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

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Discovery Research Portal

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
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 casual 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 \(\times 10^{-4}\)). 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\(^1\). 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, \(P = 5.7 \times 10^{-3}\) 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\(^7,8\).
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\(^5\) and the Metachip 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 (\(q < 0.05\)) 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 (\(q < 0.05\)) 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)
were strikingly chromosome ideograms with centromeres represented by pink bars.

**Annotation and ENCODE analysis**

Functional annotations were assigned to the 9.4 million variants studied in the CAD additive meta-analysis using ANNOVAR software (Supplementary Table 8). The 202 FDR variants were depleted in intergenic regions (P = 2.5 × 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 × 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 (r² > 0.8) showed CAD associations (P < 1 × 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; REST encodes a transcription factor that suppresses the expression of voltage-dependent sodium and potassium channels; it has been shown to maintain vascular smooth muscle cells (VSMCs) in a quiescent, non-proliferative state and is itself downregulated in neointimal hyperplasia; in the 1000 Genomes Project 1 v3 training set met the restrictive power 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.

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 × 10⁻⁴) in a candidate gene meta-analysis based on 15,600
Table 1 Ten new CAD-associated loci

<table>
<thead>
<tr>
<th>Lead variant</th>
<th>Locus name</th>
<th>Chr.</th>
<th>A1/A2</th>
<th>Effect allele (A1) freq.</th>
<th>Imputation quality</th>
<th>Heterogeneity P</th>
<th>n studies*</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17087335</td>
<td>REST-NOA1</td>
<td>4</td>
<td>T/G</td>
<td>0.21</td>
<td>0.99</td>
<td>0.20</td>
<td>0.11</td>
<td>1.06 (1.04–1.09)</td>
<td>48</td>
</tr>
<tr>
<td>rs3918226</td>
<td>NOS3</td>
<td>7</td>
<td>T/C</td>
<td>0.06</td>
<td>0.78</td>
<td>0.15</td>
<td>0.19</td>
<td>1.14 (1.09–1.19)</td>
<td>45</td>
</tr>
<tr>
<td>rs10840293</td>
<td>SWAP70</td>
<td>11</td>
<td>A/G</td>
<td>0.55</td>
<td>0.94</td>
<td>0.17</td>
<td>0.16</td>
<td>1.06 (1.04–1.08)</td>
<td>47</td>
</tr>
<tr>
<td>rs56062135</td>
<td>SMAD3</td>
<td>15</td>
<td>C/T</td>
<td>0.79</td>
<td>0.98</td>
<td>0.00</td>
<td>0.67</td>
<td>1.07 (1.05–1.10)</td>
<td>46</td>
</tr>
<tr>
<td>rs8042271</td>
<td>MGAT2-ABHD2</td>
<td>15</td>
<td>G/A</td>
<td>0.9</td>
<td>0.93</td>
<td>0.16</td>
<td>0.21</td>
<td>1.06 (1.04–1.14)</td>
<td>46</td>
</tr>
<tr>
<td>rs7121798</td>
<td>BCAS3</td>
<td>22</td>
<td>G/T</td>
<td>0.97</td>
<td>0.86</td>
<td>0.00</td>
<td>0.67</td>
<td>1.06 (1.04–1.08)</td>
<td>46</td>
</tr>
<tr>
<td>rs11830157</td>
<td>KSR2</td>
<td>12</td>
<td>G/T</td>
<td>0.36</td>
<td>0.99</td>
<td>0.14</td>
<td>0.22</td>
<td>1.04 (1.02–1.06)</td>
<td>42</td>
</tr>
<tr>
<td>rs12976411</td>
<td>ZNF507-LOC400684</td>
<td>19</td>
<td>T/A</td>
<td>0.09</td>
<td>0.93</td>
<td>0.50</td>
<td>5.09 × 10⁻⁴</td>
<td>3.30 × 10⁻⁸</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Association results are presented for two inheritance models; results from the discovery association model are shown in bold. Chr., chromosome; A1, effect allele; A2, non-effect allele; freq., frequency; I², 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.

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. SMAD3 was tentatively associated with CAD in an earlier GWAS, although the lead SNP (rs17228212) in that association is in linkage
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 (MFGES-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 (r²) 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 (MFGES-ABHD2) locus, the lead intergenic SNP rs8042271 maps 117 kb upstream of MFGES (milk fat globule-EGF factor 8) and 57 kb upstream of ABHD2 (abhydrolase domain-containing protein 2) (Fig. 4e). MFGES (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 engulfment.15 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 x 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 SPECCIL 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.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 intrinsic 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, 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.44 However, the additive model is systematically underpowered to detect recessively inherited traits, particularly with lower-frequency alleles.45 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 traits48 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. We confirmed that ABO is particularly associated with risk of myocardial infarction, 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. 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 few directly target vessel wall processes. Detailed investigating cessation), but the large majority of existing drug treatments for strategies target the vessel wall (control of blood pressure and smoking). Links to vascular biology can be established. A number of preventative require substantial further investigation and validation before firm (signaling) ABHD2. Their gene products are involved in diverse processes, including cell genes so implicated have well-documented roles in vessel wall biology. eQTL associations suggest a plausible candidate gene for CAD, the newly identified loci where genomic data, biological precedent and so at very large scale. In this respect, it is notable that, for some of the completely unbiased approach and, to our knowledge, is the first to do based CAD GWAS meta-analysis. 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 (REST20), TGF-β signaling (SMAD3 (refs. 56,57)), anti-inflammatory and infarct-sparing effects (ADORA2A58 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.


© 2015 Nature America, Inc. All rights reserved.


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. 3Welcome 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. 6Center for 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. 9Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, Lübeck, Germany. 10DZHK (German Center 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. 12Perelman 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 (FIMM), University of Helsinki, Finland, 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. 27State Key Laboratory 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. 34Welcome 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 Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. 41Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland. 42Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA. 43Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 44Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, California, USA. 45Department of Anesthesiology, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. 46Department of Medical Informatics, Statistics and Epidemiology, Medical Faculty, University of Leipzig, Leipzig, Germany. 47Friedrich Schiller University Jena, School of Computer Science and Engineering, Jena, 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. 60INSERM, UMR S1138, Centre de Recherche des Cordeliers, Paris, France. 61Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. 62School of Medicine, Lebanese American University, Beirut, Lebanon. 63Hypertension Division, Fujui 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.
These authors contributed equally to this work. These authors jointly supervised this work. Correspondence should be addressed to H.W., S. Kathiresan, R.M.C., or M.F.
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 method62 (Supplementary Table 15), and adjusted values were submitted for meta-analysis. Varians that were retained in at least 60% of the studies were submitted for meta-analysis using the GWAMA program63. Following an inverse variance-weighted fixed-effects meta-analysis, heterogeneity was assessed by Cochran’s Q statistic64 and the I² inconsistency index65, and variants showing marked heterogeneity were reanalyzed using a random-effects model66. 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 p.adjust function (R package). The FDR was assessed using a step-up procedure coded in the R package. Supplementary Fig. 7). This comparison showed that 95% of the β values (regression coefficients) and standard errors were accurately approximated. The –log10(P values) from the two analyses were positively correlated (0.86 < p < 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 model67 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 x 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 downloads71–73 and the National Human Genome Research Institute (NHGRI) GWAS catalog74 (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 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 Guevadis Data Browser (accessed June 2014) and other published data22,28,29,75–79.

Annotation and ENCODE analysis. Varians 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 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 Funegen 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=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 metabolism68), vascular endothelial cells (atherosclerosis41) and myoblasts (injury and repair69) 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 project has previously mapped 4,492 significant GWAS SNPs from the NHGRI catalog24 (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.