiung C-reactive protein and fibrinogen levels; thrombosis/blood coagulation),	*PROCR**, *EIF6**, *ITGB4BP**	EIF6, ITGB4BP	*PROCR*
rs6088590 C>T			*PROCR**, *GGT7**, *MAP1LC3A**, *ACSS2**, *TRPC4AP**	GGT7	

455

456 * indicates that the eQTL is identified in one of blood (including peripheral blood mononuclear cells) heart, aorta/coronary artery or live. Note the *PCNX3* region also
457 encompasses *AP5B1, ARL2, CAPN1, CDC42EP2, CDCA5, CFL1, CTSW, DPF2, EFEMP2, EHBP1L1, FAM89B, FAU, FRMD8, KAT5, KCNK7, LTBP3, MAP3K11, MRPL49,*
458 *MUS81, NAALADL1, OVOL1.* The *DHX38* region also encompasses, *IST1, MARVELD3, PHLPP2, PKD1L3, PMFBP1, TAT, TXNL4B, ZFHX3, ZNF19, ZNF23, ZNF821*. The

PROCR region also includes: *FAM83C, UQCC1, GDF5, SPAG4, CEP250, C20orf173, ERGIC3, FER1L4, CPNE1, RBM12, NFS1, ROMO1, RBM39, SCAND1, CNBD2, EPB41L1, LINC00657, AAR2, DLGAP4*

Online Methods

Study participants

A full description of the component studies with *de novo* genotyping is given in the Supplementary Information and Supplementary Table 1. In brief, the European (EUR) 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 (SAS) 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 9 sites in Pakistan, and the Bangladesh Risk of Acute Vascular Events (BRAVE) study based in Dhaka, Bangladesh. The East Asian (EA) studies comprised 4,129 CAD cases and 6,369 controls recruited from 7 studies across Taiwan that collectively comprise the TAIwan metaboCHIp (TAICHI) Consortium. The African American (AA) 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 customised version of the Illumina CardioMetabochip (referred to as the "Metabochip+", Illumina, San Diego, USA), 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 Hudson-Alpha and Cedars Sinai (TAICHI[50], WHI, ARIC[51]) and the Broad Institute (MIGen).

Each collection was genotyped and underwent QC separately (Supplementary Tables 1 and 2). In brief, studies genotyped on the Metabochip+ had genotypes assigned using the Illumina GenCall software in Genome Studio. Samples were removed if they had a call rate < 0.97, average heterozygosity >±3 standard deviations away from the overall mean heterozygosity or their genotypic sex did not match their reported sex. One of each pair of duplicate samples and first degree relatives (assessed with a kinship co-efficient > 0.2) were removed.

Across all studies, SNP exclusions were based on minor allele frequency (MAF) < 0.01, $P < 1x10^{-6}$ for Hardy Weinberg Equilibrium or call rate (CR) less than 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 the Metabochip+, namely, CR < 0.99, $P_{HWE} < 1x10^{-4}$ 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 an OR=1.05 for a SNP with MAF=0.2 at $\alpha=5x10^{-8}$, assuming a multiplicative model on the OR scale. For a lower MAF of 0.1 the study would have 0.93 power to detect OR=1.07 at $\alpha=5x10^{-8}$, assuming a multiplicative model. Power calculations were performed using Quanto.

Association with CAD was assessed in studies with de novo genotyping from EUR, SAS, and EA, 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 0/1) was the outcome variable, up to five principal components and the test SNP, coded additively, were included as fixed effects. *P*-values from the score test are reported. The AA studies were analysed using a logistic model in PLINK, with CAD as the outcome variable and SNP coded additively as predictor. The covariates used by each study, including the number of principal components are reported in the Supplementary Information. Genomic inflation was at most 5% for any given study (Supplementary Table 3, Supplementary Fig. 1). A subset of the PROMIS study and EPIC-CVD consortium were 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 been contributed to the CARDIoGRAMplusC4D study and before meta-analyzing our results with the results from CARDIoGRAMplusC4D consortium. The CARDIoGRAMplusC4D SNP association results were converted onto the plus strand of GRh37, checked for heterogeneity and checked to ensure allele frequencies were consistent with EUR 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-ethnic study designs. However, for our studies the data were broadly consistent with the results from METAL (Table 1, Supplementary Table 4) and we therefore primarily report the fixed effect meta-analysis.

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 in CARDIoGRAMplusC4D, EUR, SAS, and EAS respectively and the results were combined using an inverse-variance-weighted fixed effects meta-analysis approach. The conditional analyses were not performed in AA, because the SNP-level case-control counts were not made available for ARIC, MIGen, and WHI. 1000Genome Phase3 v5 ethnic-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 3 ancestries considered in the analyses. As approximately 9% of CARDIoGRAMplusC4D samples were SAS and the remainder EUR, in order to calculate LD for this dataset, we sampled with replacement the genotypes of 50 individuals from the 1000Genome SAS reference panel and combined them with the genotypes of the 503 EUR individuals available in 1000 Genomes. To identify SNPs that are 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 EUR, SAS, EAS and CARIoGRAMplusC4D. The SNPs were identified as independent signals for a specific region, if the conditional $P \leq 1x10^{-4}$. In each region, we performed several rounds of conditional analyses until the conditional *P*-values $>1x10^{-4}$ for all SNPs in the region.

eQTL and epigenetic analyses

The MuTHER dataset 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 14 newly identified CAD regions (IMPUTE info score >0.8) were extracted from the MuTHER project dataset for each of the tissues, LCL (n=777), adipose (n=776) and skin (n=667).

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

566 From each resource, we report eQTL signals, which reach the resource-specific thresholds for
567 significance described above, for SNPs that are in LD (r^2>0.8) with our sentinel SNP.

568 In addition to the publicly available MuTHER and GTeX databases imputed to HapMap and
569 1000Genomes, respectively, we used a curated database of over 100 distinct eQTL datasets to
570 determine whether our lead CAD-associated SNPs or SNPs in high LD with them ($r^2 > 0.8$ in
571 Europeans from HapMap or 1000G) were associated with the expression of one or more nearby genes
572 in cis[57]. Our collated eQTL datasets meet criteria for statistical thresholds for SNP-gene transcript
573 associations as described in the original studies. [57] In total, more than 30 different cells/tissues were
574 queried including, circulating white blood cells of various types, liver, adipose, skin, brain, breast,
575 heart and lung tissues. Complete details of the datasets and tissues queried in the current work can be
576 found in the Supplement Information and Supplementary Table 10, and a general overview of a subset
577 of over 50 eQTL studies has been published[57]. We first identified all sets of eQTLs in perfect LD (r^2
578 =1 among Europeans in HapMap or 1000G) with each other for each unique combination of study,
579 tissue, and transcript. We then determined whether any of these sets of eQTL were either in perfect (r^2
580 = 1) or high LD ($1>r^2> 0.8$) with our lead CAD SNP (Supplementary Table 10).

581 We required that any eQTL had $P<5x10^{-8}$ for association with expression levels to be included in the
582 eQTL tables.

583

584 We examined chromatin state maps of 23 relevant primary cell types and tissues. Chromatin states are
585 defined as spatially coherent and biologically meaningful combinations of specific chromatin marks.
586 These are computed by exploiting the correlation of such marks, including DNA methylation,
587 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 1000Genomes + 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 QC (e.g. coefficient of variation<20%) were log transformed and age, sex, duration between venepuncture and sample processing and the first 3 principal components of genetic ancestry were regressed out. Residuals were then rank-inverse normalized before genomewide association testing using an additive model accounting for imputation uncertainty.

Enrichment analyses

Ingenuity pathway analyses

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

Modified MAGENTA

Given 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 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

613 out of the association results such that $r^2 < 0.2$ for all pairs of SNPs (based on 1,000 Genomes Project
614 data[62]; Supplementary Table 6) prior to implementation of the modified MAGENTA. The null set was
615 defined as the 1,000 remaining QT interval SNPs with the largest *P*-values (least evidence) for
616 association with CAD. The associated set was defined as variants (after LD pruning) that showed
617 evidence of association $P < 1x10^{-6}$. This approach was adopted to select the null and associated sets so
618 as to limit the number of variants included in the hypergeometric cumulative mass function, as a large
619 number of variants results in an intractable calculation for the binomial coefficients. The observed *P*-
620 value from the hypergeometric test is compared to the *P*-values obtained from 10,000 random sets to
621 compute an empirical enrichment *P*-value.

622 *Haploreg: H3K27ac-based tissue enrichment analysis*

623 The associated set as defined for MAGENTA was used for Haploreg analyses and compared to a
624 background set of 12,000 SNPs previously associated with any trait at $P<1x10^{-5}$ (taken from sources
625 such as NHGRI-EBI GWAS catalogue). Using data from HaploReg[15] we counted the number of SNPs
626 with an H3K27ac annotation, or in high LD ($r^2 > 0.8$ from the SNiPA[63] EUR 1000 Genomes maps)
627 with a SNP with an H3K27ac annotation. The significance of the enrichment in H3K27ac marks from
628 a particular tissue was determined by comparing the fraction of associated SNPs with that mark, to the
629 fraction of background SNPs with that same mark. A hypergeometric test was used to assign a *P*-
630 value to the enrichment.

631

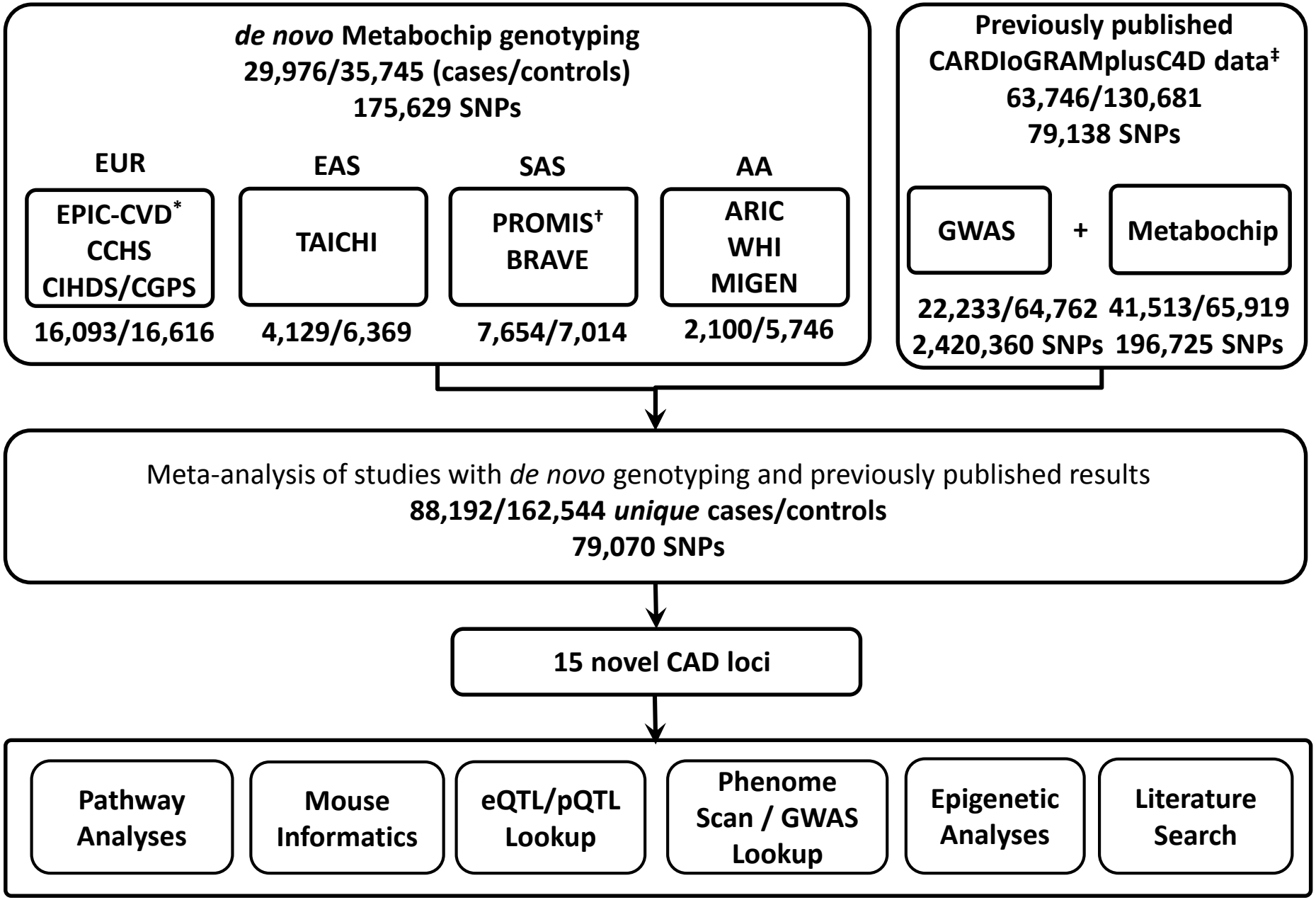
632 ***Data availability***

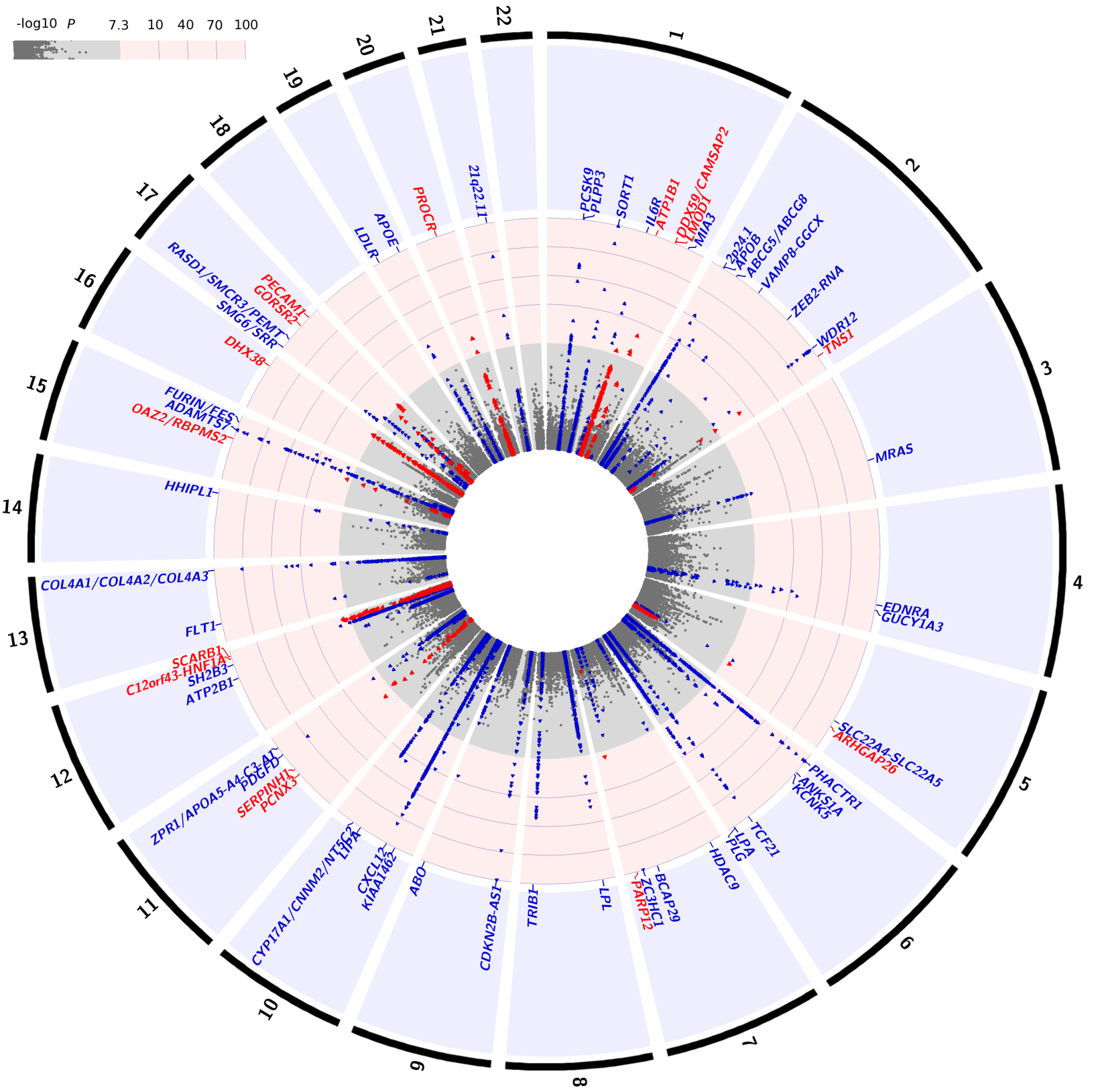
633 The full set of results data from the trans-ancestry meta-analysis and the EUR meta-analysis from this
634 report is available through www.phenoscanner.medschl.cam.ac.uk upon publication.

635

REFERENCES

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**, 559-75 (2007).
53. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-4 (2012).
54. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
55. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* **35**, 809-22 (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**, 369-75, S1-3 (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**, 817-25 (2010).
59. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (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 e19 (2016).
62. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
63. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).


de novo Metabochip genotyping
29,976/35,745 (cases/controls)
175,629 SNPs
EUR
EPIC-CVD*
CCHS
CIHDS/CGPS
16,093/16,616
EAS
TAICHI
4,129/6,369
SAS
PROMIS†
BRAVE
7,654/7,014
AA
ARIC
WHI
MIGEN
2,100/5,746
Previously published
CARDIoGRAMplusC4D data‡
63,746/130,681
79,138 SNPs
GWAS
+
Metabochip
22,233/64,762
2,420,360 SNPs
41,513/65,919
196,725 SNPs
Meta-analysis of studies with *de novo* genotyping and previously published results
88,192/162,544 *unique* cases/controls
79,070 SNPs
15 novel CAD loci
Pathway Analyses
Mouse Informatics
eQTL/pQTL Lookup
Phenome Scan / GWAS Lookup
Epigenetic Analyses
Literature Search


-log10 P
7.3
10
40
70
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
PCSK9
PLPP3
SORT1
IL6R
ATP1B1
PDX59/CAMSAP2
LMOD1
MIA3
2p24.1
APOB
ABCG5/ABCG8
VAMP8-GGCX
ZEB2-RNA
WDR12
TNS1
MRAS
EDNRA
GUCY1A3
SLC22A4-SLC22A5
ARHGAP26
PHACTR1
ANKS1A
KCNK5
TCF21
LPA
PLG
HDAC9
BCAP29
ZC3HC1
PARP12
LPL
TRIB1
CDKN2B-AS1
ABO
CXCL12
KIAA1462
LIPA
CYP17A1/CNNM2/NT5C2
ZPR1/APOA5-A4-C3-A1
PDGFD
SERPINH1
PCNX3
COL4A1/COL4A2/COL4A3
FLT1
SCARB1
C12orf43-HNF1A
SH2B3
ATP2B1
HHIPL1
FURIN/FES
ADAMTS7
OAZ2/RBPMS2
DHX38
PECAM1
GORSR2
RASD1/SMCR3/PEMT
SMG6/SRR
LDLR
APOE
PROCR
21q22.11