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

Nikpay, Majid; Goel, Anuj; Won, Hong-Hee; Hall, Leanne M.; Willenborg, Christina; Kanoni, Stavroula

Published in:
Nature Genetics

DOI:
10.1038/ng.3396

Publication date:
2015

Citation for published version (APA):
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 archetypical common complex disease with both genetic and environmental determinants. 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). The 1000 Genomes Project 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 × 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, which previously showed no association in the Metabochip meta-analysis of Europeans and South Asians. 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, also showed strong associations (LPA: rs10455872, P = 5.7 × 10^{-9} and rs3798220, P = 4.7 × 10^{-9}; APOE: rs7412, P = 8.2 × 10^{-11}). The LPA SNPs have previously been shown to be strongly associated with CAD in candidate gene studies based on experimental genotype data.

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; doi:10.1038/ng.3396
The minor allele of SNP rs7412 encodes the e2 allele of APOE, and it has been well documented that carriers of the e2 allele have lower cholesterol levels; significant protection from CAD by this allele was confirmed in a large meta-analysis\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{9}. 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\textsuperscript{17} to perform an approximate joint association analysis (Online Methods), we identified 202 FDR variants (\(q < 0.05\)) 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\textsuperscript{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 (\(q < 0.05\)) with a common variant. Twenty of the 202 FDR variants (9.9%) were

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1}
\caption{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.}
\end{figure}
were strikingly chromosome ideograms with centromeres represented by pink bars.

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\(^\text{[20]}\). 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.

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;\(^\text{[19]}\) REST encodes a transcription factor that suppresses the expression of voltage-dependent sodium and potassium channels;\(^\text{[20]}\); it has been shown to maintain vascular smooth muscle cells (VSMCs) in a quiescent, non-proliferative state and is itself downregulated in neointimal hyperplasia.\(^\text{[21]}\) SNP rs17087335 showed a cis-eQTL signal for REST in lung cells (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 freq. | Imputation quality | \(I^2\) | Heterogeneity \(P\) | \(n\) studies | Additive Association model |
|--------------|------------|-----|-------|---------------------|-----------------|------|-------------|--------------|----------------------|------------------|
| rs17087335   | REST-NOAI  | 4   | T/G   | 0.21               | 0.99            | 0.20 | 0.11        | 48           | 1.06 (1.04–1.09) | 4.60 × 10\(^{-8}\) |
| rs3918226    | NOS3       | 7   | T/C   | 0.06               | 0.78            | 0.15 | 0.19        | 45           | 1.14 (1.09–1.19) | 1.70 × 10\(^{-8}\)  |
| rs10840293   | SWAP70     | 11  | A/G   | 0.55               | 0.94            | 0.17 | 0.16        | 47           | 1.06 (1.04–1.08) | 1.30 × 10\(^{-8}\)  |
| rs56062135   | SMEAD3     | 8   | C/T   | 0.79               | 0.98            | 0.00 | 0.67        | 46           | 1.07 (1.05–1.10) | 4.50 × 10\(^{-9}\)  |
| rs8042277    | MEIS1-ABHD2| 15  | G/A   | 0.9                | 0.93            | 0.16 | 0.19        | 46           | 1.10 (1.04–1.14) | 3.70 × 10\(^{-8}\)  |
| rs7212798    | BOC5       | 15  | C/T   | 0.15               | 0.95            | 0.14 | 0.21        | 48           | 1.08 (1.05–1.11) | 1.90 × 10\(^{-8}\)  |
| rs663129     | PMAP1-MC4R | 18  | A/G   | 0.26               | 1.00            | 0.00 | 0.67        | 47           | 1.06 (1.04–1.08) | 3.20 × 10\(^{-8}\)  |
| rs180803     | POM1219R   | 22  | G/T   | 0.97               | 0.86            | 0.00 | 0.67        | 41           | 1.20 (1.13–1.27) | 1.60 × 10\(^{-10}\) |
| rs11830157   | KSR2       | 12  | G/T   | 0.36               | 0.99            | 0.14 | 0.22        | 42           | 1.04 (1.02–1.06) | 3.90 × 10\(^{-8}\)  |
| rs12976411   | ZNFS07- LOC400684 | | 19 | T/A | 0.09 | 0.93 | 0.50 | 5.09 × 10\(^{-4}\) | 3 | 0.95 (0.92–0.99) | 9.10 × 10\(^{-3}\) |

Association results are presented for two inheritance models; results from the discovery association model are shown in bold. 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.

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). SWAP70 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\(^5\), skin\(^9\) 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.

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.

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...
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 neovascularization32, and it is secreted from activated macrophages and binds to apoptotic cells, facilitating phagocytic engulfment33. 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.

At the chromosome 18q21.32 (PAI1P1-MC4R) locus, the lead intergenic SNP rs663129 lies 266 kb downstream of PAI1P1 (phorbol-12-myristate-13-acetate–induced protein 1) and 200 kb downstream of MC4R (melanocortin 4 receptor) (Fig. 4g). PAI1P1 is a hypoxia-inducible factor (HIF)-1α–induced proapoptotic gene that mediates hypoxic cell death by the generation of reactive oxygen species. 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 × 10^{-42}) and obesity-associated risk factors, including higher triglyceride and lower high-density lipoprotein (HDL) concentrations and type 2 diabetes. 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 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 intrinsic 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.

At the chromosome 19q13.11 (ZNFS07-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 ZNFS07 (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 intrinsic to ZNFS07, 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, 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, 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. However, the additive model is systematically underpowered to detect recessively inherited traits, particularly with lower-frequency alleles. This motivated our meta-analysis using a recessive model, which identified two new CAD risk loci, KSR2 and ZNFS07-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, 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.

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 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.\textsuperscript{49} We confirmed that ABO is particularly associated with risk of myocardial infarction,\textsuperscript{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.\textsuperscript{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\textsuperscript{52} design was based on candidate genes, and the Metaobichip studies,\textsuperscript{53} drew on earlier association data with risk factor traits as well as an earlier HapMap 2-based CAD GWAS meta-analysis.\textsuperscript{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\textsuperscript{20}), TGF-β signaling (SMAD3 (refs. 56,57)), anti-inflammatory and infarct-sparing effects (ADORA2A\textsuperscript{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.

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


COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.


1Rudy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada. 2Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. 3Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. 4Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA. 5Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA. 6Department of Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. 7Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. 8Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. 9Department for Integrative und Experimentelle Genomik, Universität zu Lübeck, Lübeck, Germany. 10DZHK (German Centre for Cardiovascular Research), partner site Hamburg–Lübeck–Kiel, Lübeck, Germany. 11William Harvey Research Institute, Bart’s and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. 12Peirlem School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 13Center for Non-Communicable Diseases, Karachi, Pakistan. 14National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. 15Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK. 16Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. 17DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany. 18Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. 19Estonian Genome Center, University of Tartu, Tartu, Estonia. 20Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. 21Division of Biometrical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. 22Department of Health, National Institute for Health and Welfare, Helsinki, Finland. 23Institute for Molecular Medicine Finland (FMI), University of Helsinki, Helsinki, Finland. 24Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland. 25Division of Preventive Medicine, Brigham and Women’s Hospital, Boston, Massachusetts, USA. 26Harvard Medical School, Boston, Massachusetts, USA. 27Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. 28Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK. 29Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA. 30Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. 31Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. 32Molecular Epidemiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. 33Science for Life Laboratory, Uppsala University, Uppsala, Sweden. 34Wellcome Trust Sanger Institute, Hinxton, UK. 35National Heart, Lung and Blood Institute’s Framingham Heart Study, Framingham, Massachusetts, USA. 36Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. 37Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, Korea. 38Department of Medicine (Nephrology, Hypertension, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, Mannheim, Germany. 39Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA. 40Department of Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, New York, USA. 41Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. 42Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland. 43Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA. 44Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 45Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, California, USA. 46Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. 47Institute for Medical Informatics, Statistics and Epidemiology, Medical Faculty, University of Leipzig, Leipzig, Germany. 48Life Research Center of Civilization Diseases, Leipzig, Germany. 49Icelandic Heart Association, Kopavogur, Iceland. 50Faculty of Medicine, University of Iceland, Reykjavik, Iceland. 51Department of Public Health, University of Helsinki, Helsinki, Finland. 52Department of Epidemiology and Biostatistics, Imperial College London, London, UK. 53Department of Cardiology, Ealing Hospital National Health Service (NHS) Trust, Middlesex, UK. 54Medical Research Institute, University of Dundee, Dundee, UK. 55Population Health Research Institute, Hamilton Health Sciences, Department of Medicine, McMaster University, Hamilton, Ontario, Canada. 56Platform for Genome Analytics, Institutes of Neurogenetics and Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany. 57Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, Imperial College of Science, Technology and Medicine, London, UK. 58Heart Center Leipzig, Cardiology, University of Leipzig, Leipzig, Germany. 59Department of Dietetics-Nutrition, Harokopio University, Athens, Greece. 60INESRM, UMRS 1138, Centre de Recherche des Cordeliers, Paris, France. 61Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. 62School of Medicine, Lebanon American University, Beirut, Lebanon. 63Hypertension Division, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. 64Klinikum Rechts der Isar, Munich, Germany. 65Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany. 66Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden. 67Institut für Epidemiologie, Christian Albrechts Universität zu Kiel, Kiel, Germany. 68Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden. 69Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland. 70Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. 71Punjab Institute of Cardiology, Lahore, Pakistan. 72All India Institute of Medical Sciences, New Delhi, India.
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. 3). Overdispersion of association statistics was assessed by the genomic control method27 (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 program55. Following an inverse variance–weighted fixed-effects analysis, heterogeneity was assessed by Cochran’s Q statistic44 and the η2 inconsistency index55, and variants showing marked heterogeneity were reanalyzed using a random-effects model56. 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 qvalue Stata program57. This procedure has been reported to be well controlled under positive regression dependency conditions58; simulations based on 1,000 permuted replicates of the PROCARDS 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 software16, 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 precise59. 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. 60). 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 P values (regression coefficients) and standard errors were accurately approximated. The −log10(P values) from the two analyses were positively correlated (0.86 < r < 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 model70 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 downloads41–47,71–73 and the National Human Genome Research Institute (NHGRI) GWAS catalog48 (accessed May 2014). As previous GWAS for risk factors were mainly based on HapMap 2–imputed data sets, all SNPs in LD (r2 > 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 data22,28,29,75–79.

Annotation and ENCODE analysis. Variants were annotated using ANNOVAR software18 (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 genomic 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 genomic 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-experiment-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 genomic annotation status. Variants were grouped into three functional sets—IMs, 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 metabolism64), vascular endothelial cells (atherosclerosis81) and myoblasts (injury and repair82) 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 ENCODE23 project has previously mapped 4,492 significant GWAS SNPs from the NHGRI catalog48 (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.


75. Fehmann, 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).


