Integrative Prioritization of Causal Genes for Coronary Artery Disease

Ke Hao, ScD*; Raili Ermel, MD*; Katyayani Sukhavasi, MSc*; Haoxiang Cheng, PhD; Lijiang Ma, PhD; Ling Li, MSc; Letizia Amadori, PhD; Simon Koplev, MScEng; Oscar Franzén, PhD; Valentina d’Escamard, PhD; Nirupama Chandel, PhD; Kathryn Wolhuter, PhD; Nicole S. Bryce, PhD; Vamsidhar R.M. Venkata, PhD; Clint L. Miller, PhD; Arno Ruusalepp, MD, PhD; Heribert Schunkert, MD; Johan L.M. Björkegren, MD, PhD; Jason C. Kovacic, MBBS, PhD

BACKGROUND: Hundreds of candidate genes have been associated with coronary artery disease (CAD) through genome-wide association studies. However, a systematic way to understand the causal mechanism(s) of these genes, and a means to prioritize them for further study, has been lacking. This represents a major roadblock for developing novel disease- and gene-specific therapies for patients with CAD. Recently, powerful integrative genomics analyses pipelines have emerged to identify and prioritize candidate causal genes by integrating tissue/cell-specific gene expression data with genome-wide association study data sets.

METHODS: We aimed to develop a comprehensive integrative genomics analyses pipeline for CAD and to provide a prioritized list of causal CAD genes. To this end, we leveraged several complimentary informatics approaches to integrate summary statistics from CAD genome-wide association studies (from UK Biobank and CARDIoGRAMplusC4D) with transcriptomic and expression quantitative trait loci data from 9 cardiometabolic tissue/cell types in the STARNET study (Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task).

RESULTS: We identified 162 unique candidate causal CAD genes, which exerted their effect from between one and up to 7 disease-relevant tissues/cell types, including the arterial wall, blood, liver, skeletal muscle, adipose, foam cells, and macrophages. When their causal effect was ranked, the top candidate causal CAD genes were \( CDKN2B \) (associated with the 9p21.3 risk locus) and \( PHACTR1 \); both exerting their causal effect in the arterial wall. A majority of candidate causal genes were represented in cross-tissue gene regulatory co-expression networks that are involved with CAD, with 22/162 being key drivers in those networks.

CONCLUSIONS: We identified and prioritized candidate causal CAD genes, also localizing their tissue(s) of causal effect. These results should serve as a resource and facilitate targeted studies to identify the functional impact of top causal CAD genes.

Key Words: aorta • atherosclerosis • coronary artery disease • genomics • liver

Genome-wide association studies (GWAS) have been remarkably informative and provided lists of hundreds of variants that are associated with coronary artery disease (CAD).1-3 Based largely on proximity, researchers have somewhat arbitrarily inferred the genes that are most likely to be associated with these variants.4,5 Despite the success of GWAS, this raises a number of concerns. To begin with, these inferences assigning genes that are
Nonstandard Abbreviations and Acronyms

| Abbreviation | Description |
|--------------|-------------|
| BLOOD        | venous blood |
| COR          | coronary artery |
| CAD          | coronary artery disease |
| eQTL         | expression quantitative trait loci |
| GWAS         | genome-wide association studies |
| GRN          | gene regulatory co-expression network |
| GTEx         | Genotype-Tissue Expression |
| IGA          | integrative genomics analysis |
| MAM          | internal mammary artery |
| SF           | subcutaneous fat |
| SMR          | summary-based Mendelian randomization |
| STARNET      | Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task |
| TWAS         | transcriptome-wide association study |
| UKBB         | UK Biobank |
| VAF          | visceral abdominal fat |

associated with these variants rely on several assumptions and are not always correct. Furthermore, for most of these genes, we do not know which are truly causal, rather than just being associated with CAD. In addition, at present there is no overall prioritized ranking of these genes based upon which are the most important for causing CAD.

Yet another issue arising from GWAS is the lack of knowledge of which disease-relevant tissue(s) a given CAD-related gene exerts its effect in. For example, genes that might cause CAD can exert effect(s) in adipose, liver, inflammatory cells, the arterial wall, and other tissues/cell types. This lack of knowledge of both the prioritized importance of CAD genes, and also their tissue(s) of causal effect, is a major obstacle to scientific efforts to understand atherosclerosis and CAD. Indeed, at present, of the almost 300 single nucleotide polymorphisms known from GWAS to be associated with CAD, there are limited insights into the specific genes and tissues involved in modulating their CAD risk effect. On the contrary, a prioritized list of causal CAD genes, and knowledge of their tissues of causal effect, would be a key resource that would allow targeted studies to identify the functional impact of the top causal genes for CAD in appropriate tissues.

As an important advance, powerful techniques have emerged for integrating tissue and cell–specific data with GWAS data sets. These integrative genomics analysis (IGA) methodologies include the TWAS (Transcriptome-Wide Association Study), Summary-based Mendelian Randomization (SMR), MetaXcan, and Coloc. IGA approaches integrate GWAS data sets with gene expression measurements (eg, expression quantitative trait loci [eQTLs]), which permit the identification of specific genes and variants that are not only associated with CAD but which also directly govern aspects of disease pathobiology. Furthermore, IGA methodologies have the potential to determine causality and are well suited to the agnostic prioritization of causal mediators of disease pathobiology.

In terms of resources that could be used to undertake an IGA for CAD, as well as publicly available GWAS data sets, STARNET (Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task) is a genetics-of-gene expression study that now includes >1000 CAD subjects and >250 controls of European ancestry. From each subject, venous blood (BLOOD) as well as biopsies from atherosclerotic aortic wall, pre/early atherosclerotic mammary artery (MAM), liver, skeletal muscle, subcutaneous fat (SF), and visceral abdominal fat (VAF) were obtained, and RNA was extracted. BLOOD was also used to obtain macrophages and foam cells. The STARNET data sets have been extensively curated and already provided significant insights on CAD pathobiology and in particular on gene regulatory co-expression networks (GRNs) that contribute to CAD heritability. Here, we used next-generation RNA sequencing data from up to 9 different tissues/cell types that were collected from STARNET CAD subjects, and intersected this with CAD GWAS data sets, to develop a comprehensive IGA pipeline for CAD in a disease-relevant context. Resulting from this, and as a key scientific resource, we provide a prioritized list of 162 candidate causal CAD genes and the tissues in which they govern CAD risk.

METHODS

As a key resource in this study, the STARNET study has been extensively described. Briefly, after providing written informed consent, patients with angiographically proven CAD who were eligible for open-thorax surgery and control subjects without CAD were enrolled into this institutional review committee approved protocol (Ethics Review Committee on Human Research of the University of Tartu). The STARNET data is accessible through Database of Genotypes and Phenotypes, accession phs001203.v1.p1. The subsequent IGA incorporated 2 data sources: GWAS summary statistics from an interim release of UK Biobank (UKBB) data and CARDioGRAMplusC4D and tissue/cell-specific eQTLs from STARNET. The corresponding authors are also willing to address queries regarding the data or results upon reasonable request.

RESULTS

Proof-of-Concept Studies to Determine Causal Tissues and Cell Types for CAD

A study overview is shown in Figure 1. To ascertain the feasibility of determining the tissues/cells in which
genes identified by GWAS exert their effects in promoting CAD, we performed a linkage disequilibrium score regression analysis by leveraging publicly available data from BLUEPRINT\textsuperscript{16,17} and GTEx (the Genotype-Tissue Expression project),\textsuperscript{18} and GWAS data from either UKBB\textsuperscript{1} or CARDIoGRAMplusC4D.\textsuperscript{2} Linkage disequilibrium score regression integrates eQTL, gene expression, and epigenetic marks to identify disease-relevant tissues/cells. From the multiple diverse tissues represented in this analysis, the majority of which are not related to the heart or vasculature, we identified a clear tissue enrichment signal that the pathobiology of CAD is predominantly driven by tissues/cells of the cardiovascular and immune systems (Tables S2 and S3). This unbiased analysis indicates that it is possible to determine the tissues/cells that promote CAD by integrating GWAS and epigenomic data sets.

**IGA Identifies and Prioritizes Candidate Causal Genes for CAD**

Our IGA pipeline incorporated 2 sources of data: GWAS summary statistics (from both UKBB and CARDioGRAMplusC4D) and tissue/cell-specific eQTLs from STARNET. Our IGA employed 3 methods from 2 broad classes: MetaXcan and SMR (class 1) and Coloc (class 2). We intersected the results of class 1 and 2 methods to identify a set of likely causal CAD genes. In total, 197,888 class 1 tests (MetaXcan and SMR, Table S4) were conducted, on which we calibrated the false discovery rate (FDR). Findings at ≤5% FDR were further filtered by genetic co-localization posterior probability estimated by Coloc.

Using the UKBB and CARDioGRAMplusC4D GWASs, our IGA pipeline revealed 129 and 121 candidate CAD causal genes, respectively (Tables S5 and S6). Genes demonstrating the strongest MetaXcan evidence ($P<5\times10^{-8}$) were visualized in Figure 2. The STARNET eQTLs and this IGA pipeline allowed us to pinpoint the tissue-specificity of causal genes (Figure 3), and candidate causal CAD genes were identified as exerting their effect in differing numbers of tissue/cell types which ranged from 1 up to 7 types. Notably, arterial wall tissues (aorta and MAM) yielded the greatest number of candidate causal CAD genes. For example, the IGA integrating aorta eQTLs with UKBB or CARDioGRAMplusC4D GWASs both yielded 49 candidate causal genes; while the IGA involving MAM eQTLs with UKBB or CARDioGRAMplusC4D GWASs yielded 42 and 41 candidate causal CAD genes, respectively (Figure 3). These findings indicate that the arterial wall is of major importance with respect to CAD pathogenesis.

In comparing the IGA results using GWAS data from UKBB versus CARDioGRAMplusC4D, there was reasonably strong overlap for most of the 9 tissue/cell types (Figure 4A). In addition, we found a high degree of concordance for $Z$ score results generated using MetaXcan alone for UKBB versus CARDioGRAMplusC4D GWAS data when integrated with STARNET.
eQTL data. Importantly, this concordance was not only in terms of the specific candidate causal genes identified, but also both the tissues in which they are likely to be causal and the directionality of their association with CAD (Figures 4B through 4D). In considering the number of candidate causal CAD genes across the IGAs performed using either UKBB or CARDioGRAMplusC4D with STARNET (129 and 121 genes, respectively), there were a total of 162 unique candidate causal CAD genes across both IGAs. These 162 candidate causal CAD genes were then ranked by 

P
value and the top 25 are presented in Table 1, with all 162 ranked genes presented in Table S7. These 162 candidate causal CAD genes were found to exert their effects across a mean of 1.9±1.4 tissue/cell types (mean±SD; Figure 5, Table S7).

Of the 163 independent CAD association peaks previously compiled by Erdmann et al,6 56 of these were identified in our IGA as being linked to causal CAD genes (Table S8). While the genes nominated by our IGA were in high agreement with this literature,6 we also identified novel candidate causal genes. For example, at a GWAS peak around rs2022938 the previously attributed gene was HDAC9.8 Our analysis clarified that rather than HDAC9, the adjacent gene TWIST1 is the likely causal CAD gene (Table S8). The reassignment of this GWAS peak from HDAC9 to TWIST1 as the likely causal candidate CAD gene is corroborated by another recent study by Nurnberg et al19 conducted in smooth muscle cells. Of importance, our IGA also pinpointed the tissue-specificity of the candidate causal genes (Figure 5, Table S7). Taking the same example, our IGA found that TWIST1 plays a causal role for CAD in aorta and MAM (Figure 5). Because the predominant cell type in aorta and MAM (ie, the arterial wall) is smooth muscle cells, this finding adds further corroborative evidence to the study by Nurnberg et al.19

Various potential pathways and aspects of CAD and atherosclerosis were represented by these 162 genes and the corresponding tissues in which they exert their effects. For example, CDKN2B (cyclin dependent kinase inhibitor 2B) residing in the strongest genetic locus for CAD, 9p21.3,20 was the top ranked candidate causal gene for CAD (Table 1). CDKN2B is known to have strong effects on vascular cells,21,22 which is consistent with the single tissue of effect for CDKN2B in this IGA being aorta (Figure 5). Other candidate causal CAD genes that involved only a single tissue included PDE5A in aorta, TNF in BLOOD, and CCDC97 in liver (Figure 5). Of the 31 genes that were associated with 2 tissue/cell types, 15 were associated with aorta and MAM (with both aorta and MAM being arterial wall) including PDGFD (platelet...
Figure 3. Summary of integrative genomics analysis (IGA) and MetaXcan results.

A, MetaXcan results based on UK Biobank (UKBB) genome-wide association studies (GWAS) data. The x axis shows different tissue/cell types, and combinations of these different tissue/cell types, from among the 9 tissue/cell types sampled in STARNET (Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task). The y axis shows the number of genes identified by MetaXcan for that combination of tissue/cell types. B, As per (A), showing MetaXcan results based on CARDIoGRAMplusC4D GWAS data. AOR indicates aorta; BLOOD, venous blood; FC, foam cells; LIV, liver; MAM, internal mammary artery; MP, macrophages; SF, subcutaneous fat; SKLM, skeletal muscle; and VAF, visceral abdominal fat.
derived growth factor D), TWIST1 (twist-related protein 1), and PHACTR1 (phosphatase and actin regulator 1), with PHACTR1 being the second top ranked candidate causal gene for CAD (Table 1). Three genes were associated with VAF and SF (both adipose tissue), including SCD (stearoyl-CoA desaturase) and IRS1 (insulin receptor substrate 1). Furthermore, 2 genes were associated with macrophages and foam cells (closely related inflammatory cell types), being SMAD3 (mothers against decapentaplegic homolog 3) and MIA3 (melanoma inhibitory activity family member 3; Figure 5).

**Validation of IGA Using an Alternate Transcriptomic Data Set**

As further validation, we substituted transcriptomic data from GTEx18 for the STARNET data set that was originally used. Although GTEx contained 48 tissues in its data sets, many of these tissues are unlikely to be related to CAD (eg, uterus, bladder, esophagus, tibial nerve). Therefore, we only considered the following GTEx tissues that have biologic plausibility for causing CAD: SF, VAF, aorta, liver, skeletal muscle, BLOOD, and coronary artery (COR—which was not obtained in STARNET). Note that while GTEx allowed us to include COR, and to also analyze SF, VAF, aorta, liver, skeletal muscle, and BLOOD that were all in STARNET, on the contrary GTEx does not have MAM, macrophages, or foam cells and therefore, these tissues/cell types were excluded from this GTEx validation analysis.

Interestingly, when GTEx was used rather than STARNET fewer causal genes were identified, with only 47 candidate causal CAD genes identified with UKBB and GTEx (Table S9) and 53 with CARDIoGRAMplusC4D and GTEx (Table S10). Despite there being less than half the number of candidate causal genes identified when GTEx was used rather than STARNET, many of the candidate causal genes identified using GTEx were also identified using STARNET (Table S11).

As stated, unlike STARNET, GTEx includes COR. Using UKBB and GTEx for the IGA, candidate causal CAD genes identified in COR were THOC5, MRAS, NBEAL1, and PHACTR1 (Table S9). As an alternative, using CARDIoGRAMplusC4D and GTEx, candidate causal CAD genes in COR were: SF3A3, FHL3, MRAS, NBEAL1, ADAMTS7, PHACTR1, and INPP5B (Table S10). Demonstrating the similarity of COR and aorta in their predisposition to atherosclerosis, the majority of these were also identified as candidate causal CAD genes using aorta in STARNET (Tables S5 and S6), with the only exceptions being ADAMTS7 and INPP5B.
Most Candidate Causal Genes Are Involved in CAD GRNs

To identify potential pathways and mechanisms of how these genes cause CAD, we queried the GRNs that have been inferred from the STARNET data sets.5,11,12,14,15 We focused on identifying GRNs where the tissue of potential causality from the IGA matched the tissue of effect for that gene in the GRN. On this basis, for the 162 candidate causal CAD genes identified in the IGA using STAR-NET (Figure 5) we found that 144 (144/162=88.9%) were represented in at least one GRN, in the same tissue (Figures 6 and 7, Table S12).

Candidate Causal CAD Genes as Key Drivers in CAD GRNs

We also explored which candidate causal CAD genes are key drivers of GRNs. From the 162 candidate causal CAD genes, there were 22 (22/162, 13.6%) that were key drivers in GRN(s) where the tissue of causality in the IGA matched the tissue of effect of that gene in the GRN (Table 2, Figure 7).

Table 1. Top 25 Prioritized Candidate Causal Genes for CAD Identified Using Our IGA Pipeline With Either UKBB1 With STARNET12 or CARDIoGRAMplusC4D2 With STARNET12

| Candidate causal CAD gene | Most significant P value | Tissue with most significant P value | GWAS used in IGA with most significant P value (UKBB or CaridoG) | Causal in that tissue in UKBB, CaridoG, or Both |
|---------------------------|--------------------------|-----------------------------------|---------------------------------------------------------------|-----------------------------------------------|
| CDKN2B                    | 2.16×10^{-47}            | AOR                               | UKBB                                                          | Both                                          |
| PHACTR1                   | 3.65×10^{-40}            | MAM                               | CardioG                                                       | Both                                          |
| TBC1D7                    | 2.40×10^{-38}            | MAM                               | CardioG                                                       | Both                                          |
| GFOG1                     | 2.84×10^{-34}            | MAM                               | CardioG                                                       | Both                                          |
| PSRC1                     | 3.40×10^{-24}            | BLOOD                             | UKBB                                                          | Both                                          |
| SORT1                     | 1.18×10^{-23}            | LIV                               | UKBB                                                          | Both                                          |
| CELSR2                    | 5.19×10^{-23}            | LIV                               | UKBB                                                          | Both                                          |
| MRPS6                     | 1.96×10^{-22}            | AOR                               | UKBB                                                          | Both                                          |
| SLC5A3                    | 1.96×10^{-22}            | AOR                               | UKBB                                                          | Both                                          |
| SARS                      | 2.42×10^{-20}            | LIV                               | UKBB                                                          | Both                                          |
| KCNE2                     | 8.19×10^{-20}            | AOR                               | UKBB                                                          | CardioG                                       |
| NBEAL1                    | 4.04×10^{-18}            | AOR                               | UKBB                                                          | Both                                          |
| ICA1L                     | 1.08×10^{-17}            | AOR                               | UKBB                                                          | Both                                          |
| CARF                      | 1.79×10^{-17}            | MP                                | UKBB                                                          | Both                                          |
| LIPA                      | 1.58×10^{-15}            | LIV                               | UKBB                                                          | Both                                          |
| GGCGX                     | 3.94×10^{-13}            | SF                                | UKBB                                                          | Both                                          |
| TWIST1                    | 3.97×10^{-13}            | AOR                               | UKBB                                                          | Both                                          |
| VAMP5                     | 1.11×10^{-12}            | MP                                | UKBB                                                          | Both                                          |
| VAMP8                     | 1.13×10^{-12}            | FC                                | UKBB                                                          | CardioG                                       |
| FES                       | 1.39×10^{-12}            | VAF                               | UKBB                                                          | Both                                          |
| MIA3                      | 2.56×10^{-12}            | FC                                | UKBB                                                          | Both                                          |
| KIAAA462                  | 5.86×10^{-12}            | MAM                               | UKBB                                                          | Both                                          |
| PDGFD                     | 6.28×10^{-12}            | MAM                               | UKBB                                                          | Both                                          |
| FURIN                     | 1.07×10^{-11}            | AOR                               | UKBB                                                          | Both                                          |
| MAT2A                     | 2.33×10^{-11}            | AOR                               | UKBB                                                          | Both                                          |

Candidate causal genes were prioritized based on the smallest P value for the class 1 analyses (MetaXcan or SMR), however, for all of these top 25 candidate causal genes the most significant P value was obtained with MetaXcan (as opposed to SMR). Full results for all 162 candidate causal genes are in Table S7. AOR indicates aorta; BLOOD, venous blood; CAD, coronary artery disease; FC, foam cells; IGA, integrative genomics analysis; LIV, liver; MAM, internal mammary artery; MP, macrophages; SF, subcutaneous fat; SMR, summary-based Mendelian randomization; STARNET, Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task; UKBB, UK Biobank; and VAF, visceral abdominal fat.
PHACTR1 Is a Top Causal Gene for CAD

CDKN2B and PHACTR1 were the top 2 candidate causal genes for CAD in this study (Table 1). While a great deal of research has been conducted on CDKN2B and the related 9p21.3 locus,20–22 much less is known about PHACTR1. Accordingly, we probed STARNEST and the GWASs explored here to gain additional insights on this gene. In STARNEST using FDR <5%, we identified 4 index eQTLs (the best associations for this gene per tissue) for PHACTR1 and 2 further independent but nonindex eQTLs by stepwise regression (Table 3). Among these, rs9349379 was an index eQTL for PHACTR1 in both MAM and aorta. Notably, the statistical significance of the index eQTLs at rs9349379 were many orders of magnitude stronger than other eQTLs for PHACTR1 in this analysis (Table 3). Apart from MAM and aorta, there were no other tissues at (FDR 5%). While there were 3 additional cis-eQTLs at rs9349379 for other genes, at FDR 5% these were of marginal significance.

As a whole, these results indicate that rs9349379 is likely to be the causal PHACTR1-associated single nucleotide polymorphism and that the CAD causal effects of rs9349379 and PHACTR1 arise in the arterial wall (ie, aorta, MAM and COR in our analyses).

PHACTR1 is known to have multiple isoforms. To understand which are potentially the most important for the arterial wall, we discuss the potential functional implications of these transcripts in the next section.
causing CAD, we queried STARNET for isoform-specific eQTLs of PHACTR1 at rs9349379 (thereby avoiding the need to correct for multiple comparisons). As shown in Table S13, we identified 15 isoform-specific eQTLs for PHACTR1 at rs9349379, with 13 of these being in aorta or MAM. Interestingly, these eQTLs coded for both protein and nonprotein coding PHACTR1 isoforms. However, the strongest eQTLs to emerge, and thus by inference the strongest causal candidate isoforms for CAD, were PHACTR1 isoforms 201, 206, and 207.

**DISCUSSION**

The pathobiology of CAD and atherosclerosis are profoundly complex, but until now there have been few insights as to which causal mechanisms are most important. This study directly addressed this concern and developed an IGA pipeline that provided a prioritized list of candidate causal CAD genes, and the tissues in which these genes exert their effect. This will enable a sharp refocusing of research efforts, both with respect to which genes are most critical for causing CAD and also where their effects are mediated.

Our IGA pipeline (Figure 1) integrated large eQTL and GWAS data sets. Several methods can be applied for this purpose, which belong to 2 broad classes. Class 1 includes TWAS, MetaXcan and SMR, while class 2 includes Coloc and eCAVIAR (only MetaXcan, SMR, and Coloc were used in this study). It has been reported that the results of these classes do not fully overlap, which was corroborated by our study. Accordingly, our methodology was conservative, requiring candidate causal genes to be identified both using Coloc and either MetaXcan or SMR. While this likely led to the exclusion of additional causal genes that did not meet these conservative criteria, as the first systematic CAD IGA it provided assurance that the candidate causal genes identified are valid and correct. Furthermore, when our IGA pipeline was applied to different GWAS data sets (UKBB versus CARDIoGRAMplusC4D) or different eQTL data sets (STARNET versus GTEx), the results were comparable. Presumably,
any differences in the candidate causal genes identified between these alternate data sets were related to differences between the subjects enrolled and their demographic features. However, another difference was that STARNET samples were from living subjects undergoing coronary artery bypass surgery and that after procurement these samples were immediately placed into solutions to stabilize RNA. Conversely, GTEx samples were obtained at autopsy, and additional factors such as end-of-life treatment modality, sequencing contamination, and other technical factors have been shown to influence gene expression in this data set.

As one of the main readouts of this IGA, we prioritized candidate causal CAD genes based on the smallest P value for the class 1 analyses (MetaXcan or SMR; Table 1 and Table S7). This is important to consider, because it means the prioritization was on the basis of the strengths of the correlations between the eQTL and GWAS results. While this gives assurance that the top ranked genes have very robust statistical associations to support their causal status, it does not imply that the top genes are those with the strongest effect on CAD. Ranking the strength of effect on CAD for the hundreds of genes identified by GWAS, across multiple different tissues, will be a major undertaking that might require added layers of data to be considered such as burden of CAD, the role of gene enhancer or promoter elements, and other aspects. At the present time, we are not aware that this has been attempted using GWAS and other large-scale data sets.

While we believe our study is the first systematic, large-scale IGA for CAD, it is important to acknowledge a recent study that undertook a more restricted analysis for the association of 51 loci with CAD based on evidence from experimental and in silico studies, but which also included an SMR analysis using GTEx. While the analytic strategy was very different from that applied here, a likely causal gene was identified for 36 of 51 loci, and several genes were validated as being potentially causal for CAD across that study and ours, including PHACTR1, FURIN, IL6R, LPL, LIPA, MRAS, KIAA1462 (also known as JCAD), GUCY1A3, SH2B3, and PDGFD.

It was reassuring in our study that CDKN2B was one of the top 2 candidate causal genes (Table 1). This is consistent with CDKN2B being among the closest coding genes to the 9p21.3 CAD risk locus and that the 9p21.3 locus influences CDKN2B expression. In turn, 9p21.3 is known to be a powerful common genetic risk factor for CAD. Our finding that CDKN2B is only potentially causal for CAD in aorta corroborates previous studies in mice and in humans whereby regulatory elements in coronary artery smooth muscle cells were linked to CDKN2B expression. These findings should guide research efforts to focus on the effects of this gene and the 9p21.3 CAD risk locus in the arterial wall, while other candidate genes at 9p21 make a significant contribution to the epidemiology of CAD.
Our results prioritized PHACTR1 as the other of the top 2 candidate causal CAD genes. As a CAD risk locus with largely unknown function, rs9349379, which resides in the third intron of the PHACTR1 gene, had already emerged as likely having a critical role in vascular pathobiology.1,2,11,30 Our results extend the knowledge-base regarding rs9349379 and PHACTR1, showing that PHACTR1 is a likely causal gene for CAD and that this causality is most likely to be mediated through the arterial wall. Furthermore, our study highlights the profound complexity of rs9349379 in terms of its regulation of the expression levels (ie, eQTLs) of at least 10 PHACTR1 isoforms, which include protein coding and noncoding isoforms (Table S13). Despite these complexities, it is clear given its ranking as among the top candidate causal CAD genes, that redoubled research efforts on PHACTR1 are justified and urgently needed.

Many other novel findings emerged from this analysis. For example, after PHACTR1 and CDKN2B, 2 of the next most significant candidate causal CAD genes were TBC1D7 (causal in aorta and MAM) and GFOID1 (causal in MAM; Table 1). Apart from the fact that they have been associated with CAD through GWAS;1,2 almost nothing is known about how these genes might be causal for CAD. Our study localized the tissue of likely causality to the arterial wall for both these genes. Furthermore, both genes are involved in GRNs; GFOID1 in STARNET GRN 82 and TBC1D7 in STARNET GRNs 167 and 217 (Table S12). As another novel finding, our study found that most candidate causal CAD genes were in CAD GRNs, but noncoding RNA ANRIL may still exert causal effects at the epigenetic or posttranscriptional levels.29

Candidate causal CAD genes in this table represent those identified in the IGA performed using STARNET and either UKBB or CARDioGRAMplusC4D, where the tissue of causality in the IGA is the same tissue where the gene is also a key driver of a GRN. Note that STARNET does not yet have curated GRNs for MP and FC. Therefore this table only considered AOR, MAM, LIV, BLOOD, VAF, SF and SKLM. AOR indicates aorta; BLOOD, venous blood; BMI, body mass index; CAD DGE, the enrichment of differential gene expression in the module between cases and controls; CRP, C-reactive protein; FC, foam cells; GRN, gene regulatory co-expression network; HBA1C, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; IGA, integrative genomics analysis; LDL-C, low-density lipoprotein cholesterol; LIV, liver; MAM, internal mammary artery; MP, macrophages; SF, subcutaneous fat; SKLM, skeletal muscle; STARNET, Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task; TG, triglycerides; UKBB, UK Biobank; VAF, visceral abdominal fat; and WHR, waist-hip ratio.
only a minority were key drivers (Table 2). The fact that only a minority of candidate causal genes were GRN key drivers is consistent with our understanding of how gene networks and their key drivers cause disease. A leading explanation is that hub nodes (governed by key drivers) tend to be essential for life and are evolutionarily conserved, and that disease genes do not typically encode hubs. Nonetheless, for the 22 candidate causal CAD genes that were found to be key drivers (Table 2), the mechanism of CAD causality appears to be at least partially evident via their key driver role in modulating the effects of those GRNs. For other candidate causal genes, it appears plausible that some participate in GRNs but in a nonkey driver role. While elucidating the precise mechanisms of effect of all causal CAD genes is beyond the scope of the present study, the many network associations of these candidate causal genes (Table 2, Table S12) is an important starting point for future research efforts.

There are certain limitations of this study. First, IGA methodologies for integrating GWAS and eQTL data continue to evolve, and with further improvements to these methodologies the causal gene list for CAD could be refined. Second, we used STARNET as our main transcriptomic data set, with GTEx as a validation data set. Because it collected samples from living individuals, STARNET does not include coronary artery samples, rather the arterial samples collected in STARNET were the atherosclerosis-prone aorta and pre/early atherosclerotic MAM. As CAD is characterized by atherosclerotic plaques in coronary arteries, atherosclerotic aortic tissue might not be the ideal arterial tissue to study CAD. However, since atherosclerosis is a systemic disease, aorta should reflect ongoing disease patterns in differing vascular beds. Furthermore, GTEx does not contain MAM, macrophages, or foam cells—therefore these tissues/cells could not be included in the validation analyses. In addition, both STARNET and GTEx used bulk (whole tissue) RNA sequencing and did not use state-of-the-art single cell RNA sequencing. Hopefully, future large-scale efforts to create CAD-relevant single cell transcriptomic data sets will bring even greater clarity to the causal genes and cell types for CAD and other diseases. As another possible limitation, CAD was defined differently across STARNET and the GWAS data sets. STARNET applied a rigorous definition using coronary angiography, and CAD cases were those with severe CAD requiring coronary artery bypass graft surgery.1,12 In contrast, for the UKBB data set a soft but inclusive CAD definition was used that incorporated self-reported angina or other evidence of chronic coronary disease, but also including more stringent definitions such as myocardial infarction and revascularization. Similarly, the CARDIoGRAMplusC4D GWAS data set also used an inclusive definition of CAD (see Table S1). The impact of these differing definitions on this study is unknown, although, the fact that STARNET applied a stringent CAD definition provides reassurance of the validity of our findings.

In conclusion, we developed an informatics pipeline and thus conducted a large-scale IGA of GWAS and transcriptomic data using advanced computational methods to generate a refined list of candidate causal genes for CAD, which also localizes the tissue of causal effect. These results should serve as an important resource, facilitating the focusing of research efforts toward the most powerful causal CAD genes, and to the tissues and mechanisms that are most critical for that causal effect.

**ARTICLE INFORMATION**

Received September 7, 2020; accepted December 9, 2021.

**Affiliations**

Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, NY (KH, H.C., L.M., L.A., J.L.M.B.). Sema4, Stamford, CT (K.H.). Department of Cardiac Surgery and The Heart Clinic, Tartu University Hospital, Estonia (R.E., K.S., A.R.). Cardiovascular Research Institute, Icahn School of Medicine at Mount Sinai, NY (M.L., Vd., N.C., J.C.K.). Department of Cardiology, Deutsches Herzzentrum München, Technische Universität München, Munich, Germany (L.L., H.S.). Center for Doctoral Studies in Informatics and its Applications, Department of Informatics, Technische Universität München, Munich, Germany (L.L.). Deutsches Zentrum für Herz- und Kreislaufforschung (DZHK), Munich Heart Alliance, Germany (L.L., H.S.). Now with New York University Cardiovascular Research Center, Department of Medicine, Leon H. Charney Division of Cardiology, New York University Grossman School of Medicine, New York University Langone Health, NY (L.A.). Cancer Research UK Cam-

**Table 3. Genome-Wide Significant eQTLs Involving PHACTR1**

| Tissue                | Locus     | Location on chromosome 6 | \( P \) value          | \( \beta \) value |
|-----------------------|-----------|---------------------------|------------------------|------------------|
| Lead eQTLs            |           |                           |                        |                  |
| Aorta                 | rs9349379 | 12903725                  | 9.37\times10^{-17}     | 0.49             |
| Internal mammary artery | rs9349379 | 12903725                  | 1.95\times10^{-55}     | 0.86             |
| Blood                 | rs413120  | 13280409                  | 1.8\times10^{-6}       | 0.41             |
| Subcutaneous adipose  | rs386406198 | 13060791             | 3.63\times10^{-4}      | 0.45             |
| Nonlead eQTLs         |           |                           |                        |                  |
| Aorta                 | rs6458568 | 12961440                  | 1.33\times10^{-4}      |                  |
| Subcutaneous adipose  | rs20499   | 13294772                  | 5.09\times10^{-5}      |                  |

At genome-wide significance, four lead eQTLs for PHACTR1 were identified (the best associations for this gene per tissue), with stepwise regression revealing 2 additional nonlead eQTLs. eQTL indicates expression quantitative trait loci.
References

1. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell 2016;167:1145–1149. doi: 10.1016/j.cell.2016.10.026

2. GTEX Consortium. Human genomics. The Genotype-Tissue Expression (GTEX) pilot analysis: multifaceted gene regulation in humans. Science 2015;348:648–660. doi: 10.1126/science.1262110

3. Missmohl N, Tisken CA, Fraulein C, et al. Integrative prioritization of causal genes for coronary artery disease based on cumulative multiomic network model links lipid metabolism to glucose regulation in coronary artery disease. Nat Commun 2021;12:547. doi: 10.1038/s41467-020-15821-9

4. Zhang J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, Evans DM, Smith GD. Recent developments in Mendelian randomization studies. Curr Epidemiol Rep 2017;4:330–345. doi: 10.1007/s40471-017-0128-6

5. Friedman O, Ermel R, Cohan A, Akers NK, Dinardo A, Talukdar HA, Foroughi-Asl H, Giambartolomei C, Fullard JF, Sukhavasi K, et al. Cardiometabolic risk loci share downstream cis- and trans–gene regulation across tissues and diseases. Science 2016;353:827–830. doi: 10.1126/science.aad6970

6. Zeng L, Talukdar HA, Kovacek S, Giannarelli C, Ivert T, Gan LM, Ruusalepp A, Schadt EE, Kovacic JC, Lusis AJ, et al. Contribution of gene regulatory networks to heritability of coronary artery disease. J Am Coll Cardiol 2019;73:2946–2957. doi: 10.1016/j.jacc.2019.03.0520

7. Michels KC, Nomura-Kitabayashi A, Lecce L, Frangostin SK, Oikev S, Xu Y, Santini MF, D’Escamard V, Lee JL, Fuster V, et al. CD90 identifies adventitial mesenchymal progenitor cells in adult human medium- and large-sized arteries. Stem Cell Reports 2018;11:242–257. doi: 10.1016/j.stem.2018.06.001

8. Hartman RJG, Ovsianny K, Ma L, Kovacek S, Hao K, Slenders L, Civelek M, Mokry M, Kovacic JC, Pasterkamp G, et al. Sex-stratified gene regulatory networks reveal female key driver genes of atherosclerosis involved in smooth muscle cell phenotype switching. Circulation 2021;143:719–726. doi: 10.1161/CIRCULATIONAHA.120.051231

9. Choy AT, Barrington WT, Jordan DM, Beckmann ND, Argmann CA, Houten SM, Charney AW, Ermel R, Sukhavasi K, Franzen O, et al. An integrative multi-omic network model links lipid metabolism to glucose regulation in coronary artery disease. Nat Commun 2021;12:547. doi: 10.1038/s41467-020-20570-8

10. Nguyen V, Shadrina AS, Shashkova TI, Torgasheva AA, Sharapov SZ, Klarić L, Nieuwenhuis TO, Yang SY, Verma RX, Pillalamarri V, Arking DE, Denissenkho A, et al. Genetic prioritization of causal genes for coronary artery disease based on cumulative multiomic network model links lipid metabolism to glucose regulation in coronary artery disease. Nat Commun 2021;12:547. doi: 10.1038/s41467-020-20570-8

11. Galis GS, Asahara T, Gage F, et al. The international Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. Cell 2016;167:1145–1149. doi: 10.1016/j.cell.2016.11.007

12. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martin D, Watt S, Yan Y, Kudnu K, Ecker S, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell 2016;167:1190–1194.e24. doi: 10.1016/j.cell.2016.10.026

13. Gränsbo K, Almgren P, Spigberg M, Smith JG, Engström G, Hedblad B, Melander O. Chromosome 9p21 genetic variation explains 13% of cardiovascular disease incidence but does not improve risk prediction. J Intern Med 2013;273:234–240. doi: 10.1111/joim.12063

14. Visel A, Zhu Y, May G, Altal F, Gong E, Attanasio C, Blow MJ, Cohen JC, Rudolph EM, Pennachio LA. Target deletion of the 9p21 non-coding coronary artery disease risk interval in mice. Nature 2010;464:409–412. doi: 10.1038/nature08801

15. Miller CL, Pijanic M, Wang T, Nguyen T, Cohan A, Lee JD, Pericir L, Hedin U, Kundu RK, Majmudar D, et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. Nat Commun 2016;7:12092. doi: 10.1038/ncomms12092

16. Nishiyama S, Deyssenroth MA, Dinardo A, Afshari H, Zhang H, Zang T, Lambertini L, Ruusalepp A, Kovacic JC, Bjorgjekn JLM, Marsil J, et al. Genetic regulation of the placental transcriptome underlies birth weight and risk of childhood obesity. PLoS Genet 2018;14:e1007799. doi: 10.1371/journal.pgen.1007799

17. McCall MN, Ilii PB, Halushka MK. Complex sources of variation in tissue expression data: analysis of the GTEX lung transcriptome. Am J Hum Genet 2016;99:624–635. doi: 10.1016/j.ajhg.2016.07.007

18. Nieuwenhuis TO, Yang SY, Verma RX, Pillalamarri V, Arking DE, Rosenburg AG, McCall MN, Halushka MK. Consistent RNA sequencing contamination in GTEX and other data sets. Nat Commun 2020;11:1933. doi: 10.1038/s41467-020-15821-9

19. Boix CA, James BT, Park YP, Meuleman W, Kellis M. Regulatory genetic circuitry of human disease loci by integrative epigenomics. Nature 2018;558:300–307. doi: 10.1038/s41586-020-0314a

20. Chadrina AS, Shashkova T, Torgasheva AA, Sharapov SZ, Klarić L, Pakhomov ED, Alexeev DG, Wilson JF, Tsepilov YA, Joshi PK, et al. Prioritization of causal genes for coronary artery disease based on cumulative evidence from experimental and in silico studies. Sci Rep 2020;10:10486. doi: 10.1038/s41598-020-67001-w

Supplemental Materials

Tables S1–S3

References25–43
28. Lo Sardo V, Chubukov P, Ferguson W, Kumar A, Teng EL, Duran M, Zang L, Cost G, Engler AJ, Urnov F, et al. Unveiling the role of the most impactful cardiovascular risk locus through haplotype editing. Cell 2018;175:1796–1810.e20. doi: 10.1016/j.cell.2018.11.014
29. Holdt LM, Teupser D. Long noncoding RNA ANRIL: Lnc-ing genetic variation at the chromosome 9p21 locus to molecular mechanisms of atherosclerosis. Front Cardiovasc Med. 2016;5:145. doi: 10.3389/fcmv.2018.00145
30. Gupta RM, Hadyaya J, Trehan A, Zekavat SM, Roselli C, Klarin D, Emdin CA, Hilering CRE, Bianchi V, Mueller C, et al. A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. Cell 2017;170:522–533.e15. doi: 10.1016/j.cell.2017.06.049
31. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12:56–68. doi: 10.1038/nrg2918
32. Moore KJ, Koplev S, Fisher EA, Tabas I, Björkegren JLM, Doran AC, Kovacic JC. Macrophage trafficking, inflammatory resolution, and genomics in atherosclerosis: JACC macrophage in CVD series (Part 2). J Am Coll Cardiol 2018;72:2181–2197. doi: 10.1016/j.jacc.2018.08.2147
33. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21. doi: 10.1093/bioinformatics/bts635
34. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28:1353–1358. doi: 10.1093/bioinformatics/bts163
35. Talukdar HA, Foroughi Asl H, Jain RK, Ermel R, Ruusalepp A, Franzen O, Kidd BA, Readhead B, Giannarelli C, Kovacic JC, et al. Cross-tissue regulatory gene networks in coronary artery disease. Cell Syst. 2016;2:196–208. doi: 10.1016/j.cels.2016.02.002
36. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. doi: 10.1186/1471-2105-9-559
37. Huynh-Thu VA, Irthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS One. 2010;5:e12776. doi: 10.1371/journal.pone.0012776
38. Arneson D, Bhattacharya A, Shu L, Mäkinen VP, Yang X. Mergeomics: a web server for identifying pathological pathways, networks, and key regulators via multidimensional data integration. BMC Genomics. 2016;17:722. doi: 10.1186/s12866-016-3095-8
39. Shao Q, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–2504. doi: 10.1101/gr.1239303
40. Finucane HK, Bullik-Sullivan B, Gusev A, Tynka G, Reshef Y, Loh PR, Aittila V, Xu H, Zang C, Farh K, 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
41. Hormozdiari F, Gazal S, van de Geijn B, Finucane HK, Ju CJ, Loh PR, Schoech A, Reshef Y, Liu X, O’Connor L, et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nat Genet. 2018;50:1041–1047. doi: 10.1038/s41588-018-0148-2
42. Ehret GB, Ferreira T, Chasman DI, Jackson AU, Schmidt EM, Johnson T, Thorleifsson G, Luan J, Donnelly LA, Kanoni S, et al.; CHARGE-EchoGen consortium; CHARGE-HF consortium; Wellcome Trust Case Control Consortium. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat Genet. 2016;48:1171–1184. doi: 10.1038/ng.3667
43. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: an R package for rapidly calculating linkage disequilibrium statistics in diverse populations. Front Genet. 2020;11:157. doi: 10.3389/fgene.2020.00157
