

Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease

Pim van der Harst, Niek Verweij

Rationale: Coronary artery disease (CAD) is a complex phenotype driven by genetic and environmental factors. Ninety-seven genetic risk loci have been identified to date, but the identification of additional susceptibility loci might be important to enhance our understanding of the genetic architecture of CAD.

Objective: To expand the number of genome-wide significant loci, catalog functional insights, and enhance our understanding of the genetic architecture of CAD.

Methods and Results: We performed a genome-wide association study in 34 541 CAD cases and 261 984 controls of UK Biobank resource followed by replication in 88 192 cases and 162 544 controls from CARDIoGRAMplusC4D. We identified 75 loci that replicated and were genome-wide significant ($P < 5 \times 10^{-8}$) in meta-analysis, 13 of which had not been reported previously. Next, to further identify novel loci, we identified all promising ($P < 0.0001$) loci in the CARDIoGRAMplusC4D data and performed reciprocal replication and meta-analyses with UK Biobank. This led to the identification of 21 additional novel loci reaching genome-wide significance ($P < 5 \times 10^{-8}$) in meta-analysis. Finally, we performed a genome-wide meta-analysis of all available data revealing 30 additional novel loci ($P < 5 \times 10^{-8}$) without further replication. The increase in sample size by UK Biobank raised the number of reconstituted gene sets from 4.2% to 13.9% of all gene sets to be involved in CAD. For the 64 novel loci, 155 candidate causal genes were prioritized, many without an obvious connection to CAD. Fine mapping of the 161 CAD loci generated lists of credible sets of single causal variants and genes for functional follow-up. Genetic risk variants of CAD were linked to development of atrial fibrillation, heart failure, and death.

Conclusions: We identified 64 novel genetic risk loci for CAD and performed fine mapping of all 161 risk loci to obtain a credible set of causal variants. The large expansion of reconstituted gene sets argues in favor of an expanded omnigenic model view on the genetic architecture of CAD. (*Circ Res.* 2018;122:433-443. DOI: 10.1161/CIRCRESAHA.117.312086.)

Key Words: computational biology ■ coronary artery disease ■ genetics ■ genome-wide association study ■ sample size

Coronary artery disease (CAD) is the predominant cause of ischemic heart disease often leading to myocardial infarction and a leading cause of death. Globally, deaths because of ischemic heart disease increased by 16.6% from 2005 to 2015 to 8.9 million deaths. However, the age-standardized mortality rates are decreasing (fell by 12.8%)¹ because of preventive and treatment strategies established on evolving knowledge of the underlying pathophysiology of CAD.

Editorial, see p 391
In This Issue, see p 385

CAD is a complex disease, resulting from numerous additive and interacting contributions in an individual's

environment and lifestyle in combination with their underlying genetic architecture. Since the first genome-wide association studies (GWAS) for CAD in 2007,²⁻⁴ multiple additional studies with progressively larger sample sizes identified 97 genome-wide significant genetic loci associated with CAD⁵⁻¹⁰ at the time of analysis. The continuous effort to identify additional loci associated with CAD and share these early with the scientific community is important, especially to enhance our understanding of the biological underpinnings of CAD and to catalyze the development of drugs. A comprehensive understanding of the genetic architecture of CAD is also essential to enable precision medicine approaches by identifying subgroups of patients at increased risk of CAD or its

Original received September 18, 2017; revision received November 26, 2017; accepted December 4, 2017. In November 2017, the average time from submission to first decision for all original research papers submitted to *Circulation Research* was 11.99 days.

From the Department of Cardiology (P.v.d.H., N.V.) and Department of Genetics (P.v.d.H.), University Medical Center Groningen, University of Groningen, The Netherlands; and Durrer Center for Cardiogenetic Research, Netherlands Heart Institute, Utrecht (P.v.d.H.).

The online-only Data Supplement is available with this article at <http://circres.ahajournals.org/lookup/suppl/doi:10.1161/CIRCRESAHA.117.312086/-/DC1>.

Correspondence to Pim van der Harst, MD, PhD, Department of Cardiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands. E-mail p.van.der.harst@umcg.nl

© 2017 The Authors. *Circulation Research* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution Non-Commercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is noncommercial, and no modifications or adaptations are made.

Novelty and Significance

What Is Known?

- Coronary artery disease (CAD) is a multifactorial disease with a substantial heritable component.
- Genome-wide association studies in the past decade have identified 96 loci associated with CAD and are believed to provide biological insights into key pathways under the presumption of a polygenetic model.

What New Information Does This Article Contribute?

- We have identified 64 additional loci, which were associated with CAD. We fine mapped all new and known loci to provide evidence in support of the causal role of the genetic variants or genes in CAD.
- Network analyses suggest a complex genetic architecture of CAD, which might not be fully captured by the prevailing polygenetic model of CAD.
- This work lends support to the omnigenetic model, proposing that associated genetic variants might not necessarily lay in key disease pathways. Instead, all gene-regulatory networks maybe sufficiently interrelated such that all genes expressed, including those outside key disease pathways, may influence key disease-related genes.

CAD—a leading cause of death—is a complex multifactorial disease. Genome-wide association studies of CAD have offered new biological insights and added to risk prediction and identification of drugable targets. We performed a large systematic meta-analysis of genome-wide association studies, involving 122 733 cases and 424 528 controls and identified 64 new genetic loci that were associated with CAD. Fine mapping of all known and novel CAD loci highlighted potential causal single-nucleotide polymorphism–gene mechanisms. A large proportion of all biological pathways and a plethora of human tissues were found to be associated with CAD for no obvious reason. This finding could indicate that the polygenetic model may not uphold with ever increasing sample sizes for CAD genetics, and the omnigenetic model may be more appropriate to accommodate the increasing complexity. This study underscores the importance of tissue-specific dedicated mechanistic studies. New methods and tools are required to advance our understanding of genetic mechanisms influencing the development and progression of CAD.

Nonstandard Abbreviations and Acronyms

CAD	coronary artery disease
CCM2	CCM2 scaffolding protein
DEPICT	data-driven expression-prioritized integration for complex
eQTL	expression quantitative trait locus
FNDC3B	fibronectin type III domain containing 3B
GWAS	genome-wide association study
SNP	single-nucleotide polymorphism
TRIM5	tripartite motif containing 5

complications and might identify those with a specific driving pathophysiology in whom a particular therapeutic or preventive approach would be most useful.¹¹

To further our knowledge of the genetic architecture of CAD, we performed a de novo GWAS of the UK Biobank resource and meta-analyses with CARDIoGRAMplusC4D data. Our approach led to the identification of 64 novel loci associated with CAD, expanding the grand total to 161. These loci were interrogated using bioinformatic approaches to catalog and interpret the potential biological relevance of our findings. We also performed network and gene-set analyses and proposed the omnigenetic model to explain our findings. This expanding resource is now available for other investigators to help to further elucidate the underlying biology and relevance.

Methods

The data that support the findings of this study are available from the corresponding author on reasonable request. The de novo GWAS analysis and meta-analysis have been posted on Mendeley (doi:10.17632/2zdd47c94h.1; doi:10.17632/gbbsrpx6bs.1). A summary of the methods is provided below, and a more detailed description of the experimental procedures is provided in the [Online Data Supplement](#).

Study Design and Samples

The study design consisted of a reciprocal 2-stage sequential discovery and replication approach (Online Figure I) providing the most robust

statistical evidence followed by an overall meta-analysis of all available data for which currently no replication data were available in this study. First, using the UK Biobank resource, we conducted a GWAS to discover single-nucleotide polymorphisms (SNPs) associated with CAD. In stage 2, we took forward all promising SNPs reaching nominal significance ($P < 0.0001$) for replication in CARDIoGRAMplusC4D data. Replicating SNPs ($P < 0.05$ after Bonferroni adjustment) were meta-analyzed and considered true when surpassing the genome-wide significance threshold ($P < 5 \times 10^{-8}$). The reciprocal stage 1 entailed the identification for all promising SNPs ($P < 0.0001$) in CARDIoGRAMplusC4D and replication in UK Biobank ($P < 0.05$ after Bonferroni adjustment) followed by meta-analysis. Again, SNPs replicating and surpassing the genome-wide significance threshold were considered true. A sentinel SNP in a locus was defined as the most significant variant in a 1-mb region that was independent from other sentinel SNPs ($r^2 < 0.1$). A locus was defined as a region of 1 mb at either side of the sentinel SNP. A locus was considered novel if the sentinel SNP was not within a 1-mb window (at either side) of earlier reported genome-wide significant SNPs (Online Table I). Finally, we performed a genome-wide meta-analysis of the UK Biobank resource and CARDIoGRAMplusC4D to identify additional CAD-associated loci ($P < 5 \times 10^{-8}$ in meta-analysis). A potential sample overlap between the UK Biobank and cohorts of CARDIoGRAMplusC4D was estimated to be $< 0.1\%$; no evidence was found that this biased the test statistics ([Online Data Supplement](#)).

Candidate Genes and Insights in Biology

Candidate causal genes at each of the loci were prioritized based on proximity, expression quantitative trait locus (eQTL) data, DEPICT analyses (Data-Driven Expression-Prioritized Integration for Complex Traits),¹² and long-range chromatin interactions of variants with gene promoters ([Online Data Supplement](#)).^{8,13} Summary information of genes was obtained via queries in GeneCards, UniProt, and Tocris. The Mouse Genomic Informatics database was used for obtaining insights into mammalian phenotypes associated with disruption of candidate genes. DEPICT was also used to test for enrichment of gene sets and identify relevant tissues and cell types. Ingenuity pathway analysis (June 2017 release) was performed to strengthen the biological relevancy of the novel loci.

Insights in Loci by Associations With Other Phenotypes

The GWAS catalog was queried and a phenome scan was performed by intersecting the identified loci with the GWAS catalog and by testing the association of the newly identified SNPs with a wide range of

phenotypes using linear or logistic regression analysis in UK Biobank (Online Data Supplement). Genetic risk scores (GRS) were constructed using effect estimates obtained from the CARDIoGRAMplusC4D data as described previously.⁸ Multivariable Cox proportional hazards models were fitted for quintiles of the GRS in the UK Biobank resource, to assess the extent to which the GRS could predict new-onset atrial fibrillation/flutter and heart failure.

Regulatory DNA and Fine Mapping of Probable Causal Variants

To systematically characterize the functional, cellular, and regulatory contribution of genetic variation, we used GARFIELD,¹⁴ analyzing the enrichment of genome-wide association summary statistics in tissue-specific functional elements at given significance thresholds. Probabilistic Annotation Integrator was used to fine-map loci by integrating genetic association signal strength with genomic functional annotation data.¹⁵ We explored the potential target genes of these candidate causal variants by determining their direct effects on protein function (missense variants) and evidence connecting the causal variant in an untranslated region (Utr)-3' region to gene expression (eQTL) or physical interactions (Hi-C) with the promotor of an eQTL gene. Determination of potential causal mechanisms of the potential causal variants based on (1) missense variation, (2) chromatin interaction between the causal variant and the promotor of a gene for which the causal variant was also significantly associated with gene expression by eQTL analyses, or (3) Utr3' overlapping variants that were also significantly associated with gene expression of the same gene corresponding to the Utr3' position. In addition, for genes/mechanisms to be prioritized by eQTL analyses and chromatin interactions or Utr3', the respective causal variant was required to be in an enhancer region.

Results

Genome-Wide Analyses of 34 541 Cases and 261 984 Controls

The stage 1 GWAS analysis in UK Biobank (34 541 cases and 261 984 controls; Online Table II) of 7 947 838 SNPs revealed 630 suggestive SNPs ($P < 0.0001$) in 442 loci (Online Table III). Eighty-six independent SNPs in 75 loci both replicated ($P < 0.05$ Bonferroni adjusted) in stage 2 in $\leq 88 192$ cases and 162 544 controls of CARDIoGRAMplusC4D, and achieved genome-wide significance ($P < 5 \times 10^{-8}$) with no evidence of heterogeneity of effects ($P_{\text{het}} \geq 0.10$). Thirteen of the 75 loci are not established CAD-associated loci (Table 1).

Next, we reanalyzed the data from the MetaboChip meta-analysis of CARDIoGRAMplusC4D,⁹ the CARDIoGRAMplusC4D 1000 Genomes meta-analysis,⁷ and the CARDIoGRAM Exome array data¹⁶ to identify the promising SNPs ($P < 0.0001$). We identified 568 promising SNPs located in 375 loci (Online Table IV). One hundred and thirteen independent SNPs in 96 loci both replicated ($P < 0.05$ Bonferroni adjusted) in stage 2, UK Biobank, and achieved genome-wide significance in meta-analysis ($P < 5 \times 10^{-8}$), including 21 additional novel loci (Table 1; Online Table V).

Finally, we performed a meta-analysis of CARDIoGRAMplusC4D⁹ and the CARDIoGRAMplusC4D 1000 Genomes meta-analysis⁷ with UK Biobank and identified 30 additional loci for which no replication test was available (Table 1; Online Table VI) increasing the total number of genome-wide significant CAD loci to 161 (Online Figure II). The novel variants were common ($>5\%$, except for 1, rs112635299 near *SERPINA1*). Online Figure III shows the regional association plot of each novel locus. For some variants, a dominant or recessive linkage model appears to be a better fit compared

with an additive model (Online Table VII). Complete summary statistics of all SNPs in UK Biobank and the UK Biobank CARDIoGRAMplusC4D meta-analysis are available as download on www.cardiomics.net.

Candidate Genes and Deeper Insights Into Biology

To disentangle whether associations were driven more by acute myocardial infarction as opposed to stable CAD, we performed multinomial logistic regression analyses for all genome-wide significant ($P < 5 \times 10^{-8}$) loci in UK Biobank. In total, 17 666 of 34 541 CAD individuals were diagnosed with myocardial infarction. None of the novel loci and only 2 previously identified variants (rs9349379 and rs10947789) appear to be mainly driven by its association with myocardial infarction rather than stable CAD (false discovery rate [FDR], $P < 0.05$; Online Table VIII).

We further explored the potential biology of the 64 novel CAD-associated loci by prioritizing 155 candidate causal genes in these loci: 69 genes were in proximity (the nearest gene and any additional gene within 10 kb) of the lead variant, 9 genes contained coding genetic variation in linkage disequilibrium ($r^2 > 0.8$) with the lead variant (Online Table IX), 50 genes were selected based on eQTL analyses (Online Table X), 64 genes showed significant chromatin interactions (Hi-C) between the genetic variant and promoter of the gene (Online Table XI), and 60 genes were prioritized based on DEPICT analyses (Online Table XII). Of the 155 candidate genes, 63 were prioritized by multiple methods of identification, which may be used to prioritize candidate causal genes. A summary of the current function annotation of each novel candidate gene is provided in Online Table XIII, and knowledge on pharmacological compounds and nutrients influencing these genes is provided in Online Table XIV. Next, we performed a systematic search in the Mouse Genome Informatics database to identify the effect of mutations in orthologous genes for these candidate causal genes (details in Online Table XV). In brief, we identified 34 genes that expressed at least 1 cardiovascular system phenotype (*AGT*, *ARHGAP42*, *BACH1*, *CALCRL*, *CASQ2*, *CCM2*, *CDC123*, *CDKN1A*, *FIGN*, *FOXC1*, *GITI1*, *GNPAT*, *HCRT*, *HSD17B12*, *MAP1S*, *MAP3K1*, *MSANTD1*, *NGF*, *NPHP3*, *PCIFI1*, *PDS5B*, *PLCG1*, *PLEKHA1*, *PPP2R3A*, *PRDM16*, *PRKCE*, *RAC1*, *SEMA5A*, *SH3PXD2A*, *TFPI*, *TIPARP*, *TMEM106B*, *VEGFA*, and *ZFPM2*) and 34 genes that affected other potentially plausible traits linked to CAD, including metabolic/lipid/adipose/weight abnormalities (*AGT*, *CORO6*, *FIGN*, *GITI1*, *KAT2A*, *NGF*, *PPP2R3A*, *NPHH3*, *SH3PXD2A*, *TMEM106B*, *VEGFA*, *ZHX3*, *OPTN*, *FAM213A*, *DNAJC7*, and *COPRS*), abnormalities in inflammation or white blood cells (*DHX58*, *FHL3*, *HNRNPD*, *PLCG2*, *PRDM16*, *TFPI*, *VEGFA*, *ZNF335*, *PRKCE*, *MYO1G*, *RAC1*, and *ARID4A*), and abnormalities in platelets or coagulation (*FHL3*, *PLCG2*, *TFPI*, *VEGFA*, *DST*, and *KLF4*).

Novel Insights From Pathway Analyses

Ingenuity pathway analysis restricted to the 155 candidate causal genes confirmed that these are enriched for effects on the cardiovascular system and cell cycle functions (Online Table XVI). Pathway insights provided by the DEPICT framework identified 1525 reconstituted gene sets that could be captured

Table 1. Sixty-Four Novel Genome-Wide Significant CAD Loci

Cytoband	Position	Lead SNP	A1	A2	Freq	Variant Function	Candidate Genes	Resource	OR (95% CI)	P Value
1p36.33	2252205	rs36096196	T	C	0.15	Downstream	<i>MORN1</i> ,* <i>SKIT</i> †	MA	1.05 (1.03–1.06)	1.3×10 ⁻⁸
1p36.32	3325912	rs2493298	A	C	0.14	Intronic	<i>PRDM16</i> ,*†‡ <i>PEX10</i> ,‡ <i>PLCH2</i> ,‡ <i>RER1</i> †	UK	1.06 (1.04–1.08)	1.9×10 ⁻⁹
1p34.3	38461319	rs61776719	A	C	0.53	Intergenic	<i>FHL3</i> ,*†§ <i>UTP11</i> ,§ <i>SF3A3</i> ,§ <i>MANEAL</i> ,§ <i>INPP5B</i> §	UK	1.04 (1.03–1.06)	1.1×10 ⁻⁹
1p13.2	115753482	rs11806316	A	G	0.37	Intergenic	<i>NGF</i> ,*†‡ <i>CASQ2</i> ‡	CA	0.96 (0.95–0.97)	4.9×10 ⁻¹⁰
1q32.2	210468999	rs60154123	T	C	0.15	Intergenic	<i>HHAT</i> ,*§ <i>SERTAD4</i> ,†‡ <i>DIEXF</i> †	MA	1.05 (1.03–1.06)	2.5×10 ⁻⁸
1q42.2	230845794	rs699	A	G	0.58	Missense	<i>AGT</i> ,*§‡ <i>CAPN9</i> ,‡ <i>GNPAT</i> †	CA	0.96 (0.95–0.98)	2.1×10 ⁻⁸
2p21	45896437	rs582384	A	C	0.53	Intronic	<i>PRKCE</i> ,* <i>TMEM247</i> ‡	MA	1.03 (1.02–1.05)	7.6×10 ⁻⁹
2q24.3	164957251	rs12999907	A	G	0.82	Intergenic	<i>FIGN</i> *†	UK	1.06 (1.04–1.07)	2.4×10 ⁻¹¹
2q32.1	188196469	rs840616	T	C	0.35	Intergenic	<i>CALCRL</i> ,*†‡§ <i>TFP1</i> §	CA	0.96 (0.95–0.97)	3.0×10 ⁻⁹
2q37.3	238223955	rs11677932	A	G	0.32	Intergenic	<i>COL6A3</i> *†‡	MA	0.97 (0.96–0.98)	2.6×10 ⁻⁸
3p21.31	46688562	rs7633770	A	G	0.41	Intergenic	<i>ALS2CL</i> ,*†‡§ <i>RTP3</i> ‡	MA	1.03 (1.02–1.04)	1.1×10 ⁻⁸
3p21.31	48193515	rs7617773	T	C	0.67	Intergenic	<i>CDC25A</i> ,* <i>SPINK8</i> ,§ <i>MAP4</i> ,§ <i>ZNF589</i> §	UK	1.04 (1.03–1.05)	2.3×10 ⁻¹¹
3q22.1	132257961	rs10512861	T	G	0.14	Downstream	<i>DNAJC13</i> ,*† <i>NPHP3</i> ,† <i>ACAD11</i> ,‡ <i>UBA5</i> ‡	CA	0.96 (0.94–0.97)	1.5×10 ⁻⁸
3q22.3	136069472	rs667920	T	G	0.78	Intronic	<i>STAG1</i> ,*†‡ <i>MSL2</i> ,† <i>NCK1</i> ,† <i>PPP2R3A</i> †	MA	1.05 (1.04–1.06)	6.0×10 ⁻¹⁵
3q25.31	156852592	rs4266144	C	G	0.68	Intergenic	<i>CCNL1</i> ,* <i>TIPARP</i> †§	MA	0.97 (0.95–0.98)	1.4×10 ⁻⁸
3q26.31	172115902	rs12897	A	G	0.59	UTR3	<i>FNDC3B</i> *†§	CA	0.96 (0.95–0.97)	1.9×10 ⁻¹⁰
4p16.3	3449652	rs16844401	A	G	0.07	Missense	<i>HGFAC</i> ,*†‡ <i>RGS12</i> ,* <i>MSANTD1</i> ‡	CA	1.07 (1.04–1.10)	4.0×10 ⁻⁸
4q21.1	77416627	rs12500824	A	G	0.36	Intronic	<i>SHROOM3</i> ,*†‡§ <i>SEPT11</i> ,‡ <i>FAM47E</i> ,‡ <i>STBD1</i> †	UK	1.04 (1.03–1.05)	4.1×10 ⁻¹⁰
4q21.22	82587050	rs11099493	A	G	0.69	Intergenic	<i>HNRNP</i> ,* <i>RASGEF1B</i> †	UK	1.04 (1.03–1.06)	5.1×10 ⁻¹⁰
4q22.3	96117371	rs3775058	A	T	0.23	Intronic	<i>UNC5C</i> *†	MA	1.04 (1.03–1.05)	7.6×10 ⁻⁹
4q32.3	169687725	rs7696431	T	G	0.51	Intronic	<i>PALLD</i> ,*†‡ <i>DDX60L</i> ‡	UK	1.04 (1.02–1.05)	2.7×10 ⁻⁸
5p15.31	9556694	rs1508798	T	C	0.81	Intergenic	<i>SEMA5A</i> ,*†§ <i>TAS2R1</i> †	CA	1.05 (1.04–1.07)	4.8×10 ⁻¹³
5q11.2	55860781	rs3936511	A	G	0.82	Intronic	<i>MAP3K1</i> ,*†‡ <i>MIER3</i> †	MA	0.96 (0.95–0.98)	3.7×10 ⁻⁸
6p25.3	1617143	rs9501744	T	C	0.13	Intergenic	<i>FOXC1</i> *†	CA	0.95 (0.93–0.96)	2.2×10 ⁻⁸
6p21.2	36638636	rs1321309	A	G	0.49	Intergenic	<i>CDKN1A</i> ,*†§ <i>PI16</i> †	MA	1.03 (1.02–1.04)	3.4×10 ⁻⁸
6p21.1	43758873	rs6905288	A	G	0.57	Intergenic	<i>VEGFA</i> ,*† <i>MRPL14</i> ,‡ <i>TMEM63B</i> ‡	UK	1.05 (1.03–1.06)	1.9×10 ⁻¹²
6p11.2	57160572	rs9367716	T	G	0.32	Intergenic	<i>PRIM2</i> ,* <i>RAB23</i> ,§ <i>DST</i> ,‡ <i>BEND6</i> ‡	CA	0.96 (0.95–0.97)	9.6×10 ⁻¹⁰
6q14.1	82612271	rs4613862	A	C	0.53	Intergenic	<i>FAM46A</i> *†‡	MA	1.03 (1.02–1.04)	6.5×10 ⁻¹⁰
6q22.32	126717064	rs1591805	A	G	0.49	Intergenic	<i>CENPW</i> *§	UK	1.04 (1.03–1.06)	2.1×10 ⁻¹⁰
6q25.1	150997401	rs17080091	T	C	0.08	Intronic	<i>PLEKHG1</i> ,*† <i>IYD</i> †	MA	0.95 (0.93–0.96)	6.0×10 ⁻⁹
7p22.3	1937261	rs10267593	A	G	0.20	Intronic	<i>MAD1L1</i> *§	MA	0.96 (0.95–0.98)	1.8×10 ⁻⁸
7p22.1	6486067	rs7797644	T	C	0.23	Intronic	<i>DAGLB</i> ,*§‡ <i>RAC1</i> ,†‡§ <i>FAM220A</i> ,§ <i>KDEL2</i> †	MA	0.96 (0.95–0.98)	2.1×10 ⁻⁸
7p21.3	12261911	rs11509880	A	G	0.36	Intronic	<i>TMEM106B</i> ,*§ <i>THSD7A</i> †	CA	1.04 (1.02–1.05)	2.8×10 ⁻⁸
7p13	45077978	rs2107732	A	G	0.09	Missense	<i>CCM2</i> ,*§‡ <i>MYO1G</i> †	MA	0.94 (0.93–0.96)	3.6×10 ⁻⁸
7q31.2	117332914	rs975722	A	G	0.60	Intergenic	<i>CTTNBP2</i> ,* <i>CFTR</i> ,† <i>ASZ1</i> †	MA	0.97 (0.96–0.98)	4.1×10 ⁻⁸
8p22	18286997	rs6997340	T	C	0.31	Intergenic	<i>NAT2</i> *†	CA	1.04 (1.02–1.05)	4.9×10 ⁻⁹
8p21.3	22033615	rs6984210	C	G	0.94	Intronic	<i>BMP1</i> ,*†§ <i>SFTPC</i> ,† <i>DMTN</i> ,§ <i>PHYHIP</i> ,§ <i>DOK2</i> ,‡ <i>XPO7</i> †	UK	0.92 (0.90–0.94)	2.1×10 ⁻¹¹
8q23.1	106565414	rs10093110	A	G	0.42	Intronic	<i>ZFPM2</i> *†‡	MA	0.97 (0.96–0.98)	1.8×10 ⁻⁸
9q31.2	110517794	rs944172	T	C	0.72	Intergenic	<i>KLF4</i> *†	UK	0.96 (0.95–0.97)	1.1×10 ⁻¹¹

(Continued)

Table 1. Continued

Cytoband	Position	Lead SNP	A1	A2	Freq	Variant Function	Candidate Genes	Resource	OR (95% CI)	P Value
9q33.2	124420173	rs885150	T	C	0.73	Intronic	<i>DAB2IP</i> *†‡§	MA	0.97 (0.95–0.98)	7.8×10 ⁻¹⁰
10p13	12303813	rs61848342	T	C	0.64	Intergenic	<i>CDC123</i> ,* <i>NUDT5</i> ,§ <i>OPTN</i> ‡	MA	0.96 (0.95–0.98)	6.3×10 ⁻¹⁰
10q23.1	82251514	rs17680741	T	C	0.72	Intronic	<i>TSPAN14</i> ,*†§ <i>MAT1A</i> ,† <i>FAM213A</i> ‡	UK	1.05 (1.03–1.06)	2.3×10 ⁻¹¹
10q24.33	105693644	rs4918072	A	G	0.27	Intergenic	<i>STN1</i> ,*§ <i>SH3PX2</i> †	UK	1.04 (1.03–1.06)	2.6×10 ⁻⁹
10q26.13	124237612	rs4752700	A	G	0.55	Intronic	<i>HTRA1</i> ,*† <i>PLEKHA1</i> ††	MA	0.97 (0.96–0.98)	8.0×10 ⁻¹¹
11p15.4	5701074	rs11601507	A	C	0.07	Missense	<i>TRIM5</i> ,*‡; <i>TRIM22</i> ,* <i>TRIM6</i> ,‡ <i>OR52N1</i> ,‡ <i>OR52B6</i> ‡	UK	1.09 (1.06–1.11)	2.1×10 ⁻¹²
11p11.2	43696917	rs7116641	T	G	0.69	Intergenic	<i>HSD17B12</i> *	MA	0.97 (0.96–0.98)	1.0×10 ⁻⁸
11q22.1	100624599	rs7947761	A	G	0.72	Intronic	<i>ARHGAP42</i> *†	CA	0.96 (0.95–0.97)	3.0×10 ⁻⁹
12p13.31	7175872	rs11838267	T	C	0.87	Intronic	<i>C1S</i> *†‡	MA	1.05 (1.04–1.07)	6.1×10 ⁻¹⁰
12q22	95355541	rs7306455	A	G	0.10	Intergenic	<i>NDUFA12</i> ,* <i>FGD6</i> †	CA	0.95 (0.93–0.96)	1.0×10 ⁻⁸
13q13.1	33058333	rs9591012	A	G	0.34	Intronic	<i>N4BP2L2</i> ,‡§ <i>PDS5B</i> §†	CA	0.96 (0.94–0.97)	7.0×10 ⁻¹¹
13q34	113631780	rs1317507	A	C	0.26	Intronic	<i>MCF2L</i> ,*§ <i>PCID2</i> ,‡ <i>CUL4A</i> ‡	CA	1.04 (1.03–1.06)	8.4×10 ⁻¹²
14q23.1	58794001	rs2145598	A	G	0.58	Intronic	<i>ARID4A</i> ,*†§ <i>PSMA3</i> ‡	MA	0.97 (0.96–0.98)	4.2×10 ⁻⁸
14q32.13	94838142	rs112635299	T	G	0.02	Intergenic	<i>SERPINA2</i> ,*† <i>SERPINA1</i> *‡	MA	0.87 (0.84–0.91)	8.4×10 ⁻¹⁰
15q26.2	96146414	rs17581137	A	C	0.75	Intergenic		MA	1.04 (1.02–1.05)	1.2×10 ⁻⁸
16q23.3	81906423	rs7199941	A	G	0.40	Intronic	<i>PLCG2</i> ,*†‡ <i>CENPN</i> §	CA	1.04 (1.03–1.05)	9.2×10 ⁻¹³
17q11.2	27941886	rs13723	A	G	0.51	UTR3	<i>CORO6</i> ,*§ <i>ANKRD13B</i> ,*§ <i>GIT1</i> ,§ <i>SSH2</i> ,†‡§ <i>EFCAB5</i> ‡	CA	0.96 (0.95–0.97)	5.6×10 ⁻¹⁰
17q11.2	30033514	rs76954792	T	C	0.22	Intergenic	<i>COPRS</i> ,*§ <i>RAB11FIP4</i> ‡	MA	1.04 (1.03–1.05)	1.1×10 ⁻⁸
17q21.2	40257163	rs2074158	T	C	0.82	Missense	<i>DHX58</i> ,*§‡; <i>KAT2A</i> ,*† <i>RAB5C</i> ,†§ <i>NKIRAS2</i> ,‡ <i>DNAJC7</i> ,‡ <i>KCNH4</i> ,‡ <i>HCRT</i> ,‡ <i>GHDC</i> ‡	CA	0.95 (0.93–0.96)	2.2×10 ⁻¹⁰
18q21.1	47229717	rs9964304	A	C	0.72	Intergenic	<i>ACAA2</i> ,*† <i>RPL17</i> ‡	MA	0.96 (0.95–0.97)	1.1×10 ⁻⁹
19p13.11	17855763	rs73015714	C	G	0.80	Intergenic	<i>MAP1S</i> ,*†† <i>FCHO1</i> ,*§ <i>COLGALT1</i> ‡	CA	0.94 (0.93–0.96)	8.3×10 ⁻¹⁴
20q12	39924279	rs6102343	A	G	0.25	Intronic	<i>ZHX3</i> ,* <i>PLCG1</i> ,†‡§ <i>TOP1</i> ††	MA	1.04 (1.02–1.05)	1.1×10 ⁻⁸
20q13.12	44586023	rs3827066	T	C	0.14	Intronic	<i>PCIF1</i> ,*† <i>ZNF335</i> ,*† <i>NEURL2</i> ,§ <i>PLTP</i> †§	MA	1.04 (1.03–1.06)	4.4×10 ⁻⁹
20q13.32	57714025	rs260020	T	C	0.13	Intergenic	<i>ZNF831</i> *	MA	1.05 (1.04–1.07)	7.9×10 ⁻¹⁰
21q21.3	30533076	rs2832227	A	G	0.82	Intronic	<i>MAP3K7CL</i> ,*§ <i>BACH1</i> †	UK	0.96 (0.94–0.97)	1.7×10 ⁻⁹

List of novel CAD-associated replicating ($P < 0.05$ Bonferroni adjusted, direction of effect consistent) and surpassing the genome-wide significance threshold in meta-analysis of UK Biobank and CARDIoGRAMplusC4D. Full details are shown in Online Tables I and III through XII. Results are shown for the discovery, replication, and combined meta-analysis. CA indicates CARDIoGRAMplusC4D as discovery and UK Biobank as replication; CAD, coronary artery disease; CI, confidence interval; Freq, frequency of the effect allele; GWAS, genome-wide association study; MA, discovery by being genome-wide significant in the GWAS meta-analysis; OR, odds ratio; SNP, single-nucleotide polymorphism; UK, UK Biobank as discovery and CARDIoGRAMplusC4D as replication; and Utr, untranslated region.

*Nearest.

†Depict.

‡Hi-C.

§eQTL.

‡Coding variants.

in 156 meta gene sets (Online Table XVII). The 4 most significant metaset were complete embryonic lethality during organogenesis, blood vessel development, anemia, and SRC PPI subnetwork. The platelet α -granule lumen, SRC PPI subnetwork, blood vessel development, and hemostasis had the largest betweenness centrality—an indicator of a node's centrality in the network. The tissue enrichment analyses by DEPICT indicated blood vessels as the most relevant tissue ($P = 4 \times 10^{-7}$); 41 additional tissues or cell types were significantly enriched

at $FDR < 0.05$ (Online Table XVIII). We compared the contribution of novel information with previous work. The previous CARDIoGRAMplusC4D analysis led to 457 reconstituted gene sets (at $FDR < 0.05$); the addition of the intermediate data set UK Biobank of 150 k individuals identified a total of 889 significant gene sets, substantially less than the current 1525 gene sets (Figure 1; Online Table XVII). Considering all 10 968 possible gene sets, this study represents an increase from 4.16% to 13.90% of all gene sets involved in CAD since

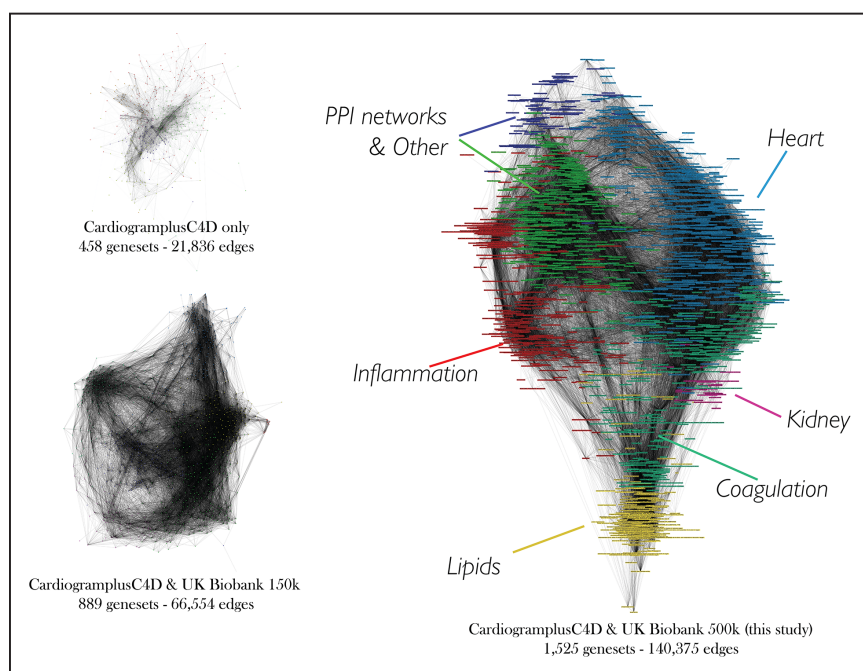


Figure 1. Network analyses of reconstituted gene sets. The total number of significant gene sets involved in coronary artery disease (CAD) increased to 13.90% since the 1000 Genome genome-wide association studies of CARDIoGRAMplusC4D, considering all possible gene sets. Clustering by modularity using Gephi software indicated that pathways specific for cardiovascular/heart development, inflammation, lipids, kidney and coagulation clustered together. PPI networks & Other indicates a remaining bin predominantly populated by protein-protein interaction networks.

the 1000 Genomes analysis of CARDIoGRAMplusC4D in 2015. Genes implicated by DEPICT on the $FDR < 0.05$ level are 94 in the previous data, which has increased to 540 genes.

Insights in Loci by Associations With Other Phenotypes

To increase our understanding of potentially mediating mechanisms at the genetic variant level, we searched the GWAS catalog for previously reported variants. Of the 64 novel loci, 23 loci were in linkage disequilibrium ($r^2 > 0.6$) with genetic variants previously reported to be associated with other traits surpassing the genome-wide significant ($P < 5 \times 10^{-8}$) threshold (Online Table XIX). We found associations with anthropometric measurements (rs6905288, rs1591805, rs3936511, and rs840616), antineutrophil antibody-associated vasculitis (rs112635299), angiotensinogen measurements (rs699), coffee consumption (rs13723), C-reactive protein (rs667920), pulmonary function (rs61848342, rs13723, and rs112635299), fibrinogen levels (rs67920, rs16844401, and rs2074158), glomerular filtration rate (rs12500824), high-density lipoprotein cholesterol (rs667920, rs10512861, and rs6905288), low-density lipoprotein cholesterol (rs10512861), total cholesterol (rs6997340), triglycerides (rs667920, rs3936511, rs6905288, and rs6997340), diabetes mellitus (rs1591805 and rs3936511), blood pressure indices (rs260020, rs17080091, rs61776719, rs7696431, and rs1317507), transferrin levels (rs6997340), QRS amplitude (rs13723), abdominal aortic aneurysm (rs885150 and rs3827066), adiponectin measurements (rs6905288), and age at menarche (rs1591805); full details can be found in Online Table XIX. We also explored the association of the 64 lead SNPs with a range of traits in UK Biobank resource. Consistent with the GWAS-catalog search and in keeping with earlier observations in established CAD loci, several of our novel loci were associated with hyperlipidemia, blood pressure traits, diabetes mellitus, and anthropometric traits (Figure 2). For example, rs6905288 (*VEGFA*)

was also associated with waist-to-hip ratio and hyperlipidemia, and rs61776719 (*FHL3* and *UTP11L*) was also closely associated with pulse pressure in UK Biobank. Interestingly, we observed that 15 of 64 loci were associated with platelet counts.

Genetic Risk for CAD, and Association With CAD Risk Factors and Outcome

To explore potential clinical relevance, we constructed a GRS, weighted for their effects in CARDIoGRAMplusC4D by multiplying the effect sizes with the number of effect variants of each variant in each individual, and divided this GRS into quintiles. The associations with many different traits and diseases from the UK Biobank are visualized in Figure 2. The risk of a future diagnosis of atrial fibrillation and heart failure in UK Biobank participants was higher in quantile 5 individuals as compared with quantile 1 (hazard ratio, 1.18 [95% confidence interval, 1.10–1.27; $P = 1.2 \times 10^{-6}$] and 1.59 [95% confidence interval, 1.43–1.77; $P = 3.3 \times 10^{-18}$], respectively; Online Figure IV). In addition, all-cause mortality and especially cardiovascular mortality was higher in individuals of quantile 5 compared with quantile 1 (hazard ratio, 1.12 [95% confidence interval, 1.06–1.19; $P = 4 \times 10^{-4}$] and 1.94 [95% confidence interval, 1.70–2.21; $P = 2 \times 10^{-23}$], respectively; Online Figure IV).

Role of Regulatory DNA and Fine Mapping of Candidate Causal Variants

Across the genome, virtually all tissues showed significant enrichment of DNase I hypersensitivity sites providing limited indications for involved biology (Figure 3A and 3B). Minimal differential enrichment of functional elements for the identified genetic loci was observed in blood vessels and liver. To facilitate future functional studies directed at causal variants and molecular mechanisms, we prioritized variants via the probabilistic framework of Probabilistic Annotation Integrator. Because no clear differential enrichment was observed for tissue-specific functional elements, we focused on DNA annotations from the

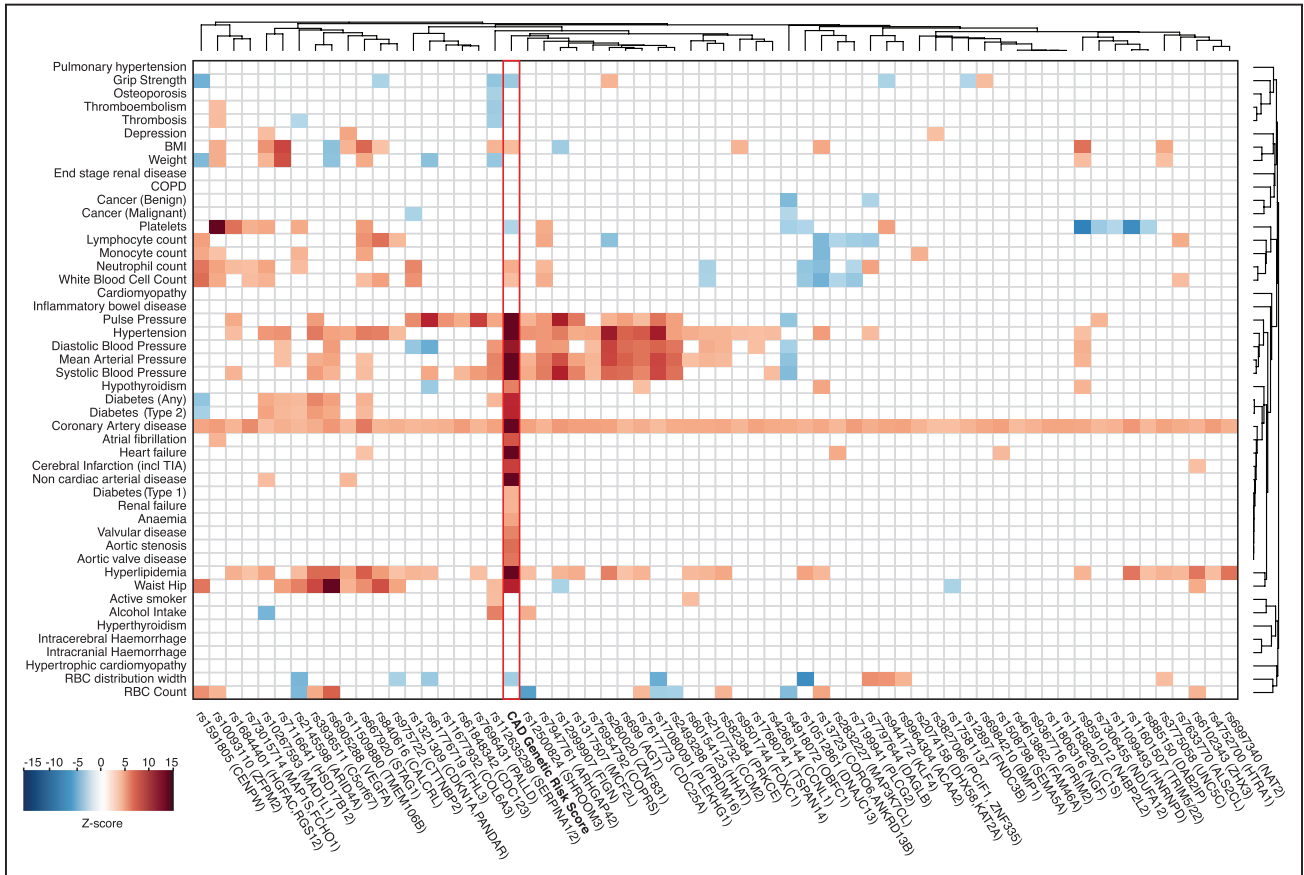


Figure 2. Heatmap of associations in UK Biobank with novel loci. Heatmap of z scores for different diseases and phenotypes in UK Biobank, aligned to increased risk of coronary artery disease. Only significant associations (false discovery rate<0.01) are shown. The genetic risk score constructed with the known and novel loci, weighted using coefficients of CARDIoGRAMplusC4D, is highlighted by the red rectangle. BMI indicates body mass index; COPD, chronic obstructive pulmonary disease; RBC, red blood cell; and TIA, transient ischemic attack.

study of Finucane et al¹⁷ that are not specific for tissue or cell types. Probabilistic Annotation Integrator determined the significance of each annotation to be causal (Figure 3C and 3D), and a model was constructed using linkage disequilibrium information, *P* value distribution, and information on coding variation, conservation and H3K4me1 sites to prioritize potential causal SNPs of all 161 (known and novel) loci. This analysis yielded 28 variants $\geq 95\%$ confidence level for which we prioritized candidate genes (Online Table XX; Table 2).

For example, rs974819 was prioritized as causal variant and could be linked to PDGFD by Hi-C evidence and eQTL data in relevant tissues (Online Figure V). In total, 15 of the 28 fine-mapped loci could be pinpointed to 1 single potential causal mechanism implicating a single gene. For 2 loci, there were 2 potential causal mechanisms (*TRPC4AP/PROCR* and *MRPS6/SLC5A3*) with equal evidence.

Discussion

The present study is the largest genetic association study of CAD performed to date. We report on the primary results and downstream bioinformatic analyses of the meta-analysis of de novo GWAS data derived from the UK Biobank combined with existing data from CARDIoGRAMplusC4D, leading to the inclusion of $\leq 122\,733$ cases and 424 528 controls. This study contributes to the existing literature by reporting 64 novel genetic

loci representing 38% of all 161 GWAS-identified CAD loci to date.¹⁸ For the novel loci, a detailed catalog of 155 candidate genes (based on proximity, gene-expression data, coding variation, and physical chromatin interaction) is provided. We demonstrate that the increase in significantly associated CAD loci results in a large expansion of implicated reconstituted gene networks, from 4% to almost 14%. Finally, by integrating genetic association strength, linkage disequilibrium, and functional annotation data, we performed fine mapping of all 161 CAD loci, providing a novel credible list of causal variants and plausible genes to be prioritized for functional validation.

The 64 novel genetic loci reported in this single article are exceptionally large compared with previous articles, including those of CARDIoGRAMplusC4D and others reporting on 10 to 15 novel loci each.²⁻¹⁰ Thirty-four of the 64 loci are significant in a robust reciprocal replication strategy between CARDIoGRAMplusC4D and the UK Biobank, but another 30 are genome-wide significant in the overall meta-analysis as is commonly considered sufficient evidence.^{7,10} The obvious reason for the large number of novel loci is the considerable number of novel CAD cases and non-CAD controls compared with these earlier efforts combined with less heterogeneity in samples, collection, and definitions used. By increasing the sample size, more loci can be identified, more genes can be implicated, and more gene networks or pathways can

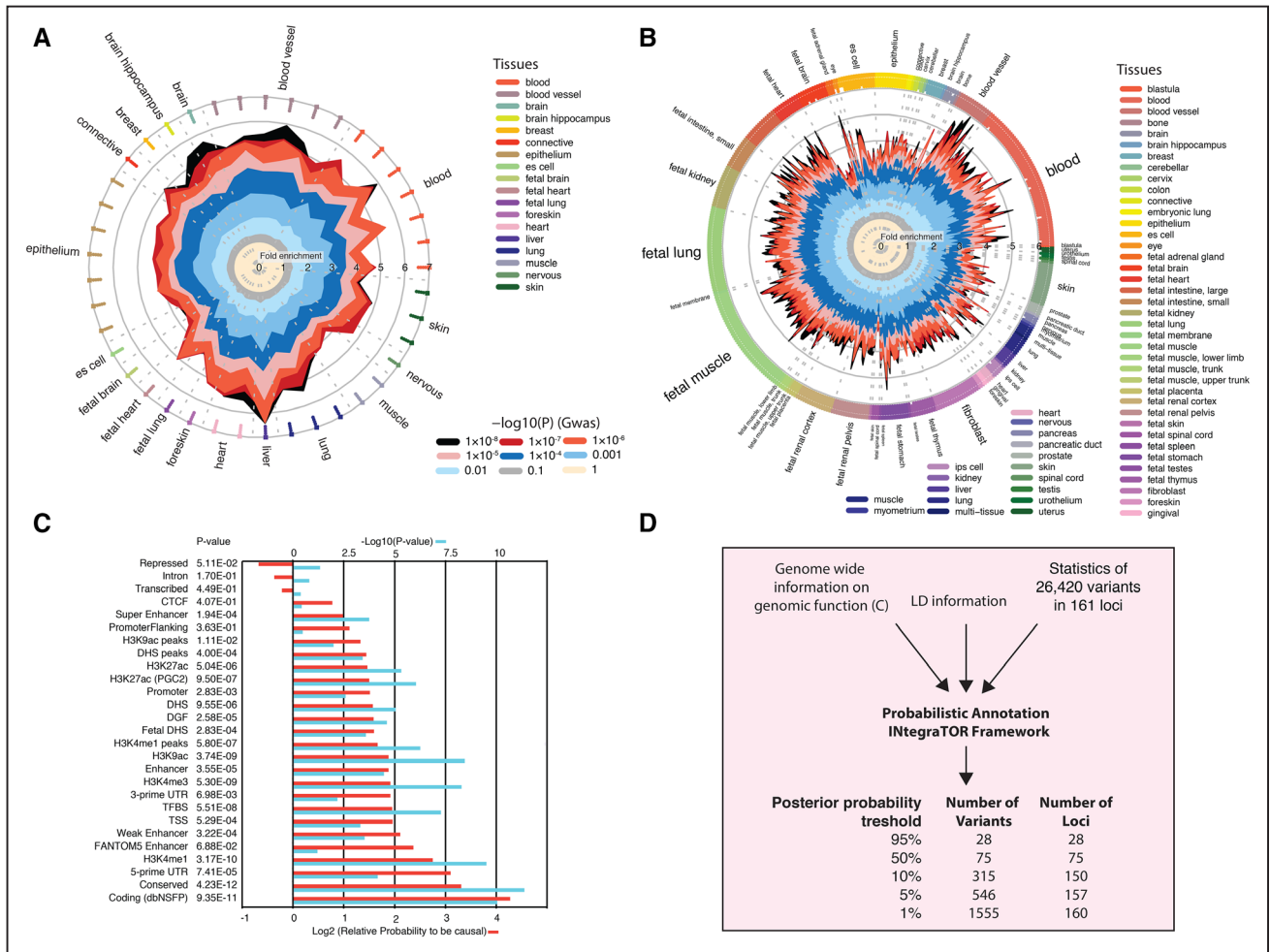


Figure 3. The role of regulatory DNA underlying coronary artery disease (CAD)-associated single-nucleotide polymorphisms (SNPs). Enrichment of genome-wide association study (GWAS) threshold SNPs were significantly enriched in DHS footprints (A) and hot spots (B) across many different tissues and cell types. The fold enrichment was highly significant for most tissues and cell types ($P < 1 \times 10^{-6}$) as indicated by the 4 colored circles next to the labels, 3 colored circles indicate $P < 1 \times 10^{-7}$. Label sizes of tissue types were downsized because of space limitations; tissue types may be represented by multiple samples, indicated by hash marks of the same color. C, Subsequent prioritization of potential causal annotations underlying the 161 CAD loci also suggested that regions of DHS may be underlying the associations, but coding variants, conservation, 5' untranslated region (UTR), and H3K4me1 annotations were more likely to be causal. D, Posterior probabilities for causality for each variant in the 164 CAD loci were calculated by an empirical Bayes approach implemented in the Probabilistic Annotation Integrator Framework, taking into account linkage disequilibrium (LD), association statistics, and the potentially causal annotations and summarized in Table 2 and Online Table XX. CTCF indicates transcriptional repressor CTCF; DGF, digital genomic footprint by DNase1 hypersensitivity; FANTOM5, functional annotation of the mammalian genome V5; TFBS, transcription factor binding site; and TSS, transcription start site.

be constructed. Not only is the increase of associated loci in the past decade rapidly outpacing functional validation, even understanding biological networks seems to insufficiently accommodate the increased amount of GWAS hits under the conceptual polygenic model. This can be illustrated by the large increase of reconstituted gene networks observed in our study. For the first time, we show that almost 14% of all existing gene networks are involved in the complex CAD trait (Figure 1), and this will only increase when further samples are added to the GWAS study making it increasingly more difficult to consider these all to be key pathways. In our data, we also observed genetic association signals to be spread across most of the genome, and many of the novel 155 candidate genes do not have an obvious connection to CAD. In addition, virtually all cell types showed significant enrichment of DNase I hypersensitivity and other functional elements. These

notions are all supportive of the omnigenic model, which has recently been proposed by the Pritchard team suggesting that prevailing conceptual models for complex diseases are incomplete. The omnigenic model hypothesizes that all gene-regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells can influence the function of core disease-related genes and a major proportion of heritability can be explained by effects of genes outside key pathways.¹⁹ To further our knowledge, it is questionable whether further increasing the GWAS sample size will resolve the outstanding issues concerning our incomplete understanding of cellular regulatory networks and our ability to differentiate core genes from peripheral genes. If the omnigenic model is indeed correct, detailed mapping of cell-specific regulatory networks will be essential to understand CAD.

Table 2. For 28 Loci, the 95% Credible Set of Causal Variants Consisted of a Single Coronary Artery Disease Variant

Cytoband	Causal Variant	No. of SNPs in Locus	MAF (EUR)	GWAS <i>P</i> Value	Posterior <i>P</i> Value	Annotation	Candidate Gene/Mechanism
1p32.3	rs11591147	4	0.02	1.9×10 ⁻²²	1.00	Missense (T)	<i>PCSK9</i> †
1p13.3	rs602633	67	0.21	3.6×10 ⁻⁵⁸	1.00	Downstream	<i>SORT1</i> †‡§ ^(51.1) , <i>SARS</i> † ^(11.6) , <i>PSRC1</i> † ⁽¹⁵²⁾ , <i>CELSR2</i> † ⁽¹⁰⁸⁾ , <i>ATXN7L2</i> † ^(11.7)
1q32.1	rs6700559	83	0.49	1.8×10 ⁻⁰⁸	0.97	Intronic	<i>CAMSAP2</i> †‡§ ^(9.5) § ^(23.9) , <i>DDX59</i> † ^(42.0)
2p24.1	rs16986953	66	0.07	1.1×10 ⁻¹⁶	1.00	Intergenic	...
2q35	rs2571445	50	0.40	1.6×10 ⁻¹²	0.97	Missense (T)	<i>TNS1</i> *†‡ ^(121.5) , <i>DIRC3</i> † ^(7.9)
3p21.31	rs7633770	49	0.44	1.1×10 ⁻⁰⁸	0.97	Intergenic	<i>ALS2CL</i> † ^(8.6) , <i>RTP3</i> § ^(49.6) , <i>LTF5</i> § ^(49.6)
3q26.31	rs12897	67	0.41	1.2×10 ⁻⁰⁹	1.00	UTR3	<i>FNDCC3B</i> †‡¶ ^(6.8)
4q21.22	rs11099493	54	0.37	2.5×10 ⁻¹⁰	1.00	Intergenic	...
5q31.3	rs246600	15	0.50	6.5×10 ⁻¹⁷	1.00	Intronic	<i>HMHBI</i> § ^(159.2)
6p24.1	rs9349379	268	0.41	2.7×10 ⁻⁷⁶	1.00	Intronic	<i>EDN1</i> †‡ ^(2.2) § ^(23.9) , <i>TBC1D7</i> † ^(15.9) , <i>PHACTR1</i> † ^(55.7) , <i>GFOD1</i> † ^(8.1)
7p21.1	rs2107595	71	0.18	1.3×10 ⁻²⁴	1.00	Intergenic	<i>TWIST1</i> † ^(36.8)
7p13	rs2107732	11	0.10	3.6×10 ⁻⁰⁸	0.98	Missense (T)	<i>CCM2</i> *†§ ^(9.6) , <i>MYO1G</i> § ^(22.9)
7q32.2	rs11556924	95	0.38	1.4×10 ⁻²³	1.00	Missense (D)	<i>ZC3HC1</i> *†, <i>NRF1</i> § ⁽³⁸⁾ , <i>KLF14</i> § ^(216.9)
7q36.1	rs3918226	25	0.09	1.4×10 ⁻²⁰	1.00	Intronic	<i>NOS3</i> † ^(6.0)
9p21.3	rs4977574	161	0.49	8.8×10 ⁻²²³	1.00	Intronic	<i>CDKN2B</i> †‡ ^(4.7) § ⁽¹³³⁾ , <i>MTAP5</i> † ⁽¹⁶⁸⁾
11p15.4	rs11601507	3	0.06	5.6×10 ⁻¹³	1.00	Missense (D)	<i>TRIM5</i> †, <i>OR52N1</i> § ⁽⁴⁵⁾ , <i>TRIM6</i> § ⁽⁴⁹⁾ , <i>OR52B6</i> § ⁽⁴⁹⁾
11q13.1	rs3741380	207	0.48	2.8×10 ⁻¹¹	0.95	Missense (T)	<i>EHBP1L1</i> *†‡ ^(51.5)
11q13.5	rs590121	122	0.28	1.5×10 ⁻¹⁰	0.98	Intronic	<i>SERPINH1</i> † ^(5.5) , <i>KLHL35</i> § ^(23.7)
11q22.3	rs974819	428	0.24	1.1×10 ⁻²⁸	0.99	Intergenic	<i>PDGFD1</i> † ^(20.6) , § ^(87.0)
13q34	rs11617955	19	0.11	6.9×10 ⁻¹⁸	1.00	Intronic	...
13q34	rs1317507	94	0.25	8.2×10 ⁻¹²	1.00	Intronic	<i>PCID2</i> § ^(22.5) , <i>CUL4A</i> § ^(22.5)
15q25.1	rs7173743	367	0.46	5.5×10 ⁻³⁶	0.96	Intergenic	<i>RASGRF1</i> † ^(4.2) , <i>ADAMTS7</i> † ^(33.3)
16q23.3	rs7500448	257	0.22	1.6×10 ⁻¹⁶	1.00	Intronic	<i>CDH13</i> †‡ ^(70.2) § ^(64.8)
17q21.32	rs17608766	178	0.17	8.2×10 ⁻¹⁰	1.00	UTR3	<i>GOSR2</i>
19p13.2	rs116843064	7	0.03	3.6×10 ⁻¹⁰	1.00	Missense (D)	<i>ANGPTL4</i> *†
19q13.32	rs7412	39	0.07	2.1×10 ⁻³⁵	1.00	Missense (D)	<i>APOE</i> *†, <i>APOC2</i> § ^(64.1) , <i>CLPTM1</i> § ^(64.1) , <i>APOC4</i> § ^(64.1)
20q11.22	rs867186	>500	0.10	6.8×10 ⁻¹²	0.97	Missense (T)	<i>PROCR</i> *†‡ ^(20.3) , <i>TRPC4AP1</i> †‡§ ⁽¹¹⁶⁾ § ⁽¹¹⁶⁾ , <i>GGT7</i> † ^(4.6) , <i>EDEM2</i> † ^(7.9) , <i>NCOA6</i> § ^(75.1) , <i>HMG3P1</i> § ^(75.1)
21q22.11	rs28451064	104	0.13	2.6×10 ⁻³³	1.00	Intergenic	<i>MRPS61</i> † ^(17.4) § ^(238.6) , <i>SLC5A3</i> †‡§ ^(32.1) § ^(238.6)

Online Table XX contains full details of the loci on the variant level. The number in brackets indicates the significance $-\log_{10}$ of the eQTL or chromatin interaction. D indicates deleterious (*SIF7*); eQTL, expression quantitative trait locus; EUR, European; GWAS, genome-wide association studies; LD, linkage disequilibrium; MAF, minor allele frequency; SNP, single-nucleotide polymorphism; T, tolerated (*SIF7*); and UTR, untranslated region.

*Gene with a missense causal variant.

†Genes indicate converging evidence of a potential functional SNP-gene mechanism, further described in the methods.

‡eQTL gene.

§Chromatin interaction between the causal variant and the promoter of the gene.

||SNPs of novel loci.

¶Gene of which the 3' untranslated region overlaps with the causal variant.

Number of SNPs in locus correspond to the number of SNPs with a $P < 0.01$ that are in low LD ($r^2 > 0.1$) with the sentinel SNP.

To facilitate functional research based on our findings, we not only provided extensive bioinformatic analyses of coding variation, gene expression, and chromatin interactions for the 64 novel loci but also performed novel fine mapping and

presented statistically convincing arguments for causal genetic variants at 28 loci, linking 19 genes in the 161 CAD loci. In the known loci, these genes included *APOE*, *PCSK9*, *ANGPTL4*, and *SORT1*, all implicated as core genes in lipid metabolism.

Recently, *PCSK9* has been validated in clinical trials,²⁰ and functional studies are also supporting a key role for *SORT1*.²¹ More recently, *EDNI* has indeed been identified as the likely causal gene in the pathogenesis of CAD instead of the nearby *PHACTR*.²² In the novel loci, we found evidence for causal variants linked to *FNDC3B* (Fibronectin Type III Domain Containing 3B), *CCM2* (*CCM2* Scaffolding Protein), and *TRIM5* (Tripartite Motif Containing 5). Indeed, the functional link between these genes and CAD is not obvious and remains to be determined. *FNDC3B* has been suggested to function as a positive regulator of adipogenesis.²³ *CCM2* has been implicated in abnormal vascular morphogenesis in the brain, leading to cerebral cavernous malformations²⁴ but is also expressed in the heart. Although its effect in the coronary arteries has not been investigated, *Ccm2* knockdown in the mouse brain endothelial cells leads to increased monolayer permeability, decreased tubule formation, and reduced cell migration after wound healing.²⁵ *TRIM5* has been suggested to promote innate immune signaling, and its activity is amplified by retroviral infections.²⁶ All SNP-gene mechanisms proposed in this article should be experimentally sought out. Also, the analyses were restricted to variants available in the Haplotype Reference Consortium imputation panel. Although this is the largest imputation panel to date, it only comprised SNPs; future fine-mapping efforts are necessary that include non-SNPs as well, such as indels, to cover the additional aspects of the human variation landscape. However, a 95% credible set that contains just 1 potential causal variant per locus provides a first starting point for generating new hypotheses and scientific explorations.

In our current work, we validated our previous finding that these genetic variants of CAD also predict the risk of atrial fibrillation, heart failure,⁸ and extended it to all-cause death. We also aimed to differentiate between stable CAD and acute myocardial infarction by performing multinomial logistic regression analyses. Most loci were not driven by 1 clinical presentation specifically. However, for 2 previously identified loci (rs9349379 [*EDNI*] and rs10947789 [*KCNK5*]), we found statistical evidence that these loci may be driven by acute myocardial infarction and not stable CAD. Also, for this observation, functional hypotheses are to be developed and tested. Our variants might be driven mainly by nonfatal CAD, and different variants might exist for fatal heart disease.

Some limitations of the current work are to be acknowledged. This work is based on statistical evidence and does not provide functional experimental validation. The genetic variants identified and the genes prioritized require further direct investigations in future studies to elucidate their role, and function, in the development and progression of CAD. However, in the short term, these data open up new possibilities to improve quantitative measures of genetic risk prediction. Recent data suggests that instead of operating in a deterministic fashion, high genetic risk is indeed modifiable by lifestyle,²⁷ pharmacotherapy,²⁸ and also by incorporation of genetic risk into shared decision-making sessions with patients.²⁹

In conclusion, our GWAS, meta-analyses, and bioinformatic analyses provide several novel insights into the biology of CAD. We report 64 novel loci, link 155 candidate genes, and performed fine mapping of all old and novel loci, providing a credible list of causal genetic variants. However,

with the ever-increasing sample size, our work is the first to indicate that an omnigenic model may be more appropriate to accommodate the complex genetic architecture of CAD, compared with a polygenic model. In addition to an expanded view, it also suggests new methods and tools are required to further our understanding of CAD biology through genetics.

Acknowledgments

This research has been conducted using the UK Biobank resource under application number 12006 and 15031. We thank the CARDIoGRAMplusC4D investigators for making their data publicly available. We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high-performance computing cluster.

Sources of Funding

N. Verweij is supported by Marie Skłodowska-Curie GF (call: H2020-MSCA-IF-2014; project identifier: 661395) and an NWO VENI grant (016.186.125). We acknowledge the support from the Netherlands Cardiovascular Research Initiative—an initiative with support of the Dutch Heart Foundation, CVON2015-17 EARLY-SYNERGY.

Disclosures

None.

References

1. Wang H, Naghavi M, Allen C, Barber R, Bhutta ZA, Carter C, Casey C, Charlson F, Chen C, Coates M, Dandona H, Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388:1459–1544.
2. Samani NJ, Erdmann J, Hall AS, et al; WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443–453. doi: 10.1056/NEJMoa072366.
3. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491–1493. doi: 10.1126/science.1142842.
4. McPherson R, Pertsemlidis A, Kavavlar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488–1491. doi: 10.1126/science.1142447.
5. Schunkert H, König IR, Kathiresan S, et al; Cardiogenics; CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333–338. doi: 10.1038/ng.784.
6. Deloukas P, Kanoni S, Willenborg C, et al; CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25–33. doi: 10.1038/ng.2480.
7. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121–1130. doi: 10.1038/ng.3396.
8. Verweij N, Eppinga RN, Hagemeyer Y, van der Harst P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Sci Rep*. 2017;7:2761. doi: 10.1038/s41598-017-03062-8.
9. Howson JMM, Zhao W, Barnes DR, et al; CARDIoGRAMplusC4D; EPIC-CVD. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet*. 2017;49:1113–1119. doi: 10.1038/ng.3874.
10. Nelson CP, Goel A, Butterworth AS, et al; EPIC-CVD Consortium; CARDIoGRAMplusC4D; UK Biobank CardioMetabolic Consortium CHD Working Group. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*. 2017;49:1385–1391. doi: 10.1038/ng.3913.
11. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet*. 2017;18:331–344. doi: 10.1038/nrg.2016.160.

12. Pers TH, Karjalainen JM, Chan Y, et al; Genetic Investigation of ANthropometric Traits (GIANT) Consortium. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun.* 2015;6:5890. doi: 10.1038/ncomms6890.
13. van der Harst P, van Setten J, Verweij N, et al. 52 genetic loci influencing myocardial mass. *J Am Coll Cardiol.* 2016;68:1435–1448. doi: 10.1016/j.jacc.2016.07.729.
14. Iotchkova V, Huang J, Morris JA, et al; UK10K Consortium. Discovery and refinement of genetic loci associated with cardiometabolic risk using dense imputation maps. *Nat Genet.* 2016;48:1303–1312. doi: 10.1038/ng.3668.
15. Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, Kraft P, Pasaniuc B. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* 2014;10:e1004722. doi: 10.1371/journal.pgen.1004722.
16. Stitzel NO, Stirrups KE, Masca NG, et al; Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N Engl J Med.* 2016;374:1134–1144. doi: 10.1056/NEJMoa1507652.
17. Finucane HK, Bulik-Sullivan B, Gusev A, et al; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; RACI Consortium. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015;47:1228–1235. doi: 10.1038/ng.3404.
18. Klarin D, Zhu QM, Emdin CA, et al; CARDIoGRAMplusC4D Consortium. Genetic analysis in UK biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet.* 2017;49:1392–1397. doi: 10.1038/ng.3914.
19. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell.* 2017;169:1177–1186. doi: 10.1016/j.cell.2017.05.038.
20. Ridker PM, Revkin J, Amarenco P, et al; SPIRE Cardiovascular Outcome Investigators. Cardiovascular efficacy and safety of bococizumab in high-risk patients. *N Engl J Med.* 2017;376:1527–1539. doi: 10.1056/NEJMoa1701488.
21. Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature.* 2010;466:714–719. doi: 10.1038/nature09266.
22. Gupta RM, Hadaya J, Trehan A, et al. A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell.* 2017;170:522.e15–533.e15. doi: 10.1016/j.cell.2017.06.049.
23. Kishimoto K, Kato A, Osada S, Nishizuka M, Imagawa M. Fad104, a positive regulator of adipogenesis, negatively regulates osteoblast differentiation. *Biochem Biophys Res Commun.* 2010;397:187–191. doi: 10.1016/j.bbrc.2010.05.077.
24. Liquori CL, Berg MJ, Siegel AM, et al. Mutations in a gene encoding a novel protein containing a phosphotyrosine-binding domain cause type 2 cerebral cavernous malformations. *Am J Hum Genet.* 2003;73:1459–1464. doi: 10.1086/380314.
25. Crose LE, Hilder TL, Sciaky N, Johnson GL. Cerebral cavernous malformation 2 protein promotes smad ubiquitin regulatory factor 1-mediated RhoA degradation in endothelial cells. *J Biol Chem.* 2009;284:13301–13305. doi: 10.1074/jbc.C900009200.
26. Pertel T, Hausmann S, Morger D, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature.* 2011;472:361–365. doi: 10.1038/nature09976.
27. Khera AV, Emdin CA, Drake I, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med.* 2016;375:2349–2358. doi: 10.1056/NEJMoa1605086.
28. Mega JL, Stitzel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet.* 2015;385:2264–2271. doi: 10.1016/S0140-6736(14)61730-X.
29. Kullo IJ, Jouni H, Austin EE, Brown SA, Krusselbrink TM, Isseh IN, Haddad RA, Marroush TS, Shameer K, Olson JE, Broeckel U, Green RC, Schaid DJ, Montori VM, Bailey KR. Incorporating a genetic risk score into coronary heart disease risk estimates: effect on low-density lipoprotein cholesterol levels (the MI-GENES clinical trial). *Circulation.* 2016;133:1181–1188. doi: 10.1161/CIRCULATIONAHA.115.020109.