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

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

PROCR encodes the endothelial protein C receptor (EPCR). We found the G allele at rs867186 (which codes for the glycine residue at p.Ser219Gly) in *PROCR* confers protection from CAD (OR[95%CI]=0.93[0.91-0.96]; Table 1, 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 the EPCR more susceptible to proteolytic cleavage, resulting in increased shedding of membrane-bound EPCR from the endothelial surface[30] 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 1000G EUR samples; Supplementary Fig. 8), an intronic SNP in *NCOA6* with the T allele conferring increased risk of CAD (conditional on rs867186, conditional P=1.14x10^{-5}, OR[95% CI]=0.97[0.95-0.98]). No additional SNPs were associated with CAD after conditioning on rs867186 and rs6088590 (P>0.01).

Five of the novel CAD 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, β=0.11, P=1.43x10^{-11}). The protein is found in smooth muscle cells (SMC)[32,33]. *In vitro* and transgenic mouse studies demonstrate an essential requirement for CArG elements in the expression of LMOD1 through both serum response factor (SRF) and myocardin (MYOCD)[35]. Myocardin has emerged as 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 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, 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, AM, JBW, 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).

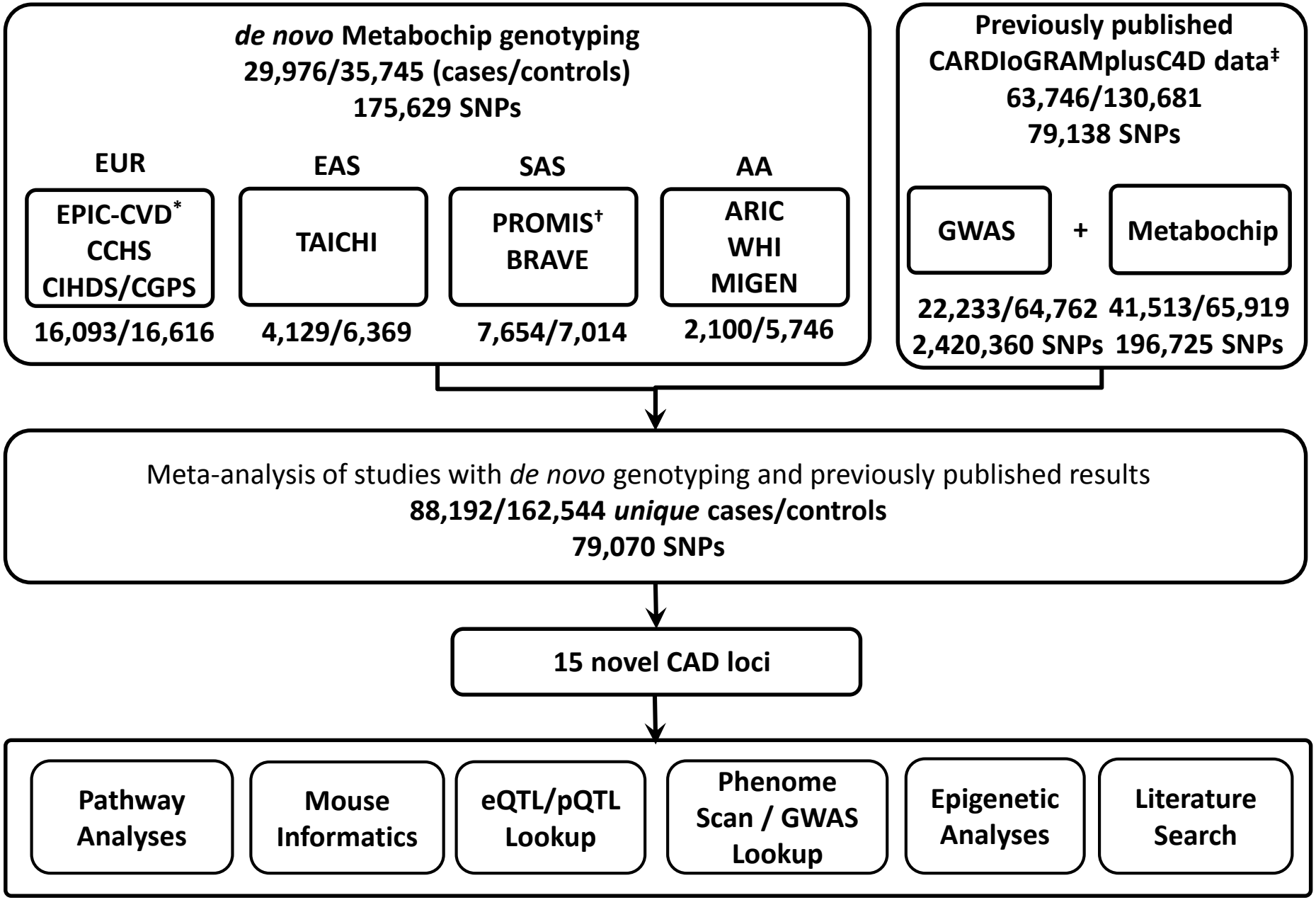
23. Johansson, A. *et al.* Identification of genetic variants influencing the human plasma proteome. *Proc Natl Acad Sci U S A* **110**, 4673-8 (2013).
24. Holme, I., Aastveit, A.H., Hammar, N., Jungner, I. & Walldius, G. Haptoglobin and risk of myocardial infarction, stroke, and congestive heart failure 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 ($-\log_{10}(P) \sim 7.3$).


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

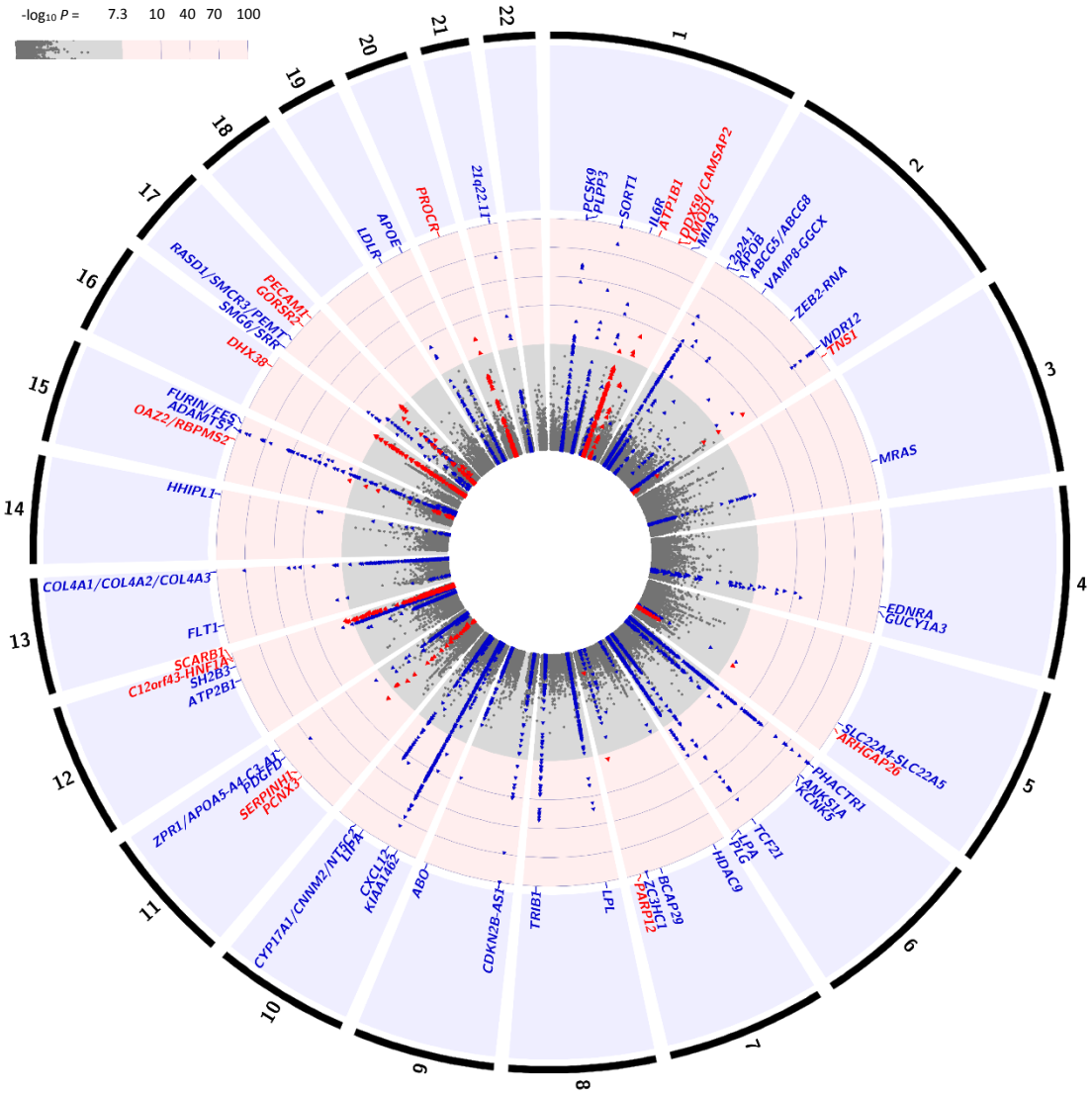

-log10 P = 7.3 10 40 70 100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
PCSK9
PLPP3
SORT1
ATP1B1
PDYS9/CAMSAP2
MIA3
2q24.1
APOB
ABCC5/ABCG8
VAMP8-GGCX
ZEB2-RNA
WDR12
TNS1
MRAS
EDNRA
GUCY1A3
SLC22A4-SLC22A5
ARHGAP26
PHACTR1
ANKS1A
KCNK5
TCF21
LPA
LPG
HDAC9
BCAP29
FAM213A
PARP12
LPL
TRIB1
CDKN2B-AS1
ABO
CXCL12
KIAA1462
LIPA
CYP17A1/CNNM2/NT5C2
ZPR1/APOA5-A4-C3-A1
PDGFD
SERPINH1
PCNA3
C12orf43
HNF1A
SCARB1
SH2B3
ATP2B1
FLT1
COL4A1/COL4A2/COL4A3
HHIPL1
FURIN/FES
ADAMTS7
OAZ2/RBPMS2
DHX38
RASD1/SMCR3/PEMT
SMG6/SRR
PECAM1
GOSR2
LDLR
APOE
PROCR
21q22.11

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. 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**

[a]These are nonsynonymous SNPs.

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 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 across meta-analysed datasets are provided in Supplementary Table 4 and I^2 statistics in Supplementary Fig. 3.

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 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 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: [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 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 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, 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 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	

***** 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 encompasses *AP5B1, ARL2, CAPN1, CDC42EP2, CDCA5, CFL1, CTSW, DPF2, EFEMP2, EHBP1L1, FAM89B, FAU, FRMD8, KAT5, KCNK7, LTBP3, MAP3K11, MRPL49, 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).

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

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

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

We required that any eQTL had $P<5x10^{-8}$ 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 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

out of the association results such that $r^2 < 0.2$ for all pairs of SNPs (based on 1,000 Genomes Project data[62]; Supplementary Table 6) prior to implementation of the modified MAGENTA. The null set was defined as the 1,000 remaining QT interval 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 $P < 1x10^{-6}$. 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 $P<1x10^{-5}$ (taken from sources such as NHGRI-EBI GWAS catalogue). 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] EUR 1000 Genomes 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 set of results data from the trans-ancestry meta-analysis and the EUR meta-analysis from this report is available through www.phenoscanner.medschl.cam.ac.uk upon publication.

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).

Supplementary Information

Supplementary Note

Supplementary Tables

Supplementary Table 1: (a) Study-specific *sample* quality control exclusions and baseline characteristics of the studies with *de novo* genotyping. (b) Study-specific definitions of disease outcome (CAD).

Supplementary Table 2: Summary of study-specific *SNP genotype* quality control of the studies with *de novo* genotyping.

Supplementary Table 3: Inflation factors for the studies with *de novo* genotyping.

Supplementary Table 4: Results of CAD association tests from the European and All ancestry meta-analyses for SNPs with $P<5x10^{-8}$ at the new loci.

Supplementary Table 5: Results of meta-analyses across the studies (including all samples) with *de novo* genotyping at the new CAD-associated SNPs reported in Supplementary Table 4.

Supplementary Table 6 Summary of *P*-values for the null and associated sets used in the modified MAGENTA pathway analyses with the hypergeometric tests.

Supplementary Table 7 Pathways with enrichment of CAD associated variants from the all ancestry meta-analyses identified using modified MAGENTA.

Supplementary Table 8 Ingenuity pathway analysis results. (a) identified canonical pathways (b) identified upstream regulators.

Supplementary Table 9 Co-ordinates for the genomics regions used for each new CAD locus

Supplementary Table 10 eQTL look ups of the 15 new CAD-associated regions.

Supplementary Table 11 Haploreg enrichment analyses of H3K27ac enhancer marks.

Supplementary Table 12: Association results for the established CAD loci.

Supplementary Figures

Supplementary Figure 1 QQ plots illustrating array-wide inflation for each of the studies with *de novo* genotyping.

Supplementary Figure 2 Manhattan plot showing the association of ~79,000 variants with CAD from the European meta-analysis in up to ~221,000 individuals.

Supplementary Figure 3 Forest plots from the all studies meta-analysis for the 15 sentinel CAD-associated SNPs.

Supplementary Figure 4 Regional association plots for novel CAD associated loci (a) from the all ancestry meta-analysis for 13 loci and the European meta-analysis for *GOSR* and *SERPINH1*. (b) from the publicly available CARDIoGRAMplusC4D 1000G imputed GWAS results[1].

Supplementary Figure 5 Manhattan plot for the association of the Metabochip SNPs in the studies with *de novo* genotyping.

Supplementary Figure 6 *SCARB1* regional association plots with (a) CAD (b) HDL[2] (c) LDL[2] and (d) triglycerides[2]. Physical position is given for GRCh37.

Supplementary Figure 7 Annotation of the *SCARB1* region using publicly available transcriptomic and epigenomic reference data sets

Supplementary Figure 8 Association of the PROCR gene region

Study descriptions

Baseline characteristics of the contributing studies are summarized in **Supplementary Table 1.**

The Copenhagen Ischaemic Heart Disease Study (CIHDS)

This study comprised 2,724 cases with myocardial infarction and other major acute coronary syndromes and 2,815 controls matched by age and sex from the Copenhagen General Population Study (CGPS) described below. The cases were recruited from Copenhagen University Hospital during the period from 1991 to 2009. In addition to a diagnosis of acute coronary syndrome, these cases also had stenosis or atherosclerosis on coronary angiography and/or positive results on exercise electrocardiography. Cases were classified by World Health Organization International Classification of Diseases-Eighth Revision, codes 410 to 414; International Classification of Diseases-Tenth Revision, codes I20 to I25, and through review of all hospital admissions and diagnoses entered in the national Danish Patient Registry and all causes of death entered in the national Danish Causes of Death Registry, as previously described[3].

The Copenhagen General Population Study (CGPS)

The CGPS is a population-based prospective study initiated in 2003 with ongoing enrolment[3]. Participants were selected on the basis of the national Danish Civil Registration System to reflect the adult Danish population age 20 to ≥80 years. Data were obtained from a questionnaire, a physical examination, and blood samples including deoxyribonucleic acid extraction. Follow-up was 100% complete; that is, no participant was lost to follow-up. As noted above, individuals free of coronary heart disease at the time of examination were selected to serve as controls for CIHDS (Copenhagen Ischemic Heart Disease Study).

Copenhagen City Heart Study (CCHS)

CCHS is a population-based prospective study initiated in 1976 with follow-up examinations from 1981 to 1983, 1991 to 1994, and 2001 to 2003[4]. Selection of individuals for the CCHS was based on the same criteria as for the CGPS. Information on diagnosis of CAD (defined as WHO ICD 8 410 to 414 and WHO-ICD 10 I20 to I25) was collected and verified from 1976 until 2010 by reviewing all hospital admissions and diagnoses entered in the national Danish Patient Registry, and by reviewing all causes of death entered in the national Danish Causes of Death Registry[4,5]. Again, follow-up was 100% complete for both non-fatal coronary outcomes and mortality.

European Investigation into Cancer and Nutrition-CVD (EPIC-CVD)

EPIC is a multi-centre prospective cohort study[6] of 519,978 participants (366,521 women and 153,457 men, mostly aged 35–70 years) recruited between 1992 and 2000 in 23 centres located in 10 European countries. Participants were invited mainly from population-based registers (Denmark, Germany, certain Italian centres, the Netherlands, Norway, Sweden, UK)[7]. Other sampling frameworks included: blood donors (Spain and Turin and Ragusa in Italy); screening clinic attendees (Florence in Italy and Utrecht in the Netherlands); people in health insurance programmes (France); and health conscious individuals (Oxford, UK)[7]. About 97% of the participants were of white European ancestry. Prevalent CAD was ascertained through self-reported history of MI or angina, or registry-ascertained CAD event prior to baseline. EPIC-CVD employs a nested case-cohort design, analogous to the EPIC-InterAct study for type-2 diabetes[8] which established a common set of referents through selection of a random sample of the entire cohort ("subcohort"). Incident CAD cases

have been defined as fatal and non-fatal MI and other major acute coronary events, according to ICD-10 codes I20-I25. All centres have recorded cause-specific mortality through mortality registries and/or active follow-up, and have ascertained and validated incident fatal and non-fatal CAD through a combination of methods (eg, morbidity registers, general practice records, MONICA registries, self-report, clinical records[7]).

Bangladesh Risk of Acute Vascular Events (BRAVE)

BRAVE is a retrospective case-control study of first-ever confirmed acute myocardial infarction (MI) in Bangladesh. Patients (male or female; age between 30-80 years) admitted to the emergency rooms of the collaborating hospital in Dhaka, Bangladesh were eligible for inclusion as MI cases if they fulfilled all of the following criteria: i) presented within 24 hours of the onset of sustained clinical symptoms suggestive of MI lasting longer than 20 minutes, including chest pain and breathlessness; ii) had ECG changes indicative of MI (new pathologic Q waves, at least 1 mm ST elevation in any 2 or more contiguous limb leads or a new left bundle branch block, or new persistent ST-T wave changes diagnostic of a non-Q wave MI) with a subsequent confirmation by troponin-I measurements; and iii) had no previous cardiovascular diseases; defined as self-reported history of angina, MI, coronary revascularisation, transient ischaemic attack, stroke or evidence of CAD on prior ECG or in other medical records. Participants were not recruited into BRAVE if any of the following features had been evident: i) a previous history of cardiovascular disease (including self-reported MI, angina, coronary revascularization, stroke, transient ischaemic attack, or peripheral vascular disease, and, in cases, presence of cardiogenic shock); ii) a history of a viral or bacterial infection in the previous 2 weeks; iii) current hospitalization for acute cerebrovascular events; iv) MI secondary to any surgery; v) documented chronic conditions, such as malignancy, any chronic infection, leprosy, malaria or other bacterial/parasitic infections, chronic inflammatory disorders, hepatitis or renal failure on past medical history; vi) pregnancy or related conditions; or vii) unable to provide consent. Controls were hospital based and frequency-matched to cases on age (within 5 year age bands) and sex, and without a self-reported history of cardiovascular disease.

Pakistan Risk of Myocardial Infarction Study (PROMIS)

PROMIS is an ongoing retrospective case-control study of first-ever confirmed acute MI in Pakistan. Since 2005, the study has enrolled close to 18,500 MI cases and equivalent number of controls; the present investigation has included all MI cases and controls that had been enrolled until 2011. Patients aged 30-80 years who were admitted to the emergency rooms of nine recruitment centres across Pakistan [9] were eligible for inclusion as cases if they fulfilled all of the following criteria: symptoms within 24 hours of hospital presentation; typical ECG changes; and positive troponin-I test. To identify referents from approximately the same source population as the cases, controls were identified contemporaneously in the same hospitals as the index cases and selected from among people who had no history of CVD and who were: visitors of patients attending the outpatient department; patients attending outpatient departments for routine non-cardiac complaints; or non-blood relatives visiting index MI cases. Controls were frequency-matched to MI cases by sex and age (5-year bands). People with recent illnesses or infections were not eligible.

ARIC

The ARIC study is a multi-center cohort and community surveillance investigation in predominantly bi-racial populations (white and African Americans)[10]. ARIC recruited 15,792 individuals of which,

4,266 were African Americans. Individuals were aged 45-64 years and from four communities in Forsyth County, N.C., Jackson, M.S., Minneapolis, M.N., and Washington County, M.D. Baseline examination occurred between 1987-1989, with four follow-up examinations. Annual follow-up and community surveillance identified CAD events including hospitalizations and deaths which were then classified by an expert panel of physicians based on review of hospital records, death certificates and interviews of next of kin[10]. CAD events were defined as acute hospitalized MI (definitive or probable), definite fatal CAD, or ECG diagnosis of MI. Acute MI was defined based on criteria that included cardiac pain, cardiac markers and ECG readings. Events through December 31st, 2007 are included. After genotyping quality control and exclusion of prevalent CAD cases, 3204 African American participants 366 of which had incident CAD events were included in this study. All participants included in these analyses gave consent for genetic studies and data sharing.

WHI

WHI is a prospective study investigating post-menopausal women's health in the U.S[11]. A total of 161,838 women aged 50–79 years old were recruited from 40 US clinical centers between 1993 and 1998 to participate in an observational study (OS) and in three clinical trials (CT). Annual (OS) and semi-annual (CT) follow-up identified self-reported events which were then classified by an expert panel of physicians based on review of hospital records, death certificates and interviews of next of kin[12]. A subset of 2,200 WHI African American women was selected to be genotyped with the CardioMetaboChip by the Population Architecture using Genomics and Epidemiology (PAGE) study[13] investigators. Women were selected for genotyping on the basis of DNA and biomarker availability, and consent. CAD was defined as acute hospitalized MI (definitive or probable) and definite fatal CAD. Acute MI was defined based criteria that included cardiac pain, cardiac markers and ECG readings. Follow-up of events in WHI were through August 2009. The final sample after genotyping quality control and exclusion of prevalent self-reported CAD was up to 1954 with 99 incident CAD events. All participants included in these analyses gave consent for genetic studies and data sharing. Additional study descriptions are shown in Supplementary Table 1.

MIGen

Involves a conglomerate of six MIGen studies focused exclusively on African American (AA) ancestry and included: 565 from Multi-ethnic Study of Atherosclerosis (MESA); 700 from the Cleveland Clinic GeneBank; 410 from the International Verapamil SR/Trandolapril Study (INVEST); 324 from Translational Research Investigating Underlying Disparities in Acute Myocardial Infarction Patients' Health Status (TRIUMPH); 469 from Penn Medicine Biobank, and 315 from Emory GeneBank .[14]

TAIwan metaboCHIp Consortium (TAICHI)

The TAICHI consortium is formed of seven studies through a collaborative effort between investigators based in the U.S. and Taiwan. The main U.S academic sites participating in the TAICHI consortium include Stanford University School of Medicine in Stanford, California; Hudson-Alpha Biotechnology Institute in Huntsville, Alabama; and Harbor-UCLA in Los Angeles, California. The main academic sites in Taiwan include National Health Research Institute (NHRI); National Taiwan University Hospital (NTUH); Taipei and Taichung Veteran's General Hospitals (VGH) and Tri-Service General Hospital (TSGH). These investigators have assembled a large, well-phenotyped sample set consisting of >13,000 Han Chinese from seven existing studies[15-19]. The consortium aims to identify genetic determinants of atherosclerosis and diabetes related traits in East Asians and to fine map validated loci identified in other race/ethnic groups.

A majority of coronary artery disease (CAD) cases in TAICHI were ascertained through hospital based studies enrolling subjects admitted for coronary angiography and/or clinical complications of CAD. These subjects were labelled as a case if a chart review by a qualified MD (most often a cardiologist) revealed that the subject either currently or in the past was suffering from a myocardial infarction, an acute coronary syndrome, angina, or demonstrated at least one epicardial coronary artery obstruction of >50% on coronary angiogram. A small minority of cases were identified among the non-hospital based prospective cohort studies through a self-report of either having suffered an MI, having undergone one or more procedures related to clinical complications of CAD, or having an ECG diagnostic of a prior q wave myocardial infarction or an ongoing ST-segment elevation MI based on the Minnesota Code[20]. Subjects who had no previous history of clinical CAD who were found to have sub-occlusive disease on angiogram (*i.e.* some evidence of atherosclerosis but no epicardial coronary artery obstruction of >50%) were excluded (*i.e.* they were neither considered a case or a control). All other subjects were considered controls.

1. **Taiwan Coronary Artery Disease GENetic (TCAGEN) study (PI - Dr. Jyh-Ming Juang)** is an ongoing cohort study that has been enrolling patients undergoing coronary angiography or percutaneous intervention at the National Taiwan University Hospital (NTUH) in the setting of either stable angina pectoris or prior myocardial infarction[19]. Participants are from both the north of Taiwan where the main NTU medical school/hospital is located, and from the Yulin branch of NTUH, located in south/central Taiwan. The hospital uses an elaborate electronic medical record system that provides access to clinic visit notes, diagnostic codes of clinic encounters, prescriptions, and laboratory data in a searchable form. Fasting blood samples were collected before cardiac catheterization while peripheral blood was collected in the catheter lab specifically for buffy coat isolation and DNA extraction.
2. **Taichung CAD (TCAD) study (PI - Dr. Wayne Huey-Herng Sheu)** includes patients with a variety of cardiovascular diseases receiving care at the Taichung Veterans General Hospital. Specifically, individuals who were hospitalized for diagnostic and interventional coronary angiography examinations and treatment are included in TAI CHI[16]. Also included in TAI CHI are subjects with a history of myocardial infarction or revascularization of any type (percutaneous coronary intervention or coronary artery bypass).
3. **TAiwan Coronary and Transcatheter intervention (TACT) cohort study (PI Dr. Tzung-Dau Wang)** enrolled patients with angina pectoris and objective documentation of myocardial ischemia who underwent diagnostic coronary angiography and/or revascularization any time after October 2000 at the National Taiwan University Hospital (NTUH)[18]. This cohort is very similar to TCAGEN but was collected independently. Participants provided clinically relevant information including use of cardiovascular related medication through a standardized questionnaire. Clinically relevant information is also available through a comprehensive electronic medical records database that includes information on drug use and surgical interventions. Fasting blood samples were collected before cardiac catheterization.
4. **Taiwan Diabetes and RelAted Genetic COmplicatioN (Taiwan DRAGON) cohort study (PI - Dr. Wayne Huey-Herng Sheu)** of type 2 diabetes (T2D) at the Veteran's General Hospital in Taichung, Taiwan (Taichung VGH)[17]. Participants include individuals with either newly diagnoses or established diabetes who visit the diabetes outpatient clinic on a regular basis. Subjects with hyperglycemia who do not meet criteria for T2D defined by IDF are not included. Individuals participate in a health examination program at Taichung VGH are also interviewed. Specialized tests include an oral glucose tolerance tests (OGTT) in subjects without an established diagnosis of diabetes.

5. **Taiwan USA Diabetes Retinopathy (TUDR) cohort study (PI - Dr. Wayne Huey-Herng Sheu)** enrolled subjects with T2D receiving care at Taiching Veteran's General Hospital, a small number of subjects were included from Tri-Service General Hospital (TSGH)[17]. All TUDR subjects underwent a complete fundoscopic examination to carefully document the presence and extent of retinopathy. To date, a total of 2,222 unrelated T2D subjects with and without retinopathy were ascertained and have undergone metabochip genotyping. Of the 2,222 subjects, 1,201 were T2D without eye diseases, 479 were T2D with NPDR and 542 T2D with PDR. In addition to DNA and buffy coats, fasting blood for future measurement of serum/plasma biomarkers has also been banked. A variety of additional clinical related phenotypes are available. All 2,222 overlaps with the Taiwan Dragon Study.
6. **Healthy Aging Longitudinal Study in Taiwan (HALST) (PI – Dr. Agnes Chao Hsiung)** is a population based multi-site cohort study of ambulatory adults aged > 55 years living in 7 major geographic regions of Taiwan, established by the NHRI[21]. The aim of the study is to investigate the multidimensional determinants, including lifestyle, genetic, metabolic, and inflammatory factors, of an older Asian population. These 7 locations include both urban and rural areas: two are in the north (Taipei's Shilin District and Taoyuan County's Yangmei Township), two in central Taiwan (Miaoli City in Miaoli County and Changhua City in Changhua County), two in the south (Puzi, Chiayi County, and Kaohsiung's Lingya District), and one in the east (Hualien City/County). The only exclusion criteria are presence of highly contagious diseases, advanced illnesses with limited life span or bedridden status, dementia, other advanced neurological deficit, severe hearing loss, and institutionalization in a chronic care facility for any reason. Over 5000 subjects have been recruited over a five-year period (2008-2012) from seven recruitment sites across the country. Follow-up in person visits are currently ongoing and will continue throughout a second 5-year study cycle scheduled that began in 2013 (~1000 subjects / year). Within each wave, participants are to be followed up by telephone contact every year for vital status and for updates on health-related conditions. Medical records are requested to confirm the development of any new health conditions. Vital status, health claims, health care utilization data are being collected for the cohort on a regular basis by linking to the National Death Registry Database and the National Health Insurance Database. HALST served as one the main "control" cohorts for this study after exclusions of subjects with a self-report of CAD or a ECG diagnostic of prior MI.
7. **Stanford-Asian Pacific Program in Hypertension and Insulin Resistance (SAPPHIRe) family based study (PIs – Dr. Thomas Quertermous, Agnes Chao Hsiung, and Wayne Huey-Herng Sheu**) was established in 1995 with an initial goal of identifying major genetic loci underlying hypertension and insulin resistance through linkage in East Asian populations. SAPPHIRe was also one of four networks participating the NHLBI's Family Blood Pressure Program (FBPP)[15]. At the outset, SAPPHIRe involved recruitment sites in the San Francisco Bay Area, Hawaii, and Taiwan. However, a majority of the ~1,700 sibpairs in SAPPHIRe were recruited from 3 centers in Taiwan (NTUH, Taipei VGH and Taichung VGH) with NHRI being the DCC. Sibpairs were either highly concordant or discordant for blood pressure and a subset underwent an insulin suppression test. Many metabolic variables associated with blood pressure and insulin resistance were examined in the first 5-year investigative cycle funded by the NIH (1995-2000). Further extensive phenotyping through return visits and regular follow ups occurred between 2001 and 2008 in the Taiwanese SAPPHIRe participants which included echocardiographic and multi-detector row CT imaging procedures. These efforts were facilitated by a programmatic collaboration between the NHLBI's FBPP and the National Health Research Institute in Taiwan. Like

HALST, SAPPHIRe served predominantly as a “control” cohort in this study. Only one sib per family was included as a control in this study.

Two of the TAICHI studies (Taiwan DRAGON and TUDR) were T2D cohorts and so T2D cases that had been diagnosed with CAD were included as cases in the CAD analyses, while the remaining T2D samples were included as controls.

Pathway and network analyses

Modified MAGENTA

Given the Metabochip comprises a select set of SNPs and lacks complete genomic coverage[22], 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[23]. 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 out of the association results such that $r^2 < 0.2$ for all pairs of SNPs (based on 1,000 Genomes Project data[24]; www.1000genomes.org; Supplementary Table 6) prior to implementation of the modified MAGENTA. The null set was defined as the 1,000 remaining QT interval 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 $P < 1x10^{-6}$. 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.

An analysis of European, and all ancestry meta-analyses are reported. A total of 47,468 SNPs (of which 2,937 were QT interval SNPs) mapped to 11,190 genes and could be included in the European analysis, whilst 61,223 SNPs (3,403 of which were QT interval SNPs) mapped to 11,904 genes were included in the all ancestry analysis. Within the null set of the European analysis 873 genes were covered by the 1,000 null SNPs, whilst within the associated set 73 genes were covered by 76 SNPs. For the all ancestry analysis, 887 genes were covered by the 1,000 null SNPs and 78 genes were covered by 85 SNPs in the associated set. Sensitivity analyses to specific parameters used in the modified MAGENTA analyses were assessed. Sensitivity to the *P*-value threshold for inclusion in the associated set of variants was tested at $P<10^{-5}$, $P < 10^{-7}$; the number of variants included in the null set of variants was set to 900 and 1,100; known CAD regions (identified in the NIH Catalog of Published Genome-Wide Association Studies, https://www.genome.gov/26525384) were removed; the newly identified CAD loci were removed; the *COL4A1* and *COL4A2* genes that appear in the associated sets for several enriched pathways were excluded and the number of random sets used to calculate the empirical enrichment *P*-value by was changed to 1,000 and 100,000 random sets.

Seven databases (BioCarta www.biocarta.com/genes/indexasp, Kyoto Encyclopedia of Genes and Genomes [KEGG], www.genome.jp.kegg, Ingenuity, www.ingenuity.com, Panther, Panther Biological Processes and Panther Molecular Functions www.pantherdb.org, and Reactome, www.reactome.org) comprising 1,558 pathways, were tested for enrichment of genes associated with CAD. There were 23 pathways (18 independent) with $P<0.01$ from the European only pathway analysis and 19 pathways (16 independent) with $P<0.01$ from the all ancestry pathway analysis. (A more stringent significance threshold of $p<0.01$ was used rather than the more conventional $P\leq0.05$ so as to minimise the number of enriched pathways identified.) Ten pathways were in common between these analyses. Independence of pathways were determined by pathway gene content, if a pathway was a subset of another then it was deemed dependent. For example, the Reactome cell surface interactions at the vascular wall pathway (93 genes) is contained in the Reactome hemostasis pathway (272 genes). The chylomicron mediated lipid transport pathway (17 genes) is contained in the

Reactome lipoprotein metabolism pathway (27 genes), which is itself contained in the Reactome metabolism of lipids and lipoproteins pathway (228 genes).

The strongest evidence for enrichment in the European only analysis was for the KEGG glycerolipid metabolism pathway (49 genes, $P<3x10^{-5}$). The strongest evidence for enrichment from the all ancestry analysis was shown by the Reactome lipoprotein metabolism pathway (27 genes, $P<3x10^{-5}$). Generally pathways involved in lipid metabolism were the most enriched (11 of the 32 with $P<0.01$ in the European or all ancestry analysis).

The sensitivity analyses revealed that changing the number of random sets to 1,000 or 100,000 instead of 10,000 as used in the main analysis, resulted in the same pathways being identified in the European and the all ancestry analyses ($P<0.01$). Exclusion of *COL4A1/2* from the associated set resulted in fewer pathways being enriched, however, all but one of those enriched for the all ancestry pathway analysis were identified by the main analyses. Inclusion of less ($P<10^{-7}$) associated variants resulted in the loss of several pathways that were identified by the main analysis. Use of less associated variants ($P<1x10^{-7}$) identified fewer pathways as expected, however, most of these were identified by the main analysis. A more liberal *P*-value threshold ($P<1x10^{-5}$) for inclusion of variants in the associated set produced more enriched pathways than the main analysis, with most of the pathways identified by the main analysis being detected. The all ancestry sensitivity analysis with more associated variants detected several unique pathways. Use of less null variants (N=900) generally identified the same pathways as the main analysis. However, inclusion of more variants in the null set (N=1,100) resulted in an attenuation of pathways that were enriched. Most pathways identified under this sensitivity were detected by the main analysis. Removal of known and novel CAD loci generally resulted in less pathways being identified. The pathways found to be enriched from the sensitivity analyses that were not detected by the main analyses did not give additional insights into novel biological pathways involved with CAD. Three positive control pathways were also tested for enrichment of variants associated with CAD (Supplementary Table 7). The CAC and CAD pathways were significantly enriched for variants associated with CAD (CAC pathway *P*-value range: $0.00\text{-}1x10^{-5}$; CAD pathway $P = 0.00$ for all 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. IPA mapped 41,480 of the ~78,954 SNPs in our meta-analysis to ~8,894 RefSeq genes (*i.e.* the reference set of genes). Given the ~79,000 SNPs examined were primarily preselected candidate SNPs for association with CAD or its risk factors[22] and CAD has a complex genetic architecture the appropriate *P*-value cut-off to select SNPs for inclusion in the pathway analysis was unclear[25,26]. Therefore, six *P*-value thresholds ($5x10^{-7}$, $5x10^{-6}$, $5x10^{-5}$, 0.0005, 0.005, and 0.05) were considered. The number of focus genes increased as the *P*-value threshold was lowered (76, 142, 228, 402, and 909, 2,439 for the six *P*-value thresholds). IPA uses a right-tailed Fisher Exact Test to test for statistically significant over-representation of focus genes in a given canonical pathway among the genes with SNPs with low *P*-values compared to the reference set of genes.

We also used the IPA to identify potential upstream regulators of genes with SNP(s) with low *P*-values

[http://pages.ingenuity.com/rs/ingenuity/images/0812%20upstream_regulator_analysis_whitepaper.pdf]. Upstream regulators were not necessarily represented by SNPs on the Metabochip but could still be expected to play an important role in the pathogenesis of CAD.

The use of a liberal *P*-value thresholds in IPA revealed evidence of enrichment for the, sildenafil, PPARα/RXRα Activation, Protein Kinase A signaling, and the Axonal Guidance Signaling pathways (Supplementary Table 8). We note analyses of the CARDIoGRAM GWAS data with only partial overlap with subjects examined here and using a different gene-set enrichment analysis algorithm also identified the axonal guidance pathways as relevant to CAD[27]. While axon guidance pathways modulate diverse biological phenomena within the nervous system, there is growing evidence that neural guidance cues play important roles outside the nervous system. For example, Netrin-1 is secreted by macrophage foam cells in atherosclerotic plaques and acts to inhibit emigration of these cells out of lesions by causing dysregulation of the actin cytoskeleton[28] and semaphorin 3A, expressed in coronary artery endothelial cells, potently inhibits chemokine-directed migration of human monocytes[29,30].

The tests for enrichment of genes associated with CAD[23] (Supplementary Tables 6 and 7) and analyses using the Ingenuity Pathway Analysis (IPA) software (Supplementary Table 8) identified known CAD associated pathways, such as metabolism of lipids and lipoproteins, farnesoid X receptor (FXR)/retinoid X receptor (RXR) activation and liver X receptor (LXR)/RXR activation. Evidence of enrichment using IPA with a P=0.05 cut-off was obtained for a number of pathways including the PPARα/RXRα Activation, and Protein Kinase A signaling pathways (P=3.98x10^{-11} and 3.16x10^{-11}, respectively) which could indicate new areas of biology to investigate.

Further information on the new CAD gene regions

ATP1B1

The sentinel CAD associated SNP, rs1892094, is located at intron 3 of the *ATP1B1* gene, encoding for the Na^+/K^+ ATPase beta subunit 1. Several GWASs have identified common variants at this locus associated with electrocardiographic parameters, including QT interval[31-33]. However, the implicated variants are not in LD with rs1892094 ($r^2<0.2$, 1000 Genomes EUR). The CAD SNP is however, associated with expression of *ATP1B1* in atherosclerotic root ($P=5.24x10^{-24}$)[34]. A recent study has reported an association of the region with pulse pressure[35] with a SNP (rs7519279) that is in LD with the CAD associated SNP ($r^2=0.44$, D'=0.98 in 1000Genomes EUR). In mouse, mutations in this gene have resulted in increased heart mass and cardiac hypertrophy, which could suggest a common mechanism. Together these findings make *ATP1B1* an interesting candidate gene.

Expression studies, however, also highlight *NME7* as a possible candidate. The CAD SNP rs1892094 is associated with *NME7* gene expression levels in LCLs ($P=4.82x10^{-13}$) adipose ($P=9.46x10^{-14}$)[36], aorta ($P=2.39x10^{-14}$)[37], peripheral blood mononuclear cells ($P=7.98x10^{-18}$)[38] and monocytes ($P=1.1\times10^{-11}$)[39]. *NME7* encodes the protein NME/NM23 family member 7 that is found in high abundance in many tissues including liver and kidney. However, there is no compelling cardiovascular phenotypes reported for this gene in mouse and its possible gene function with regards to the pathobiology of cardiovascular disease is elusive.

TNS1

The non-synonymous CAD associated SNP, rs2571445 (W1197R) has previously been associated with pulmonary function[40-42]. Repapi et al. (ref [40]), also showed *TNS1* was expressed in lung tissue, bronchial epithelial cells, airway smooth muscle cells and peripheral blood mononuclear cells in human. The CAD risk allele is associated with increased expression of TNS1 in adipose ($\beta=0.12$, $P=8.88x10^{-10}$)[36] and peripheral blood ($P=1.81x10^{-24}$)[43]. The encoded protein is found in many tissues including smooth muscle cells and heart. Animal models have shown that this gene causes abnormal kidney morphology, kidney failure and abnormal renal glomerulus morphology and decreased renal plasma flow rate.

ARHGAP26

The CAD risk allele, rs246600-T ($P=1.51x10^{-8}$; OR[95%CI]=1.04[1.03-1.06]) maps to an intron of Rho GTPase-Activating Protein 26 (*ARHGAP26*) a region that has been associated with triglycerides, type 2 diabetes and BMI. These variants however are not in LD with rs246600, the CAD associated SNP. However, this gene remains a very interesting candidate because the protein encoded by this gene is a GTPase activating protein that binds to focal adhesion kinase (FAK), a protein involved in the signaling cascades that regulate the organization of the actin-cytoskeleton, and mediates the activity of the GTP binding proteins RhoA and Cdc42[44], which represent proteins involved in the regulation and timing of cell division, morphology, migration and endocytosis. These processes may be relevant to the migration of fibroblasts and smooth muscle cells in the arterial vessel wall in response to the

deposition of vessel wall plaque as has been recently shown for the *TCF21* CAD susceptibility locus[45].

PARP12

We have shown that rs10237377-T confers protection from CAD (P=1.75x10^{-8}, OR[95%CI]=0.95[0.93-0.97]). This variant (or a tag r^2>0.8) has not been associated with another trait as GWS to date but it is an eQTL for Thromboxane A Synthase 1 (*TBXAS1*) in whole blood (P=3.09x10^{-71})[43]. TBXAS1, catalyzes the conversion of the prostaglandin endoperoxide into thromboxane A2, a potent vasoconstrictor and inducer of platelet aggregation[46]. *TBXAS1*, has been implicated in reduction of CAD complications in a recent trial[47]. The gene has also been associated with thromboxane synthetase deficiency a rare bleeding disorder (OMIM). In mouse, mutations in *TBXAS1* have resulted in increased bleeding and decreased platelet aggregation. Together these findings suggest that *TBXAS1* could be a candidate causal gene in the region through platelet aggregation mechanisms.

SERPINH1

rs590121-T maps to an intron of *SERPINH1* and we show is associated with increased risk of CAD (Table 1; OR=1.05[1.03-1.07], P=1.54x10^{-8}). This SNP is in LD with rs6704 (D'=1, r^2=0.86) which is an eQTL for *SERPINH1* in whole blood (P=3.3x10^{-22}). This gene encodes a member of the serpin superfamily of serine proteinase inhibitors. The encoded protein is found in smooth muscle cells, is localized to the endoplasmic reticulum and plays a role in collagen biosynthesis as a collagen-specific molecular chaperone. Autoantibodies to the encoded protein have been found in patients with rheumatoid arthritis.

The CAD associated SNP is also an eQTL for a neighbouring gene, *GDPD5*, in whole blood (P=8.69x10^{-10}) [43] and peripheral blood mononuclear cells (P=2.22x10^{-14})[38], however, the LD between the CAD associated SNP and the top eQTL is low r^2<0.1. GDPD5 protein is found in many tissues including liver and kidney.

C12orf43/HNF1A

The *C12orf43* region harbours three SNPs in the EUR studies and four in the all ancestry (five in total) that are associated with CAD at genome-wide significance (Supplementary Table 4). rs2258287, the sentinel SNP in EUR is located about 2Kb upstream of *C12orf43*. The A allele increases risk of CAD (OR[95% CI]=1.05[1.03-1.06], P=6x10^{-9}) and has previously been associated with increased LDL-C and total cholesterol levels (P=6.66x10^{-17})[2].

rs2258287 is also associated in the all ancestry analyses (P=2.18x10^{-8}) however rs2244608 has a modestly smaller P-value (P=1.57x10^{-8}). These SNPs are not in LD in African ancestries (r^2=0.12, D'=0.65, 1000G AFR), East Asians (r^2=0.14, D'=0.8, 1000G) or South Asians (r^2=0.38, D'=0.7 1000G SAS) and are in moderate LD in Europeans (r^2=0.68, D'=0.84 in 1000G EUR). These SNPs or strong proxies (r^2>0.8) were associated with decreased C-reactive protein (P=6.66x10^{-17}) [48,49], increased gamma glutamlytransferase levels (P=8.30x10^{-38}) [50] and activity (P=6.66x10^{-17})[51]. rs2244608 is intronic in the *HNF1A* gene. *HNF1A* encodes hepatocyte nuclear factor 1 homeobox A, a transcription factor highly expressed in the

digestive system and liver, which regulates many genes involved in a wide range of biological processes, including lipid and glucose transport and metabolism, and coagulation pathways. However, *HNF1A* is perhaps better known as a gene containing low-frequency variants causing maturity onset diabetes of the young (MODY3), a Mendelian form of diabetes caused by low-frequency dominant mutations. A tightly correlated missense variant in *HNF1A* (rs1169288, r^2=0.96, D'=0.99 with rs2244608 in 1000G EUR) is predicted to have functional effects on *HNF1A.*

SCARB1

The CAD and HDL associations in the *SCARB1* region are likely to be independent as neither of our CAD associated SNPs in *SCARB1* (rs11057830 and rs11057841) were in LD with the sentinel HDL-C associated SNP, rs838880, (r^2=0.02, D'=0.6 in 1000 Genomes CEU samples; Supplementary Fig. 6).

To further test the CAD and HDL associations, the summary statistics for major lipids[2] (joint analysis of metabochip and GWAS data http://csg.sph.umich.edu//abecasis/public/lipids2013/) made available by the Global Lipids Genetics Consortium were downloaded and used for the conditional analyses at the *SCARB1* region. The association of rs11057830 with CAD remained after conditioning on the HDL signal (P=1.30x10^{-8}: note, rs838880, a SNP in strong LD with the sentinel HDL SNP, r^2=0.83, D'=0.95 in the 1000G CEU samples, was used as rs838876 was not genotyped on the Metabochip). The association of rs838876 with HDL remained (P=1.15x10^{-35}, β=-0.049) after conditioning on the CAD associated SNP, rs11057830.

<u>The unconditional associations of the above mentioned SNPs with CAD, HDL, LDL and TG.</u>

	CAD SNP rs11057830 A/G	**Reported HDL SNP rs838876** G/A	**Metabochip Tag of HDL SNP rs838880** T/C	**Top TG SNP rs10846744** C/G
	β (*P*-value)			
CAD	0.0623 (1.34x10^{-8})	--	0.0153 (0.055)	0.0524 (5.857x10^{-7})
HDL	-0.0181 (0.0018)	-0.049 (7.33x10^{-33})	-0.048 (6.38x10^{-32})	-0.0145 (0.009)
LDL	0.0253 (2.58x10^{-5})	0.003 (0.44)	0.0006 (0.88)	0.0253 (4.654x10^{-5})
TG	0.0220 (8.34x10^{-5})	0.0052 (0.38)	0.0059 (0.31)	0.0236 (2.218x10^{-5})

Note the effect allele/non-effect alleles are listed after the SNP name.

In contrast, there is no evidence of association in the region after conditioning on the top CAD SNP rs11057830, which is also the top LDL SNP, in this region and is in high LD with the top TG SNP rs10846744 (r^2=0.94 in 1000 Genome phase 3 EUR samples). Given there is evidence of association with LDL-C and triglycerides at the CAD associated SNPs, this suggests that the *SCARB1* CAD association may be mediated via pro-atherogenic lipids.

DHX38

The *DHX38* region has previously been associated with increased total and LDL cholesterol[52]. Indeed, rs2000999-A, the cholesterol associated SNP, was associated with CAD in our data, but with less evidence (P=6.8x10^{-7}, OR[95% CI]=1.04[1.03-1.06]) than the SNPs that map to *DHX38* and was not convincingly associated with CAD after conditioning on rs1050362 (P>0.001). In addition to the cholesterol associations, the *DHX38* region has been reported to be associated with metabolites (tyrosine, phenylalanine/tyrosine ratio and glycoprotein)[53,54], ischemic stroke[55], atrial fibrillation[56] and Kawasaki disease[57], however the SNPs involved are not in LD with the CAD associated SNPs (r^2<0.15) suggesting these associations act through different causal pathways to the CAD association.

GOSR2

Within the *GOSR2* region, the CAD risk increasing allele rs17608766-C (OR[95% CI]=1.07[1.04-1.09]) has previously been reported to be associated with increased SBP[58] and increased pulse pressure.[59] It has also been associated with expression of *GOSR2* in liver[60,61] and reduced expression in brain, cerebellum and temporal cortex.[62] The association with CAD is likely to be through blood pressure and so the neighboring gene *WNT9B* also makes an interesting candidate. The kidney has an important role in blood pressure regulation. WNT9B protein shows highest expression in kidney (human protein atlas) and is implicated in kidney development. In mouse, mutation in the orthologous gene result in abnormal kidney development. Canonical Wnt9b signaling balances progenitor cell expansion and differentiation during kidney development.

PROCR

The CAD-associated SNP, rs867186 (or a SNP in strong LD r^2>0.8 in 1000G EUR) is associated with expression of *PROCR* across a range of tissues including, atherosclerotic aortic root[34], liver[60], skin and subcutaneaous adipose tissue[36] and transformed fibroblasts[37] (Supplementary Table 8, Supplementary Figure 8). While it is also in LD with the top eQTL for *EIF6*[63] and *ITGB4BP*[39] in monocytes, *PROCR* remains a plausible candidate gene for the CAD association.

The complexity underpinning the *PROCR* pathway is highlighted by its apparently paradoxical effects to reduce activity of the protein C pathway and increase risk of venous thrombosis, but *decrease* risk of CAD. Future studies will seek to elucidate this pathway, noting that previous studies have also highlighted a role of the EPCR in influencing vascular permeability and inflammation[64], which may be independent of its thrombotic effects.

References

1. CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121-30 (2015).
2. Global Lipids Genetics Consortium *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).
3. Nordestgaard, B.G., Benn, M., Schnohr, P. & Tybjaerg-Hansen, A. Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. *JAMA* **298**, 299-308 (2007).
4. Kamstrup, P.R., Tybjaerg-Hansen, A., Steffensen, R. & Nordestgaard, B.G. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA* **301**, 2331-9 (2009).
5. Varbo, A. *et al.* Remnant cholesterol as a causal risk factor for ischemic heart disease. *J Am Coll Cardiol* **61**, 427-36 (2013).
6. Riboli, E. *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* **5**, 1113-24 (2002).
7. Danesh, J. *et al.* EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. *Eur J Epidemiol* **22**, 129-41 (2007).
8. InterAct, C. *et al.* Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* **54**, 2272-82 (2011).
9. Saleheen, D. *et al.* The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other determinants of myocardial infarction in South Asia. *Eur J Epidemiol* **24**, 329-38 (2009).
10. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
11. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* **19**, 61-109 (1998).
12. Curb, J.D. *et al.* Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Ann Epidemiol* **13**, S122-8 (2003).
13. Matise, T.C. *et al.* The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. *Am J Epidemiol* **174**, 849-59 (2011).
14. Myocardial Infarction Genetics, C. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-41 (2009).
15. Chang, Y.C. *et al.* Common ALDH2 genetic variants predict development of hypertension in the SAPPHIRe prospective cohort: Gene-environmental interaction with alcohol consumption. *BMC Cardiovasc Disord* **12**, 58 (2012).
16. Liang, K.W. *et al.* Diabetes exacerbates angiographic coronary lesion progression in subjects with metabolic syndrome independent of CRP levels. *Clin Chim Acta* **388**, 41-5 (2008).
17. Lee, I.T., Huang, C.N., Lee, W.J., Lee, H.S. & Sheu, W.H. Aggravation of albuminuria by metabolic syndrome in type 2 diabetic Asian subjects. *Diabetes Res Clin Pract* **81**, 345-50 (2008).
18. Wang, T.D. *et al.* Association of epicardial adipose tissue with coronary atherosclerosis is region-specific and independent of conventional risk factors and intra-abdominal adiposity. *Atherosclerosis* **213**, 279-87 (2010).

19. Tsai, C.T. *et al.* Polygenetic regression model of renin-angiotensin system genes and the risk of coronary artery disease in a large angiographic population. *Clin Chim Acta* **412**, 619-24 (2011).
20. Prineas, R.J., Crow, R.S. & Blackburn, H.W. *The Minnesota code manual of electrocardiographic findings : standards and procedures for measurement and classification*, xiii, 328 p. (Springer, London, 2010).
21. Wu, I.C. *et al.* Association between dietary fiber intake and physical performance in older adults: a nationwide study in Taiwan. *PLoS One* **8**, e80209 (2013).
22. 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).
23. 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).
24. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
25. Speed, D., Hemani, G., Johnson, M.R. & Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **91**, 1011-21 (2012).
26. Goldstein, B.A., Yang, L., Salfati, E. & Assimes, T.L. Contemporary Considerations for Constructing a Genetic Risk Score: An Empirical Approach. *Genet Epidemiol* **39**, 439-45 (2015).
27. Ghosh, S. *et al.* Systems Genetics Analysis of Genome-Wide Association Study Reveals Novel Associations Between Key Biological Processes and Coronary Artery Disease. *Arterioscler Thromb Vasc Biol* **35**, 1712-22 (2015).
28. Oksala, N. *et al.* Association of neuroimmune guidance cue netrin-1 and its chemorepulsive receptor UNC5B with atherosclerotic plaque expression signatures and stability in human(s): Tampere Vascular Study (TVS). *Circ Cardiovasc Genet* **6**, 579-87 (2013).
29. Wanschel, A. *et al.* Neuroimmune guidance cue Semaphorin 3E is expressed in atherosclerotic plaques and regulates macrophage retention. *Arterioscler Thromb Vasc Biol* **33**, 886-93 (2013).
30. van Gils, J.M. *et al.* Endothelial expression of guidance cues in vessel wall homeostasis dysregulation under proatherosclerotic conditions. *Arterioscler Thromb Vasc Biol* **33**, 911-9 (2013).
31. Pfeufer, A. *et al.* Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat Genet* **41**, 407-14 (2009).
32. Sano, M. *et al.* Genome-wide association study of electrocardiographic parameters identifies a new association for PR interval and confirms previously reported associations. *Hum Mol Genet* **23**, 6668-76 (2014).
33. Arking, D.E. *et al.* Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet* **46**, 826-36 (2014).
34. Franzen, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827-30 (2016).
35. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* (2016).
36. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-9 (2012).
37. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
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. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
40. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* **42**, 36-44 (2010).
41. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* **43**, 1082-90 (2011).
42. Hancock, D.B. *et al.* Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet* **8**, e1003098 (2012).
43. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).
44. Doherty, G.J. *et al.* The endocytic protein GRAF1 is directed to cell-matrix adhesion sites and regulates cell spreading. *Mol Biol Cell* **22**, 4380-9 (2011).
45. Nurnberg, S.T. *et al.* Coronary Artery Disease Associated Transcription Factor TCF21 Regulates Smooth Muscle Precursor Cells That Contribute to the Fibrous Cap. *PLoS Genet* **11**, e1005155 (2015).
46. Chase, M.B., Baek, S.J., Purtell, D.C., Schwartz, S. & Shen, R.F. Mapping of the human thromboxane synthase gene (TBXAS1) to chromosome 7q34-q35 by two-color fluorescence in situ hybridization. *Genomics* **16**, 771-3 (1993).
47. Kopylov, F.Y. *et al.* [Positive effect of low-activity thromboxane A synthase gene on prognosis in coronary heart disease]. *Ter Arkh* **87**, 59-65 (2015).
48. Shah, T. *et al.* Gene-centric analysis identifies variants associated with interleukin-6 levels and shared pathways with other inflammation markers. *Circ Cardiovasc Genet* **6**, 163-70 (2013).
49. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* **41**, 35-46 (2009).
50. Chambers, J.C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* **43**, 1131-8 (2011).
51. Middelberg, R.P. *et al.* Loci affecting gamma-glutamyl transferase in adults and adolescents show age x SNP interaction and cardiometabolic disease associations. *Hum Mol Genet* **21**, 446-55 (2012).
52. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-13 (2010).
53. Inouye, M. *et al.* Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet* **8**, e1002907 (2012).
54. Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet* **44**, 269-76 (2012).
55. Traylor, M. *et al.* Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol* **11**, 951-62 (2012).
56. Ellinor, P.T. *et al.* Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet* **44**, 670-5 (2012).
57. Burgner, D. *et al.* A genome-wide association study identifies novel and functionally related susceptibility Loci for Kawasaki disease. *PLoS Genet* **5**, e1000319 (2009).
58. 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).
59. Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
60. 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).

61. Schadt, E.E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* **6**, e107 (2008).
62. Zou, F. *et al.* Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet* **8**, e1002707 (2012).
63. Fairfax, B.P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
64. Danese, S., Vetrano, S., Zhang, L., Poplis, V.A. & Castellino, F.J. The protein C pathway in tissue inflammation and injury: pathogenic role and therapeutic implications. *Blood* **115**, 1121-30 (2010).

Acknowledgements

The TAICHI study was supported by a grant from the NHLBI grant HL087647. The Taiwan Dragon (TAICHI) was supported by grants from National Science Council (NSC 98-2314-B-075A-002-MY3) and Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1013001C;TCVGH-1013002D). The TCAD (TAICHI) was supported by grants from National Science Council (NSC 98-2314-B-075A-002-MY3) and Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1013001C;TCVGH-1013002D). The SAPPHIRe (TAICHI) was supported by grants from National Health Research Institutes (BS-094-PP-01 & PH-100-PP-03). The HALST (TAICHI) was supported by grants from National Health Research Institutes (PH-100-SP-01)

The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI), and U01HG004801 (Coordinating Center), and their respective NHGRI ARRA supplements. The complete list of PAGE members can be found at http://www.pagestudy.org.

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. This manuscript was prepared in collaboration with investigators of the WHI, and has been approved by the WHI. WHI investigators are listed at http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf.

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

JE funding is provided by the German Federal Ministry of Education and Research (BMBF) in the context of the e:Med program (e:AtheroSysMed and sysINFLAME) , the FP7 European Union project CVgenes@target (261123) and a grant from the Fondation Leducq (CADgenomics: Understanding Coronary Artery Disease Genes, 12CVD02. This study was also supported through the Deutsche Forschungsgemeinschaft (DFG) cluster of excellence 'Inflammation at Interfaces'.

We are thankful to all the study participants in Pakistan. Recruitment in PROMIS was funded through grants available to investigators at the Center for Non-Communicable Diseases, Pakistan (Danish Saleheen and Philippe Frossard) and investigators at the University of Cambridge, UK (Danish Saleheen and John Danesh). Field-work, genotyping, and standard clinical chemistry assays in PROMIS were principally supported by grants awarded to the University of Cambridge from the British Heart Foundation, UK Medical Research Council,

Wellcome Trust, EU Framework 6-funded Bloodomics Integrated Project, Pfizer. We would like to acknowledge the contributions made by the following individuals who were involved in the field work and other administrative aspects of the study: Mohammad Zeeshan Ozair, Usman Ahmed, Abdul Hakeem, Hamza Khalid, Kamran Shahid, Fahad Shuja, Ali Kazmi, Mustafa Qadir Hameed, Naeem Khan, Sadiq Khan, Ayaz Ali, Madad Ali, Saeed Ahmed, Muhammad Waqar Khan, Muhammad Razaq Khan, Abdul Ghafoor, Mir Alam, Riazuddin, Muhammad Irshad Javed, Abdul Ghaffar, Tanveer Baig Mirza, Muhammad Shahid, Jabir Furqan, Muhammad Iqbal Abbasi, Tanveer Abbas, Rana Zulfiqar, Muhammad Wajid, Irfan Ali, Muhammad Ikhlaq, Danish Sheikh and Muhammad Imran. The genetic and biomarker assays in this study (consumables and labour) were supported by grants from the British Heart Foundation (PG/09/035/27378), US National Institutes of Health (RC2HL101834 and RC1TW008485), Fogarty International Center (RC1TW008485), the Wellcome Trust (084711/Z/08/Z), and Pfizer, as well as by underpinning support from the UK Medical Research Council (MR/L003120/1), British Heart Foundation (RG/13/13/30194), and National Institute for Health Research Cambridge Biomedical Research Centre

We thank participants and staff of the Copenhagen City Heart Study, Copenhagen Ischemic Heart Disease Study, and the Copenhagen General Population Study for their important contributions.

CAD case ascertainment and validation, genotyping, and clinical chemistry assays in EPIC-CVD were principally supported by grants awarded to the University of Cambridge from the EU Framework Programme 7 (HEALTH-F2-2012-279233), the UK Medical Research Council (G0800270) and British Heart Foundation (SP/09/002), and the European Research Council (268834). We thank all EPIC participants and staff for their contribution to the study, the laboratory teams at the Medical Research Council Epidemiology Unit for sample management and Cambridge Genomic Services for genotyping, Sarah Spackman for data management, and the team at the EPIC-CVD Coordinating Centre for study coordination and administration.

The BRAVE study genetic epidemiology working group is a collaboration between the Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, UK, the Centre for Control of Chronic Diseases, icddr,b, Dhaka, Bangladesh and the National Institute of Cardiovascular Diseases, Dhaka, Bangladesh.

EPIC-CVD consortium study PIs

Kim Overvad[1,2], Anne Tjønneland[3], Francoise Clavel-Chapelon[4], Rudolf Kaaks[5], Heiner Boeing[6], Antonia Trichopoulou[7,8], Pietro Ferrari[9], Domenico Palli[10], Vittorio Krogh[11], Salvatore Panico[12], Rosario Tumino[13], Giuseppe Matullo[14,15], Jolanda Boer[16], Yvonne van. der Schouw[17], Elisabete Weiderpass[18,19,20,21], J. Ramon Quiros[22], María-José Sánchez[23,24], Carmen Navarro[25], Conchi Moreno-Iribas[26], Larraitz Arriola[27], Olle Melander[28], Patrik Wennberg[29], Nicholas J. Wareham[30], Timothy J. Key[31], Elio Riboli[32], Adam S. Butterworth[33,34], John Danesh[33,34,35]

1. Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark
2. Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark
3. Diet, Genes and Environment, Danish Cancer Society Research Center, Copenhagen, Denmark
4. INSERM, Centre for Research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones, and Women's Health Team, Institut Gustave Roussy, Villejuif, France
5. Division of Cancer Genetic Epidemiology, German Cancer Research Centre (DKFZ), im Neuenheimer Feld 581, 69120 Heidelberg, Germany
6. Department of Epidemiology, German Institute of Human Nutrition (DIfE), Potsdam-Rehbrücke, Germany
7. WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece
8. Hellenic Health Foundation, Athens, Greece
9. International Agency for Research on Cancer, Lyon, France
10. Molecular and Nutritional Epidemiology Unit, Centro per lo Studio e la Prevenzione Oncologica-Scientific Institute of Tuscany, Florence, Italy
11. Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
12. Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy
13. Cancer Registry and Histopathology Unit, Civic- M.P.Arezzo Hospital, ASP Ragusa, Italy
14. Human Genetics Foundation, Turin, Italy
15. Department of Medical Sciences, University of Turin, Italy
16. Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
17. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands
18. Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway
19. Department of Research, Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway
20. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
21. Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland
22. Public Health Directorate, Asturias, Spain
23. Escuela Andaluza de Salud Pública, Instituto de Investigación Biosanitaria

ibs.GRANADA. Hospitales Universitarios de Granada/Universidad de Granada, Granada, Spain
24. CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
25. Epidemiology Department, Murcia Health Authority, Murcia, Spain
26. Public Health Institute of Navarra, Pamplona, Spain
27. Public Health Division of Gipuzkoa, Instituto Bio-Donostia, Basque Government, CIBERESP, Spain
28. Department of Clinical Sciences, Hypertension & Cardiovascular Disease, Clinical Research Centre, Malmö University Hospital, Malmö, Sweden
29. Department of Public Health and Clinical Medicine, Family Medicine, Umeå University, Umeå, Sweden
30. Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK
31. Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, UK
32. School of Public Health, Imperial College London, UK
33. Cardiovascular Epidemiology Unit, Department of Public Health & Primary Care, University of Cambridge, UK
34. The National Institute for Health Research Blood and Transplant Unit (NIHR BTRU) in Donor Health and Genomics at the University of Cambridge, UK
35. Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK

Supplementary Tables

Supplementary Table 1 (a) Study-specific sample quality control exclusions and baseline characteristics of the studies with *de novo* genotyping. **(b)** Study-specific definitions of disease outcome (CAD).

Collection	Recruitment Country	Study design	Disease Outcome	N samples failed QC	N related samples removed	N ancestry outliers removed	N cases (% male)	N controls (% male)	Mean (SD) age
European									
EPIC-CVD	UK, Germany, Netherlands, Sweden, Norway, France, Spain, Greece, Italy	Case-cohort	CAD	231	475	616	11,391 (60) **	7,251 (35) **	59.3 (8.9) [56.5 (10.1)]*
CCHS	Denmark (Copenhagen)	Prospective	CAD	122	212	13	1,999 (52)	6,562 (42)	66.1 (10.8) [57.9 (15.1)]*
CIHDS	Denmark (Copenhagen)	Case series	ACS	106	163	8	2,703 (73)	NA	60.4 (11.8)*
CGPS	Denmark (Copenhagen)	Prospective	N/A	120	58	6	NA	2,803 (44)	58.0 (12.6)
South Asian									
PROMIS	Pakistan (8 centres)	Case/Control	AMI	385	264	2	5,833 (84) **	5,369 (81) **	53.2 (12.3)

BRAVE	Bangladesh (Dhaka)	Case/Control	AMI	63	111	9	1,821 (88)	1,645 (89)	48.5 (15.9)
African American									
ARIC (females)	USA	Prospective	CAD	NA	NA	NA	192 (0)	1840	53.3 (5.7)
ARIC (males)	USA	Prospective	CAD	NA	NA	NA	174 (100)	998	53.6 (6.0)
WHI	USA	Prospective	CAD	NA	0	2	99 (0)	1855	60.8 (6.8)
MIGen	USA	Case/Control	CAD	85	60	0	1,635 (NA)	1,053 (NA)	NA
East Asian									
TAICHI	Taiwan	Case/Control	CAD	797	1,312	4	4,129 (78)	6,369 (53)	66.6 (11.3) [64.4 (12.5)][*]

[*]Mean age at diagnosis of CAD rather than baseline age at recruitment is given. [**] These are the numbers of samples genotyped and passing QC. A subset of the PROMIS samples (3,704 CAD cases and 3,433 controls) and the EPIC-CVD samples (1,830 CAD cases and 449 controls) had been included in the CARDIoGRAMplusC4D discovery effort and were therefore not included in the meta-analyses with CARDIoGRAMplusC4D. Note, 21 samples were dropped from CIHDS and 12 from CGPS as they were identical to samples found in CCHS.

Supplementary Table 1(b)

Study	Case definition
MIGen	CAD defined as acute myocardial infarction, >50% stenosis in coronary artery on coronary angiography, abnormal stress test, or unstable angina diagnosis
ARIC	CAD, defined as acute hospitalized MI (definitive or probable), definite fatal CAD, or ECG diagnosis of MI, validated by review of hospital records, death certificates and interviews of next of kin
BRAVE	Acute MI within the preceding 24 hours
CCHS	CAD, defined as ICD10 I20-I25, from morbidity and mortality registries
CGPS	CAD, defined as ICD10 I20-I25, from morbidity and mortality registries
CIHDS	MI or other major acute coronary syndromes plus stenosis or atherosclerosis on coronary angiography and/or positive results on exercise electrocardiography.
EPIC-CVD	CAD, defined as ICD10 I20-I25, ascertained and validated through various methods (morbidity registers, general practice records, MONICA registries, self- report, clinical records)
PROMIS	Acute MI within the preceding 24 hours
TAICHI	CAD, defined as either currently or in the past suffering from an MI, an ACS, angina, or demonstrated at least one epicardial coronary artery obstruction of >50% on coronary angiogram.
WHI	CAD, defined as acute hospitalized MI (definitive or probable), definite fatal CAD, or ECG diagnosis of MI, validated by review of hospital records, death certificates and interviews of next of kin

ACS = acute coronary syndrome; CAD = coronary artery disease; ECG = electrocardiogram; MI = myocardial infarction. Similar CAD definitions were used by the CARDIoGRAMplusC4D: Supplementary Table 2 of reference 3.

Supplementary Table 2 Summary of study specific SNP genotype quality control for studies with *de novo* genotyping. The CardioMetabochip+ genotypes 209,818 SNPs, of which 209,529 map to the autosomes, while the CardioMetabochip includes 196,725 SNPs of which 196,479 map to the autosomes.

Collection	**Genotyping array**	**HWE *P*-value threshold**	**SNP call rate threshold**	**#SNPs with no calls**	**# monomorphic snps**	**Number of SNPs removed call rate/HWE/MAF<0.01**	**Number of SNPs passing QC**
European							
EPIC-CVD	CardioMetabo+	$1x10^{-6}$	0.97	1,403	25,192	48,322	134,612
CCHS	CardioMetabo+	$1x10^{-6}$	0.97	1,374	37,152	35,093	135,910
CIHDS/CGPS	CardioMetabo+	$1x10^{-6}$	0.97	1,387	42,708	29,307	136,127
South Asian							
PROMIS	CardioMetabo+	$1x10^{-6}$	0.97	1,149	21,302	55,491	131,587
BRAVE	CardioMetabo+	$1x10^{-6}$	0.97	1,019	52,407	25,485	130,618

African American							
ARIC (males)	CardioMetabo	$1x10^{-6}$	0.95	NA	NA	NA	143,615
ARIC (females)	CardioMetabo	$1x10^{-6}$	0.95	NA	NA	NA	143,473
WHI	CardioMetabo	$1x10^{-6}$	0.95	NA	NA	NA	145,132
MIGen	CardioMetabo	$1x10^{-6}$	0.97	535	10,940	37,019	148,231
East Asian							
TAICHI	CardioMetabo	$1x10^{-6}$	0.97	5,787 (<95%)	46,543	39,092	105,834

Supplementary Table 3 Inflation factors for studies with *de novo* genotyping. Lambda represents the inflation of the test statistics across all variants that passed QC in a study. Given that the CardioMetabochip was a customised genotyping array that included fine-mapping of previously established CAD loci and not a random selection of SNPs from across the genome, we also calculated inflation factors having excluded known CAD regions. The Lambda (noCAD) have had the variants that map to one of the 47 previously published CAD regions (or within 1Mb) of that region. QQ plots of the association statistics are provided in Supplementary Figure 1.

Collection	Association model	Lambda	Lambda (noCAD)	# PCs
European				
EPIC-CVD	LMM	1.03	0.99	5
CCHS	LMM	1.00	0.98	0
CIHDS/CGPS	LMM	1.05	0.99	1
South Asian	LMM	1.10	1.03	3
PROMIS	Logistic	1.07	1.03	1
BRAVE	Logistic	1.06	1.04	1
African American				
WHI	Logistic	0.97	0.97	10
MIGen	Logistic	1.06	1.05	10
ARIC (males)	Logistic	1.03	1.01	10
ARIC (females)	Logistic	1.04	1.04	10
East Asian				
TAICHI	LMM	1.09	1.05	5

LMM = linear mixed model as implemented in GEMMA. Logistic = logistic regression model

Supplementary Table 4 Results of CAD association tests from the European and All ancestry meta-analyses for SNPs with $P<5x10^{-8}$ at the new loci.

Closest gene(s)	SNP	Chr:Position	Effect allele (AF)	European collections			All collections			
				OR [95% CI]	*P*	P_{het}	OR [95% CI]	*P*	P_{het}	$log_{10}BF$
ATP1B1	rs1892094C>T	1:169094459	T (0.50;0.48)	**0.96 [0.94-0.97]**	**$3.99x10^{-8}$**	**0.86**	0.96 [0.94-0.97]	$2.25x10^{-8}$	0.83	6.33
	rs10919065G>T	1:169093557	T (0.43;0.43)	**1.05 [1.03-1.06]**	**$1.57x10^{-8}$**	**0.72**	1.04 [1.02-1.05]	$9.28x10^{-7}$	0.11	5.06
	rs1200159C>T	1:169100241	T (0.43;0.42)	**1.05 [1.03-1.06]**	**$3.40x10^{-8}$**	**0.72**	1.04 [1.02-1.05]	$1.90x10^{-6}$	0.20	4.68
DDX59/CAMSAP 2	rs6700559C>T	1:200646073	T (0.47;0.47)	**0.96 [0.94-0.97]**	**$2.50x10^{-8}$**	**0.14**	0.96 [0.95-0.97]	$1.13x10^{-8}$	0.51	6.68
LMOD1	rs2820315C>T	1:201872264	T (0.30;0.29)	**1.05 [1.03-1.07]**	**$4.14x10^{-9}$**	**0.01**	1.05 [1.03-1.07]	$7.70x10^{-10}$	0.02	7.72
	rs2819348T>C	1:201884952	C (0.34;0.33)	**1.05 [1.03-1.06]**	**$2.83x10^{-8}$**	**0.02**	1.05 [1.03-1.06]	$1.77x10^{-8}$	0.01	6.42
(nsSNP) *TNS1*	rs2571445G>A	2:218683154	A (0.39;0.39)	1.04 [1.02-1.06]	$3.58x10^{-6}$	0.86	**1.05 [1.03-1.06]**	**$4.55x10^{-10}$**	**0.01**	**8.41**
ARHGAP26	rs246600C>T	5:142516897	T (0.48;0.46)	**1.05 [1.03-1.06]**	**$1.29x10^{-8}$**	**0.41**	1.04 [1.03-1.06]	$1.51x10^{-8}$	0.36	6.39
PARP12	rs10237377G>T	7:139757136	T (0.35;0.38)	0.95 [0.93-0.97]	$1.70x10^{-7}$	0.13	**0.95 [0.93-0.97]**	**$1.74x10^{-8}$**	**0.34**	**6.32**
PCNX3	rs12801636G>A	11:65391317	A (0.23;0.25)	0.95 [0.93-0.97]	$1.00x10^{-7}$	0.22	**0.95 [0.94-0.97]**	**$9.72x10^{-9}$**	**0.48**	**6.64**
SERPINH1	rs590121G>T	11:75274150	T (0.30;0.31)	**1.05 [1.03-1.07]**	**$1.54x10^{-8}$**	**0.47**	1.04 [1.03-1.06]	$9.32x10^{-8}$	0.05	5.80
C12orf43/HNF1A	**rs2258287C>A**	**12:121454313**	**A (0.34;0.37)**	**1.05 [1.03-1.06]**	**$6.00x10^{-9}$**	**0.10**	1.04 [1.03-1.06]	$2.18x10^{-8}$	0.13	6.40
	rs2708081C>T	12:121463288	T (0.48;0.47)	**0.96 [0.94-0.97]**	**$1.02x10^{-8}$**	**0.32**	0.96 [0.95-0.98]	$1.56x10^{-7}$	0.28	4.99
	rs3213545G>A	12:121471337	A (0.31;0.32)	**1.04 [1.03-1.07]**	**$2.50x10^{-8}$**	**0.22**	1.04 [1.03-1.06]	$4.81x10^{-8}$	0.43	6.13

	rs2244608A>G	**12:121416988**	**G (0.34;0.34)**	1.04 [1.03-1.06]	$1.96x10^{-7}$	0.30	**1.04 [1.03-1.06]**	**$1.57x10^{-8}$**	**0.71**	**6.42**
	rs1169288A>C	12:121416650	C (0.34;0.34)	1.05 [1.03-1.06]	$3.44x10^{-7}$	0.20	**1.05 [1.03-1.06]**	**$4.53x10^{-8}$**	**0.68**	**5.94**
SCARB1	rs11057830G>A	12:125307053	A (0.16;0.15)	**1.07 [1.05-1.10]**	**$5.65x10^{-9}$**	**0.56**	1.06 [1.04-1.09]	$1.34x10^{-8}$	0.78	6.49
	rs11057841C>T	12:125316743	T (0.15;0.16)	**1.07 [1.04-1.09]**	**$1.19x10^{-8}$**	**0.60**	1.05 [1.03-1.08]	$7.52x10^{-7}$	0.23	4.92
OAZ2/*RBPMS2*	rs6494488A>G	15:65024204	G (0.18;0.21)	0.95 [0.93-0.97]	$1.43x10^{-6}$	0.44	**0.95 [0.93-0.97]**	**$2.09x10^{-8}$**	**0.50**	**6.41**
DHX38	rs1050362C>A	16:72130815	A (0.38;0.39)	1.04 [1.03-1.06]	$2.32x10^{-7}$	0.59	**1.04 [1.03-1.06]**	**$3.52x10^{-8}$**	**0.60**	**6.16**
	rs2072142C>T	16:72132713	T (0.37;0.38)	1.05 [1.03-1.06]	$2.44x10^{-7}$	0.66	**1.05 [1.03-1.06]**	**$4.26x10^{-8}$**	**0.65**	**5.75**
GOSR2	rs17608766T>C	17:45013271	C (0.14;0.14)	**1.07 [1.04-1.09]**	**$4.14x10^{-8}$**	**0.99**	1.06 [1.04-1.09]	$2.10x10^{-7}$	0.74	5.30
PECAM1	rs1867624T>C	17:62387091	C (0.39;0.38)	0.96 [0.94-0.97]	$1.14x10^{-7}$	0.70	**0.96 [0.95-0.97]**	**$3.98x10^{-8}$**	**0.36**	**6.03**
	rs9892152C>T	17:62401965	T (0.47;0.46)	0.96 [0.95-0.98]	$2.73x10^{-7}$	0.41	**0.96 [0.95-0.98]**	**$5.00x10^{-8}$**	**0.75**	**5.92**
(nsSNP) *PROCR*	rs867186A>G	20:33764554	G (0.11;0.11)	**0.93 [0.91-0.96]**	**$1.26x10^{-8}$**	**0.61**	0.93 [0.91-0.96]	$2.70x10^{-9}$	0.74	7.11

Chr:Position = chromosome:position (build 37). AF= allele frequency in Europeans; allele frequency averaged across All ancestries. OR [95% CI] = odds ratio [95% confidence interval]. *P* = CAD association *P*-value. P_{het} is the *P*-value for heterogeneity from the meta-analysis. $\log_{10}$BF is the log base 10 of the Bayes factors obtained from the MANTRA analyses ($\log_{10}$BF≥6 is considered significant)

Supplementary Table 6 Summary of *P*-values for the null and associated sets used in the modified MAGENTA pathway analyses with the hypergeometric tests.

	Null set						**Associated set**					
		P-value distribution						*P*-value distribution				
Meta-analysis	*N*	*Min*	Q_1	*Median*	Q_3	*Max*	*N*	*Min*	Q_1	*Median*	Q_3	*Max*
EUR	1,000	0.6530	0.7370	0.8195	0.9041	0.9998	53	5.82×10^{-14}	1.585×10^{-6}	1.958×10^{-5}	4.296×10^{-5}	9.890×10^{-5}
EUR+SAS+AA+EAS+CG	1,000	0.6871	0.7696	0.8388	0.9153	0.9999	85	8.40×10^{-97}	4.86×10^{-11}	9.72×10^{-9}	8.04×10^{-8}	1.00×10^{-6}

N: number of variants contained in the set. *Min*: minimum. Q_1: first quartile. Q_3: third quartile. *Max* = maximum.

Supplementary Table 7 Pathways with enrichment of CAD associated variants from the all ancestry meta-analyses identified using modified MAGENTA.

				Europeans				All ancestry			
Category	**Database**	**Pathway**	**Genes**	*k*	*n*	P_{obs}	P_{enr}	*k*	*n*	P_{obs}	P_{enr}
Lipids / lipoproteins	Reactome	Lipoprotein metabolism	27	3	3	0.00034	0.0004	4	4	0.00004	0.0000
	KEGG	Glycerolipid metabolism	49	4	4	0.00002	0.0000	3	3	0.00047	0.0002
	Reactome	Metabolism of lipids and lipoproteins	228	6	11	0.00004	0.0001	7	15	0.00005	0.0003
	Reactome	Chylomicron mediated lipid transport	17	1	1	0.07063	1.0000	3	3	0.00047	0.0003
	Ingenuity	FXR/RXR activation	57	3	6	0.00581	0.0047	4	6	0.00047	0.0006
	Panther BP	Lipid and fatty acid transport	111	3	11	0.03701	0.0378	5	11	0.00083	0.0016
	Ingenuity	LXR/RXR activation	40	1	2	0.13634	1.0000	3	4	0.00176	0.0020
	Panther MF	Apolipoprotein	23	1	2	0.13634	1.0000	2	2	0.00607	0.0050
	Panther MF	Lipase	19	1	1	0.07063	1.0000	2	2	0.00607	0.0069
	KEGG	Glycerophospholipid metabolism	77	3	5	0.00306	0.0029	1	5	0.33548	1.0000
	Reactome	HDL mediated lipid transport	11	2	2	0.00493	0.0041	1	1	0.07834	1.0000
Immune system / thrombosis	Reactome	Signaling by platelet derived growth factor (PDGF)	64	4	10	0.00349	0.0026	4	7	0.00103	0.0009
	BioCarta	Platelet amyloid precursor protein (APP)	14	3	5	0.00306	0.0034	3	4	0.00176	0.0019

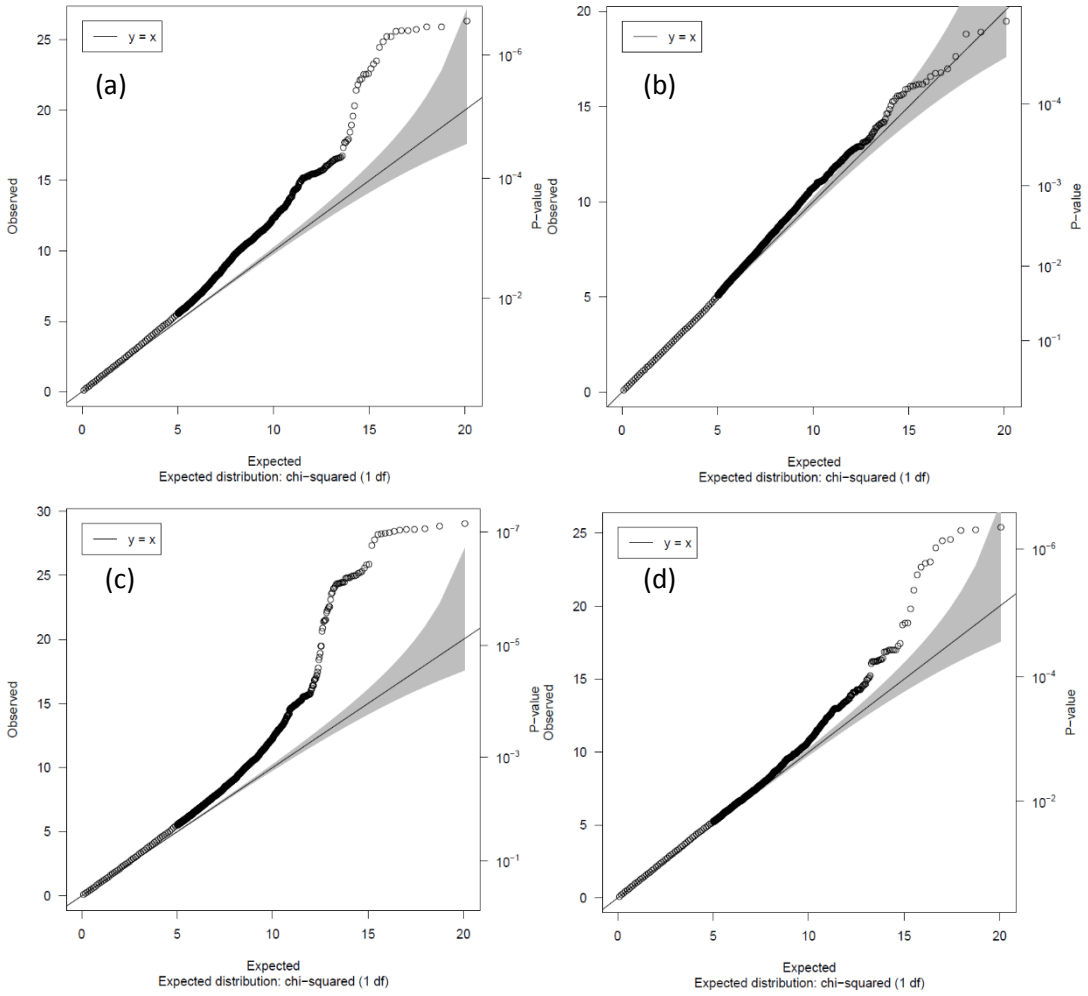
	BioCarta	Intrinsic prothrombin activation	23	2	2	0.00493	0.0049	2	3	0.01728	0.0167
	Reactome	Formation of platelet plug	185	5	16	0.00364	0.0045	3	15	0.10636	0.1037
	Reactome	Platelet activation	166	5	16	0.00364	0.0033	3	15	0.10636	0.1048
	Reactome	G alpha Q signalling events	155	4	12	0.00737	0.0072	2	8	0.12471	0.1172
Heart / cardiac function	BioCarta	Acute myocardial infarction (AMI)	20	3	4	0.00129	0.0004	3	4	0.00176	0.0011
	BioCarta	Angiotensin-converting enzyme (ACE) 2	13	2	2	0.00493	0.0041	2	2	0.00607	0.0073
Blood	Panther	Endothelin signaling pathway	19	3	4	0.00129	0.0013	3	4	0.00176	0.0019
	Reactome	Hemostasis	272	7	21	0.00035	0.0006	5	21	0.01959	0.0206
Vitamin C	BioCarta	Vitamin C in the brain	11	2	2	0.00493	0.0047	2	2	0.00607	0.0052
Phosphatase	Panther MF	Phosphatase modulator	19	1	1	0.07063	1.0000	2	2	0.00607	0.0058
DNA/RNA modification	Reactome	Elongation and processing of capped transcripts	133	1	1	0.07063	1.0000	2	2	0.00607	0.0069
Cell structure / interactions	Panther MF	Cation transporter	112	3	11	0.03701	0.0345	4	11	0.00757	0.0076
	Reactome	mRNA splicing	107	1	1	0.07063	1.0000	2	2	0.00607	0.0077
	Reactome	Integrin cell surface interactions	81	4	6	0.00031	0.0006	3	8	0.01948	0.0169
	Reactome	Cell surface interactions at the vascular wall	93	4	8	0.00130	0.0006	3	8	0.01948	0.0227
	KEGG	ECM receptor interaction	84	3	6	0.00581	0.0057	2	7	0.09840	0.0959

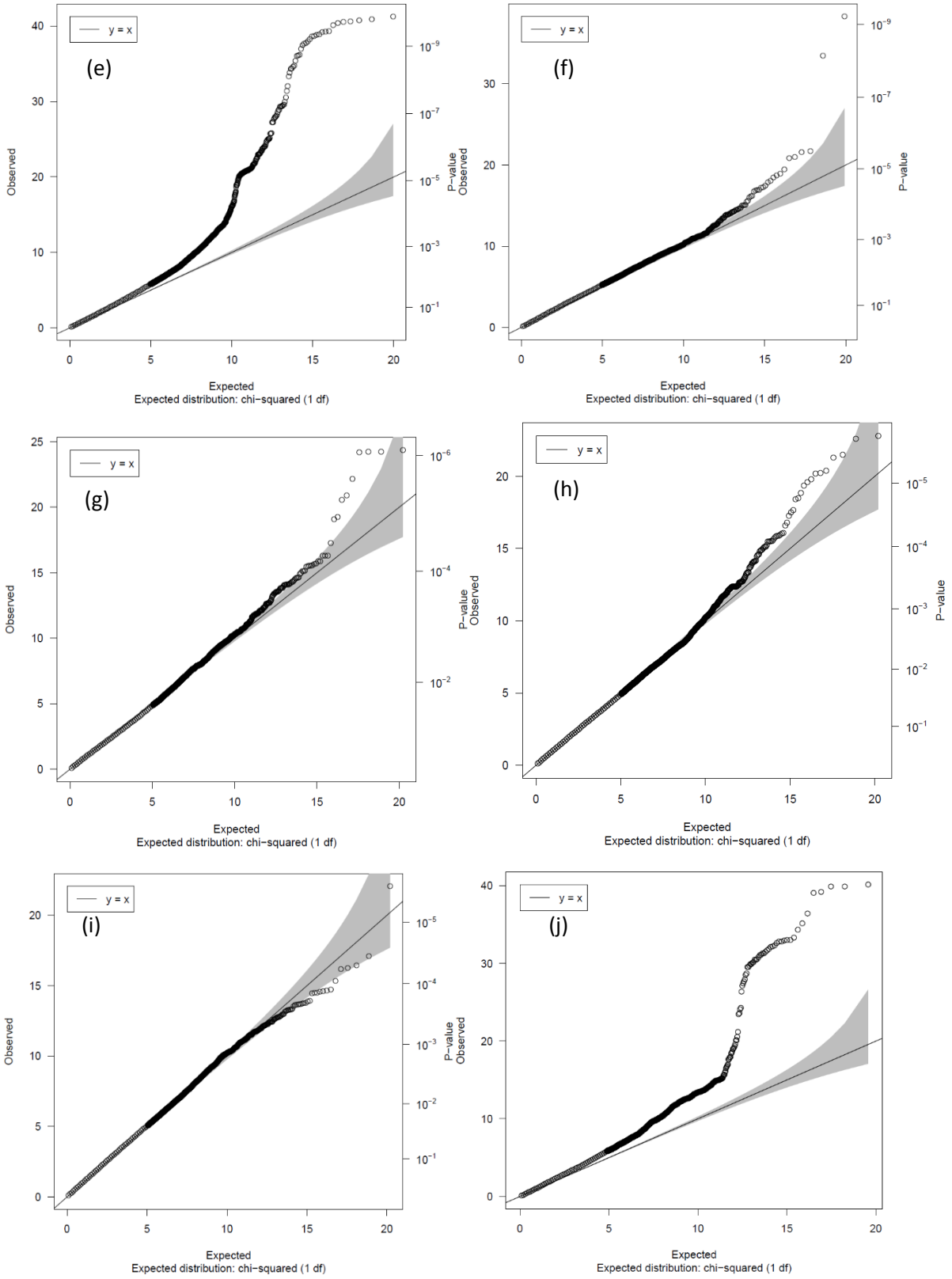
SNARE protein	Panther MF	SNARE protein	36	2	2	0.00493	0.0041	2	3	0.01728	0.0168
	KEGG	SNARE interactions in vesicular transport	37	2	2	0.00493	0.0040	2	3	0.01728	0.0173
Liver	Ingenuity	Hepatic fibrosis / hepatic stellate cell activation	83	4	11	0.00519	0.0068	3	10	0.03723	0.0397

For each hypergeometric test the empirical *P*-values displayed were calculated based on comparing the observed *P*-value to those obtained from 10,000 random sets. Seventy six SNPs formed the associated set for the European ancestry analysis and 85 for the All ancestry analysis. Genes: the number of genes that are listed for that pathway in the database. n: number of analyses for which this pathway showed evidence of enrichment at $P < 0.01$. *k:*number of variants in the associated set that were mapped to a gene listed in the pathway. [§] $P < 0.0001$, occurs when no random set hypergeometric test *P*-values are less than or equal to the observed *P*-value.

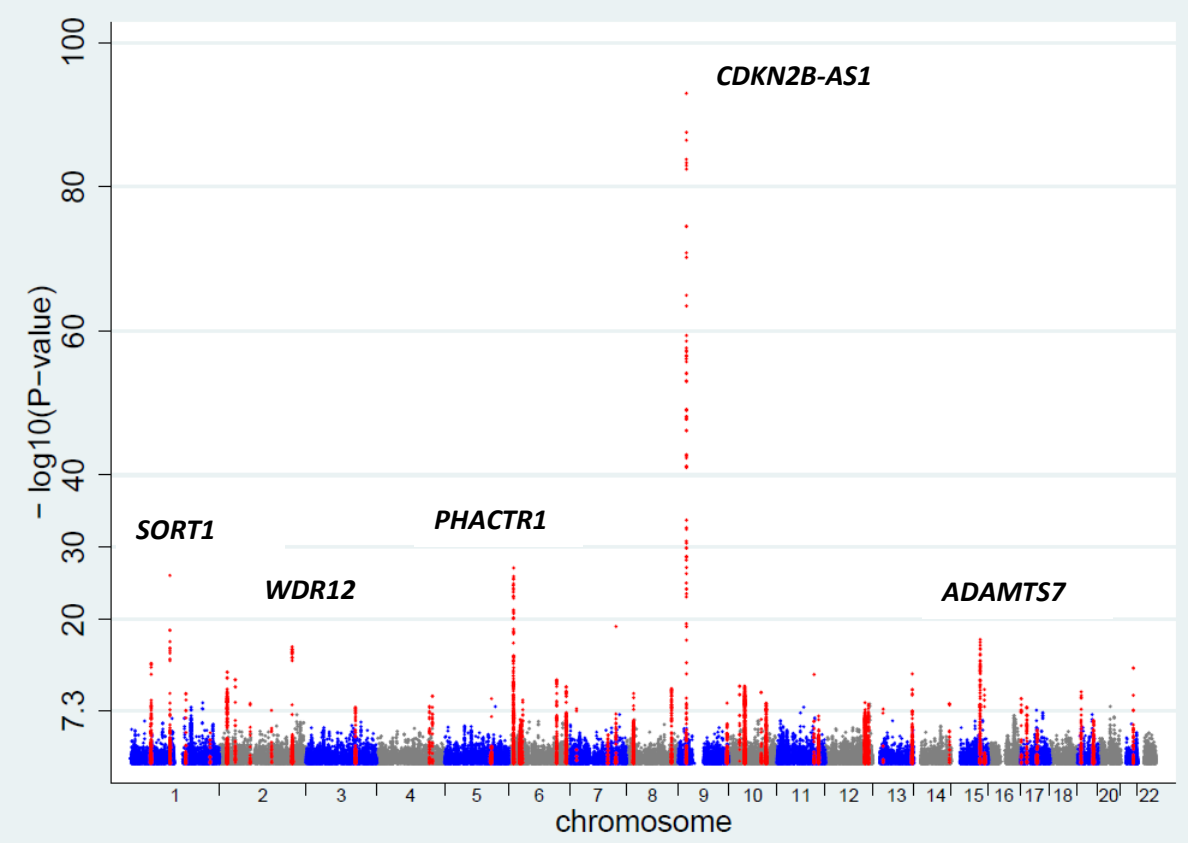
Supplementary Figures

Supplementary Figure 1: QQ plots illustrating array-wide inflation for each of the studies with *de novo* genotyping. **(a)** CIHDS/CGPS studies analysed using a mixed effects model at 136,127 SNPs **(b)** CCHS study analysed at 135,910 SNPs **(c)** EPIC-CVD analysed using a mixed model at 134,533 SNPs **(d)** EPIC-CVD-Umea analysed using a mixed model at 133,849 SNPs **(e)** South Asian studies PROMIS and BRAVE combined in a mixed model analysis of 127,114 SNPs **(f)** MIGen analysed using a logistic regression model at 123,885 SNPs (note the two SNPs with $P < 1x10^{-8}$, only passed QC in MIGen and consequently are likely to be genotype clustering artefacts and were excluded from all meta analyses) **(g)** WHI analysed at 145,132 SNPs **(h)** ARIC males analysed using a logistic regression model at 143,615 SNPs **(i)** ARIC females analysed using logistic regression at 143,473 SNPs (**j**) TAICHI using linear a mixed model at 103,238 SNPs. Inflation factors are reported in Supplementary Table 3.

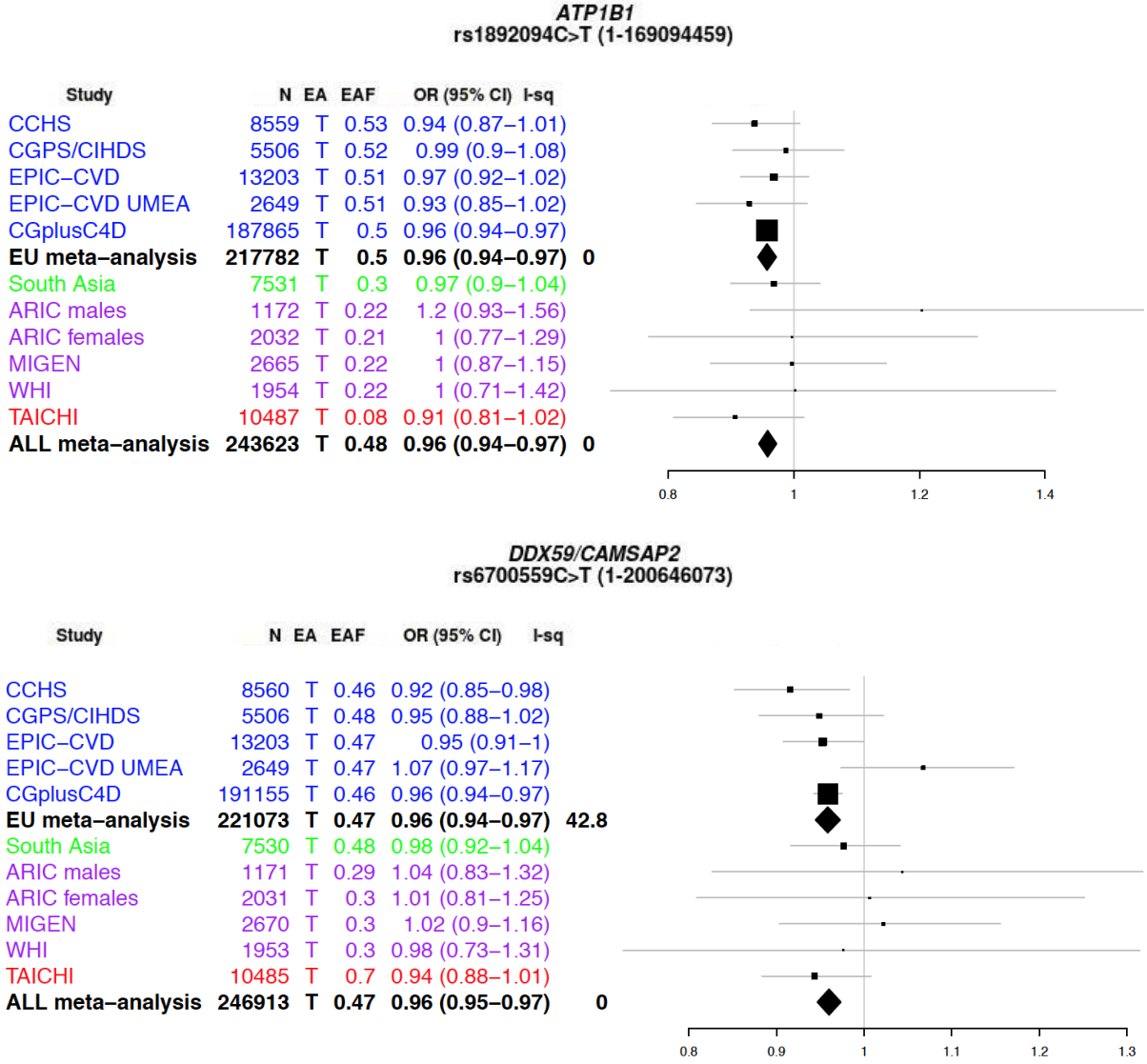



(e)
y = x
Observed
P−value
Observed
Expected
Expected distribution: chi−squared (1 df)
(f)
y = x
P−value
Expected
Expected distribution: chi−squared (1 df)
(g)
y = x
Observed
P−value
Observed
Expected
Expected distribution: chi−squared (1 df)
(h)
y = x
P−value
Expected
Expected distribution: chi−squared (1 df)
(i)
y = x
Observed
P−value
Observed
Expected
Expected distribution: chi−squared (1 df)
(j)
y = x
P−value
Observed
Expected
Expected distribution: chi−squared (1 df)

Supplementary Figure 2: Manhattan plot showing the association of ~79,000 variants with CAD from the European meta-analysis in up to ~221,000 individuals. Red dots represent SNPs that map to LD blocks that include the previously published (known) CAD regions. The SNP with the most evidence of association in this meta-analysis was rs133045 in the 9p21 region (P=1x10^{-93}). -log(P=5x10^{-8})~7.3

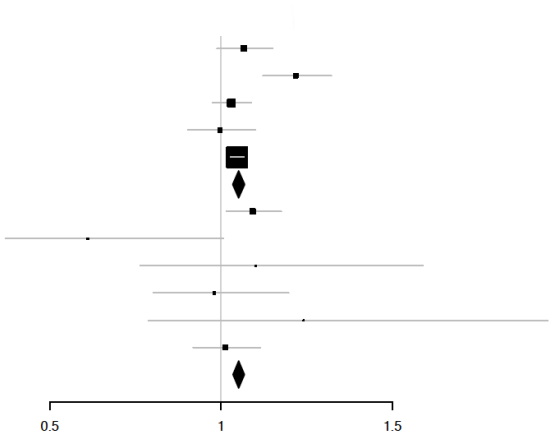


Supplementary Figure 3: Forest plots from the all studies meta-analysis for the 15 sentinel CAD-associated SNPs. N = number of subjects, EA= effect allele, EAF= effect allele frequency, OR = odds ratio, CI = confidence interval.



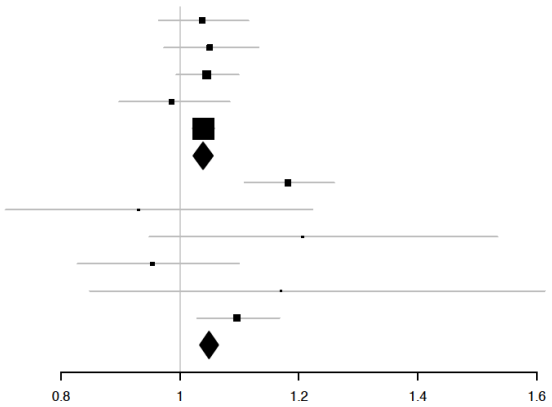
LMOD1
rs2820315C>T (1-201872264)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8558	T	0.31	1.07 (0.99–1.15)	
CGPS/CIHDS	5503	T	0.3	1.22 (1.12–1.32)	
EPIC–CVD	13202	T	0.28	1.03 (0.97–1.09)	
EPIC–CVD UMEA	2649	T	0.31	1 (0.9–1.1)	
CGplusC4D	184932	T	0.31	1.05 (1.03–1.07)	
EU meta–analysis	**214844**	**T**	**0.3**	**1.05 (1.03–1.07)**	**72.3**
South Asia	7528	T	0.25	1.09 (1.01–1.18)	
ARIC males	1172	T	0.08	0.61 (0.37–1.01)	
ARIC females	2031	T	0.08	1.1 (0.76–1.59)	
MIGEN	2670	T	0.09	0.98 (0.8–1.2)	
WHI	1953	T	0.1	1.24 (0.79–1.95)	
TAICHI	10487	T	0.11	1.01 (0.92–1.11)	
ALL meta–analysis	**240685**	**T**	**0.29**	**1.05 (1.03–1.07)**	**53.8**



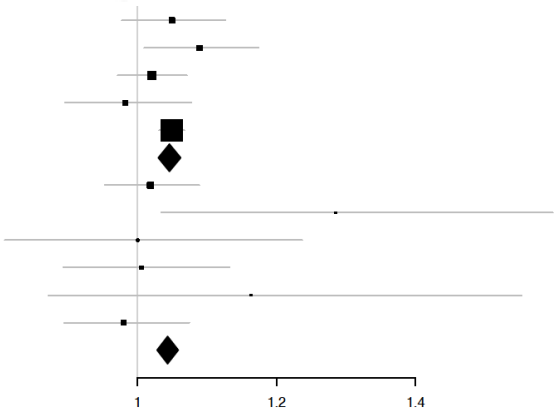
TNS1
rs2571445G>A (2-218683154)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8561	A	0.38	1.04 (0.96–1.11)	
CGPS/CIHDS	5506	A	0.39	1.05 (0.97–1.13)	
EPIC–CVD	13190	A	0.38	1.05 (0.99–1.1)	
EPIC–CVD UMEA	2649	A	0.41	0.99 (0.9–1.08)	
CGplusC4D	164348	A	0.4	1.04 (1.02–1.06)	
EU meta–analysis	**194254**	**A**	**0.39**	**1.04 (1.02–1.06)**	**0**
South Asia	7525	A	0.44	1.18 (1.11–1.26)	
ARIC males	1161	A	0.19	0.93 (0.71–1.22)	
ARIC females	2011	A	0.19	1.21 (0.95–1.53)	
MIGEN	2655	A	0.2	0.95 (0.83–1.1)	
WHI	1954	A	0.21	1.17 (0.85–1.61)	
TAICHI	10487	A	0.36	1.1 (1.03–1.17)	
ALL meta–analysis	**220047**	**A**	**0.39**	**1.05 (1.03–1.06)**	**54.8**



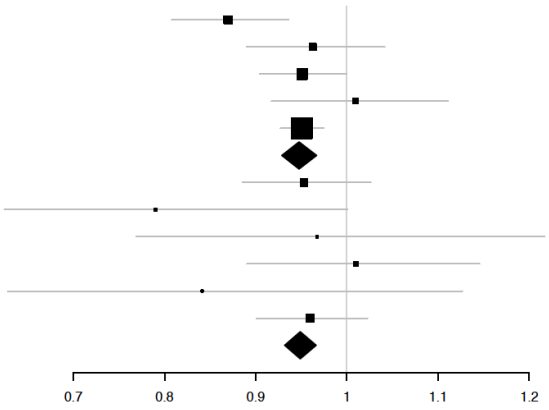
ARHGAP26
rs246600C>T (5-142516897)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8559	T	0.5	1.05 (0.98–1.13)	
CGPS/CIHDS	5506	T	0.5	1.09 (1.01–1.17)	
EPIC–CVD	13205	T	0.5	1.02 (0.97–1.07)	
EPIC–CVD UMEA	2649	T	0.5	0.98 (0.9–1.08)	
CGplusC4D	180461	T	0.48	1.05 (1.03–1.07)	
EU meta–analysis	**210380**	**T**	**0.48**	**1.05 (1.03–1.06)**	**0**
South Asia	7530	T	0.35	1.02 (0.95–1.09)	
ARIC males	1172	T	0.32	1.29 (1.03–1.6)	
ARIC females	2030	T	0.33	1 (0.81–1.24)	
MIGEN	2669	T	0.32	1.01 (0.89–1.13)	
WHI	1954	T	0.34	1.16 (0.87–1.55)	
TAICHI	10488	T	0.12	0.98 (0.89–1.07)	
ALL meta–analysis	**236223**	**T**	**0.46**	**1.04 (1.03–1.06)**	**8.5**



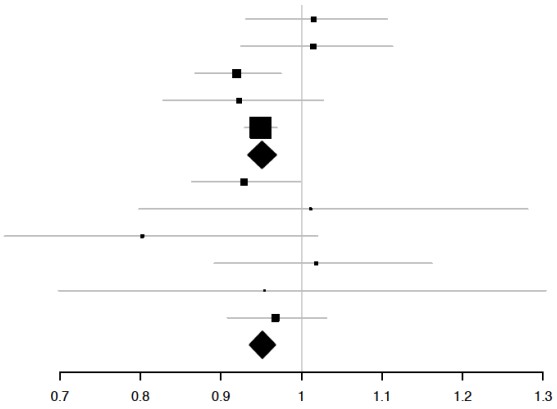
PARP12
rs10237377G>T (7-139757136)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8557	T	0.36	0.87 (0.81–0.94)	
CGPS/CIHDS	5506	T	0.36	0.96 (0.89–1.04)	
EPIC–CVD	13204	T	0.38	0.95 (0.9–1)	
EPIC–CVD UMEA	2649	T	0.35	1.01 (0.92–1.11)	
CGplusC4D	151643	T	0.34	0.95 (0.93–0.97)	
EU meta–analysis	**181559**	**T**	**0.35**	**0.95 (0.93–0.97)**	**44.1**
South Asia	7529	T	0.24	0.95 (0.89–1.03)	
ARIC males	1172	T	0.73	0.79 (0.62–1)	
ARIC females	2031	T	0.72	0.97 (0.77–1.22)	
MIGEN	2667	T	0.71	1.01 (0.89–1.15)	
WHI	1954	T	0.7	0.84 (0.63–1.13)	
TAICHI	10487	T	0.6	0.96 (0.9–1.02)	
ALL meta–analysis	**207399**	**T**	**0.38**	**0.95 (0.93–0.97)**	**11**



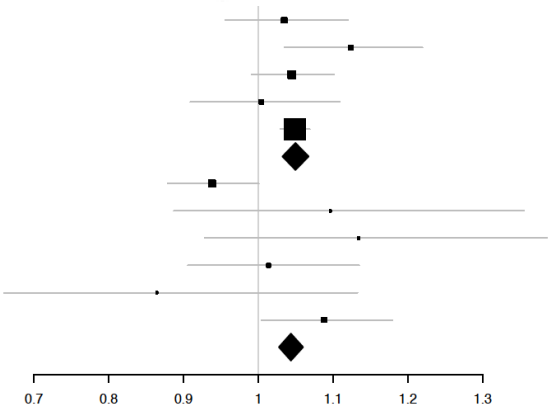
PCNX3
rs12801636G>A (11-65391317)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8559	A	0.22	1.01 (0.93–1.11)	
CGPS/CIHDS	5506	A	0.21	1.01 (0.92–1.11)	
EPIC–CVD	13188	A	0.22	0.92 (0.87–0.97)	
EPIC–CVD UMEA	2649	A	0.24	0.92 (0.83–1.03)	
CGplusC4D	181250	A	0.23	0.95 (0.93–0.97)	
EU meta–analysis	**211152**	**A**	**0.23**	**0.95 (0.93–0.97)**	**29.8**
South Asia	7530	A	0.27	0.93 (0.86–1)	
ARIC males	1171	A	0.27	1.01 (0.8–1.28)	
ARIC females	2029	A	0.26	0.8 (0.63–1.02)	
MIGEN	2663	A	0.25	1.02 (0.89–1.16)	
WHI	1954	A	0.25	0.95 (0.7–1.3)	
TAICHI	10486	A	0.4	0.97 (0.91–1.03)	
ALL meta–analysis	**236985**	**A**	**0.25**	**0.95 (0.94–0.97)**	**0**



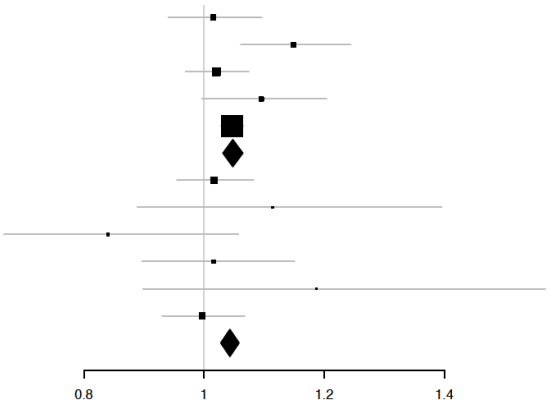
SERPINH1
rs590121G>T (11-75274150)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8555	T	0.28	1.03 (0.96–1.12)	
CGPS/CIHDS	5498	T	0.28	1.12 (1.03–1.22)	
EPIC–CVD	13175	T	0.3	1.04 (0.99–1.1)	
EPIC–CVD UMEA	2646	T	0.33	1 (0.91–1.11)	
CGplusC4D	177552	T	0.3	1.05 (1.03–1.07)	
EU meta–analysis	**207426**	**T**	**0.3**	**1.05 (1.03–1.07)**	**0**
South Asia	7526	T	0.41	0.94 (0.88–1)	
ARIC males	1168	T	0.46	1.1 (0.89–1.36)	
ARIC females	2029	T	0.45	1.13 (0.93–1.39)	
MIGEN	2661	T	0.46	1.01 (0.91–1.13)	
WHI	1954	T	0.45	0.86 (0.66–1.13)	
TAICHI	10482	T	0.17	1.09 (1–1.18)	
ALL meta–analysis	**233246**	**T**	**0.31**	**1.04 (1.03–1.06)**	**45.8**



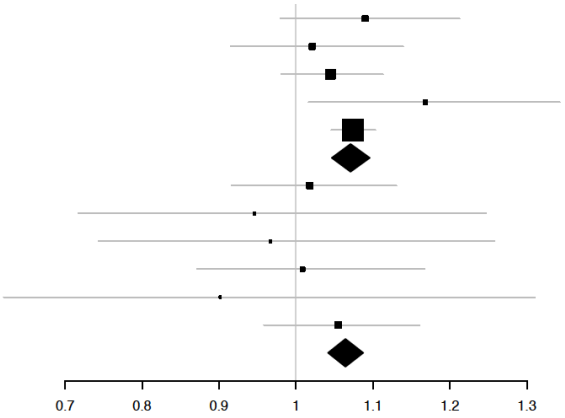
C12orf43
rs2258287C>A (12-121454313)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8558	A	0.32	1.02 (0.94–1.1)	
CGPS/CIHDS	5504	A	0.32	1.15 (1.06–1.24)	
EPIC–CVD	13200	A	0.33	1.02 (0.97–1.07)	
EPIC–CVD UMEA	2649	A	0.39	1.1 (1–1.2)	
CGplusC4D	191157	A	0.34	1.05 (1.03–1.07)	
EU meta–analysis	**221068**	**A**	**0.34**	**1.05 (1.03–1.06)**	**48.4**
South Asia	7525	A	0.5	1.02 (0.95–1.08)	
ARIC males	1172	A	0.3	1.11 (0.89–1.4)	
ARIC females	2031	A	0.29	0.84 (0.67–1.06)	
MIGEN	2667	A	0.3	1.02 (0.9–1.15)	
WHI	1954	A	0.3	1.19 (0.9–1.57)	
TAICHI	10484	A	0.74	1 (0.93–1.07)	
ALL meta–analysis	**246901**	**A**	**0.37**	**1.04 (1.03–1.06)**	**33.9**



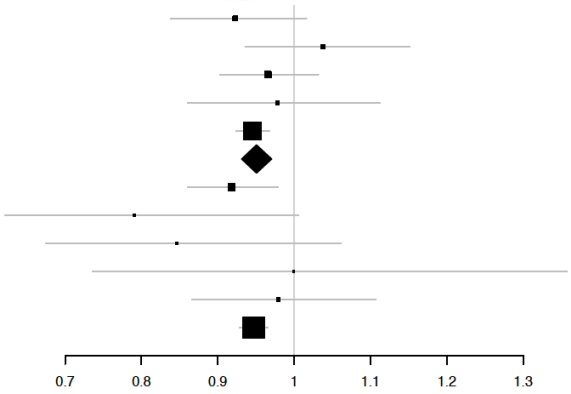
SCARB1
rs11057830G>A (12-125307053)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8561	A	0.13	1.09 (0.98–1.21)	
CGPS/CIHDS	5506	A	0.14	1.02 (0.91–1.14)	
EPIC–CVD	13205	A	0.18	1.04 (0.98–1.11)	
EPIC–CVD UMEA	2649	A	0.13	1.17 (1.02–1.34)	
CGplusC4D	147629	A	0.16	1.07 (1.05–1.1)	
EU meta–analysis	**177550**	**A**	**0.16**	**1.07 (1.05–1.1)**	**0**
South Asia	7531	A	0.1	1.02 (0.92–1.13)	
ARIC males	1172	A	0.18	0.95 (0.72–1.25)	
ARIC females	2032	A	0.18	0.97 (0.74–1.26)	
MIGEN	2667	A	0.18	1.01 (0.87–1.17)	
WHI	1954	A	0.16	0.9 (0.62–1.31)	
TAICHI	10488	A	0.11	1.05 (0.96–1.16)	
ALL meta–analysis	**203394**	**A**	**0.15**	**1.06 (1.04–1.09)**	**0**



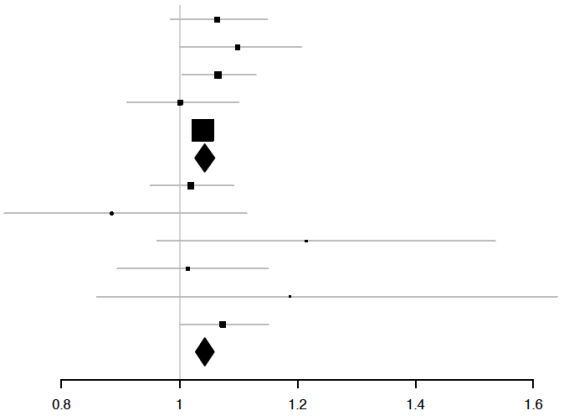
OAZ2/RBPMS2
rs6494488A>G (15-65024204)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8560	G	0.16	0.92 (0.84–1.02)	
CGPS/CIHDS	5503	G	0.16	1.04 (0.94–1.15)	
EPIC–CVD	13179	G	0.16	0.97 (0.9–1.03)	
EPIC–CVD UMEA	2648	G	0.15	0.98 (0.86–1.11)	
CGplusC4D	175520	G	0.18	0.95 (0.92–0.97)	
EU meta–analysis	**205410**	**G**	**0.18**	**0.95 (0.93–0.97)**	**0**
South Asia	7527	G	0.46	0.92 (0.86–0.98)	
ARIC males	1171	G	0.69	0.79 (0.62–1.01)	
ARIC females	2030	G	0.69	0.85 (0.68–1.06)	
WHI	1954	G	0.65	1 (0.74–1.36)	
TAICHI	10486	G	0.06	0.98 (0.87–1.11)	
ALL meta–analysis	228578	G	0.21	0.95 (0.93–0.97)	



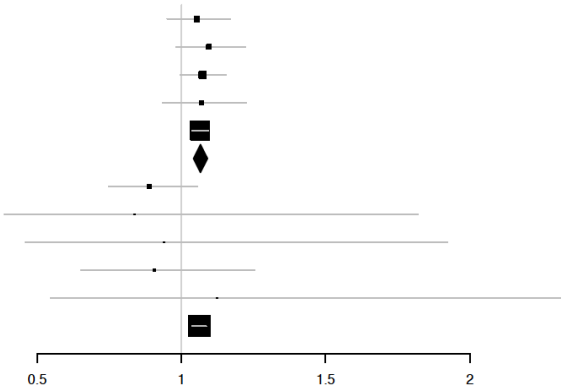
DHX38
rs1050362C>A (16-72130815)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8558	A	0.37	1.06 (0.99–1.15)	
CGPS/CIHDS	5504	A	0.37	1.1 (1–1.21)	
EPIC–CVD	13203	A	0.37	1.06 (1–1.13)	
EPIC–CVD UMEA	2647	A	0.42	1 (0.91–1.1)	
CGplusC4D	186113	A	0.39	1.04 (1.02–1.06)	
EU meta–analysis	**216025**	**A**	**0.38**	**1.04 (1.03–1.06)**	**0**
South Asia	7526	A	0.47	1.02 (0.95–1.09)	
ARIC males	1171	A	0.69	0.89 (0.7–1.11)	
ARIC females	2031	A	0.7	1.22 (0.96–1.53)	
MIGEN	2666	A	0.69	1.01 (0.89–1.15)	
WHI	1954	A	0.7	1.19 (0.86–1.64)	
TAICHI	10485	A	0.26	1.07 (1–1.15)	
ALL meta–analysis	**241858**	**A**	**0.39**	**1.04 (1.03–1.06)**	**0**



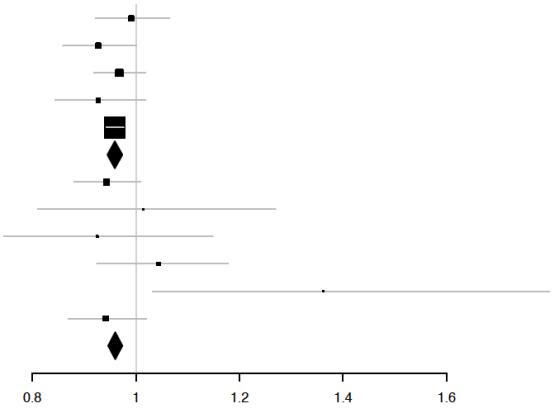
GOSR2
rs17608766T>C (17-45013271)

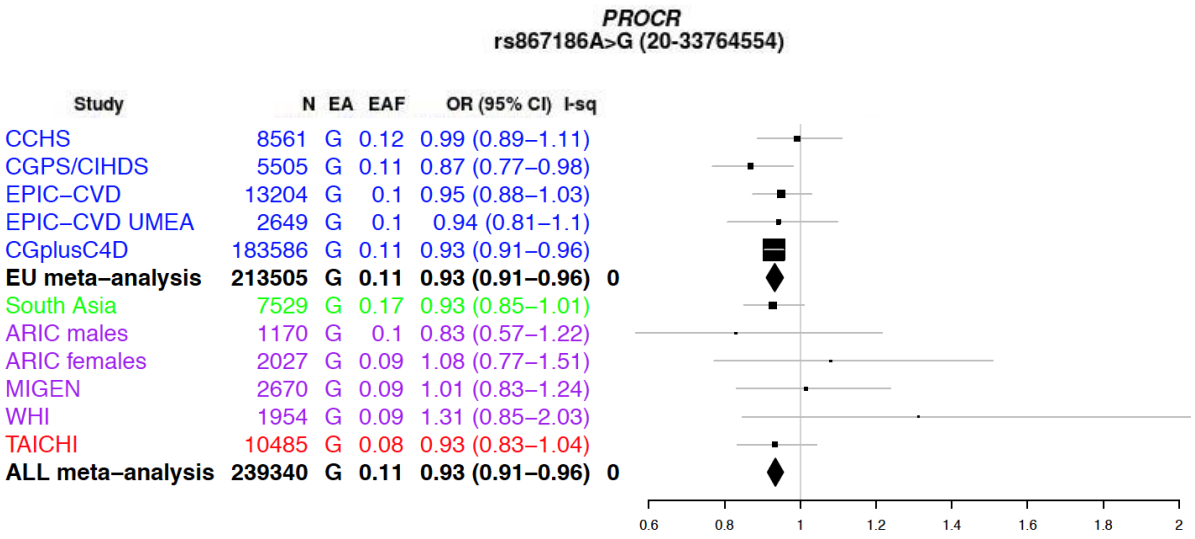
Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8559	C	0.14	1.05 (0.95–1.17)	
CGPS/CIHDS	5500	C	0.14	1.09 (0.98–1.22)	
EPIC–CVD	13199	C	0.13	1.07 (1–1.16)	
EPIC–CVD UMEA	2649	C	0.14	1.07 (0.93–1.22)	
CGplusC4D	185950	C	0.14	1.06 (1.04–1.09)	
EU meta–analysis	**215857**	**C**	**0.14**	**1.07 (1.04–1.09)**	**0**
South Asia	7530	C	0.04	0.89 (0.75–1.05)	
ARIC males	1172	C	0.02	0.84 (0.38–1.82)	
ARIC females	2031	C	0.02	0.94 (0.46–1.92)	
MIGEN	2669	C	0.03	0.9 (0.65–1.25)	
WHI	1954	C	0.03	1.12 (0.55–2.31)	
ALL meta–analysis	231213	C	0.14	1.06 (1.04–1.09)	



PECAM1
rs1867624T>C (17-62387091)

Study	N	EA	EAF	OR (95% CI)	I-sq
CCHS	8561	C	0.41	0.99 (0.92–1.07)	
CGPS/CIHDS	5503	C	0.4	0.93 (0.86–1)	
EPIC–CVD	13202	C	0.36	0.97 (0.92–1.02)	
EPIC–CVD UMEA	2649	C	0.38	0.93 (0.84–1.02)	
CGplusC4D	190916	C	0.39	0.96 (0.94–0.98)	
EU meta–analysis	**220831**	**C**	**0.39**	**0.96 (0.94–0.97)**	**0**
South Asia	7527	C	0.33	0.94 (0.88–1.01)	
ARIC males	1172	C	0.3	1.01 (0.81–1.27)	
ARIC females	2032	C	0.31	0.92 (0.74–1.15)	
MIGEN	2670	C	0.31	1.04 (0.92–1.18)	
WHI	1954	C	0.32	1.36 (1.03–1.8)	
TAICHI	10488	C	0.18	0.94 (0.87–1.02)	
ALL meta–analysis	**246674**	**C**	**0.38**	**0.96 (0.95–0.97)**	**8.6**

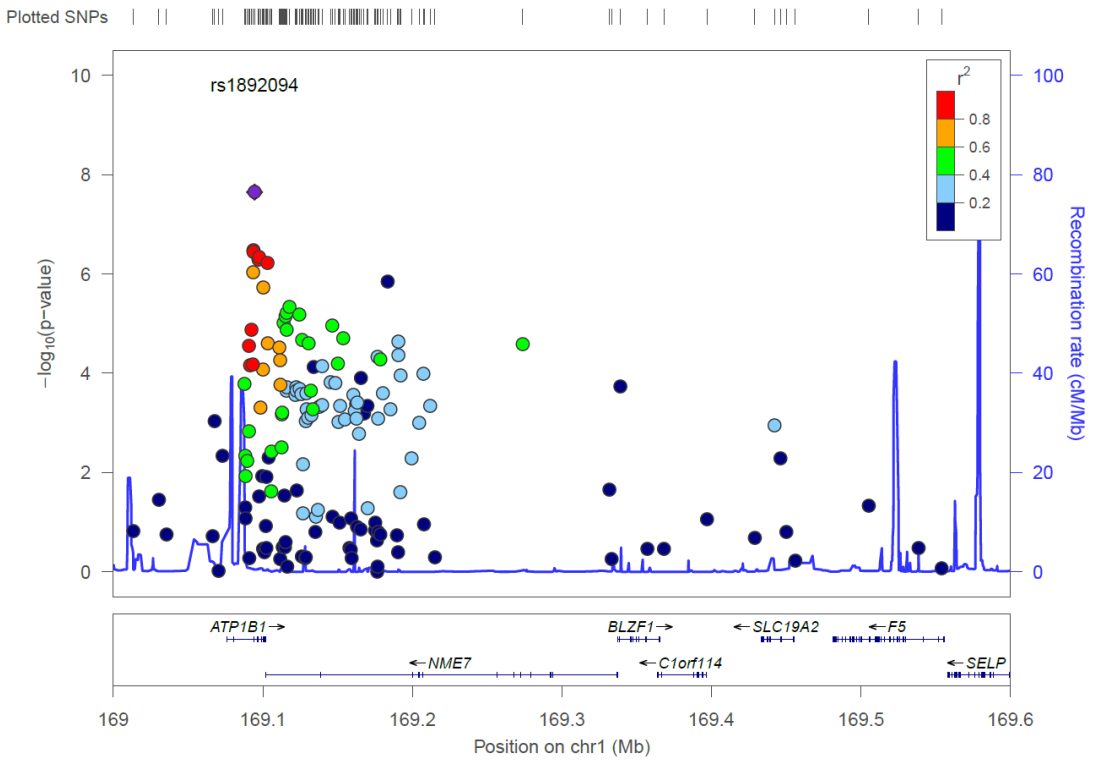



PROCR
rs867186A>G (20-33764554)
Study N EA EAF OR (95% CI) I-sq
CCHS 8561 G 0.12 0.99 (0.89–1.11)
CGPS/CIHDS 5505 G 0.11 0.87 (0.77–0.98)
EPIC–CVD 13204 G 0.1 0.95 (0.88–1.03)
EPIC–CVD UMEA 2649 G 0.1 0.94 (0.81–1.1)
CGplusC4D 183586 G 0.11 0.93 (0.91–0.96)
EU meta–analysis 213505 G 0.11 0.93 (0.91–0.96) 0
South Asia 7529 G 0.17 0.93 (0.85–1.01)
ARIC males 1170 G 0.1 0.83 (0.57–1.22)
ARIC females 2027 G 0.09 1.08 (0.77–1.51)
MIGEN 2670 G 0.09 1.01 (0.83–1.24)
WHI 1954 G 0.09 1.31 (0.85–2.03)
TAICHI 10485 G 0.08 0.93 (0.83–1.04)
ALL meta–analysis 239340 G 0.11 0.93 (0.91–0.96) 0
0.6
0.8
1
1.2
1.4
1.6
1.8
2

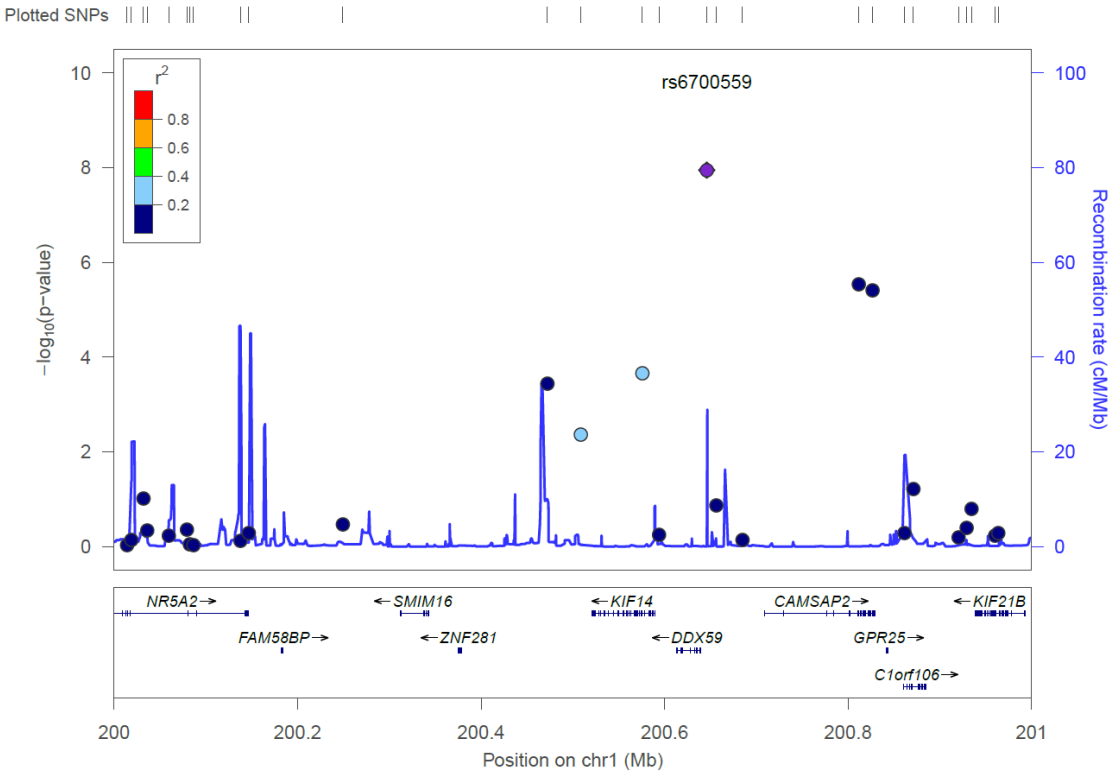
Supplementary Figure 4: Regional association plots for novel CAD associated loci (a) from the all ancestry meta-analysis for 13 loci and the European meta-analysis for *GOSR* and *SERPINH1*. (b) from the publicly available CARDIoGRAMplusC4D 1000G imputed GWAS results[1]. The r^2 information was calculated from the phased genotypes of 1000 Genome phase3 v5 (11/04/2014) super-populations (*PROCR* is given in Supplementary Figure 8).

Supplementary Figure 4(a)

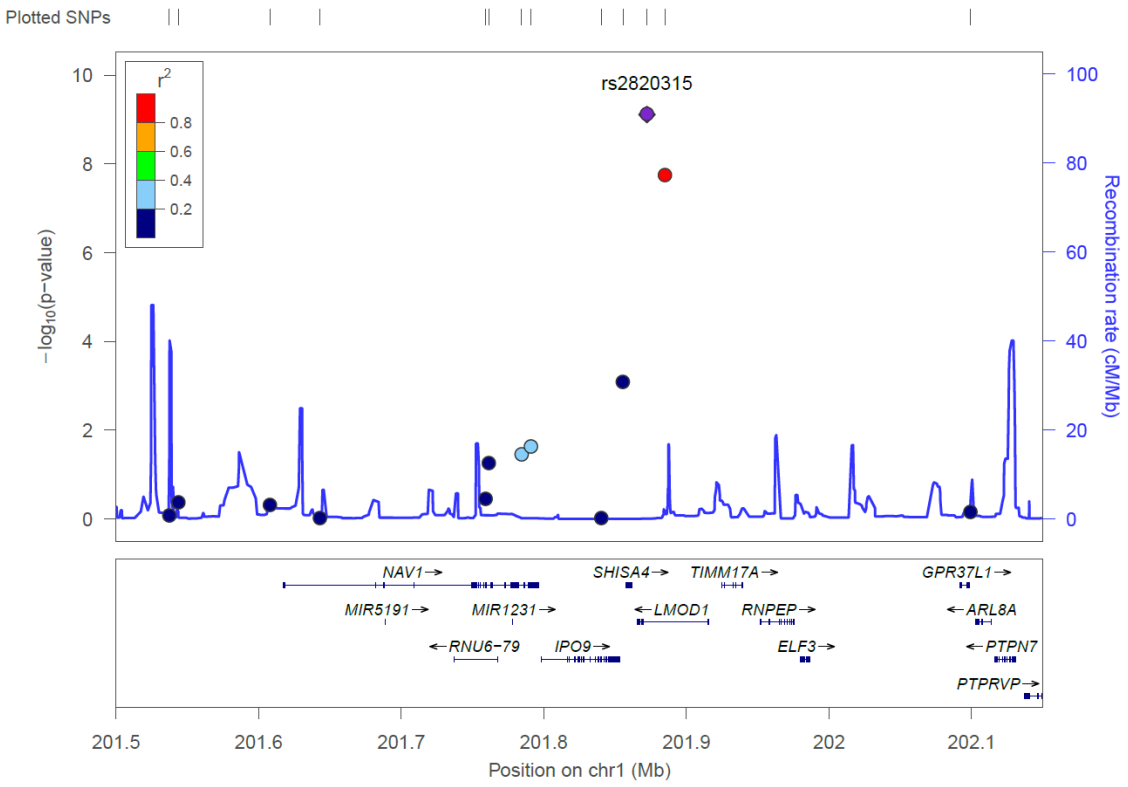
ATP1B1



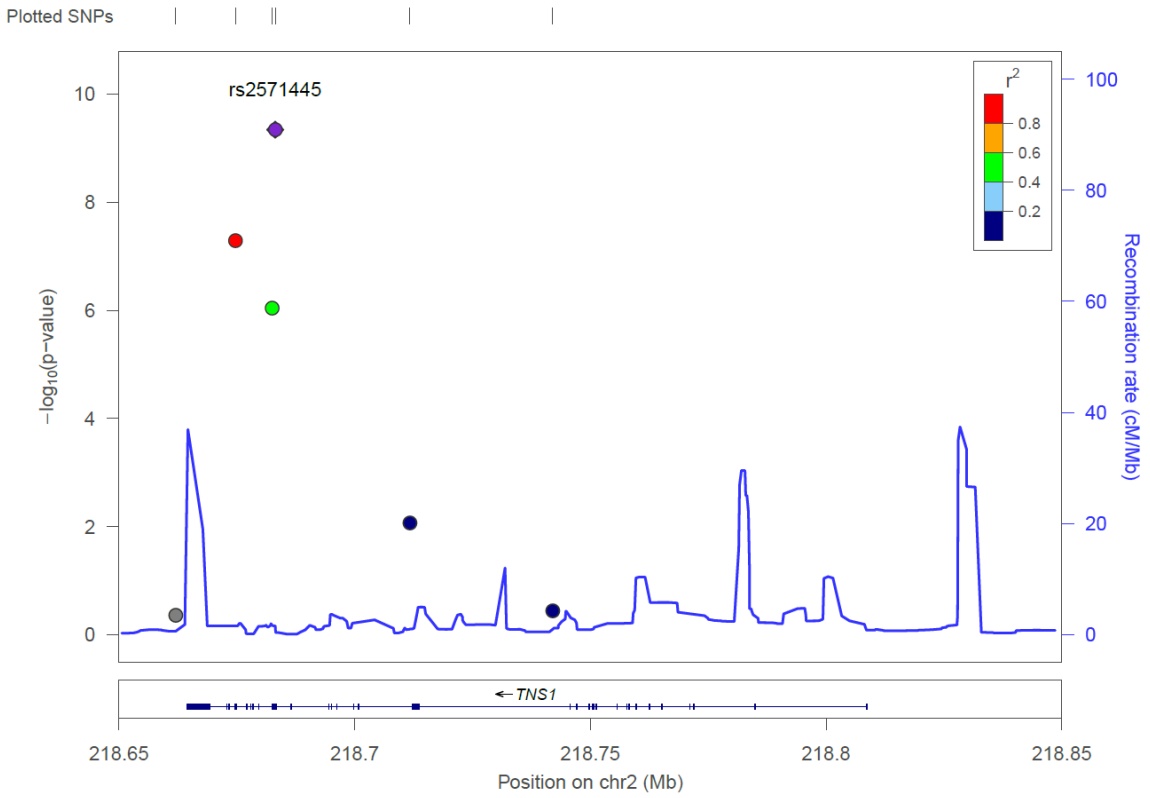
DDX59/CAMSAP2



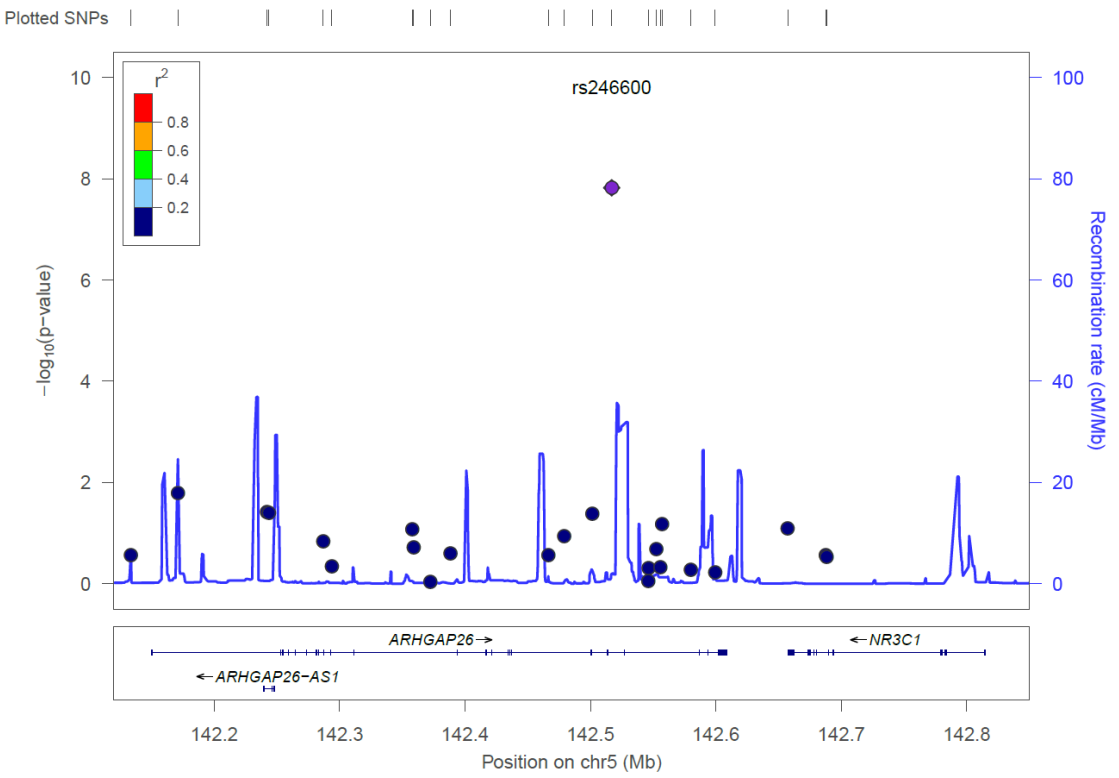
LMOD1



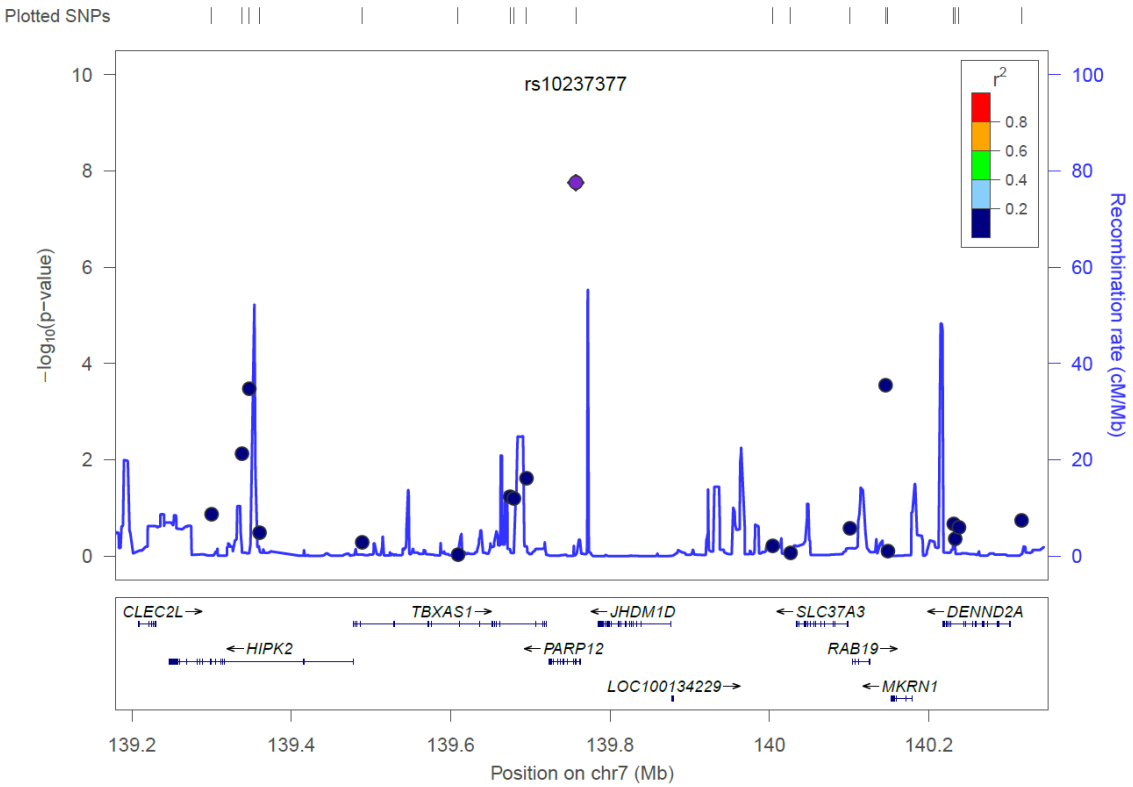
TNS1



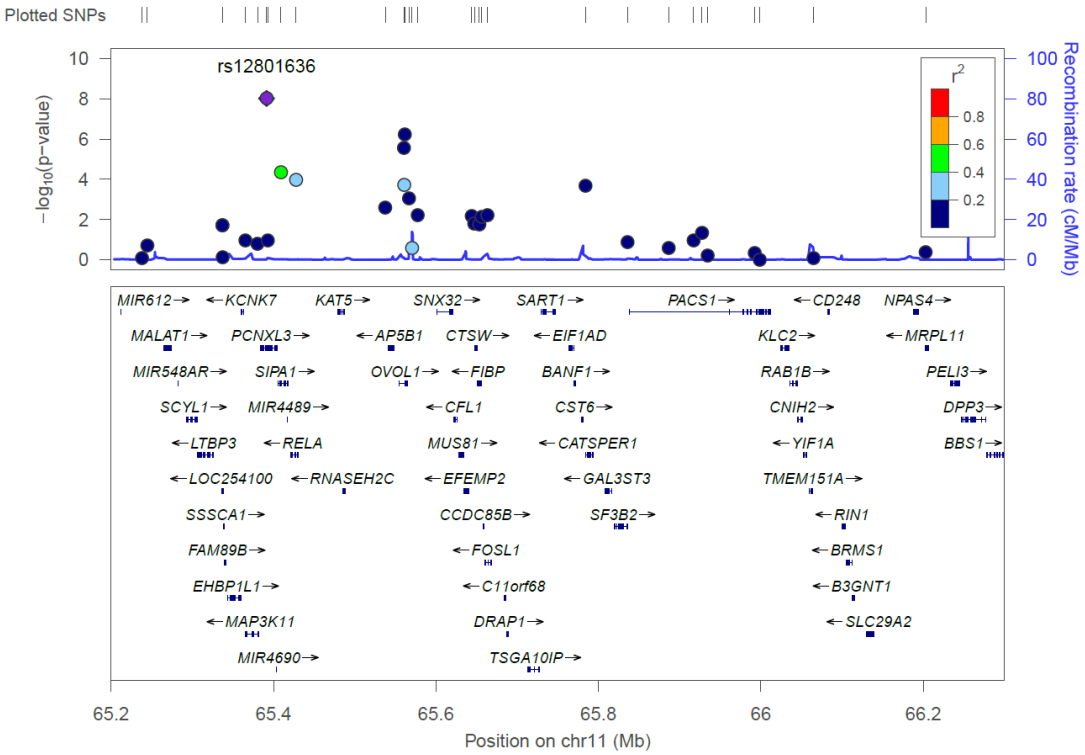
ARHGAP26



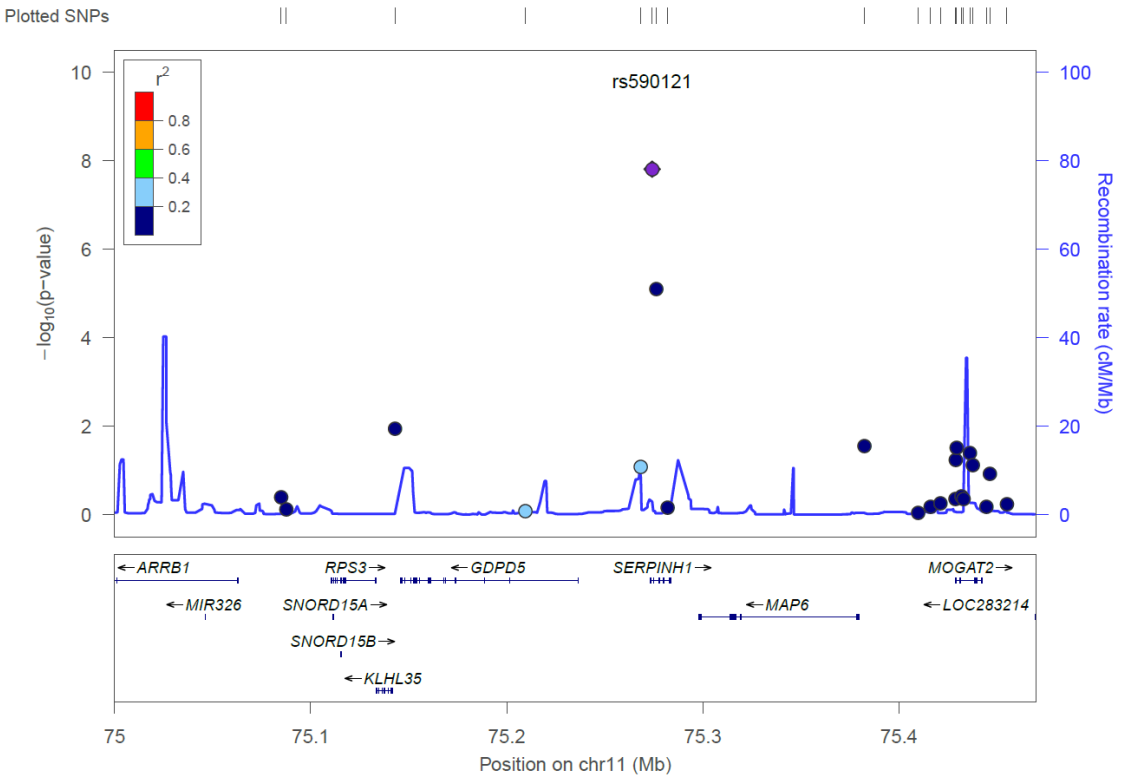
PARP12


Plotted SNPs
rs10237377
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
CLEC2L→
TBXAS1→
←JHDM1D
←SLC37A3
←DENND2A
←HIPK2
←PARP12
RAB19→
LOC100134229→
←MKRN1
Position on chr7 (Mb)

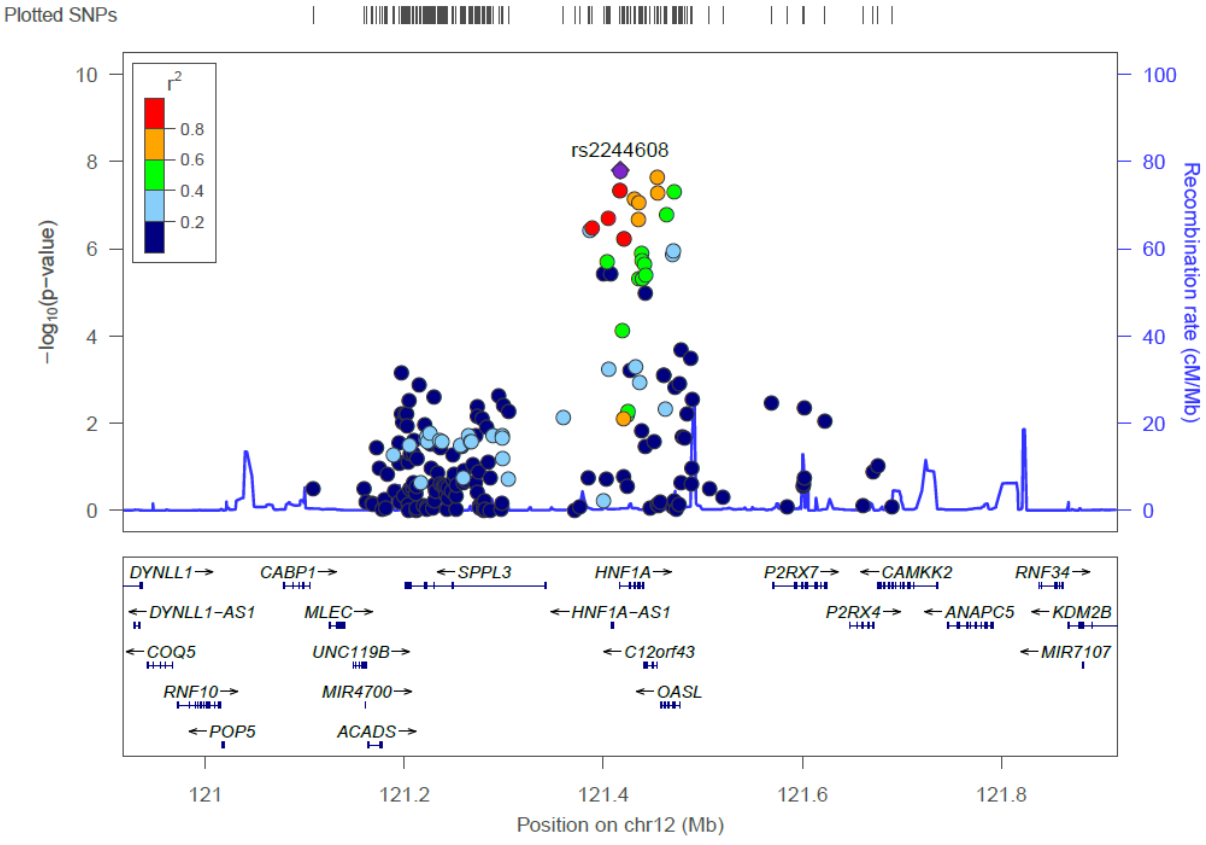
PCNX3


Plotted SNPs
rs12801636
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
MIR612→
←KCNK7
KAT5→
SNX32→
SART1→
PACS1→
←CD248
NPAS4→
MALAT1→
PCNXL3→
←AP5B1
CTSW→
←EIF1AD
KLC2→
←MRPL11
MIR548AR→
SIPA1→
OVOL1→
←FIBP
BANF1→
RAB1B→
PELI3→
SCYL1→
MIR4489→
←CFL1
CST6→
CNIH2→
DPP3→
←LTBP3
←RELA
MUS81→
←CATSPER1
←YIF1A
BBS1→
←LOC254100
←RNASEH2C
←EFEMP2
←GAL3ST3
TMEM151A→
SSSCA1→
CCDC85B→
SF3B2→
←RIN1
FAM89B→
←FOSL1
←BRMS1
EHBP1L1→
←C11orf68
←B3GNT1
←MAP3K11
DRAP1→
←SLC29A2
MIR4690→
TSGA10IP→
Position on chr11 (Mb)

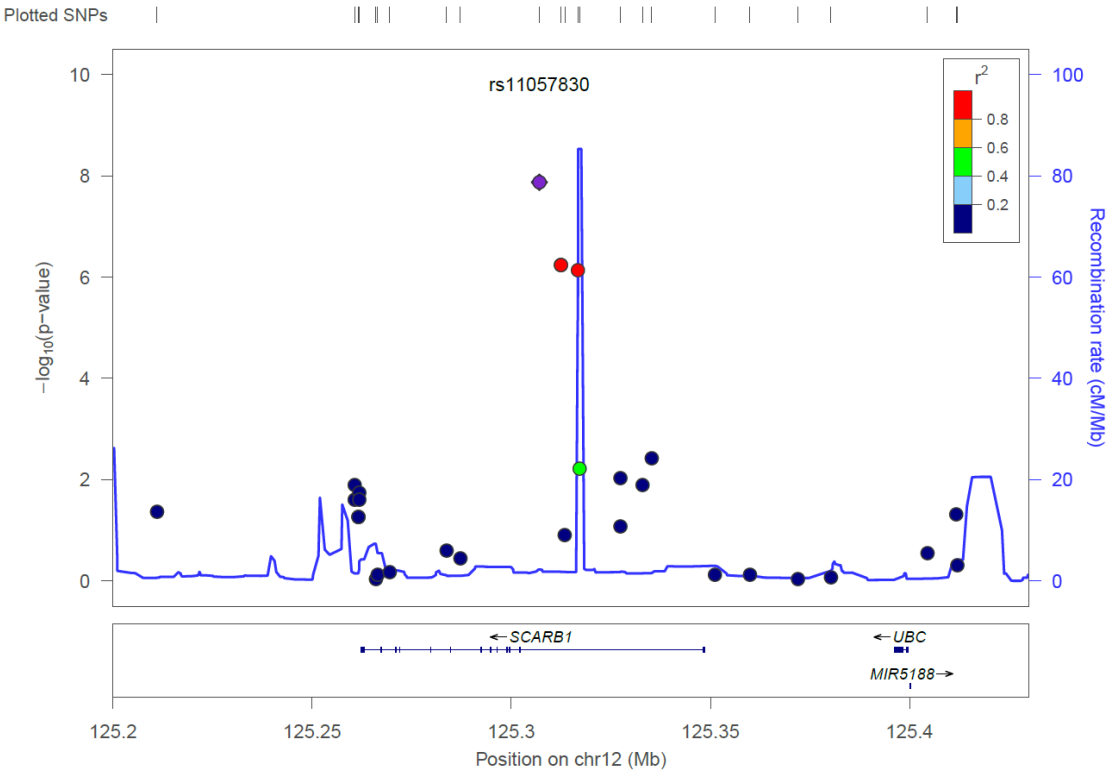
SERPINH1



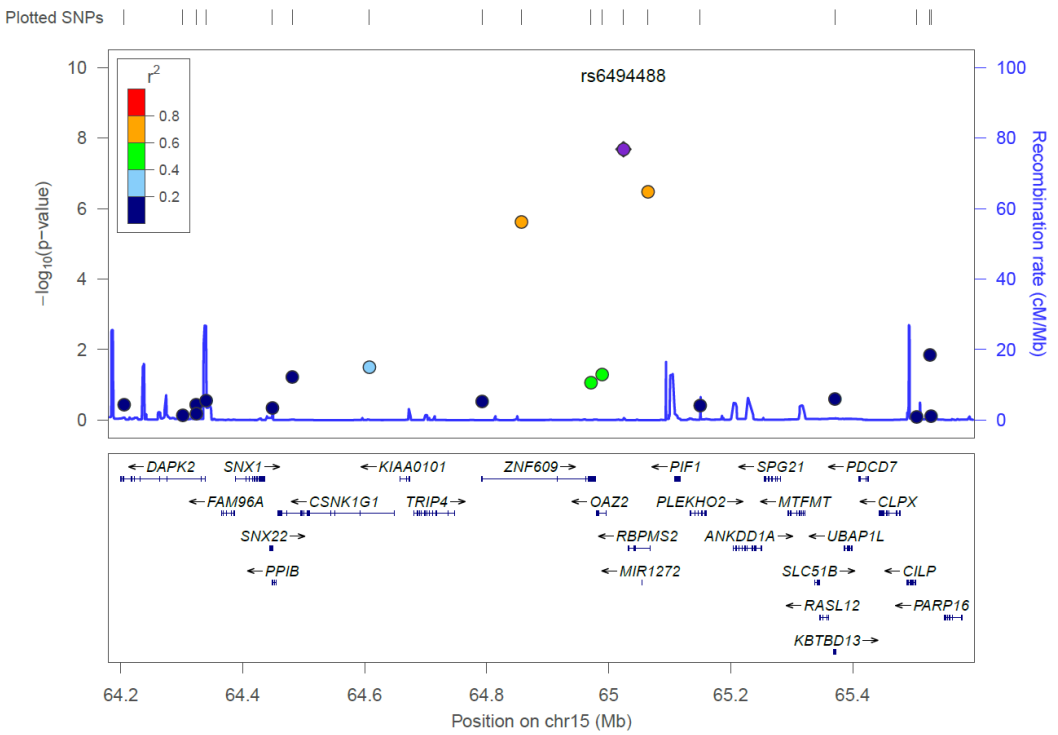
C12orf43/HNF1A



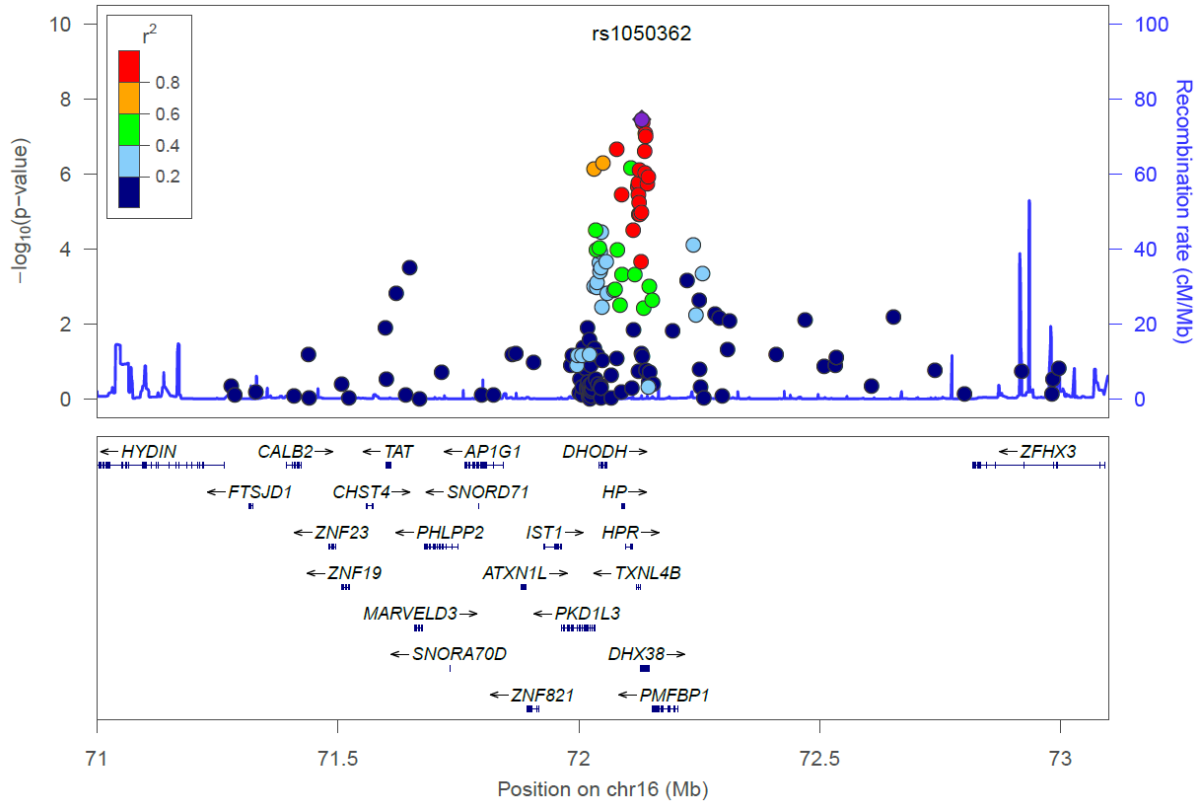
SCARB1



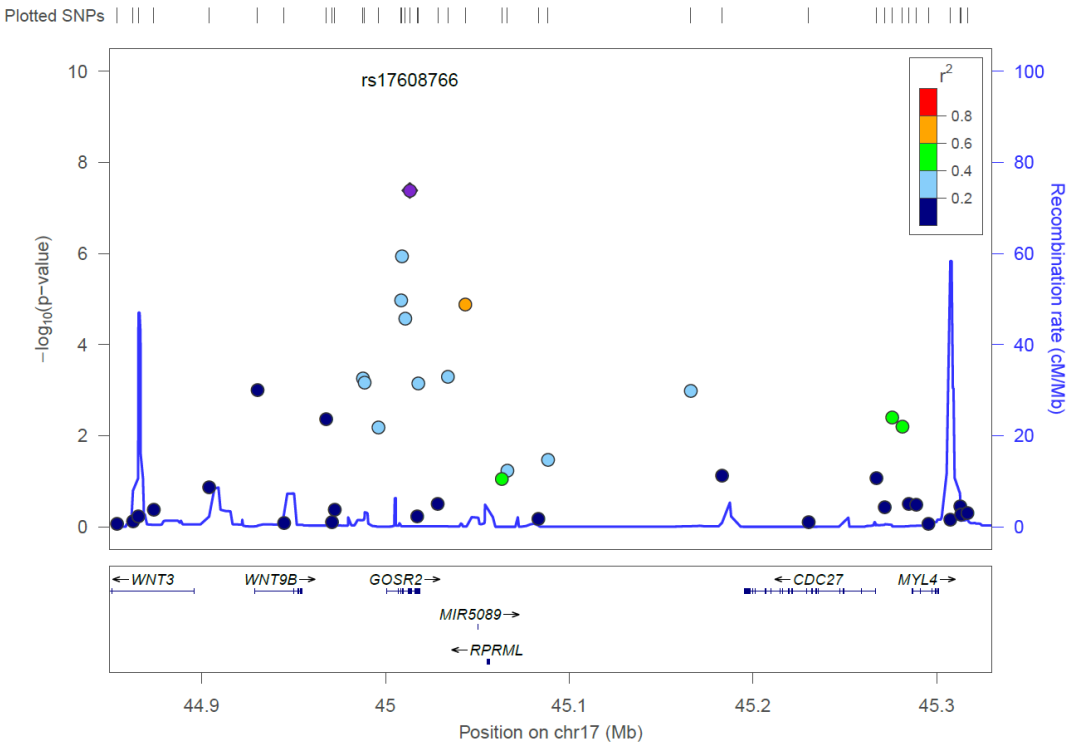
OAZ2, RBPMS2



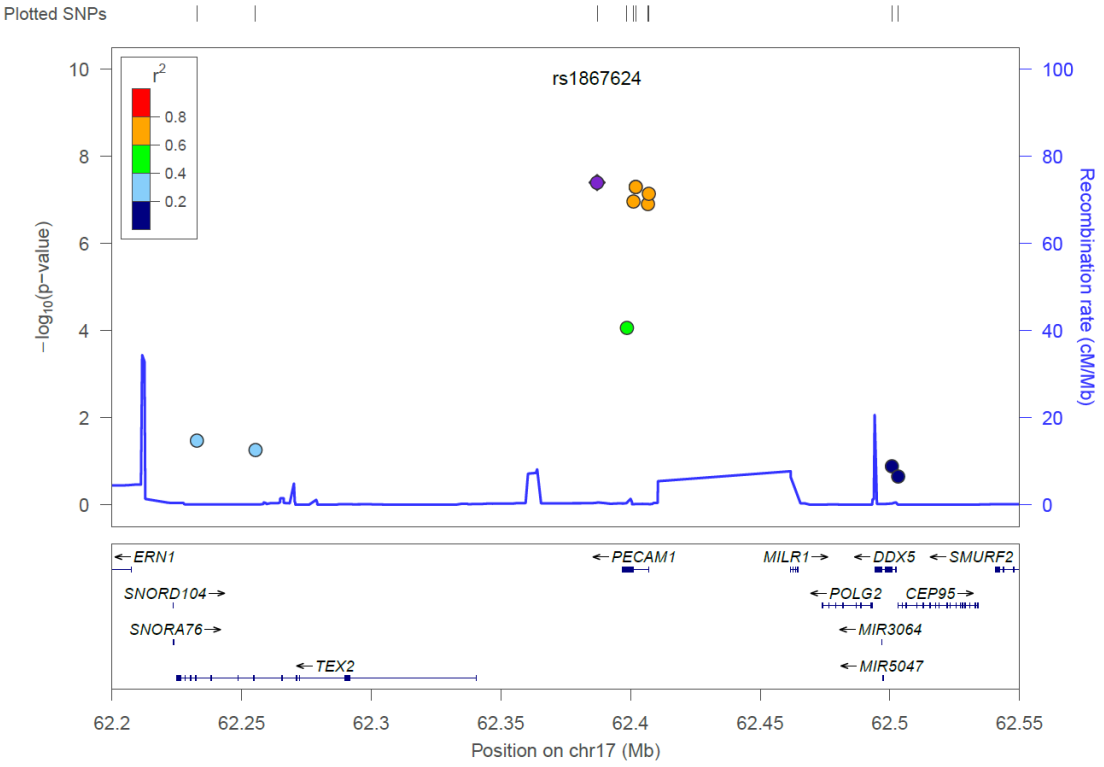
DHX38



GOSR2

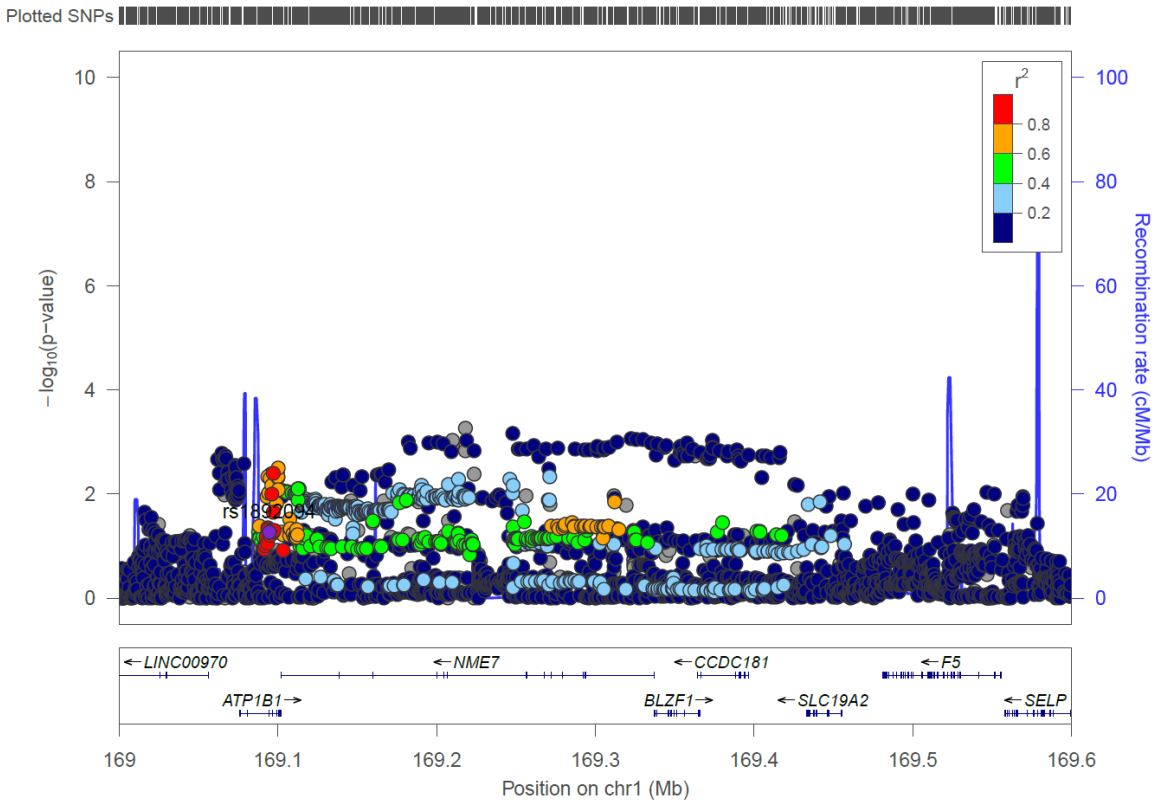


PECAM1

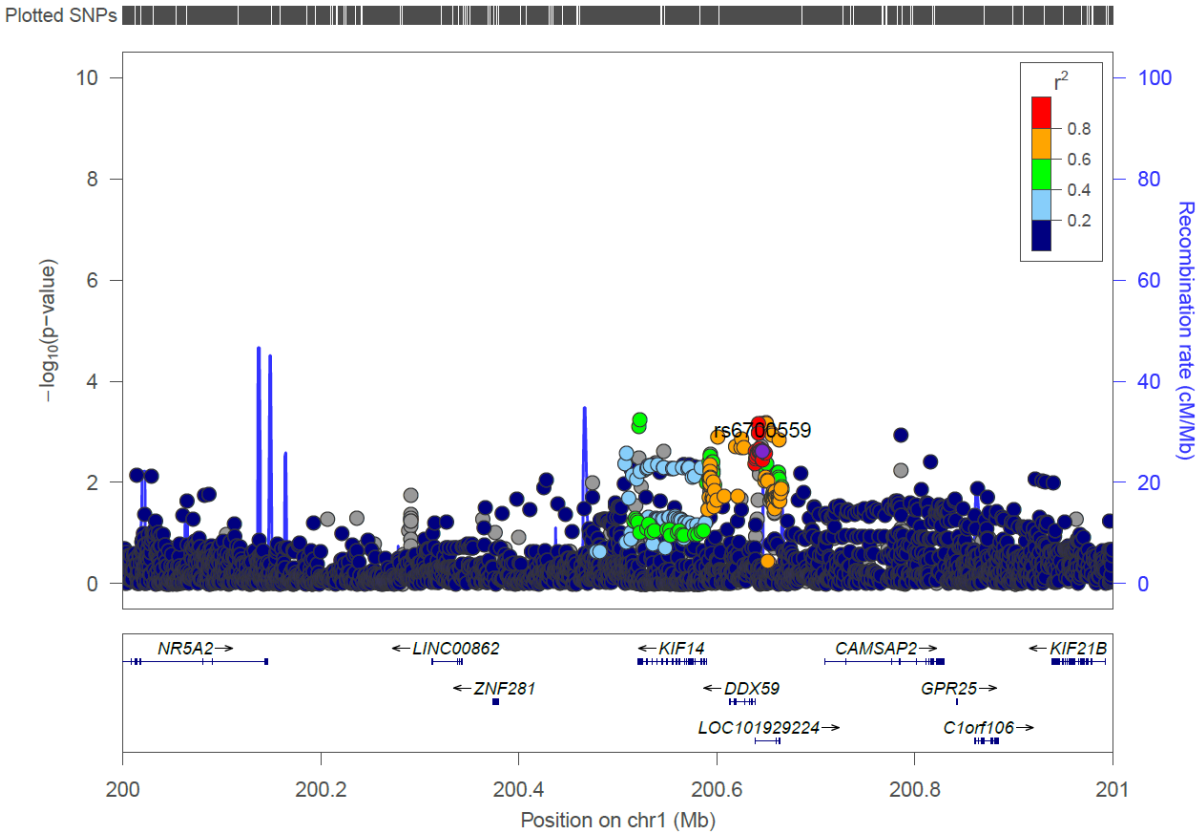


Supplementary Figure 4(b)

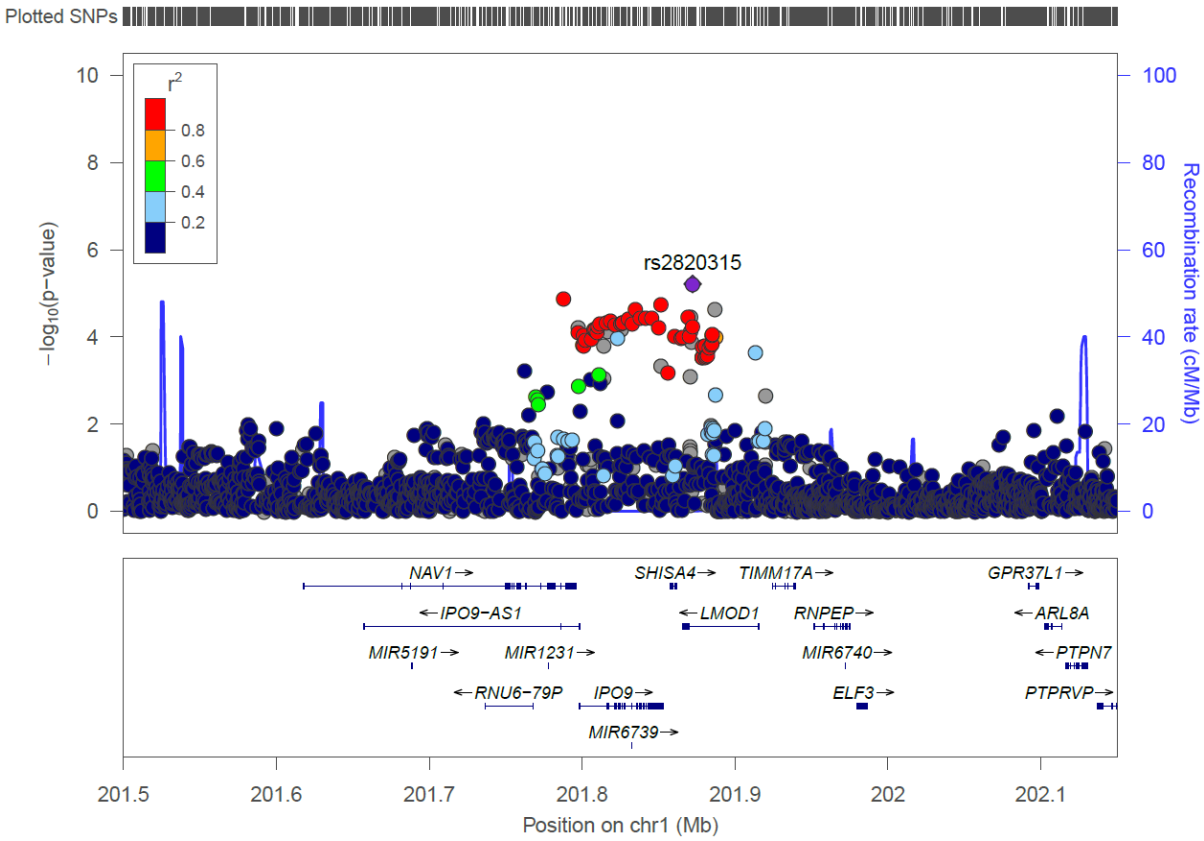
ATP1B1 (1000G)



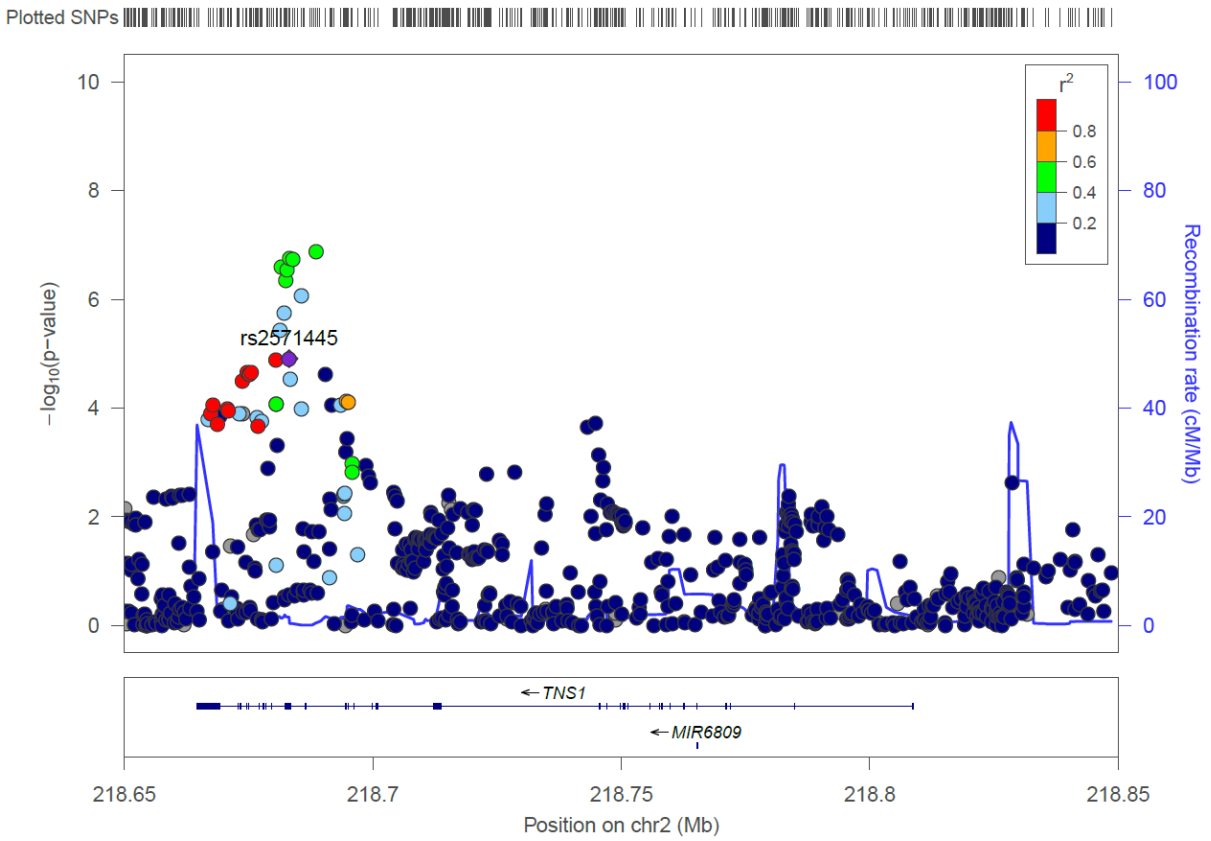
***DDX59/CAMSAP2* (1000G)**



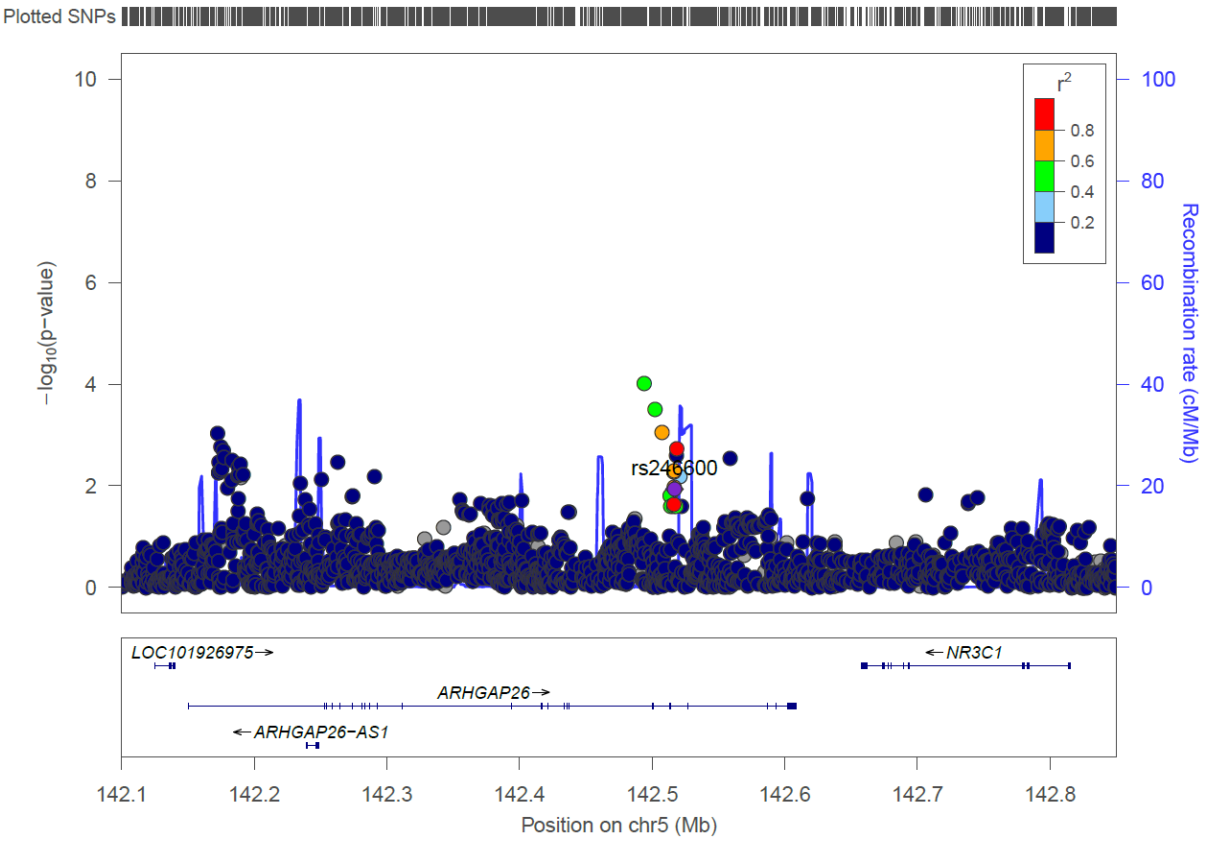
***LMOD1* (1000G)**



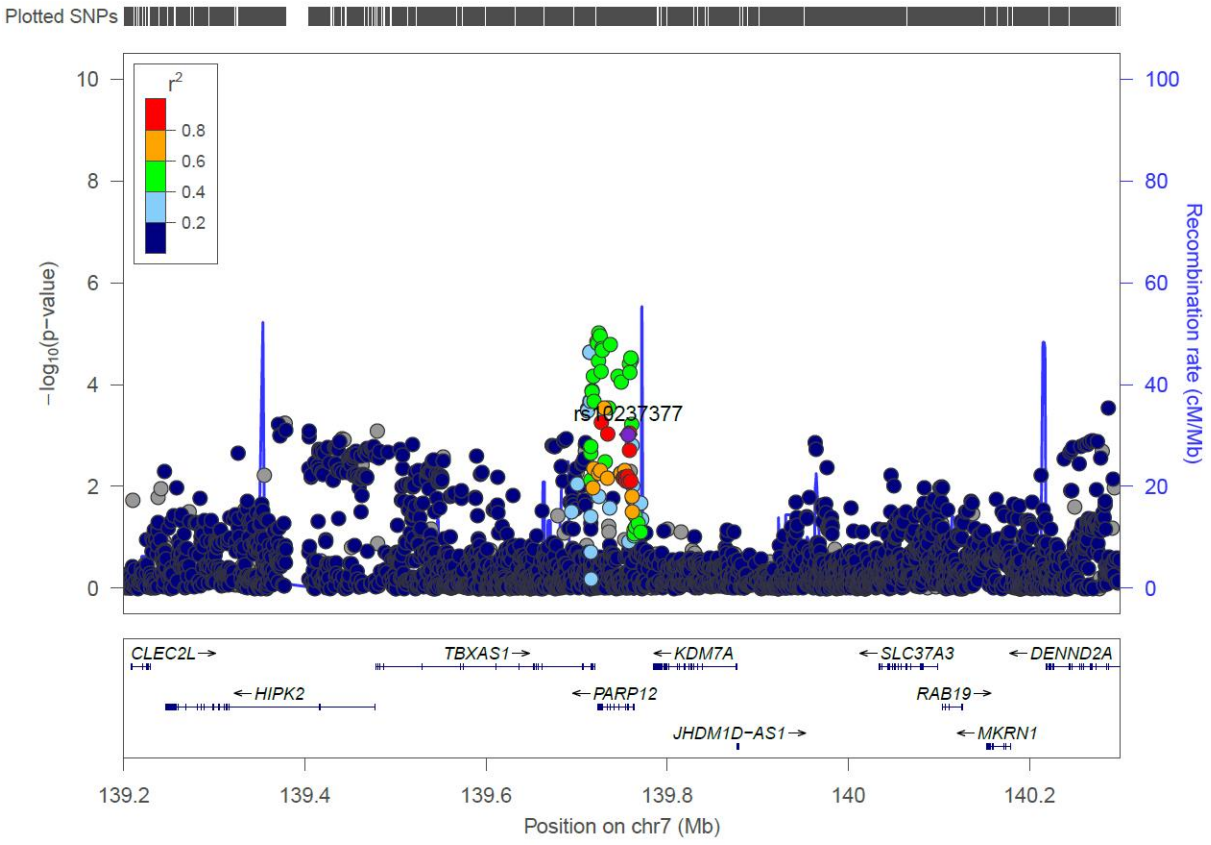
***TNS1* (1000G)**



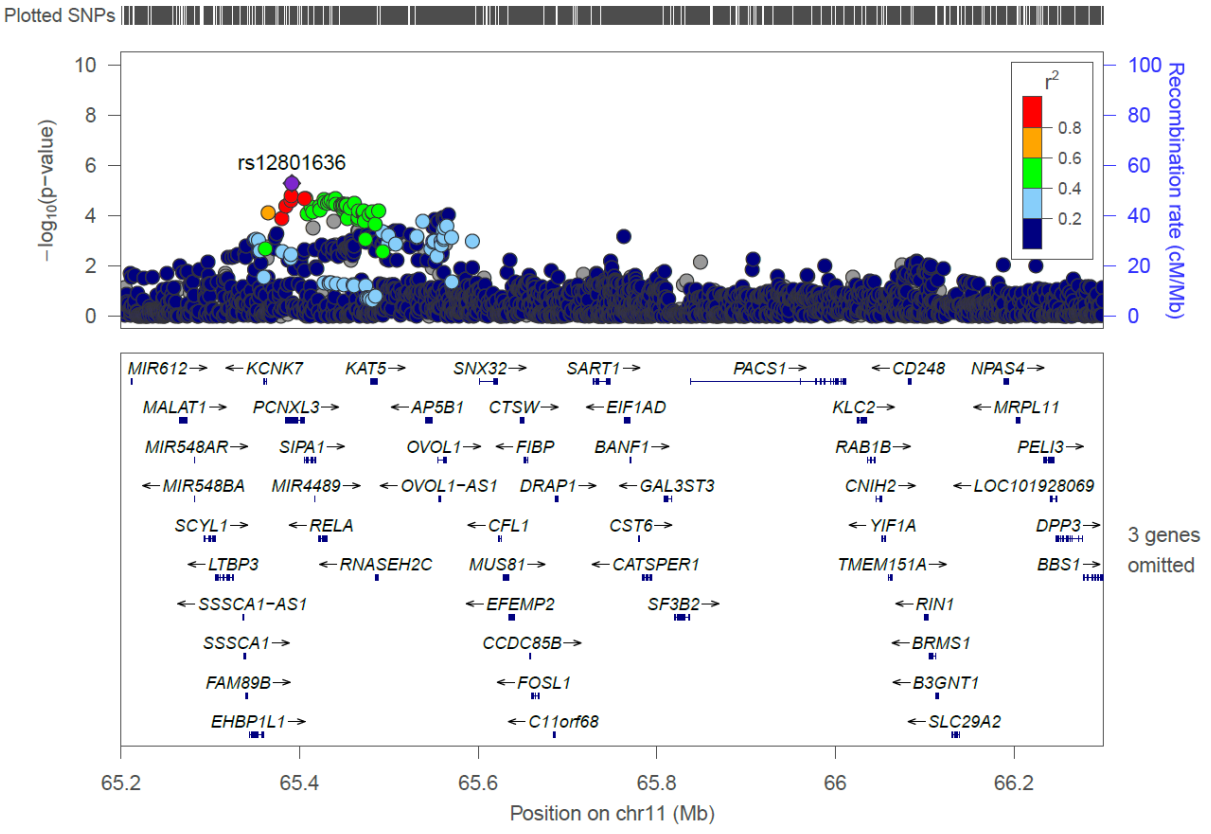
***ARHGAP26* (1000G)**



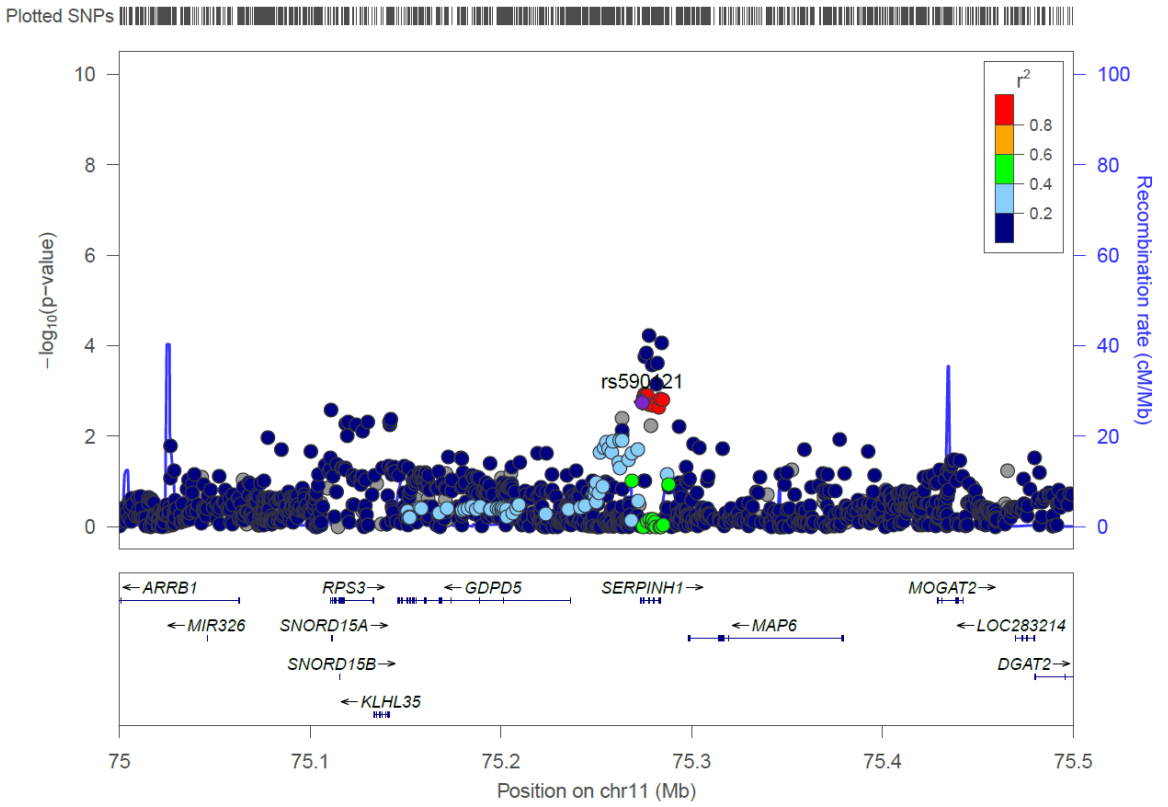
***PARP12* (1000G)**



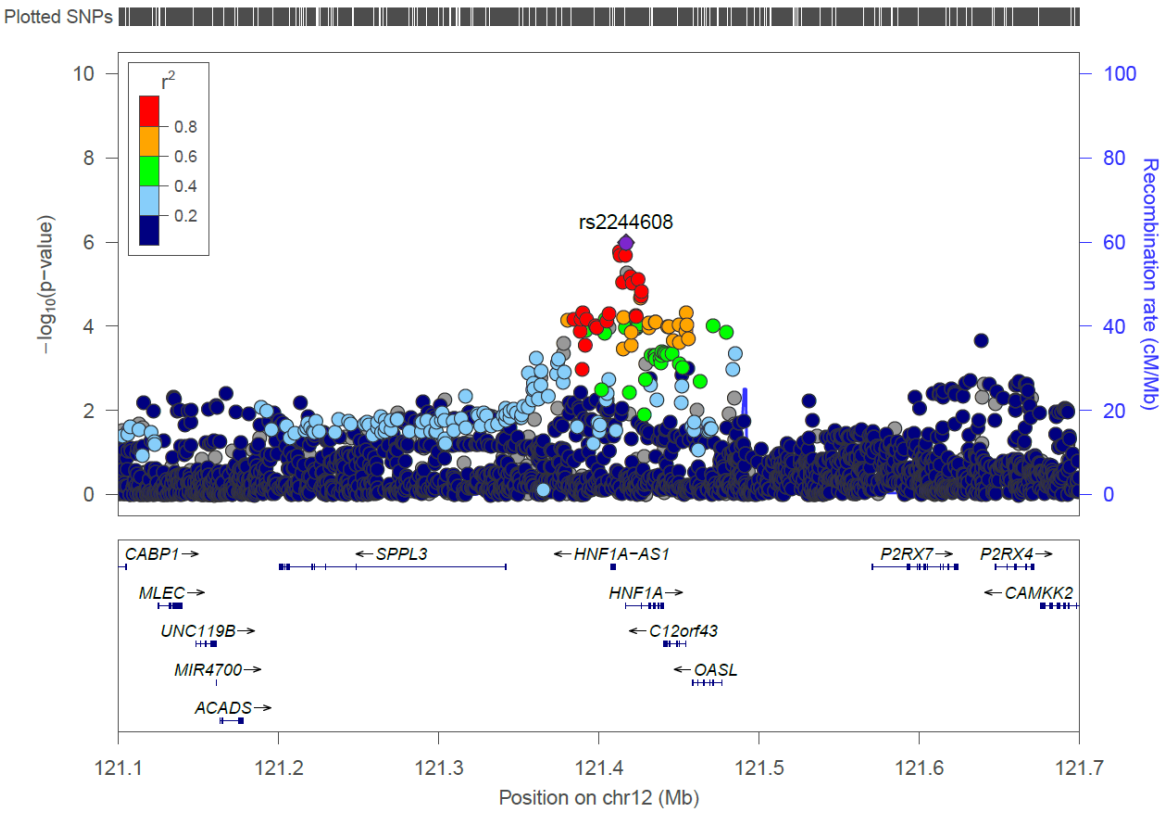
***PCNX3* (1000G)**



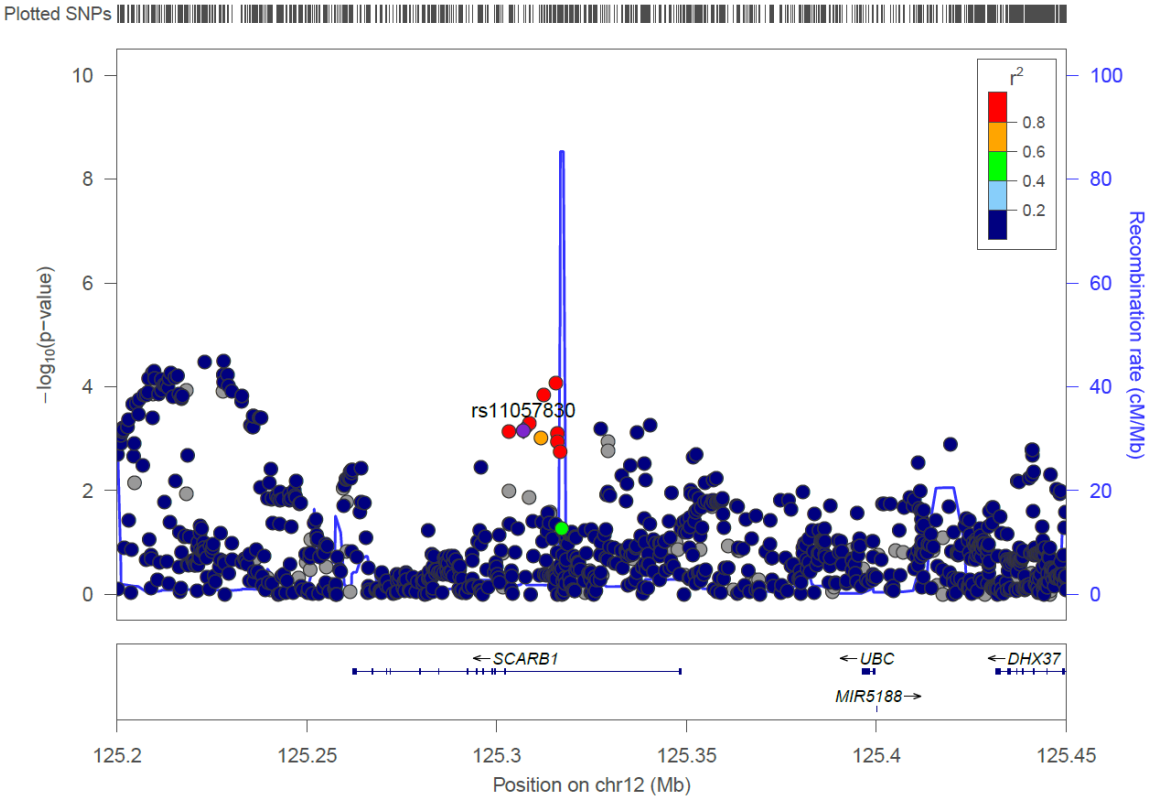
SERPINH1 (1000G)



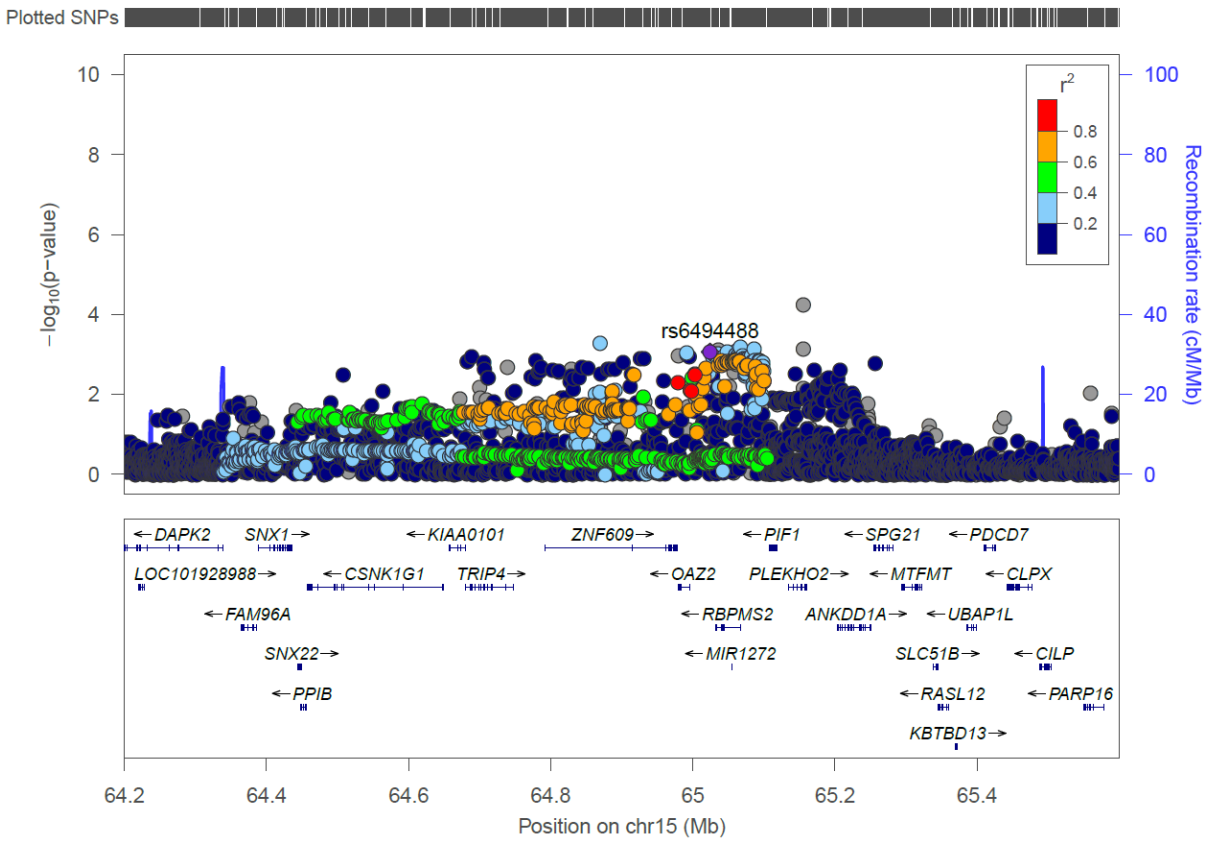
C12orf43/HNF1A (1000G)



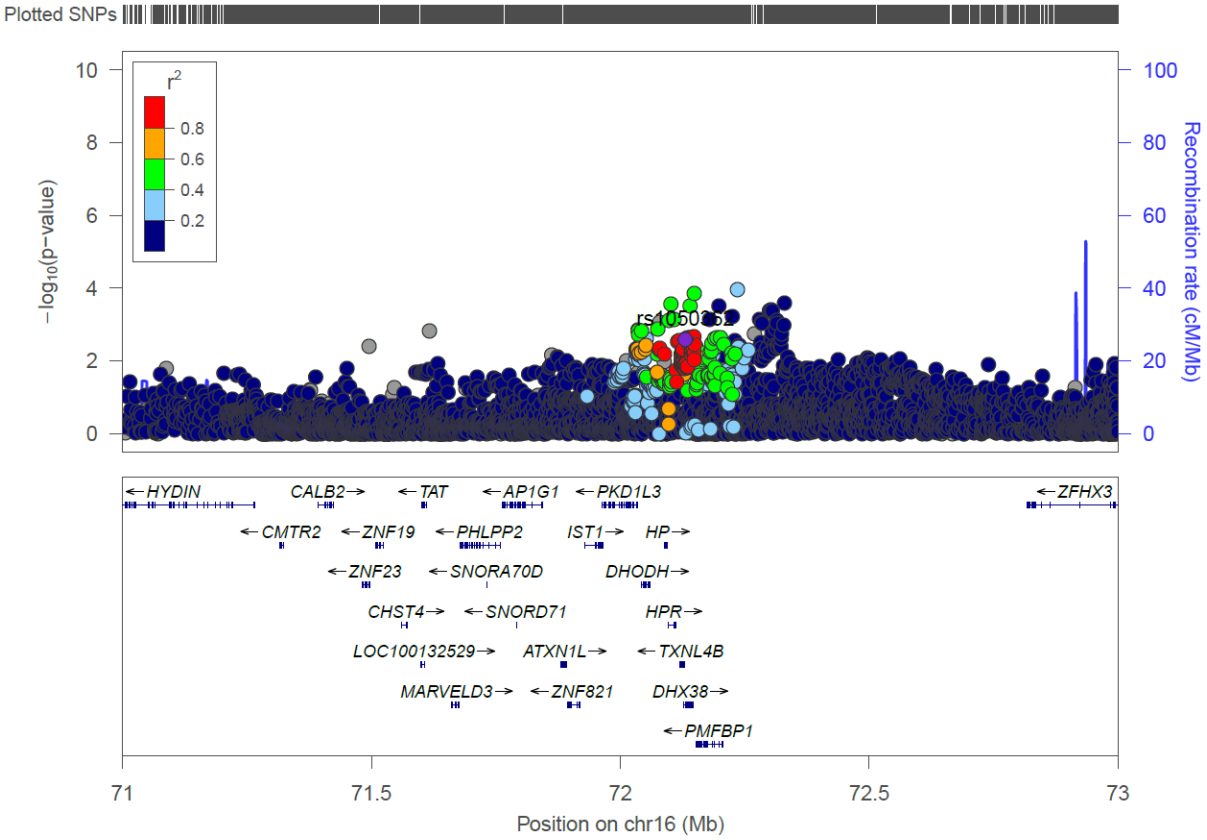
***SCARB1* (1000G)**



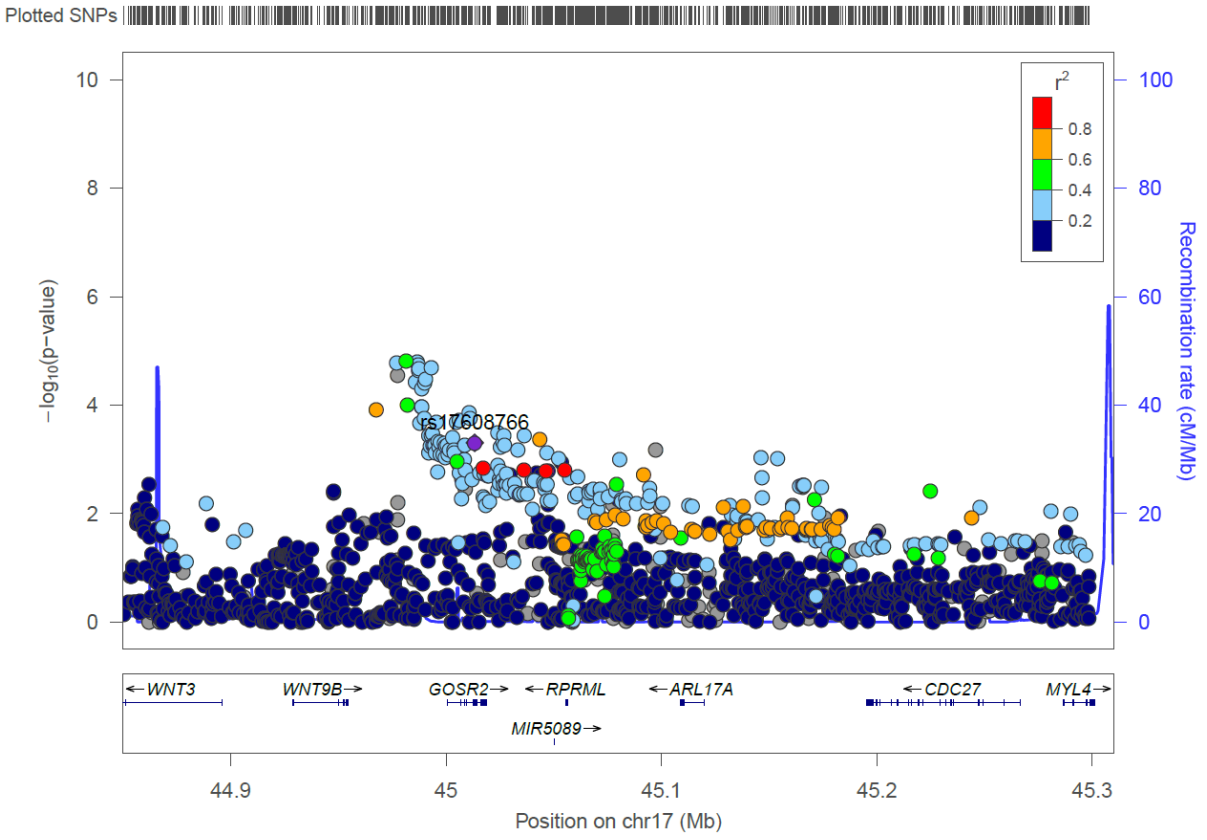
***OAZ2, RBPMS2* (1000G)**



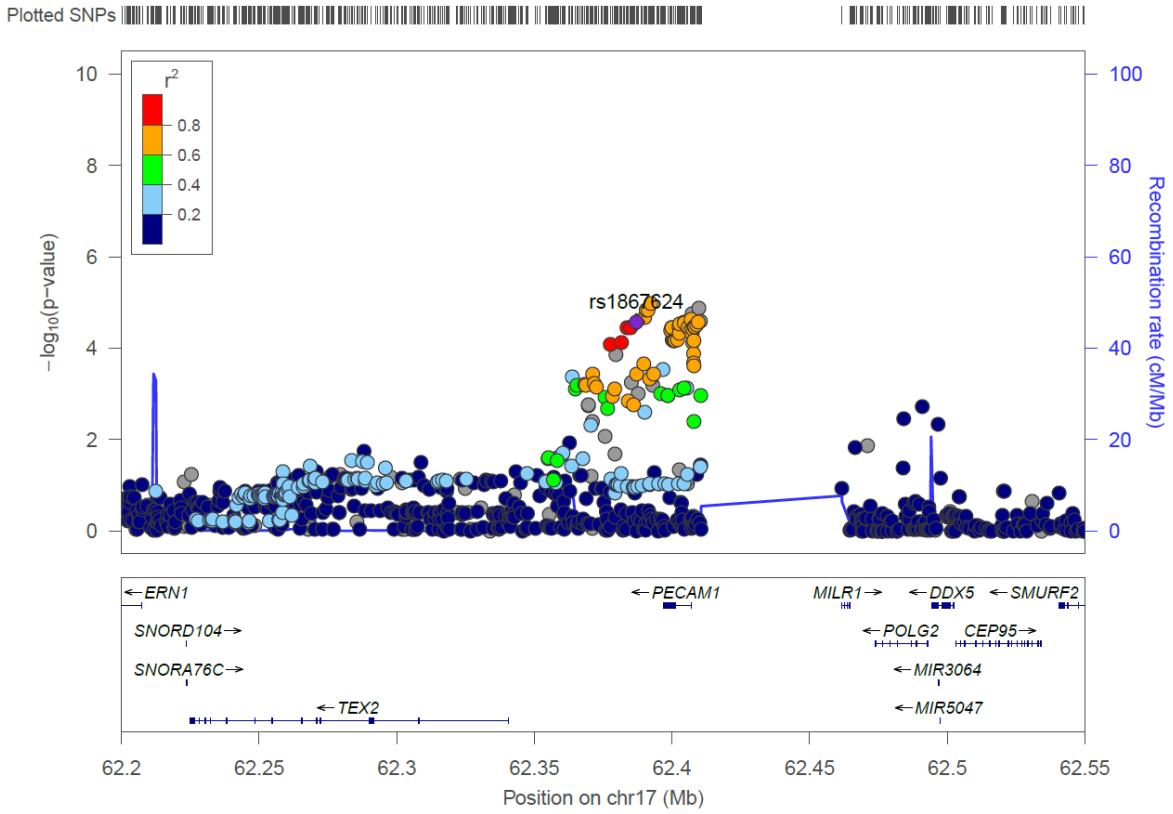
***DHX38* (1000G)**



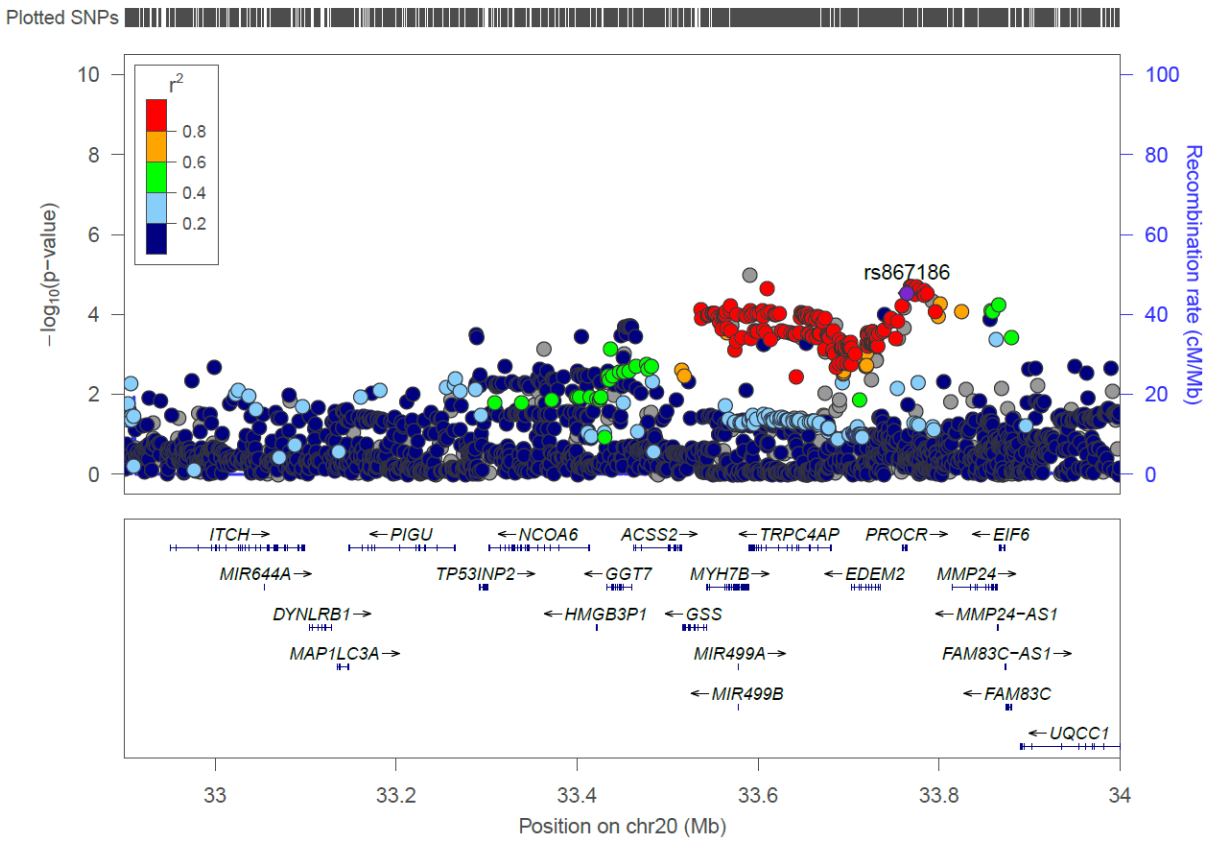
***GOSR2* (1000G)**



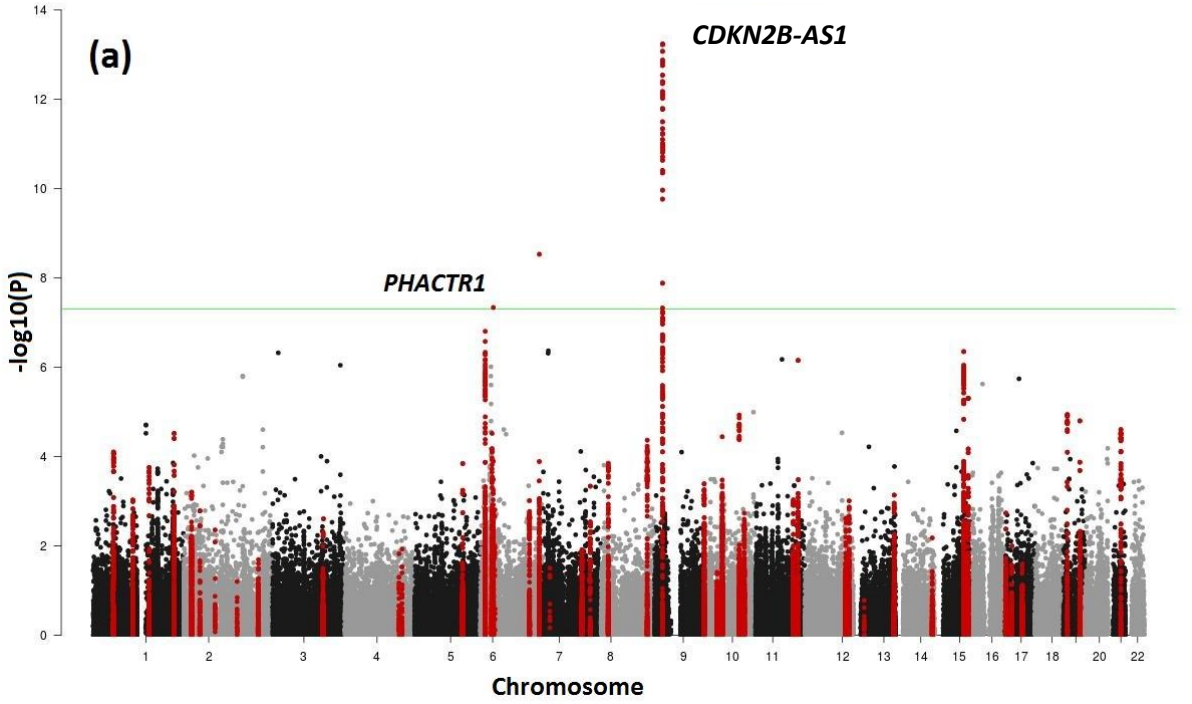
***PECAM1* (1000G)**

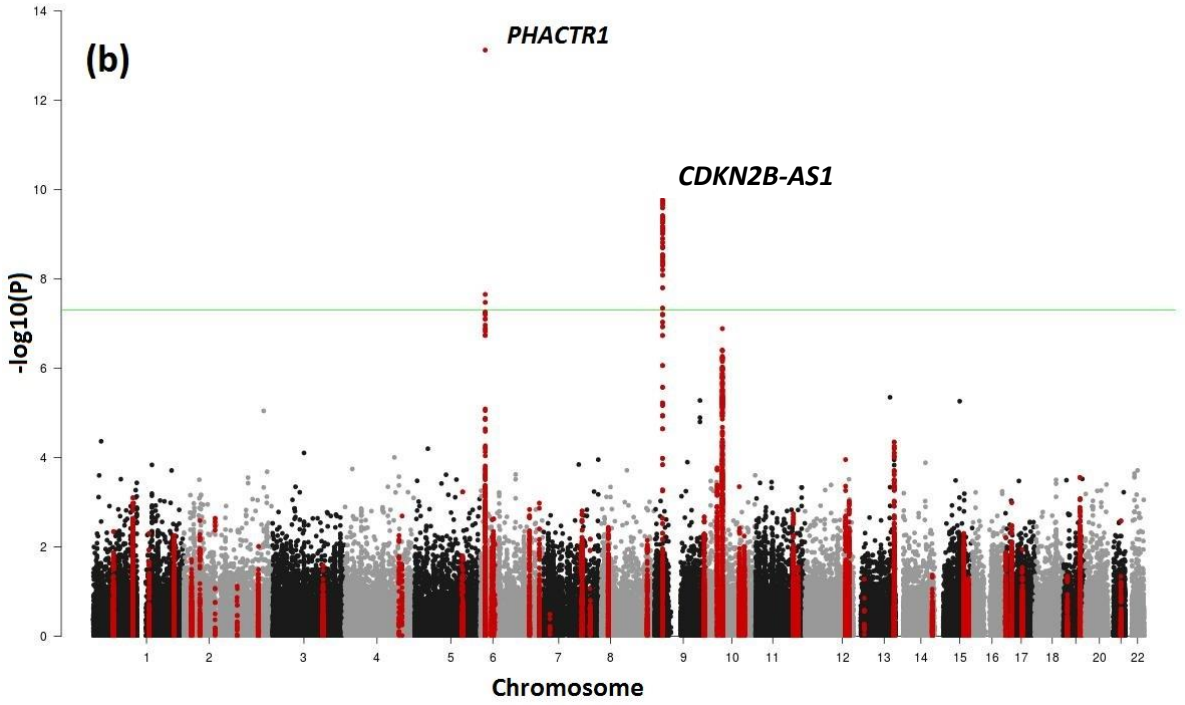


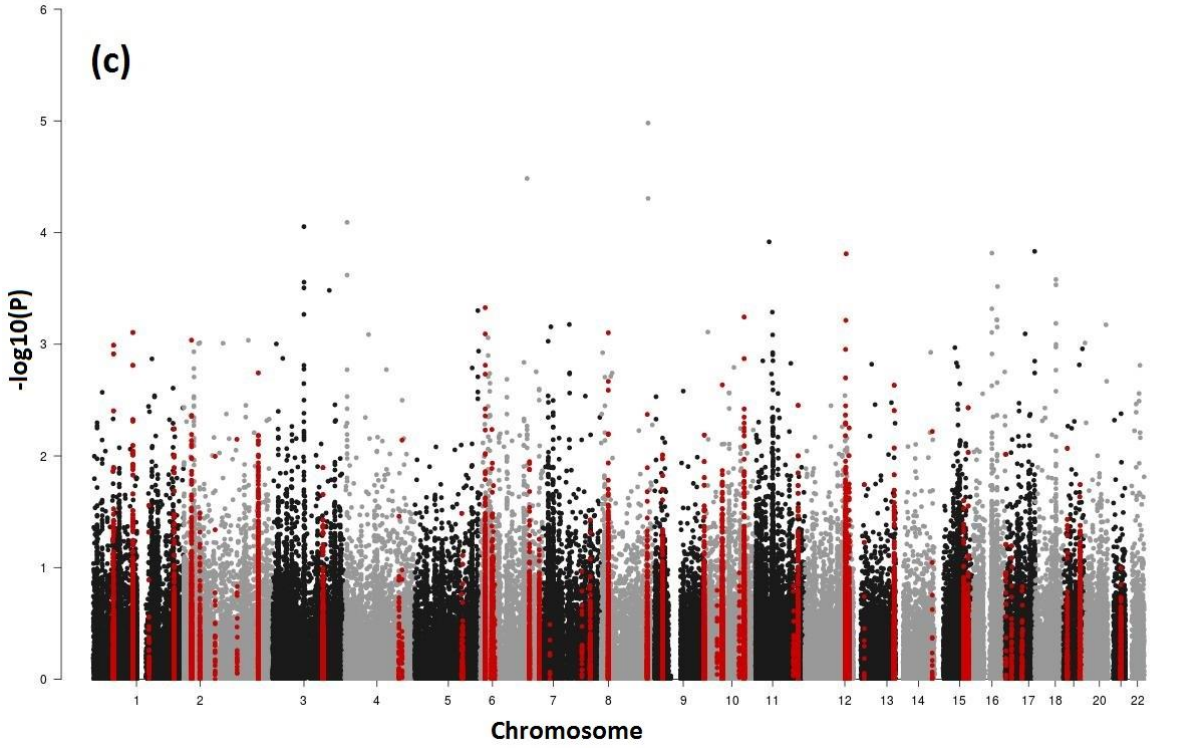
***PROCR* (1000G)**

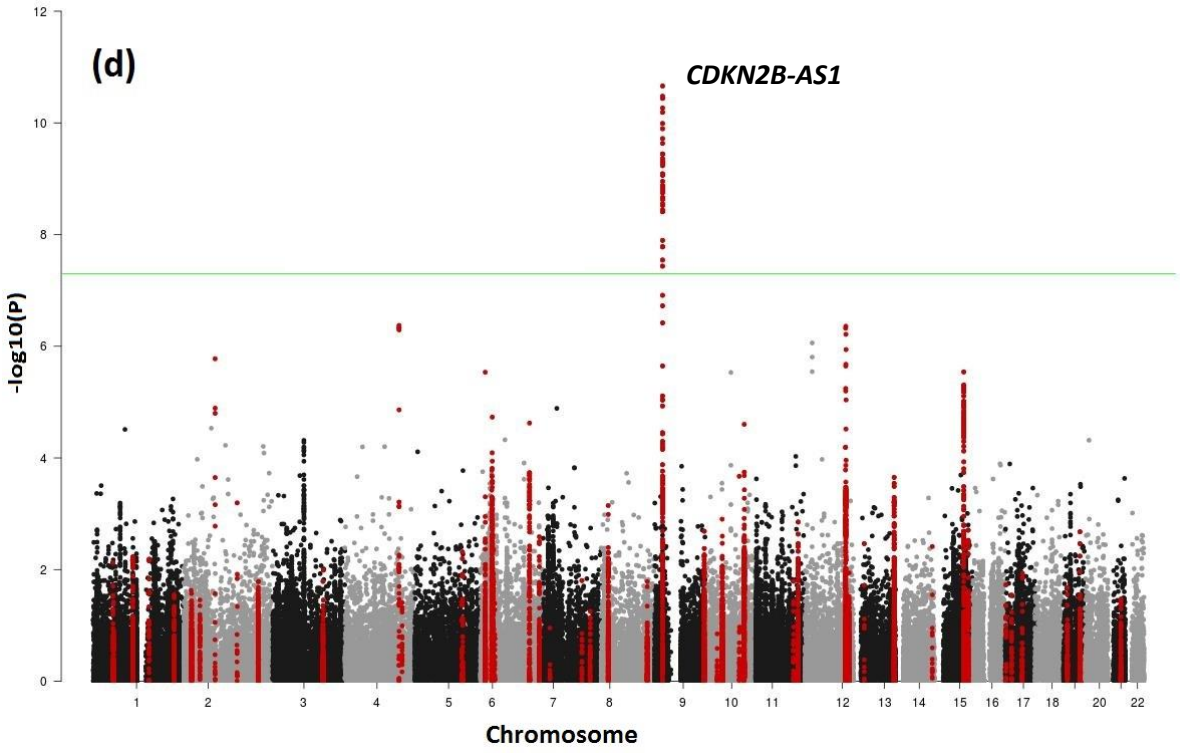


Supplementary Figure 5: Manhattan plot for the association of the Metabochip SNPs in the studies with *de novo* genotyping (a) European studies, CGPS/CIHDS, CCHS, EPIC-CVD, EPIC-Umea (b) the South Asian studies, BRAVE and PROMIS (c) the African American samples from MIGEN, WHI, and ARIC (d) the East Asian studies, TAICHI. -log(P=5x10^{-8})~7.3. Note these plots are across the whole CardioMetabochip and excluded the published CARDIoGRAMplusC4D data.

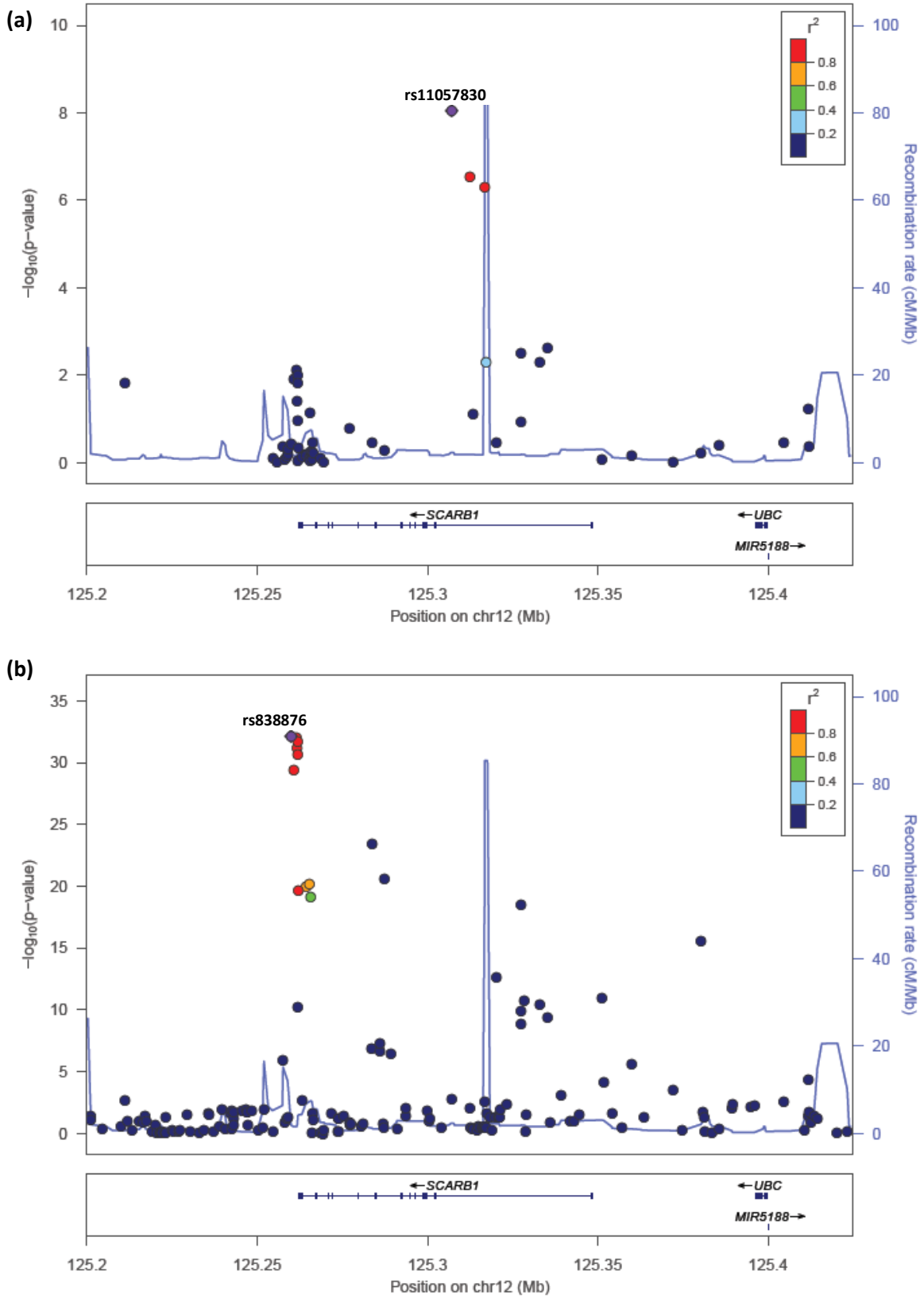



(b)
PHACTR1
CDKN2B-AS1
-log10(P)
Chromosome

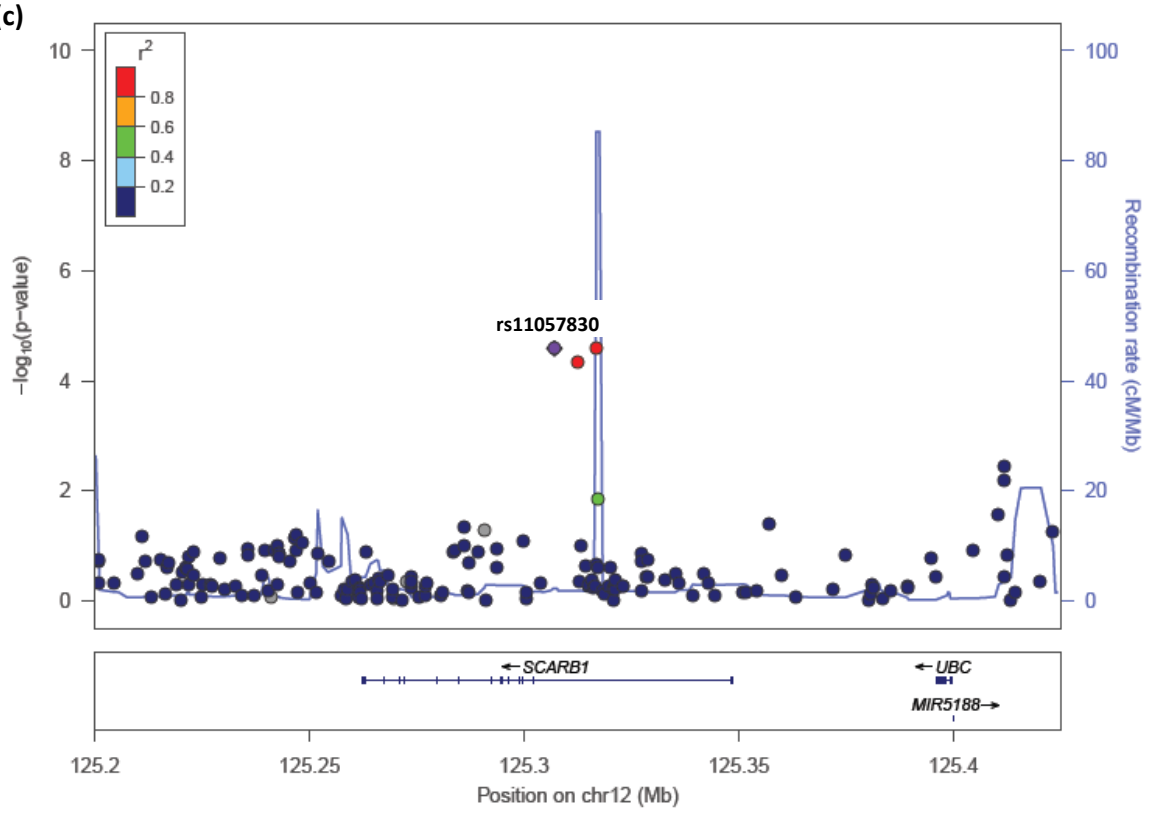

(c)
-log10(P)
Chromosome


(d)
CDKN2B-AS1
-log10(P)
Chromosome
0
2
4
6
8
10
12
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22

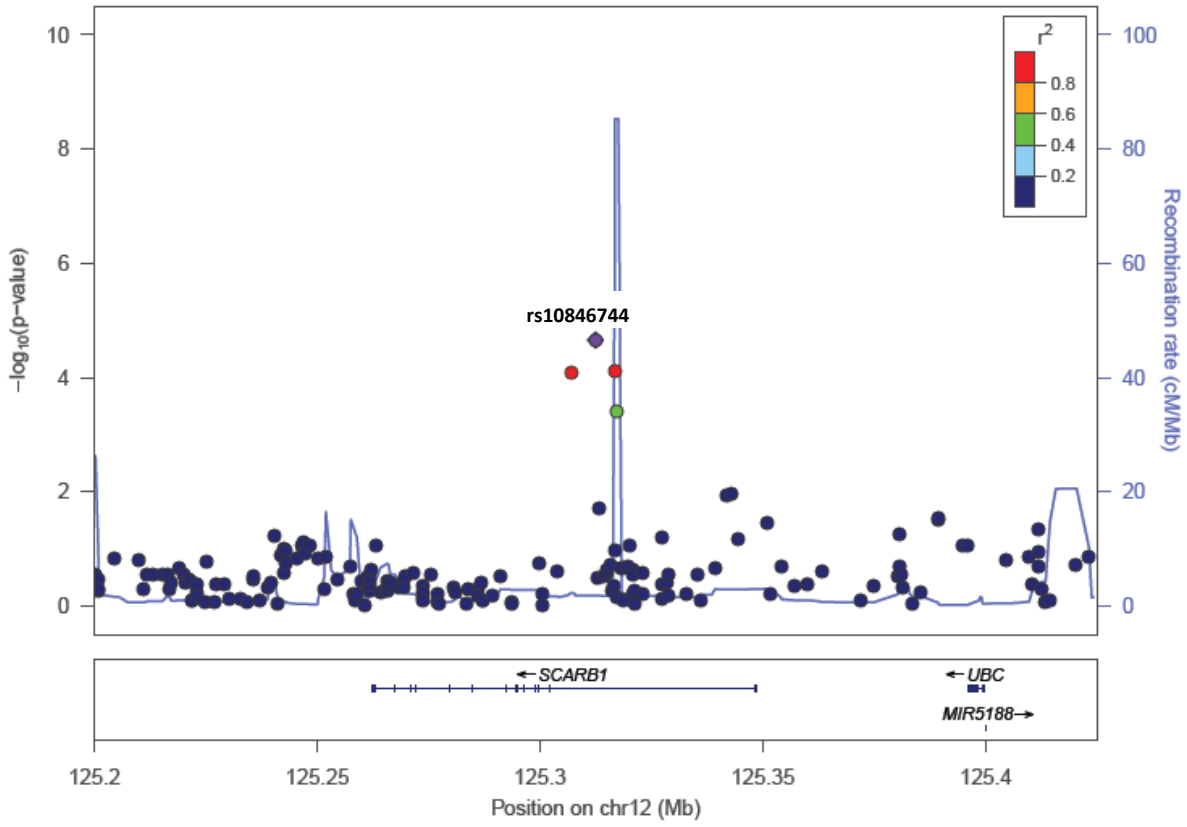
Supplementary Figure 6 *SCARB1* regional association plots with (a) CAD (b) HDL[2] (c) LDL[2] and (d) triglycerides[2]. Physical position is given for GRCh37. The r^2 information was from the 1000 Genome phase3 v5 EUR samples.



(c)

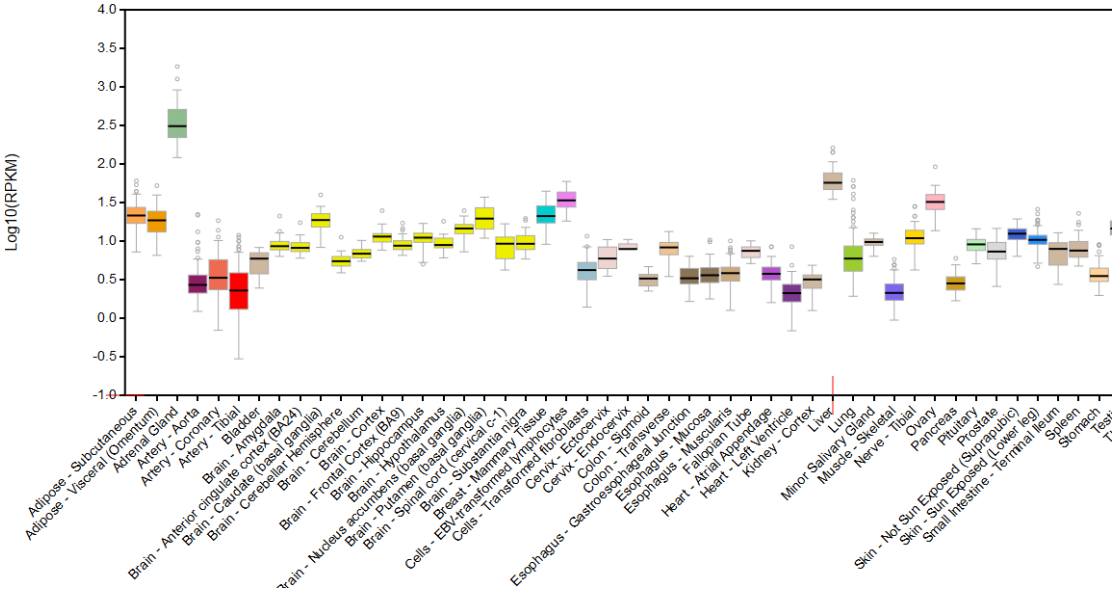

r²
0.8
0.6
0.4
0.2
rs11057830
−log10(p−value)
Recombination rate (cM/Mb)
SCARB1
UBC
MIR5188
125.2
125.25
125.3
125.35
125.4
Position on chr12 (Mb)

(d)


r²
0.8
0.6
0.4
0.2
rs10846744
−log10(p−value)
Recombination rate (cM/Mb)
SCARB1
UBC
MIR5188
125.2
125.25
125.3
125.35
125.4
Position on chr12 (Mb)

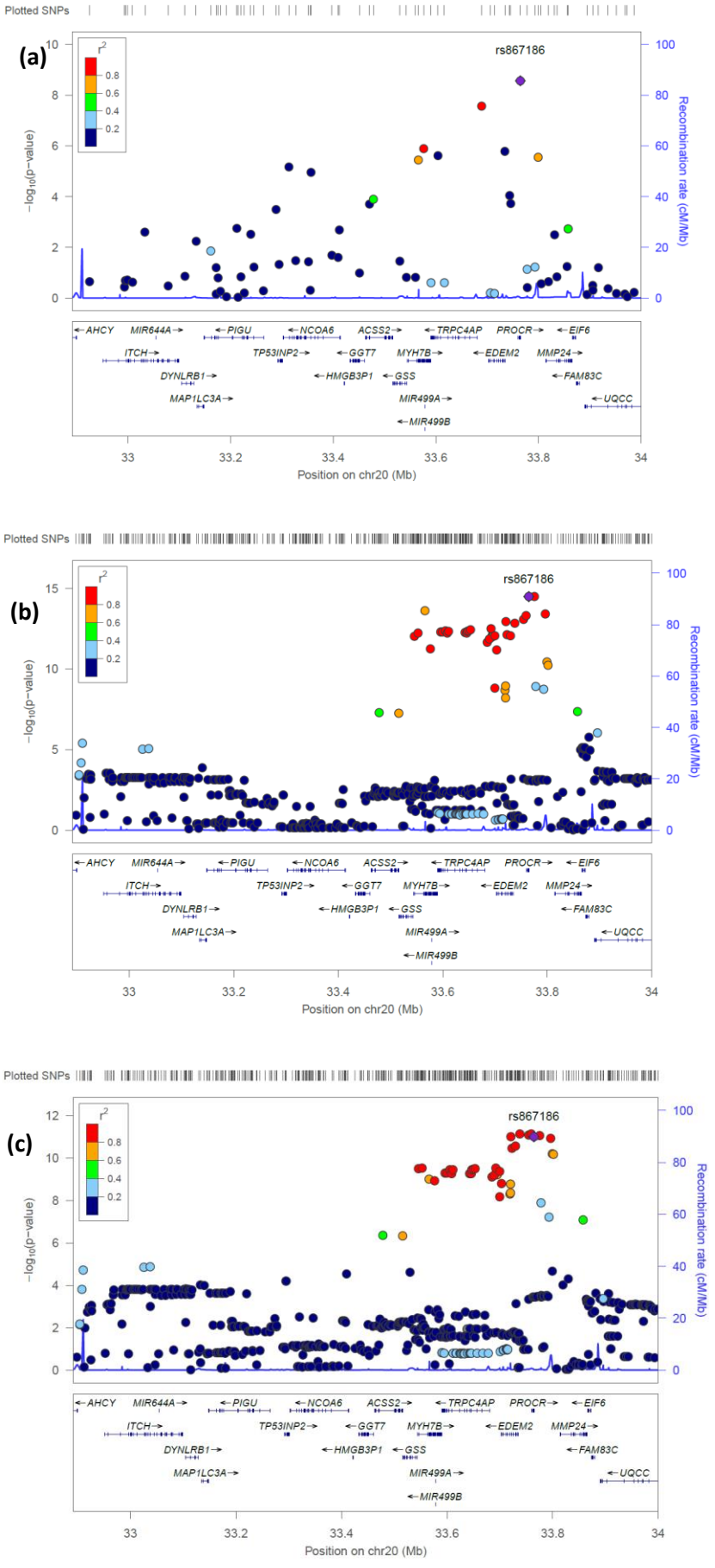
Supplementary Figure 7 Annotation of the *SCARB1* gene locus using publicly available transcriptomic and epigenomic reference data sets. **(a)** Gene expression profile of *SCARB1* in the GTEx data set (release V4; dbGaP accession phs000424.v4.p1). Among the profiled tissues, *SCARB1* is most highly expressed in adrenal gland and liver tissues. **(b)** Annotation of epigenomic features at the *SCARB1* locus (chr12:125,259,174–125,348,519; hg19) using the WashU Epigenome Browser v40.0.0 (http://epigenomegateway.wustl.edu/browser/). In the top panel, we show the two correlated variants rs11057830 and rs11057841 associated with CAD, as well as the variant rs838880 associated the HDL levels. RefSeq genes are shown at the bottom panel. A total of 23 epigenomic reference tracks (i.e. chromatin state maps) provided by the NIH Roadmap Epigenomics Project are displayed. Specifically, we show primary chromatin state maps in all available adult cell types/tissues (blood, bone, brain, fat and muscle tissues were excluded). All three highlighted genetic variants map to enhancers active in primary liver tissue.

A

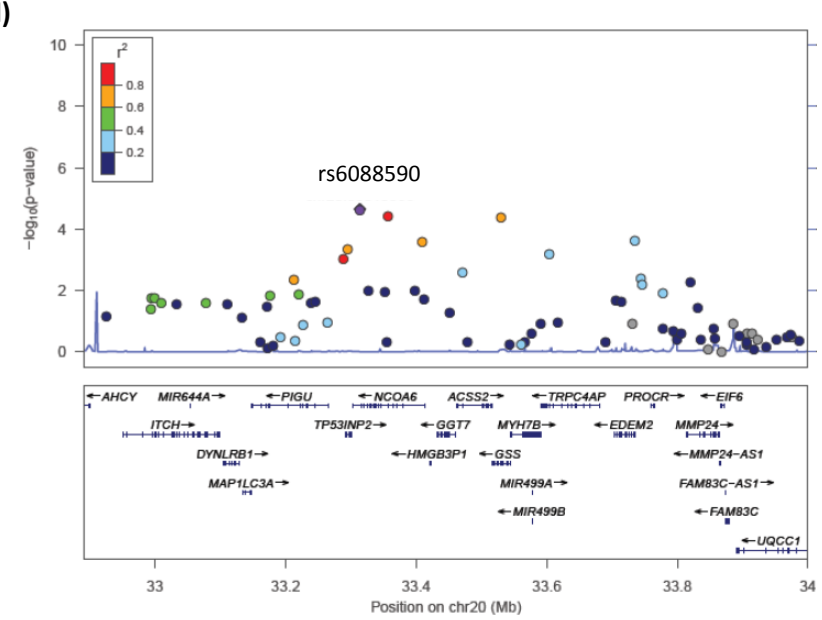


B

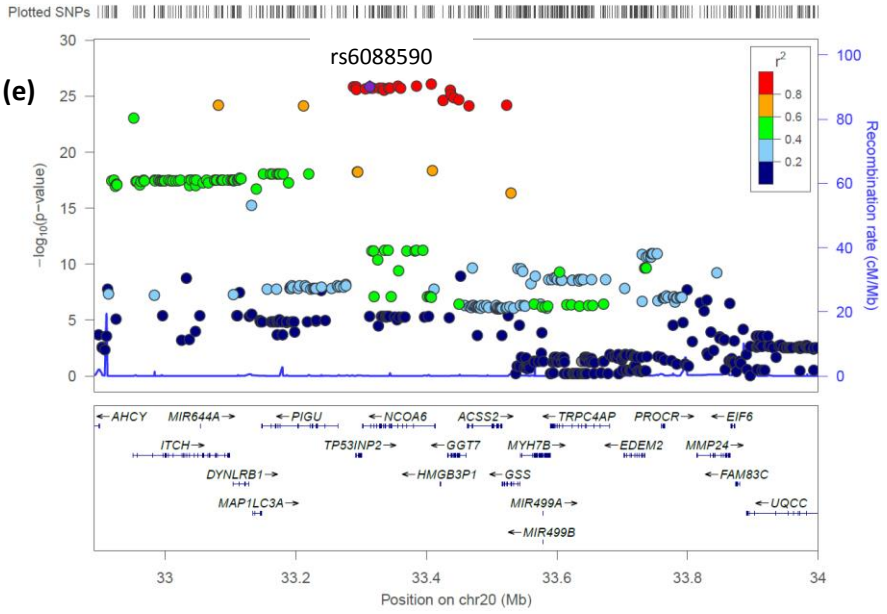
Supplementary Figure 8 Association of the PROCR gene region. (a) with CAD (b) with *PROCR* expression QTLs in subcutaneous adipose tissue from MuTHER (c) *PROCR* expression QTLs in skin tissue from MuTHER (d) CAD-association of the *PROCR* region conditional on the sentinel SNP, rs867186 (e) *GGT7* expression QTLs in subcutaneous adipose tissue from MuTHER (f) *GGT7* expression QTLs in skin tissue from MuTHER. Physical position is given for GRCh37. r^2 is calculated using 1000G EUR samples and reported relative to the sentinel CAD SNP, rs867186, in (a), (b) & (c) and to the second CAD-associated SNP, rs6088590, in (d), (e) and (f).



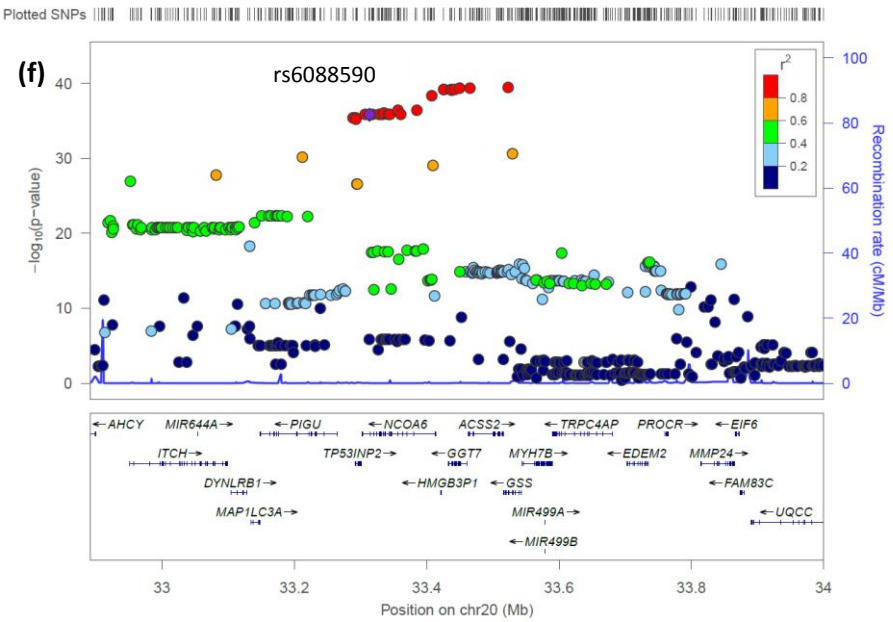
(d)


rs6088590
r²
0.8
0.6
0.4
0.2
−log10(p−value)
AHCY
MIR644A
PIGU
NCOA6
ACSS2
TRPC4AP
PROCR
EIF6
ITCH
TP53INP2
GGT7
MYH7B
EDEM2
MMP24
DYNLRB1
HMGB3P1
GSS
MMP24-AS1
MAP1LC3A
MIR499A
FAM83C-AS1
MIR499B
FAM83C
UQCC1
Position on chr20 (Mb)

(e)


Plotted SNPs
rs6088590
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
AHCY
MIR644A
PIGU
NCOA6
ACSS2
TRPC4AP
PROCR
EIF6
ITCH
TP53INP2
GGT7
MYH7B
EDEM2
MMP24
DYNLRB1
HMGB3P1
GSS
FAM83C
MAP1LC3A
MIR499A
UQCC
MIR499B
Position on chr20 (Mb)

(f)


Plotted SNPs
rs6088590
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
AHCY
MIR644A
PIGU
NCOA6
ACSS2
TRPC4AP
PROCR
EIF6
ITCH
TP53INP2
GGT7
MYH7B
EDEM2
MMP24
DYNLRB1
HMGB3P1
GSS
FAM83C
MAP1LC3A
MIR499A
UQCC
MIR499B
Position on chr20 (Mb)