# A Systems Biology Framework Identifies Molecular Underpinnings of Coronary Heart Disease

Tianxiao Huan, Bin Zhang, Zhi Wang, Roby Joehanes, Jun Zhu, Andrew D. Johnson, Saixia Ying, Peter J. Munson, Nalini Raghavachari, Richard Wang, Poching Liu, Paul Courchesne, Shih-Jen Hwang, Themistocles L. Assimes, Ruth McPherson, Nilesh J. Samani, Heribert Schunkert, Coronary ARteryDIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) Consortium, International Consortium for Blood Pressure GWAS (ICBP), Qingying Meng, Christine Suver, Christopher J. O'Donnell, Jonathan Derry, Xia Yang, Daniel Levy

***Objective***—Genetic approaches have identified numerous loci associated with coronary heart disease (CHD). The molecular mechanisms underlying CHD gene–disease associations, however, remain unclear. We hypothesized that genetic variants with both strong and subtle effects drive gene subnetworks that in turn affect CHD.

***Approach and Results***—We surveyed CHD-associated molecular interactions by constructing coexpression networks using whole blood gene expression profiles from 188 CHD cases and 188 age- and sex-matched controls. Twenty-four coexpression modules were identified, including 1 case-specific and 1 control-specific differential module (DM). The DMs were enriched for genes involved in B-cell activation, immune response, and ion transport. By integrating the DMs with gene expression–associated single-nucleotide polymorphisms and with results of genome-wide association studies of CHD and its risk factors, the control-specific DM was implicated as CHD causal based on its significant enrichment for both CHD and lipid expression–associated single-nucleotide polymorphisms. This causal DM was further integrated with tissue-specific Bayesian networks and protein–protein interaction networks to identify regulatory key driver genes. Multitissue key drivers (*SPIB* and *TNFRSF13C*) and tissue-specific key drivers (eg, *EBF1*) were identified.

***Conclusions***—Our network-driven integrative analysis not only identified CHD-related genes, but also defined network structure that sheds light on the molecular interactions of genes associated with CHD risk. (***Arterioscler Thromb Vasc Biol*. 2013;33:1427-1434.**)

**Key Words:** coexpression network ■ coronary heart disease ■ gene expression ■ systems biology



Atherosclerotic coronary heart disease (CHD) is a multifactorial disease with a prominent inflammatory vascular component that remains the leading cause of death in most developed countries and will soon become the leading cause of death in developing countries.[1] Atherosclerosis involves an interaction between modified lipoproteins, monocyte-derived macrophages, T cells, and the normal cellular elements of the arterial wall, resulting in the formation of atherosclerotic plaques. Plaque rupture and thrombosis can result in myocardial infarction (MI) and stroke.[2] Despite rapid advances in the genetic and genomic analysis of CHD in recent years, the molecular elements and the mechanisms involved in CHD pathogenesis remain unclear. Understanding the molecular basis of this disease can help identify biomarkers for more accurate clinical diagnosis and point to candidate drug targets for more effective therapy.

At the genetic level, recent genome-wide association studies (GWAS) have identified 27 loci associated with CHD and

Received on: October 4, 2012; final version accepted on: March 4, 2013.

From the National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA (T.H., R.J., A.D.J., P.C., S.-J.H., D.L.); the Center for Population Studies and the Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD (T.H., R.J., A.D.J., S.-J.H., C.J.O'D., D.L.); Department of Genetics and Genomic Sciences (B.Z., J.Z.), Institute of Genomics and Multiscale Biology (B.Z., J.Z.), and Graduate School of Biological Sciences (B.Z., J.Z.), Mount Sinai School of Medicine, New York, NY; Sage Bionetworks, Seattle, WA (Z.W., C.S., J.D.); School of Life Sciences, Center for Evolutionary Medicine and Informatics, the Biodesign Institute, Arizona State University, Tempe, AZ (Z.W.); Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Bethesda, MD (R.J., S.Y., P.J.M.); Genomics Core facility Genetics and Developmental Biology Center, National Heart, Lung, and Blood Institute, Bethesda, MD (N.R., R.W., P.L.); Department of Medicine, Stanford University School of Medicine, Stanford, CA (T.L.A.); Departments of Medicine and Biochemistry, University of Ottawa, Ottawa, ON (R.McP.); Department of Cardiovascular Sciences, University of Leicester, and National Institute for Health Research Leicester Cardiovascular Biomedical Research Unit, Leicester, United Kingdom (N.J.S.); Medizinische Klinik II, Universitätzu Lübeck, Lübeck, and Deutsches Zentrumfür Herz-Kreislauf-Forschung, Universitätzu Lübeck, Lübeck, Germany (H.S.); Department of Integrative Biology and Physiology, University of California, Los Angeles, CA (Q.M., X.Y.).

This manuscript was sent to Qingbo Xu, Consulting Editor, for review by expert referees, editorial decision, and final disposition.

**The online-only Data Supplement is available with this article at http://atvb.ahajournals.org/lookup/suppl/doi:10.1161/ATVBAHA.112.300112/-/DC1.**

Correspondence to Xia Yang, PhD, Department of Integrative Biology and Physiology, University of California, Los Angeles, CA 90095 (e-mail xyang123@ucla.edu); or Daniel Levy, MD, Framingham Heart Study, Center for Population Studies, National Heart, Lung, and Blood Institute, 73 Mt. Wayte Ave, Suite 2, Framingham, MA 01702 (e-mail Levyd@nih.gov).

MI, with many representing novel risk loci.[3,4] Identifying the casual genes and mechanisms underlying GWAS signals, however, remains a challenging task. In addition, even large-scale GWAS involving tens of thousands of cases and controls are unlikely to explain >15% to 20% of the heritability of CHD and MI.[5] Therefore, further experiments and integrative analyses are needed to elucidate functional mechanisms underlying CHD risk loci and to identify the missing heritability.

At the gene expression level, peripheral blood has been investigated widely because of its ease of collection and its pathological relevance to CHD, because multiple blood constituents seem to play a role in plaque accumulation, rupture, and thrombus formation.[6] Gene expression changes in whole blood of patients with CHD may reflect the interaction of genetic predisposition, disease activity, immunity, thrombosis, metabolism, and environmental modifiers underlying disease mechanisms. In recent studies,[7–9] mRNA profiling of tens to >100 CHD cases has yielded lists of differentially expressed genes in an effort to identify effective markers of disease severity. Most of these studies, however, identified relatively few significantly differentially expressed genes between cases and controls at a stringent statistical cutoff, and there was little overlap in the gene signatures of different CHD expression profiling studies. For instance, there is no common CHD gene signature among 160 genes reported by Sinnaeve et al,[7] 50 genes reported by Wingrove et al,[8] and 23 genes identified by Rosenberg et al.[9] This failure to replicate may be attributable to the small sample sizes, differences in phenotypes or experimental platforms, and the low levels of differential expression of single genes.

To address the limitations of previous studies, we designed a systems biology framework to explore the molecular underpinnings of CHD via integration of whole blood gene expression profiles of RNA collected from Framingham Heart Study (FHS) participants with network approaches, GWAS, and genetics of gene expression (Figure 1). The FHS case–control study includes 188 pairs of prevalent CHD cases and age- and sex-matched controls. In our companion article, we identified 35 differentially expressed genes between CHD cases and controls at false discovery rate<50%.[10] To extend on results of analysis of differential gene expression, in this investigation, we first constructed coexpression networks and compared the robust coexpression patterns of cases with controls to identify differential coexpression modules (differential modules [DMs]) that demonstrate gene coregulation in cases but not in controls or vice versa. The CHD DMs were then integrated with single-nucleotide polymorphisms (SNPs) associated with altered gene expression (expression-associated SNPs [eSNPs] or eQTLs) from CHD-relevant tissues as well as CHD and CHD risk factor GWAS association results to look for enrichment of CHD risk eSNPs among the DM genes. This integration allowed us to differentiate causal (ie, upstream) from reactive (ie, downstream) DMs. Last, using a key driver (KD) analysis that leverages DMs and graphical networks—including external tissue-specific Bayesian networks (BNs) and protein–protein interaction (PPI) networks—we identified putative regulators of the DMs and derived a CHD subnetwork.

## Materials and Methods

Materials and Methods are available in the online-only Supplement.

## Results

### Clinical Characteristics of the Study Sample

The case–control study included 188 CHD cases and 188 matched controls. Demographic data and the clinical and laboratory characteristics of the study participants are summarized in Table 1. CHD cases had a higher prevalence of hypertension than controls ($P$=0.019).

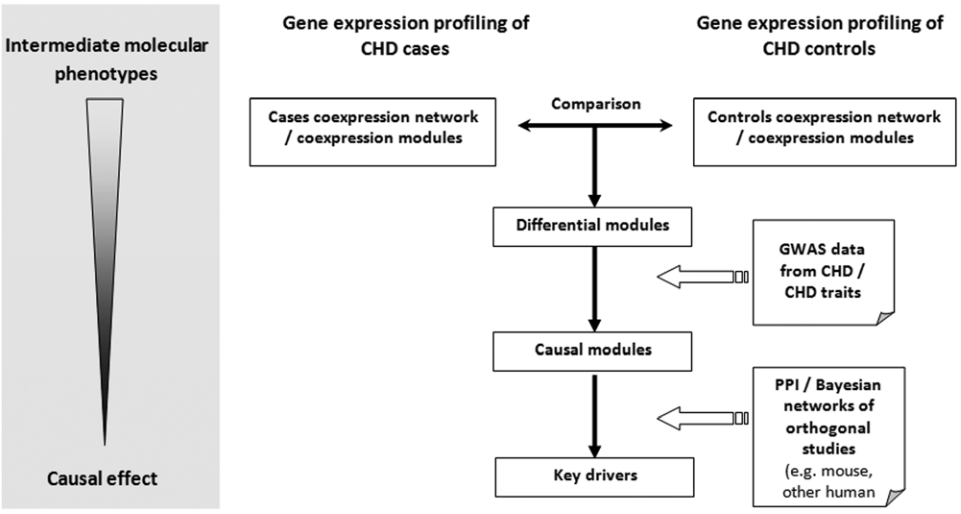


**Figure 1.** Systems biology analysis flow chart depicting the process of identifying coronary heart disease (CHD) causal modules and key drivers (KDs). The overall analysis includes 3 steps. First, coexpression networks are constructed from CHD cases and controls separately. CHD differential modules (DMs) are then identified by comparing the network structures between cases and controls. Step 2, the DMs are then integrated with genome-wide association studies (GWAS) of CHD and related traits using a single-nucleotide polymorphism set enrichment analysis to identify causal DMs. Third, key regulatory genes or KDs are identified for the causal DMs based on directional networks derived from independent studies using a KD analysis algorithm. PPI indicates protein–protein interaction.

**Table 1. Clinical Characteristics of the Case–Control Study Groups**

| | CHD Cases n=188 | CHD Controls n=188 |
|---|---|---|
| Age, y | 71±7.9 | 71±8.0 |
| Sex, male (female) | 140 (48) | 140 (48) |
| Smoking, n (%) | 15 (8) | 2 (1) |
| Alcohol, n (%) | 31 (16) | 34 (18) |
| Glucose, mg/dL | 115±33 | 111±26 |
| Diabetes mellitus, n (%) | 51 (27) | 37 (20) |
| Hypertension, n (%) | 158 (84) | 130 (69) |
| BMI, kg/m$^2$ | 28.9±4.7 | 28.7±4.8 |
| Systolic BP, mm Hg | 127±18 | 129±17 |
| Diastolic BP, mm Hg | 68±10 | 72±10 |
| Total cholesterol, mg/dL | 151±33 | 174±31 |
| Triglycerides, mg/dL | 126±85 | 116±70 |
| HDL cholesterol, mg/dL | 48±14 | 53±15 |
| Hypertension treatment, n (%) | 154 (82) | 109 (58) |
| Diabetes mellitus treatment, n (%) | 41 (22) | 35 (19) |
| Lipid treatment, n (%) | 175 (93) | 125 (66) |

Values are mean±SD, sample number or percentage. BMI indicates body mass index; BP, blood pressure; CHD, coronary heart disease; and HDL, high-density lipoprotein.

## Construction of Coexpression Networks for CHD and Identification of DMs

Using differential expression analysis, we identified 35 genes that were differentially expressed between CHD cases and controls at a false discovery rate <50%; these results are summarized in our companion report.[10] We hypothesized that alterations in the gene regulatory network architecture may better reflect the underlying molecular differences between groups. Therefore, we compared coexpression networks between CHD cases and controls to identify differential coexpression modules (DMs).

Using the previously developed Weighted Gene Coexpression Network Analysis method,[11] we constructed gene coexpression networks for CHD cases and their corresponding controls, separately. A total of 12 coexpression network modules were identified for CHD cases and 12 for controls (Figure III in the online-only Data Supplement). We tested whether modules were conserved between case and control networks by examining the overlap in gene memberships using Fisher exact test. Modules that overlapped at $P<2.0\times10^{-3}$ (corresponding to Bonferroni-corrected $P<0.05$) were defined as conserved between networks. If a module in one network did not have a corresponding conserved module in the other network, it was considered a DM. Two DMs—the CHD_tan module and the control_tan module containing 38 and 79 genes, respectively—differed between CHD cases and controls (Figure IV and Table III in the online-only Data Supplement). The genes in these 2 CHD DMs did not show any overlap with the 35 differential expression signatures,[10] suggesting that the altered gene coexpression pattern and differential expression of individual genes captured different biological signals.

## Functional Annotation of the DMs

To understand the biological pathways and functional categories of the DMs, we conducted functional enrichment analysis. As shown in Table 2, the top enriched gene ontology biology process term in the control_tan module is B-cell activation, with 9 of 51 module genes (existing in the gene ontology database) overlapping with 103 genes in this pathway (19.8-fold enrichment, Bonferroni corrected enrichment $P=5.6\times10^{-10}$). These B-cell–related genes demonstrated strong coexpression in controls, but not in CHD cases, indicating that B-cell dysregulation may be a major factor in atherosclerotic CHD. In contrast to the control_tan module, the CHD_tan module was identified in CHD cases but not in controls. The CHD_tan module is enriched for genes involved in ion transport (gene ontology category; Bonferroni-corrected enrichment $P=0.025$). These results suggest a reorganization and shift of immune response and ion transporter pathways, with the former disrupted and the later enhanced in CHD.

## DMs Are Enriched for CHD Risk eSNPs

By virtue of the case–control design of this study, blood samples from the CHD cases used for gene expression profiling

**Table 2. Gene Ontology Enrichment of CHD Differential Modules**

| Ontology Category | Overlap | Fold Enrichment | *P* Value | Corrected *P* Value* |
|---|---|---|---|---|
| Control_tan module | | | | |
| B-cell activation | 9 | 19.8 | $5.65\times10^{-10}$ | $5.65\times10^{-7}$ |
| B-cell differentiation | 7 | 23.6 | $1.53\times10^{-8}$ | $1.53\times10^{-5}$ |
| B-cell receptor signaling pathway | 6 | 22.6 | $2.30\times10^{-7}$ | $2.30\times10^{-4}$ |
| Immune response | 12 | 6.3 | $2.86\times10^{-6}$ | $2.86\times10^{-3}$ |
| Lipid transport | 5 | 9.6 | $1.66\times10^{-4}$ | 0.166 |
| Hemopoiesis | 9 | 4.2 | $2.23\times10^{-4}$ | 0.229 |
| CHD_tan module | | | | |
| Ion transport | 10 | 4.7 | $2.53\times10^{-5}$ | 0.025 |
| Metal ion transport | 7 | 6.0 | $1.21\times10^{-4}$ | 0.121 |
| Cell motility | 7 | 3.0 | 0.007 | 1 |

*Bonferroni correction for testing of multiple pathways and functional categories in the GO biology process and Kyoto Encyclopedia of Genes and Genomes pathway databases. CHD indicates coronary heart disease; and GO, gene ontology.

were drawn after the CHD events, thus some of the DMs we identified are potentially downstream of CHD rather than being causal. To identify causal DMs, we integrated the CHD DMs with CHD-related GWAS databases to look for enrichment of risk SNPs among DMs, with the rationale that genes harboring disease-risk SNPs are putatively causal for the disease of interest,[12,13] because genotype does not change in response to disease occurrence.

We used SNP set enrichment analysis (SSEA; see description in the Methods section) to test whether the DMs for CHD are likely to play causal roles in disease development by evaluating the overall association *P* value distribution of putative functional SNPs among DM genes in comparison with the null distribution. All 5 control CHD gene sets from different resources (details in Materials and Methods) were highly enriched with low-*P* value CHD association eSNPs in the Coronary ARteryDIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) CHD GWAS[3] (Table IV in the online-only Data Supplement). These results confirm the sensitivity of SSEA to identify gene sets suspected of being causal of CHD.

The control-specific DM, control_tan, was not only enriched for risk eSNPs for CHD in the CARDIoGRAM CHD GWAS (1.6-fold enrichment; $P$=0.04 and $1.5\times10^{-5}$ from Kolmogorov–Smirnov [KS] and Fisher tests, respectively), but also is enriched for low-*P* value associations for lipid traits, including high-density lipoprotein cholesterol (KS $P=4.8\times10^{-12}$ and Fisher $P=7.8\times10^{-9}$), low-density lipoprotein cholesterol (KS $P=9.0\times10^{-19}$ and Fisher $P=1.9\times10^{-26}$), and total cholesterol (KS $P=4.2\times10^{-21}$and Fisher $P=7.5\times10^{-28}$) in the Global Lipid Genetics Consortium lipid GWAS,[14] suggesting that this DM may play a role in CHD pathogenesis via effects on atherogenic lipids (Table IV in the online-only Data Supplement SSEA results for CHD-related gene sets, and Table V in the online-only Data Supplement CHD risk eSNPs for the control_tan module). The CHD_tan DM, on the contrary, did not show enrichment for CHD risk SNPs (enrichment $P$=0.49 and 0.85 from Fisher and KS tests, respectively), and thus may be reactive rather than causal. The differential gene expression signatures identified at false discovery rate<50% also failed to show any enrichment for CHD risk SNPs.

## Identification of Key Drivers of the CHD Causal DM and Their Associated Subnetworks via Network Models Derived From Orthogonal Studies

We speculated that the CHD causal DM we identified (control_tan) was driven by key regulatory genes with broad impact on this network module when perturbed. Such key regulatory genes or KDs can serve as candidate genes for therapeutic intervention. To identify KDs, we projected the genes within the CHD causal DM onto the precompiled directional networks, including tissue-specific BNs (Table II in the online-only Data Supplement)[15,16] and nondirectional PPI networks[17] from previous studies. At a Bonferroni-corrected $P$<0.05, we identified 59 unique KDs for the CHD causal module in the 6 tissue-specific BNs (59 KDs for blood, 0 for liver, 1 for fat, 1 for kidney, 0 for heart, and 0 for muscle) and 1 PPI network (0 KD). The human blood BN gave rise to the largest number of KDs (n=59) in contrast to the limited number of KDs in other tissue-specific networks. This is not surprising given that the DMs were identified from human blood expression profiles, confirming the tightly coregulated relationships among the DM genes in blood rather than in other tissues. Furthermore, as the DMs were from the FSH and the blood BN used for KD identification was constructed from an independent Icelandic cohort,[15] the consistency between the 2 networks cross-validated our results. Among the 59 KDs, 27 KDs contained eSNPs showing association with CHD or its risk factors at $P$<0.05; 7 contained eSNPs at $P<1.0\times10^{-3}$, and 3 contained eSNPS at $P<1.0\times10^{-4}$ in GWAS (Table VI in the online-only Data Supplement). Two of the KDs (*CD79B* and *SPIB*) overlapped with the 23 gene signature of CHD reported by Rosenberg et al[9] (enrichment $P=5.5\times10^{-5}$).

We further prioritized the 59 KDs based on 4 elements: (1) consistency of the KD across multiple directional networks, that is, a higher priority is given to KDs driving the CHD causal DM in >1 tissue-specific/PPI network; (2) whether the KD is a DM gene itself; (3) whether the KD contains eSNPs showing evidence of association with CHD and its risk factors at $P<1.0\times10^{-3}$ (the liberal GWAS *P* value cutoff was used here because the GWAS information was not considered alone but integrated with several levels of additional data to prioritize genes and reduce false discoveries); and (4) the *P* value of the KD based on the KD analysis algorithm (Table VI in the online-only Data Supplement). Among the top 20 KDs (Table 3), *TNFRSF13C* and *SPIB* appeared in 2 networks—*TNFRSF13C* in adipose and blood and *SPIB* in kidney and blood—and were considered as multitissue KDs; the remaining were blood-specific KDs. In fact, Sage et al[18,19] reported that in Ldlr knockout mice, deletion of the B-cell activating factor receptor (*BAFFR*, encoded by the multitissue KD *TNFRSF13C*) reduced aortic root atherosclerosis significantly. In addition, Fretz et al[20] observed that knockout of another top KD, *EBF1*, in mice induced defects in lipid metabolism, adipose tissue deposition, and impaired glucose mobilization.[20] Therefore, some of the top KDs we identified are supported by existing experimental evidence for their relevance to metabolism, atherosclerosis, or other CHD-related phenotypes.

As shown in Figure 2, these top KDs constitute a tightly linked subnetwork with certain elements being tissue-specific and others showing cross-tissue connections. The KDs are at the center of the subnetwork and hence represent master regulators of this subnetwork. Not surprisingly, this subnetwork, which includes 152 genes depicted in Figure 2, is significantly enriched for immune response (gene ontology functional category) and B-cell receptor signaling pathways (Bonferroni-corrected $P=1.9\times10^{-6}$ and $P=5.1\times10^{-6}$, respectively). In addition, the subnetwork is highly enriched for genes whose eSNPs show low-*P* value associations for both CHD in the CARDIoGRAM GWAS (enrichment $P=9.2\times10^{-3}$ and $3.8\times10^{-3}$ from Fisher and KS tests, respectively) and CHD risk traits, including systolic blood pressure and diastolic blood pressure in the International Consortium of Blood Pressure GWAS, and lipid traits (high-density lipoprotein, low-density lipoprotein, and total cholesterol) in the Global Lipid Genetics Consortium GWAS (Table IV in the online-only Data Supplement). Furthermore, 15 genes were candidate cardiovascular genes curated in the Genetic Association Database (http://geneticassociationdb.nih.gov/).

**Table 3. Top 20 Key Drivers Derived From the CHD Causal Module From Tissue-Specific Bayesian and Protein–Protein Interaction Networks**

| Rank | KD | Tissue | Is DM Gene | Adjusted KD *P* Value by Subnet Size | GWAS Source Where KD eSNPs Show Association at $P<1\times10^{-3}$ | Literature Evidence for CHD Association |
|---|---|---|---|---|---|---|
| 1 | SPIB | Blood/kidney | 0 | Blood $P=1.76\times10^{-25}$/kidney $P=0.045$ | | One of the CHD biomarkers of the 23 gene signatures[9] |
| 2 | TNFRSF13C | Blood/adipose | 0 | Blood $P=1.94\times10^{-24}$/adipose $P=0.041$ | | Deletion of BAFF receptor reduced atherosclerosis in Ldlr knockout mice[18,19] |
| 3 | PPAPDC1B | Blood | 1 | $2.1\times10^{-30}$ | ICBP_SBP | |
| 4 | EBF1 | Blood | 1 | $3.2\times10^{-18}$ | ICBP_DBP | EBF1$^{-/-}$ mice resulted in lipid abnormalities[20] |
| 5 | RALGPS2 | Blood | 1 | $1.8\times10^{-16}$ | GLGC_HDL | |
| 6 | CD72 | Blood | 1 | $4.6\times10^{-25}$ | | |
| 7 | RASGRP3 | Blood | 1 | $1.7\times10^{-22}$ | | |
| 8 | CD200 | Blood | 1 | $3.2\times10^{-22}$ | | |
| 9 | TNFRSF17 | Blood | 1 | $5.8\times10^{-22}$ | | |
| 10 | BACH2 | Blood | 1 | $1.0\times10^{-04}$ | | |
| 11 | COBLL1 | Blood | 0 | $1.2\times10^{-28}$ | GLGC_HDL, TC, TG | |
| 12 | CD79B | Blood | 0 | $1.6\times10^{-28}$ | GIANT_BMI | One of the CHD biomarkers of the 23 gene signatures[9] |
| 13 | SAV1 | Blood | 0 | $8.8\times10^{-22}$ | GLGC_TC | |
| 14 | FAM3C | Blood | 0 | $2.2\times10^{-16}$ | CARDIoGRAM | |
| 15 | LARGE | Blood | 0 | $9.3\times10^{-32}$ | | |
| 16 | PKIG | Blood | 0 | $2.6\times10^{-31}$ | | |
| 17 | TSPAN13 | Blood | 0 | $2.9\times10^{-30}$ | | |
| 18 | CELSR1 | Blood | 0 | $4.3\times10^{-30}$ | | |
| 19 | CTGF | Blood | 0 | $3.0\times10^{-29}$ | | Increased expression in animals and patients with atrial fibrillation[21] |
| 20 | BANK1 | Blood | 0 | $4.1\times10^{-29}$ | | |

The key drivers (KDs) were ranked based on 4 elements: (1) consistency of the KD across multiple directional networks; (2) whether the KD is a DM gene itself; (3) whether the KD contains eSNPs showing evidence of association with CHD and its risk factors at $P<1.0\times10^{-3}$; and (4) the *P* value of the KD based on the Key driver analysis algorithm. BAFF indicates B-cell activating factor; BMI, body mass index; CABG, coronary artery bypass grafting; CARDIoGRAM, Coronary ARteryDIsease Genome wide Replication and Meta-analysis; CHD, coronary heart disease; CVD, cardiovascular disease; DBP, diastolic blood pressure; DIAGRAM, DIAbetes Genetics Replication and Meta-analysis; eSNP, expression-associated single-nucleotide polymorphism; FHS, Framingham Heart Study; GIANT, Genetic Investigation of Anthropometric Traits; GLGC, Global Lipid Genetics Consortium; GO, gene ontology; GWAS, genome-wide association studies; HDL, high-density lipoprotein; ICBP, International Consortium of Blood Pressure; MI, myocardial infarction; PPI, protein–protein interaction; PTCA, percutaneous transluminal coronary angioplasty with or without a stent; SBP, systolic blood pressure; SSEA, SNP set enrichment analysis; TC, total cholesterol; TG, triglycerides; and WGCNA, Weighted Gene Coexpression Network Analysis.

All these lines of evidence support the relevance of the KDs and their related subnetwork to atherosclerotic CHD.

## Discussion

We conducted a systems biology investigation of gene expression in a case–control study of CHD using network approaches. To our knowledge, this is the largest gene expression study to date of CHD incorporating multidimensional genome-wide data from multiple sources, including GWAS, gene expression, eSNPs, and molecular networks.

Extending a traditional approach that targeted differentially expressed individual genes,[10] we compared the coexpression network structure in CHD cases with that in matched controls to derive differential network modules. The advantage of a network approach is that it provides a contextual framework to explain how genes interact with each other differently in CHD cases versus controls. Previous research has reported strong correlations of gene coexpression patterns with causal disease mechanisms. For example, a conserved coexpression module was discovered in liver and adipose tissues in mouse and human, were found to be correlated with multiple metabolic traits,[15,23] and its causal relationship with metabolic diseases was further validated experimentally.[23–25]

Using coexpression network analyses, we identified 2 DMs that demonstrated differential coexpression patterns between cases and controls. The 79 genes in the control_tan module (Table III in the online-only Data Supplement) showed a tightly coregulated pattern in controls, but a disrupted

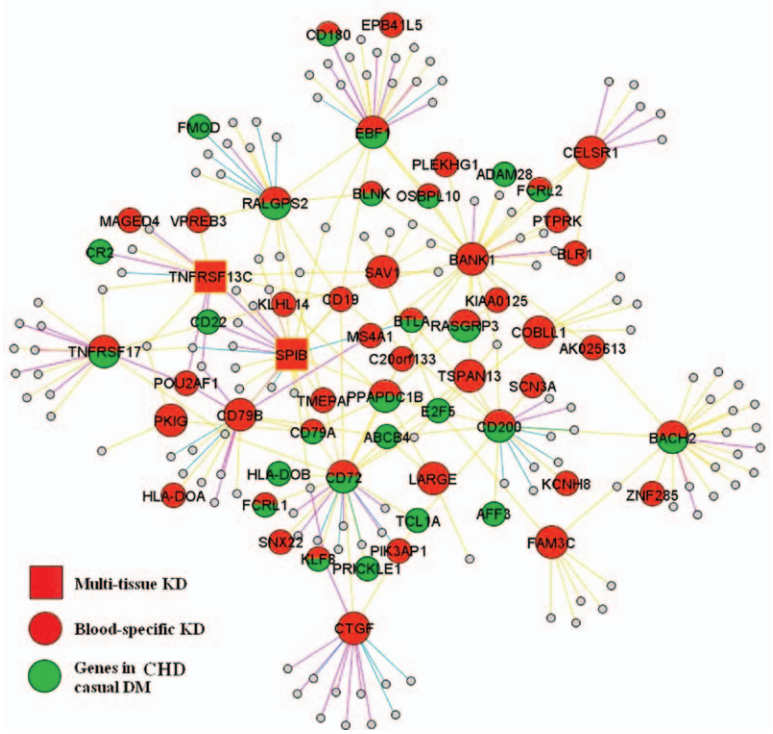


**Figure 2.** Top key drivers (KDs) of the coronary heart disease (CHD) causal module and the associated subnetwork. The subnetwork was derived from the top 20 KDs using the tissue-specific networks from which the KDs were identified. The nodes of the largest size in the network are the top 20 KDs, and the middle-sized nodes represent other KDs and genes in the CHD causal model. Red rectangular, red circular, and green circular nodes are multitissue KDs, blood-specific KDs, and causal differential module (DM) genes, respectively. The network graph was drawn by ProteoLens.[22]

coexpression pattern in CHD cases. B-cell–centered immune genes were highly enriched in this module, along with a suggestive enrichment for lipid transport genes (*ABCA6, ABCA9, ABCB4, APOD,* and *OSBPL10*). This is consistent with the known roles of immune response and lipid-related pathways in atherogenesis. At a cellular level, atherosclerosis involves lipid accumulation in the artery wall, which can lead to plaque rupture, platelet aggregation, and coagulation, leading to occlusive thrombus and MI. Immune responses are involved in multiple phases of atherogenesis.[6,26] From the inception of atherogenesis, multiple immunocytes, including macrophages, T cells, and mast cells, are recruited by adhesion molecules or leukocyte receptors expressed on the surface of endothelial cells. The precise role of B cells in atherosclerosis, however, is relatively less well studied. In the traditional view, B cells are thought to play a protective role against atherosclerosis.[27] Kyaw et al[28] found that innate-like B1 cells mediate protection from atherosclerosis via the secretion of natural IgM antibodies. The tightly coregulated pattern of B-cell–related genes in controls and the disrupted pattern in CHD cases found in our study are consistent with an atheroprotective role of B cells. A recent study, however, showed reduced atherosclerosis in mice after B-cell depletion,[29] suggesting that B-cell subpopulations may have contrasting roles in the pathogenesis of atherosclerosis.[30]

It is of note that the blood samples used in this study were collected after the onset of clinical CHD events. Therefore, the differential gene signatures and DMs identified here could be a consequence of CHD or its treatment (ie, downstream signals) rather being causal of CHD (ie, upstream signals). GWAS has identified scores of loci associated with CHD and CHD risk factors, with ≈27 loci for CHD or MI,[3,4] 30 for blood pressure,[31–33] 95 for lipids,[14] 28 for body mass index,[34] and 27 for type 2 diabetes mellitus.[35] Traditional GWAS, however, can only detect common loci yielding stringent *P* value associations, typically $P<5\times10^{-8}$, and susceptibility loci with more subtle effects are missed. In an attempt to separate the causal and reactive effects of the DMs that differed between CHD cases and controls, we used SSEA to integrate our gene expression findings with publicly available GWAS results for CHD and its risk factors based on the reasoning that genetic variation is unaltered over time and that converging genetic associations with CHD gene expression signatures will highlight causal genes and pathways. SSEA is a complementary approach to single-point analysis, because it tests the overall association of a set of SNPs derived from a given gene set to a trait rather than a single SNP or gene association with a trait,[12] thus empowering the identification of biological pathways, functional categories, or networks. While individual genes may only exert subtle effects on a trait, collectively they may contribute to disease susceptibility on a larger scale. Another advantage of SSEA lies in the fact that it uses eSNPs (SNPs associated with the expression levels of genes) from disease-relevant tissues to map genes to their functional SNPs. This is in contrast to distance-based gene-to-SNP mapping used in other studies.[36–38]

The control_tan DM (enriched for genes involved in B-cell activation/differentiation and immune function), but not the CHD_tan DM (enriched for ion transport genes), was identified as a causal module showing enrichment for eSNPs with low-*P* value associations from CHD GWAS in CARDIoGRAM[3] and lipid traits GWAS in Global Lipid Genetics Consortium.[14] These results indicate that B-cell–related immune function may exert a causal effect on CHD and on lipids. As discussed above, previous studies support a causal role of B cells in atherogenesis. In addition, recent research by Fretz et al[20]

revealed direct evidence that B-cell dysfunction (Ebf1$^{-/-}$ mice) results in lipid abnormalities. Therefore, our results not only support previous phenotypic observations linking B cells to both atherogenesis and lipid dysregulation, but also provide network and genetic support at the molecular level.

We further identified and prioritized putative key regulatory genes or KDs for the CHD causal DM based on the global networks derived from tissue-specific BNs and PPI networks. We identified 59 KDs for the control_tan module, 2 from multiple tissue networks and 57 from the blood-specific network. After prioritization of the 59 KDs, we retrieved a subnetwork derived from the top KD and found that the KDs and other control_tan DM genes form a tightly linked subnetwork. Perturbations of the multitissue KD, *TNFRSF13C*, and a tissue-specific KD, *EBF1*, in animal models have been shown to affect aortic root atherosclerosis and other CHD-related phenotypes.[18,20] Therefore, the subnetwork structure and the KDs we uncovered provide both known CHD candidate genes with experimental support and novel candidate genes that warrant further validation.

In summary, we systematically compared the gene expression features of 188 CHD cases and 188 matched controls from the FHS via network modeling. Orthogonal studies incorporating CHD and CHD risk factor GWAS and data-driven gene regulatory networks from human and mouse models were integrated in this analysis to infer casual effects and possible pathological molecular mechanism. We found B-cell–centered immune function to be related to CHD pathogenesis. We hypothesize that disruption of gene coregulation of a B-cell–related DM via KDs contributes to CHD susceptibility. Our results provide further evidence supporting a critical role of adaptive immunity in atherosclerosis, especially aspects involving B-cell function. Consequently, we speculate that immunomodulatory therapy may be useful to prevent or treat atherosclerotic CHD.[39]



## Acknowledgments

D. Levy and X. Yang designed, directed, and supervised the experiment. D. Levy was responsible for funding of the project. T. Huan, X. Yang, and D. Levy drafted the article. P. Courchesne organized the experiment material and data exchange. All authors participated in revising and editing the articles. All authors have read and approved the final version of the article. We thank Xavier Schildwachter, Jeanette Erdmann, and Christina Willenborg for technical assistance.

## Sources of Funding

The study was supported by the Intramural Research Program of National Heart, Lung, and Blood Institute (D. Levy, Principal Investigator). The Framingham Heart Study is supported by National Institutes of Health Contract N01-HC-25195. N.J. Samani holds a Chair funded by the British Heart Foundation and is a National Institute for Health Research Senior Investigator.

## Disclosures

None.

## References

1. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095–2128.
2. Glass CK, Witztum JL. Atherosclerosis. the road ahead. *Cell*. 2001;104:503–516.
3. Schunkert H, König IR, Kathiresan S, et al; Cardiogenics; CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333–338.
4. Lotta LA, Peyvandi F. Addressing the complexity of cardiovascular disease by design. *Lancet*. 2011;377:356–358.
5. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet*. 2010;42:570–575.
6. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. *Nature*. 2011;473:317–325.
7. Sinnaeve PR, Donahue MP, Grass P, Seo D, Vonderscher J, Chibout SD, Kraus WE, Sketch M Jr, Nelson C, Ginsburg GS, Goldschmidt-Clermont PJ, Granger CB. Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. *PLoS One*. 2009;4:e7037.
8. Wingrove JA, Daniels SE, Sehnert AJ, Tingley W, Elashoff MR, Rosenberg S, Buellesfeld L, Grube E, Newby LK, Ginsburg GS, Kraus WE. Correlation of peripheral-blood gene expression with the extent of coronary artery stenosis. *Circ Cardiovasc Genet*. 2008;1:31–38.
9. Rosenberg S, Elashoff MR, Beineke P, et al; PREDICT (Personalized Risk Evaluation and Diagnosis in the Coronary Tree) Investigators. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. *Ann Intern Med*. 2010;153:425–434.
10. Joehanes R, Ying S, Huan T, Johnson AD, Raghavachari N, Wang R, Liu P, Woodhouse KA, Tanriverdi K, Courchesne P, Freedman JE, O'Donnell CJ, Levy D, Munson PJ. Gene expression signatures of coronary heart disease. *Arterioscler Thromb Vasc Biol*. March 28, 2013. DOI: 10.1161/ATVBAHA.112.301169. http://atvb.ahajournals.org/content/early/2013/03/28/ATVBAHA.112.301169.long. Accessed April 2, 2013.
11. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
12. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet*. 2010;86:581–591.
13. Zhong H, Beaulaurier J, Lum PY, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet*. 2010;6:e1000932.
14. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707–713.
15. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. *Nature*. 2008;452:423–428.
16. Derry JM, Zhong H, Molony C, et al. Identification of genes and networks driving cardiovascular and metabolic phenotypes in a mouse F2 intercross. *PLoS One*. 2010;5:e14319.
17. Mishra GR, Suresh M, Kumaran K, et al. Human protein reference database–2006 update. *Nucleic Acids Res*. 2006;34:D411–D414.
18. Sage AP, Harrison JE, Baker LA, Masters L, Tsiantoulas D, Binder CJ, Mallat Z. Depletion of mature B cells by deletion of BAFF receptor reduces atherosclerosis in ldlr knockout mice. *Circulation*. 2011; 124:A9224
19. Sage AP, Tsiantoulas D, Baker L, Harrison J, Masters L, Murphy D, Loinard C, Binder CJ, Mallat Z. BAFF receptor deficiency reduces the development of atherosclerosis in mice–brief report. *Arterioscler Thromb Vasc Biol*. 2012;32:1573–1576.
20. Fretz JA, Nelson T, Xi Y, Adams DJ, Rosen CJ, Horowitz MC. Altered metabolism and lipodystrophy in the early B-cell factor 1-deficient mouse. *Endocrinology*. 2010;151:1611–1621.
21. Nalls MA, Couper DJ, Tanaka T, et al. Multiple loci are associated with white blood cell phenotypes. *PLoS Genet*. 2011;7:e1002113.
22. Huan T, Sivachenko AY, Harrison SH, Chen JY. ProteoLens: a visual analytic tool for multi-scale database-driven biological network data mining. *BMC Bioinformatics*. 2008;9:S5.
23. Chen Y, Zhu J, Lum PY, et al. Variations in DNA elucidate molecular networks that cause disease. *Nature*. 2008;452:429–435.
24. Yang X, Deignan JL, Qi H, et al. Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nat Genet*. 2009;41:415–423.
25. Yang X, Peterson L, Thieringer R, et al. Identification and validation of genes affecting aortic lesions in mice. *J Clin Invest*. 2010;120:2414–2422.
26. Hansson GK, Libby P. The immune response in atherosclerosis: a double-edged sword. *Nat Rev Immunol*. 2006;6:508–519.
27. Caligiuri G, Nicoletti A, Poirier B, Hansson GK. Protective immunity against atherosclerosis carried by B cells of hypercholesterolemic mice. *J Clin Invest*. 2002;109:745–753.

28. Kyaw T, Tay C, Krishnamurthi S, Kanellakis P, Agrotis A, Tipping P, Bobik A, Toh BH. B1a B lymphocytes are atheroprotective by secreting natural IgM that increases IgM deposits and reduces necrotic cores in atherosclerotic lesions. *Circ Res*. 2011;109:830–840.
29. Ait-Oufella H, Herbin O, Bouaziz JD, Binder CJ, Uyttenhove C, Laurans L, Taleb S, Van Vré E, Esposito B, Vilar J, Sirvent J, Van Snick J, Tedgui A, Tedder TF, Mallat Z. B cell depletion reduces the development of atherosclerosis in mice. *J Exp Med*. 2010;207:1579–1587.
30. Perry HM, McNamara CA. Refining the role of B cells in atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2012;32:1548–1549.
31. Newton-Cheh C, Johnson T, Gateva V, et al; Wellcome Trust Case Control Consortium. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet*. 2009;41:666–676.
32. Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, Huang H, Zhou J, Lashley K, Chen Y, Christman M, Rotimi C. A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet*. 2009;5:e1000564.
33. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009;41:677–687.
34. Speliotes EK, Willer CJ, Berndt SI, et al; MAGIC; Procardis Consortium. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010;42:937–948.
35. Voight BF, Scott LJ, Steinthorsdottir V, et al; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*. 2010;42:579–589.
36. Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D; DIAGRAM Consortium; MAGIC investigators. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet*. 2010;6:e1001058.
37. Weng L, Macciardi F, Subramanian A, Guffanti G, Potkin SG, Yu Z, Xie X. SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinformatics*. 2011;12:99.
38. Holden M, Deng S, Wojnowski L, Kulle B. GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics*. 2008;24:2784–2785.
39. Lahoute C, Herbin O, Mallat Z, Tedgui A. Adaptive immunity in atherosclerosis: mechanisms and future therapeutic targets. *Nat Rev Cardiol*. 2011;8:348–358.


## Significance

Atherosclerotic coronary heart disease (CHD) as a multifactorial disease is influenced by both genetic and environmental factors. This study used a systems biology framework to identify genes and networks associated with CHD via integration of whole blood gene expression profiles with network approaches, genome-wide association studies, and genetics of gene expression. This study demonstrated a CHD differential coexpression module (differential module), involving B-cell–centered immune functions with a disrupted coexpression pattern in CHD cases but not in matched controls. This differential module was implicated as causal based on its significant enrichment for both CHD and lipid expression–associated single-nucleotide polymorphisms. This causal differential module was further integrated with networks to identify regulatory key driver genes and their driving subnetwork structure. These findings provided further evidence of a critical role of adaptive immunity in atherosclerosis, especially aspects involving B-cell function, and may help to identify new drug targets for the treatment or prevention of CHD.