THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease

OPEN ACCESS

Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease

Graphical Abstract

Highlights

- We reconstruct regulatory gene networks across seven vascular and metabolic tissues

- Integrative analysis using GWASs reveals 30 networks causally related to CAD

- 12 CAD-causal networks are indicated to be evolutionarily conserved from mouse

- An atherosclerotic arterial wall RNA-processing network affects foam cell formation
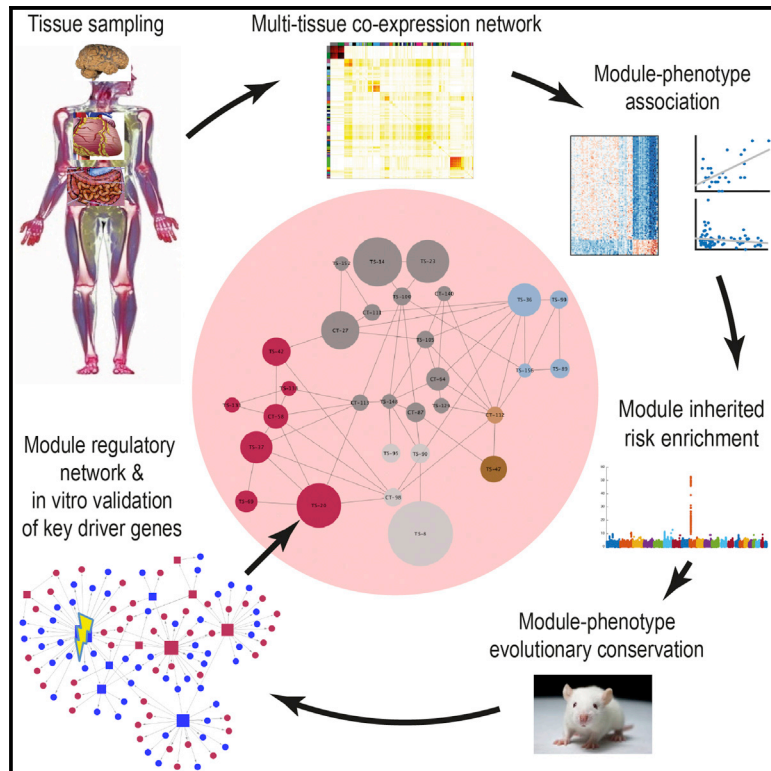
Authors

Husain A. Talukdar,
Hassan Foroughi Asl, Rajeev K. Jain, ...,
Josefin Skogsberg, Tom Michoel,
Johan L.M. Björkegren

Correspondence

johan.bjorkegren@mssm.edu

In Brief

Well-established atherosclerosis risk factors and pathways are shown to operate through regulatory gene networks, active both within and across vascular and metabolic tissues, to cause coronary artery disease (CAD). Within these CAD-causal networks, the hierarchical order and connectivity patterns of both established and new genes in CAD, including so-called key disease driver genes, advance not only our global understanding of the molecular landscape in CAD but also reveal new candidate genes that may serve as suitable drug targets.

CellPress

Article

Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease

Husain A. Talukdar,[1] Hassan Foroughi Asl,[1] Rajeev K. Jain,[2] Raili Ermel,[2,3] Arno Ruusalepp,[2,3,4] Oscar Franzén,[4,5] Brian A. Kidd,[5] Ben Readhead,[5] Chiara Giannarelli,[5,6] Jason C. Kovacic,[6] Torbjörn Ivert,[7,8] Joel T. Dudley,[5] Mete Civelek,[9] Aldons J. Lusis,[9] Eric E. Schadt,[4,5] Josefin Skogsberg,[1] Tom Michoel,[4,10,12] and Johan L.M. Björkegren[1,2,4,5,11,12,*]

[1]Cardiovascular Genomics Group, Division of Vascular Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm, Sweden
[2]Department of Physiology, Institute of Biomedicine and Translation Medicine, University of Tartu, 51014 Tartu, Estonia
[3]Department of Cardiac Surgery, Tartu University Hospital, 51014 Tartu, Estonia
[4]Clinical Gene Networks AB, 114 44 Stockholm, Sweden
[5]Department of Genetics & Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[6]The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[7]Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 77 Stockholm, Sweden
[8]Department of Thoracic Surgery, Karolinska University Hospital, 171 76 Stockholm, Sweden
[9]Departments of Medicine, Cardiology, Human Genetics, Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles, Los Angeles, CA 90024, USA
[10]Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Edinburgh EH25 9RG, UK
[11]Cardiovascular Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[12]Co-senior author
*Correspondence: johan.bjorkegren@mssm.edu
http://dx.doi.org/10.1016/j.cels.2016.02.002

SUMMARY

Inferring molecular networks can reveal how genetic perturbations interact with environmental factors to cause common complex diseases. We analyzed genetic and gene expression data from seven tissues relevant to coronary artery disease (CAD) and identified regulatory gene networks (RGNs) and their key drivers. By integrating data from genome-wide association studies, we identified 30 CAD-causal RGNs interconnected in vascular and metabolic tissues, and we validated them with corresponding data from the Hybrid Mouse Diversity Panel. As proof of concept, by targeting the key drivers *AIP*, *DRAP1*, *POLR2I*, and *PQBP1* in a cross-species-validated, arterial-wall RGN involving RNA-processing genes, we re-identified this RGN in THP-1 foam cells and independent data from CAD macrophages and carotid lesions. This characterization of the molecular landscape in CAD will help better define the regulation of CAD candidate genes identified by genome-wide association studies and is a first step toward achieving the goals of precision medicine.

INTRODUCTION

Coronary artery disease (CAD) is a heritable complex disease caused by the interactions of multiple genetic and environmental risk factors that change the molecular landscape of vascular and metabolic tissues to accelerate atherosclerosis. Despite lifestyle improvements and the successful targeting of CAD risk factors, such as hypercholesterolemia (Samani and de Bono, 1996) and hypertension (Bangalore et al., 2012), CAD still accounts for the majority of cardiovascular diseases. In fact, the clinical manifestations of CAD and atherosclerosis—myocardial infarction (MI) and stroke—are responsible for nearly 50% of all deaths globally (Mensah et al., 2014). New research strategies are urgently needed to battle CAD. One promising strategy is systems genetics (Barabási et al., 2011; Björkegren et al., 2015; Civelek and Lusis, 2014; Schadt, 2009; Schadt and Björkegren, 2012), which will help achieve a global understanding of how regulatory gene networks (RGNs) act within and across tissues to cause CAD. Such knowledge is central to tailor therapies to the specific molecular pathology of the individual patient (Collins and Varmus, 2015).

An important part of systems genetics is genome-wide association studies (GWASs), the predominant approach to genetic analysis of complex diseases for the last decade, which have led to the discovery of more than 150 genetic risk loci for CAD alone (Deloukas et al., 2013; Peden and Farrall, 2011;). However, these information-rich datasets have been analyzed only from the perspective of single DNA variants and, therefore, are a largely untapped resource to further elucidate the genetic basis of complex diseases. We and others (Barabási et al., 2011; Björkegren et al., 2015; Civelek and Lusis, 2014; Schadt and Björkegren, 2012) propose to use systems genetics to integrate the analyses of GWASs with functional genomic datasets where the combined effects of many, sometimes subtle, genetic and environmental influences are captured within molecular networks. In this study, we applied a systems genetics pipeline (Figures 1 and S1), including integrative multi-tissue, GWAS, and cross-species analyses to robustly identify RGNs in CAD. In
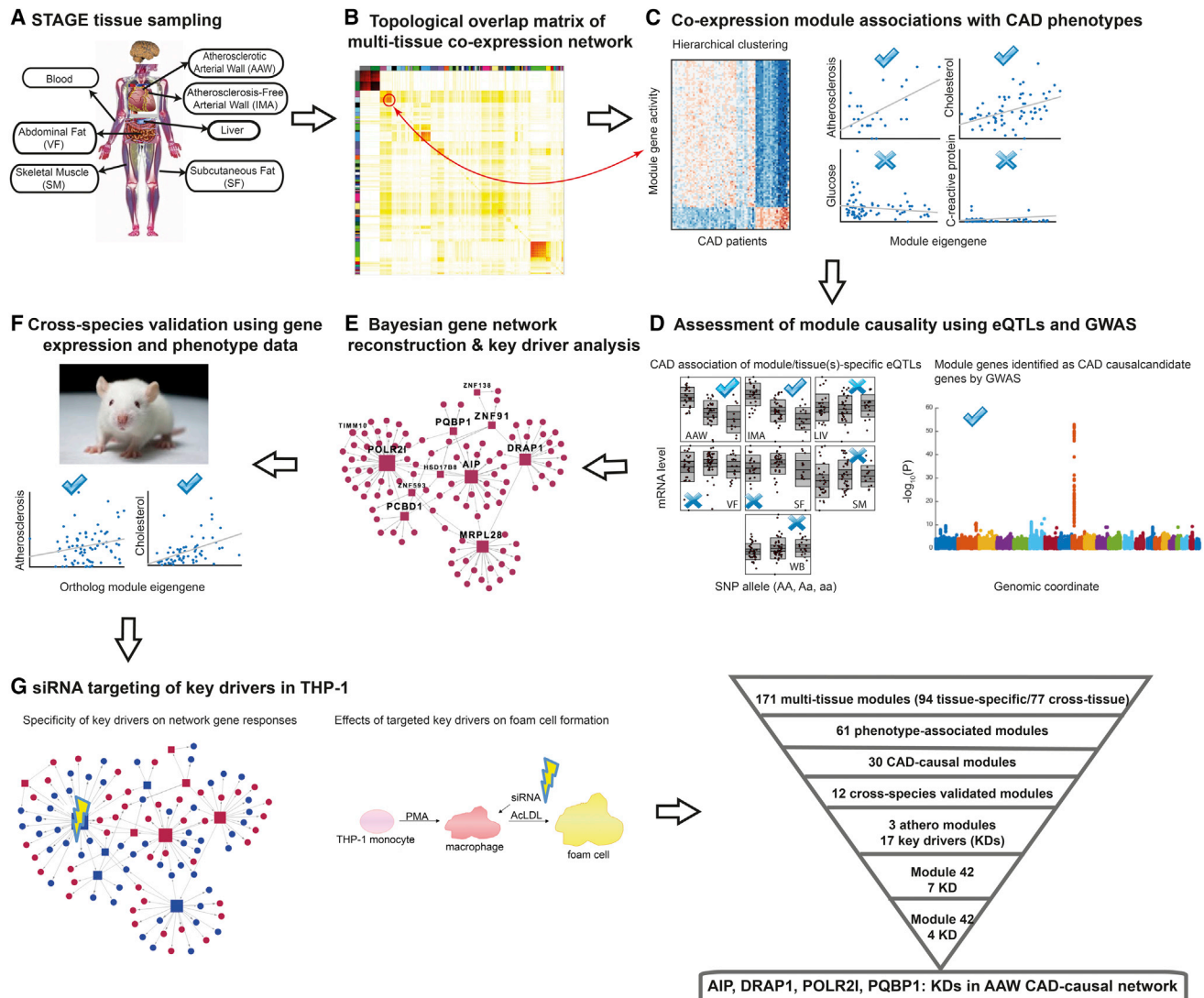
Figure 1. Schematic Flow of Analytic Steps

(A) STAGE tissue sampling. The STAGE study was undertaken at the Karolinska University Hospital, Stockholm, Sweden. Patients eligible for coronary artery bypass grafting were included, and seven metabolic and vascular tissues were biopsied during surgery. RNA samples were screened with Affymetrix Gene Chips rendering expression values for 19,610 genes. DNA was screened for 909,622 SNPs (Experimental Procedures).

(B) Weighted gene co-expression network analysis (WGCNA) to generate tissue-specific and cross-tissue co-expression modules. Gene expression values across all seven tissues were considered together to identify tissue-specific and cross-tissue co-expression modules. In the illustrated topological overlap matrix (TOM, also in Figure 2A), rows and columns are genes expressed in specific tissues (i.e., the same gene may occur up to seven times), and the seven tissues are intermixed. The diagonal indicates modules of highly co-expressed genes (Experimental Procedures).

(C) Module association with CAD phenotypes. To define modules of possible relevance for CAD, we sought both linear and nonlinear relationships between the activity of the identified co-expression modules and four main CAD phenotypes as follows: (1) the extent of coronary atherosclerosis as assessed in preoperative angiograms (atherosclerosis modules), (2) plasma measures of cholesterol (i.e., total, VLDL, LDL, or HDL cholesterol levels; cholesterol modules), (3) glucose metabolism (i.e., fasting glucose, HbA1c, insulin, or pro-insulin; glucose modules), and (4) high-sensitivity C-reactive protein (CRP modules). Heatmap (left) shows an example of how genes in a second-step clustering segregate the patients into two groups that differ phenotypically. Shown are examples of regression plots (right) between individual eigengene values (x axis) and individual phenotypic values (y axis).

(D) Assessment of module causality with STAGE eQTLs and GWASs. Causal co-expression modules were sought by (1) assessing the associations of module eQTLs (or of SNPs within ±500 kb of transcription start or end site if less than ten eQTLs) with CAD using CARDIoGRAM data and (2) the module contents for CAD candidate genes identified by GWASs. The eQTLs were sought to match the tissue-specific gene expression (left); GWA genes are those mapped to regions of genome-wide significant loci as exemplified in the Manhattan plot (right).

(E) Bayesian network reconstruction of RGNs and key driver analysis. Co-expression modules indicate correlations between genes without directions. In contrast, inference of RGNs considers the causal roles of eQTLs, transcription factors, and CAD GWA candidate genes to construct directed gene-gene connections. As a consequence, the hierarchical order among network genes can be determined, and the genes most essential for network regulation can be identified in the form of key drivers.

(legend continued on next page)

sum, we identified 30 CAD-causal RGNs harboring 59 CAD-related GWA candidate genes (Deloukas et al., 2013), whereof 26 RGNs were validated in corresponding gene expression and phenotypic data from the Hybrid Mouse Diversity Panel (HMDP) (Bennett et al., 2010). As proof of concept, key drivers in CAD-causal RGNs active in both the human and mouse atherosclerotic arterial wall (AAW) were further evaluated in a THP-1 foam cell model (Figure 1G) and in independent data from primary CAD macrophages and carotid lesions.

RESULTS

Inference of Co-expression Network Modules
In the first step of our analysis (Figure 1), we used weighted gene co-expression network analysis (WGCNA) (Zhang and Horvath, 2005) to infer functionally related genes in the form of co-expression modules from 612 newly generated gene expression profiles from seven tissues—AAW and non-AAW (internal mammary artery [IMA]), liver, skeletal muscle (SM), visceral fat (VF), subcutaneous fat (SF), and whole blood from the late-stage CAD patients of the Stockholm Atherosclerosis Gene Expression (STAGE) study (Hägg et al., 2009; Table S1; Figures 1A, 1B, and S1). We identified 94 tissue-specific and 77 cross-tissue modules (Figure 2A; Tables S2 and S3). The relevance of these 171 modules for CAD was reflected in the top ten biological processes and molecular functions according to gene ontology (Table S3; Supplemental Experimental Procedures). Moreover, we found 2,467 genes previously related to CAD/atherosclerosis (33%, $p = 1.5 \times 10^{-24}$; Supplemental Experimental Procedures) and 147 genes of 309 candidate genes proposed for risk loci identified by GWASs (Table S4; GWA genes) for CAD (n = 33 of 53 GWA genes) (Deloukas et al., 2013), plasma lipids (n = 87 of 183), and measures of plasma glucose metabolism (n = 22 of 43), including type 2 diabetes (n = 45 of 98) (Welter et al., 2014).

To test the robustness of cross-tissue modules against the contribution of gene expression in individual tissues, we removed gene expression data from one tissue and re-clustered the data. The 171 modules could largely be retrieved, except when data from dominant tissues in cross-tissue modules were removed (Supplemental Experimental Procedures).

Identification of Modules Associated with Main CAD Phenotypes
In the third step of the analysis pipeline (Figures 1C and S1), we used regression analysis and second-step clustering to identify linear and non-linear associations between module gene expression and CAD phenotypes (Supplemental Experimental Procedures). We found 61 modules that were associated with at least one of the following four major CAD phenotypes: (1) extent of coronary atherosclerosis as measured in preoperative angiograms (Hägg et al., 2009) (resulting in 14 atherosclerosis modules); (2) plasma levels of total cholesterol (TC), very-low-density lipoprotein (VLDL), low-density lipoprotein (LDL), or high-density lipoprotein (HDL) (29 cholesterol modules); (3) plasma glucose (fasting), HbA1c, insulin, or pro-insulin levels (14 glucose modules); and (4) the inflammatory marker C-reactive protein (CRP) (15 CRP modules) (Figure 2B). Ten modules were associated with more than one of the four phenotypes (Figure 2B; Table S2). Manual analysis of the atherosclerosis-, cholesterol-, glucose-, and CRP-associated modules suggested that, depending on phenotype association, the modules were more likely to be tissue specific (atherosclerosis, cholesterol, and CRP modules) or cross-tissue (glucose modules). Moreover, some modules were associated with more than one of the four major CAD phenotypes (Supplemental Experimental Procedures).

Inferring Module Causality by Using GWAS of CAD
The associations between the 61 co-expression modules and the four CAD phenotypes do not reveal whether the module activity alters the phenotypic values (causal) or is caused by changes in phenotypic values (reactive). Therefore, in the fourth step of our analysis (Figures 1D and S1), we assessed if the phenotype-associated modules were CAD causal or not. As in previous studies (Emilsson et al., 2008; Foroughi Asl et al., 2015; Schadt and Björkegren, 2012), we defined causal modules as those containing either of the following: (1) cis-acting expression quantitative trait loci (cis-eQTLs) (or, if a module had <10 eQTLs, cis-located SNPs) enriched in association with CAD according to CAD GWAS (Deloukas et al., 2013; Figure 2B, column 2); or (2) one or more CAD candidate genes (Brænne et al., 2015) already identified for CAD GWAS risk loci (Figure 2B, column 3; Table S4).

Using these definitions, we identified 30 CAD-causal modules (Figure 2B): eight atherosclerosis modules (of 14), ten cholesterol modules (of 29), five glucose modules (of 14), four CRP modules (of 15), and three additional CAD-causal modules that were associated with more than one of the four CAD phenotypes (Figure 2B). By integrating GWASs also for CAD risk factors (Welter et al., 2014), we found that the 30 CAD-causal modules contained 59 unique GWA genes for CAD (n = 19), plasma cholesterol (n = 39), and glucose metabolism (n = 14, including type 2 diabetes GWA genes) (Table S4) and 42% CAD/atherosclerosis genes ($p = 1.2 \times 10^{-5}$). Of the CAD-causal modules, 18 were also causal for plasma cholesterol levels (n = 15) or glucose metabolism (n = 12), based on their eQTL risk enrichment or the presence of GWA genes (Table S2).

Inference of RGNs and Key Drivers from the Co-expression Modules
In the next step of the analysis (Figures 1E and S1), to define the directions of edges between genes in the 30 CAD-causal

---

(F) Cross-species validation with the HMDP. Many metabolic and vascular processes are believed to be conserved between mice and humans. To validate the human RGN modules, we related gene expression in corresponding mouse tissues to corresponding mouse phenotypes (Experimental Procedures).

(G) The siRNA targeting of key drivers in THP-1 macrophages and foam cells. As proof of concept, we tested key drivers of the cross-species-validated atherosclerosis RGN modules in human THP-1 monocytes differentiated into THP-1 macrophages and foam cells. First, to test key driver specificity in activating the network module, we used custom microarrays to assess the response in network gene activity in THP-1 macrophages (left). Then we determined whether siRNA targeting of the key drivers also affected THP-1 foam cell formation (right). (Bottom right) The same analytical flow is presented in a triangle, ranging from WGCNA to THP-1 foam cells.
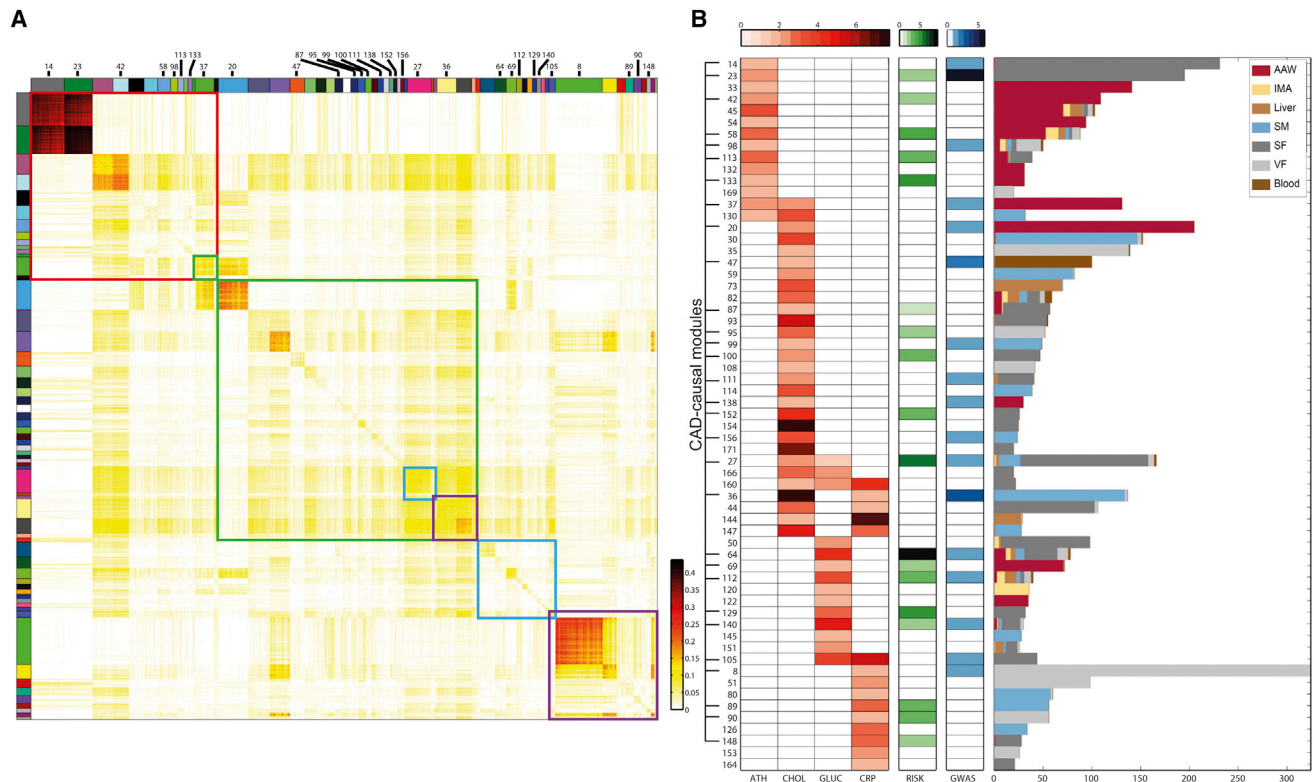
Figure 2. Co-expression Modules Associated with CAD Phenotypes

(A) TOM of gene activity sorted according to co-expression within and across tissues forming co-expression modules. The TOM shows the 20% most strongly co-expressed genes within and across tissues in 61 CAD-associated modules. Color-coded bars on the y axis and x axis indicate individual modules. The red rectangle encompasses modules related to coronary atherosclerosis; the two green rectangles encompass the modules related to measures of plasma cholesterol; the two blue rectangles encompass the modules related to measures of plasma glucose metabolism; and the two purple rectangles encompass the modules related to plasma CRP. Overlapping rectangles reflect modules associated with more than one of the four main phenotypes. The yellow-red color code (bottom right) indicates the strength of TOM associations (i.e., a weighted combination of direct co-expression and shared co-expression to other genes in the network).

(B) The sixty-one modules (module ID/row) were associated with the extent of coronary atherosclerosis assessed in preoperative angiograms (first column, ATH); plasma levels of total, VLDL, LDL, and HDL cholesterol and triglycerides (second column, CHOL); fasting glucose, HbA1c, insulin, or pro-insulin levels (third column, GLUC); and CRP level (fourth column, CRP). The module-phenotype association was assessed by second-step clustering of the STAGE patients, based on expression of module genes, and by correlating eigengene module values for each STAGE patient with phenotypic values; red colors indicate strength of associations ($-\log_{10}(P)$) (Experimental Procedures). The second and third columns indicate enrichment in disease association (risk) determined from the CARDIoGRAM dataset of module eQTLs (RISK, from 1- to 9-fold, green) and the presence of established GWA genes for CAD (GWAS, from 1–6 GWA genes, blue), respectively. (Right) The lengths of the horizontal bars indicate module size (i.e., gene numbers); the colors indicate the tissues where the module is active.

modules, we inferred RGNs including key drivers by Bayesian network modeling, using transcription factors, eQTL genes, and GWA genes as prior candidate regulators (Figures 3 and S2–S5; Tables S5 and S6; Supplemental Experimental Procedures).

Of the RGNs inferred from the 30 CAD-causal modules, eight associated with the extent of atherosclerosis (i.e., atherosclerosis modules) are shown in Figure 3. A nearly SF-specific RGN related to steroid and lipid metabolism (n = 231 genes, including one liver and one AAW gene) contained ten GWA genes identified for CAD, cholesterol, or glucose traits ($APOA1^{CAD,LDL,HDL,TC}$, $ANGPTL3^{LDL,TC}$, $HPR^{LDL,TC}$, $LIPC^{HDL,TC}$, $LIPG^{HDL,TC}$, $NPC1L1^{LDL,TC}$, $PPP1R3B^{LDL,HDL,TC,insulin,glucose}$, $PROX1^{glucose}$, $SLC2A2^{glucose}$, and $TMEM195^{glucose}$). Of these 10 GWA genes, only $APOA1$ was a key driver; the other key drivers were $CYP4F2$, $F11$, $KLKB1$, KMO, $MYCL1$, and $ZGPAT$. This RGN also was highly enriched in association with fasting glucose (8.0-fold, p < 3.3 × 10^{-272}) and TC (8.7-

fold, p < 1.0 × 10^{-300}) (Figure 3A). Another truly SF-specific RGN, related to metabolic processes and oxidoreductase activity (n = 195 genes), contained as many as 15 GWA genes ($ABCG5^{LDL,CAD,TC}$, $ABCG8^{LDL,CAD,TC}$, $ANXA9^{LDL}$, $APOA5^{LDL,HDL,CAD,TC-TG}$, $APOC4^{LDL}$, $CPS1^{HDL}$, $CYP7A1^{LDL,TC}$, $FOXA2^{glucose}$, $GCKR^{TC,glucose,insulin}$, $HNF1A^{TC,LDL}$, $HNF1B^{T2DM}$, $LPA^{LDL,CAD,TC}$, $NR0B2^{LDL,HDL}$, PLG^{CAD}, and $PCSK9^{LDL,CAD,TC}$). Again, besides PLG, the key drivers of this network were not these GWA genes; they were $FOXA2$, $NR1I2$, $ONECUT2$, and $UROC1$ (Figure 3B).

Three CAD-causal atherosclerosis-related RGNs (Figures 3C, 3D, and 3H) were AAW specific. The first, an RGN (n = 131 genes) enriched in immune system processes, contained the macrophage lipid and inflammatory regulators and GWA genes $APOE^{CAD}$ and $NR1H3^{HDL}$ as its key drivers, together with $CD80$, $E2F2$, $HCLS1$, and $PDE8B$ (Figure 3C). The second, an RGN (n = 109 genes) also linked to HDL cholesterol levels,

contained RNA-processing genes highly enriched in CAD association (2.1-fold, p < 1.0 × 10^{-300}) with seven key drivers (*AIP*, *DRAP1*, *MRPL28*, *PCBD1*, *POLR2I*, *PQBP1*, and *ZNF91*) and one GWA gene (*ARF5*^{T2DM}) (Figure 3D). The third AAW-specific RGN was small (n = 31 genes), even further enriched in CAD association (4.6-fold, p < 1.0 × 10^{-300}), and mainly contained genes involved in inflammatory responses and the GWA gene *ADRA2A*^{glucose} (Figure 3H).

The remaining three CAD-causal, atherosclerosis-related RGNs contained genes from multiple tissues (Figures 3E–3G). The first was dominated by AAW genes (n = 53/81), but it had at least three genes active in all of the CAD-relevant tissues except in blood. This RGN contained two GWA genes (*ABCA1*^{HDL,TC} and *CETP*^{LDL,HDL,TG-TC}); was highly enriched in association with CAD (3.8-fold, p < 8.8 × 10^{-49}), plasma triglyceride levels (5.3-fold, p < 2.3 × 10^{-103}), and fasting glucose levels (1.9-fold, p < 5.6 × 10^{-06}); and had seven key drivers (four in AAW and three in liver) with diverse functions, including transmembrane signaling (Figure 3E). The second RGN was smaller (n = 44 genes) and active across all tissues except liver (Figure 3F). Besides the Chr9p21 risk locus GWA gene *CDKN2A* (Welter et al., 2014) in VF, it had three additional VF key drivers: *CLDN4*, *CTNNAL1*, and *MYSM1*. The third multi-tissue RGN (n = 37 genes), dominated by SF and AAW genes (n = 12 and 22, respectively), was linked to HbA1c levels and enriched in associations with CAD (4.1-fold, p < 3.4 × 10^{-16}), HbA1c levels (3.3-fold, p < 6.8 × 10^{-10}), and insulin levels (3.0-fold, p < 1.8 × 10^{-8}) (Figure 3G). This RGN had three key drivers (*CLEC12A*, *CTSS*, and *FOLR3*) implicated in cell adhesion, cell-cell signaling, glycoprotein turnover, inflammation, and immune responses (e.g., antigen presentation). The RGNs of the remaining 22 CAD-causal cholesterol, glucose, and CRP modules (Figure 2B; Table S2) are shown in Figures S2–S5; the characteristics and directions of gene-gene interactions of the remaining RGNs inferred from all 171 modules are described in Tables S5 and S6.

The CAD GWA genes (Brænne et al., 2015) were included as regulatory priors in the RGN inference. However, compared to eQTLs and transcription factors, their role as transcriptional regulators can be debated. Of 25 GWA genes in 17 of the 30 CAD-causal modules, 13 were identified as key driver genes. To test whether the RGN topology and the identification of the non-GWA key drivers were dependent on including GWA genes as regulatory priors, we repeated the RGN inference without these; 16 RGNs were re-identified and contained, on average, 89% of the same edges and 91% of the same non-GWA key drivers (Supplemental Experimental Procedures).

Cross-Species Validation of CAD-Causal RGNs

Examples of the successful use of mouse models to study atherosclerosis in relation to plasma levels of cholesterol (Björkegren et al., 2014; Skogsberg et al., 2008), glucose (Guo, 2014), and CRP (Kovacs et al., 2007) are many, suggesting that some vascular and metabolic processes leading to CAD/atherosclerosis are shared with humans. This might be particularly true for RGNs thought to be well conserved throughout evolution (Hinman et al., 2003). Therefore, in the next step of the pipeline (Figures 1F and S1), we tested the phenotypic associations of 26 of the 30 CAD-causal RGNs (Figures 3, S2, S3, and

S5) against the HMDP (Bennett et al., 2010); the four CRP RGNs (Figure S4) were not validated, as CRP was not assessed in the HMDP.

The HMDP consists of up to 105 different strains of mice with considerable phenotypic variation and global gene expression from relevant tissues—such as aortic atherosclerotic lesions, liver, heart (used for STAGE SM), and adipose tissue—as well as phenotypic characteristics, similar to the STAGE study (Supplemental Experimental Procedures). In examining mouse orthologs of the 26 CAD-causal RGN genes (Table 1), we found that the activity in three of seven CAD-causal, atherosclerosis-specific modules (Figures 3D–3F) also segregated the 105 strains of mice according to their extent of atherosclerosis. Similarly, three of ten cholesterol-specific modules (Figures S2C, S2D, and S2F), three of five glucose-specific modules (Figures S3A, S3C, and S3E), and at least one phenotype in three modules associated with multiple phenotypes (Figure S5) also segregated the 105 mice of the HMDP according to the corresponding mouse phenotypes (Table 1). Thus, 46% (12/26) of the CAD-causal RGNs/modules were related to corresponding phenotypes in mice.

In addition, we used aortic root lesion eQTLs and SNP p values from a GWAS of the extent of aortic root lesions in the HMDP (Supplemental Experimental Procedures) to assess the extent to which the three cross-species-validated, atherosclerosis-specific modules (Figures 3D–3F) also were causal for the mouse atherosclerosis phenotype. The mouse eQTLs of orthologs of genes in RGN 42 (Figure 3D), but not in RGN 58 (0.2-fold, p = not applicable) and RGN 98 (0.0-fold, p = not applicable) (Figures 3E and 3F), were highly enriched in association with mouse atherosclerosis (2.5-fold, p < 4.3 × 10^{-84}) and, thus, causal for mouse atherosclerosis.

The Role of Cross-Species-Validated Atherosclerosis RGNs in THP-1 Foam Cells

Foam cells form continuously throughout all stages of atherosclerosis development (Lusis, 2000) and, thus, may be affected by CAD-causal RGNs related to the extent of human and mouse atherosclerosis. Accordingly, in the final step of our analysis (Figures 1G and S1), we used differential gene expression and cholesterol-ester accumulation measures (Supplemental Experimental Procedures) to examine the effects of small interfering RNA (siRNA) silencing of key drivers in the three cross-species-validated CAD-causal atherosclerosis-related RGNs (Figures 3D–3F; modules 42, 58, and 98 in Table S2) in an in vitro model of atherosclerosis. THP-1 cells differentiated into macrophages and incubated with acetylated LDL to generate foam cells (Skogsberg et al., 2008). In RGN 42, which is enriched in RNA-processing genes (Figure 3D), individual silencing of four of seven key drivers (*AIP*, *DRAP1*, *POLR2I*, and *PQBP1*) specifically activated RGN 42 network genes in THP-1 macrophages and reduced cholesterol-ester accumulation in THP-1 foam cells (Table 2). In contrast, silencing of key drivers in RGNs 58 and 98 did not specifically activate genes in these RGNs (Supplemental Experimental Procedures; Table S7).

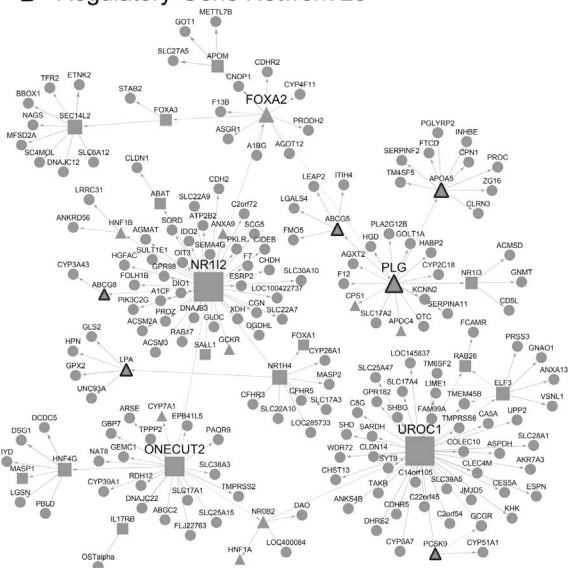## Re-identification of RGN 42 in Independent Gene Expression Data

Since RGN 42 was the sole RGN to emerge from the analysis pipeline (Figures 1A–1G and S1), we used independent gene
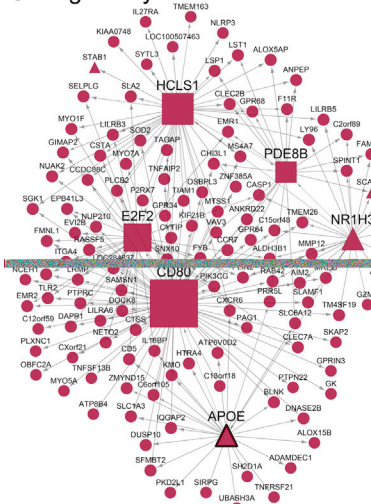
A Regulatory-Gene Network 14
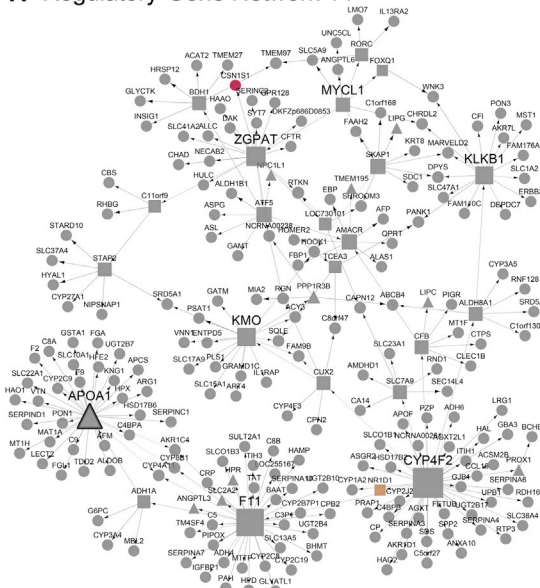
231 genes (7 KDs, 10 GWAS genes, 1 CAD)
Steroid and lipid metabolism

B Regulatory-Gene Network 23

195 genes (5 KDs, 15 GWAS genes, 6 CAD)
1.8-fold enriched with CAD association
Metabolic process and oxidoreductase activity

C Regulatory-Gene Network 37

131 genes (6 KDs, 4 GWAS genes, 1 CAD)
Immune system process

D Regulatory-Gene Network 42
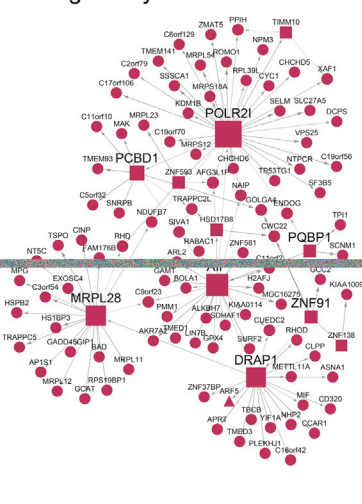
109 genes (7 KDs, 1 GWAS gene)
2.1-fold enriched with CAD association
RNA processing

E Regulatory-Gene Network 58
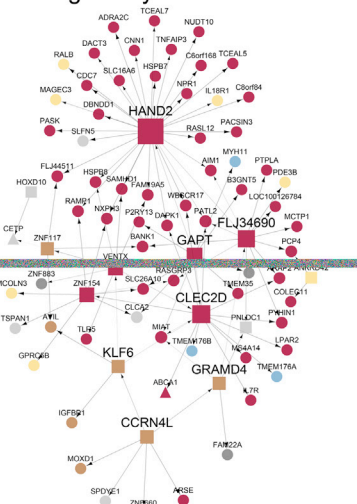
81 genes (7 KDs, 2 GWAS genes)
3.8-fold enriched with CAD association
Regulation of nucleotide biosynthetic process

F Regulatory-Gene Network 98
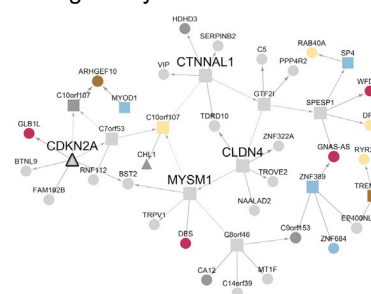
44 genes (4 KDs, 2 GWAS genes, 1 CAD)
Endopeptidase regulator activity

G Regulatory-Gene Network 113
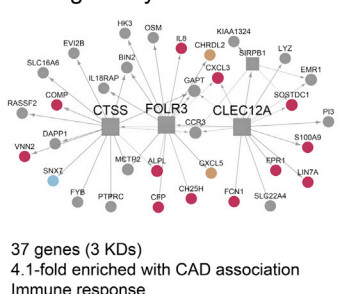
37 genes (3 KDs)
4.1-fold enriched with CAD association
Immune response

H Regulatory-Gene Network 133

31 genes (2 KDs, 1 GWAS gene)
4.6-fold enriched with CAD association
Regulation of inflammatory response

Legend: AAW, IMA, Liver, SM, SF, VF, Blood; TF / eQTL gene, GWAS gene, CAD GWAS gene

(legend on next page)

Table 1. RGN Modules Causally Related to CAD Segregating the STAGE Patients and up to 105 Strains of Mice of the HMDP According to Similar Phenotypes

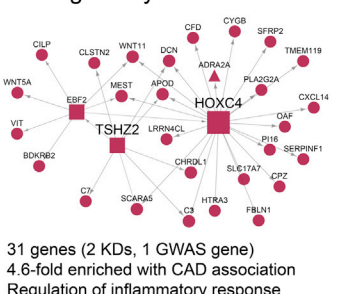| Module ID	STAGE Phenotype	STAGE Tissue	STAGE Combined p Value	HMDP Phenotype	HMDP Tissue	HMDP Combined p Value
Atherosclerosis modules						
42 (Figure 3D)	coronary atherosclerosis	AAW	0.006	atherosclerosis	aorta	0.01
58 (Figure 3E)	coronary atherosclerosis	AAW	0.001	atherosclerosis	aorta	0.0007
98 (Figure 3F)	coronary atherosclerosis	AAW	0.03	atherosclerosis	aorta	0.005
Cholesterol modules						
87 (Figure S2C)	plasma-cholesterol	SF	0.01	LDL/TC	adipose	0.008
95 (Figure S2D)	HDL cholesterol	VF	0.002	LDL cholesterol	adipose	0.009
100 (Figure S2F)	HDL cholesterol	SF	0.006	LDL cholesterol	adipose	0.0002
Glucose modules						
64 (Figure S3A)	pro-insulin	SF	7.0×10^{-05}	insulin	adipose	0.02
112 (Figure S3C)	pro-insulin	liver	0.0007	glucose	liver	0.001
140 (Figure S3E)	blood insulin	SF	2.0×10^{-05}	insulin	adipose	0.009
Multi-phenotype modules						
27 (Figure S5A)	LDL cholesterol	SF	0.006	LDL cholesterol	adipose	0.0001
36 (Figure S5B)	plasma-cholesterol	SM	5.0×10^{-08}	TC	heart	0.02
105 (Figure S5C)	blood insulin	SF	8.0×10^{-05}	insulin	adipose	0.02

AAW expression of mouse orthologs for genes in the STAGE atherosclerosis modules were tested against the extent of aortic root lesions in up to 105 strains of mice cross-bred with the ApoE$_{Leiden}$ transgenic pro-atherogenic mouse model. Adipose tissue, heart, and liver expression of orthologs of genes in the STAGE cholesterol and glucose modules were tested against corresponding phenotypes in up to 105 strains of wild-type mice. The extent of human coronary atherosclerosis was assessed from coronary angiograms; mouse atherosclerosis was assessed by Oil-Red-O staining of the aortic roots. The p values are combined from second-step clustering and eigengene module correlations in both STAGE and HMDP (Supplemental Experimental Procedures).

expression datasets to examine the extent to which it could be re-identified in the following: (1) in-house global gene expression datasets from primary blood macrophage samples (n = 36) and carotid lesions (n = 25) isolated from patients before and during carotid surgery, respectively (Hägg et al., 2009); and (2) datasets generated by others from lipopolysaccharide-stimulated monocytes (n = 18) and macrophages (n = 18) from CAD patients (Schirmer et al., 2009). Reassuringly, the 109 genes in RGN 42 were more strongly co-expressed than sets of 109 random genes, especially in the carotid lesions (9.5-fold higher connectivity of RGN 42 genes than of genes in 10,000 random sets, $p < 1.0 \times 10^{-300}$) and also in macrophages (3.7-fold, $p = 6.2 \times 10^{-37}$) (Hägg et al., 2009), CAD lipopolysaccharide-stimulated monocytes (2.2-fold, $p = 1.1 \times 10^{-13}$), and CAD macrophages (2.0-fold, $p = 1.1 \times 10^{-11}$ (Schirmer et al., 2009; Supplemental Experimental Procedures).

CAD-Causal Arterial Wall RGNs and Gene Connectivity

Finally, we analyzed the CAD phenotype-associated modules outside the main analysis pipeline (Figures 1A–1G and S1). First, we utilized a unique feature of the STAGE cohort, a nearly atherosclerosis-free arterial wall sample (i.e., IMA), as an internal control for the AAW. When molecular networks reflecting physiological events are challenged with the effects of genetic and environmental disease perturbations, gene-gene connectivity increases for disease-relevant processes (Zhang et al., 2013). Accordingly, to look for changes in connectivity in the STAGE modules, we compared AAW (disease) and IMA (healthy).

For the 13 AAW/IMA-specific or dominant modules associated with CAD phenotypes (Figure 2B), we calculated the gain/loss in module connectivity (Figure 4). Interestingly, all five arterial wall modules with >2-fold increases in connectivity also were CAD causal (Figure 4). Of the eight arterial wall modules with <2-fold increases in connectivity (seven AAW and one IMA), only two were CAD causal (both in AAW). In both of these, gene connectivity was already high in IMA (e.g., module 42, Figure 4). Thus, CAD-causal modules appear to be characterized by high or increased connectivity, whereas CAD-reactive modules either do not change or lose connectivity in the arterial wall as CAD develops.

To identify potential causes of the connectivity differences, we analyzed cell type enrichment and transcription factor-binding site enrichment (Roider et al., 2009) for the seven CAD-causal modules with >2-fold increases in connectivity (modules 20, 37, 42, 58, 69, 133, and 138) (Figure 4). For the cell type enrichment,

Figure 3. CAD-Causal RGNs Associated with Coronary Atherosclerosis

(A–H) Bayesian RGNs with key drivers inferred (Experimental Procedures) from CAD-causal modules (i.e., module eQTLs are risk enriched or contain CAD GWA candidate genes) linked to extent of coronary atherosclerosis. Fold enrichment for CAD association was assessed for network eQTLs/SNPs using the CARDIoGRAM dataset. The molecular process with the strongest functional enrichment assessed by gene ontology is indicated. KD, key driver; GWA genes, total and CAD candidate genes identified in GWASs of CAD, plasma lipid/glucose levels, and type 2 diabetes.

Table 2. Effects of siRNA Targeting of the AAW Key Drivers in RGN Module 42 (Figure 3D) on Network Gene Activity and Cholesterol-Ester Accumulation in THP-1 Macrophages

Key Drivers in Module 42	Effects of siRNA Inhibition of Key Drivers on Network Genes (n = 109) in THP-1 Macrophages				Effects of siRNA Inhibition of Key Drivers on CE Accumulation in THP-1 Foam Cells			
	Total Affected Genes	Hyper-geometric Test	Total Affected Genes	Hyper-geometric Test	Control (Relative CE Levels)	siRNA (Relative CE Levels)	CE Content (% Relative Control)	p Value
	Nominal p Value <0.05		FDR < 10%					
POLR2I	62	2.2×10^{-06}	59	1.0×10^{-06}	100 ± 48[a]	38 ± 24[a]	−62	0.001
PQBP1	41	0.0002	24	0.002	100 ± 15[a]	69 ± 11[a]	−31	7.0×10^{-05}
AIP	25	0.0005	9	0.003	100 ± 15[a]	149 ± 81[a]	+49	0.02
DRAP1	32	0.003	12	0.02	100 ± 15[a]	70 ± 17[a]	−30	0.0007
MRPL28	15	0.03	1	0.44	–	–	–	–
PCBD1	13	0.03	0	1.0	–	–	–	–
ZNF91	24	0.07	0	1.0	–	–	–	–

All siRNA experiments achieved >80% gene silencing except for POLR2I (>50%). CE, cholesterol ester; FDR, false discovery rate; –, not tested.
[a]Values are mean ± SD.

we examined 321 of 409 genes with matches in CTen (Shoemaker et al., 2012) that were responsible for the increased connectivity in these top five modules. CD14-positive monocytes stood out with an enrichment score of 65 (CD33-positive myeloid cells had the second highest enrichment score, 49), suggesting that connectivity is increased by the infiltration and possibly differentiation of circulating monocytes into/in the atherosclerotic plaque.

For the transcription factor-binding site enrichment, genes in module 58 were enriched for binding sites of *MEF2A*, a transcriptional activator implicated in CAD (Bhagavatula et al., 2004; Wang et al., 2013). Interestingly, among *MEF2A*'s target genes were the first (*CLEC2D*) and fifth (*CCRN4L*) order key drivers in module 58; *CCRN4L* has been linked to cardiovascular diseases (Binns et al., 2009).

A Super-Network of 30 CAD-Causal RGNs

The second analysis we performed outside the main pipeline (Figures 1A–1G and S1) was to identify a super-network of the 30 CAD-causal RGNs/modules by calculating the correlation values between the module eigengenes (Supplemental Experimental Procedures; Langfelder and Horvath, 2007). We found that all 30 CAD-causal RGNs/modules were interconnected and together formed a super-network connected across all seven tissues (r > 0.4, p < 0.05, Figure 5). Within this super network, SF dominated with 13 modules, AAW with seven, and SM and VF with four modules each; notably, blood and liver had only one CAD-causal module each. Nine modules were connected only within a given tissue as follows: modules 14, 23, 148, and 152 in SF; module 8 in VF; modules 89 and 99 in SM; and modules 69 and 133 in AAW (Figure 5). In contrast, the cross-tissue modules 98 (dominated by VF genes including its key drivers) and 113 (dominated by SF genes including its key drivers) appeared to act as hub modules, mediating all but one (module 27 in SF) connection from SM and liver (module 98) and from SF (module 113) to the AAW (none of the CAD-causal modules were identified in the IMA).

The notion that cross-tissue modules can serve as hubs in this super-network of CAD-causal modules was reinforced by the greater and stronger connectivity in the super-network of

cross-tissue modules than of tissue-specific modules (mean number of connections: 5.33 ± 1.73 versus 3.43 ± 2.04, p < 0.02; mean connectivity strength: 2.96 ± 0.98 versus 1.98 ± 1.20, p < 0.03). Notably, module 98 contained genes involved in endopeptidase activity (Table S3) previously implicated in atherosclerosis (Kugiyama et al., 1996), and one of its four key drivers, *CDKN2A*, is a CAD candidate gene for the Chr9p21 risk locus (Welter et al., 2014; Table S2). Interestingly, the allelic variation of the Chr9p21 risk SNP (rs4977574) was associated with increased connectivity of the VF genes in module 98 (risk genotype G/G, n = 20 STAGE patients; mean connectivity ± SD, 2.34 ± 1.01), the non-risk heterozygotes (A/G, n = 47 STAGE patients, mean connectivity 1.37 ± 0.33); and the non-risk homozygotes (A/A, n = 21 STAGE patients, mean connectivity 1.43 ± 0.34) (p = 0.002, Kruskal-Wallis test).

DISCUSSION

In this study, we used systems genetics integrating the analysis of (1) DNA genotypes and gene expression profiles in seven CAD-relevant tissues from CAD patients (STAGE cohort) and (2) CAD GWA datasets (primarily Coronary Artery Disease Genome-wide Replication and Meta-analysis [CARDIoGRAM]). Our findings provide a preliminary view of the regulatory landscape of causal molecular processes active within and across a majority of tissues believed to be central to advanced CAD. The identified RGNs included both established (Brænne et al., 2015) and previously unreported CAD candidate genes in the form of key drivers. These candidate genes participate in diverse molecular processes and established pathways of atherosclerosis, cholesterol and glucose metabolism, and acute inflammation, and they were regulated in both tissue-specific and cross-tissue networks. Importantly, we found that nearly half of the RGNs (where similar mouse phenotypes were available, 12/26, 46%) were evolutionarily conserved, as judged from validation against the HMDP (Bennett et al., 2010). As proof of concept, in one cross-species-validated, mouse atherosclerosis- and CAD-causal network active in AAW and involving RNA-processing genes, four key drivers (*AIP*, *DRAP1*, *POLR2I*, and *PQBP1*) specifically activated the same network genes

and affected THP-1 foam cell formation, and the entire network also was re-identified in independent gene expression data from both CAD macrophages and carotid lesions.

It is increasingly recognized that to defeat complex diseases like CAD it is not sufficient to focus on individual DNA variants or CAD candidate genes. In parallel, we need to use systems genetics to understand how these disease genes operate in complex regulatory networks (Schadt, 2009). In examining the 171 RGNs inferred from all STAGE modules (Figures 3 and S2–S5; Tables S5 and S6) in greater detail, we found a good representation of established genes, pathways, and molecular functions in CAD. For instance, these modules contained 147 of the 309 GWA genes thus far proposed for CAD and CAD risk factors, such as lipid and glucose metabolism, including type 2 diabetes (Table S4). Thus, our results should be useful in gaining a better understanding of the tissue-specific or cross-tissue molecular contexts of established CAD candidate genes identified by GWASs. Our results will be particularly useful in providing clues about their upstream regulations by key disease drivers, which frequently were not the GWA genes themselves.

Although it is vital to identify individual CAD candidate genes in the context of regulatory networks, RGNs do not act in isolation but interact. Indeed, we identified a super-network containing all 30 CAD-causal RGNs across all the main CAD tissues. The tissue distribution of RGN interactions is significant, as it links a good portion of disease-driving molecular processes in CAD, including their relation to key metabolic risk factors (e.g., cholesterol and glucose metabolism), and it provides a preliminary overview of the gene regulatory landscape in CAD. Mapping the regulatory framework of complex diseases in this fashion provides a starting point to assess the overall molecular status of individual patients (Björkegren et al., 2015). In fact, more detailed versions of regulatory maps like the one presented here (Figure 5) are required to achieve the goals of precision medicine.

Notably, most of the CAD-causal RGNs in this super-network were active in nonhepatic peripheral organs, particularly SF (Figure 5); only one was in the liver. The paucity of hepatic modules/RGNs and the unexpectedly high number of fat modules/RGNs may be consistent with the emerging notion that key hepatic processes in cardiometabolic diseases are governed by metabolic processes and gene expression activity in nonhepatic peripheral tissues (Lomonaco et al., 2012). In fact, the connectivity map of the super-network of CAD-causal modules suggests that causal effects of metabolic processes in the SM and liver on the CAD-causal modules of the AAW are largely mediated by a single cross-tissue RGN primarily active in VF. This RGN also was found to contain CDKN2A, a candidate gene for the well-established Chr9p21 CAD risk locus, as one of its four key drivers (Figure 5). Moreover, the connectivity of this RGN was found to be associated with the allele frequency of the Chr9p21 CAD risk locus. Similarly, effects on the CAD-causal modules of the AAW of the many RGNs identified in SF also were predominantly mediated by one cross-tissue module (Figure 5).

Although a systems genetics approach has the significant advantage of simultaneously considering many genes (i.e., RGNs) in relation to disease, there are also challenges (Civelek and Lusis, 2014). In brief, cell diversity in the tissue biopsies can be a problem, as changes in gene activity also will reflect changes in cell type content. Furthermore, the inference of causal (i.e., directed) RGNs from omics data can be hampered by data noise, limited sample size, and the inherent limitations of representing complex biological processes by statistical models, which inevitably leads to false positives even among the most significant predictions (Marbach et al., 2010). Thus, as with gene-by-gene approaches, RGNs inferred from genetics of gene expression studies must be examined and validated extensively.
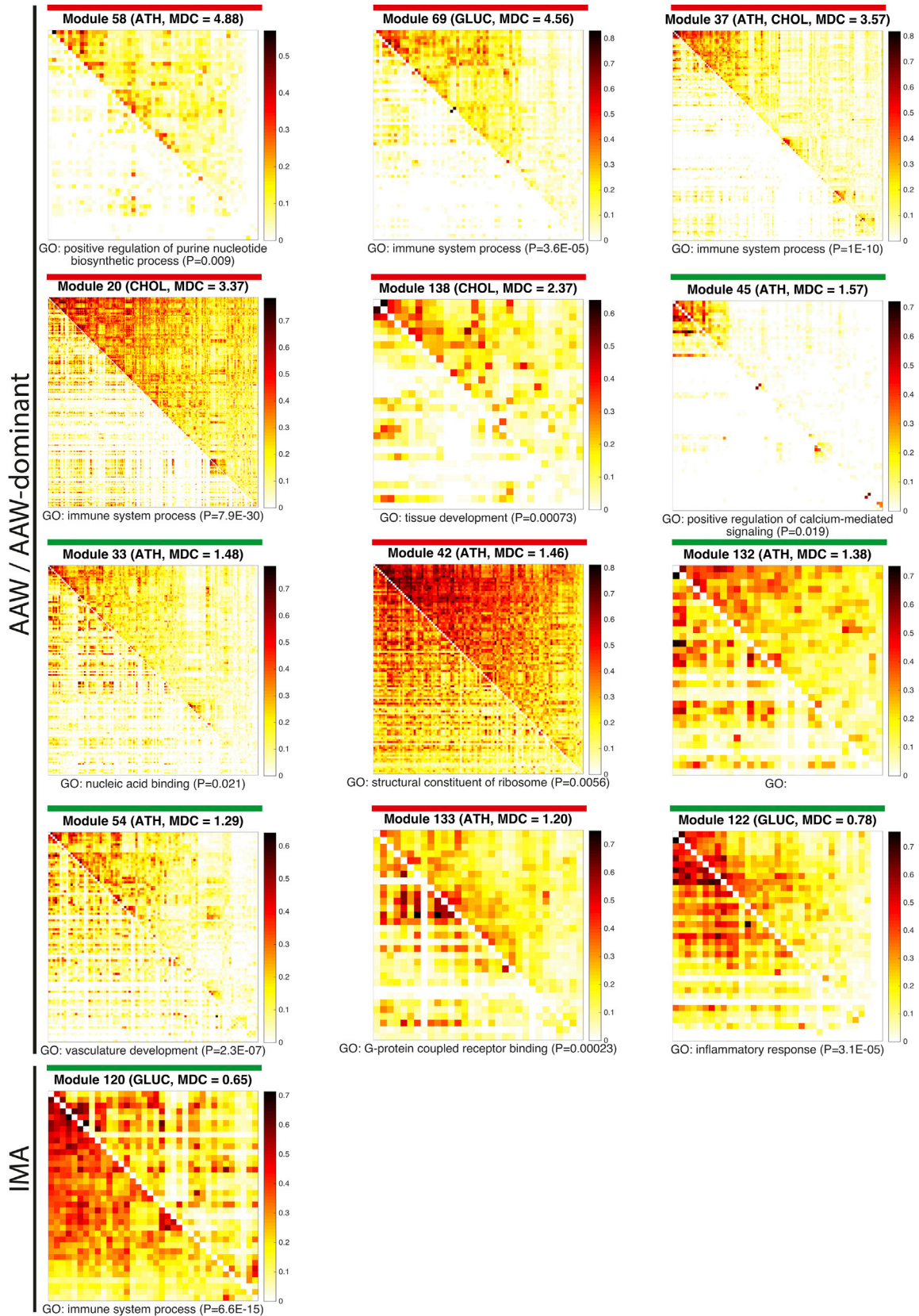
To this point, we validated three atherosclerosis-related RGNs using an in vitro model of atherosclerosis (Skogsberg et al., 2008; Figures 1 and S1, Step G). Silencing of the key drivers of RGN 42, AIP, DRAP1, POLR2I, and PQBP1, which involve RNA processing, specifically reactivated RGN 42 genes in THP-1 macrophages and affected cholesterol-ester accumulation (i.e., foam cell formation) (Table 2). Although none of these four key drivers have been directly linked to CAD, atherosclerosis, or foam cell formation, AIP affects the expression of peroxisome proliferator-activated receptors α and β (Sumanasekera et al., 2003a, 2003b) and, thus, may prevent foam cell formation via the HDL/ABCA1 pathway (Chinetti et al., 2001) or the CD36 scavenger receptor pathway (Li et al., 2004). Similarly, DRAP1 represses TGF-β signaling by interfering with the binding of FoxH1-Smad2/3/4 to its DNA targets during transcription (Iratni et al., 2002). TGF-β signaling prevents foam cell formation by activating ABC1 and reverse cholesterol transport (Kozaki et al., 1997). POLR2I and PQBP1, however, have not even been indirectly associated with atherosclerosis or foam cell formation, returning only 93 and 69 hits, respectively, in PubMed. POLR2I encodes a subunit of RNA polymerase II, the polymerase responsible for synthesizing mRNA in eukaryotes and also targeted by HIV-1 during lytic and latent viral stages (Nilson and Price, 2011). PQBP1 is a key regulator of mRNA processing and gene transcription. In fact, mutations in PQBP1 have been reported in several X chromosome-linked intellectual disability disorders, including Golabi-Ito-Hall syndrome (Sudol et al., 2012). PQBP1 also may be linked to fat metabolism (Takahashi et al., 2009), and, as a splicing factor, it alters fibroblast growth factor signaling (Wang et al., 2013). The fact that RGN 42 is related to RNA processing may be significant in view of the gene-regulatory roles of noncoding RNAs in general (Mihailescu, 2015) and in CAD (Bronze-da-Rocha, 2014).

In summary, our identification of several high-hierarchy, evolutionarily conserved, strongly inherited risk-enriched, and reproducible RGNs is a breakthrough in our understanding of the molecular landscape causing CAD across tissues. Specifically, we have shown for the first time that RNA-processing genes appear central in causing CAD. Key regulatory genes in this and other RGNs may be proven useful as targets for novel CAD therapies.
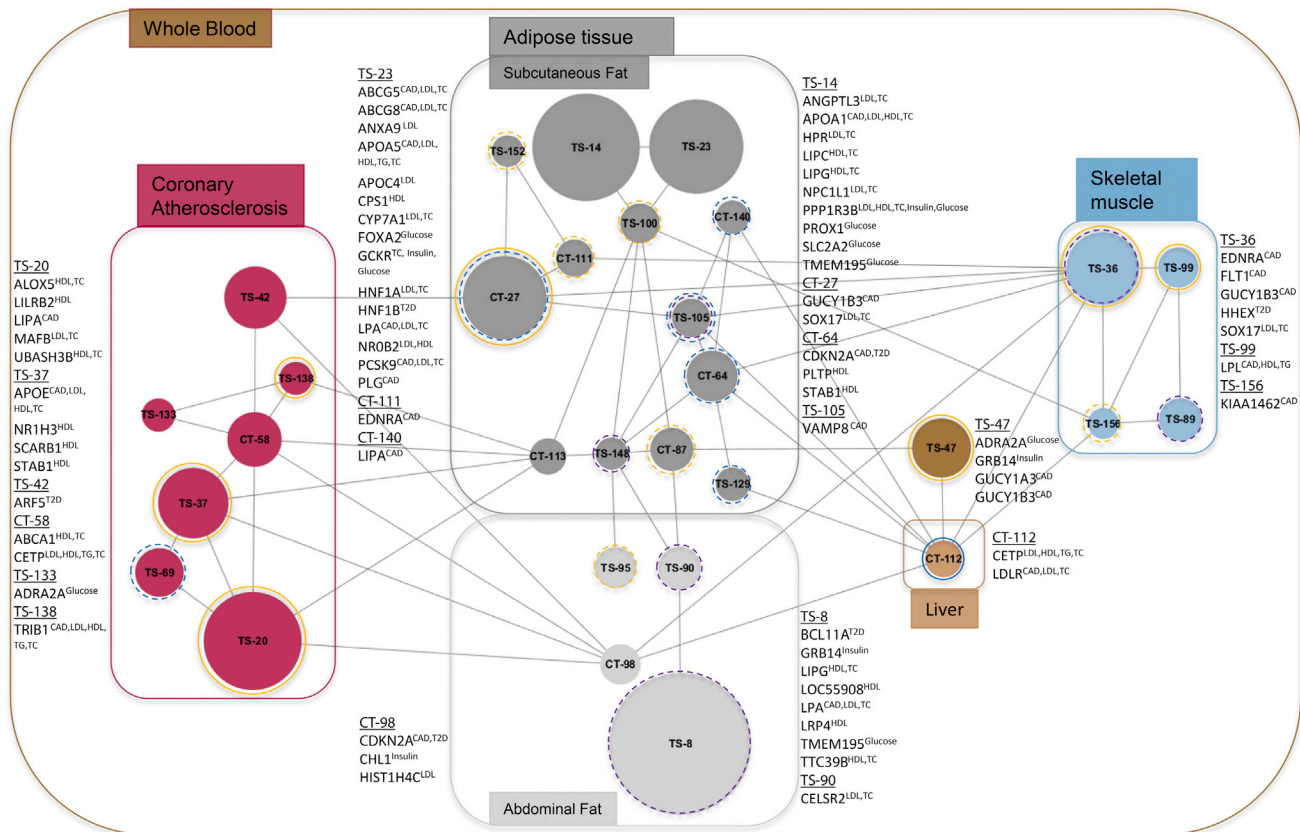
EXPERIMENTAL PROCEDURES

The STAGE Genetics of Gene Expression Study

In the STAGE study, 612 tissue samples were obtained during coronary artery bypass grafting surgery from AAW (n = 73), IMA (n = 88), liver (n = 87), SM (n = 89), SF (n = 72), and VF (n = 98) of well-characterized CAD patients (Hägg et al., 2009). Fasting whole blood was obtained for isolation of DNA (n = 109) and RNA (n = 105) and biochemical analyses. RNA samples were used for gene expression profiling with a custom Affymetrix array (HuRSTA-2a520709). Blood DNA was genotyped for 909,622 SNPs with the Affymetrix

AAW / AAW-dominant

Module 58 (ATH, MDC = 4.88)
GO: positive regulation of purine nucleotide biosynthetic process (P=0.009)

Module 69 (GLUC, MDC = 4.56)
GO: immune system process (P=3.6E-05)

Module 37 (ATH, CHOL, MDC = 3.57)
GO: immune system process (P=1E-10)

Module 20 (CHOL, MDC = 3.37)
GO: immune system process (P=7.9E-30)

Module 138 (CHOL, MDC = 2.37)
GO: tissue development (P=0.00073)

Module 45 (ATH, MDC = 1.57)
GO: positive regulation of calcium-mediated signaling (P=0.019)

Module 33 (ATH, MDC = 1.48)
GO: nucleic acid binding (P=0.021)

Module 42 (ATH, MDC = 1.46)
GO: structural constituent of ribosome (P=0.0056)

Module 132 (ATH, MDC = 1.38)
GO:

Module 54 (ATH, MDC = 1.29)
GO: vasculature development (P=2.3E-07)

Module 133 (ATH, MDC = 1.20)
GO: G-protein coupled receptor binding (P=0.00023)

Module 122 (GLUC, MDC = 0.78)
GO: inflammatory response (P=3.1E-05)

IMA

Module 120 (GLUC, MDC = 0.65)
GO: immune system process (P=6.6E-15)

(legend on next page)

Figure 5. A Super-Network of CAD-Causal Modules
Eigengene associations (r > 0.4, p < 0.05) were used to link the 30 CAD-causal modules (Experimental Procedures). RGNs/modules (circles) are oriented according to dominating tissue-belonging of genes and are color coded accordingly. Numbers in circles are the module IDs. Circle size corresponds to the number of RGN/module genes. Next to each box, the names of GWA genes in the RGNs/modules and related trait (in superscript) are indicated. TS, tissue-specific RGN/module; CT, cross-tissue RGN/module; colored circle circumferences, phenotype associations; dotted lines, RGNs/modules that are reactive to indicated phenotype; solid line, RGNs/modules that are causal for indicated phenotype (i.e., GWA gene[s] for the phenotype or eQTLs enriched with association for the phenotype); orange, plasma cholesterol measures; blue, plasma glucose metabolism measures; purple, plasma CRP levels.

GenomewideSNP_6 array. The carotid lesion and CAD macrophage gene expression data were generated with the same custom Affymetrix arrays mentioned above (Shang et al., 2014).

Multi-tissue Weighted Gene Co-expression Network Reconstruction
WGCNA (Zhang and Horvath, 2005) was used to identify co-expression network modules that simultaneously captured gene-gene relations within and across tissues (Table S8), using the following most variably expressed genes (SD > 0.5 across samples): 3,369 in AAW; 2,761 in IMA; 1,588 in liver; 2,492 in SM; 3,596 in SF; 5,215 in VF; and 1,891 in blood. In WGCNA, values of $\beta = 6$ and $\beta = 3$ were used to weigh within-tissue and across-tissue correlations, respectively, to ensure that both tissue-specific and cross-tissue subnetworks, as well as whole networks, were scale free.

Associations of Modules and CAD Phenotypes
Module associations with four main CAD phenotypes (Results) were tested as follows: (1) by a rank-sum test of the mean phenotype values of the two most distinct patient subsets, defined by clustering the gene activity within each

module; and (2) by calculating Pearson correlations between phenotype and module eigengene values (Langfelder and Horvath, 2007). Integrated p values from both tests and from multiple measurements for each main phenotype (e.g., TC, LDL, VLDL, and HDL for plasma cholesterol levels) were computed with a weighted data integration method (Hwang et al., 2005) and corrected for multiple testing by estimating a false discovery rate (FDR < 0.2) (Storey, 2002).

eQTL Enrichment for Risk Associations and CAD Candidate Genes According to GWASs
To examine the extent to which modules were CAD causal, we used enrichment in inherited CAD association or GWAS candidate genes (Deloukas et al., 2013). In brief, disease assocations of cis-eQTLs detected in the STAGE patients (or, if the number of cis-eQTLs was less than ten, SNPs located within ±500 kb of the transcription start or end site of each module gene) were used to calculate fold enrichment and statistical significance compared to 10,000 random SNP groups corrected for chromosome, gene density, and major allele frequency (Foroughi Asl et al., 2015). CAD candidate genes (n = 53)

Figure 4. Differential Connectivity in Arterial Wall Modules with and without Atherosclerosis
Differential connectivity plots for AAW- and IMA-dominant, phenotype-associated modules. Each plot shows the adjacency values (weighted correlations) between all pairs of module genes in IMA (lower triangle) and AAW (upper triangle). Modules are sorted from high to low according to module differential connectivity (MDC) (ratios of sum of connectivity values in AAW to IMA). Red bars indicate CAD-causal modules. Green bars indicate non-causal modules. MDC values >1 correspond to modules with a gain of connectivity in AAW versus IMA.

were obtained from CARDIoGRAM. Similarly eQTL disease associations and additional GWA candidate genes for CAD risk factors were computed using additional GWASs (Supplemental Experimental Procedures) and data from the HMDP.

RGN Reconstruction

Gaussian Bayesian networks were reconstructed for each module by imposing a prior that only genes with eQTLs, transcription factors, or CAD GWA genes could be parents of other genes. The Bayesian Information Criterion (BIC) was used to score models; a multiple-restart greedy hill-climbing algorithm, with edge additions, deletions, and reversals, was used to search locally optimal models (Schmidt et al., 2007).

Key Driver Analysis and Validation

Key drivers were identified from the Bayesian networks as described previously (Zhang and Zhu, 2013), and they were validated by siRNA silencing in THP-1 macrophages incubated with Ac-LDL to induce foam cell formation (Table S9), as described previously (Skogsberg et al., 2008). Gene expression data from siRNA-treated and control THP-1 cells were generated by Agilent Human Custom Gene Expression Microarray 8×15 and compared to assess the effect of a key driver on its RGN, as described previously (Björkegren et al., 2014).

Mouse Phenotype Associations

To validate the phenotype associations of CAD-causal modules, we assessed the association of ortholog genes in corresponding mouse tissues and phenotypes in the HMDP (Bennett et al., 2010), using the same methods described.

Module Connectivity and Differential Connectivity

Total module connectivity was calculated as the sum of adjacency values (i.e., weighted correlation coefficients) between all pairs of genes in a module. Module differential connectivity between AAW and IMA was calculated as the ratio of total module connectivity in AAW to total module connectivity in IMA.

ACCESSION NUMBERS

The accession number for the STAGE data reported in this paper is GEO: GSE40231. The accession numbers for the HMDP data reported in this paper are GEO: GSE66570, GSE64768, and GSE64769.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and nine tables and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2016.02.002.

AUTHOR CONTRIBUTIONS

Conceptualization, H.A.T., T.M., and J.L.M.B.; Methodology, H.A.T., J.S., T.M., and J.L.M.B.; Software, H.A.T., H.F.A., and T.M.; Formal Analysis, H.A.T., H.F.A., O.F., and T.M.; Investigation, H.A.T., R.K.J., J.S., T.M., and J.L.M.B.; Resources, R.E., A.R., B.A.K., B.R., C.G., T.I., J.T.D., M.C., A.J.L., J.S., and J.L.M.B.; Writing – Original Draft, H.A.T, J.S., T.M., and J.L.M.B.; Writing – Review & Editing, H.A.T, J.C.K., J.S., T.M., and J.L.M.B.; Supervision, E.E.S., J.S., T.M., and J.L.M.B.

ACKNOWLEDGMENTS

REFERENCES

Bangalore, S., Steg, G., Deedwania, P., Crowley, K., Eagle, K.A., Goto, S., Ohman, E.M., Cannon, C.P., Smith, S.C., Zeymer, U., et al.; REACH Registry Investigators (2012). β-Blocker use and clinical outcomes in stable outpatients with and without coronary artery disease. JAMA 308, 1340–1349.

Barabási, A.L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat. Rev. Genet. 12, 56–68.

Bennett, B.J., Farber, C.R., Orozco, L., Kang, H.M., Ghazalpour, A., Siemers, N., Neubauer, M., Neuhaus, I., Yordanova, R., Guan, B., et al. (2010). A high-resolution association mapping panel for the dissection of complex traits in mice. Genome Res. 20, 281–290.

Bhagavatula, M.R., Fan, C., Shen, G.Q., Cassano, J., Plow, E.F., Topol, E.J., and Wang, Q. (2004). Transcription factor MEF2A mutations in patients with coronary artery disease. Hum. Mol. Genet. 13, 3181–3188.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., and Apweiler, R. (2009). QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 25, 3045–3046.

Björkegren, J.L., Hägg, S., Talukdar, H.A., Foroughi Asl, H., Jain, R.K., Cedergren, C., Shang, M.M., Rossignoli, A., Takolander, R., Melander, O., et al. (2014). Plasma cholesterol-induced lesion networks activated before regression of early, mature, and advanced atherosclerosis. PLoS Genet. 10, e1004201.

Björkegren, J.L., Kovacic, J.C., Dudley, J.T., and Schadt, E.E. (2015). Genome-wide significant loci: how important are they? Systems genetics to understand heritability of coronary artery disease and other common complex disorders. J. Am. Coll. Cardiol. 65, 830–845.

Brænne, I., Civelek, M., Vilne, B., Di Narzo, A., Johnson, A.D., Zhao, Y., Reiz, B., Codoni, V., Webb, T.R., Foroughi Asl, H., et al.; Leducq Consortium CAD Genomics‡ (2015). Prediction of causal candidate genes in coronary artery disease loci. Arterioscler. Thromb. Vasc. Biol. 35, 2207–2217.

Bronze-da-Rocha, E. (2014). MicroRNAs expression profiles in cardiovascular diseases. BioMed Res. Int. 2014, 985408.

Chinetti, G., Lestavel, S., Bocher, V., Remaley, A.T., Neve, B., Torra, I.P., Teissier, E., Minnich, A., Jaye, M., Duverger, N., et al. (2001). PPAR-alpha and PPAR-gamma activators induce cholesterol removal from human macrophage foam cells through stimulation of the ABCA1 pathway. Nat. Med. 7, 53–58.

Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nat. Rev. Genet. 15, 34–48.

Collins, F.S., and Varmus, H. (2015). A new initiative on precision medicine. N. Engl. J. Med. 372, 793–795.

Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T.L., Thompson, J.R., Ingelsson, E., Saleheen, D., Erdmann, J., Goldstein, B.A., et al.; CARDIoGRAMplusC4D Consortium; DIAGRAM Consortium; CARDIOGENICS Consortium; MuTHER Consortium; Wellcome Trust Case Control Consortium (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. Nat. Genet. 45, 25–33.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. Nature 452, 423–428.

Foroughi Asl, H., Talukdar, H.A., Kindt, A.S., Jain, R.K., Ermel, R., Ruusalepp, A., Nguyen, K.D., Dobrin, R., Reilly, D.F., Schunkert, H., et al.; CARDIoGRAM Consortium (2015). Expression quantitative trait Loci acting across multiple tissues are enriched in inherited risk for coronary artery disease. Circ Cardiovasc Genet 8, 305–315.

Guo, S. (2014). Insulin signaling, resistance, and the metabolic syndrome: insights from mouse models into disease mechanisms. J. Endocrinol. 220, T1–T23.

Hägg, S., Skogsberg, J., Lundström, J., Noori, P., Nilsson, R., Zhong, H., Maleki, S., Shang, M.M., Brinne, B., Bradshaw, M., et al. (2009). Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: the Stockholm Atherosclerosis Gene Expression (STAGE) study. PLoS Genet. 5, e1000754.

Hinman, V.F., Nguyen, A.T., Cameron, R.A., and Davidson, E.H. (2003). Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. Proc. Natl. Acad. Sci. USA 100, 13356–13361.

Hwang, D., Rust, A.G., Ramsey, S., Smith, J.J., Leslie, D.M., Weston, A.D., de Atauri, P., Aitchison, J.D., Hood, L., Siegel, A.F., and Bolouri, H. (2005). A data integration methodology for systems biology. Proc. Natl. Acad. Sci. USA 102, 17296–17301.

Iratni, R., Yan, Y.T., Chen, C., Ding, J., Zhang, Y., Price, S.M., Reinberg, D., and Shen, M.M. (2002). Inhibition of excess nodal signaling during mouse gastrulation by the transcriptional corepressor DRAP1. Science 298, 1996–1999.

Kovacs, A., Tornvall, P., Nilsson, R., Tegnér, J., Hamsten, A., and Björkegren, J. (2007). Human C-reactive protein slows atherosclerosis development in a mouse model with human-like hypercholesterolemia. Proc. Natl. Acad. Sci. USA 104, 13768–13773.

Kozaki, K., Akishita, M., Eto, M., Yoshizumi, M., Toba, K., Inoue, S., Ishikawa, M., Hashimoto, M., Kodama, T., Yamada, N., et al. (1997). Role of activin-A and follistatin in foam cell formation of THP-1 macrophages. Arterioscler. Thromb. Vasc. Biol. 17, 2389–2394.

Kugiyama, K., Sugiyama, S., Matsumura, T., Ohta, Y., Doi, H., and Yasue, H. (1996). Suppression of atherosclerotic changes in cholesterol-fed rabbits treated with an oral inhibitor of neutral endopeptidase 24.11 (EC 3.4.24.11). Arterioscler. Thromb. Vasc. Biol. 16, 1080–1087.

Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. 1, 54.

Li, A.C., Binder, C.J., Gutierrez, A., Brown, K.K., Plotkin, C.R., Pattison, J.W., Valledor, A.F., Davis, R.A., Willson, T.M., Witztum, J.L., et al. (2004). Differential inhibition of macrophage foam-cell formation and atherosclerosis in mice by PPARalpha, beta/delta, and gamma. J. Clin. Invest. 114, 1564–1576.

Lomonaco, R., Ortiz-Lopez, C., Orsak, B., Webb, A., Hardies, J., Darland, C., Finch, J., Gastaldelli, A., Harrison, S., Tio, F., and Cusi, K. (2012). Effect of adipose tissue insulin resistance on metabolic parameters and liver histology in obese patients with nonalcoholic fatty liver disease. Hepatology 55, 1389–1397.

Lusis, A.J. (2000). Atherosclerosis. Nature 407, 233–241.

Marbach, D., Prill, R.J., Schaffter, T., Mattiussi, C., Floreano, D., and Stolovitzky, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. Proc. Natl. Acad. Sci. USA 107, 6286–6291.

Mensah, G.A., Moran, A.E., Roth, G.A., and Narula, J. (2014). The global burden of cardiovascular diseases, 1990-2010. Glob. Heart 9, 183–184.

Mihailescu, R. (2015). Gene expression regulation: lessons from noncoding RNAs. RNA 21, 695–696.

Nilson, K.A., and Price, D.H. (2011). The role of RNA polymerase II elongation control in HIV-1 gene expression, replication, and latency. Genet. Res. Int. 2011, 726901.

Peden, J.F., and Farrall, M. (2011). Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. Hum. Mol. Genet. 20 (R2), R198–R205.

Roider, H.G., Manke, T., O'Keeffe, S., Vingron, M., and Haas, S.A. (2009). PASTAA: identifying transcription factors associated with sets of co-regulated genes. Bioinformatics 25, 435–442.

Samani, N.J., and de Bono, D.P. (1996). Prevention of coronary heart disease with pravastatin. N. Engl. J. Med. 334, 1333–1334, author reply 1334–1335.

Schadt, E.E. (2009). Molecular networks as sensors and drivers of common human diseases. Nature 461, 218–223.

Schadt, E.E., and Björkegren, J.L. (2012). NEW: network-enabled wisdom in biology, medicine, and health care. Sci. Transl. Med. 4, 115rv1.

Schirmer, S.H., Fledderus, J.O., van der Laan, A.M., van der Pouw-Kraan, T.C., Moerland, P.D., Volger, O.L., Baggen, J.M., Böhm, M., Piek, J.J., Horrevoets, A.J.G., and van Royen, N. (2009). Suppression of inflammatory signaling in monocytes from patients with coronary artery disease. J. Mol. Cell. Cardiol. 46, 177–185.

Schmidt, M., Niculescu-Mizil, A., and Murphy, K. (2007). Learning graphical model structure using L1-regularization paths. In Proceedings of the 22nd National Conference on Artificial Intelligence, 2, A. Cohn, ed (AAAI Press), pp. 1278–1283.

Shang, M.M., Talukdar, H.A., Hofmann, J.J., Niaudet, C., Asl, H.F., Jain, R.K., Rossignoli, A., Cedergren, C., Silveira, A., Gigante, B., et al. (2014). Lim domain binding 2: a key driver of transendothelial migration of leukocytes and atherosclerosis. Arterioscler. Thromb. Vasc. Biol. 34, 2068–2077.

Shoemaker, J.E., Lopes, T.J.S., Ghosh, S., Matsuoka, Y., Kawaoka, Y., and Kitano, H. (2012). CTen: a web-based platform for identifying enriched cell types from heterogeneous microarray data. BMC Genomics 13, 460.

Skogsberg, J., Lundström, J., Kovacs, A., Nilsson, R., Noori, P., Maleki, S., Köhler, M., Hamsten, A., Tegnér, J., and Björkegren, J. (2008). Transcriptional profiling uncovers a network of cholesterol-responsive atherosclerosis target genes. PLoS Genet. 4, e1000036.

Storey, J.D. (2002). A direct approach to false discovery rates. J. R. Stat. Soc. Series B Stat. Methodol. 64, 479–498.

Sudol, M., McDonald, C.B., and Farooq, A. (2012). Molecular insights into the WW domain of the Golabi-Ito-Hall syndrome protein PQBP1. FEBS Lett. 586, 2795–2799.

Sumanasekera, W.K., Tien, E.S., Davis, J.W., 2nd, Turpey, R., Perdew, G.H., and Vanden Heuvel, J.P. (2003a). Heat shock protein-90 (Hsp90) acts as a repressor of peroxisome proliferator-activated receptor-alpha (PPARalpha) and PPARbeta activity. Biochemistry 42, 10726–10735.

Sumanasekera, W.K., Tien, E.S., Turpey, R., Vanden Heuvel, J.P., and Perdew, G.H. (2003b). Evidence that peroxisome proliferator-activated receptor alpha is complexed with the 90-kDa heat shock protein and the hepatitis virus B X-associated protein 2. J. Biol. Chem. 278, 4467–4473.

Takahashi, K., Yoshina, S., Masashi, M., Ito, W., Inoue, T., Shiwaku, H., Arai, H., Mitani, S., and Okazawa, H. (2009). Nematode homologue of PQBP1, a mental retardation causative gene, is involved in lipid metabolism. PLoS ONE 4, e4104.

Wang, Q., Moore, M.J., Adelmant, G., Marto, J.A., and Silver, P.A. (2013). PQBP1, a factor linked to intellectual disability, affects alternative splicing associated with neurite outgrowth. Genes Dev. 27, 615–626.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42, D1001–D1006.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4, Article 17.

Zhang, B., and Zhu, J. (2013). Identification of key causal regulators in gene networks. In Proceedings of the World Congress on Engineering & Computer Science II, 1309–1312.

Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 153, 707–720.