# UCLA UCLA Previously Published Works

# Title

Shared Molecular Pathways and Gene Networks for Cardiovascular Disease and Type 2 Diabetes Mellitus in Women Across Diverse Ethnicities

**Permalink** https://escholarship.org/uc/item/64w0d9vv

**Journal** Circulation Genomic and Precision Medicine, 7(6)

**ISSN** 1942-325X

# Authors

Chan, Kei Hang K Huang, Yen-Tsung Meng, Qingying <u>et al.</u>

**Publication Date** 

2014-12-01

# DOI

10.1161/circgenetics.114.000676

Peer reviewed

# Shared Molecular Pathways and Gene Networks for Cardiovascular Disease and Type 2 Diabetes Mellitus in Women Across Diverse Ethnicities

Kei Hang K. Chan, PhD; Yen-Tsung Huang, MD, ScD; Qingying Meng, PhD; Chunyuan Wu, MS; Alexander Reiner, MD; Eric M. Sobel, PhD; Lesley Tinker, PhD; Aldons J. Lusis, PhD; Xia Yang, PhD; Simin Liu, MD, ScD

- *Background*—Although cardiovascular disease (CVD) and type 2 diabetes mellitus (T2D) share many common risk factors, potential molecular mechanisms that may also be shared for these 2 disorders remain unknown.
- *Methods and Results*—Using an integrative pathway and network analysis, we performed genome-wide association studies in 8155 blacks, 3494 Hispanic American, and 3697 Caucasian American women who participated in the national Women's Health Initiative single-nucleotide polymorphism (SNP) Health Association Resource and the Genomics and Randomized Trials Network. Eight top pathways and gene networks related to cardiomyopathy, calcium signaling, axon guidance, cell adhesion, and extracellular matrix seemed to be commonly shared between CVD and T2D across all 3 ethnic groups. We also identified ethnicity-specific pathways, such as cell cycle (specific for Hispanic American and Caucasian American) and tight junction (CVD and combined CVD and T2D in Hispanic American). In network analysis of gene–gene or protein–protein interactions, we identified key drivers that included *COL1A1, COL3A1*, and *ELN* in the shared pathways for both CVD and T2D. These key driver genes were cross-validated in multiple mouse models of diabetes mellitus and atherosclerosis.
- *Conclusions*—Our integrative analysis of American women of 3 ethnicities identified multiple shared biological pathways and key regulatory genes for the development of CVD and T2D. These prospective findings also support the notion that ethnicity-specific susceptibility genes and process are involved in the pathogenesis of CVD and T2D. (*Circ Cardiovasc Genet.* 2014;7:911-919.)

Key Words: cardiovascular diseases ■ diabetes mellitus ■ ethnology ■ genetics ■ genome-wide association study ■ women

It has long been known that cardiovascular disease (CVD) and type 2 diabetes mellitus (T2D) share many common risk factors and pathophysiological intermediaries including obesity, dyslipidemia, insulin resistance, and proinflammatory and prothrombotic states.<sup>1–3</sup> However, the key molecular drivers underlying these highly correlated phenotypes as well as the potential regulatory networks shared in the pathogenesis of CVD and T2D remain poorly understood.<sup>4</sup>

#### **Clinical Perspective on p 919**

Of the  $\approx 60$  genetic loci identified for CVD and T2D in large-scale genome-wide association studies (GWAS),<sup>5</sup> only 2 significant loci (*TCF7L2* and *VEGFA*) are shared between the 2 diseases. There are  $\geq 3$  fundamentally important questions that remain to be answered. First, are there additional genetic risks, either alone or concentrated in specific regulatory networks that may explain the pathophysiological connections between CVD and T2D? Second, are there any ethnicityspecific genetic mechanisms for these 2 common vascular diseases, given that there are drastic ethnicity-specific differences in their risk and linkage disequilibrium patterns.<sup>6-8</sup> Third, although common genetic loci have been detected across different populations (eg, the *Chr9p21* locus for CVD and the *TCF7L2* locus for T2D) supporting the presence of common pathological paths across ethnicity, to what degree molecular mechanisms are shared across ethnicities?

To answer these 3 questions, we performed 2 GWAS for both CVD and T2D using an integrative pathway and network analysis in the national Women's Health Initiative SNP Health Association Resource (WHI-SHARe) and the Genomics and

The Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.114.000676/-/DC1.

*Circ Cardiovasc Genet* is available at http://circgenetics.ahajournals.org

Received March 18, 2014; accepted September 23, 2014.

From the Department of Epidemiology (K.H.K.C., Y.-T.H., S.L.) and Division of Endocrinology, Department of Medicine (S.L.), Warren Alpert Medical School of Brown University, Providence, RI; Department of Integrative Biology and Physiology (K.H.K.C., Q.M., X.Y.), Department of Human Genetics (E.M.S.), Department of Medicine/Division of Cardiology, David Geffen School of Medicine (A.J.L.), and Departments of Medicine and Obstetrics and Gynecology, David Geffen School of Medicine (S.L.), University of California Los Angeles; Biostatistics Division (C.W.), Public Health Sciences Division (L.T.), Fred Hutchinson Cancer Research Center, Seattle, WA; and Department of Epidemiology, University of Washington, Seattle (A.R.).

Correspondence to Xia Yang, PhD, Department of Integrative Biology and Physiology, University of California Los Angeles, Los Angeles, CA 90095. E-mail xyang123@ucla.edu or Simin Liu, MD, ScD, Department of Epidemiology, Brown University, Providence, RI 02912. E-mail simin\_liu@brown.edu © 2014 American Heart Association, Inc.

Randomized Trials Network (WHI-GARNET). These cohorts provide unique opportunities to examine both CVD and T2D, alone or in combination, across multiple ethnicities to allow interdisease and interethnicity comparisons.

#### Methods

#### **Study Participants**

A detailed description of study participants of both WHI-SHARe and WHI-GARNET is given in the Data Supplement and Table I in the Data Supplement. In brief, the WHI-SHARe included 8155 blacks and 3494 Hispanic American (HA) women. The WHI-GARNET involved 3697 Caucasian American (CA) women. The research protocol was approved by the institutional review board and that all human participants gave written informed consent.

#### **Definition of Clinical End Points**

In WHI-SHARe, incident cases of CVD were classified based on any event of myocardial infarction, stroke, deep vein thrombosis, and pulmonary embolism during follow-up. Incident cases of T2D were identified on the basis of those clinical cases that had no history of T2D at baseline and diagnosed during the follow-up period. Those women in the cohort who were free of T2D or CVD were used as controls. In WHI-GARNET, CVD cases were identified during the Hormone Therapy (HT) trial based on clinical diagnosis of acute myocardial infarction that required overnight hospitalization, silent myocardial infarction determined from serial electrocardiograms obtains every 3 years, or death because of coronary heart disease. Cases of T2D were also identified during the HT trial. Controls were free of coronary heart disease, stroke, venous thromboembolism, and T2D by the end of the HT trial.

#### **Genetic Data**

Genome-wide genotyping of the WHI-SHARe participants was performed using the Affymetrix 6.0 array (Affymetrix, Inc, Santa Clara, CA). Genotyping for WHI-GARNET participants was performed using the Illumina HumanOmnil-Quad SNP platform (Illumina, Inc, San Diego, CA). Details about genotyping methods and quality control are given in the Data Supplement.

#### **Standard SNP Analysis**

We performed standard SNP association analysis for 3 end points, that is, CVD, T2D, and combined CVD+T2D adjusting for principal components for global ancestry and matching factors (detailed in the Data Supplement). Demographic and lifestyle factors do not influence germline genetic variants and as such were not treated as confounders and were not adjusted for in these models. Given that WHI-SHARe included 8155 blacks and 3494 HA, statistical power seems excellent (>80%) to detect an odds ratio of 1.25 for minor allele frequency >0.25 in blacks and an odds ratio of 1.5 for minor allele frequency >0.13 in HA. Power estimate among 3697 WHI-GARNET CA is almost identical to that in HA.

## Pathway and Network-Based Integrative Analysis

Accumulating evidence supports that multiple genes involved in biological pathways or gene networks, rather than individual isolated genes, coordinate together to contribute to disease risks.<sup>9-14</sup> To uncover the hidden mechanisms that are not obvious from the individual top GWAS hits alone, we augmented the standard GWAS analysis with pathway and network approaches. Functionally related genes involved in metabolic and signaling pathways were obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome. We tested each pathway for enrichment of genetic signals for CVD and T2D, alone or in combination, using 5 well-established methodologies (Meta-Analysis Gene-set Enrichment of VariaNT Associations [MAGENTA],<sup>12</sup> gene set analysis-SNP [GSA-SNP],<sup>11</sup> Network Interface Miner for Multigenic Interactions [NIMMI],<sup>9</sup> Pathway and Network-Oriented GWAS Analysis [PANOGA],<sup>10,15</sup> and expression SNP [eSNP]<sup>14</sup>; detailed in the Data Supplement and Table II in the Data Supplement) that investigate whether functionally related genes are enriched for both strong and subtle genetic risks (ie, not limited to top, genome-wide significant loci) of diseases. We chose to use multiple methodologies to avoid potential bias from any particular method. We ran the 5 methodologies separately and a statistical cutoff of false-discovery rate<sup>16</sup> <5% (implemented in MAGENTA, GSA-SNP, NIMMI, and eSNP approaches) or Bonferroni-corrected P<0.05 (implemented in PANOGA) as provided by each method was considered significant. Pathways that showed significance in ≥2 methods were chosen as the top pathways to report.

### Identification of Key Regulatory Genes for the Disease-Associated Pathways Using Gene Regulatory Networks and Protein–Protein Interaction Networks

As hundreds of genes are involved in the biological pathways, we seek to identify important regulators of the top significant pathways as a means to prioritize genes and uncover novel regulatory mechanisms. We integrated the 169 shared genes involved in the top 8 pathways with graphical networks (Bayesian networks and protein–protein interaction<sup>17</sup>; sources detailed in Table III in the Data Supplement) to identify key regulators of the 169 shared genes using a key driver (KD) analysis method.<sup>18,19</sup> KD analysis takes a set of genes (G) and a gene network N as input. For every node K in N, the subnetwork (N<sub>k</sub> of K) was determined by 3-edge expansion and then tested for enrichment of genes in G using Fisher exact test. Nodes whose neighborhood subnetwork shows significant enrichment at Bonferronicorrected *P*<0.05 were termed KDs. As multiple KD lists were generated using multiple networks, we ranked the KDs using a normalized rank score to summarize the consistency and strength, where

 $NRS = \frac{C_{KD}}{N_{BN} + PPI} \times \sum_{i=1}^{C_{KD}} Rank_{KDi}; C_{KD}$  is the count of network models from which a KD was identified among all networks used including 8 BNs from the 8 tissue types (ie, adipose, liver, blood, heart, brain, islet, kidney, and muscle) and protein–protein interaction;  $C_{KD}$ is then normalized by the total number of networks  $(N_{BN+PPI})$  to represent the consistency of a KD across all networks tested; the KD strength is represented by summing the normalized statistical rank in each network  $i(Rank_{KDi})$  across all networks from which the KD is identified;  $Rank_{KDi}$  was calculated by dividing the rank of a KD based on the *P* values of the Fisher exact test in descending order divided by

pathway genes and high consistency across networks tested. To cross-validate the top KD genes from top disease pathways identified, we searched for multiple mouse databases that include (1) genes tested causal for CVD and T2D phenotypes,<sup>20</sup> (2) the phenotypic changes in genetically modified mouse models with individual genes perturbed,<sup>21</sup> and (3) genes identified for CVD and T2D phenotypes in the hybrid mouse diversity panel (>100 strains of inbred or recombinant inbred mouse).<sup>22</sup>

the total number of KDs identified from a network *i*. KDs with high

normalized rank score were those with high network enrichment for

#### **Results**

The descriptive statistics on demographics and lifestyle factors of each study population are shown in Table 1. The blacks, HA, and CA women in WHI-SHARe and WHI-GARNET differed significantly in age, body mass index, current smoking, alcohol drinking, hormone usage, physical activity, and family history of T2D (P<0.001).

### Identification of Significant Genetic Loci Using Standard GWAS Analysis

Four genomic loci reached genome-wide significance (P < 5e - 8) in the standard GWAS analysis, including

Characteristic	Blacks in WHI-SHARe	HA in WHI-SHARe	CA in WHI-GARNET
n	8155	3494	3697
Age (mean±SD), y	61.6±7.03	60.3±6.69	65.7±6.90
BMI (mean±SD)	31.0±6.37	28.9±5.60	29.7±6.13
Current smoking, %	11.7	6.95	11.1
Current alcohol drinking, %	54.4	66.8	73.3
Current hormone user, %	25.5	35.4	7.87
Physical activity: total metabolic equivalents (METS)/wk (mean±SD)	9.67±12.7	10.8±13.8	10.2±12.7
T2D case, %	19.4	17.8	32.4
Family history of T2D, %	51.3	45.0	36.6
CVD case, %	7.59	4.33	20.4

Table 1. Baseline Characteristics" of Participants in WHI-SHARE and WHI-GARNET Stratified by Ethr
---

BMI indicates body mass index; CA, Caucasian American; CVD, cardiovascular disease; HA, Hispanic American; T2D, type 2 diabetes mellitus; WHI-GARNET, Women's Health Initiative-Genomics and Randomized Trials Network; and WHI-SHARe, Women's Health Initiative SNP Health Association Resource.

\*Significant difference (P<0.001) seen between the 3 populations for every above variable. P value is calculated using ANOVA test for continuous variables (ie, age, BMI, and physical activity) and  $\chi^2$  test for categorical variables (ie, current smoking, current alcohol drinking, current hormone user, family history of T2D, CVD case in WHI-SHARe or CVD case in WHI-GARNET, and T2D case).

rs11885576 (*KLHL29* at 2p24) for CVD in CA, rs2805429 (*RYR2* at 1q43) for T2D in blacks, and rs17591786 (*FLJ45721* at 4p25) and rs7825609 (*NAT2* at 8p22) for combined CVD+T2D in CA (summarized in Table 2). In addition, several previously established loci including *Chr9p21*, *TCF7L2*, and *CDKAL1* for CVD or T2D reached P < 5e - 3 in our study (summarized in Table 3; Table IV in the Data Supplement).

### Identification of Biological Pathways Using Integrative Pathway and Network Analysis

Thirty six of the 1501 pathways from the KEGG and Reactome databases were found to be associated with  $\geq 1$  of the 3 end points (ie, CVD, T2D, and combined CVD+T2D) in  $\geq 1$  ethnic group (Tables V and VI in the Data supplement). Of these 36 pathways identified, 8 including focal adhesion, hypertrophic cardiomyopathy (HCM), extracellular matrix–receptor interaction signaling, dilated cardiomyopathy, arrhythmogenic right ventricular cardiomyopathy (ARVC), calcium signaling, axon guidance, and cell adhesion molecules were commonly enriched for genetic signals for all 3 end points (CVD, T2D, and combined CVD+T2D) across all 3 ethnic groups. These pathways were also significantly enriched in the C4D and CARDIOGRAM GWAS (Table VII in the Data Supplement).

There were 638 unique genes involved in these common pathways (Figure 1; genes listed in Table V in the Data Supplement). Forty-five of these genes have been implicated previously in CVD (24 genes)<sup>24-26</sup> and T2D (21 genes).<sup>14,27,28</sup> Five of the 8 commonly enriched pathways, namely, HCM, dilated cardiomyopathy, ARVC, focal adhesion, and extracellular matrix–receptor interaction signaling, were also found to be highly interconnected as demonstrated by a shared common set of 117 genes among them (Figure 1).

Besides the shared pathways across diseases and ethnicities, we also identified disease- and ethnicity-specific pathways (Table VIII in the Data Supplement). By disease, the apoptosis pathway was significantly associated with CVD+T2D, but not T2D; acute amyloid leukemia was associated with T2D, but not for CVD. By ethnicity, the cell cycle pathway was significant for all 3 diseases for HA and CA, but not blacks: the WNT signaling pathway was significant for HA; pathways in cancer was not significant for CA; and dorsoventral axis formation and prion diseases were only specific to blacks. Certain pathways demonstrated both disease and ethnicity specificity. For instance, the adipocytokine signaling pathway was only significant for T2D in CA; the prostate cancer and melanogenesis pathways were only significant for T2D or CVD+T2D in HA or blacks, but not for CVD in any of the 3 ethnicities.

Table 2. Genome-Wide Significant SNPs for CVD, T2D, and Combined CVD+T2D in Women's Health Initiative Women

End Point	Population	Chromosomal Region	Top SNP in Region*	Position Hg18†	Candidate Gene	Minor/Major Allele‡	Minor Allele Frequency	Odds Ratio (95% Cl)	<i>P</i> Value
CVD	CA	2p24	rs11885576	23526223	KLHL29	G/ <u>T</u>	0.04	0.43 (0.32–0.58)	3.5e–8
T2D	Blacks	1q43	rs2805429	235750901	RYR2	<u>G</u> /C	0.47	1.23 (1.15–1.33)	4.0e-8
CVD+T2D	CA	4p15	rs17591786	26828037	FLJ45721	G/ <u>A</u>	0.40	1.29 (1.18–1.41)	1.8e-8
CVD+T2D	CA	8p22	rs7825609	18290501	NAT2	C/T	0.01	0.35 (0.25–0.49)	6.0e-10

CA indicates Caucasian American; CVD, cardiovascular disease; and T2D, type 2 diabetes mellitus.

\*The top SNP with the smallest *P* values among the genotyped SNP for each locus. These are novel associations (SNPs not found in the Catalog of published genomewide association studies, however, the *KLHL29* and *NAT2* loci have been reported before).

†Positions of the SNPs were derived from dbSNP build 136.

‡The coded allele used to calculate the effect size was underlined.

Disease	Genes	Region	Reference SNP (RS) ID	Observed <i>P</i> Value in WHI			Association With CVD+T2D in WHI		
				Blacks	HA	CA	Blacks	HA	CA
T2D	CDKAL1	6p22.3	rs10946398	0.1	0.0002*	0.09	0.34	0.0006*	0.13
T2D	CDKAL1	6p22.3	rs7754840	0.08	0.0002*	0.10	0.28	0.0006*	0.13
T2D	CDKAL1	6p22.3	rs7756992	0.59	0.0001*	0.04	0.75	0.0001*	0.05
T2D	TCF7L2	10q25.2	rs4506565	0.001*	0.63	0.04	0.08	0.94	0.08
T2D	TCF7L2	10q25.2	rs7901695	0.002*	0.47	0.02	0.14	0.82	0.06
CVD	Intergenic	9p21.3	rs1333042	0.64	0.46	0.001*	0.54	0.99	0.0003*
CVD	Intergenic	9p21.3	rs4977574	0.84	0.28	0.005*	0.01	0.74	0.002*

CA indicates Caucasian American; CVD, cardiovascular disease; HA, Hispanic American; T2D, type 2 diabetes mellitus; and WHI, Women's Health Initiative. \**P* values <5×10<sup>-3</sup>.

## Identification and Validation of Putative Key Regulatory Genes for the Shared Pathways Across Diseases and Ethnicities

To identify potential KD genes among the significant pathways shared between diseases and ethnicities, we integrated the pathway genes with 9 different regulatory or interaction networks that capture gene-gene or protein-protein interactions. These KD genes represent central network genes which, when perturbed, can potentially affect a large number of genes involved in the CVD and T2D pathways and thus exert stronger impact on diseases. The 10 top KD genes included COLIAI, COL3AI, ELN, COL4AI, CD93, FN1, MMP2, SPARC, COL2A1, and THBS2 in multiple networks (Figure 2; Table IX in the Data Supplement). These KD genes were also confirmed in multiple mouse data sets that documented their modulating impact on risk of T2D and CVD (Table IX in the Data Supplement). For example, the gene expression levels or the SNPs regulating the expression levels of the COL4A1 gene were tested causal for 14 CVD and T2D traits in 4 different tissues in 7 mouse F2 cross data sets and a mouse data set comprised >100 inbred or recombinant inbred strains.

Interestingly, the KDs themselves are not among the GWAS hits from the current and previous GWAS for CVD and T2D, although the genes within the pathways that these KDs seem to regulate are enriched for disease risk SNPs. We speculate that genetic polymorphisms that strongly perturb KDs may impose evolutionary constraints, which may explain the lack of strong GWAS hits in the KDs, whereas subtle genetic polymorphisms that affect KDs may still be enriched for disease risks. To this end, we analyzed the risk enrichment for the top 10, 30, and 100 KDs, respectively. Our results indeed indicated that the top KDs, especially the top 30 and 100 KDs, were significantly enriched for genetic risks of CVD, T2D, and combined CVD+T2D (Table X in the Data Supplement).

#### Discussion

In this genome-wide assessment of 8155 blacks, 3494 HA, and 3697 CA women who participated in the WHI-SHARe and the WHI-GARNET, we identified 4 independent genetic loci and 36 pathways to be significantly associated with CVD, T2D, and CVD+T2D in one or more ethnicities. Among the significant signals, the *FLJ45721* and *NAT2* loci were associated with the combined CVD+T2D end point in CA and 8 pathways were

consistently associated with both types of vascular diseases across all 3 ethnicities. These results suggest the presence of core mechanisms underlying both CVD and T2D. Ethnicityand disease-specific pathways were also identified. We further uncovered potential novel regulators of these shared pathways supporting their pleiotropic and causal impact on CVD and T2D.

Our standard GWAS analysis of 3 ethnic populations identified several biologically plausible disease loci including 2 previously implicated loci (KLHL29 and NAT2) for cardiometabolic diseases and 2 novel loci (RYR2 and FLJ45721). The KLHL29 locus was found to be associated with CVD in CA in our study and it was also previously implicated in CVD in blacks<sup>29</sup> and obesity in HA<sup>30</sup>. These lines of evidence support its importance for multiple cardiometabolic diseases. The KLHL29 locus is highly complex and seems to encode multiple proteins containing BTB and kelch motifs but with poorly annotated functions. NAT2 was significant for the joint CVD+T2D end point in our study and was previously detected as a GWAS signal for several important cardiometabolic traits including total cholesterol, triglyceride,<sup>31</sup> and insulin sensitivity (GENESIS consortium, personal communication) that are relevant for both CVD and T2D. NAT2 (N-acetyltransferase 2) is a well-known pharmacogenetic gene responsible for O- and N-acetylation of arylamine and hydrazine drugs and carcinogens but the mechanisms linking NAT2 to cardiometabolic traits are unknown. Along with NAT2, the FLJ45721 locus was a novel signal for the combined CVD+T2D end point in our study. However, there is currently limited knowledge about this locus and the candidate genes in this region. An additional novel locus, RYR2, was found to be associated with T2D in blacks in our analysis. RYR2 encodes ryanodine receptor 2, a calcium channel. As calcium is critical for insulin secretion and sensitivity,<sup>32</sup> RYR2 may contribute to T2D by affecting insulin levels and activities. Indeed, a recent study supported a role of *RYR2* in islet  $\beta$ -cell function, insulin secretion, and glucose tolerance, all key processes in T2D.33

Our findings also add to the body of literatures linking the involvement of multiple regulatory gene networks in the pathogenesis of complex cardiometabolic diseases, although individual genes may only exert subtle effects.<sup>14,24,34</sup> Our integrative pathway-based analysis revealed 8 consistent pathways between CVD and T2D across the 3 ethnic groups. Four pathways, calcium signaling, axon guidance, focal adhesion,



**Figure 1.** Network of 8 top pathways enriched for cardiovascular disease (CVD), type 2 diabetes mellitus (T2D), and combined CVD+T2D among blacks, Hispanic American, and Caucasian American Women's Health Initiative women. The color codes are: salmon, hypertrophic cardiomyopathy (HCM); green, dilated cardiomyopathy (DC); yellow, arrhythmogenic right ventricular cardiomyopathy (ARVC); light blue, calcium signaling pathway (Ca+); orange, axon guidance (axon); magenta, cell adhesion molecules (CAMs); brown, focal adhesion (FA); and purple, extracellular matrix –receptor interaction (ECM). The diamond nodes represent pathway and the round nodes denote gene, and the edge shows the interaction, that is, the association between a gene and a pathway. Genes involved in  $\geq$ 2 pathways were put with larger font, label and the nodes are in light green. The figure was created using cytoscape.<sup>23</sup>

and extracellular matrix–receptor interaction signaling, have been implicated previously in CAD and T2D.<sup>14,24</sup> Although the axon guidance pathway is mainly involved in localization and neuronal extension during embryogenesis,<sup>35</sup> genes within the axon guidance pathway have been connected to both CVD and T2D. For instance, a family of secreted proteins known as repulsive axon guidance cues (SLIT) and roundabout axon guidance receptors in the pathway have been found to reduce cytokine and thapsigargin-induced cell death under hyperglycemic conditions. Particularly, SLIT also triggered a release of endoplasmic reticulum luminal Ca<sup>2+</sup>, which suggested a molecular mechanism that defends  $\beta$  cells from endoplasmic reticulum stress–induced apoptosis. Therefore, local SLIT secretion may play a role in the survival and function of pancreatic  $\beta$  cells. Because of the fact that T2D results from a deficiency in functional  $\beta$ -cell mass, the axon



**Figure 2.** Network key drivers and gene subnetworks of the top 8 cardiovascular disease (CVD)/type 2 diabetes mellitus (T2D) pathways. Top 10 ranked multitissue key drivers (bigger nodes in yellow) of the top 8 CVD/T2D pathways in the protein–protein interaction network (edge color, green) and Bayesian network of adipose (orange), liver (yellow), blood (red), heart (brown), brain (blue), islet (pink), kidney (purple), and muscle a (light blue) tissues. The genes within the top 8 CVD/T2D pathways were highlighted in pink.

guidance pathway especially SLIT may contribute to therapeutic approaches for improving  $\beta$ -cell survival and function.<sup>36</sup> However, netrins, a family of proteins involved in axon guidance during embryogenesis, were found to be involved in angiogenesis and ischemia–reperfusion injury.<sup>37</sup> Four additional pathways, including HCM, dilated cardiomyopathy, ARVC, and cell adhesion molecules, were not reported previously. The HCM pathway involves genes that increase the calcium sensitivity of cardiac myofilaments leading to imbalance in energy supply and demand in the heart under severe stress, which may contribute to the development of CVD.<sup>38</sup> Calcium sensitivity is also important for T2D as discussed earlier. The dilated cardiomyopathy pathway involves genes, when altered that pose defect residing within the cytoskeleton or sarcomere, within the mitochrondria that causes deficient energy generation, or in the calcium cycling resulting in inefficient force activation and insulin secretion, which are processes related to CVD and T2D.<sup>39</sup> Genes that involve in the ARVC pathway includes *RYR2*, which is involved in calcium and insulin activities as discussed above, processes important for both cardiac function and insulin activities. A common mechanism among HCM pathway, dilated cardiomyopathy pathway, and ARVC pathway seem to be calcium homeostasis and sensitivity. The cell adhesion molecules are glycoproteins expressed on the cell surface and play an important role in a wide range of biological processes that includes homeostasis, immune response, and inflammation. Soluble intercellular adhesion molecules and vascular cell adhesion molecules have been associated with the development of coronary heart disease in the Health Professional Follow-up Study.<sup>40</sup> Higher levels of intercellular adhesion molecule-1 were also consistently associated with increased T2D risk in the WHI-Observational Study.<sup>3</sup> Therefore, these pathways appear to link to CVD and T2D via diverse mechanisms. The fact that these pathways were consistently identified across multiple ethnicities in our study highlights their central role in the joint mechanisms between CVD and T2D.

Importantly, the significant pathways were found to be highly connected through a large number of shared genes involved in extracellular matrix (collagens and laminins), cytoskeleton (actins), cell adhesion (integrins), calcium channels, and adenylate cyclases. In our further investigation of these genes using KD analysis network approach, the top 10 KDs were found to be expressed in almost all tissues or cell types involved in CVD and T2D including islet, liver, adipose, muscle, and kidney in our mouse tissue-specific data (details shown in Table VI in the Data Supplement). Perturbations of these genes and pathways that are critical for cell integrity and cellular communications in multiple tissues will likely affect vascular functions and subsequently CVD and T2D.

In addition to these shared mechanisms, we also identified disease- and ethnicity-specific pathways. For instance, several cancer-related pathways including melanoma, bladder cancer, and pathways in cancer were found to be specific for HA. However, the robustness of these signals awaits further validation in independent, ethnic-specific cohorts.

In the current study, only knowledge-driven biological pathways from existing pathway databases were used for disease risk signal enrichment analysis and we did not include data-driven networks from protein-protein interaction experiments or large-scale genomic data sets. Although data-driven networks may represent a more unbiased source to uncover previously unknown functional pathways, we focused on knowledge-driven pathways for the following reasons. First, these pathways represent a straightforward means to clearly define functionally related gene sets. In contrast, implementing data-driven networks requires more sophisticated considerations such as how to handle species, tissue, and sex specificity, how to clearly define gene sets of reasonable size based on large networks, and how to deal with network inconsistencies between data sets. Second, the results from canonical pathways are easily interpretable as they are largely derived from experimentally tested biochemical reactions, signaling cascades, and functional categories. In contrast, interpreting the results from the data-driven networks requires extra steps of annotation. Some of the gene subnetworks cannot be easily annotated by known knowledge, further complicating the result interpretation. Third, limiting the analysis to canonical pathways reduces the number of gene sets tested and thus helps reduce statistical penalty from multiple testing. Nonetheless, we acknowledge the power of data-driven approaches to detect novel insights, as demonstrated by our recent comprehensive investigation of coronary artery disease where a sophisticated analytic pipeline was used to include both knowledge-driven and data-driven approaches.<sup>34</sup> We will further pursue additional novel biological insights in the WHI cohorts using data-driven approaches in the future.

As hundreds of genes are likely involved in the core pathways identified, it is important to prioritize on KD genes, which, when perturbed, should have major impact on the pathways and hence the eventual vascular outcome of interest. Indeed, multiple KD genes, such as COL4A1, CD93, MMP2, and SPARC, were found in our network analysis reflecting gene-gene regulatory relations or protein-protein interactions. Interestingly, the KD genes were not identified in standard GWAS analysis of single SNPs, suggesting that important regulatory genes may not harbor common susceptibility polymorphisms because of evolutionary constraints.<sup>41</sup> Our further investigations of the top 10 KD genes yielded convincing evidence in support of the notion that perturbations of these KD genes in multiple mouse studies affect both CVD and T2D phenotypes. Their actions seem to be important in multiple tissues and consistent across multiple mouse studies of diabetes mellitus and CVD. In addition, these genes or their proteins have also been associated with obesity,<sup>30</sup> diabetes mellitus,42 CAD,43 and CVD44 traits in the literature. Of note, all of the top 10 KDs either encode extracellular matrix proteins or are involved in cell-matrix interactions, which places extracellular matrix at the central intersection of CVD and T2D pathogenesis.

The observations from us and others that important regulators are rarely GWAS hits and that GWAS candidate genes are mostly peripheral nodes in gene networks support that GWAS SNPs may serve as subtle modifiers of disease predisposition during a life period. Such subtle effects may help explain (1) their low selection pressure and hence commonality in the general population and (2) each GWAS locus only explains a small fraction of genetic heritability of complex diseases. These lines of available evidence suggest that the GWAS candidates may not serve as the best candidates for therapeutic interventions, although we acknowledge their importance in informing the biological pathways and processes involved in complex diseases and the possibility that rare mutations with strong effects in these genes may exist. However, KD or regulatory genes, although lack genetic variations that can be detected through GWAS, have hub properties in the networks and may behave like master switches that exert strong effects on disease networks and therefore may be better candidates from a therapeutic perspective.

Our study has several unique features. First, the comparison across multiple ethnicities allowed detection of both robust, shared mechanisms across populations, and potential ethnicity-specific signals. Second, apart from studying CVD and T2D separately, we also treated CVD and T2D together as a combined end point to increase the sensitivity to capture shared risk and pathology. Indeed, we identified genome-wide significant loci as well as pathways to be significant only for the combined end point. Finally, using a systems biology framework that integrates GWAS, pathways, gene expression, networks, and phenotypic information from both human and mouse populations, we were able to derive novel mechanistic insights and identify potential therapeutic targets.

Through a multiethnic GWAS augmented with comprehensive screening of ≈1500 curated biological pathways to capture processes that are genetically related to CAD and T2D, we identified many of the previously implicated biological pathways as well as novel genetic loci and pathways. We regard the identification of many suspected signals as encouraging and confirmatory. Importantly, our analyses imply causal involvement of the identified pathways or processes using genetics as the anchor, which represents an important step forward. Such causal inference is generally not possible with classic epidemiological studies of biomarkers, which may reflect consequences of disease rather than causal mechanisms underlying diseases. Furthermore, to our knowledge, this is the first time that these processes are found to be genetically linked to both CVD and T2D in multiple ethnicities, which points to shared mechanisms of vascular diseases that can be targeted for future therapeutic interventions. Moreover, we further explored the gene-gene interactions as well as the potential regulatory mechanisms to better understand the relationships between genes within and between pathways. The network topology revealed and the potential novel regulators identified provide deeper insights into the close connections, coordinated actions, and regulation of the pathways. The novel regulators identified may serve as more effective drug targets because of their central role in the regulatory networks. We think that these progresses made through our study are important for not only improving our understanding of the causal disease mechanisms but also for future development of more effective therapies.

In conclusion, our integrative analysis of American women of 3 ethnicities identified multiple shared biological pathways and key regulatory genes for the development of CVD and T2D. These prospective findings also support the notion that ethnicity-specific susceptibility genes and process are involved in the complex pathogenesis of CVD and T2D.

#### Acknowledgments

We thank Yi-Hsiang Hsu, ScD, Yiqing Song, MD, ScD, and Andrea Hevener, PhD, for reviewing the draft of the article; Ville-Petteri Mäkinen, Dsc, for his help with the eSNP methodology, particularly for generating the SNP set enrichment analysis analysis pipeline; and Calvin Pan for his help with putting together the mouse in silico validation data sets. We thank the Women's Health Initiative (WHI) investigators and staff for their dedication and the study participants for making the program possible. A listing of WHI investigators can be found at http://www.whiscience.org/publications/WHI\_investigators\_shortlist\_2010-2015.pdf.

#### Sources of Funding

The Women's Health Initiative (WHI) program is funded by the National Heart, Lung, and Blood Institute; National Institutes of Health; and the United States Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. Dr Yang is funded by the American Heart Association and the Leducq Foundation.

#### None.

## Disclosures

#### References

 Stern MP. Do non-insulin-dependent diabetes mellitus and cardiovascular disease share common antecedents? Ann Intern Med. 1996;124(1 Pt 2):110–116.

- Liu S, Tinker L, Song Y, Rifai N, Bonds DE, Cook NR, et al. A prospective study of inflammatory cytokines and diabetes mellitus in a multiethnic cohort of postmenopausal women. *Arch Intern Med.* 2007;167:1676–1685.
- Song Y, Manson JE, Tinker L, Rifai N, Cook NR, Hu FB, et al. Circulating levels of endothelial adhesion molecules and risk of diabetes in an ethnically diverse cohort of women. *Diabetes*. 2007;56:1898–1904.
- Manolio TA. Cohort studies and the genetics of complex disease. Nat Genet. 2009;41:5–6.
- Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009;106:9362–9367.
- Mosca L, Barrett-Connor E, Wenger NK. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. *Circulation*. 2011;124:2145–2154.
- Ding EL, Song Y, Malik VS, Liu S. Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and metaanalysis. JAMA. 2006;295:1288–1299.
- Rasmussen-Torvik LJ, Guo X, Bowden DW, Bertoni AG, Sale MM, Yao J, et al. Fasting glucose GWAS candidate region analysis across ethnic groups in the Multiethnic Study of Atherosclerosis (MESA). *Genet Epidemiol.* 2012;36:384–391.
- Akula N, Baranova A, Seto D, Solka J, Nalls MA, Singleton A, et al; Bipolar Disorder Genome Study (BiGS) Consortium; Wellcome Trust Case-Control Consortium. A network-based approach to prioritize results from genome-wide association studies. *PLoS One*. 2011;6:e24220.
- Bakir-Gungor B, Sezerman OU. A new methodology to associate SNPs with human diseases according to their pathway related context. *PLoS One.* 2011;6:e26277.
- Nam D, Kim J, Kim SY, Kim S. GSA-SNP: a general approach for gene set analysis of polymorphisms. *Nucleic Acids Res.* 2010;38(Web Server issue):W749–W754.
- 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.
- Wang K, Li M, Hakonarson H. Analysing biological pathways in genomewide association studies. *Nat Rev Genet*. 2010;11:843–854.
- 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.
- Bakir-Gungor B, Egemen E and Sezerman OU. PANOGA: a web server for identification of SNP-targeted pathways from genome-wide association study data. *Bioinformatics*. 2014;30:1287–9.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol*. 1995;57:289–300.
- Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, et al. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* 2003;13:2363–2371.
- Wang IM, Zhang B, Yang X, Zhu J, Stepaniants S, Zhang C, et al. Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers. *Mol Syst Biol.* 2012;8:594.
- Yang X, Zhang B, Molony C, Chudin E, Hao K, Zhu J, et al. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. *Genome Res.* 2010;20:1020–1036.
- Yang X, Peterson L, Thieringer R, Deignan JL, Wang X, Zhu J, et al. Identification and validation of genes affecting aortic lesions in mice. *J Clin Invest.* 2010;120:2414–2422.
- Maddatu TP, Grubb SC, Bult CJ, Bogue MA. Mouse Phenome Database (MPD). Nucleic Acids Res. 2012;40(Database issue): D887–D894.
- Parks BW, Nam E, Org E, Kostem E, Norheim F, Hui ST, et al. Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab.* 2013;17:141–152.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–2504.
- de las Fuentes L, Yang W, Dávila-Román VG, Gu C. Pathway-based genome-wide association analysis of coronary heart disease identifies biologically important gene sets. *Eur J Hum Genet*. 2012;20:1168–1173.
- Yang X. Use of functional genomics to identify candidate genes underlying human genetic association studies of vascular diseases. *Arterioscler Thromb Vasc Biol.* 2012;32:216–222.

- Gottesman O, Drill E, Lotay V, Bottinger E, Peter I. Can genetic pleiotropy replicate common clinical constellations of cardiovascular disease and risk? *PLoS One*. 2012;7:e46419.
- Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* 2011;21:1109–1121.
- Perry JR, McCarthy MI, Hattersley AT, Zeggini E, Weedon MN, Frayling TM; Wellcome Trust Case Control Consortium. Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. *Diabetes*. 2009;58:1463–1467.
- Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, et al. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet*. 2011;7:e1001300.
- Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One*. 2012;7:e51954.
- Suhre K, Wallaschofski H, Raffler J, Friedrich N, Haring R, Michael K, et al. A genome-wide association study of metabolic traits in human urine. *Nat Genet.* 2011;43:565–569.
- 32. Ma B, Lawson AB, Liese AD, Bell RA, Mayer-Davis EJ. Dairy, magnesium, and calcium intake in relation to insulin sensitivity: approaches to modeling a dose-dependent association. *Am J Epidemiol.* 2006;164:449–458.
- Dixit SS, Wang T, Manzano EJ, Yoo S, Lee J, Chiang DY, et al. Effects of CaMKII-mediated phosphorylation of ryanodine receptor type 2 on islet calcium handling, insulin secretion, and glucose tolerance. *PLoS One*. 2013;8:e58655.
- 34. Mäkinen VP, Civelek M, Meng Q, Zhang B, Zhu J, Levian C, et al; Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Consortium. Integrative genomics reveals novel

molecular pathways and gene networks for coronary artery disease. *PLoS Genet*. 2014;10:e1004502.

- 35. Tang H, Wei P, Duell EJ, Risch HA, Olson SH, Bueno-de-Mesquita HB, et al. Axonal guidance signaling pathway interacting with smoking in modifying the risk of pancreatic cancer: a gene- and pathway-based interaction analysis of GWAS data. *Carcinogenesis*. 2014;35:1039–1045.
- Yang YH, Manning Fox JE, Zhang KL, MacDonald PE, Johnson JD. Intraislet SLIT-ROBO signaling is required for beta-cell survival and potentiates insulin secretion. *Proc Natl Acad Sci USA*. 2013;110:16480–16485.
- Bongo JB, Peng DQ. The neuroimmune guidance cue netrin-1: a new therapeutic target in cardiovascular disease. J Cardiol. 2014;63:95–98.
- Sorajja P, Elliott PM, McKenna WJ. The molecular genetics of hypertrophic cardiomyopathy: prognostic implications. *Europace*. 2000;2:4–14.
- Osterziel KJ, Perrot A. Dilated cardiomyopathy: more genes means more phenotypes. *Eur Heart J.* 2005;26:751–754.
- Shai I, Pischon T, Hu FB, Ascherio A, Rifai N, Rimm EB. Soluble intercellular adhesion molecules, soluble vascular cell adhesion molecules, and risk of coronary heart disease. *Obesity (Silver Spring)*. 2006;14:2099–2106.
- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci USA*. 2007;104:8685–8690.
- Kos K, Wilding JP. SPARC: a key player in the pathologies associated with obesity and diabetes. *Nat Rev Endocrinol*. 2010;6:225–235.
- 43. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al; Cardiogenics; CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333–338.
- 44. Yamada Y, Kato K, Oguri M, Fujimaki T, Yokoi K, Matsuo H, et al. Genetic risk for myocardial infarction determined by polymorphisms of candidate genes in a Japanese population. *J Med Genet*. 2008;45:216–221.

## **CLINICAL PERSPECTIVE**

Cardiovascular disease (CVD) and type 2 diabetes mellitus (T2D) are highly heritable, share many common risk factors, and demonstrate ethnic-specific prevalence, yet a comprehensive molecular-level understanding of these observations is currently lacking. In this study, we seek to explore 3 clinically relevant questions: (1) whether there are additional genetic risks on top of the 60 identified genetic loci for CVD and T2D that may explain the pathophysiological link between CVD and T2D; (2) whether there are any ethnicity-specific genetic mechanisms for the 2 diseases; and (3) to what extent molecular mechanisms are shared across ethnicities for CVD and T2D. Using integrative pathway and network approaches, we conducted genome-wide association studies for both CVD and T2D in 3 ethnic populations, blacks, Caucasian Americans, and Hispanic Americans, in the national Women's Health Initiative. We identified 8 pathways and gene networks related to cardiomyopathy, calcium signaling, axon guidance, cell adhesion, and extracellular matrix that seemed to be commonly shared between CVD and T2D across all 3 ethnic groups. Potential key drivers of these shared pathways, such as *COL1A1*, *COL3A1*, and *ELN*, were also unraveled and cross-validated. We also identified ethnicity-specific pathways such as cell cycle (specific for Hispanic Americans and Caucasian Americans) and tight junction (specific for Hispanic Americans). These findings not only suggest the existence of major mechanistic pathways and key regulatory genes underlying the development of both CVD and T2D but also support the notion that ethnicity-specific mechanisms play a role in the complex pathogenesis of CVD and T2D.