nature genetics

A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease

Existing knowledge of genetic variants affecting risk of coronary artery disease (CAD) is largely based on genome-wide association study (GWAS) analysis of common SNPs. Leveraging phased haplotypes from the 1000 Genomes Project, we report a GWAS meta-analysis of ~185,000 CAD cases and controls, interrogating 6.7 million common (minor allele frequency (MAF) > 0.05) and 2.7 million low-frequency (0.005 < MAF < 0.05) variants. In addition to confirming most known CAD-associated loci, we identified ten new loci (eight additive and two recessive) that contain candidate causal genes newly implicating biological processes in vessel walls. We observed intralocus allelic heterogeneity but little evidence of low-frequency variants with larger effects and no evidence of synthetic association. Our analysis provides a comprehensive survey of the fine genetic architecture of CAD, showing that genetic susceptibility to this common disease is largely determined by common SNPs of small effect size.

CAD is the main cause of death and disability worldwide and represents an archetypal common complex disease with both genetic and environmental determinants[1,2]. Thus far, 48 genomic loci have been found to harbor common SNPs in genome-wide significant association with the disease. Previous GWAS of CAD have tested the common disease–common variant hypothesis, with meta-analyses typically based on HapMap imputation training sets or tagging SNP arrays with up to 2.5 million SNPs (85% with MAF >0.05)[3,4]. The 1000 Genomes Project[5] has considerably expanded the coverage of human genetic variation, especially for lower-frequency variants and insertion-deletions (indels). We assembled 60,801 cases and 123,504 controls from 48 studies for a GWAS meta-analysis of CAD; 34,997 (57.5%) of the cases and 49,512 (40.1%) of the controls had been previously included in our Metabochip-based CAD meta-analysis (**Supplementary Fig. 1**) (ref. 3). Imputation was based on the 1000 Genomes Project phase 1 v3 training set with 38 million variants, of which over half are low frequency (MAF < 0.005) and one-fifth are common (MAF > 0.05). The majority (77%) of the participants were of European ancestry; 13% and 6% were of South Asian (India and Pakistan) and East Asian (China and Korea) ancestry, respectively, with smaller samples of Hispanic and African Americans (**Supplementary Table 1**). Case status was defined by an inclusive CAD diagnosis (for example, myocardial infarction, acute coronary syndrome, chronic stable angina or coronary stenosis of >50%). After selecting variants that met the allele frequency (MAF > 0.005) and imputation quality control criteria in at least 29 (>60%) of the studies, 8.6 million SNPs and 836,000 (9%) indels were included in the meta-analysis (**Fig. 1**); of these variants, 2.7 million (29%) were low frequency (0.005 < MAF < 0.05).

RESULTS

Scanning for additive associations

The results of an additive genetic model meta-analysis are summarized in Manhattan plots (**Fig. 2** and **Supplementary Fig. 2**). In total, 2,213 variants (7.6% indels) showed significant associations ($P < 5 \times 10^{-8}$) with CAD with a low false discovery rate (FDR q value < 2.1×10^{-4}). When these 2,213 variants were grouped into loci, 8 represented regions not previously reported as being associated with CAD at genome-wide levels of significance (**Fig. 2** and **Table 1**). Of the 48 loci previously reported at genome-wide levels of significance, 47 showed nominally significant associations (**Supplementary Table 2**). The exception was rs6903956, the lead SNP for the *ADTRP-C6orf105* locus detected in Han Chinese[6], which previously showed no association in the Metabochip meta-analysis of Europeans and South Asians[3]. Thirty-six previously reported loci showed genome-wide significance (**Supplementary Table 2**). Monte Carlo simulations, guided by published effect sizes, suggest that our study was powered to detect 34 of the previously reported loci (95% confidence interval (CI) = 31–41 loci) at genome-wide significance. Hence, our findings are fully consistent with the previously identified CAD-associated loci.

The majority of the loci showing GWAS significance in the present analysis were well imputed (82% with imputation quality >0.9) (**Fig. 3a**) and had small effect sizes (odds ratio (OR) < 1.25) (**Fig. 3b**). An exception was the lead SNP in the newly associated chromosome 7q36.1 (*NOS3*) locus, rs3918226, which was only moderately well imputed (quality of 0.78), but the validity of this association was supported by existing genotype data, as rs3918226 was present on the HumanCVD BeadChip for which data were available for some of the cohorts used in the present analysis, thereby allowing directly measured genotypes to be compared with imputed genotypes (**Supplementary Table 3**) (ref. 7). Three additional lower-frequency and moderately well-imputed SNPs in *LPA* and *APOE* (**Fig. 3a**), which were not previously reported in CAD GWAS[3,4], also showed strong associations (*LPA*: rs10455872, $P = 5.7 \times 10^{-39}$ and rs3798220, $P = 4.7 \times 10^{-9}$; *APOE*: rs7412, $P = 8.2 \times 10^{-11}$). The *LPA* SNPs have previously been shown to be strongly associated with CAD in candidate gene studies based on experimental genotype data[7,8].

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

Received 13 January; accepted 14 August; published online 7 September 2015; corrected online 14 September 2015 (details online); doi:10.1038/ng.3396

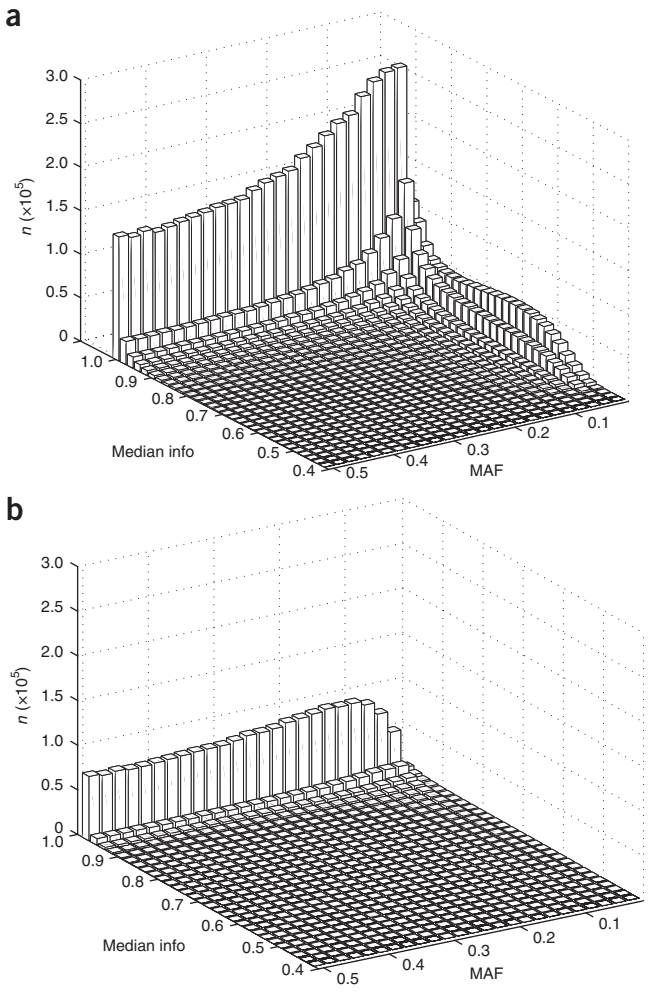

Figure 1 Comparing the 1000 Genomes Project and HapMap imputation training sets. Spectra of MAFs and median imputation quality (median info) scores showing the number (*n*) of variants in each bin. (**a**) The distribution for the 9.4 million 1000 Genomes Project phase 1 v3 variants. (**b**) The distribution for 2.5 million HapMap 2 SNPs. Imputation quality was calculated as the median of the respective values in up to 48 contributing studies; the imputation quality for genotyped variants was set equal to 1.0. The 1000 Genomes Project training set includes more low-frequency variants, many of which have imputation qualities >0.9.

The minor allele of SNP rs7412 encodes the ε2 allele of *APOE*, and it has been well documented that carriers of the ε2 allele have lower cholesterol levels; significant protection from CAD by this allele was confirmed in a large meta-analysis[9] and the Metabochip study ($P = 0.0009$) (ref. 3). However, rs7412 is not present on most commercially available genome-wide genotyping arrays and cannot be imputed using HapMap reference panels, highlighting the value of the expanded coverage of the 1000 Genomes Project reference panels. Finally, SNP rs11591147 in *PCSK9*, which encodes the low-frequency (MAF = 0.01) p.Arg46Leu substitution that has been associated with low LDL (low-density lipoprotein) cholesterol levels and cardioprotection[10–13], was imperfectly imputed (imputation quality = 0.61). Nonetheless, these data provide the strongest evidence yet for a protective effect of this variant in CAD ($P = 7.5 \times 10^{-6}$).

Scanning for non-additive associations

Few GWAS of CAD have systematically scanned for associations that include dominance effects, and few truly recessive loci have been reported[14,15]. We used a recessive inheritance model to search for susceptibility effects conferred by homozygosity for the minor (less frequent) allele. Two new recessive susceptibility loci were identified with MAF = 0.09 and 0.36 and genotypic OR = 0.67 and 1.12, respectively (**Fig. 2** and **Table 1**); these loci showed very little evidence of association under an additive model (**Table 1**). A supplementary analysis applying a dominant model identified multiple strong associations with variants, all of which overlapped with loci identified in the analysis applying an additive model (**Supplementary Table 4**).

Myocardial infarction subphenotype analysis

Subgroup analysis in cases with a reported history of myocardial infarction (~70% of the total number of cases) did not identify any additional associations reaching genome-wide significance. The association results for the myocardial infarction subphenotype for the 48 previously known CAD-associated loci and the 8 new additive CAD-associated loci discovered in this study are shown in **Supplementary Table 5**. The odds ratios for the lead SNPs at 56 loci for the broader CAD phenotype (full cohort) and the myocardial infarction subphenotype are compared in **Supplementary Figure 3**. Although, as expected, the odds ratios were very similar for most of the loci, the odds ratios for the *ABO* and *HDAC9* loci were sufficiently distinct in the two cohorts for their 95% confidence intervals to lie away from the line of equality, suggesting that the *ABO* locus preferentially associates with myocardial infarction and the *HDAC9* locus preferentially associates with stable coronary disease but not myocardial infarction *per se*.

FDR and heritability analysis

We performed a joint association analysis to search for evidence of synthetic associations[16], where multiple low-frequency susceptibility variants at a locus might be in LD with a common variant discovered as the lead variant in a GWAS, and to compile an FDR-defined list of informative variants for annotation and heritability analysis[3]. Variants that showed suggestive additive association ($P < 5 \times 10^{-5}$) were assigned to 214 putative susceptibility loci of 2 cM centered on each lead variant, and all variants in these loci were examined; consequently, the search space for the joint analysis included 1,399,533 variants. Using GCTA software[17] to perform an approximate joint association analysis (Online Methods), we identified 202 FDR variants (q value < 0.05) in 129 loci (**Supplementary Table 6**) with multiple (2–14) tightly linked variants, corresponding to 57% of the putative CAD susceptibility loci. The 202 FDR variants were mostly common (median MAF = 0.22) and well imputed (median imputation quality = 0.97). Ninety-five variants (explaining 13.3 ± 0.4% of CAD heritability) mapped to 44 significant loci from GWAS, and 93 variants (explaining 12.9 ± 0.4% of CAD heritability) mapped to loci that included a previously reported significant variant from GWAS analysis. One hundred nine variants (explaining a further 9.3 ± 0.3% of CAD heritability) mapped to other loci. Fifteen low-frequency (MAF < 0.05) variants explained only 2.1 ± 0.2% of CAD heritability, indicating that our study was ~90% powered to detect OR >1.5 with low-frequency variants (**Supplementary Table 7**).

Common variants showing typical GWAS signals might be coupled with one or more low-frequency variants with relatively large effects[16]. We found no evidence for such synthetic associations in the joint association analysis; that is, all low-frequency variants were either a lead variant or were jointly associated (q value < 0.05) with a common variant. Twenty of the 202 FDR variants (9.9%) were indels (4–14 bp in size) as compared to 8.8% of all the variants in the meta-analysis ($P = 0.60$). Low-frequency variants (MAF < 0.05)

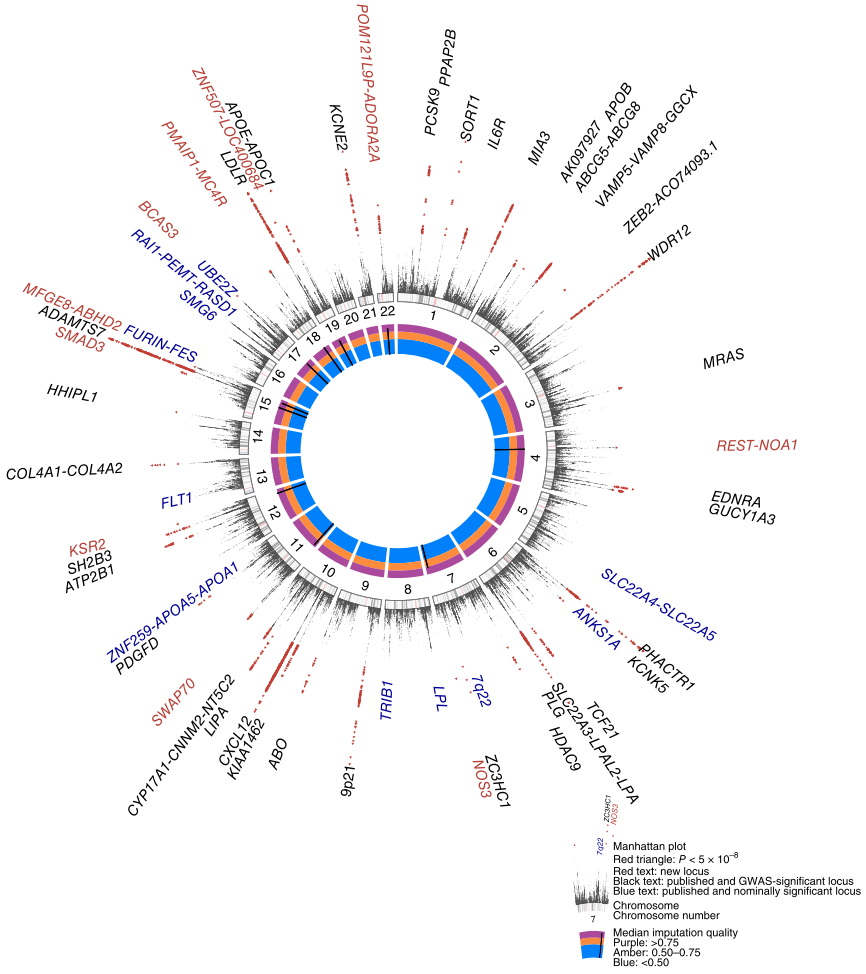

Figure 2 A circular Manhattan plot summarizing the 1000 Genomes Project CAD association results. The meta-analysis statistics were adjusted for overdispersion (before applying double genomic control, $\lambda = 1.18$); overdispersion is predicted to be a regular feature in GWAS under a polygenic inheritance model[60]. The association statistics were capped at $P = 1 \times 10^{-20}$. Genome-wide significant variants ($P < 5 \times 10^{-8}$) are indicated by red triangles. New CAD-associated loci are indicated by red text (**Table 1**). Previously reported loci showing genome-wide significant association are indicated by black text, and those showing nominal significance ($P < 0.05$) in our meta-analysis are indicated by blue text (**Supplementary Table 2**). The inner track shows the imputation quality scores of the lead variants in the new loci. The middle track shows numbered chromosome ideograms with centromeres represented by pink bars.

were strikingly under-represented (6.9% versus 29.0%; $P = 4.9 \times 10^{-12}$), which may reflect on the statistical power to detect the modest effects associated with these variants.

Annotation and ENCODE analysis

Functional annotations were assigned to the 9.4 million variants studied in the CAD additive meta-analysis using ANNOVAR software[18] (**Supplementary Table 8**). The 202 FDR variants were depleted in intergenic regions ($P = 2.5 \times 10^{-7}$) and enriched in introns ($P = 0.00035$). Variants were also assigned to three sets of ENCODE (Encyclopedia of DNA Elements) features, namely histone/chromatin modifications (HMs), DNase I–hypersensitive sites (DHSs) and transcription factor binding sites (TFBSs) (**Supplementary Table 9**). The FDR variants showed independent enrichment across 11 cell types for the HM ($P = 2.8 \times 10^{-6}$) and DHS ($P = 0.0003$) ENCODE feature sets and with genic annotation status ($P = 0.0013$) (**Supplementary Tables 10** and **11**). These associations were also evident in three cell types selected for maximal CAD relevance, with a 2.6-fold enrichment for DHSs, a 2.2-fold enrichment for HMs and a 1.6-fold enrichment for genic status (**Supplementary Tables 12** and **13**). These findings suggest that the 202 FDR variants are enriched for functional variants with potential relevance to CAD pathogenesis.

Post-hoc power calculations

Of the 9.4 million variants analyzed, 8.2 million (87%) were highly powered (>90%) to detect an OR ≥1.3 (**Supplementary Table 7**). The number of variants with power of ≥90% to detect associations varied systematically with allele frequency and imputation quality (results for OR = 1.3 shown in **Supplementary Fig. 4**); 1.5 million of the 2.7 million (55%) low-frequency variants (0.005 < MAF < 0.05) in the meta-analysis were adequately powered to detect an OR ≥1.3, as most of these variants were accurately imputed (median imputation quality = 0.94, interquartile range = 0.88–0.98). Of the more common variants (MAF > 0.05), almost all (99.8%) were highly powered to detect an OR ≥1.3. However, in terms of total coverage of low-frequency variation, only 15.3% of the 9.3 million low-frequency variants (0.005 < MAF < 0.05) in the 1000 Genomes Project phase 1 v3 training set met the allele frequency and imputation quality entry criteria in the 60% of the studies required for inclusion in the meta-analysis and were predicted to be adequately powered to detect significant associations; 100% of these variants were highly powered (>90%) to detect an OR ≥3.15.

Interrogation of ten newly identified additive and recessive loci

We examined whether there were any expression quantitative trait loci (eQTLs), associations with known cardiovascular risk factors or prior evidence of the involvement of genes with atherosclerotic processes in each of the newly identified loci to define putative mechanisms by which the loci might affect risk of CAD.

At the chromosome 4q12 (*REST-NOA1*) locus, the lead SNP rs17087335 lies within an intron of the *NOA1* gene (nitric oxide–associated 1); 23 SNPs in LD ($r^2 > 0.8$) showed CAD associations ($P < 1 \times 10^{-6}$) across the *NOA1* and *REST* (repressor element-1 silencing transcription factor) genes (**Fig. 4a**). *NOA1* encodes a GTP-binding protein involved in the regulation of mitochondrial respiration and apoptosis[19]. *REST* encodes a transcription factor that suppresses the expression of voltage-dependent sodium and potassium channels[20]; it has been shown to maintain vascular smooth muscle cells (VSMCs) in a quiescent, non-proliferative state and is itself downregulated in neointimal hyperplasia[21]. SNP rs17087335 showed a *cis*-eQTL signal for *REST* in lung[22] (**Supplementary Table 14**).

At the chromosome 7q36.1 (*NOS3*) locus, the lead SNP rs3918226 (MAF = 0.07) lies in the first intron of *NOS3* (nitric oxide synthase 3) (**Fig. 4b**). This SNP was tentatively associated with CAD (OR = 1.14, $P = 1.4 \times 10^{-4}$) in a candidate gene meta-analysis based on 15,600

Table 1 Ten new CAD-associated loci

Lead variant	Locus name	Chr.	A1/A2	Effect allele (A1) freq.	Imputation quality	I^2	Heterogeneity *P*	*n* studies[a]	Association model: Additive OR (95% CI)	Additive *P*	Recessive OR (95% CI)	Recessive *P*
rs17087335	*REST-NOA1*	4	T/G	0.21	0.99	0.20	0.11	48	**1.06 (1.04–1.09)**	**4.60×10^{-8}**	1.11 (1.05–1.17)	3.30×10^{-4}
rs3918226	*NOS3*	7	T/C	0.06	0.78	0.15	0.19	45	**1.14 (1.09–1.19)**	**1.70×10^{-9}**	1.26 (0.99–1.60)	5.96×10^{-2}
rs10840293	*SWAP70*	11	A/G	0.55	0.94	0.17	0.16	47	**1.06 (1.04–1.08)**	**1.30×10^{-8}**	1.05 (1.02–1.09)	1.51×10^{-3}
rs56062135	*SMAD3*	15	C/T	0.79	0.98	0.00	0.67	46	**1.07 (1.05–1.10)**	**4.50×10^{-9}**	1.17 (1.10–1.25)	8.88×10^{-7}
rs8042271	*MFGE8-ABHD2*	15	G/A	0.9	0.93	0.16	0.19	46	**1.10 (1.06–1.14)**	**3.70×10^{-8}**	1.25 (1.13–1.37)	7.27×10^{-6}
rs7212798	*BCAS3*	17	C/T	0.15	0.95	0.14	0.21	48	**1.08 (1.05–1.11)**	**1.90×10^{-8}**	1.17 (1.07–1.28)	6.12×10^{-4}
rs663129	*PMAIP1-MC4R*	18	A/G	0.26	1.00	0.00	0.6	47	**1.06 (1.04–1.08)**	**3.20×10^{-8}**	1.11 (1.06–1.17)	7.15×10^{-6}
rs180803	*POM121L9P-ADORA2A*	22	G/T	0.97	0.86	0.00	0.67	41	**1.20 (1.13–1.27)**	**1.60×10^{-10}**	NA	NA
rs11830157	*KSR2*	12	G/T	0.36	0.99	0.14	0.22	42	1.04 (1.02–1.06)	3.90×10^{-4}	**1.12 (1.08–1.16)**	**2.12×10^{-9}**
rs12976411	*ZNF507-LOC400684*	19	T/A	0.09	0.93	0.50	5.09×10^{-4}	34	0.95 (0.92–0.99)	9.10×10^{-3}	**0.67 (0.60–0.74)**	**1.18×10^{-14}**

Association results are presented for two inheritance models; results from the discovery association model are shown in bold. *P* values were adjusted for overdispersion following meta-analysis. Heterogeneity *P* values are for the respective discovery association model. Chr., chromosome; A1, effect allele; A2, non-effect allele; freq., frequency; I^2, heterogeneity inconsistency index; OR, odds ratio; CI, confidence interval; NA, not available owing to insufficient numbers (<60%) of studies having reliable results.
[a]The number of studies that participated in the discovery result, where up to 48 studies participated in the additive model meta-analysis and up to 43 studies participated in the recessive model meta-analysis.

cases and 35,000 controls genotyped with the HumanCVD BeadChip[7] and was firmly associated with essential hypertension (OR = 1.34, $P = 1.0 \times 10^{-14}$) (ref. 23). *NOS3* is involved in the production of nitric oxide (NO), a potent vascular smooth muscle relaxant, and is a well-studied candidate gene for CAD. Indeed, the genes encoding the components of the NO receptor (soluble guanylyl cyclase) display both linkage and genome-wide association with CAD[3,24]. There are several overlapping ENCODE features in intron 1 of *NOS3*, suggesting a functional role for rs3918226. However, there are 30 genes neighboring *NOS3* within a 2-cM window centered on this variant, and the current data do not allow the candidacy of these genes to be excluded. A nonsynonymous SNP, rs1799983, in *NOS3* previously associated with cardiovascular phenotypes[25] is in weak LD with rs3918226 but did not achieve significance in the additive or joint association analysis.

At the chromosome 11p15.4 (*SWAP70*) locus, SNP rs10840293 is intronic to *SWAP70* (switch-associated protein-70) (**Fig. 4c**). SWAP-70 is a signaling molecule involved in the regulation of filamentous actin networks[26] in cell migration and adhesion. SNP rs10840293 and other SNPs in strong LD are *cis* eQTLs for *SWAP70* in naive and challenged monocytes[27], with SNP rs93138 showing strong association with CAD ($P = 5.5 \times 10^{-8}$) and being a *cis* eQTL for *SWAP70* in naive and challenged monocytes[28], fat[29], skin[29] and lung[22] (**Supplementary Table 14**); three of the linked SNPs (rs93138, rs173396 and rs472109) are intronic and lie within ENCODE regulatory functional elements. Although this CAD-associated locus includes 33 genes, the eQTL and ENCODE data implicate *SWAP70* as a plausible causal gene and suggest putative causal SNPs.

At the chromosome 15q22.33 (*SMAD3*) locus, the lead SNP rs56062135 is intronic to *SMAD3* and the CAD association is tightly localized between two recombination hot spots (**Fig. 4d**). Mice lacking *Smad3*, a major downstream mediator of transforming growth factor (TGF)-β signaling, show enhanced neointimal hyperplasia with decreased matrix deposition in response to vascular injury[30]. *SMAD3* was tentatively associated with CAD in an earlier GWAS[31], although the lead SNP (rs17228212) in that association is in linkage

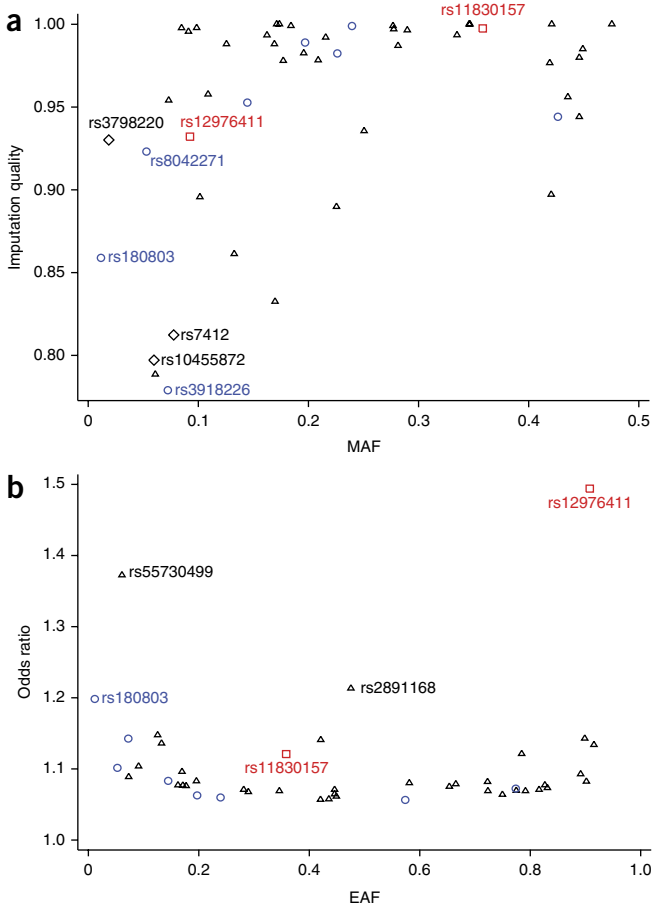

Figure 3 The imputation quality and effect size of lead variants at 46 genome-wide significant loci. (**a**) The imputation quality and MAF for the lead variants at 46 genome-wide significant susceptibility loci. Blue circles, new additive loci; red squares, new recessive loci; black triangles, previously mapped additive loci; black diamonds, key SNPs in *LPA* and *APOE*. Imputation quality and MAF were each calculated as the median of the respective values in up to 48 contributing studies; the imputation quality for studies with genotype data was fixed at 1.0. (**b**) The odds ratio and effect allele frequency (EAF) for the lead variants at 46 genome-wide significant loci. Blue circles, new additive loci; red squares, new recessive loci; black triangles, previously mapped additive loci. SNPs rs55730499 and rs2891168 are lead variants in the *LPA* and chromosome 9p21 susceptibility loci, respectively. EAF was calculated as the median value in up to 48 contributing studies.

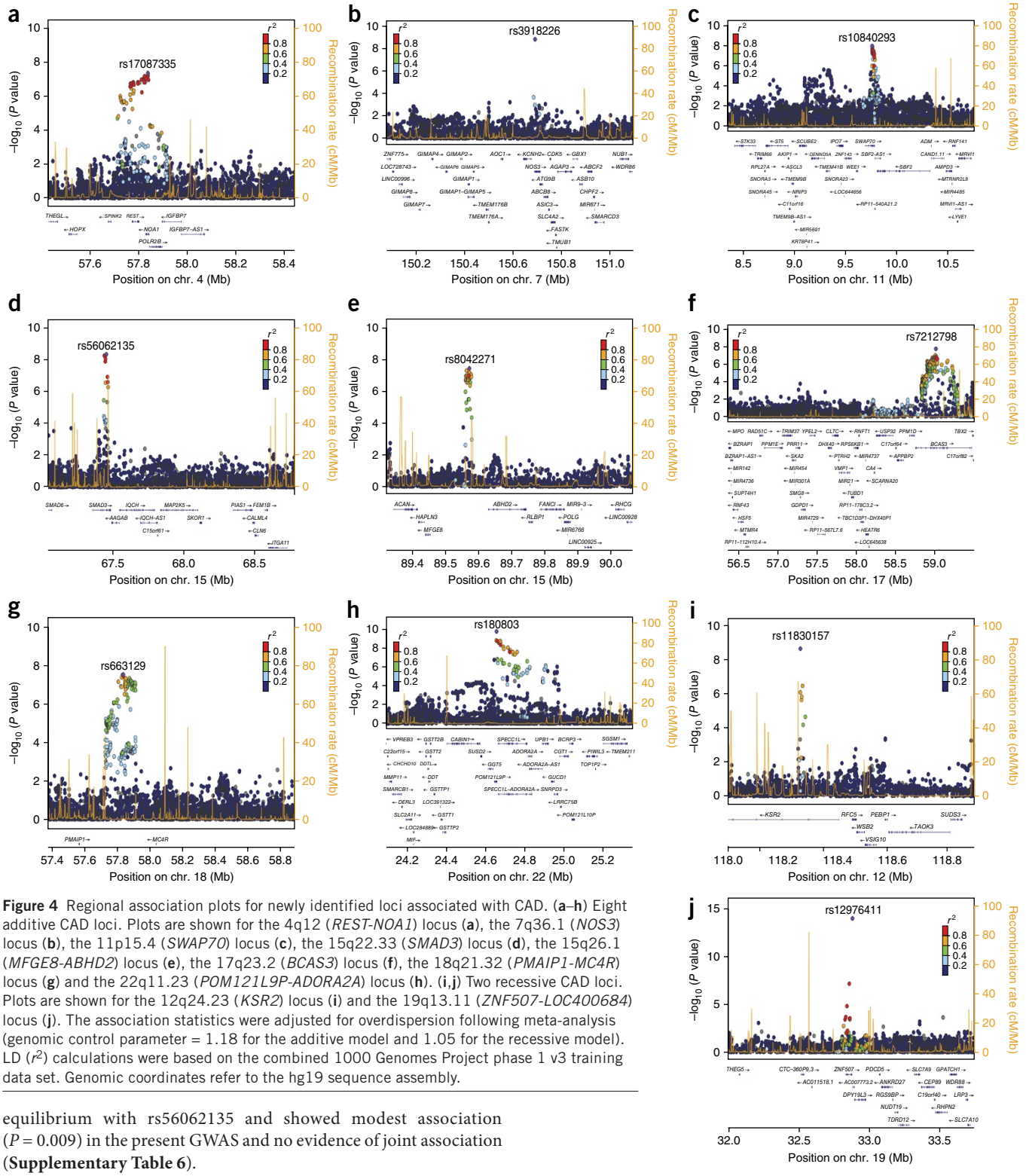

Figure 4 Regional association plots for newly identified loci associated with CAD. (**a**–**h**) Eight additive CAD loci. Plots are shown for the 4q12 (*REST-NOA1*) locus (**a**), the 7q36.1 (*NOS3*) locus (**b**), the 11p15.4 (*SWAP70*) locus (**c**), the 15q22.33 (*SMAD3*) locus (**d**), the 15q26.1 (*MFGE8-ABHD2*) locus (**e**), the 17q23.2 (*BCAS3*) locus (**f**), the 18q21.32 (*PMAIP1-MC4R*) locus (**g**) and the 22q11.23 (*POM121L9P-ADORA2A*) locus (**h**). (**i**,**j**) Two recessive CAD loci. Plots are shown for the 12q24.23 (*KSR2*) locus (**i**) and the 19q13.11 (*ZNF507-LOC400684*) locus (**j**). The association statistics were adjusted for overdispersion following meta-analysis (genomic control parameter = 1.18 for the additive model and 1.05 for the recessive model). LD (r^2) calculations were based on the combined 1000 Genomes Project phase 1 v3 training data set. Genomic coordinates refer to the hg19 sequence assembly.

equilibrium with rs56062135 and showed modest association ($P = 0.009$) in the present GWAS and no evidence of joint association (**Supplementary Table 6**).

At the chromosome 15q26.1 (*MFGE8-ABHD2*) locus, the lead intergenic SNP rs8042271 maps 117 kb upstream of *MFGE8* (milk fat globule–EGF factor 8) and 57 kb upstream of *ABHD2* (abhydrolase domain–containing protein 2) (**Fig. 4e**). MFGE8 (lactadherin) has a crucial role in vascular endothelial growth factor (VEGF)-dependent neovascularization[32], and it is secreted from activated macrophages and binds to apoptotic cells, facilitating phagocytic engulfment[33]. *ABHD2* (ref. 34) has been shown to be expressed in human atherosclerotic lesions, with higher levels in patients with unstable angina. There were no overlapping risk factor quantitative trait locus (QTL), eQTL or ENCODE features in this locus to guide the nomination of a putative causal gene.

At the chromosome 17q23.2 (*BCAS3*) locus, the lead intronic SNP rs7212798 lies in *BCAS3* (breast carcinoma amplified sequence 3) (**Fig. 4f**). Multiple variants in LD with rs7212798 map to *BCAS3* introns and showed strong association with CAD. *BCAS3* encodes

the Rudhira protein, which has been shown to activate Cdc42 to affect actin organization and control cell polarity and motility in endothelial cells, thus contributing to angiogenesis[35].

At the chromosome 18q21.32 (*PMAIP1-MC4R*) locus, the lead intergenic SNP rs663129 lies 266 kb downstream of *PMAIP1* (phorbol-12-myristate-13-acetate–induced protein 1) and 200 kb downstream of *MC4R* (melanocortin 4 receptor) (**Fig. 4g**). *PMAIP1* is a hypoxia-inducible factor (HIF)-1α–induced proapoptotic gene that mediates hypoxic cell death by the generation of reactive oxygen species[36]. *MC4R* is a well-studied obesity-related locus, and the variant (and corresponding proxy variants) that were associated with higher CAD risk are also associated with body mass index (BMI) ($P = 6 \times 10^{-42}$) and obesity-associated risk factors, including higher triglyceride and lower high-density lipoprotein (HDL) concentrations and type 2 diabetes[37–41]. However, we found no eQTL data or ENCODE features for the lead or proxy SNPs to further implicate *MC4R* as the causal gene underlying CAD susceptibility.

At the chromosome 22q11.23 (*POM121L9P-ADORA2A*) locus, the lead SNP rs180803 lies in *POM121L9P* (encoding the noncoding RNA POM121 transmembrane nucleoporin–like 9, pseudogene). A 2-cM region centered on this variant spans 1.2 Mb and includes 21 variants that were associated with CAD at genome-wide significance, most of which are in LD ($r^2 > 0.6$) with the lead SNP and map to intronic regions of the *SPECC1L* and *ADORA2A* genes (**Fig. 4h**).

At the chromosome 12q24.23 (*KSR2*) locus, the lead SNP rs11830157 (MAF = 0.36) associated with CAD risk in a recessive model (genotypic OR = 1.12) is intronic to *KSR2* (kinase suppressor of ras 2) (**Fig. 4i**) and overlaps with ENCODE functional elements. KSR2 interacts with multiple proteins, including AMP-activated protein kinase (AMPK), and rare loss-of-function coding variants in *KSR2* are associated with severe obesity, hyperphagia and insulin resistance, a phenotype recapitulated in *Ksr2*-null mice[42].

At the chromosome 19q13.11 (*ZNF507-LOC400684*) locus, the lead SNP rs12976411 (MAF = 0.09) lies in a gene for an uncharacterized noncoding RNA (*LOC400684*) and is 3.4 kb downstream of *ZNF507* (**Fig. 4j**). The minor allele showed a protective effect in CAD (genotypic OR = 0.69) in the recessive model. ENCODE analysis of this locus suggests that several SNPs, including rs12981453 and rs71351160, which are in strong LD ($r^2 > 0.8$) and are intronic to *ZNF507*, overlap with ENCODE functional elements.

DISCUSSION

We demonstrate that the ability of GWAS to investigate the genetic architecture of complex traits is enhanced by the 1000 Genomes Project. Analysis with this reference set has allowed us to conclude that low-frequency variants of larger effect, synthetic associations and indel polymorphisms are unlikely to explain a significant portion of the missing heritability for CAD. Rather, all ten newly identified CAD associations found in the present analysis, as well as all but one of the previously identified loci, are represented by risk alleles with a frequency of >5%. Thus, this comprehensive analysis strongly supports the common disease–common variant hypothesis[43], given that it was powered to detect variants with MAF <0.05 having OR >1.5. Moreover, risk-associated alleles are significantly clustered within or close to genes and are enriched in regions with functional annotations. Finally, genes implicated by this unbiased approach suggest hypotheses that explore the biology of the arterial vessel wall as a critical component of CAD pathogenesis.

The success of the GWAS meta-analysis strategy in mapping common, small-effect susceptibility variants for complex diseases has leaned heavily on genotype imputation with publically available training sets. The 1000 Genomes Project provides a substantial step-up from the HapMap era in terms of coverage of lower-frequency variants and the integration of indel polymorphism (**Fig. 1**). The lead SNPs for four of the ten newly identified CAD loci were either absent or imperfectly tagged ($r^2 < 0.8$) in the HapMap 2 training set, which reduced the power of discovering these loci in previous GWAS meta-analyses. Although lower-frequency variants often show geographical differentiation[5], the 1000 Genomes Project phase 1 v3 training set includes numerous low-MAF variants that are tractable to a global meta-analysis that includes ancestry groups from multiple continents. Key SNPs in *APOE* and *PCSK9*, which mediate their effects on CAD via LDL cholesterol–linked mechanisms, showed strong associations and reinforce the sensitivity of our 1000 Genomes Project analysis in detecting lower-frequency, imperfectly imputed susceptibility variants that were missed in HapMap-based GWAS.

Association analysis under the customary additive inheritance model widely used in GWAS is optimally powered to detect traits with no dominance variance but conveniently has adequate power to also detect dominantly inherited traits[44]. However, the additive model is systematically underpowered to detect recessively inherited traits, particularly with lower-frequency alleles[44]. This motivated our meta-analysis using a recessive model, which identified two new CAD risk loci, *KSR2* and *ZNF507-LOC400684*, that escaped detection in a conventional additive association scan.

Our GWAS explores two potential sources of missing heritability for CAD, as it includes indels and an extended panel of lower-frequency variants. Although there was no evidence that indels were systematically enriched for CAD association, they represented 10% of the 202 variants with an FDR q value <5%. In terms of surveying the totality of human genetic variation, the 1.5 million of the 2.7 million lower-frequency variants included in the meta-analysis with power to detect alleles of moderate penetrance (OR > 1.3) might seem modest. Yet the relative paucity of significant associations for these variants and the finding that 15 variants with MAF <0.05 explained 2% of CAD heritability and provided no evidence of synthetic associations will temper expectations for the role of low-frequency variants in CAD susceptibility, specifically with respect to risk prediction in a population-based setting. It is important to acknowledge that GWAS analysis based on SNP array data has limited power to resolve genes with rare mutation burdens. For example, *LDLR*[45], *APOA5* (ref. 45), *APOC3* (ref. 46) and *NPC1L1* (ref. 47) are loaded with risk-conferring or protective mutations for CAD. These mutations were only discovered by whole-exome sequencing studies in large series of cases and controls and explain less than 1% of the missing heritability for CAD[45].

Annotation analysis showed that the CAD-associated variants were significantly clustered within or close to genes. Furthermore, there was strong and independent enrichment for overlap of the CAD associations with ENCODE features, particularly in cell types relevant to CAD pathogenesis. This phenomenon has previously been reported for other diseases and traits[48] and can guide candidate gene nomination and the design of future functional studies. We found few suggestions of overlap with risk factor QTLs or eQTLs in available data sets; this may in part reflect that the use of proxy variants can be limiting in cross-referencing the 1000 Genomes Project and HapMap association databases.

Coronary atherosclerosis underlies the development of the vast majority of myocardial infarction cases; therefore, the two are intimately related. However, additional factors, such as plaque vulnerability and the extent of the thrombotic reaction to plaque disruption, may predispose to myocardial infarction in the presence

of CAD[49]. We confirmed that *ABO* is particularly associated with risk of myocardial infarction[50], suggesting that this locus may specifically increase the risk of plaque rupture and/or thrombosis. In contrast, *HDAC9* showed a stronger association with CAD than with myocardial infarction, suggesting that it might predispose to atherosclerosis but not the precipitant events leading to a myocardial infarction. However, *HDAC9* shows even stronger association with ischemic strokes involving thrombosis or embolism due to atherosclerosis of a large artery[51]. Although further epidemiological as well as experimental data are required to substantiate these findings, they suggest that certain loci may affect distinct mechanisms related to the development and progression of CAD.

Several of the genes implicated thus far in large-scale analyses of CAD susceptibility encode proteins with a known role in the biology of risk factors for CAD, notably circulating lipid levels and the metabolism of lipoproteins; other susceptibility genes are related to other known atherosclerosis risk factors, including genes implicated in systemic inflammation and hypertension. Such findings are unsurprising, partly because of the undoubted importance of these known risk factors in the etiology of CAD but also because some of the previous analyses particularly targeted genes involved in risk factor traits; for example, HumanCVD BeadChip[52] design was based on candidate genes, and the Metabochip studies[3,53] drew on earlier association data with risk factor traits as well as an earlier HapMap 2–based CAD GWAS meta-analysis[54]. The current experiment adopts a completely unbiased approach and, to our knowledge, is the first to do so at very large scale. In this respect, it is notable that, for some of the newly identified loci where genomic data, biological precedent and eQTL associations suggest a plausible candidate gene for CAD, the genes so implicated have well-documented roles in vessel wall biology. Their gene products are involved in diverse processes, including cell adhesion and leukocyte and VSMC migration (*SWAP70* (ref. 26) and *ABHD2* (ref. 55)), VSMC phenotypic switching (*REST*[20]), TGF-β signaling (*SMAD3* (refs. 56,57)), anti-inflammatory and infarct-sparing effects (*ADORA2A*[58] and *MFGE8* (ref. 59)), angiogenesis (*BCAS3* (ref. 35)) and NO signaling (*NOS3* (ref. 24)).

It is important to note that these putative new susceptibility genes require substantial further investigation and validation before firm links to vascular biology can be established. A number of preventative strategies target the vessel wall (control of blood pressure and smoking cessation), but the large majority of existing drug treatments for lowering CAD risk operate through manipulation of circulating lipid levels and few directly target vessel wall processes. Detailed investigation of new aspects of vessel wall biology that are implicated by genetic association but have not previously been explored in atherosclerosis may provide new insights into the complex etiology of disease and, hence, identify new targets.

URLs. Ensembl database, http://www.ensembl.org/; University of Chicago eQTL browser, http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/; Genotype-Tissue Expression (GTEx) Portal, http://www.gtexportal.org/home/; Geuvadis Data Browser, http://www.ebi.ac.uk/Tools/geuvadis-das/; CARDIoGRAMplusC4D Consortium, http://www.cardiogramplusc4d.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

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

ACKNOWLEDGMENTS

We sincerely thank the participants and the medical, nursing, technical and administrative staff in each of the studies who have contributed to this project. We are grateful for support from our funders; more detailed acknowledgments are included in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Cohort oversight: D.A., E.B., I.B.B., E.P.B., J.E.B., J.C.C., R. Collins, L.A.C., J.D., I.D., R.E., S.E.E., T.E., M.F.F., O.H.F., M.G.F., C.B.G., D. Gu, V.G., A.S.H., A. Hamsten, T.B.H., S.L.H., C.H., A. Hofman, E.I., C.I., J.W.J., P.J.K., B.-J.K., J.S.K., I.J.K., T.L., R.J.F.L., O.M., A.M., W.M., C.N.P., M.P., T.Q., D.J.R., P.M.R., S.R., R.R., V.S., D.K.S., S.M.S., U.S., A.F.S., D.J.S., J.T., P.A.Z., C.J.O'D., M.P.R., T.L.A., J.R.T., J.E., H.W., S. Kathiresan, R.M., P.D., H.S., N.J.S. and M.F. Cohort genotyping: H.-H.W., S. Kanoni, D.S., J.C.H., Jie Huang, M.E.K., Y.L., L.-P.L., A.U., S.S.A., L.B., G.D., D. Gauguier, A.H.G., M.H., B.-G.H., S.J., L. Lind, C.M.L., M.-L.L., P.K.M., A.P.M., M.S.N., N.L.P., J.S., K.E.S., S.T., L.W., I.B.B., J.C.C., R. Collins, M.F.F., A. Hofman, E.I., J.S.K., T.L., R.R., D.K.S., A.F.S., R. Clarke, P.D. and N.J.S. Cohort phenotyping: D.S., J.C.H., A.D., M.A., K.A., Y.K.K., E.M., L.M.R., S.S.A., F.B., G.D., P.F., A.H.G., O.G., Jianfeng Huang, T. Kessler, I.R.K., L. Lannfelt, W.L., L. Lind, C.M.L., P.K.M., N.H.M., N.M., T.M., F.-ur-R.M., A.P.M., N.L.P., A.P., L.S.R., A.R., M. Samuel, S.H.S., K.S.Z., D.A., J.E.B., J.C.C., R. Collins, R.E., C.B.G., V.G., A.S.H., A. Hamsten, S.L.H., E.I., J.W.J., P.J.K., J.S.K., I.J.K., O.M., A.M., M.P., R.R., D.K.S., A.F.S., D.J.S., P.A.Z., M.P.R., R. Clarke, S. Kathiresan, H.S. and N.J.S. Cohort data analyst: M.N., A.G., H.-H.W., L.M.H., C.W., S. Kanoni, D.S., T. Kyriakou, C.P.N., J.C.H., T.R.W., L.Z., A.D., M.A., S.M.A., K.A., A.B., D.I.C., S.C., I.F., N.F., C. Gieger, C. Grace, S.G., Jie Huang, S.-J.H., Y.K.K., M.E.K., K.W.L., X.L., Y.L., L.-P.L., E.M., A.C.M., N.P., L.Q., L.M.R., E.S., R.S., M. Scholz, A.V.S., E.T., A.U., X.Y., W. Zhang, W. Zhao, M.d.A., P.S.d.V., N.R.v.Z., M.F.F., J.R.T. and M.F. Meta-analysis: M.N., A.G., H.-H.W., L.M.H., C.P.N., J.R.T. and M.F. Variant annotation: M.N., A.G., H.-H.W., T. Kyriakou, J.C.H. and T.R.W. Manuscript drafting: M.N., A.G., H.-H.W., L.M.H., T. Kyriakou, J.C.H., H.W., S. Kathiresan, R.M., H.S., N.J.S. and M.F. Project steering committee: M.N., A.G., H.-H.W., L.M.H., S. Kanoni., J.C.H., D.I.C., M.E.K., N.R.v.Z., C.N.P., R.R., C.J.O'D., M.P.R., T.L.A., J.R.T., J.E., R. Clarke, H.W., S. Kathiresan, R.M., P.D., H.S., N.J.S. and M.F. (secretariat: J.C.H. and R. Clarke). CARDIoGRAMplusC4D executive committee: J.D., D. Gu, A. Hamsten, J.S.K., R.R., H.W., S. Kathiresan, P.D., H.S. and N.J.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Kessler, T., Erdmann, J. & Schunkert, H. Genetics of coronary artery disease and myocardial infarction—2013. *Curr. Cardiol. Rep.* **15**, 368 (2013).
2. O'Donnell, C.J. & Nabel, E.G. Genomics of cardiovascular disease. *N. Engl. J. Med.* **365**, 2098–2109 (2011).
3. CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25–33 (2013).
4. Coronary Artery Disease Genetics (C4D) Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* **43**, 339–344 (2011).
5. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
6. Wang, F. *et al.* Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat. Genet.* **43**, 345–349 (2011).
7. IBC 50K CAD Consortium. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet.* **7**, e1002260 (2011).
8. Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* **361**, 2518–2528 (2009).
9. Bennet, A.M. *et al.* Association of apolipoprotein E genotypes with lipid levels and coronary risk. *J. Am. Med. Assoc.* **298**, 1300–1311 (2007).
10. Benn, M., Nordestgaard, B.G., Grande, P., Schnohr, P. & Tybjaerg-Hansen, A. PCSK9 R46L, low-density lipoprotein cholesterol levels, and risk of ischemic heart disease: 3 independent studies and meta-analyses. *J. Am. Coll. Cardiol.* **55**, 2833–2842 (2010).
11. Cohen, J.C., Boerwinkle, E., Mosley, T.H. Jr. & Hobbs, H.H. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
12. Myocardial Infarction Genetics Consortium. A *PCSK9* missense variant associated with a reduced risk of early-onset myocardial infarction. *N. Engl. J. Med.* **358**, 2299–2300 (2008).
13. Peloso, G.M. *et al.* Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am. J. Hum. Genet.* **94**, 223–232 (2014).
14. Davies, R.W. *et al.* A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ Cardiovasc Genet* **5**, 217–225 (2012).

15. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
16. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D.B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
17. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
18. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
19. Tang, T. *et al.* hNOA1 interacts with complex I and DAP3 and regulates mitochondrial respiration and apoptosis. *J. Biol. Chem.* **284**, 5414–5424 (2009).
20. Chong, J.A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949–957 (1995).
21. Cheong, A. *et al.* Downregulated REST transcription factor is a switch enabling critical potassium channel expression and cell proliferation. *Mol. Cell* **20**, 45–52 (2005).
22. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* **8**, e1003029 (2012).
23. Salvi, E. *et al.* Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension* **59**, 248–255 (2012).
24. Erdmann, J. *et al.* Dysfunctional nitric oxide signalling increases risk of myocardial infarction. *Nature* **504**, 432–436 (2013).
25. Casas, J.P. *et al.* Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. *Am. J. Epidemiol.* **164**, 921–935 (2006).
26. Chacón-Martínez, C.A. *et al.* The switch-associated protein 70 (SWAP-70) bundles actin filaments and contributes to the regulation of F-actin dynamics. *J. Biol. Chem.* **288**, 28687–28703 (2013).
27. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
28. Fairfax, B.P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
29. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
30. Ashcroft, G.S. *et al.* Mice lacking Smad3 show accelerated wound healing and an impaired local inflammatory response. *Nat. Cell Biol.* **1**, 260–266 (1999).
31. Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
32. Silvestre, J.S. *et al.* Lactadherin promotes VEGF-dependent neovascularization. *Nat. Med.* **11**, 499–506 (2005).
33. Hanayama, R. *et al.* Identification of a factor that links apoptotic cells to phagocytes. *Nature* **417**, 182–187 (2002).
34. Miyata, K. *et al.* Elevated mature macrophage expression of human *ABHD2* gene in vulnerable plaque. *Biochem. Biophys. Res. Commun.* **365**, 207–213 (2008).
35. Jain, M., Bhat, G.P., Vijayraghavan, K. & Inamdar, M.S. Rudhira/BCAS3 is a cytoskeletal protein that controls Cdc42 activation and directional cell migration during angiogenesis. *Exp. Cell Res.* **318**, 753–767 (2012).
36. Kim, J.Y., Ahn, H.J., Ryu, J.H., Suk, K. & Park, J.H. BH3-only protein Noxa is a mediator of hypoxic cell death induced by hypoxia-inducible factor 1α. *J. Exp. Med.* **199**, 113–124 (2004).
37. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
38. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
39. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
40. Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
41. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
42. Pearce, L.R. *et al. KSR2* mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* **155**, 765–777 (2013).
43. Schork, N.J., Murray, S.S., Frazer, K.A. & Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **19**, 212–219 (2009).
44. Lettre, G., Lange, C. & Hirschhorn, J.N. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet. Epidemiol.* **31**, 358–362 (2007).
45. Do, R. *et al.* Exome sequencing identifies rare *LDLR* and *APOA5* alleles conferring risk for myocardial infarction. *Nature* **518**, 102–106 (2015).
46. TG and HDL Working Group of the Exome Sequencing Project. Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. *N. Engl. J. Med.* **371**, 22–31 (2014).
47. Myocardial Infarction Genetics Consortium Investigators. Inactivating mutations in *NPC1L1* and protection from coronary heart disease. *N. Engl. J. Med.* **371**, 2072–2082 (2014).
48. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
49. Libby, P., Ridker, P.M. & Hansson, G.K. Progress and challenges in translating the biology of atherosclerosis. *Nature* **473**, 317–325 (2011).
50. Reilly, M.P. *et al.* Identification of *ADAMTS7* as a novel locus for coronary atherosclerosis and association of *ABO* with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383–392 (2011).
51. Dichgans, M. *et al.* Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke* **45**, 24–36 (2014).
52. Keating, B.J. *et al.* Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS ONE* **3**, e3583 (2008).
53. 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).
54. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
55. Miyata, K. *et al.* Increase of smooth muscle cell migration and of intimal hyperplasia in mice lacking the α/β hydrolase domain containing 2 gene. *Biochem. Biophys. Res. Commun.* **329**, 296–304 (2005).
56. Bobik, A. Transforming growth factor-βs and vascular disorders. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1712–1720 (2006).
57. Mallat, Z. *et al.* Inhibition of transforming growth factor-β signaling accelerates atherosclerosis and induces an unstable plaque phenotype in mice. *Circ. Res.* **89**, 930–934 (2001).
58. Yang, Z. *et al.* Infarct-sparing effect of A2A-adenosine receptor activation is due primarily to its action on lymphocytes. *Circulation* **111**, 2190–2197 (2005).
59. Aziz, M., Jacob, A., Matsuda, A. & Wang, P. Review: milk fat globule–EGF factor 8 expression, function and plausible signal transduction in resolving inflammation. *Apoptosis* **16**, 1077–1086 (2011).
60. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).

Majid Nikpay[1,130], Anuj Goel[2,3,130], Hong-Hee Won[4–7,130], Leanne M Hall[8,130], Christina Willenborg[9,10,130], Stavroula Kanoni[11,130], Danish Saleheen[12,13,130], Theodosios Kyriakou[2,3], Christopher P Nelson[8,14], Jemma C Hopewell[15], Thomas R Webb[8,14], Lingyao Zeng[16,17], Abbas Dehghan[18], Maris Alver[19,20], Sebastian M Armasu[21], Kirsi Auro[22–24], Andrew Bjonnes[4,6], Daniel I Chasman[25,26], Shufeng Chen[27], Ian Ford[28], Nora Franceschini[29], Christian Gieger[17,30,31], Christopher Grace[2,3], Stefan Gustafsson[32,33], Jie Huang[34], Shih-Jen Hwang[35,36], Yun Kyoung Kim[37], Marcus E Kleber[38], King Wai Lau[15], Xiangfeng Lu[27], Yingchang Lu[39,40], Leo-Pekka Lyytikäinen[41,42], Evelin Mihailov[19], Alanna C Morrison[43], Natalia Pervjakova[19,22–24], Liming Qu[44], Lynda M Rose[25], Elias Salfati[45], Richa Saxena[4,6,46], Markus Scholz[47,48], Albert V Smith[49,50], Emmi Tikkanen[23,51], Andre Uitterlinden[18], Xueli Yang[27], Weihua Zhang[52,53], Wei Zhao[12], Mariza de Andrade[21], Paul S de Vries[18], Natalie R van Zuydam[3,54], Sonia S Anand[55], Lars Bertram[56,57], Frank Beutner[48,58], George Dedoussis[59], Philippe Frossard[13], Dominique Gauguier[60], Alison H Goodall[14,61], Omri Gottesman[39], Marc Haber[62], Bok-Ghee Han[37], Jianfeng Huang[63], Shapour Jalilzadeh[2,3], Thorsten Kessler[16,64], Inke R König[10,65], Lars Lannfelt[66], Wolfgang Lieb[67], Lars Lind[68], Cecilia M Lindgren[3,4], Marja-Liisa Lokki[69], Patrik K Magnusson[70],

Nadeem H Mallick[71], Narinder Mehra[72], Thomas Meitinger[17,73,74], Fazal-ur-Rehman Memon[75], Andrew P Morris[3,76], Markku S Nieminen[77], Nancy L Pedersen[70], Annette Peters[17,30], Loukianos S Rallidis[78], Asif Rasheed[13,75], Maria Samuel[13], Svati H Shah[79], Juha Sinisalo[77], Kathleen E Stirrups[11,80], Stella Trompet[81,82], Laiyuan Wang[27,83], Khan S Zaman[84], Diego Ardissino[85,86], Eric Boerwinkle[43,87], Ingrid B Borecki[88], Erwin P Bottinger[39], Julie E Buring[25], John C Chambers[52,53,89], Rory Collins[15], L Adrienne Cupples[35,36], John Danesh[34,90], Ilja Demuth[91,92], Roberto Elosua[93], Stephen E Epstein[94], Tõnu Esko[4,19,95,96], Mary F Feitosa[88], Oscar H Franco[18], Maria Grazia Franzosi[97], Christopher B Granger[79], Dongfeng Gu[27], Vilmundur Gudnason[49,50], Alistair S Hall[98], Anders Hamsten[99], Tamara B Harris[100], Stanley L Hazen[101], Christian Hengstenberg[16,17], Albert Hofman[18], Erik Ingelsson[3,32,33,45], Carlos Iribarren[102], J Wouter Jukema[81,103,104], Pekka J Karhunen[41,105], Bong-Jo Kim[37], Jaspal S Kooner[53,89,106], Iftikhar J Kullo[107], Terho Lehtimäki[41,42], Ruth J F Loos[39,40,108], Olle Melander[109], Andres Metspalu[19,20], Winfried März[38,110,111], Colin N Palmer[54], Markus Perola[19,22–24], Thomas Quertermous[45,112], Daniel J Rader[113,114], Paul M Ridker[25,26], Samuli Ripatti[23,34,51], Robert Roberts[115], Veikko Salomaa[116], Dharambir K Sanghera[117–119], Stephen M Schwartz[120,121], Udo Seedorf[122], Alexandre F Stewart[1], David J Stott[123], Joachim Thiery[48,124], Pierre A Zalloua[62,125], Christopher J O'Donnell[35,126,127], Muredach P Reilly[114], Themistocles L Assimes[45,112], John R Thompson[128], Jeanette Erdmann[9,10], Robert Clarke[15], Hugh Watkins[2,3,131], Sekar Kathiresan[4–7,131], Ruth McPherson[1,131], Panos Deloukas[11,129,131], Heribert Schunkert[16,17,131], Nilesh J Samani[8,14,131] & Martin Farrall[2,3,131] for the CARDIoGRAMplusC4D Consortium

[1]Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada. [2]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [4]Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA. [5]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA. [6]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [7]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [8]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. [9]Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, Lübeck, Germany. [10]DZHK (German Research Center for Cardiovascular Research), partner site Hamburg-Lübeck-Kiel, Lübeck, Germany. [11]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [12]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [13]Center for Non-Communicable Diseases, Karachi, Pakistan. [14]National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. [15]Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK. [16]Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. [17]DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany. [18]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. [19]Estonian Genome Center, University of Tartu, Tartu, Estonia. [20]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [21]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. [22]Department of Health, National Institute for Health and Welfare, Helsinki, Finland. [23]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [24]Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland. [25]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [26]Harvard Medical School, Boston, Massachusetts, USA. [27]State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [28]Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK. [29]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA. [30]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [31]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [32]Molecular Epidemiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [33]Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [34]Wellcome Trust Sanger Institute, Hinxton, UK. [35]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [36]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [37]Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, Korea. [38]Vth Department of Medicine (Nephrology, Hypertensiology, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, Mannheim, Germany. [39]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [40]Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [41]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. [42]Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland. [43]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA. [44]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [45]Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, California, USA. [46]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [47]Institute for Medical Informatics, Statistics and Epidemiology, Medical Faculty, University of Leipzig, Leipzig, Germany. [48]LIFE Research Center of Civilization Diseases, Leipzig, Germany. [49]Icelandic Heart Association, Kopavogur, Iceland. [50]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [51]Department of Public Health, University of Helsinki, Helsinki, Finland. [52]Department of Epidemiology and Biostatistics, Imperial College London, London, UK. [53]Department of Cardiology, Ealing Hospital National Health Service (NHS) Trust, Middlesex, UK. [54]Medical Research Institute, University of Dundee, Dundee, UK. [55]Population Health Research Institute, Hamilton Health Sciences, Department of Medicine, McMaster University, Hamilton, Ontario, Canada. [56]Platform for Genome Analytics, Institutes of Neurogenetics and Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany. [57]Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, Imperial College of Science, Technology and Medicine, London, UK. [58]Heart Center Leipzig, Cardiology, University of Leipzig, Leipzig, Germany. [59]Department of Dietetics-Nutrition, Harokopio University, Athens, Greece. [60]INSERM, UMRS 1138, Centre de Recherche des Cordeliers, Paris, France. [61]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. [62]School of Medicine, Lebanese American University, Beirut, Lebanon. [63]Hypertension Division, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [64]Klinikum Rechts der Isar, Munich, Germany. [65]Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany. [66]Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden. [67]Institut für Epidemiologie, Christian Albrechts Universität zu Kiel, Kiel, Germany. [68]Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden. [69]Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland. [70]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [71]Punjab Institute of Cardiology, Lahore, Pakistan. [72]All India Institute of Medical Sciences, New Delhi, India.

[73]Institut für Humangenetik, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [74]Institute of Human Genetics, Technische Universität München, Munich, Germany. [75]Red Crescent Institute of Cardiology, Hyderabad, Pakistan. [76]Department of Biostatistics, University of Liverpool, Liverpool, UK. [77]Department of Cardiology, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland. [78]Second Department of Cardiology, Attikon Hospital, School of Medicine, University of Athens, Athens, Greece. [79]Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA. [80]Department of Haematology, University of Cambridge, Cambridge, UK. [81]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands. [82]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. [83]National Human Genome Center at Beijing, Beijing, China. [84]National Institute of Cardiovascular Diseases, Karachi, Pakistan. [85]Division of Cardiology, Azienda Ospedaliero Universitaria di Parma, Parma, Italy. [86]Associazione per lo Studio della Trombosi in Cardiologia, Pavia, Italy. [87]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. [88]Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [89]Imperial College Healthcare NHS Trust, London, UK. [90]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [91]Berlin Aging Study II; Research Group on Geriatrics, Charité Universitätsmedizin Berlin, Berlin, Germany. [92]Institute of Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany. [93]Grupo de Epidemiología y Genética Cardiovascular, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain. [94]MedStar Heart and Vascular Institute, MedStar Washington Hospital Center, Washington, DC, USA. [95]Division of Endocrinology and Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA. [96]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [97]Department of Cardiovascular Research, Istituto di Ricerca e Cura a Carattere Scientifico (IRCCS) Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy. [98]Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK. [99]Cardiovascular Genetics and Genomics Group, Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. [100]Laboratory of Epidemiology, Demography and Biometry, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. [101]Department of Cellular and Molecular Medicine, Cleveland Clinic, Cleveland, Ohio, USA. [102]Kaiser Permanente Division of Research, Oakland, California, USA. [103]Durrer Center for Cardiogenetic Research, Amsterdam, the Netherlands. [104]Interuniversity Cardiology Institute of the Netherlands, Utrecht, the Netherlands. [105]Department of Forensic Medicine, University of Tampere School of Medicine, Tampere, Finland. [106]Cardiovascular Science, National Heart and Lung Institute, Imperial College London, London, UK. [107]Division of Cardiovascular Diseases, Department of Medicine, Mayo Clinic, Rochester, Minnesota, USA. [108]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [109]Department of Clinical Sciences, Hypertension and Cardiovascular Disease, Lund University, University Hospital Malmö, Malmö, Sweden. [110]Synlab Academy, Synlab Services, Mannheim, Germany. [111]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. [112]Stanford Cardiovascular Institute, Stanford University, Stanford, California, USA. [113]Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [114]Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [115]University of Ottawa Heart Institute, Ottawa, Ontario, Canada. [116]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [117]Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA. [118]Department of Pharmaceutical Sciences, College of Pharmacy, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA. [119]Oklahoma Center for Neuroscience, Oklahoma City, Oklahoma, USA. [120]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [121]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [122]Department of Prosthetic Dentistry, Center for Dental and Oral Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [123]Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow, UK. [124]Institute for Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital Leipzig, Medical Faculty, Leipzig, Germany. [125]Harvard School of Public Health, Boston, Massachusetts, USA. [126]National Heart, Lung, and Blood Institute Division of Intramural Research, Bethesda, Maryland, USA. [127]Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts, USA. [128]Department of Health Sciences, University of Leicester, Leicester, UK. [129]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. [130]These authors contributed equally to this work. [131]These authors jointly supervised this work. Correspondence should be addressed to H.W. (hugh.watkins@rdm.ox.ac.uk), S. Kathiresan (sekar@broadinstitute.org), R.M. (rmcpherson@ottawaheart.ca) or M.F. (martin.farrall@well.ox.ac.uk).

ONLINE METHODS

Association analysis. Three models of heritable disease susceptibility were analyzed by logistic regression: (i) an additive model where the log(genotype risk ratio) (log(GRR)) for a genotype was proportional to the number of risk alleles; (ii) a recessive model where the log(GRR) for homozygotes for the minor allele was compared with a reference risk in pooled heterozygotes and homozygotes for the major allele; and (iii) a dominant model where the log(GRR) for homozygotes for the minor allele pooled with heterozygotes was compared with a reference risk in homozygotes for the major allele. Minor and major alleles were identified by reference to allele frequencies in the pooled populations (all continents) of 1000 Genomes Project phase 1 v3 data. For the recessive and dominant analyses, genotype probabilities were analyzed by all contributing studies to allow for variable imputation quality; for the additive analysis, genotype probabilities or allelic dosages were used (**Supplementary Table 1**).

Data quality control. Association data for each contributing study were individually filtered for MAF >0.005 (estimated in cases and controls combined) and an imputation quality metric, rsq >0.3 for Minimac or info_proper >0.4 for IMPUTE2 (ref. 61). Allele frequencies for each study were binned and compared with those from other studies to detect systematic flipping of alleles (**Supplementary Fig. 5**). Overdispersion of association statistics was assessed by the genomic control method[62] (**Supplementary Table 15**), and adjusted values were submitted for meta-analysis. Variants that were retained in at least 60% of the studies were submitted for meta-analysis using the GWAMA program[63]. Following an inverse variance–weighted fixed-effects meta-analysis, heterogeneity was assessed by Cochran's Q statistic[64] and the I^2 inconsistency index[65], and variants showing marked heterogeneity were reanalyzed using a random-effects model[66]. Overdispersion in the resulting meta-analysis statistics was adjusted for by a second application of the genomic control procedure (**Supplementary Fig. 6**).

FDR estimation. FDR was assessed using a step-up procedure coded in the qqvalue Stata program[67]. This procedure has been reported to be well controlled under positive regression dependency conditions[68]; simulations based on 1,000 permuted replicates of the PROCARDIS imputed data demonstrated that the FDR was conservatively controlled (theoretical q value = 0.05, empirical q value = 0.026, 95% CI = 0.017–0.038) in the context of the LD patterns prevalent in the 1000 Genomes Project phase 1 v3 training set.

GCTA and heritability analysis. Joint association analysis of the CAD additive meta-analysis results was performed using GCTA software[17], which fits an approximate multiple regression model on the basis of summary association statistics and LD information derived from a reference genotype database (here the 1000 Genomes Project phase 1 v3 training set for all continents and populations that includes genotypes for 1,092 individuals). In this analysis, the lead variant is not necessarily retained in the final joint association model in situations where there might be multiple associated variants in strong LD. The accuracy of this analysis depends on appropriate ancestry matching as well as the sample size of the reference genotype panel to ensure that estimated LD correlations are unbiased and acceptably precise[69]. Simulations suggest that the expected correlation between P values based on the GCTA method using a reference panel of 1,000 genotyped samples and P values from 'exact' multiple regression based on experimental genotypes will be between 0.90 and 0.95 (ref. 69). We investigated the empirical accuracy of the GCTA joint association analysis by comparing GCTA joint association results with those for a standard multiple–logistic regression analysis in four contributing studies (**Supplementary Fig. 7**). This comparison showed that 95% of the β values (regression coefficients) and standard errors were accurately approximated. The $-\log_{10}$ (P values) from the two analyses were positively correlated ($0.86 < \rho < 0.93$), with the GCTA method showing an insignificant trend ($P > 0.20$) toward yielding slightly inflated values.

Heritability calculations were based on a multifactorial liability-threshold model[70] assuming that the disease prevalence was 5% and that the total heritability of CAD was 40% (ref. 3); multiple regression estimates of allelic effect sizes were used following the GCTA joint association analysis. The standard errors for the heritability estimates were generated by Monte Carlo sampling with 1,000 replicates (for each variant, β values are drawn randomly from the variant's β value ± s.e.m. estimate, heritability is calculated for each β value by replicate draw, heritability is summed across n variants within each replicate and, finally, the standard error of the heritability estimates is calculated across the 1,000 replicates).

Power calculations. Power to detect genetic associations depends on the magnitude of the genetic risk (effect size), the type I error rate, the risk allele frequency and imputation quality, and the sample size. Non-centrality parameter calculations were based on double–genomic controlled standard error estimates from the additive model meta-analysis; these estimates integrate information on allele frequency, imputation quality and sample size, which typically vary across studies. The type I error was set at 5×10^{-8}, and an additive risk model was assumed.

Risk factor QTL survey. The ten newly identified CAD-associated loci were scanned for associations with heritable risk factors for CAD using publically available resources, including large-scale GWAS consortium data downloads[37–41,71–73] and the National Human Genome Research Institute (NHGRI) GWAS catalog[74] (accessed May 2014). As previous GWAS for risk factors were mainly based on HapMap 2–imputed data sets, all SNPs in LD ($r^2 > 0.8$ based on the 1000 Genomes Project phase 1 v3 ALL reference panel) with the new variants were examined for risk factor associations. The newly associated loci were cross-referenced with known *cis*- and *trans*-eQTL associations from the University of Chicago eQTL browser (accessed July 2014), the GTEx Portal (accessed June 2014), the Geuvadis Data Browser (accessed June 2014) and other published data[22,28,29,75–79].

Annotation and ENCODE analysis. Variants were annotated using ANNOVAR software[18] (version August 2013) based on a GRCh37/hg19 gene annotation database. Upstream or downstream status was assigned to variants that mapped ≤1 kb from the transcript start or end, respectively. Variants without intergenic annotation were assigned a genic annotation status (42%). The annotation status of the 9.4 million variants included in the CAD additive meta-analysis is shown in **Supplementary Table 8**; 86% of the genic variants map to introns.

ENCODE features were downloaded from the Ensembl database using the Funcgen Perl API module (release 75). The list of ENCODE experiments stored in the Ensembl database can be browsed at http://Feb2014.archive.ensembl.org/Homo_sapiens/Experiment/Sources?db=core;ex=project-ENCODE-. This list summarizes 100 different types of functional evidence in 11 different cell types for a total of 379 ENCODE experiments that identified 6,099,034 features. Variants that overlapped one or more of these features were cross-tabulated with their ANNOVAR annotation status (**Supplementary Table 10**); 50% of variants mapped to one or more ENCODE features, and variants in ENCODE features were strongly enriched for genic annotation status. Variants were grouped into three functional sets—HMs, DHSs and TFBSs (**Supplementary Table 9**). Cell types were grouped into CAD-relevant types and others (**Supplementary Table 12**) on the basis of their potential roles in CAD pathophysiology; hepatocytes (for example, lipid metabolism[80]), vascular endothelial cells (atherosclerosis[81]) and myoblasts (injury and repair[82]) were selected as being the most relevant to the CAD phenotype. Multiway contingency tables reporting ENCODE feature and ANNOVAR annotation status with inclusion in the list of variants with FDR <5% (FDR202 status) are summarized for 11 ENCODE cell types in **Supplementary Table 11** and for the 3 CAD-relevant cell types in **Supplementary Table 13**. Contingency table counts were modeled by a logistic multiple regression model predicting FDR202 status with the independent explanatory variables HM, DHS, TFBS and genic/intergenic status. The ENCODE[83] project has previously mapped 4,492 significant GWAS SNPs from the NHGRI catalog[74] (accessed June 2011) to transcription factor (12%) and DHS (34%) features in an extended data set of 1,640 experiments. The 202 FDR variants were slightly less prevalent in these feature groups (10.4% transcription factor and 19.8% DHS features), which could reflect a CAD-specific issue or a more general consequence of our analysis being based on a subset of the ENCODE data retrieved from the Ensembl database.

61. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
62. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
63. Mägi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
64. Cochran, W.G. The combination of estimates from different experiments. *Biometrics* **10**, 101–129 (1954).
65. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
66. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).
67. Newson, R.B. Frequentist *q*-values for multiple-test procedures. *Stata J.* **10**, 568–584 (2010).
68. Benjamini, Y. & Yekutieli, D. The control of the false-discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
69. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
70. So, H.C., Gui, A.H., Cherny, S.S. & Sham, P.C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.* **35**, 310–317 (2011).
71. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
72. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
73. Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* **43**, 1005–1011 (2011).
74. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
75. Fehrmann, R.S. *et al.* *Trans*-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7**, e1002197 (2011).
76. Garnier, S. *et al.* Genome-wide haplotype analysis of *cis* expression quantitative trait loci in monocytes. *PLoS Genet.* **9**, e1003240 (2013).
77. Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
78. Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res.* **23**, 716–726 (2013).
79. Westra, H.J. *et al.* Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
80. Busch, S.J., Barnhart, R.L., Martin, G.A., Flanagan, M.A. & Jackson, R.L. Differential regulation of hepatic triglyceride lipase and 3-hydroxy-3-methylglutaryl-CoA reductase gene expression in a human hepatoma cell line, HepG2. *J. Biol. Chem.* **265**, 22474–22479 (1990).
81. Park, H.J. *et al.* Human umbilical vein endothelial cells and human dermal microvascular endothelial cells offer new insights into the relationship between lipid metabolism and angiogenesis. *Stem Cell Rev.* **2**, 93–102 (2006).
82. Durrani, S., Konoplyannikov, M., Ashraf, M. & Haider, K.H. Skeletal myoblasts for cardiac repair. *Regen. Med.* **5**, 919–932 (2010).
83. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).