

# Large-scale association analysis identifies new risk loci for coronary artery disease

The CARDIoGRAMplusC4D Consortium<sup>1</sup>

Coronary artery disease (CAD) is the commonest cause of death. Here, we report an association analysis in 63,746 CAD cases and 130,681 controls identifying 15 loci reaching genome-wide significance, taking the number of susceptibility loci for CAD to 46, and a further 104 independent variants ( $r^2 < 0.2$ ) strongly associated with CAD at a 5% false discovery rate (FDR). Together, these variants explain approximately 10.6% of CAD heritability. Of the 46 genome-wide significant lead SNPs, 12 show a significant association with a lipid trait, and 5 show a significant association with blood pressure, but none is significantly associated with diabetes. Network analysis with 233 candidate genes (loci at 10% FDR) generated 5 interaction networks comprising 85% of these putative genes involved in CAD. The four most significant pathways mapping to these networks are linked to lipid metabolism and inflammation, underscoring the causal role of these activities in the genetic etiology of CAD. Our study provides insights into the genetic basis of CAD and identifies key biological pathways.

Coronary artery disease and its main complication, myocardial infarction, is the leading cause of death worldwide. Although, epidemiological studies have identified many risk factors for CAD, including plasma lipid concentrations, blood pressure, smoking, diabetes and markers of inflammation, a causal role has been proven only for some (for example, low-density lipoprotein (LDL) cholesterol and blood pressure), primarily through randomized clinical trials of drug therapy directed at the risk factor<sup>1</sup>. Twin and family studies have documented that a significant proportion (40–50%) of susceptibility to CAD is heritable (for a review, see ref. 2). Because genotypes are not confounded by environmental exposures, genetic analysis has the potential to define which risk factors are indeed causal and to identify pathways and therapeutic targets<sup>3,4</sup>. To date, genome-wide association studies (GWAS) have collectively reported a total of 31 loci, associated with CAD risk at genome-wide significance ( $P < 5 \times 10^{-8}$ )<sup>5–13</sup>. However, variants at these loci explain less than 10% of the heritability of CAD. One likely reason for this is that, given the polygenic nature of complex traits and the relatively small observed effect sizes of the loci identified, many genuinely associated variants do not reach the stringent  $P$ -value threshold for genome-wide significance. Indeed, there is increasing evidence that the genetic architecture of common traits involves a large number of causative alleles with very small effects<sup>14</sup>. Addressing this will require the discovery of additional loci while leveraging large-scale genomic data to identify the molecular pathways underlying the pathogenesis of CAD. Such discovery is facilitated by building molecular networks, on the basis of DNA, RNA and protein interactions, which have nodes of known biological function that also show evidence of association with risk variants for CAD and related metabolic traits.

In the largest GWAS meta-analysis of CAD undertaken to date by the Coronary ARtery Disease Genome-wide Replication and

Meta-analysis (CARDIoGRAM) Consortium<sup>5</sup>, which involved 22,233 cases and 64,762 controls, in addition to loci reported at genome-wide significance, a linkage disequilibrium (LD)-pruned set of 6,222 variants achieved a nominal association  $P$  value of less than 0.01. Here, we test these 6,222 SNPs in a meta-analysis of over 190,000 individuals, with the primary aim of identifying additional susceptibility loci for CAD. To this end, we used the Metachip array<sup>15</sup>, which is a custom iSELECT chip (Illumina) containing 196,725 SNPs, designed to (i) follow-up putative associations in several cardiometabolic traits, including CAD, and (ii) fine map confirmed loci for these traits. All SNPs on the array with data in the CARDIoGRAM study were considered for analysis (79,138 SNPs, of which 6,222 were the replication SNPs and 20,876 were fine-mapping SNPs in the 22 CAD susceptibility loci identified at the time at which the array was designed; the remaining SNPs were submitted by the other consortia contributing to the Metachip array<sup>15</sup>). In addition, we assess whether the genome-wide significant CAD risk alleles act through traditional risk factors by considering the available large GWAS for these traits<sup>16–20</sup>. Finally, we identify a broader set of SNPs passing a conservative FDR threshold for association with CAD and use this set to undertake network analysis to find key biological pathways underlying the pathogenesis of CAD.

## RESULTS

### Study design

We expanded the CARDIoGRAM discovery data set (22,233 cases and 64,762 controls<sup>5</sup>, stage 1) with 34 additional CAD sample collections (stage 2) of European or south Asian descent comprising 41,513 cases and 65,919 controls (study descriptions and sample characteristics are given in **Supplementary Tables 1a** and **2a**, respectively) and undertook a 2-stage meta-analysis to test SNPs on the Metachip array

<sup>1</sup>A full list of authors and affiliations appears at the end of the paper.

Received 24 April; accepted 2 November; published online 2 December 2012; doi:10.1038/ng.2480

for disease association in a total of 63,746 cases and 130,681 controls. A further set of 3,630 cases and 11,983 controls from 4 independent studies was used for replication of SNPs that reached  $5 \times 10^{-8} < P < 1 \times 10^{-6}$  in combined stage 1 and 2 analysis (stage 3; **Supplementary Tables 1b** and **2b**). An overview of the study design is provided in **Supplementary Figure 1**. Cases were selected for inclusion following the standard criteria for CAD and myocardial infarction used in the CARDIoGRAM study<sup>5</sup> (details for the stage 2 and 3 cohorts are given in **Supplementary Table 2**). Collections were typed with either the MetaboChip array (60% of samples) or provided GWAS data imputed using HapMap (Supplementary Table 3). We applied standard quality control criteria to each study and corrected for population stratification if  $\lambda_{GC}$  was  $\geq 1.05$  (estimated for samples typed on the MetaboChip using 4,310 SNPs associated with long QT syndrome and located at least 5 Mb away from established CAD risk loci; Online Methods). Case-control association analyses were adjusted for sex and age. For the 79,138 SNPs on the MetaboChip with both stage 1 and 2 data, we combined (2-sided) *P* values from stage 1 with their respective (1-sided) *P* values for stage 2 using Fisher's method (Online Methods). In stage 3, we validated SNPs at  $5 \times 10^{-8} < P < 1 \times 10^{-6}$  and combined evidence across all stages (1–3) using a sample size-weighted meta-analysis.

### Genome-wide significant loci

We first examined the 30 CAD risk loci previously reported in individuals of European ancestry at genome-wide significance (the *ADTRP* (*C6orf105*) locus has been reported only in Chinese)<sup>12</sup> in the stage 2 samples. For the 26 loci in which we could test the known lead SNP or a suitable proxy ( $r^2 > 0.8$ ), we found highly significant associations in the stage 2 samples (**Table 1**). Notably, in four of these loci (*CDKN2B-AS1*, *COL4A2*, *CXCL12* and *APOE*), we detected additional SNPs not in LD ( $r^2 < 0.5$ ) with the lead SNP, which also reached genome-wide significance and were conditionally independent when analyzed with GCTA software<sup>21</sup>. The additional SNP in the *APOE* locus, rs445925 ( $P = 9.42 \times 10^{-11}$ ;  $r^2 = 0.015$  with rs207560 in 1000 Genomes Project data), is located near *APOC1*, a gene previously suggested to confer risk for CAD<sup>22</sup>. The  $r^2$  value between rs445925 ( $P = 9.42 \times 10^{-11}$ ;  $n = 31$  studies) and rs7412 ( $P = 8.86 \times 10^{-4}$ ;  $n = 21$  studies), which tags the *APOE* e2 allele, is 0.588. The *LIPA* locus also harbors a strong independent signal, which, however, did not reach genome-wide significance. Findings for the strongest associated variant available on the MetaboChip for the other four loci (*MIA3*, 7q22, *ZNF259-APOA5-APOA1* and *ADAMTS7*) for which we did not have a good proxy for the previously reported lead SNP are also given (**Table 1**). Notably, for *ADAMTS7*, rs7173743 ( $r^2 = 0.38$  with rs3825807, the published lead SNP) also achieved genome-wide significance.

We next examined the association of the 6,222 SNPs with  $P < 0.01$  in CARDIoGRAM (we excluded SNPs in all loci listed in **Table 1**). Distribution of the absolute *z* scores for these SNPs in the stage 2 samples showed strong enrichment in positive scores corresponding to SNPs with directionally consistent signals between stages 1 and 2 under the null distribution, which is defined by mean = 0 and s.d. = 1 (4,260 SNPs observed versus 3,111 SNPs expected; binomial 2-sided  $P = 7.5 \times 10^{-187}$ ) (**Supplementary Fig. 2**). In total, 19 loci showed association at  $P < 1 \times 10^{-6}$  in the combined stage 1 and 2 analysis, with 13 of them reaching genome-wide significance, namely *IL6R*, *APOB*, *VAMP5-VAMP8-GGCX*, *SLC22A4-SLC22A5*, *ZEB2-AC074093.1*, *GUCY1A3*, *KCNK5*, *LPL*, *PLG*, *TRIB1*, *ABCG5-ABCG8*, *FURIN-FES* and *FLT1* (**Table 2**; Forest and regional association plots are given in **Supplementary Figs. 3** and **4**, respectively). The 6 loci with associations not reaching  $P < 5 \times 10^{-8}$  were further validated (stage 3) in 4

independent studies (3,630 cases and 11,983 controls; **Supplementary Table 1b**). Two loci, *EDNRA* and *HDAC9* replicated at  $P < 0.05$  and reached genome-wide significance in a combined analysis of stages 1–3 (**Table 2**); findings for those SNPs not meeting the above criteria are shown in **Supplementary Table 4**.

Of the newly associated loci reaching genome-wide significance, *TRIB1* and *ABCG5-ABCG8* were recently reported to reach study-wide significance ( $P < 3 \times 10^{-6}$ ) in a large candidate gene (IBC array) study of CAD<sup>13</sup>. The same study reported rs2706399 in the *IL5* locus, which is located 200,349 bp away from the SNP we detected in the *SLC22A4-SLC22A5* locus (rs273909; **Table 2**). Although located in the same recombination interval, these SNPs are not in LD ( $r^2 = 0.02$ ), and conditional analysis in a subset of 85,136 samples (up to 19,200 cases) suggested that the 2 signals are conditionally independent; when conditioning on rs2706399 (*IL5* locus), the *P* value for rs273909 (*SLC22A4* locus) was  $5.54 \times 10^{-3}$  ( $1.33 \times 10^{-3}$  initially), whereas the converse conditioning gave a *P* value of  $3.34 \times 10^{-2}$  for rs2706399 (*IL5*;  $7.55 \times 10^{-3}$  initially). We also detected a second signal in the *FES* locus (rs2521501;  $P = 1.31 \times 10^{-9}$ ); conditional analysis with rs17514846 and rs2521501 ( $r^2 = 0.43$  in 1000 Genomes Project data) showed the two signals not only to be independent but to also increase in strength upon conditioning (rs17514846 associated at  $P = 1.07 \times 10^{-25}$  when conditioned on rs2521501; conversely, the *P* value for rs2521501 was  $9.24 \times 10^{-26}$ ).

### Subgroup analyses

Genetic risk of CAD could vary by age and gender and could also specifically influence the risk of its main adverse outcome, myocardial infarction<sup>23</sup>. We therefore undertook exploratory association analyses in subgroups partitioned by either gender, age at event (with individuals of <50 years of age being defined as young cases) or history of myocardial infarction (Online Methods). For the 46 genome-wide significant CAD risk loci, we observed no trend for higher odds ratios (ORs) in any of the subgroup analyses (**Supplementary Table 5**). However, one new locus reached genome-wide significance in males and in young CAD cases (rs16986953;  $P = 1.89 \times 10^{-8}$  and  $1.67 \times 10^{-8}$ , respectively), which is located in a gene desert (with nearest transcript AK097927), 1.3 Mb away from the *APOB* gene. Interaction analysis conducted in a subset of studies ( $n = 12$ ) where we had individual-level data provided suggestive evidence of an association with age ( $P = 0.033$ ) but not with sex ( $P = 0.708$ ); further studies are required to confirm this finding.

### Wider MetaboChip content

In addition to SNPs provided by the CARDIoGRAM Consortium, the MetaboChip array contains a further 113,248 SNPs submitted for a range of cardiometabolic traits<sup>15</sup> other than CAD itself (associated at  $P > 0.01$  with CAD in CARDIoGRAM samples or not tested). For these SNPs, we did not detect any new locus reaching genome-wide significance in our data set (including stage 1 and 3 data, when available). In total, therefore, we discovered 15 newly associated loci at genome-wide significance, increasing the total number of genome-wide significant loci to 45 in individuals of European and south Asian ancestry.

### Localizing candidate CAD genes

To identify potential causal CAD-associated genes at the 15 new susceptibility loci identified in our study, we first analyzed genome-wide expression quantitative trait locus (eQTL) data in multiple tissues (circulating monocytes, liver, fat, skin, omentum, aortic media and adventitia, mammary artery and lymphoblastoid cell lines (LCLs)). We found that the lead SNP or a proxy in high LD ( $r^2 \geq 0.8$ ) in three of the new loci was associated in *cis* with variable expression levels of the *GGCX-VAMP8*, *PLG* and *FES* genes (**Supplementary Table 6**).

**Table 1 Association findings for known CAD susceptibility loci**

Known loci <sup>a</sup>	Published lead SNP or proxy	New SNP ( $r^2$ with lead SNP)	Chr.	Effect/non-effect allele (frequency)	Stage 2 OR	Stage 2 $P$	Combined $P$	Combined OR
<i>SORT1</i> <sup>b</sup>	rs602633 (tagging rs599839; $r^2 = 1.00$ )		1	C/A (0.77)	1.13	$2.19 \times 10^{-18}$	$1.47 \times 10^{-25}$	1.12
<i>PCSK9</i>	rs11206510		1	T/C (0.84)	1.04	$5.09 \times 10^{-3}$	$1.79 \times 10^{-5}$	1.06
<i>WDR12</i>	rs6725887		2	C/T (0.11)	1.10	$5.29 \times 10^{-8}$	$1.16 \times 10^{-15}$	1.12
<i>MRAS</i>	rs9818870		3	T/C (0.14)	1.05	$1.83 \times 10^{-3}$	$2.62 \times 10^{-9}$	1.07
<i>TCF21</i>	rs12190287		6	C/G (0.59)	1.04	$6.48 \times 10^{-4}$	$4.94 \times 10^{-13}$	1.07
<i>SLC22A3-LPAL2-LPA</i>	rs3798220		6	C/T (0.01)	1.28	$4.90 \times 10^{-5}$	N/A	N/A
		rs2048327 (0.03)	6	G/A (0.35)	1.05	$1.09 \times 10^{-5}$	$6.86 \times 10^{-11}$	1.06
<i>ZC3HC1</i>	rs11556924		7	C/T (0.65)	1.08	$1.45 \times 10^{-9}$	$6.74 \times 10^{-17}$	1.09
<i>CDKN2BAS1</i>	rs1333049		9	C/G (0.47)	1.21	$1.08 \times 10^{-34}$	$1.39 \times 10^{-52}$	1.23
		rs3217992 (0.50)	9	A/G (0.38)	1.14	$7.27 \times 10^{-32}$	$7.75 \times 10^{-57}$	1.16
<i>ABO</i>	rs579459		9	G/T (0.21)	1.04	$2.13 \times 10^{-2}$	$2.66 \times 10^{-8}$	1.07
<i>CYP17A1-CNNM2-NT5C2</i>	rs12413409		10	C/A (0.89)	1.08	$4.12 \times 10^{-3}$	$6.26 \times 10^{-8}$	1.10
<i>KIAA1462</i>	rs2505083		10	C/T (0.42)	1.06	$2.82 \times 10^{-7}$	$1.35 \times 10^{-11}$	1.06
<i>PDGFD</i>	rs974819		11	A/G (0.29)	1.08	$2.03 \times 10^{-9}$	$3.55 \times 10^{-11}$	1.07
<i>SH2B3</i>	rs3184504		12	T/C (0.40)	1.07	$6.13 \times 10^{-7}$	$5.44 \times 10^{-11}$	1.07
<i>COL4A1-COL4A2</i>	rs4773144		13	G/A (0.42)	1.06	$2.34 \times 10^{-6}$	$1.43 \times 10^{-11}$	1.07
		rs9515203 (0.01)	13	T/C (0.74)	1.08	$1.13 \times 10^{-8}$	$5.85 \times 10^{-12}$	1.08
<i>HHIPL1</i>	rs2895811		14	C/T (0.43)	1.04	$1.18 \times 10^{-4}$	$4.08 \times 10^{-10}$	1.06
<i>RAI1-PEMT-RASD1</i>	rs12936587		17	G/A (0.59)	1.04	$2.06 \times 10^{-4}$	$1.24 \times 10^{-9}$	1.06
<i>LDLR</i>	rs1122608		19	G/T (0.76)	1.06	$3.72 \times 10^{-6}$	$6.33 \times 10^{-14}$	1.10
Gene desert ( <i>KCNE2</i> )	rs9982601		21	T/C (0.13)	1.10	$8.69 \times 10^{-9}$	$7.67 \times 10^{-17}$	1.13
<i>PPAP2B</i>	rs17114036		1	A/G (0.91)	1.09	$2.68 \times 10^{-5}$	$5.80 \times 10^{-12}$	1.11
<i>ANKS1A</i>	rs12205331 (tagging rs17609940; $r^2 = 0.85$ )		6	C/T (0.81)	1.01	$4.36 \times 10^{-1}$	$4.18 \times 10^{-5}$	1.04
<i>PHACTR1</i>	rs9369640 (tagging rs12526453; $r^2 = 0.90$ )		6	A/C (0.65)	1.09	$1.11 \times 10^{-12}$	$7.53 \times 10^{-22}$	1.09
<i>CXCL12</i>	rs501120		10	A/G (0.83)	1.06	$7.13 \times 10^{-5}$	$1.79 \times 10^{-9}$	1.07
		rs2047009 (0.05)	10	C/A (0.48)	1.05	$9.66 \times 10^{-6}$	$1.59 \times 10^{-9}$	1.05
<i>LIPA</i>	rs2246833 (tagging rs1412444; $r^2 = 0.98$ )		10	T/C (0.38)	1.04	$2.76 \times 10^{-2}$	$9.49 \times 10^{-6}$	1.06
		rs11203042 (0.39)	10	T/C (0.44)	1.03	$9.86 \times 10^{-3}$	$6.08 \times 10^{-6}$	1.04
<i>UBE2Z</i>	rs15563 (tagging rs46522; $r^2 = 0.93$ )		17	C/T (0.52)	1.01	$2.44 \times 10^{-1}$	$9.37 \times 10^{-6}$	1.04
<i>SMG6</i>	rs2281727 (tagging rs216172; $r^2 = 0.96$ )		17	C/T (0.36)	1.04	$8.46 \times 10^{-4}$	$7.83 \times 10^{-9}$	1.05
<i>ApoE-ApoC1</i>	rs2075650		19	G/A (0.14)	1.11	$5.86 \times 10^{-11}$	N/A	N/A
		rs445925 (0.03)	19	C/T (0.90)	1.13	$8.76 \times 10^{-9}$	N/A	N/A
<i>MIA3</i>	N/A	rs17464857 (0.18)	1	T/G (0.87)	1.02	$1.56 \times 10^{-1}$	$6.06 \times 10^{-5}$	1.05
7q22	N/A	rs12539895 (0.64)	7	A/C (0.19)	1.02	$4.00 \times 10^{-2}$	$5.33 \times 10^{-4}$	1.08
<i>ZNF259-APOA5-APOA1</i>	N/A	rs9326246 (0.63)	11	C/G (0.10)	1.04	$2.90 \times 10^{-2}$	$1.51 \times 10^{-7}$	1.09
<i>ADAMTS7</i>	N/A	rs7173743 (0.38)	15	T/C (0.58)	1.06	$2.46 \times 10^{-7}$	$6.74 \times 10^{-13}$	1.07

Chr., chromosome.

<sup>a</sup>Locus *C6orf105*, which has been reported only in Chinese and has no good proxy SNP (Utah residents of Northern and Western European ancestry (CEU) or Han Chinese in Beijing, China (CHB)) on the Metachip. The best available proxy is rs9348953 ( $r^2 = 0.01$ ), with combined  $P = 2.81 \times 10^{-3}$ . <sup>b</sup>rs12740374, which was reported as a functional variant in this locus and has  $r^2 = 0.895$  with rs599839, has combined  $P = 8.25 \times 10^{-18}$  (OR = 1.135) based on the random-effects model used ( $P$  in stage 2 alone was  $6.48 \times 10^{-21}$  under the fixed-effect model).

We then assessed allele-specific expression data in monocytes, fibroblasts and LCLs and found three loci where the lead SNP was associated with an imbalance in expression of either *LPL*, *GGCX* or *FES*; *IL6R* showed some evidence of allele-specific expression in the fibroblast sample (Supplementary Table 6). Finally, we examined the new CAD risk loci for genes with relevant disease trait associations in mouse knockout models; six loci harbor a gene for which a mouse knockout model has a relevant cardiovascular phenotype, namely *ABCG8*, *APOB*, *GUCY1A3*, *PLG*, *LPL* and *FES* (Supplementary Table 7). *PLG* is adjacent to *LPA*, and, although the *PLG* risk variant rs4252120[T] was strongly associated with elevated Lp(a) lipoprotein levels ( $P = 5 \times 10^{-24}$ ) in 3,698 PROCARDIS cases, it was associated with CAD independent of the *LPA*-linked variant at rs3798220. A detailed discussion of the genes in each locus is provided in the

**Supplementary Note.** Of the 30 previously reported CAD susceptibility loci in individuals of European and south Asian ancestry, mouse knockout models for the candidate genes *PEMT*, *APOE*, *LDLR*, *COL4A1*, *LIPA*, *APOA1-APOA5*, *PPAP2B* and *PCSK9* also show phenotypic characteristics directly relevant to disease (Supplementary Table 7). In total, approximately a third of the 45 CAD loci contain a known functionally relevant candidate gene.

### Overlap with traditional risk factors

We assessed both the known and new CAD susceptibility loci for overlap of associations with a number of relevant traits for which summary statistics have been made available: lipid levels (GLGC)<sup>16</sup>, blood pressure (ICBPG)<sup>17</sup>, diabetes (DIAGRAM)<sup>18</sup>, glucometabolic traits (fasting insulin and fasting glucose concentrations, HOMA-B

**Table 2 Additional loci showing genome-wide significant association with CAD**

SNP	Chr.	Nearest gene(s)	Effect/non-effect allele (frequency)	Stage 1 (18,014 cases and 40,925 controls) <sup>a</sup>		Stage 2 (40,365 cases and 63,714 controls)		Combined (stages 1 and 2)	Stage 3 (5,055 cases and 5,617 controls)		Combined (stages 1–3)	Biological relevance <sup>b</sup>
				OR	P	OR	P	P	OR	P	P	
<b>New</b>												
rs4845625	1	<i>IL6R</i>	T/C (0.47)	1.06	$4.84 \times 10^{-5}$	1.04	$3.46 \times 10^{-5}$	$3.55 \times 10^{-8}$	1.09	$1.58 \times 10^{-3}$	$3.64 \times 10^{-10}$	2
rs515135	2	<i>APOB</i>	G/A (0.83)	1.07	$8.63 \times 10^{-4}$	1.08	$2.17 \times 10^{-8}$	$4.80 \times 10^{-10}$	1.03	$4.02 \times 10^{-1}$	$2.56 \times 10^{-10}$	1
rs2252641	2	<i>ZEB2-AC074093.1</i>	G/A (0.46)	1.06	$1.37 \times 10^{-5}$	1.04	$1.27 \times 10^{-4}$	$3.66 \times 10^{-8}$	1.00	$9.54 \times 10^{-1}$	$5.30 \times 10^{-8}$	
rs1561198	2	<i>VAMP5-VAMP8-GGCX</i>	A/G (0.45)	1.06	$7.47 \times 10^{-5}$	1.05	$2.57 \times 10^{-6}$	$4.48 \times 10^{-9}$	1.07	$1.75 \times 10^{-2}$	$1.22 \times 10^{-10}$	A,1
rs7692387	4	<i>GUCY1A3</i>	G/A (0.81)	1.08	$1.04 \times 10^{-5}$	1.06	$1.89 \times 10^{-5}$	$4.57 \times 10^{-9}$	1.13	$5.47 \times 10^{-4}$	$2.65 \times 10^{-11}$	1
rs273909	5	<i>SLC22A4-SLC22A5</i>	C/T (0.14)	1.07	$3.24 \times 10^{-3}$	1.09	$2.00 \times 10^{-7}$	$1.43 \times 10^{-8}$	1.11	$2.43 \times 10^{-2}$	$9.62 \times 10^{-10}$	A,1
rs10947789	6	<i>KCNK5</i>	T/C (0.76)	1.07	$6.07 \times 10^{-5}$	1.06	$1.22 \times 10^{-5}$	$1.63 \times 10^{-8}$	1.01	$7.03 \times 10^{-1}$	$9.81 \times 10^{-9}$	3
rs4252120	6	<i>PLG</i>	T/C (0.73)	1.07	$1.18 \times 10^{-5}$	1.06	$1.82 \times 10^{-5}$	$5.00 \times 10^{-9}$	1.07	$9.58 \times 10^{-2}$	$4.88 \times 10^{-10}$	1
rs264	8	<i>LPL</i>	G/A (0.86)	1.11	$2.99 \times 10^{-7}$	1.05	$7.30 \times 10^{-4}$	$5.06 \times 10^{-9}$	1.06	$1.60 \times 10^{-1}$	$2.88 \times 10^{-9}$	1
rs9319428	13	<i>FLT1</i>	A/G (0.32)	1.06	$7.88 \times 10^{-5}$	1.05	$5.70 \times 10^{-6}$	$1.01 \times 10^{-8}$	1.10	$1.37 \times 10^{-3}$	$7.32 \times 10^{-11}$	1
rs17514846	15	<i>FURIN-FES</i>	A/C (0.44)	1.07	$2.37 \times 10^{-5}$	1.05	$7.35 \times 10^{-7}$	$4.49 \times 10^{-10}$	1.04	$3.02 \times 10^{-1}$	$9.33 \times 10^{-11}$	A,1
<b>Previously reported at array-wide level of significance (<math>P &lt; 3 \times 10^{-6}</math>)</b>												
Rs2954029	8	<i>TRIB1</i>	A/T (0.55)	1.06	$2.79 \times 10^{-5}$	1.04	$7.75 \times 10^{-5}$	$4.53 \times 10^{-8}$	1.05	$8.56 \times 10^{-2}$	$4.75 \times 10^{-9}$	4
Rs6544713	2	<i>ABCG5-ABCG8</i>	T/C (0.30)	1.06	$2.22 \times 10^{-4}$	1.06	$1.57 \times 10^{-7}$	$8.72 \times 10^{-10}$	0.96	$3.56 \times 10^{-1}$	$2.12 \times 10^{-9}$	1
<b>New (stage 3 replication)</b>												
Rs1878406	4	<i>EDNRA</i>	T/C (0.15)	1.10	$2.37 \times 10^{-6}$	1.06	$3.54 \times 10^{-3}$	$1.65 \times 10^{-7}$	1.09	$2.01 \times 10^{-2}$	$2.54 \times 10^{-8}$	1
Rs2023938	7	<i>HDAC9</i>	G/A (0.10)	1.08	$6.81 \times 10^{-4}$	1.07	$5.25 \times 10^{-5}$	$6.49 \times 10^{-7}$	1.13	$4.09 \times 10^{-2}$	$4.94 \times 10^{-8}$	1

<sup>a</sup>Total sample sizes do not include the CHARGE sample sizes. <sup>b</sup>A, *cis* eQTL in LCLs; 1, mouse model available with cardiovascular phenotype; 2, mouse model has homeostatic and immune phenotypes; 3, mouse model has respiratory, nervous system, mortality, aging, growth and renal phenotypes; 4, mouse model has growth and immune phenotypes.

(homeostatic model assessment- $\beta$  score) and HOMA-IR (insulin resistance); MAGIC<sup>19</sup> and anthropometric traits (GIANT)<sup>20,24</sup>. After applying a Bonferroni correction for the 51 independent CAD-associated alleles tested (44 loci; no data available for rs16986953 and rs2521501), 12 loci showed evidence of association ( $P < 1 \times 10^{-4}$ ) between the lead CAD risk SNP and 1 or more plasma lipid trait (total cholesterol, LDL cholesterol, high-density lipoprotein (HDL) cholesterol and triglyceride concentration) in the expected direction (the CAD risk allele was associated with higher total cholesterol, LDL cholesterol and triglyceride concentrations and lower HDL cholesterol concentration). These lead SNPs were most strongly associated with LDL cholesterol concentration at eight loci (*APOB*, *ABCG5-ABCG8*, *PCSK9*, *SORT1*, *ABO*, *LDLR*, *APOE* and *LPA*), with triglyceride concentration at two loci (*TRIB1* and the *APOA5* cluster) and with HDL cholesterol concentration at one locus (*ANKS1A*). There was near-equivalent association for triglyceride and HDL cholesterol concentrations at one locus (*LPL*). All loci except *LPA* and *ANKS1A* showed genome-wide significance for association with a lipid trait. These results underscore the importance of LDL cholesterol as a causal CAD risk factor (Supplementary Table 8). At the *SH2B3* locus, the CAD risk allele for rs3184504 was associated with both lower LDL cholesterol ( $P = 1.73 \times 10^{-9}$ ) and HDL cholesterol ( $P = 4.97 \times 10^{-6}$ ) concentration; one likely explanation is the presence of independent variants for CAD and LDL cholesterol. Two known CAD risk loci (*CYP17A1-NT5C2* and *SH2B3*) and two of the new CAD susceptibility loci (*GUCY1A3* and *FES*) have previously been associated with systolic (SBP) and diastolic (DBP) blood pressure<sup>17</sup>. Significant evidence for association with DBP was also observed for *ZC3HC1* (Supplementary Table 8). In contrast to the results for lipid concentration and blood pressure, there was no significant association of any of the loci tested with type 2 diabetes (T2D). Consistent with this observation, none of the assessed glucometabolic traits (fasting insulin and fasting glucose concentrations, HOMA-B and HOMA-IR) were related to these CAD variants (at the *ANKS1A* locus, it was not

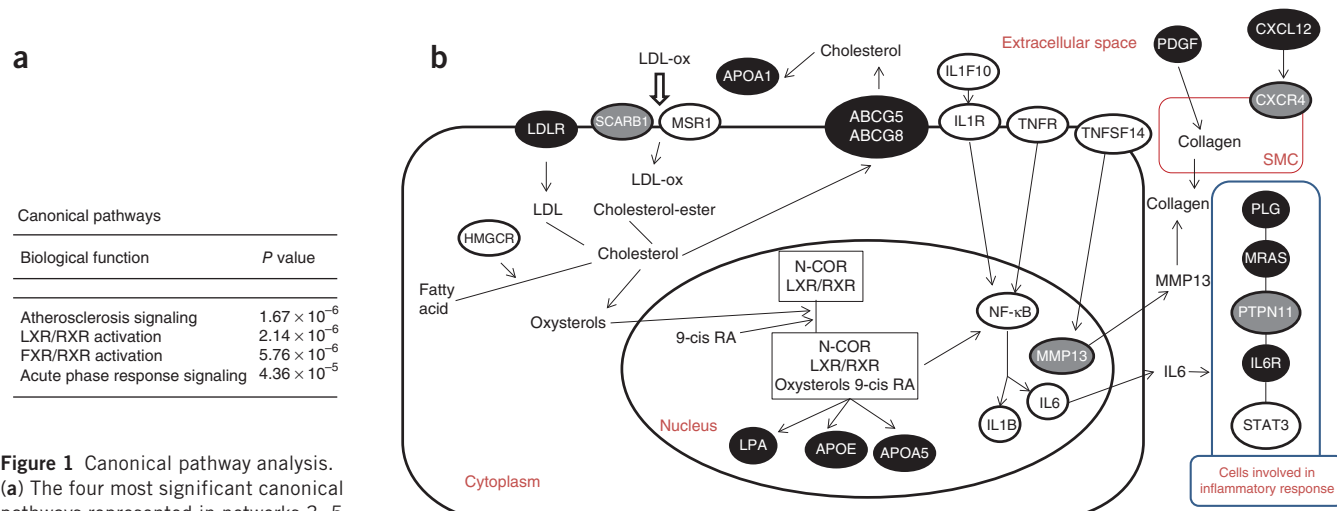
the CAD risk SNP that was associated with fasting insulin concentration and HOMA-IR). Suggestive associations ( $P < 1 \times 10^{-4}$ ) with body mass index (BMI) and waist-hip ratio were observed in the *CYP17A1-CNNM2-NT5C2* and *RAI1-PEMT-RASD1* loci, respectively.

### Additional suggestive associations

The genome-wide significance threshold,  $P < 5 \times 10^{-8}$ , we used is the accepted criterion for reporting individual association signals, as for each experiment it controls the error rate among common variants to less than 5%. However, SNPs showing suggestive association with a phenotype but not meeting this genome-wide threshold are likely to include additional true positive signals in well-powered studies (Supplementary Fig. 1). Such SNPs may also be informative in predicting CAD risk and in constructing CAD-associated biological networks. To identify such variants, we undertook an FDR analysis to assess the proportion of false positive signals in a set of (nominally) significant SNPs<sup>25</sup>. The MetaboChip array contains both SNPs with priors in terms of association to CAD (CARDIoGRAM study  $P < 0.01$ ) and blocks of highly correlated SNPs in fine-mapping regions. Therefore, to normalize the distribution of SNPs considered for FDR analysis, we (i) removed all SNPs in the CAD fine-mapping regions and LD-pruned ( $r^2 < 0.2$ ) SNPs in the non CAD fine-mapping regions and (ii) adjusted the combined  $P$  values of all SNPs with priors in stage 1 ( $P < 0.01$ ) using fixed-effect inverse variance-weighted meta-analysis  $P$  values for all other SNPs (Online Methods). In addition, we obtained 104 SNPs at an FDR threshold of 5% and LD threshold of  $r^2 < 0.2$  (Supplementary Table 9). The median OR for CAD for these SNPs was 1.054 (interquartile range of 0.0199) per risk allele (Supplementary Fig. 5).

On the basis of a heritability estimate of 40% for CAD, the combination of the known and newly associated SNPs within the 45 susceptibility loci (Tables 1 and 2) explains approximately 6% of the additive genetic variance of CAD. The addition of the 104 SNPs from FDR analysis increased the fraction explained to 10.6% (Online Methods).





**Figure 1** Canonical pathway analysis. **(a)** The four most significant canonical pathways represented in networks 3, 5 and 9, and overlapping networks ON1 (includes networks 1, 2, 6 and 8) and ON2 (includes networks 4 and 7); all molecules are listed by network in **Supplementary Table 10**. **(b)** Schematic showing parts of the atherosclerosis signaling, LXR/RXR activation and acute phase response signaling canonical pathways (Ingenuity) that are involved in both lipid metabolism and inflammation. Genes in confirmed CAD susceptibility loci (including both previously and newly reported) and in loci showing suggestive association with an FDR of  $<10\%$  are depicted as black and gray ovals, respectively. Other key genes are depicted as white ovals; notably, some of them, such as *IL1F10-IL1B*, *STAT3* and *HMGCR*, have SNPs ranking in the top 1,000 in the FDR analysis. The process leading to myocardial infarction involves multiple cell types that are depicted in this schematic as a composite cell (large oval) and its nucleus (inner oval) in the extracellular space; the smooth muscle cell is shown separately (SMC; red oval), whereas the blue oval depicts cell types involved in the inflammatory response.

## Network analysis

In contrast to estimating heritability where we want to keep the false positive rate as low as possible, in network analysis, we want to maximize the representation of potential network nodes in the gene set used. Thus, to perform network analysis, we selected the top 222 SNPs defined by the FDR analysis (10% FDR; final  $P < 6.6 \times 10^{-4}$ ) at an LD threshold of  $r^2 \leq 0.7$  and assigned 239 candidate genes on the basis of either eQTL data or physical proximity (**Supplementary Table 10**). We mapped 238 of the 239 genes in the Ingenuity Knowledge Base and considered 233 for network construction (Online Methods) on the basis of available data on interactions in humans, mice and/or rats (51 genes within the 46 genome-wide significant loci (set A) and 182 genes within the loci selected at FDR  $< 10\%$  (set B)). Including neighboring genes, Ingenuity generated 9 networks comprising 553 nodes; these included 48 (94.1%) of the genes in set A and 156 (85.7%) of those in set B (**Supplementary Table 10**). We obtained 2 overlapping networks: ON1, which included networks 1, 2, 6 and 8, comprising the majority of genes in both sets (33 and 83 in sets A and B, respectively), and ON2, which included networks 4 and 7 (**Supplementary Table 10**). The nine networks were strongly enriched for genes (query set) known to be involved in lipid metabolism ( $P = 1.48 \times 10^{-9}$ ), cellular movement (blood and endothelial cells;  $P = 1.35 \times 10^{-7}$ ) and processes such as tissue morphology (size and area of atherosclerotic lesion, quantity of leukocytes, macrophages and smooth muscle cells;  $P = 9.66 \times 10^{-10}$ ) and immune cell trafficking (migration and adhesion;  $P = 1.12 \times 10^{-7}$ ). As a negative control in the network analysis, we used a set of 368 genes selected from the least significant SNPs in the FDR analysis; the resulting networks showed no significant enrichment in relevant molecular functions and process (results described in detail in the **Supplementary Note**).

We then assessed how genes in the networks overlap with canonical pathways in the Ingenuity database. The four most significant canonical pathways represented in these networks are shown in **Figure 1a**. The top three pathways, atherosclerosis signaling, liver X receptor

(LXR)/retinoid X receptor (RXR) activation and farnesoid X receptor (FXR)/RXR activation, all harbor genes involved in lipid metabolism, including ten CAD risk loci (*ABCG5-ABCG8*, *APOA1*, *APOA5*, *APOB*, *APOE*, *CXCL12*, *LDLR*, *LPA*, *LPL* and *PDGFD*). This is in agreement with our finding that 12 CAD risk loci are associated with lipid levels at  $P < 1 \times 10^{-4}$  (**Supplementary Table 8**). Notably, three of the top four pathways also contain genes involved in inflammation. In addition to the atherosclerosis signaling and LXR/RXR activation pathways, the acute phase response signaling (AAPRS) pathway, which includes four CAD risk loci (*APOA1*, *MRAS*, *IL6R* and *PLG*), is involved in inflammation and, more specifically, the rapid inflammatory response that is triggered, among other factors, by tissue injury. Genes from both the lipid metabolism and inflammation-related pathways map to all networks, except network 9, which harbors only two genes (**Supplementary Table 10**). As shown for overlapping network ON1 (**Supplementary Fig. 6**), genes in lipid metabolism and inflammation are interconnected and include both CAD-associated loci reaching genome-wide significance and candidate loci at FDR  $< 10\%$ . Key interactions between CAD susceptibility genes (known, new and the FDR set) involved in lipid metabolism and inflammation are shown in **Figure 1b**; macrophages take up oxidized LDL (ox-LDL) through their cell surface scavenger receptors to form foam cells. Foam cells secrete proinflammatory cytokines, such as interleukin (IL)-1, IL-6 and matrix metalloproteinases, which can amplify the local inflammatory response and stimulate smooth muscle cell proliferation and initial migration toward the lesion<sup>26</sup>. Regulation of collagen secretion by smooth muscle cells in the extracellular matrix is regulated by matrix metalloproteinases. Reduction of collagen in the extracellular matrix will destabilize the plaque. Both *COL4A1* and *COL4A2* encode subunits of type IV collagen, which is the major structural component of basement membranes lining the inner surface of blood vessels. Metalloproteinases have a role in the maintenance of the extracellular matrix and remodeling, contributing to the transition of plaques from stable to vulnerable states (**Fig. 1b**).

## DISCUSSION

Here, we report the largest genetic study to date assessing the impact of common variation on CAD risk. On the basis of analyses involving 63,746 CAD cases and 130,681 controls, we identified 15 new risk alleles at genome-wide significance, bringing the total number of confirmed CAD susceptibility loci in individuals of European and south Asian ancestry to 45. We also identified a further set of 104 likely independent ( $r^2 < 0.2$ ) SNPs associated at an FDR of 5% with ORs between 1.031 and 1.126 per risk allele. In total, we estimate that these variants explain approximately 10.6% of the additive genetic variance of CAD (although we note that this may be an overestimate, given that it was not obtained in an independent sample). Our data also support the presence of additional true signals among the tested common SNPs that are likely to further contribute in explaining heritability; for example, the *P*-value adjustment we applied in the FDR analysis penalized the replication SNPs.

Among the 45 loci in individuals of European and south Asian ancestry that were confirmed to be associated with CAD, we found that 12 were significantly associated with the concentrations of blood lipids (mainly with LDL cholesterol), and 5 were associated with blood pressure. These data support the known etiological relationships of plasma lipids and blood pressure with CAD. People with T2D seem to have a 1.5- to 2-fold higher risk of CAD than those without diabetes<sup>27</sup>, but none of the 45 risk loci were associated with diabetes status or with continuous levels of various glucometabolic traits. We note that, for the binary variable of T2D status, inability to show associations with CAD risk loci may reflect limited statistical power. The temporal relationship for comorbidity with both diabetes and CAD is complex: individuals with CAD without diabetes at diagnosis often subsequently develop T2D<sup>28</sup>. Furthermore, despite clear benefits in preventing microvascular disease (for example, retinopathy and nephropathy), intensive glucose control in diabetics reduces the risk of cardiovascular disease relatively modestly<sup>29</sup>. However, before a final conclusion can be reached, as many cohorts contributing to this meta-analysis focused by design on early disease manifestation or excluded diabetic individuals, a formal testing of the relationship of T2D and CAD in Mendelian randomization experiments will be necessary. To this end, the large genetic association data set on CAD assembled here will also facilitate testing of the causal relationship of other putative risk factors for CAD.

A desirable clinical goal is to integrate genetic information into a risk score for CAD in an attempt to provide improved predictive power over traditional risk factors in asymptomatic subjects, such that preventative measures, where available, can be more appropriately targeted. Our findings provide an appropriate framework of 153 CAD risk variants (at those established as susceptibility loci meeting the genome-wide significance threshold and additional suggestive loci with an FDR of <5%) for assessing a genetic risk score in well-powered prospective studies to determine whether they are sufficiently informative and independent predictors to have potential for use in day-to-day practice.

Allowing for inherent limitations in selecting likely candidate genes at each locus, our network analysis identified lipid metabolism and inflammation as key biological pathways involved in the genetic pathogenesis of CAD. Indeed, there was significant crosstalk between the lipid metabolism and inflammation pathways identified (Fig. 1). The emergence of lipid metabolism as a key pathway provides a positive control for the network and pathway analysis. On the other hand, this analysis provides strong new evidence at the molecular level in support of the causal involvement of inflammatory mechanisms in the pathogenesis of coronary atherosclerosis<sup>30</sup>. The role of inflammation in atherosclerosis is well documented in the literature<sup>26</sup>;

for example, risk factors such as fat diet, smoking, hypertension, hyperglycemia, obesity and insulin resistance can trigger the expression of adhesion molecules (upregulated by atherogenic lipoproteins such as ox-LDL, very-low-density lipoprotein (VLDL) and Lp(a) lipoprotein) by endothelial cells, leading to the attachment of monocytes to the arterial wall. Although our analysis identified as significant the rapid inflammatory response pathway (mediated by NF- $\kappa$ B, MAPK and JAK-STAT signaling) that is primarily involved in innate immunity, many of the effector pathways in innate and adaptive immunity are heavily overlapping, and both are likely to have a role in CAD pathogenesis<sup>26</sup>. The five CAD-related networks constitute a useful framework for further functional and mechanistic studies to elucidate the biological processes underlying CAD pathogenesis and to investigate gene-environment interactions.

**URLs.** QVALUE software for FDR analysis, <http://genomics.princeton.edu/storeylab/qvalue/>; coronary heart disease statistics, <http://www.bhf.org.uk/publications/view-publication.aspx?ps=1002097>; top 10 causes of death fact sheet 310, <http://www.who.int/mediacentre/factsheets/fs310/en/index.html>; Uppsala Platform, <http://molmed.medsci.uu.se/SNP+SEQ+Technology+Platform/Genotyping>.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Summary statistics for the 79,138 SNPs considered in this study for association with CAD (SNPs with stage 1 and stage 2 data) are available at <ftp://ftp.sanger.ac.uk/pub/cardiogramplusc4d/>.

*Note: Supplementary information is available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank the personnel of the Wellcome Trust Sanger Institute (WTSI) Genotyping Facility, in particular S. Edkins, for supervising the genotyping of the AMC-PAS, Cardiogenics, GLACIER, MORGAM, PROMIS, THISEAS, and WTCCC cohorts.

## AMC-PAS/SANQUIN.

We thank A.A. Soussan for technical assistance.

We thank personnel from the Estonian Genome Center of the University of Tartu (EGCUT) and the Estonian Biocentre, especially M. Hass and V. Soo, for data generation.

## FINCAVAS.

We thank the staff of the Department of Clinical Physiology for collecting the exercise test data.

## The GLACIER Study.

The GLACIER study is a nested study within the Northern Sweden Health and Disease Study; phenotyping was conducted as part of the Västerbotten Intervention Project.

We thank the participants and the investigators from these studies for their valuable contributions, with specific thanks to L. Weinehall, Å. Agren, K. Enquist and T. Johansson.

## GoDARTS Dundee.

We are grateful to all the participants who took part in this study, to the general practitioners, to the Scottish School of Primary Care for their help in recruiting the participants and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We acknowledge the support of the Health Informatics Centre at the University of Dundee in managing and supplying the anonymized data and National Health Service (NHS) Tayside, the original data owner.

## Heart Protection Study.

The study was designed and conducted by the Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU) at the University of Oxford. Genotyping was supported by a grant to Oxford University and Centre National de Genotypage (CNG) from Merck. The funders had no role in the design of the study or in the data collection or analysis. We especially acknowledge the participants in the study,

the Steering Committee and our collaborators. J.C.H. acknowledges support from the British Heart Foundation (BHF) Centre of Research Excellence.

#### LOLIPOP.

We thank the participants and research staff who made the study possible.

#### MORGAM study.

We thank the contributing sites and key personnel, as detailed below.

**Finland:** We thank FINRISK, National Institute for Health and Welfare, Helsinki: V.S. (principal investigator), A. Juolevi, E. Vartiainen and P. Jousilahti; Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) study, National Institute for Health and Welfare, Helsinki: J. Virtamo (principal investigator) and H. Kilpeläinen; the MORGAM Data Centre, National Institute for Health and Welfare, Helsinki: K. Kuusasmaa (responsible person), Z. Cepaitis, A. Haukijärvi, B. Joseph, J. Karvanen, S. Kulathinal, M. Niemelä and O. Saarela; and the MORGAM Central Laboratory, National Institute for Health and Welfare, Helsinki: M.P. (responsible person), P. Laiho and M. Sauramo.

**France:** We thank the National Coordinating Centre, National Institute of Health and Medical Research (U258), Paris: P. Ducimetière (national coordinator) and A. Bingham; Prospective Epidemiological Study of Myocardial Infarction (PRIME)/Strasbourg, Department of Epidemiology and Public Health, EA 3430, University of Strasbourg, Faculty of Medicine, Strasbourg: D. Arveiler (principal investigator), B. Haas and A. Wagner; PRIME/Toulouse, Department of Epidemiology, Toulouse University School of Medicine, Toulouse: J.F. (principal investigator), J.-B. Ruidavets, V. Bongard, D. Deckers, C. Saulet and S. Barrere; PRIME/Lille, Department of Epidemiology and Public Health, INSERM U744–Université Lille Nord de France–Institut Pasteur de Lille, Lille: P. Amouyel (principal investigator), M. Montaye, L. Lemaire, S. Beauchant, D. Cottel, C. Graux, N. Marecaux, C. Steclebout and S. Szeremeta; and the MORGAM Laboratory, INSERM U937, Paris: F.C. (responsible person), L. Tiret and V. Nicaud.

**Italy:** We thank Centro Ricerche EPIMED–Epidemiologia e Medicina Preventiva, Dipartimento di Medicina Clinica e Sperimentale; Università dell'Insubria, Varese: M.M.F. (principal investigator) and G. Veronesi; and Research Centre on Public Health, University of Milano–Bicocca, Monza: G. Cesana.

**UK:** We thank PRIME/Belfast, Queen's University Belfast, Belfast: F.K. (principal investigator), A.E. (former principal investigator), J. Yarnell and E. Gardner; and the MORGAM Coordinating Centre, Queen's University Belfast, Belfast: A.E. (MORGAM coordinator), S. Cashman and F.K. MORGAM management group: A.E. (chair), S.S.B., F.C., M.M.F., K. Kuusasmaa, A. Palotie, M.P., A.P., V.S., H. Tunstall-Pedoe and P.G. Wiklund. Previous members: K. Asplund, L. Peltonen, D. Shields and B. Stegmayr. The PRIME Study is organized under an agreement between INSERM and the Merck, Sharpe and Dohme-Chibret Laboratory, with the following participating laboratories: The Strasbourg MONICA Project, Laboratoire d'Epidémiologie et de Santé Publique, and the Université de Strasbourg, Strasbourg, France (D. Arveiler and B. Haas); The Toulouse MONICA Project, UMR INSERM 1027, and the Department of Epidemiology, Toulouse University School of Medicine, Université Paul Sabatier, Toulouse, France (J.F. and J.-B. Ruidavets); The Lille MONICA Project, INSERM U744, Institut Pasteur de Lille and Université Lille Nord de France, Lille, France (P. Amouyel and M. Montaye); The Department of Epidemiology and Public Health, Queen's University, Belfast, Belfast, UK (A.E., J. Yarnell and F.K.); The Department of Atherosclerosis, INSERM U545, Institut Pasteur de Lille, Faculté de Médecine and Université Lille Nord de France, Lille, France (G. Luc and J.-M. Bard); The Laboratory of Haematology, INSERM U626, and Hôpital La Timone, Marseille, France (I. Juhan-Vague and P. Morange); The Laboratory of Endocrinology, INSERM U563, Toulouse, France (B. Perret); The Vitamin Research Unit, The University of Bern, Bern, Switzerland (F. Gey); The Nutrition and Metabolism Group, Centre for Public Health, Queen's University Belfast, Belfast, UK (J. Woodside and I. Young); The DNA Bank, INSERM/Université Pierre et Marie Curie (UPMC), Paris Université Unite Mixte de Recherche (UMRS) 937, Paris (F.C.); The Coordinating Centre, Institut Fédératif de Recherche Santé Publique (IFR 69), Villejuif, France (P. Ducimetière); and INSERM U970, Villejuif, France, and University Paris V, Paris Cardiovascular Research Centre (PAARC), Paris (A. Bingham).

#### PIVUS/Swedish Twin Registry.

We thank the SNP&SEQ Technology Platform in Uppsala (see URLs) for genotyping, in particular T. Axelsson, A.-C. Wiman and C. Pöntinen for excellent assistance.

#### Ulm (EMIL).

We thank the Centre of Excellence Baden-Wuerttemberg Metabolic Disorders.

#### WTCCC.

We thank the BHF Family Heart Study Research Group for the collection of the cases.

#### AUTHOR CONTRIBUTIONS

**Writing committee:** P. Deloukas, S. Kanoni, C.W., M.F., T.L.A., J.R.T., E.I., D. Saleheen, J.E., M.P. Reilly, R. Collins, S. Kathiresan, A.H., U.T., J.S.K., J.D., C.N.A.P., R.R., H.W., H.S. and N.J.S. **Steering committee:** P. Deloukas, S. Kanoni, C.W., M.F., T.L.A., J.R.T., E.I., D. Saleheen, J.E., M.P. Reilly, R. Collins, S. Kathiresan, A.H., U.T., J.S.K., J.D., C.N.A.P., R.R., H.W., H.S., N.J.S., S.S.B., B.O.B., J.C.C., R. Clarke, G.D., P.W.F., C.H., G.K.H., Jong-Young Lee, T.L., W.M., A.M., M.S.N., C.O., M.P., S. Ripatti, M.S.S., S.S., A. Siegbahn, C.J.W. and P.A.Z. **Analysis committee:** B.A.G., K. Stirrups, I.R.K., J.-B.C., Å.J., T.E., L.F., A.G., A.S. Havulinna, W.K.H., J.C.H., N.E., M.E.K., K. Kristiansson, P.L., L.-P.L., S. Rafelt, D. Shungin, R.J.S., G. Thorleifsson, E.T., N.V.Z., B.F.V., L.L.W., W.Z. and A.Z. **Genotyping:** D. Absher, I.B., C.B., S.C.-B., DIAGRAM Consortium, N.E.M., K.F., P.F., B.G., L.G., S.G., J.H., B.-G.H., S.E.H., T.K., J.W.K., C. Langenberg, C. Langford, M.I.M., M.M.-N., K.N., J.F.P., S. Rosinger, D.R., M.P. Rumpf, A. Schäfer, A.F.R.S., P.J.W. and Wellcome Trust Case Control Consortium. **Array design:** H.M.K. and N.W.R. **Functional analyses:** E.G., P.E., A.F.-C., A.L., O.M., S.M., MuTHER Consortium, T.-P.Y., A.H.G., E.S., T.P. and A.-C.S. **Samples and phenotyping:** (ADVANCE) A.S.G., C.I. and T.Q.; (AMC-PAS/SANQUIN) C.E.v.d.S. and H.B.; (Angio-Lüb/KORA) P. Diemert; (CADomics) P.S.W.; (CARDIOGENICS) F.C. and W.H.O.; (CHARGE) E.B., A.L.C., A.D. and V.G.; (Corogene) M.-L.L. and J.S.; (deCODE) G. Thorgeirsson, H.H. and K. Stefansson; (EPIC-NORFOLK) N.W.; (Estonian Biobank) E.M.; (FGENTCARD) D.G.; (FINCAVAS) M.K.; (FINRISK 2007/DILGOM) V.S.; (FRISCI) L.W.; (GerMIFS) T.I., C.M., K. Stark and M.E.Z.; (GLACIER) G.H.; (GoDARTS Dundee) A.S.F.D. and A.D.M.; (HPS) S.P.; (Korean GenRIC) Y.J., H.-S.K., Ji-Young Lee and J.E.P.; (LOLIPOP) S.-T.T.; (LURIC/AtheroRemo) R.L. and W. Koenig; (METSIM) J.K., M.B. and M.L.; (MIGen) R.D.; (MORGAM) K. Kuusasmaa, J.V., P.A., D. Arveiler, J.F., D.-A.T., N.K., A.P., P.B., M.M.F., A.E. and F.K.; (Ottawa Heart Genetics Study) G.A.W., S.L.H. and S.H.S.; (PennCATH/MedStar) S.E.E. and D.J.R.; (Pfizer-Broad-Malmo) D. Altshuler and D.C.; (PIVUS/Swedish Twin Registry) C.S., L.L. and N.L.P.; (PROMIS) A.R.; (SHEEP-SCARF) K.L. and U.d.F.; (THISEAS) M.D., G.K.; (Ulm-EMIL) W. Kratzer; and (WTCCC) A.J.B., P.S.B., M.S. and A.S. Hall.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/ng.2480>.

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

- Berry, J.D. *et al.* Lifetime risks of cardiovascular disease. *N. Engl. J. Med.* **366**, 321–329 (2012).
- Peden, J.F. & Farrall, M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum. Mol. Genet.* **20** R2, R198–R205 (2011).
- Kingsmore, S.F., Lindquist, I.E., Mudge, J., Gessler, D.D. & Beavis, W.D. Genome-wide association studies: progress and potential for drug discovery and development. *Nat. Rev. Drug Discov.* **7**, 221–230 (2008).
- Stein, E.A. *et al.* Effect of a monoclonal antibody to PCSK9 on LDL cholesterol. *N. Engl. J. Med.* **366**, 1108–1118 (2012).
- Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
- Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* **43**, 339–344 (2011).
- Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* **361**, 2518–2528 (2009).
- Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
- Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.* **41**, 280–282 (2009).
- Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
- Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* **41**, 1182–1190 (2009).
- Wang, F. *et al.* Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat. Genet.* **43**, 345–349 (2011).
- The IBC 50K CAD Consortium. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet.* **7**, e1002260 (2011).
- Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
- Voight, B.F. *et al.* The Metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
- Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).



17. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
18. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
19. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
20. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
21. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
22. Ken-Dror, G., Talmud, P.J., Humphries, S.E. & Dreno, F. *APOE/C1/C4/C2* gene cluster genotypes, haplotypes and lipid levels in prospective coronary heart disease risk among UK healthy men. *Mol. Med.* **16**, 389–399 (2010).
23. Reilly, M.P. *et al.* Identification of *ADAMTS7* as a novel locus for coronary atherosclerosis and association of *ABO* with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383–392 (2011).
24. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
25. Storey, J.D. & Tibshirani, R. Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol. Biol.* **224**, 149–157 (2003).
26. Hansson, G.K., Libby, P., Schönbeck, U. & Yan, Z.-Q. Innate and adaptive immunity in the pathogenesis of atherosclerosis. *Circ. Res.* **91**, 281–291 (2002).
27. Emerging Risk Factors Collaboration. Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N. Engl. J. Med.* **364**, 829–841 (2011).
28. Mozaffarian, D. *et al.* Incidence of new-onset diabetes and impaired fasting glucose in patients with recent myocardial infarction and the effect of clinical and lifestyle risk factors. *Lancet* **370**, 667–675 (2007).
29. Turnbull, F.M. *et al.* Intensive glucose control and macrovascular outcomes in type 2 diabetes. *Diabetologia* **52**, 2288–2298 (2009).
30. Ross, R. Atherosclerosis is an inflammatory disease. *Am. Heart J.* **138**, S419–S420 (1999).

### The authors of this paper are:

Panos Deloukas<sup>1,126</sup>, Stavroula Kanoni<sup>1,126</sup>, Christina Willenborg<sup>2,126</sup>, Martin Farrall<sup>3,4,126</sup>, Themistocles L Assimes<sup>5,126</sup>, John R Thompson<sup>6,126</sup>, Erik Ingelsson<sup>7,126</sup>, Danish Saleheen<sup>8–10,126</sup>, Jeanette Erdmann<sup>2,126</sup>, Benjamin A Goldstein<sup>5</sup>, Kathleen Stirrups<sup>1</sup>, Inke R König<sup>11</sup>, Jean-Baptiste Cazier<sup>4</sup>, Åsa Johansson<sup>12</sup>, Alistair S Hall<sup>13</sup>, Jong-Young Lee<sup>14</sup>, Cristen J Willer<sup>15,16</sup>, John C Chambers<sup>17</sup>, Tõnu Esko<sup>18,19</sup>, Lasse Folkersen<sup>20,21</sup>, Anuj Goel<sup>3,4</sup>, Elin Grundberg<sup>22</sup>, Aki S Havulinna<sup>23</sup>, Weang K Ho<sup>10</sup>, Jemma C Hopewell<sup>24,25</sup>, Niclas Eriksson<sup>12</sup>, Marcus E Kleber<sup>26,27</sup>, Kati Kristiansson<sup>23</sup>, Per Lundmark<sup>28</sup>, Leo-Pekka Lyytikäinen<sup>29,30</sup>, Suzanne Rafelt<sup>31</sup>, Dmitry Shungin<sup>32–34</sup>, Rona J Strawbridge<sup>20,21</sup>, Gudmar Thorleifsson<sup>35</sup>, Emmi Tikkanen<sup>36,37</sup>, Natalie Van Zuydam<sup>38</sup>, Benjamin F Voight<sup>39</sup>, Lindsay L Waite<sup>40</sup>, Weihua Zhang<sup>17</sup>, Andreas Ziegler<sup>11</sup>, Devin Absher<sup>40</sup>, David Altshuler<sup>41–44</sup>, Anthony J Balmforth<sup>45</sup>, Inês Barroso<sup>1,46</sup>, Peter S Braund<sup>31,47</sup>, Christof Burgdorf<sup>48</sup>, Simone Claudi-Boehm<sup>49</sup>, David Cox<sup>50</sup>, Maria Dimitriou<sup>51</sup>, Ron Do<sup>41,43</sup>, DIAGRAM Consortium<sup>52</sup>, CARDIOGENICS Consortium<sup>52</sup>, Alex S F Doney<sup>38</sup>, NourEddine El Mokhtari<sup>53</sup>, Per Eriksson<sup>20,21</sup>, Krista Fischer<sup>18</sup>, Pierre Fontanillas<sup>41</sup>, Anders Franco-Cereceda<sup>54</sup>, Bruna Gigante<sup>55</sup>, Leif Groop<sup>56</sup>, Stefan Gustafsson<sup>7</sup>, Jörg Hager<sup>57</sup>, Göran Hallmans<sup>58</sup>, Bok-Ghee Han<sup>14</sup>, Sarah E Hunt<sup>1</sup>, Hyun M Kang<sup>59</sup>, Thomas Illig<sup>60</sup>, Thorsten Kessler<sup>48</sup>, Joshua W Knowles<sup>5</sup>, Genovefa Kolovou<sup>61</sup>, Johanna Kuusisto<sup>62</sup>, Claudia Langenberg<sup>63</sup>, Cordelia Langford<sup>1</sup>, Karin Leander<sup>55</sup>, Marja-Liisa Lokki<sup>64</sup>, Anders Lundmark<sup>28</sup>, Mark I McCarthy<sup>3,65,66</sup>, Christa Meisinger<sup>67</sup>, Olle Melander<sup>56</sup>, Evelin Mihailov<sup>19</sup>, Seraya Maouche<sup>68</sup>, Andrew D Morris<sup>38</sup>, Martina Müller-Nurasyid<sup>69–72</sup>, MuTHER Consortium<sup>52</sup>, Kjell Nikus<sup>73</sup>, John F Peden<sup>3</sup>, N William Rayner<sup>3</sup>, Asif Rasheed<sup>9</sup>, Silke Rosinger<sup>74</sup>, Diana Rubin<sup>53</sup>, Moritz P Rumpf<sup>48</sup>, Arne Schäfer<sup>75</sup>, Mohan Sivananthan<sup>76,77</sup>, Ci Song<sup>7</sup>, Alexandre F R Stewart<sup>78,79</sup>, Sian-Tsung Tan<sup>80</sup>, Gudmundur Thorgeirsson<sup>81,82</sup>, C Ellen van der Schoot<sup>83</sup>, Peter J Wagner<sup>36,37</sup>, Wellcome Trust Case Control Consortium<sup>52</sup>, George A Wells<sup>78,79</sup>, Philipp S Wild<sup>84,85</sup>, Tsun-Po Yang<sup>1</sup>, Philippe Amouyel<sup>86</sup>, Dominique Arveiler<sup>87</sup>, Hanneke Basart<sup>88</sup>, Michael Boehnke<sup>59</sup>, Eric Boerwinkle<sup>89</sup>, Paolo Brambilla<sup>90</sup>, Francois Cambien<sup>68</sup>, Adrienne L Cupples<sup>91,92</sup>, Ulf de Faire<sup>55</sup>, Abbas Dehghan<sup>93</sup>, Patrick Diemert<sup>94</sup>, Stephen E Epstein<sup>95</sup>, Alun Evans<sup>96</sup>, Marco M Ferrario<sup>97</sup>, Jean Ferrières<sup>98</sup>, Dominique Gauguier<sup>3,99</sup>, Alan S Go<sup>100</sup>, Alison H Goodall<sup>31,47</sup>, Villi Gudnason<sup>81,101</sup>, Stanley L Hazen<sup>102</sup>, Hilma Holm<sup>35</sup>, Carlos Iribarren<sup>100</sup>, Yangsoo Jang<sup>103</sup>, Mika Kähönen<sup>104</sup>, Frank Kee<sup>105</sup>, Hyo-Soo Kim<sup>106</sup>, Norman Klopp<sup>60</sup>, Wolfgang Koenig<sup>107</sup>, Wolfgang Kratzer<sup>108</sup>, Kari Kuulasmaa<sup>23</sup>, Markku Laakso<sup>62</sup>, Reijo Laaksonen<sup>108</sup>, Ji-Young Lee<sup>14</sup>, Lars Lind<sup>28</sup>, Willem H Ouwehand<sup>1,109,110</sup>, Sarah Parish<sup>24,25</sup>, Jeong E Park<sup>111</sup>, Nancy L Pedersen<sup>7</sup>, Annette Peters<sup>67,112</sup>, Thomas Quertermous<sup>5</sup>, Daniel J Rader<sup>113</sup>, Veikko Salomaa<sup>23</sup>, Eric Schadt<sup>114</sup>, Svati H Shah<sup>115,116</sup>, Juha Sinisalo<sup>117</sup>, Klaus Stark<sup>118</sup>, Kari Stefansson<sup>35,81</sup>, David-Alexandre Tréguët<sup>68</sup>, Jarmo Virtamo<sup>23</sup>, Lars Wallentin<sup>12</sup>, Nicholas Wareham<sup>63</sup>, Martina E Zimmermann<sup>118</sup>, Markku S Nieminen<sup>117</sup>, Christian Hengstenberg<sup>118</sup>, Manjinder S Sandhu<sup>1,63</sup>, Tomi Pastinen<sup>119</sup>, Ann-Christine Syvänen<sup>28</sup>, G Kees Hovingh<sup>88</sup>, George Dedoussis<sup>51</sup>, Paul W Franks<sup>32–34,120</sup>, Terho Lehtimäki<sup>29,30</sup>, Andres Metspalu<sup>18,19</sup>, Pierre A Zalloua<sup>121</sup>, Agneta Siegbahn<sup>12</sup>, Stefan Schreiber<sup>75</sup>, Samuli Ripatti<sup>1,37</sup>, Stefan S Blankenberg<sup>94</sup>, Markus Perola<sup>23</sup>, Robert Clarke<sup>24,25</sup>, Bernhard O Boehm<sup>74</sup>, Christopher O'Donnell<sup>93</sup>, Muredach P Reilly<sup>122,126</sup>, Winfried März<sup>26,123</sup>, Rory Collins<sup>24,25,126</sup>, Sekar Kathiresan<sup>41,124,125,126</sup>, Anders Hamsten<sup>20,21,126</sup>, Jaspal S Kooner<sup>80,126</sup>, Unnur Thorsteinsdottir<sup>35,81,126</sup>, John Danesh<sup>9,126</sup>, Colin N A Palmer<sup>38,126</sup>, Robert Roberts<sup>78,79,126</sup>, Hugh Watkins<sup>3,4,126</sup>, Heribert Schunkert<sup>48,126</sup> & Nilesh J Samani<sup>31,47,126</sup>



<sup>1</sup>Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. <sup>2</sup>Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, Lübeck, Germany. <sup>3</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>4</sup>Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. <sup>5</sup>Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. <sup>6</sup>Department of Health Sciences, University of Leicester, Leicester, UK. <sup>7</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. <sup>8</sup>Center for Non-Communicable Diseases, Karachi, Pakistan. <sup>9</sup>Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>10</sup>Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>11</sup>Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Germany. <sup>12</sup>Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. <sup>13</sup>Division of Cardiovascular and Neuronal Remodelling, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK. <sup>14</sup>Center for Genome Science, Korea National Institute of Health, Korea Center for Disease Control and Prevention, Yoonje-ri, Chungwon-gun, Korea. <sup>15</sup>Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA. <sup>16</sup>Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA. <sup>17</sup>Department of Epidemiology and Biostatistics, Imperial College London, London, UK. <sup>18</sup>Estonian Genome Center, University of Tartu, Tartu, Estonia. <sup>19</sup>Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. <sup>20</sup>Atherosclerosis Research Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>21</sup>Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden. <sup>22</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. <sup>23</sup>Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. <sup>24</sup>Clinical Trial Service Unit, University of Oxford, Oxford, UK. <sup>25</sup>Epidemiological Studies Unit, University of Oxford, Oxford, UK. <sup>26</sup>Mannheim Institute of Public Health, Social and Preventive Medicine, Medical Faculty of Mannheim, University of Heidelberg, Mannheim, Germany. <sup>27</sup>Ludwigshafen Risk and Cardiovascular Health (LURIC) Study, Freiburg, Germany. <sup>28</sup>Department of Medical Sciences, Uppsala University, Uppsala, Sweden. <sup>29</sup>Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital, Tampere, Finland. <sup>30</sup>Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland. <sup>31</sup>Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. <sup>32</sup>Genetic & Molecular Epidemiology Unit, Department of Clinical Sciences, Lund University Diabetes Center, Skåne University Hospital, Malmö, Sweden. <sup>33</sup>Department of Public Health & Clinical Medicine, Genetic Epidemiology & Clinical Research Group, Section for Medicine, Umeå University, Umeå, Sweden. <sup>34</sup>Department of Odontology, Umeå University, Umeå, Sweden. <sup>35</sup>deCODE Genetics, Reykjavik, Iceland. <sup>36</sup>Institute for Molecular Medicine FIMM, University of Helsinki, Helsinki, Finland. <sup>37</sup>Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland. <sup>38</sup>Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK. <sup>39</sup>Department of Pharmacology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>40</sup>HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. <sup>41</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. <sup>42</sup>Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. <sup>43</sup>Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. <sup>44</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. <sup>45</sup>Division of Cardiovascular and Diabetes Research, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK. <sup>46</sup>University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. <sup>47</sup>National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. <sup>48</sup>Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. <sup>49</sup>Practice of Gynecology, Ulm University Medical Centre, Ulm, Germany. <sup>50</sup>Biotherapeutics and Bioinnovation Center, Pfizer, South San Francisco, California, USA. <sup>51</sup>Department of Dietetics–Nutrition, Harokopio University, Athens, Greece. <sup>52</sup>A list of members and affiliations appears in the **Supplementary Note**. <sup>53</sup>Klinik für Innere Medizin, Kreiskrankenhaus Rendsburg, Rendsburg, Germany. <sup>54</sup>Cardiothoracic Surgery Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. <sup>55</sup>Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>56</sup>Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö, Sweden. <sup>57</sup>Commissariat à l'Énergie Atomique (CEA)–Genomics Institute, National Genotyping Centre, Paris, France. <sup>58</sup>Department of Public Health & Clinical Medicine, Section for Nutritional Research, Umeå University, Umeå, Sweden. <sup>59</sup>Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. <sup>60</sup>Hannover Unified Biobank, Hannover Medical School, Hannover, Germany. <sup>61</sup>First Cardiology Department, Onassis Cardiac Surgery Center 356, Athens, Greece. <sup>62</sup>Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. <sup>63</sup>Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. <sup>64</sup>Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland. <sup>65</sup>Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. <sup>66</sup>Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK. <sup>67</sup>Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. <sup>68</sup>Institut National de la Santé et la Recherche Médicale (INSERM) Unité Mixte de Recherche (UMR) S937, Institute for Cardiometabolism and Nutrition (ICAN), Pierre and Marie Curie (Paris 6) University, Paris, France. <sup>69</sup>Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany. <sup>70</sup>Chair of Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. <sup>71</sup>Chair of Genetic Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. <sup>72</sup>Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. <sup>73</sup>Heart Centre, Department of Cardiology, Tampere University Hospital, Tampere, Finland. <sup>74</sup>Division of Endocrinology and Diabetes, Department of Internal Medicine, Ulm University Medical Centre, Ulm, Germany. <sup>75</sup>Institut für Klinische Molekularbiologie, Christian-Albrechts Universität, Kiel, Germany. <sup>76</sup>Division of Epidemiology, Multidisciplinary Cardiovascular Research Centre (MCRC) University of Leeds, Leeds, UK. <sup>77</sup>Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK. <sup>78</sup>University of Ottawa Heart Institute, Cardiovascular Research Methods Centre Ontario, Ottawa, Ontario, Canada. <sup>79</sup>Ruddy Canadian Cardiovascular Genetics Centre, Ottawa, Ontario, Canada. <sup>80</sup>National Heart and Lung Institute (NHLI), Imperial College London, Hammersmith Hospital, London, UK. <sup>81</sup>Faculty of Medicine, University of Iceland, Reykjavik, Iceland. <sup>82</sup>Department of Medicine, Landspítali University Hospital, Reykjavik, Iceland. <sup>83</sup>Department of Experimental Immunohematology, Sanquin, Amsterdam, The Netherlands. <sup>84</sup>Center for Thrombosis and Hemostasis, University Medical Center Mainz, Johannes Gutenberg University Mainz, Mainz, Germany. <sup>85</sup>Department of Medicine 2, University Medical Center Mainz, Johannes Gutenberg University Mainz, Mainz, Germany. <sup>86</sup>Institut Pasteur de Lille, INSERM U744, Université Lille Nord de France, Lille, France. <sup>87</sup>Department of Epidemiology and Public Health, EA3430, University of Strasbourg, Strasbourg, France. <sup>88</sup>Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands. <sup>89</sup>Human Genetics Center, University of Texas Health Science Center, Houston, Texas, USA. <sup>90</sup>Department of Experimental Medicine, University of Milano–Bicocca, Monza, Italy. <sup>91</sup>Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. <sup>92</sup>National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. <sup>93</sup>Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>94</sup>Clinic for General and Interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany. <sup>95</sup>Cardiovascular Research Institute, Washington Hospital Center, Washington, DC, USA. <sup>96</sup>Centre for Public Health, The Queen's University of Belfast, Belfast, UK. <sup>97</sup>Research Centre for Epidemiology and Preventive Medicine (EPIMED), Department of Clinical and Experimental Medicine, University of Insubria, Varese, Italy. <sup>98</sup>Department of Cardiology, Toulouse University School of Medicine, Rangueil Hospital, Toulouse, France. <sup>99</sup>INSERM UMR S872, Cordeliers Research Centre, Paris, France. <sup>100</sup>Division of Research, Kaiser Permanente Northern California, Oakland, California, USA. <sup>101</sup>Icelandic Heart Association, Kopavogur, Iceland. <sup>102</sup>Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA. <sup>103</sup>Cardiology Division, Department of Internal Medicine, Cardiovascular Genome Center, Yonsei University, Seoul, Korea. <sup>104</sup>Department of Clinical Physiology, Tampere University Hospital and University of Tampere, Tampere, Finland. <sup>105</sup>UK Clinical Research Collaboration (UKCRC) Centre of Excellence for Public Health (Northern Ireland), Queen's University of Belfast, Belfast, UK. <sup>106</sup>Department of Internal Medicine, Cardiovascular Center, Seoul National University Hospital, Seoul, Korea. <sup>107</sup>Department of Internal Medicine II–Cardiology, Ulm University Medical Center, Ulm, Germany. <sup>108</sup>Science Center, Tampere University Hospital, Tampere, Finland. <sup>109</sup>Department of Haematology, University of Cambridge, Cambridge, UK. <sup>110</sup>National Health Service (NHS) Blood and Transplant, Cambridge, UK. <sup>111</sup>Division of Cardiology, Samsung Medical Center, Seoul, Korea. <sup>112</sup>Munich Heart Alliance, Munich, Germany. <sup>113</sup>Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>114</sup>Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA. <sup>115</sup>Center for Human Genetics, Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA. <sup>116</sup>Division of Cardiology, Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA. <sup>117</sup>Division of Cardiology, Department of Medicine, Helsinki University Central Hospital (HUCH), Helsinki, Finland. <sup>118</sup>Klinik und Poliklinik für Innere Medizin II, Regensburg, Germany. <sup>119</sup>Department of Human Genetics, McGill University, Montréal, Québec, Canada. <sup>120</sup>Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. <sup>121</sup>Lebanese American University, Chouran, Beirut, Lebanon. <sup>122</sup>Cardiovascular Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>123</sup>Synlab Academy, Mannheim, Germany. <sup>124</sup>Cardiology Division, Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>125</sup>Cardiovascular Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>126</sup>These authors contributed equally to this work. Correspondence should be addressed to P. Deloukas ([panos@sanger.ac.uk](mailto:panos@sanger.ac.uk)) or N.J.S. ([njs@leicester.ac.uk](mailto:njs@leicester.ac.uk)).

## ONLINE METHODS

**Meta-analysis and combination of evidence across stages.** Analyses were performed in each study (**Supplementary Table 1a**) to test the following comparisons: all CAD cases with all controls, adjusted for sex and age; male CAD cases with male controls, adjusted for age; female CAD cases with female controls, adjusted for age; early-onset CAD cases with early age of onset ( $\leq 50$  years) with all CAD controls, adjusted for sex; late-onset CAD cases ( $> 50$  years) with all controls, adjusted for sex; and all myocardial infarction cases with all controls, adjusted for age and sex. Age was defined as the recruitment age for controls and the event age for cases. We used the additive genetic model and fixed-effect inverse variance-weighted meta-analysis. SNPs were excluded from the meta-analysis if present in  $< 17$  GWAS and/or Metabochip or  $< 13$  Metabochip stage 2 studies. Heterogeneity was evaluated using the Cochran's  $Q$  and  $I^2$  statistics. For SNPs with non-significant heterogeneity ( $P$  for  $Q > 0.01$ ), we report fixed-effect model results. For SNPs with significant heterogeneity ( $P$  for  $Q < 0.01$ ), we performed an outlier test comparing the results in each study with the average of all other studies. For outliers ( $P < 0.01$  or no studies with data), we excluded the most extreme study and repeated the meta-analysis. If no outliers were detected, but heterogeneity was significant, we used a random-effects model that was also used for all SNPs with significant heterogeneity in stage 1. In stage 3, we used a fixed-effect inverse variance-weighted meta-analysis.

The combination of evidence across stage 1 and stage 2 meta-analysis results was performed using Fisher's combined  $P$ -values method; using two-sided  $P$  values from stage 1 and one-sided  $P$  values from stage 2 for all SNPs with consistent direction of effect across the two stages. We estimated the stage 1 and 2 combined effect sizes for SNPs in the known loci using a fixed-effect inverse variance-weighted meta-analysis. The combination of evidence across stages 1–3 for the replication effort was performed using a sample size-weighted meta-analysis for the selected SNPs. All participants gave written consent for participation in genetic studies, and the protocol of each study was approved by the corresponding local research ethics committee or institutional review board.

**False discovery rate.** FDR control is an alternative approach to experiment-wise error rate control that allows for statistical multiple testing; identifying as many significantly associated SNPs as possible with a tolerable false positive burden. FDR analysis is useful for selecting extended panels of SNPs (and genes) for experiments on the basis of multiple signals (for example, pathway or network analyses) that are robust to contamination by a small number of false positive signals. However, given the specific design of the Metabochip array to include selected SNPs (for replication) with significant  $P$  values and several high-density regions (for fine mapping) associated with CAD and the other cardiometabolic traits, the number of SNPs significantly associated with CAD, as well as high LD, could bias the FDR analysis. Therefore, we excluded from the 79,138 SNPs with available stage 1 and 2 data all SNPs falling in a high-density region associated with CAD (**Tables 1 and 2**), as well as CAD risk SNPs associated at  $P < 5 \times 10^{-8}$ . Furthermore, we performed an LD-based SNP pruning of the remaining high-density regions ( $r^2 < 0.2$ ). In total, 54,806 SNPs were included in the FDR analysis.

We combined stage 1 and 2 data as an inverse variance-weighted average, and  $P$  values were calculated by Wald test. SNPs selected because their stage 1  $P$  values were below 0.01 had their combined  $P$  values adjusted. If  $p_0$  is the  $P$  value used as the criterion for selection in stage 1,  $z_{12}$  is the standardized test statistic obtained by combining the stages (arbitrarily assumed to be positive) and  $s_1$  and  $s_2$  are the standard errors for the two stages, then the adjusted  $P$  value is the sum of two integrals representing the two tails in which the stage 1 result might fall. The first is:

$$I_1 = \int_{z_{12}}^{\infty} \int_{\Phi^{-1}(1-p_0/2)}^{\infty} \frac{\Phi(u)\Phi\left(s_2\left[v\sqrt{\frac{1}{s_1^2} + \frac{1}{s_2^2}} - \frac{u}{s_1}\right]\right)}{p_0} s_2 \sqrt{\frac{1}{s_1^2} + \frac{1}{s_2^2}} du dv$$

and the second has the same form but is integrated from  $-\infty$  to  $\Phi^{-1}(p_0/2)$ , where  $\Phi$  is the cumulative normal function. To test the adjusted  $P$  values, a simulation was performed in which null SNPs were generated and selected in stage 1 on the basis of their  $P$  values. These were combined with random

second-stage data simulated again assuming a null effect. The adjusted  $P$  values had the expected uniform distribution between zero and one, suitable for use in the FDR analysis.

FDR analysis was performed using QVALUE software. A natural cubic spline (with 4 degrees of freedom) was fitted to provide a smoothed estimate of the proportion of null  $P$  values ( $\hat{\pi}_0$ ). A density histogram of the  $P$  values for the 54,806 SNPs is shown in **Supplementary Figure 7**. At FDR = 0.05, we obtained 138 SNPs that were combined with 73 independent SNPs from fine-mapping regions associated with CAD. The selection included the SNP with the lowest combined  $P$  value per fine-mapping region and all SNPs within these regions that met the 5% FDR criterion in a separate analysis and were unlinked ( $r^2 < 0.2$ ). Finally, all SNPs reported in **Tables 1 and 2** were added to the set of 211 SNPs (5% FDR results and CAD fine-mapped regions), resulting in 153 independent SNPs (104 identified through the FDR analysis) at  $r^2 < 0.2$ , which were used for heritability calculations (**Supplementary Table 9**).

**Heritability.** Heritability estimates were calculated locus by locus using the multifactorial liability threshold model based on OR estimates that assume that the lead SNP at a locus accurately tags the disease-causing variant, as described in ref. 12. The calculations are based on a disease prevalence estimate of 5% and an estimate of 40% for the total heritability of coronary disease.

**Expression analyses.** We interrogated the 16 new (or proxy;  $r^2 > 0.8$ ) CAD risk SNPs for *cis* eQTL expression in multiple tissues: the ASAP study<sup>31</sup> used tissue biopsies taken from patients undergoing carotid endarterectomy (plaque  $n = 117$ ) or valve surgery (liver  $n = 152$ , aorta media  $n = 117$ , aorta adventitia  $n = 103$  and mammary artery  $n = 88$ ). Expression data were generated using the Affymetrix HG-U133 plus 2.0 array (plaque) or the Affymetrix ST1.0 Exon array (liver, aorta and mammary artery); in the MuTHER study<sup>32</sup>, RNA levels were measured in LCLs ( $n = 826$ ), skin ( $n = 705$ ) and fat biopsies ( $n = 825$ ) from 850 female twins (one-third monozygotic and two-thirds dizygotic) from the TwinsUK resource using the Illumina HumanHT-12v3 array. We assessed genotype with gene expression associations, using an additive linear model (within a 1-Mb window); in Cardiogenics<sup>5</sup>, monocytes and macrophages were collected from healthy subjects and individuals with CAD, and RNA was profiled with the Illumina Human Ref-8 array. eQTL analysis was undertaken in 459 healthy individuals from Cambridge, UK, using an additive linear model (1-Mb window); in the Massachusetts General Hospital study<sup>33</sup> of liver, omentum and subcutaneous adipose tissue among subjects undergoing Roux-en-Y gastric bypass surgery, eQTL analysis was performed with a linear regression model using a 1-Mb window.

In loci with significant *cis*-eQTL signal(s) ( $P < 1 \times 10^{-4}$ ), we also identified the most strongly associated *cis*-eQTL SNP (eSNP) for the corresponding transcript and then performed conditional analyses, including in the regression model, with either the lead eSNP or the lead CAD-associated SNP. On the basis of the conditional analysis, we determined whether the same variant underlies both gene expression regulation and disease.

Finally, we interrogated the lead SNPs in the 16 new CAD susceptibility loci for allelic expression imbalance effects in LCLs, fibroblasts and monocytes ( $n = 188$ ; Cardiogenics), as described in ref. 34.

**Network analysis.** Genes for network analysis were selected using 310 SNPs (88 SNPs in known and new CAD risk loci and 222 SNPs at FDR  $< 10\%$  and LD pruned to  $r^2 \leq 0.7$ ). We first selected genes with an eQTL ( $P \leq 1 \times 10^{-6}$ ) and then on the basis of physical proximity (included overlapping genes on opposite strands or at equal distance from the SNP; genes were considered within a 40-kb window centered on the SNP). Spliced ESTs and putative transcripts were not included. Network analysis was performed using the Ingenuity Pathway Analysis software tool (IPA; Ingenuity Systems). We considered molecules and/or relationships available in The IPA Knowledge Base for human, mouse or rat and set the confidence filter to experimentally observed or high (predicted). Networks were generated with a maximum size of 70 genes, allowing up to 10 networks. Molecules in the query set with recorded interactions were 'eligible' for network construction using the IPA algorithm. Networks were ranked according to their degree of relevance to the eligible molecules in the query data set. The score takes into account the number of eligible molecules in the network and its size, as well as the total number of eligible molecules

analyzed and the total number of molecules in the Ingenuity Knowledge Base that could potentially be included in the networks. The network score is based on the hypergeometric distribution and is calculated by right-tailed Fisher's exact test. The significance *P* value associated with enrichment of functional processes was calculated using the right-tailed Fisher's exact test by considering the number of query molecules that participate in that function and the total number of molecules that are known to be associated with that function in the Ingenuity Knowledge Base.

31. Folkersen, L. *et al.* Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. *Circ. Cardiovasc. Genet.* **3**, 365–373 (2010).
32. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins: the MuTHER Study. *Nat. Genet.* **44**, 1084–1089 (2012).
33. Schadt, E.E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
34. Ge, B. *et al.* Global patterns of *cis* variation in human cells revealed by high-density allelic expression analysis. *Nat. Genet.* **41**, 1216–1222 (2009).