

Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States

Le Shu, Chan Kei Hang K., Guanglin Zhang, Tianxiao Huan, Zeyneb Kurt, Yuqi Zhao, Veronica Codoni, David-Alexandre Tregouet, Jun Yang, James G. Wilson, et al.

► **To cite this version:**

Le Shu, Chan Kei Hang K., Guanglin Zhang, Tianxiao Huan, Zeyneb Kurt, et al.. Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. PLoS Genetics, Public Library of Science, In press, <10.1371/journal.pgen.1007040>. <hal-01598144>

HAL Id: hal-01598144

<https://hal.sorbonne-universite.fr/hal-01598144>

Submitted on 29 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Shared Genetic Regulatory Networks for Cardiovascular Disease and Type 2 Diabetes in Multiple**
2 **Populations of Diverse Ethnicities in the United States**

3 Le Shu,¹ [¶] Kei Hang K. Chan,^{2,3} [¶] Guanglin Zhang,¹ Tianxiao Huan,⁴ Zeyneb Kurt,¹ Yuqi Zhao,¹
4 Veronica Codoni,^{5,6} David-Alexandre Trégouët,^{5,6} Cardiogenics Consortium, Jun Yang,^{7,8} James G.
5 Wilson,⁹ Xi Luo,¹⁰ Daniel Levy,⁴ Aldons J. Lusis,¹¹ Simin Liu,^{2,12*} Xia Yang,^{1,13,14*}

6 ¹Department of Integrative Biology and Physiology, University of California, Los Angeles, CA 90095,
7 USA

8 ²Departments of Epidemiology and Medicine and Center for Global Cardiometabolic Health, Brown
9 University, Providence, RI 02903, USA

10 ³Hong Kong Institute of Diabetes and Obesity, Department of Medicine and Therapeutics, The Chinese
11 University of Hong Kong, Hong Kong SAR, China

12 ⁴The Framingham Heart Study, Framingham, MA and the Population Sciences Branch, National Heart,
13 Lung, and Blood Institute, Bethesda, MD 30105, USA

14 ⁵Sorbonne Universités, UPMC Univ. Paris 06, INSERM, UMR_S 1166, Team Genomics &
15 Pathophysiology of Cardiovascular Diseases, Paris 75013, France

16 ⁶ICAN Institute for Cardiometabolism and Nutrition, Paris 75013, France

17 ⁷Department of Public Health, Hangzhou Normal University School of Medicine, 16 Xuelin Street,
18 Hangzhou 310016, China

19 ⁸Collaborative Innovation Center for the Diagnosis and Treatment of Infectious Diseases, The First
20 Affiliated Hospital, Zhejiang University, Hangzhou 310003, China

21 ⁹Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS
22 39216, USA

23 ¹⁰Department of Biostatistics, Brown University, Providence, RI 02903, USA

Mis en forme : Espace Après : 0 pt, Interligne : simple

Mis en forme : Police :Times, 11 pt

Mis en forme : Police :12 pt

24 ¹¹Departments of Medicine, Human Genetics, and Microbiology, Immunology, and Molecular Genetics,
25 David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

26 ¹²Department of Endocrinology, Guangdong General Hospital/Guangdong Academy of Medical Sciences,
27 Guangzhou 510080, Guangdong, China

28 ¹³[Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, CA](#)
29 [90095, USA](#)

30 ¹⁴[Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA](#)

Mis en forme : Exosant

Mis en forme : Police :Times

Mis en forme : Police :Times, Exosant

Mis en forme : Police :Times

31
32 **Short Title:** Shared Gene Networks and Regulators ~~for~~ CVD and T2D

33 [†] These authors contributed equally to this work.

34 ***Co-corresponding author for correspondence:**

35 Xia Yang Email: xyang123@ucla.edu

36 Address: Department of Integrative Biology and Physiology

37 University of California, Los Angeles

38 Terasaki Life Sciences Building 2000D

39 Los Angeles, CA 90095

40

41 Simin Liu Email: simin_liu@brown.edu

42 Center for Global Cardiometabolic Health and Department of Epidemiology

43 Brown University

44 121 South Main St.

45 Providence, RI 02912

46 **Abstract**

47 Cardiovascular diseases (CVD) and type 2 diabetes (T2D) are closely interrelated complex diseases likely
48 sharing overlapping pathogenesis driven by aberrant activities in gene networks. However, the molecular
49 circuitries underlying the pathogenic commonalities remain poorly understood. We sought to identify the
50 shared gene networks and their key intervening drivers for both CVD and T2D by conducting a
51 comprehensive integrative analysis driven by five multi-ethnic genome-wide association studies (GWAS)
52 for CVD and T2D, expression quantitative trait loci (eQTLs), ENCODE, and tissue-specific gene network
53 models (both co-expression and graphical models) from CVD and T2D relevant tissues. We identified
54 pathways regulating the metabolism of lipids, glucose, and branched-chain amino acids, along with those
55 governing oxidation, extracellular matrix, immune response, and neuronal system as shared pathogenic
56 processes for both diseases. Further, we uncovered 15 key drivers including *HMGCR*, *CAVI*, *IGF1* and
57 *PCOLCE*, whose network neighbors collectively account for approximately 35% of known GWAS hits
58 for CVD and 22% for T2D. Finally, we cross-validated the regulatory role of the top key drivers using *in*
59 *vitro* siRNA knockdown, *in vivo* gene knockout, and two Hybrid Mouse Diversity Panels each comprised
60 of >100 strains. Findings from this in-depth assessment of genetic and functional data from multiple
61 human cohorts provide strong support that common sets of tissue-specific molecular networks drive the
62 pathogenesis of both CVD and T2D across ethnicities and help prioritize new therapeutic avenues for
63 both CVD and T2D.

64

65 **Author summary**

66 Cardiovascular disease (CVD) and type 2 diabetes (T2D) are two tightly interrelated diseases that are
67 leading epidemics and causes of deaths around the world, ~~with T2D increasing the risk of CVD.~~
68 Elucidating the mechanistic connections between the two diseases will offer critical insights for the
69 development of novel therapeutic avenues to target both simultaneously. Because of the challenging

70 complexity of CVD and T2D, involving numerous risk factors, multiple tissues, and multidimensional
71 molecular alterations, few have attempted such an investigation. We herein report a comprehensive and
72 in-depth data-driven assessment of the shared mechanisms between CVD and T2D by integrating
73 genomics data from diverse human populations including African Americans, Caucasian Americans, and
74 Hispanic Americans with tissue-specific functional genomics information. We identified shared pathways
75 and gene networks informed by CVD and T2D genetic risks across populations, confirming the
76 importance of well-established processes, as well as unraveling previously under-appreciated processes
77 such as extracellular matrix, branched-chain amino acid metabolism, and neuronal system for both
78 diseases. Further incorporation of tissue-specific regulatory networks pinpointed potential key regulators
79 that orchestrate the biological processes shared between the two diseases, which were cross-validated
80 using cell culture and mouse models. This study suggests potential new therapeutic targets that warrant
81 further investigation for both CVD and T2D.

82 **Introduction**

83 Cardiovascular disease (CVD) and type 2 diabetes (T2D) are two leading causes of death in the United
84 States [1]. Patients with T2D are at two to six times higher risk of developing CVD compared to those
85 without T2D [2], indicating the importance of targeting common pathogenic pathways to improve the
86 prevention, diagnosis, and treatment for these two diseases. While decades of work has revealed
87 dyslipidemia, dysglycemia, inflammation, and hemodynamic disturbances as common pathophysiological
88 intermediates for both CVD and T2D [3-5], very few studies have directly investigated the genomic
89 architectures shared by the two diseases. While genetic factors are known to play a fundamental role in
90 the pathogenesis of both CVD and T2D [6], a direct comparison of the top risk variants between these
91 diseases has revealed few overlapping loci in genome-wide association studies (GWAS) from multiple
92 large consortia. Aside from the speculation that the strongest genetic risks may be disease-specific, the
93 agnostic approach requiring the application of strict statistical adjustment for multiple comparisons also
94 increases false negative rate because of the lack of “genome-wide significance”.

95 To meet these challenges, we and others have previously shown that hidden disease mechanisms can be
96 unraveled through the assessment of the combined activities of genetic loci with weak to moderate effects
97 on disease susceptibility by integrating GWAS with functional genomics and regulatory gene networks
98 [7-11]. Importantly, such high-level integration approaches are able to overcome substantial heterogeneity
99 between independent datasets and extract robust biological signals across molecular layers, tissue types,
100 and even species [8, 12-14]. This advantage is mainly conferred by the aggregation of genetic signals
101 from individual studies onto a comparable ground – molecular pathways and gene networks, before
102 conducting meta-analysis across studies [14, 15]. In other words, even if the genetic variants and linkage
103 architecture can be different between studies, the biological processes implicated are more reproducible
104 and comparable across studies [16]. In the current investigation, we employed a systematic data-driven
105 approach that leveraged multi-dimensional omics datasets including GWAS, tissue-specific expression
106 quantitative trait loci (eQTLs), ENCODE, and tissue-specific gene networks (**Fig 1**). GWAS datasets

107 were from three well-characterized and high-quality prospective cohorts of African Americans (AA),
108 European Americans (EA), and Hispanic Americans (HA) - the national Women's Health Initiative (WHI)
109 [8], the Framingham Heart Study (FHS) [17], and the Jackson Heart Study (JHS) [18]. To maximize the
110 reproducibility of our findings across different populations, we also incorporated meta-analyses of CVD
111 and T2D genetics from CARDIoGRAMplusC4D [19] and DIAGRAM [20]. Further, we
112 comprehensively curated functional genomics and gene networks derived from 25 tissue or cell types
113 relevant to CVD and T2D. A streamlined integration of these rich data sources using our Mergeomics
114 pipeline [14, 15] enabled the identification of shared pathways, gene subnetworks, and key regulators for
115 both CVD and T2D across cohorts and ethnicities. Finally, we validated the subnetworks using adipocyte
116 and knockout mouse models, and confirmed their associations with cardiometabolic traits in the Hybrid
117 Mouse Diversity Panel (HMDP) comprised of >100 mouse strains [21-23].

118 **Results**

119 **Identification of Co-expression Modules Genetically Associated with CVD and T2D across Cohorts**

120 We first investigated whether genetic risk variants of CVD and T2D from GWAS of each cohort/ethnicity
121 were aggregated in a functionally coherent manner by integrating GWAS with tissue-specific eQTLs or
122 ENCODE information and gene co-expression networks that define functional units of genes (**Fig 1A**).
123 Briefly, co-expression networks were constructed from an array of transcriptomic datasets of various
124 tissues relevant to CVD and T2D (details in **Methods**). These modules were mainly used to define sets of
125 functionally related genes in a data-driven manner. Genes within the co-expression modules (a module
126 captures functionally related genes) were mapped to single nucleotide polymorphisms (SNPs) that most
127 likely regulate gene functions via tissue-specific eQTLs or ENCODE information. SNPs were filtered by
128 linkage disequilibrium (LD) and then a chi-square like statistic was used to assess whether a co-
129 expression module shows enrichment of potential functional disease SNPs compared to random chance
130 using the marker set enrichment analysis (MSEA) implemented in our Mergeomics pipeline (details in

131 **Methods**) [14]. Subsequently, meta-analyses across individual MSEA results at the co-expression module
132 level were conducted using the Meta-MSEA function in Mergeomics to retrieve robust signals across
133 studies. Among the 2,672 co-expression modules tested, 131 were found to be significant as defined by
134 false discovery rate (FDR) < 5% in Meta-MSEA across studies (**Table 1, S1 Table**). Moreover, the
135 majority of the disease relevant tissues or cell types included in our analysis yielded informative signal,
136 supporting the systemic pathogenic perturbations known for CVD and T2D (**S1 Fig**). Of the significant
137 modules identified, 79 were associated with CVD and 54 with T2D. Two modules were associated with
138 both diseases, with one enriched for “carbohydrate metabolism” genes and the other over-represented
139 with “other glycan degradation; known T2D genes” (**Fig 2A, S1 Table**). Examination of these two shared
140 modules showed that the genetic signals driving the module significance were largely different between
141 CVD and T2D, with 14.8% lead SNPs overlapping for the carbohydrate metabolism module and 5.8%
142 lead SNPs overlapping for the glycan degradation module between diseases. These results indicate that
143 the GWAS signals for the two diseases in each module do not necessarily overlap, but the CVD and T2D
144 genes are likely functionally connected since they are co-expressed in the same modules and annotated
145 with coherent functions. Additionally, the majority of the CVD modules and T2D modules were
146 identified in more than one ethnic group based on MSEA analysis of individual studies, supporting
147 consistency across ethnicities (**Fig 2B**).

148 **Shared Biological Processes among the CVD/T2D-associated Co-expression Modules**

149 Apart from the two directly overlapping modules, between the CVD- and T2D-associated modules there
150 were many overlapping genes, indicating additional shared functions that contribute to both diseases (**S2**
151 **Fig**). Upon annotating the disease-associated modules using functional categories curated in Kyoto
152 Encyclopedia of Genes and Genomes (KEGG) and Reactome while correcting for the overlaps between
153 pathways (method details in **S1 Text; S3 Fig; S2 Table**), we found significant functional overlaps
154 between the CVD and T2D modules (overlap $p = 3.1e-15$ by Fisher’s exact test, **Fig 2C**). We further
155 ranked all the enriched functional categories by the number of CVD/T2D modules that were annotated

Mis en forme : Police :Non Gras

156 with each functional term (**Fig 3**), which showed a wide spectrum of biological processes shared by both
157 CVD and T2D across ethnicities and cohorts. Of the top ranked processes for the significant co-
158 expression modules identified, we observed well-established pathogenic processes such as lipid and fatty
159 acid metabolism [24], glucose metabolism [25], oxidation [26], and cytokine signaling [27]. Pathways
160 previously implicated mainly for T2D such as beta-cell function were also found to be shared for both
161 CVD and T2D. Interestingly, our completely data-driven approach also identified extracellular matrix
162 (ECM) and branched chain amino acids (BCAA) metabolism as top functional categories whose roles in
163 the development of cardiometabolic disorders have only been implicated in recent experimental work [28-
164 30]. Furthermore, our analysis also revealed under-appreciated processes involving the neuronal system
165 and transport of small molecules.

166 **Identification and Prioritization of Key Drivers (KDs) and Subnetworks for the CVD/T2D-** 167 **associated Modules**

168 The coexpression networks used above mainly served to capture coexpression patterns between genes and
169 to define data-driven gene sets that contain functionally related genes, but they do not carry detailed
170 topology information on gene-gene regulatory relationships. To dissect the gene-gene interactions within
171 and between the 131 disease-associated modules, and to identify key perturbation points shared for both
172 CVD and T2D modules, we used the GIANT networks [31] and Bayesian networks (BNs) from 25 CVD
173 and T2D relevant tissue and cell types, which provide detailed topological information on gene-gene
174 regulatory relationships necessary for the wKDA analysis. The BNs used in our study were generated
175 using similar sets of mouse and human gene expression datasets as used for the co-expression networks,
176 but additionally incorporated genetic data to model causal gene regulatory networks, whereas the GIANT
177 networks were derived based on independent gene expression datasets and protein interaction information.
178 We included both types of gene regulatory networks to increase the coverage of functional connections
179 between genes and only considered KDs identified in both to enhance the robustness of KD prediction.

180 Specifically, all genes in each of the co-expression modules genetically associated with CVD or T2D as
181 identified in our study were mapped onto the GIANT and BN graphical networks to identify KDs using
182 the weighted key driver analysis (wKDA) implemented in Mergeomics [14], where KDs were defined as
183 genes whose local network neighborhoods demonstrate significant enrichment of genes from disease-
184 associated modules (details in **Methods**; concept depicted in **S4 Fig**). Of note, wKDA gives higher
185 weight to network edges that are consistent across network models constructed from independent studies,
186 therefore alleviating potential bias caused by dataset heterogeneity. We identified 226 KDs that were
187 consistently captured in Bayesian and GIANT network at Bonferroni-corrected p-value < 0.05 (**Fig 1B**),
188 among which 162 were KDs for both CVD and T2D associated modules. Bonferroni-correction was used
189 here to focus on the strongest KDs for prioritization purposes. To further prioritize these 162 shared KDs,
190 tissue-specific subnetworks of these KDs were evaluated using Meta-MSEA to rank the magnitude of
191 their genetic association with CVD and T2D across cohorts, yielding 15 top-ranked KDs at FDR<10% in
192 Meta-MSEA for CVD and T2D separately (combined FDR<1% for both diseases simultaneously) (**Fig**
193 **1B, Table 2**). The top KD subnetworks were related to similar pathogenic processes highlighted in the
194 previous section, including cholesterol biosynthesis, respiratory electron transport, immune system and
195 ECM. We further inferred the directionality of the effects of each specific KD on both diseases using
196 GWAS signals mapped to each KD based on eQTLs or chromosomal distance (details in **Methods**;
197 results in **S5 Fig**). This analysis differentiated the KDs into those showing consistent direction of
198 association for both CVD and T2D (*ACLY, CAVI, SPARC, COL6A2, IGF1*), inverse directions with CVD
199 and T2D (*HMGCR, IDI1*), and uncertain directions (**Table 2**). Therefore, the shared KDs do not
200 necessarily affect the risks for the two diseases in the same direction.

201 **Shared KDs and Subnetworks Orchestrate Known CVD and T2D Genes**

202 The KDs and subnetworks were identified based on the full spectrum of genetic evidence (from strong to
203 moderate and subtle) from the various GWAS datasets examined in the current study. To assess whether
204 the top KD subnetworks were enriched for previously known disease genes that mostly represent the

205 strong and replicated genes as a means of cross-validation, we manually curated previously reported
206 genes associated with CVD, T2D, and intermediate metabolic traits related to CVD, T2D (glucose, insulin,
207 lipids, obesity) from DisGeNET [32] and the NHGRI GWAS Catalog [6] (**Fig 1C**, genes listed in **S3**
208 **Table**). The connection between the top 15 KDs and known genes for CVD, T2D and relevant
209 cardiometabolic traits was confirmed by the significant over-representation of the known disease genes in
210 KD subnetworks, with fold enrichment as large as 8, confirming the strong biological importance of these
211 KDs (**Fig 4A**). Further, the top 15 KDs showed direct connections to 28 GWAS hits reaching genome-
212 wide significance ($p < 5e-8$) for CVD and 16 for T2D, which account for 35% (fold = 3.35, $p = 7.18e-10$)
213 and 22% (fold = 2.16, $p = 8.08e-4$) of all reported significant GWAS signals for CVD and T2D in GWAS
214 catalog, respectively. Two of the 15 top KDs, namely *HMGCR* and *IGF1*, were previously identified as
215 signals of genome-wide significance for obesity, lipids and T2D, all risk factors of CVD. Additionally,
216 network visualization revealed tissue-specific KDs and interactions of CVD and T2D genes in many
217 disease-relevant tissues including adipose, adrenal gland, artery, blood, digestive tract (small intestine,
218 colon), hypothalamus, islet, liver, lymphocyte, skeletal muscle, thyroid, and vascular endothelium (**Fig**
219 **4B**). *PCOLCE* represents an intriguing hypothalamus-specific KD that interacts with important energy
220 homeostasis genes like leptin receptor *LEPR*, suggesting a role of neurohormonal control in CVD and
221 T2D pathogenesis. In contrast, *CAVI* appeared to interact extensively with other KDs in peripheral tissues,
222 especially in the adipose tissue.

223 **Experimental Validation of *CAVI* Subnetworks using an *in vitro* Adipocyte Model and *in vivo*** 224 **Knockout Mouse Model**

225 *CAVI* is a robust KD for CVD- and T2D-associated modules across multiple tissues, with the adipose
226 tissue subnetwork of *CAVI* containing the largest number of neighboring genes (**Fig 4B**). In addition,
227 adipose tissue is the only tissue where *CAVI* is a KD in both the Bayesian networks and GIANT networks.
228 These lines of evidence implicate the potential importance of *CAVI* adipose subnetwork in the shared
229 pathogenesis for both diseases. Indeed, *Cavi*^{-/-} mice have been shown to alter the lipid profile,

230 susceptibility to atherosclerosis, and insulin resistance [33, 34]. To assess whether perturbation of this
231 potential KD induces changes in the subnetwork genes as predicted by our network modeling, we
232 performed validation by conducting siRNA-mediated knock down of *Cav1* in differentiating mouse 3T3-
233 L1 adipocytes and by evaluating the whole transcriptome alteration in mouse gonadal adipose tissue
234 between wild type and *Cav1*^{-/-} mice [33] (**Fig 1C**; details in **Methods**). Of the 12 adipose network
235 neighbors of *Cav1* that were tested *in vitro*, 6 exhibited significant changes in expression level on day 2
236 after ~60% *Cav1* knockdown using two siRNAs against *Cav1*. In contrast, none of the 5 negative controls,
237 which were randomly selected among adipocyte genes that are not connected to *Cav1* or its first level
238 neighbors in the adipose network, were affected after *Cav1* perturbation (**Fig 5A**). *Cav1* knockdown also
239 led to decreased expression of *Pparg*, a major adipocyte differentiation regulator (**S6 Fig**), supporting a
240 role of *Cav1* in adipocyte differentiation as previously observed [35].

241 In 3-month-old *Cav1*^{-/-} mice which showed perturbed lipid and insulin sensitivity profiles, we observed
242 1,474 differentially expressed genes (DEGs) at FDR<1%. We found that the first and second level
243 neighbors of *CAVI* in our predicted subnetwork showed significant enrichment for DEGs in adipose
244 tissue induced by *Cav1* knockout, with the degree of fold enrichment increasing as the statistical cutoff
245 used to define DEGs became more stringent (**Fig 5B**; subnetwork view with DEGs in **S7 Fig**). On the
246 contrary, the third and fourth level neighbors of *CAVI* in our predicted subnetwork did not exhibit such
247 enrichment of DEGs (**Fig 5B**). These experimental findings support that *CAVI* is a key regulator of the
248 subnetwork and the network structure predicted by our network modeling is reliable, although it is
249 difficult to discern whether the network changes are related to alterations in adipocyte differentiation
250 status. We also observed strong enrichment for the focal adhesion pathway in both the predicted *Cav1*
251 adipose subnetwork (p=9.6e-14 by Fisher's exact test, fold enrichment = 6.0) and the differential adipose
252 genes in *Cav1*^{-/-} mice (p = 6.9e-9, fold enrichment = 3.5).

253 **Shared KDs Are Associated with CVD and T2D Traits in Experimental Mouse Models**

254 We further assessed the transcriptomic profiling in adipose (relevant to T2D and CVD) and aorta tissue
255 (main site of CVD) in relation to 7 cardiometabolic phenotypes including adiposity, lipid levels
256 (Triglyceride, LDL, HDL), fasting glucose, fasting insulin and HOMA-IR, across >100 mouse strains in
257 two HMDP panels [21-23]. HMDP is a systems genetics resource that comprises more than 100
258 commercially available mouse strains differing in genetic composition, and has emerged as a power tool
259 to study complex human diseases [22, 36]. The biological relevance of HMDP to human pathophysiology
260 has been reproducibly demonstrated [37-39]. Moreover, HMDP data was completely independent of the
261 human-focused genetic datasets and the network datasets used in our primary integrative analysis (**Fig**
262 **1C**). Here we selected two specific HMDP panels, high-fat (HF) and atherogenic (ATH), in which mice
263 were either fed with a high-fat high-sucrose diet or underwent transgenic expression of human APOE-
264 Leiden and CETP gene as a pro-atherogenic background, respectively. These two panels were chosen for
265 their representativeness of human T2D (the HF panel) and CVD (the ATH panel) pathology. First, we
266 investigated the correlation between the expression of 14 top KDs (no probe for KD *MSMO1* in HMDP)
267 and cardiometabolic traits in the adipose and aorta tissues assessed in HMDP. All 14 KDs displayed
268 significant trait association in HMDP, with the association for 11 KDs replicated in both the HF and ATH
269 HMDP panels (**Fig 6A**). Next, we retrieved the adipose and aorta gene-trait correlation statistics for the
270 top KD subnetwork genes, and used MSEA to test whether genes in the KD subnetworks displayed an
271 overall overrepresentation of strong trait association in HMDP. Again, the 14 KD subnetworks showed
272 significant trait association after Bonferroni correction (**Fig 6B**). These findings support that the close
273 involvement of the KDs in cardiometabolic trait perturbation we predicted based on human datasets can
274 be cross-validated in mouse models.

275 **Causal Implication of the Shared KD Subnetworks in Experimental Mouse Models**

276 *Cav1* knockout in mice led to dysregulation of the predicted subnetwork (**Figure 5B**) and significant
277 alterations in cardiometabolic phenotypes [33, 34], supporting the causal role of *CAVI* in both CVD and

278 T2D. To further investigate the potential causal role of the top KDs and their subnetworks in CVD and
279 T2D, we conducted integrative analysis of the KD subnetworks to assess their disease association using
280 GWAS results for the 7 cardiometabolic traits from HMDP and tissue-specific cis-eQTLs (**Fig 1C**). By
281 mapping GWAS signals to genes using adipose or aorta eQTLs and testing for enrichment of genetic
282 association with cardiometabolic traits within the KD subnetwork genes using MSEA, we found
283 consistent and significant association between cardiometabolic traits and the subnetworks of KDs *ACAT2*,
284 *CAVI*, *COL6A2*, *IGF1*, *PCOLCE*, and *SPARC* across adipose and aorta (**Fig 6C**). These results informed
285 by mouse GWAS support a potential causal role of these top KDs in perturbing gene networks in multiple
286 tissues to trigger CVD and T2D.

287 **Discussion**

288 CVD and T2D are highly correlated complex diseases and share many common risk factors. Multiple
289 genetic variants may individually exert subtle to strong effects on disease pathogenesis, and in aggregate
290 perturb diverse pathogenic pathways [8, 9, 13, 19, 20, 40]. In this systems-level, data-driven analysis of
291 GWAS from several large and high-quality cohorts of diverse ethnicities, integrated with functional data
292 (from ENCODE, eQTLs, tissue-specific co-expression and regulatory networks constructed from human
293 and mouse experiments), we identified both known and novel pathways and gene subnetworks that were
294 genetically linked to both CVD and T2D across cohorts and ethnicities. Further, KDs in tissue-specific
295 subnetworks appear to regulate many known disease genes for increased risk of CVD and T2D. Lastly,
296 we experimentally validated the network topology using *in vitro* adipocyte and data from *in vivo* gene
297 knockout models, and confirmed the role of the top KDs and subnetworks in both CVD and T2D traits in
298 independent sets of mouse studies.

299 The data-driven nature of the current study offers several strengths. First, we incorporated the full-scale of
300 genetic variant-disease association from multiple cohorts, ethnicities and disease endpoints, allowing for
301 the detection of subtle to moderate signals, as well as comparison and replication of results across

302 diseases and populations. More importantly, by focusing on results that demonstrate consistent
303 significance at pathway and network level, we overcome the difficulties in harmonizing independent
304 datasets that are complicated by substantial heterogeneity due to platform differences and population
305 substructure. This is because disease signals across populations are more conserved at pathway level than
306 at individual variant and gene levels [12, 14, 16]. Second, the comprehensive incorporation of tissue-
307 specific eQTLs, coupled with the use of tissue-specific networks, enhances our ability to achieve better
308 functional mapping between genetic variants and genes, and uncover systems-level regulatory circuits for
309 CVD and T2D in a tissue-specific fashion. Third, data-driven modules and networks used in this study
310 increase the potential for novel discovery as gene-gene interactions are defined by data rather than prior
311 knowledge. As the network models were from many independent studies reflecting diverse physiological
312 conditions, leveraging these datasets and network models offers more comprehensive coverage of
313 biological interactions than any given dataset can provide and has proven a valuable approach to unveil
314 novel biological insights [9, 13, 41]. While some of our findings confirmed those from previous
315 canonical pathway-based analysis on disease processes including ECM-receptor interaction and cell-
316 adhesion, and KDs such as *SPARC* [8], our data-driven approach in the current study uncovered
317 numerous novel genes, pathways, and gene subnetworks. A likely reason for the enhanced discovery
318 potential of co-expression modules is that several interacting pathways could be co-regulated in a single
319 module, or a pathway could interact with other poorly annotated processes in a module to together confer
320 disease risk. The use of modules capturing such interactions improves the statistical power, in contrast to
321 testing the pathways individually. Lastly, we conducted cross-validation studies in support of the
322 functional roles of specific KDs and subnetworks in CVD and T2D using independent experimental
323 models.

324 We acknowledge the following limitations in our study. First, our analyses were constrained by the
325 coverage of functional datasets that are currently available, which causes uneven tissue coverage between
326 data types and statistical bias towards more commonly profiled tissues such as adipose and liver, making

327 it difficult to achieve precise inference for all relevant tissues. Although we believe this does not
328 necessarily undermine the validity of the main findings from our study, we acknowledge that we likely
329 have missed relevant biology from tissues with fewer studies and smaller sample sizes. Further
330 investigation is needed when additional relevant datasets become available. Secondly, our FDR estimates
331 in MSEA do not take into consideration the gene overlap structure among co-expression modules, due to
332 the challenge in accurately adjusting for the various degrees of overlaps between module pairs. To
333 alleviate this limitation, we focus on modules and pathways demonstrating consistency across datasets
334 and merge overlapping modules subsequently. Thirdly, although we conducted validation experiments on
335 the *CAVI* subnetwork in both *in vitro* and *in vivo* models and cross-validated the importance of the
336 predicted top key drivers and subnetworks in two independent large-scale mouse population studies,
337 further experiments are warranted to thoroughly test the causality of the predicted KDs and elucidate the
338 detailed tissue-specific mechanisms of the KDs on CVD and T2D. This is particularly important
339 considering the limited overlaps in the modules and KDs identified from our study and the ones identified
340 in two recent multi-tissue network analysis of cardiometabolic diseases [10, 11]. Only 7 KDs overlapped
341 including *APOA1*, *CD2*, *CEBPD*, *CENPF*, *CSF1R*, *CTSS*, *UBE2S*. Methodological differences in network
342 inference and key driver analysis and differences in the pathophysiological conditions of the study
343 populations could explain the discrepancies. Lastly, ethnic-specific and sex-specific mechanisms await
344 future exploration.

345 There are several direct implications that can be drawn from the results of our integrative analyses of both
346 observational and experimental data. First, it appears that pathogenic pathways for CVD and T2D are
347 indeed common in ethnically diverse populations. These shared pathways capture most of the critical
348 processes that have been previously implicated in the development of either T2D or CVD, including
349 metabolism of lipids and lipoproteins, glucose, fatty acids, bile acids metabolism, biological oxidation,
350 coagulation, immune response, cytokine signaling, and PDGF signaling. Second, BCAA metabolism and
351 ECM are among the top and common pathways identified. Our finding on BCAA is consistent with recent

352 work relating serum levels of BCAA to risk of CVD and T2D in large prospective cohorts [42, 43],
353 although whether BCAA is a “pathophenotype” or strong pathogenic factor has been debated [28, 44].
354 Our findings support a causal role of BCAA because 1) both CVD and T2D risk variants were enriched in
355 the co-expression modules related to BCAA degradation, and 2) 15 genes in the BCAA pathway were
356 part of the top KD subnetworks, representing a significant enrichment of BCAA genes (fold enrichment =
357 3.02, Fisher’s exact test $p = 1.4e-5$). Of note, BCAA genes themselves carry few genetic risk variants for
358 CVD and T2D, albeit their network neighboring genes are highly enriched for disease variants, which
359 may result from negative evolutionary pressure due to the critical role of BCAA. More recently, Jang and
360 colleagues have shown BCAA catabolism can cause insulin resistance, providing further support for the
361 causal role of BCAA for both CVD and T2D [45]. Our finding on the role of ECM in both CVD and
362 T2D is also in line with recent reports [8, 13, 29, 30, 46]. In the top enriched subnetworks, ECM genes
363 appear to exert strong effect (**Fig 4B**) coordinating other processes such as cholesterol metabolism,
364 energy homeostasis, and immune response across a wide range of peripheral tissues and endocrine axis.
365 This substantiates the importance of ECM modeling as a mechanistic driver for CVD and T2D.

366 Secondly, our comprehensive network modeling identified critical disease modulators and key targets
367 whose functional roles were subsequently supported by multiple cross-validation efforts. This supports
368 the use of network modeling to unravel and prioritize promising top targets that may have high
369 pathogenic potential for both CVD and T2D. The KDs we identified can be considered as “highly
370 confident” for the following reasons: 1) they are KDs for both CVD and T2D associated modules, 2) the
371 tissue-specific subnetworks of these KDs show significant and replicable association with both diseases, 3)
372 their subnetworks are highly enriched with known CVD and T2D genes, 4) *in vitro* siRNA knockdown
373 and *in vivo* knockout mouse experiments confirm the role of a central KD *CAVI* in regulating the
374 downstream genes as predicted in our network model, and 5) both the expression levels of KDs and the
375 genetic variants mapped to the KD subnetworks are significantly associated with CVD and T2D relevant
376 traits in independent mouse populations with naturally occurring genetic variations.

377 Thirdly, most KDs are not GWAS signals reaching genome-wide significance, nor are they rare-variant
378 carrying genes, indicating that standard genetic studies miss important genes that orchestrate known CVD
379 and T2D genes. The phenomenon may reflect a negative evolutionary pressure experienced by the KDs
380 due to their crucial functions. In support of this hypothesis, we found a significant enrichment of human
381 essential genes lacking functional variations among the 162 KDs identified in our study [47] (Fold = 1.41,
382 $p = 9.02e-3$). This is consistent with previous findings [8, 9, 13] reaffirming the power and reliability of
383 our approach in uncovering hidden biological insights particularly in a systematic integrative manner.

384 The connections between KDs and other disease genes revealed by our study warrant future investigation
385 into the potential gene-gene interactions. Indeed, a closer examination of the biological functions from the
386 top shared KDs further corroborates their disease relevance. For instance, our network modeling
387 identified *HMGCR* as a top KD, consistent with its primary role as the target for cholesterol-lowering
388 HMG-CoA inhibitors, namely statins. Our directionality inference analysis indicates that *HMGCR* is
389 associated with CVD and T2D in opposite directions. This is consistent with the recent findings that
390 genetic variations in *HMGCR* that decrease CVD risk cause slightly increased T2D risk, and statin drugs
391 targeting *HMGCR* reduces CVD risk but increases T2D risk [48-50]. *CAVI* and *IGFI* represent two
392 tightly connected multi-functional KDs. *CAVI* null mice were found to have abnormal lipid levels,
393 hyperglycemia, insulin resistance and atherosclerosis [33, 34]. Consistent with these observations, we
394 found strong association of *CAVI* expression levels as well as *CAVI* network with diverse
395 cardiometabolic traits in both human studies and mouse HMDP panels. Our data-driven approach also
396 revealed the central role of *CAVI* in adipose tissue by elucidating its connection to a large number of
397 CVD and T2D GWAS genes and to genes involved in focal adhesion and inflammation (**Fig 4**), which
398 could drive adipocyte insulin resistance [51, 52]. The regulatory effect of *CAVI* on neighboring genes
399 was subsequently validated using *in vitro* adipocyte and *in vivo* mouse models. Moreover, our network
400 modeling also captured the central role of *CAVI* in muscle and artery tissues, suggesting multi-tissue
401 functions of *CAVI* in the pathogenic crossroads for CVD and T2D. The other multi-functional KD, *IGFI*,

402 is itself a GWAS hit for fasting insulin and HOMA-IR. Despite being primarily secreted in liver, in our
403 study *IGF1* demonstrated an adrenal gland and muscle specific regulatory circuit with CVD and T2D
404 genes, suggesting that it may confer risk to these diseases through the adrenal endocrine function and
405 muscle insulin sensitivity. The three ECM KDs we identified, *SPARC*, *PCOLCE* and *COL6A2*, were
406 especially interesting due to their consistent and strong impact on diverse cardiometabolic traits shown in
407 our cross-validation analyses in HMDP (**Fig 4**, **Fig 6**). *SPARC* encodes osteonectin, which is primarily
408 circulated by adipocytes. It inhibits adipogenesis and promotes adipose tissue fibrosis⁵⁰. *SPARC* is also
409 associated with insulin resistance and coronary artery lesions^{51,52}. *PCOLCE* (procollagen C-
410 endopeptidase enhancer) represents a novel regulator for hypothalamus ECM that could potentially
411 disrupt the neuroendocrine system. *COL6A2*, on the other hand, provides new insights into how collagen
412 may affect cardiometabolic disorders: in adrenal tissue *COL6A2* is connected to *IGF1R*, the direct
413 downstream effector for KD *IGF1*. Importantly, our directionality analysis suggests that while some KDs
414 such as *CAVI* may have similar directional effects on CVD and T2D, cases like *HMGCR* that show
415 opposite effects on these diseases are also present. Therefore, it is important to test the directional
416 predictions to prioritize targets that have the potential to ameliorate both diseases and deprioritize targets
417 with opposite effects on the two diseases.

418 In summary, through integration and modeling of a multitude of genetics and genomics datasets, we
419 identified key molecular drivers, pathways, and gene subnetworks that are shared in the pathogenesis of
420 CVD and T2D. Our findings offer a systems-level understanding of these highly clustered diseases and
421 provide guidance on further basic mechanistic work and intervention studies. The shared key drivers and
422 networks identified may serve as more effective therapeutic targets to help achieve systems-wide
423 alleviation of pathogenic stress for cardiometabolic diseases, due to their central and systemic role in
424 regulating scores of disease genes. Such network-based approach represents a new avenue for therapeutic
425 intervention targeting common complex diseases.

426 **Methods**

427 **Identification of qualified SNPs from GWAS of CVD and T2D**

428 Detailed GWAS information including sample size, ethnicity and genotyping platform was described in
429 **S4 Table** and **S1 Text**. Briefly, p-values of qualified single nucleotide polymorphisms (SNPs) at minor
430 allele frequency > 0.05 and imputation quality > 0.3 for CVD and T2D were collected for all available
431 GWAS datasets (WHI-SHARE, WHI-GARNET, JHS, FHS, CARDIoGRAMplusC4D [19], and
432 DIAGRAM [20]). SNPs meeting the following criteria were further filtered out: 1) ranked in the bottom
433 50% (weaker association) based on disease association p-values and 2) in strong linkage disequilibrium
434 (LD) ($r^2 > 0.5$) according to ethnicity-specific LD data from Hapmap V3 [53] and 1000 Genomes[54].
435 For each GWAS dataset, LD filtering was conducted by first ranking SNPs based on the association p
436 values and then checking if the next highest ranked SNP was in LD with the top SNP. If the r^2 was above
437 0.5, the SNP with lower rank was removed. The step was repeated by always checking if the next SNP
438 was in LD with any of the already accepted ones.

439 **Curation of Data-driven Gene Co-expression Network Modules**

440 Using the Weighted Gene Co-expression Network Analysis (WGCNA)[55], we constructed gene co-
441 expression modules capturing significant co-regulation patterns and functional relatedness among groups
442 of genes in multiple CVD- or T2D-related tissues (including aortic endothelial cells, adipose, blood, liver,
443 heart, islet, kidney, muscle and brain) obtained from nine human and mouse studies (**S5 Table**). Modules
444 with size smaller than 10 genes were excluded to avoid statistical artifacts, yielding 2,672 co-expression
445 modules. These coexpression modules were used as a collection of data-driven sets of functionally
446 connected genes for downstream analysis. The potential biological functions of each module were
447 annotated using pathway databases Reactome and KEGG, and statistical significance was determined by
448 Fisher's exact test with Bonferroni-corrected $p < 0.05$.

449 **Curation of Functional Genomics from eQTLs and ENCODE**

450 eQTLs establish biologically meaningful connections between genetic variants and gene expression, and
451 could serve as functional evidence in support of the potential causal role of candidate genes in pathogenic
452 processes[56, 57]. We therefore conducted comprehensive curation for significant eQTLs in a total of 19
453 tissues that have been identified by various consortia (including the Genotype-Tissue Expression (GTEx)
454 [58], Muthur [59] and Cardiogenics [60], and additional independent studies; **S6 Table**). Additional
455 functional genomics resources from ENCODE were also curated to complement the eQTLs for SNP-gene
456 mapping (**S1 Text**).

457 **Identification of Genetically-driven CVD and T2D Modules using Marker Set Enrichment Analysis**
458 **(MSEA)**

459 MSEA was used to identify co-expression modules with over-representation of CVD- or T2D-associated
460 genetic signals for each disease GWAS in each cohort/ethnicity in a study specific manner. MSEA takes
461 into three input: 1) Summary-level results of individual GWAS (WHI, FHS, JHS, CARDIoGRAM+C4D,
462 DIAGRAM) for the LD-filtered SNPs; 2) SNP-gene mapping information, which could be determined by
463 tissue-specific cis-eQTLs, ENCODE based functional annotation and chromosome distance based
464 annotation. Cis-eQTLs is defined as eQTLs within 1MB of the transcription starting sites of genes. For
465 ENCODE, we accessed the Regulome database and used the reported functional interactions to map SNPs
466 to genes by chromosomal distance. Only SNPs within 50kb of the gene region and have functional
467 evidence in Regulome database were kept; 3) Data-driven co-expression modules from multiple human
468 and mouse studies as described above. Tissue-specificity was determined by the tissue origins of eQTLs
469 and ethnic specificity was determined by the ethnicity of each GWAS cohort, since the human disease
470 genetic signals and human eQTL mapping were the main driving factors to determine the significance of
471 the modules. MSEA employs a chi-square like statistic with multiple quantile thresholds to assess
472 whether a co-expression module shows enrichment of functional disease SNPs compared to random
473 chance [14]. The varying quantile thresholds allows the statistic to be adoptable to studies of varying

474 sample size and statistical power. For the list of SNPs mapped to each gene-set, MSEA tested whether the
475 SNP list exhibited significant enrichment of SNPs with stronger association with disease using a chi-
476 square like statistic: $\chi = \sum_{i=1}^n \frac{O_i - E_i}{\sqrt{E_i + \kappa}}$, where n denotes the number of quantile points (we used ten
477 quantile points ranging from the top 50% to the top 99.9% based on the rank of GWAS p values), O and
478 E denote the observed and expected counts of positive findings (i.e. signals above the quantile point), and
479 $\kappa = 1$ is a stability parameter to reduce artefacts from low expected counts for small SNP sets. The null
480 background was estimated by permuting gene labels to generate random gene sets matching the gene
481 number of each co-expression module, while preserving the assignment of SNPs to genes, accounting for
482 confounding factors such as gene size, LD block size and SNPs per loci. For each co-expression module,
483 10000 permuted gene sets were generated and enrichment P-values were determined from a Gaussian
484 distribution approximated using the enrichment statistics from the 10000 permutations and the statistics of
485 the co-expression module. Finally, Benjami-Hochberg FDR was estimated across all modules tested for
486 each GWAS.

487 To evaluate a module across multiple GWAS studies, we employed the Meta-MSEA analysis in
488 Mergeomics, which conducts module-level meta-analysis to retrieve robust signals across studies. Meta-
489 MSEA takes advantage of the parametric estimation of p-values in MSEA by applying Stouffer's Z score
490 method to determine the meta-Z score, then converts it back to a meta P-value. The meta-FDR was
491 calculated using Benjamini-Hochberg method. Co-expression modules with meta-FDR < 5% were
492 considered significant and included in subsequent analyses.

493 **Identification of Key Drivers and Disease Subnetworks**

494 We used graphical gene-gene interaction networks including the GIANT networks [31] and Bayesian
495 networks (BN) from 25 CVD and T2D relevant tissue and cell types (**S7 Table, S1 Text**) to identify KDs.
496 If more than one dataset was available for a given tissue, a network was constructed for each dataset and
497 all networks for the same tissue were combined as a union to represent the network of that tissue, with the

498 consistency of each network edge across datasets coded as edge weight. The co-expression modules
499 genetically associated with CVD or T2D identified by Meta-MSEA were mapped onto these graphical
500 networks to identify KDs using the weighted key driver analysis (wKDA) implemented in Mergeomics
501 [14]. wKDA uniquely consider the edge weight information, either in the form of edge consistency score
502 in the case of BNs or edge confidence score in the case of GIANT networks. Specifically, a network was
503 first screened for suitable hub genes whose degree (number of genes connected to the hub) is in the top 25%
504 of all network nodes. Once the hubs have been defined, their local one-edge neighborhoods, or
505 “subnetworks” were extracted. All genes in each of the CVD and T2D-associated gene sets that were
506 discovered by meta-MSEA were overlaid onto the hub subnetworks to see if a particular subnetwork was
507 enriched for the genes in CVD/T2D associated gene sets. The edges that connect a hub to its neighbors
508 are simplified into node strengths (strength = sum of adjacent edge weights) within the neighborhood,
509 except for the hub itself. The test statistic for the wKDA is analogous to the one used for MSEA: $\chi =$
510 $\frac{O-E}{\sqrt{E-\kappa}}$, except that the values O and E represent the observed and expected ratios of genes from CVD/T2D
511 gene sets in a hub subnetwork. In particular, $E = \frac{N_k N_p}{N}$ is estimated based on the hub degree N_k , disease
512 gene set size N_p and the order of the full network N , with the implicit assumption that the weight
513 distribution is isotropic across the network. Statistical significance of the disease-enriched hubs,
514 henceforth KDs, is estimated by permuting the gene labels in the network for 10000 times and estimating
515 the P-value based on the null distribution. To control for multiple testing, stringent Bonferroni adjustment
516 was used to focus on the top robust KDs. KDs shared by CVD and T2D modules are prioritized based on
517 the following criteria: i) Bonferroni-corrected $p < 0.05$ in wKDA, ii) replicated by both GIANT networks
518 and Bayesian networks, and iii) the genetic association strength between the KD subnetworks (immediate
519 network neighbors of the KDs) and CVD/T2D in Meta-MSEA. Finally, Cytoscape 3.3.0 [61] was used for
520 disease subnetwork visualization.

521 **Inference of the Direction of Genetic Effects of KD subnetworks**

522 We used the genetic effect direction of KDs as a proxy for probable effect direction of the KD
523 subnetworks. For each KD, we retrieved their tissue-specific eQTLs as well as variants within 50kb of the
524 gene region, whose genetic association information was available in both CARDIoGRAMplusC4D and
525 DIAGRAM, the two large meta-consortia of GWAS for CVD and T2D. CVD/T2D association beta-
526 values of mapped variants of KDs were then extracted, and the signs of beta-values were examined to
527 ensure they were based on the same reference alleles between GWAS. Lastly, for all mapped variants on
528 each KD, the signs of the beta-value for CVD and T2D were compared and statistical significance of the
529 proportion of variants with similar or opposite effect direction between diseases was determined by z-test.

530 **Validation of KD Subnetwork Topology Using siRNA Knockdown in Adipocytes**

531 We chose to validate the predicted adipose subnetwork of a top ranked KD of both CVD and T2D, *Cav1*,
532 in 3T3-L1 adipocytes. Cells were cultured to confluence and adipocyte differentiation was induced using
533 MDI differentiation medium (S1 Text). Two days after the initiation of differentiation, cells were
534 transfected with 50 nM *Cav1* siRNAs (3 distinct siRNAs were tested and two of the strongest ones were
535 chosen) or a scrambled control siRNA. For each siRNA, two separate sets of transfection experiments
536 were conducted, with three biological replicates in each experiment. Two days after transfection, cells
537 were collected for total RNA extraction, reverse transcription and quantitative PCR measurement of 12
538 predicted *Cav1* subnetwork genes and 5 random genes not within the subnetwork as negative controls (S1
539 Text). β -actin was used to normalize the expression level of target genes.

540 **Validation of KD Subnetwork Topology Using Cav1 null mice**

541 We accessed the gonadal white tissue gene expression data of 3-month-old wild type and *Cav1*^{-/-} male
542 mice (N=3/group) from Gene Expression Omnibus (GEO accession: GSE35431). Detailed description of
543 the data collection procedures has been described previously [33]. Gene expression was profiled using
544 Illumina MouseWG-6 v2.0 expression beadchip and normalized using robust spline. Differentially

545 expressed genes (DEGs) between genotype groups were identified using linear model implemented in the
546 R package Limma and false discovery rate was estimated using the Benjamini-Hochberg procedure [62].
547 DEGs at different statistical cutoffs were compared to *CAVI* subnetwork genes at different levels (i.e., 1,
548 2, 3, or 4 edges away from *CAVI*) to assess overlap and significance of overlap was evaluated using
549 Fisher's exact test.

550 **Validation of KD Subnetworks Using Mouse HMDP Studies**

551 To further validate the role of KD subnetworks in CVD and T2D, we incorporated genetic, genomic and
552 transcriptomic data from HMDP (comprised of >100 mouse strains differing by genetic composition) [21-
553 23]. HMDP data was from two panels, one with mice fed with a high-fat diet (HF-HMDP)[22], and the
554 other with hyperlipidemic mice made by transgenic expression of human APOE-Leiden and CETP gene
555 (ATH-HMDP)[23]. For HF-HMDP, we retrieved gene-trait correlation data for adipose tissue (due to its
556 relevance to both CVD and T2D) and 7 core cardiometabolic traits including adiposity, fasting glucose
557 level, fasting insulin level, LDL, HDL, triglycerides and homeostatic model assessment-insulin resistance
558 (HOMA-IR). For ATH-HMDP, we retrieved aorta gene-trait correlation (aorta tissue is the main site for
559 CVD in mice) for all 7 traits. In addition to assessing the trait association strengths of individual KDs, we
560 also used MSEA to evaluate the aggregate association strength of the top CVD/T2D KD subnetworks
561 with the traits at both transcription and genetic levels through transcriptome-wide association (TWAS)
562 and GWAS in HF-HMDP and ATH-HMDP (**S1 Text**).

563 **Acknowledgements**

564 We thank [investigators from](#) the National Women's Health Initiative ([WHI-GWAS](#)), the Framingham
565 Heart Study, the Jackson Hearth Study, the DIAGRAM Consortium, [and](#) the CARDIoGRAMplusC4D for
566 [sharing](#) the GWAS summary statistics data. We thank the Cardiogenic Consortium for [data related to](#) the
567 macrophage and monocyte eQTLs. We thank Calvin Pan for assistance to access the HMDP datasets.

568 [Members of the Cardiogenics Consortium not included in the manuscript include:](#)

569 [Tony Attwood, Stephanie Belz, Peter Braund, Jessy Brocheton, François Cambien, Jason Cooper, Abi](#)
570 [Crisp-Hihn, Panos Deloukas, Patrick Diemert \(formerly Linsel-Nitschke\), Jeannette Erdmann, Nicola](#)
571 [Foad, Tiphaine Godefroy, Alison H. Goodall, Jay Gracey, Emma Gray, Rhian Gwilliams, Susanne](#)
572 [Heimerl, Christian Hengstenberg, Jennifer Jolley, Unni Krishnan, Heather Lloyd-Jones, Ulrika Liljedahl,](#)
573 [Ingrid Lugauer, Per Lundmark, Seraya Maouche, Jasbir S Moore, Gilles Montalescot, David Muir,](#)
574 [Elizabeth Murray, Chris P Nelson, Jessica Neudert, David Niblett, Karen O'Leary, Willem H. Ouwehand,](#)
575 [Helen Pollard, Carole Proust, Angela Rankin, Augusto Rendon, Catherine M Rice, Hendrik Sager, Nilesh](#)
576 [J. Samani, Jennifer Sambrook, Gerd Schmitz, Michael Scholz, Laura Schroeder, Heribert Schunkert,](#)
577 [Jonathan Stephens, Ann-Christine Syvannen, Stefanie Tennstedt \(formerly Gulde\), Chris Wallace,](#)

Mis en forme : Police :Gras, Non Italique

578

579 **References**

- 580 1. National Center for Health Statistics. Health, United States, 2015: with special feature on racial
581 and ethnic health disparities. 2016.
- 582 2. Grundy SM, Benjamin IJ, Burke GL, Chait A, Eckel RH, Howard BV, et al. Diabetes and
583 cardiovascular disease: a statement for healthcare professionals from the American Heart Association.
584 *Circulation*. 1999;100(10):1134-46. PubMed PMID: 10477542.
- 585 3. Beckman JA, Creager MA, Libby P. Diabetes and atherosclerosis: epidemiology,
586 pathophysiology, and management. *JAMA : the journal of the American Medical Association*.
587 2002;287(19):2570-81. PubMed PMID: 12020339.
- 588 4. Wilson PWF, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic syndrome as a
589 precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation*. 2005;112(20):3066-72. doi:
590 10.1161/Circulationaha.105.539528. PubMed PMID: WOS:000233255100008.
- 591 5. Mazzone T, Chait A, Plutzky J. Cardiovascular disease risk in type 2 diabetes mellitus: insights
592 from mechanistic studies. *Lancet*. 2008;371(9626):1800-9. doi: 10.1016/S0140-6736(08)60768-0.
593 PubMed PMID: 18502305; PubMed Central PMCID: PMC2774464.
- 594 6. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS
595 Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(Database
596 issue):D1001-6. doi: 10.1093/nar/gkt1229. PubMed PMID: 24316577; PubMed Central PMCID:
597 PMC3965119.
- 598 7. Civelek M, Lusk AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet*.
599 2014;15(1):34-48. doi: 10.1038/nrg3575. PubMed PMID: WOS:000328892200011.
- 600 8. Chan KH, Huang YT, Meng Q, Wu C, Reiner A, Sobel EM, et al. Shared molecular pathways
601 and gene networks for cardiovascular disease and type 2 diabetes mellitus in women across diverse
602 ethnicities. *Circulation Cardiovascular genetics*. 2014;7(6):911-9. Epub 2014/11/06. doi:
603 10.1161/CIRCGENETICS.114.000676. PubMed PMID: 25371518.

604 9. Makinen VP, Civelek M, Meng Q, Zhang B, Zhu J, Levian C, et al. Integrative genomics reveals
605 novel molecular pathways and gene networks for coronary artery disease. *PLoS genetics*.
606 2014;10(7):e1004502. doi: 10.1371/journal.pgen.1004502. PubMed PMID: 25033284.

607 10. Talukdar HA, Foroughi Asl H, Jain RK, Ermel R, Ruusalepp A, Franzen O, et al. Cross-Tissue
608 Regulatory Gene Networks in Coronary Artery Disease. *Cell systems*. 2016;2(3):196-208. doi:
609 10.1016/j.cels.2016.02.002. PubMed PMID: 27135365; PubMed Central PMCID: PMC4855300.

610 11. Franzen O, Ermel R, Cohain A, Akers NK, Di Narzo A, Talukdar HA, et al. Cardiometabolic risk
611 loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science*.
612 2016;353(6301):827-30. doi: 10.1126/science.aad6970. PubMed PMID: 27540175.

613 12. von Scheidt M, Zhao Y, Kurt Z, Pan C, Zeng L, Yang X, et al. Applications and Limitations of
614 Mouse Models for Understanding Human Atherosclerosis. *Cell Metab*. 2016. doi:
615 10.1016/j.cmet.2016.11.001. PubMed PMID: 27916529.

616 13. Zhao Y, Chen J, Freudenberg JM, Meng Q, Rajpal DK, Yang X. Network-Based Identification
617 and Prioritization of Key Regulators of Coronary Artery Disease Loci. *Arterioscler Thromb Vasc Biol*.
618 2016;36(5):928-41. doi: 10.1161/ATVBAHA.115.306725. PubMed PMID: 26966275.

619 14. Shu L, Zhao Y, Kurt Z, Byars SG, Tukiainen T, Kettunen J, et al. Mergeomics: multidimensional
620 data integration to identify pathogenic perturbations to biological systems. *BMC genomics*.
621 2016;17(1):874. doi: 10.1186/s12864-016-3198-9. PubMed PMID: 27814671; PubMed Central PMCID:
622 PMC5097440.

623 15. Arneson D, Bhattacharya A, Shu L, Makinen VP, Yang X. Mergeomics: a web server for
624 identifying pathological pathways, networks, and key regulators via multidimensional data integration.
625 *BMC Genomics*. 2016;17(1):722. doi: 10.1186/s12864-016-3057-8. PubMed PMID: 27612452; PubMed
626 Central PMCID: PMCPMC5016927.

627 16. Mooney MA, Nigg JT, McWeeney SK, Wilmot B. Functional and genomic context in pathway
628 analysis of GWAS data. *Trends in genetics : TIG*. 2014;30(9):390-400. doi: 10.1016/j.tig.2014.07.004.
629 PubMed PMID: 25154796; PubMed Central PMCID: PMC4266582.

630 17. Meigs JB, Manning AK, Fox CS, Florez JC, Liu C, Cupples LA, et al. Genome-wide association
631 with diabetes-related traits in the Framingham Heart Study. *BMC Med Genet.* 2007;8 Suppl 1:S16. doi:
632 10.1186/1471-2350-8-S1-S16. PubMed PMID: 17903298; PubMed Central PMCID: PMC1995610.

633 18. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, et al. Genome-wide
634 association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI
635 CARE Project. *PLoS genetics.* 2011;7(2):e1001300. doi: 10.1371/journal.pgen.1001300. PubMed PMID:
636 21347282; PubMed Central PMCID: PMC3037413.

637 19. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000
638 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics.*
639 2015;47(10):1121-30. doi: 10.1038/ng.3396. PubMed PMID: 26343387; PubMed Central PMCID:
640 PMC4589895.

641 20. Mahajan A, Go MJ, Zhang WH, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-
642 ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility.
643 *Nature genetics.* 2014;46(3):234-+. doi: 10.1038/ng.2897. PubMed PMID: WOS:000332036700007.

644 21. Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, et al. A high-resolution
645 association mapping panel for the dissection of complex traits in mice. *Genome research.* 2010;20(2):281-
646 90. doi: 10.1101/gr.099234.109. PubMed PMID: 20054062; PubMed Central PMCID: PMC2813484.

647 22. Parks BW, Nam E, Org E, Kostem E, Norheim F, Hui ST, et al. Genetic control of obesity and
648 gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab.*
649 2013;17(1):141-52. doi: 10.1016/j.cmet.2012.12.007. PubMed PMID: 23312289; PubMed Central
650 PMCID: PMC3545283.

651 23. Bennett BJ, Davis RC, Civelek M, Orozco L, Wu J, Qi H, et al. Genetic Architecture of
652 Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. *PLoS genetics.*
653 2015;11(12):e1005711. doi: 10.1371/journal.pgen.1005711. PubMed PMID: 26694027; PubMed Central
654 PMCID: PMC4687930.

655 24. Libby P, Theroux P. Pathophysiology of coronary artery disease. *Circulation*. 2005;111(25):3481-
656 8. doi: 10.1161/CIRCULATIONAHA.105.537878. PubMed PMID: 15983262.

657 25. Bornfeldt KE, Tabas I. Insulin resistance, hyperglycemia, and atherosclerosis. *Cell Metab*.
658 2011;14(5):575-85. doi: 10.1016/j.cmet.2011.07.015. PubMed PMID: 22055501; PubMed Central
659 PMCID: PMC3217209.

660 26. Ceriello A, Motz E. Is oxidative stress the pathogenic mechanism underlying insulin resistance,
661 diabetes, and cardiovascular disease? The common soil hypothesis revisited. *Arteriosclerosis, thrombosis,*
662 *and vascular biology*. 2004;24(5):816-23.

663 27. Haffner SM. The metabolic syndrome: inflammation, diabetes mellitus, and cardiovascular
664 disease. *The American journal of cardiology*. 2006;97(2A):3A-11A. doi: 10.1016/j.amjcard.2005.11.010.
665 PubMed PMID: 16442931.

666 28. Lynch CJ, Adams SH. Branched-chain amino acids in metabolic signalling and insulin resistance.
667 *Nature reviews Endocrinology*. 2014;10(12):723-36. doi: 10.1038/nrendo.2014.171. PubMed PMID:
668 25287287; PubMed Central PMCID: PMC4424797.

669 29. Williams AS, Kang L, Wasserman DH. The extracellular matrix and insulin resistance. *Trends in*
670 *endocrinology and metabolism: TEM*. 2015;26(7):357-66. doi: 10.1016/j.tem.2015.05.006. PubMed
671 PMID: 26059707; PubMed Central PMCID: PMC4490038.

672 30. Chistiakov DA, Sobenin IA, Orekhov AN. Vascular extracellular matrix in atherosclerosis.
673 *Cardiology in review*. 2013;21(6):270-88. doi: 10.1097/CRD.0b013e31828c5ced. PubMed PMID:
674 23422022.

675 31. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al.
676 Understanding multicellular function and disease with human tissue-specific networks. *Nature genetics*.
677 2015;47(6):569-76. doi: 10.1038/ng.3259. PubMed PMID: 25915600.

678 32. Pinero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M, et al.
679 DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes.
680 *Database-Oxford*. 2015. doi: ARTN bav028

681 10.1093/database/bav028. PubMed PMID: WOS:000361048600001.

682 33. Asterholm IW, Mundy DI, Weng J, Anderson RG, Scherer PE. Altered mitochondrial function
683 and metabolic inflexibility associated with loss of caveolin-1. *Cell Metab.* 2012;15(2):171-85. doi:
684 10.1016/j.cmet.2012.01.004. PubMed PMID: 22326219; PubMed Central PMCID: PMC3278712.

685 34. Frank PG, Lee H, Park DS, Tandon NN, Scherer PE, Lisanti MP. Genetic ablation of caveolin-1
686 confers protection against atherosclerosis. *Arteriosclerosis, thrombosis, and vascular biology.*
687 2004;24(1):98-105. doi: 10.1161/01.ATV.0000101182.89118.E5. PubMed PMID: 14563650.

688 35. Razani B, Combs TP, Wang XB, Frank PG, Park DS, Russell RG, et al. Caveolin-1-deficient
689 mice are lean, resistant to diet-induced obesity, and show hypertriglyceridemia with adipocyte
690 abnormalities. *J Biol Chem.* 2002;277(10):8635-47. doi: 10.1074/jbc.M110970200. PubMed PMID:
691 11739396.

692 36. Lusic AJ, Seldin MM, Allayee H, Bennett BJ, Civelek M, Davis RC, et al. The Hybrid Mouse
693 Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J Lipid*
694 *Res.* 2016;57(6):925-42. doi: 10.1194/jlr.R066944. PubMed PMID: 27099397; PubMed Central PMCID:
695 PMC4878195.

696 37. von Scheidt M, Zhao Y, Kurt Z, Pan C, Zeng L, Yang X, et al. Applications and Limitations of
697 Mouse Models for Understanding Human Atherosclerosis. *Cell Metab.* 2017;25(2):248-61. doi:
698 10.1016/j.cmet.2016.11.001. PubMed PMID: 27916529.

699 38. Rau CD, Lusic AJ, Wang Y. Genetics of common forms of heart failure: challenges and potential
700 solutions. *Current opinion in cardiology.* 2015;30(3):222-7. doi: 10.1097/HCO.0000000000000160.
701 PubMed PMID: 25768955; PubMed Central PMCID: PMC4406340.

702 39. Parks BW, Sallam T, Mehrabian M, Psychogios N, Hui ST, Norheim F, et al. Genetic architecture
703 of insulin resistance in the mouse. *Cell Metab.* 2015;21(2):334-46. doi: 10.1016/j.cmet.2015.01.002.
704 PubMed PMID: 25651185; PubMed Central PMCID: PMC4349439.

705 40. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the
706 missing heritability of complex diseases. *Nature*. 2009;461(7265):747-53. doi: 10.1038/nature08494.
707 PubMed PMID: 19812666; PubMed Central PMCID: PMC2831613.

708 41. Wang IM, Zhang B, Yang X, Zhu J, Stepaniants S, Zhang C, et al. Systems analysis of eleven
709 rodent disease models reveals an inflammatome signature and key drivers. *Molecular systems biology*.
710 2012;8:594. Epub 2012/07/19. doi: 10.1038/msb.2012.24. PubMed PMID: 22806142; PubMed Central
711 PMCID: PMC3421440.

712 42. Lee C, Watkins SM, Wagenknecht LE, Lorenzo C, Haffner SM, Hanley AJ. Plasma Branched-
713 Chain Amino Acids, Insulin Metabolism, and Incident Type 2 Diabetes-The Insulin Resistance
714 Atherosclerosis Study (IRAS). *Diabetes*. 2014;63:A382-A. PubMed PMID: WOS:000359481602148.

715 43. Bhattacharya S, Granger CB, Craig D, Haynes C, Bain J, Stevens RD, et al. Validation of the
716 association between a branched chain amino acid metabolite profile and extremes of coronary artery
717 disease in patients referred for cardiac catheterization. *Atherosclerosis*. 2014;232(1):191-6. doi:
718 10.1016/j.atherosclerosis.2013.10.036. PubMed PMID: WOS:000329209100027.

719 44. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, et al. Genetic Predisposition to an
720 Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian
721 Randomisation Analysis. *PLoS Med*. 2016;13(11):e1002179. doi: 10.1371/journal.pmed.1002179.
722 PubMed PMID: 27898682.

723 45. Jang C, Oh SF, Wada S, Rowe GC, Liu L, Chan MC, et al. A branched-chain amino acid
724 metabolite drives vascular fatty acid transport and causes insulin resistance. *Nat Med*. 2016;22(4):421-6.
725 doi: 10.1038/nm.4057. PubMed PMID: 26950361; PubMed Central PMCID: PMC4949205.

726 46. Meng Q, Ying Z, Noble E, Zhao Y, Agrawal R, Mikhail A, et al. Systems Nutrigenomics Reveals
727 Brain Gene Networks Linking Metabolic and Brain Disorders. *EBioMedicine*. 2016.

728 47. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-
729 coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-91. doi: 10.1038/nature19057.
730 PubMed PMID: 27535533; PubMed Central PMCID: PMC5018207.

731 48. Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, de Craen AJ, et al. Statins and risk of
732 incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet*. 2010;375(9716):735-
733 42. doi: 10.1016/S0140-6736(09)61965-6. PubMed PMID: 20167359.

734 49. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in
735 PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. *New England Journal of*
736 *Medicine*. 2016;375(22):2144-53.

737 50. Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular benefits and diabetes
738 risks of statin therapy in primary prevention: an analysis from the JUPITER trial. *Lancet*.
739 2012;380(9841):565-71. doi: 10.1016/S0140-6736(12)61190-8. PubMed PMID: 22883507; PubMed
740 Central PMCID: PMC3774022.

741 51. Kusminski CM, Scherer PE. Mitochondrial dysfunction in white adipose tissue. *Trends in*
742 *endocrinology and metabolism: TEM*. 2012;23(9):435-43. doi: 10.1016/j.tem.2012.06.004. PubMed
743 PMID: 22784416; PubMed Central PMCID: PMC3430798.

744 52. Coletta DK, Mandarino LJ. Mitochondrial dysfunction and insulin resistance from the outside in:
745 extracellular matrix, the cytoskeleton, and mitochondria. *American journal of physiology Endocrinology*
746 *and metabolism*. 2011;301(5):E749-55. doi: 10.1152/ajpendo.00363.2011. PubMed PMID: 21862724;
747 PubMed Central PMCID: PMC3214002.

748 53. Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site.
749 *Genome research*. 2005;15(11):1592-3. doi: 10.1101/gr.4413105. PubMed PMID: 16251469; PubMed
750 Central PMCID: PMC1310647.

751 54. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human
752 genomes. *Nature*. 2012;491(7422):56-65.

753 55. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis.
754 *Stat Appl Genet Mol Biol*. 2005;4:Article17. doi: 10.2202/1544-6115.1128. PubMed PMID: 16646834.

755 56. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating Pathway Analysis and
756 Genetics of Gene Expression for Genome-wide Association Studies. *Am J Hum Genet.* 2010;86(4):581-
757 91. doi: DOI 10.1016/j.ajhg.2010.02.020. PubMed PMID: WOS:000276716800008.

758 57. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. Genetics of gene
759 expression and its effect on disease. *Nature.* 2008;452(7186):423-8. doi: 10.1038/nature06758. PubMed
760 PMID: 18344981.

761 58. G. TEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nature genetics.*
762 2013;45(6):580-5. doi: 10.1038/ng.2653. PubMed PMID: 23715323; PubMed Central PMCID:
763 PMC4010069.

764 59. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, et al. The architecture of gene
765 regulatory variation across multiple human tissues: the MuTHER study. *PLoS genetics.*
766 2011;7(2):e1002003. doi: 10.1371/journal.pgen.1002003. PubMed PMID: 21304890; PubMed Central
767 PMCID: PMC3033383.

768 60. Garnier S, Truong V, Brocheton J, Zeller T, Rovital M, Wild PS, et al. Genome-wide haplotype
769 analysis of cis expression quantitative trait loci in monocytes. *PLoS genetics.* 2013;9(1):e1003240. doi:
770 10.1371/journal.pgen.1003240. PubMed PMID: 23382694; PubMed Central PMCID: PMC3561129.

771 61. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software
772 environment for integrated models of biomolecular interaction networks. *Genome research.*
773 2003;13(11):2498-504. doi: 10.1101/gr.1239303. PubMed PMID: 14597658; PubMed Central PMCID:
774 PMC403769.

775 62. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful
776 approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological).*
777 1995:289-300.

778

779 **Supporting Information Captions**

780 S1 Text. Supplemental methods and references.

781 S1 Fig. Number of significant co-expression modules found by different gene-SNP mapping types.

782 S2 Fig. Heatmap of pair-wise overlapping ratio (Jaccard index) between the 79 co-expression modules associated with CVD (y-axis) and 54
783 modules (x-axis) associated with T2D.

784 S3 Fig. Overlap ratio plots between co-expression modules and the annotated functional terms. All the annotated pathways reach >5% overlap
785 ratio either on the pathway side or on the module side. Specifically, a majority (251 out of 278, 90.3%) of the annotated pathways had $\geq 5\%$ of
786 genes overlapping with the modules to which they were assigned. For the 27 annotated pathways where $< 5\%$ pathway genes were represented,
787 these were the cases where the co-expression modules were small and the pathways were large, but all of them showed overlap of $> 5\%$ module
788 genes. The minimum, maximum, mean and median numbers of the overlapping genes for the annotations are 5, 170, 19 and 13, respectively.

789 S4 Fig. Concept of key driver analysis (KDA). KDA requires gene regulatory networks capturing gene-gene interactions. Hub genes that show
790 high degrees of connections to other networks genes are first identified, and their adjacent network neighbors (subnetworks) were extracted. All
791 genes in each CVD/T2D associated module are used as input and mapped onto each hub subnetwork to assess whether a hub subnetwork was
792 enriched for the genes in the input modules. The hubs whose subnetworks show significant enrichment of CVD/T2D module genes are defined as
793 potential key drivers.

794 S5 Fig. Scatter plots of the GWAS beta-values of variants mapped to the top 15 KDs. (A) Gene-variant mapping based on eQTLs only; (B) Gene-
795 variant mapping based on eQTLs and chromosomal distance. Percentage indicates the proportion of mapped variants with the same effect
796 direction between CVD and T2D. Statistical significance of the difference of the proportion from random expectation is determined by z-test.

797 S6 Fig. Expression changes in adipocyte differentiation state markers 3 days after the in vitro siRNA knockdown of Cav1. Statistical significance
798 of genes was determined by Student's t-test. N=3/group, mean \pm SEM, **p < 0.01.

799 S7 Fig. Visualization of CAV1 adipose subnetwork. Red color indicates significantly up-regulated genes (FDR < 1%) in *Cav1*^{-/-} mice, and blue
800 color indicates significantly down-regulated genes (FDR < 1%) in *Cav1*^{-/-} mice.

801 S1 Table. Summary of significant co-expression modules (FDR < 5%) associated with CVD or T2D.

802 S2 Table. Functional annotation terms of the significant co-expression modules.

803 S3 Table. List of previously reported genes associated with CVD, T2D, and intermediate metabolic traits related to CVD, T2D from DisGeNET
804 and GWAS Catalog.

805 S4 Table. Summary information of genome-wide association studies.

806 S5 Table. Data resources and references for co-expression networks.

807 S6 Table. Data resources and references for expression QTLs.

808 S7 Table. Data resources and references for gene-gene regulatory networks

809 **Figure Captions**

810 **Fig 1.** Framework of network-driven integrative genomics analyses. (A) Integration of genetics and functional genomics datasets to identify CVD
811 and T2D associated co-expression modules. The GWAS studies for CVD and T2D were derived from three independent cohorts representing three
812 ethnic populations: WHI (AA, EA, HA), FHS (EA), and JHS (AA). These independent datasets were supplemented with GWAS of coronary
813 artery disease from CARDIoGRAMplusC4D and T2D from DIAGRAM to increase power. We also curated a comprehensive list of tissue-specific
814 functional genomics datasets, including 2672 co-expression modules, human eQTLs of various tissues, and ENCODE based variants annotation.
815 The significant modules were identified by MSEA and Meta-MSEA, and then annotated to reveal shared pathways for CVD and T2D. In MSEA,
816 the co-expression modules were used to define data-driven gene sets each containing functionally related genes, tissue-specificity was determined
817 based on the tissue-origins of the human eQTLs, and ethnic specificity was determined based on the ethnicity of each GWAS cohort. (B)
818 Identification of disease key drivers and subnetworks. We utilized multi-tissue graphical networks to capture key drivers for disease associated co-
819 expression modules using wKDA, then prioritized KDs based on consistency and disease relevance of the subnetworks. (C) Validation of the top
820 key drivers and their subnetworks via intersection with known human CVD and T2D genes from DisGeNET and GWAS catalog, in vitro
821 adipocyte siRNA experiments, and cross-validation at both transcriptomic and genomic levels in the hybrid mouse diversity panels (HMDP).

822

823 **Fig 2.** Venn Diagrams of overlap in significant co-expression modules and functional categories between diseases and ethnicities. A) Count of
824 module overlaps by disease based on Meta-MSEA; B) Count of module overlaps for each disease by ethnicity based on MSEA of individual

825 studies. Co-expression modules captured in CARDIoGRAMplusC4D and DIAGRAM were not counted due to uncertain ethnic origin; C) Count
826 of independent functional category overlaps by disease based on results from Meta-MSEA in panel A.

827

828 **Fig 3.** Summary of 41 independent functional categories enriched in both CVD and T2D co-expression modules (Bonferroni-corrected $p < 0.05$
829 based on Fisher's exact test, number of direct overlapping genes > 5). Independent functional categories were defined as the categories with pair-
830 wise overlapping ratio $< 10\%$. Red and blue block indicates that the significant CVD or T2D co-expression modules identified from the study and
831 ethnicity origin are enriched for the particular functional category term. CAR+C4D: CARDIoGRAMplusC4D; M: mixed ethnicities; AA: African
832 Americans; HA: Hispanic Americans; EA: European Americans.

833

834 **Fig 4.** Subnetworks of the top 15 shared KDs orchestrate known genes for CVD, T2D, obesity and lipids. A) Fold enrichment of KD subnetwork
835 genes for known genes related to cardiometabolic traits reported in DisGeNET. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. B) Top KD
836 subnetworks with GWAS hits ($p < 1e-5$ as reported in GWAS Catalog) for cardiometabolic traits. KDs are large nodes. Edge color denotes tissue-
837 origin. Only high-confidence edges (those with weight score in the top 20%) are visualized.

838

839 **Fig 5.** Validation of *CAVI* subnetwork using *in vitro* siRNA knockdown (A) and *in vivo* knockout mouse model (B). A) Fold change of expression
840 level for *CAVI* subnetwork and negative control genes 2 days after *Cav1* knockdown using two siRNAs separately. Twelve *CAVI* neighbors were
841 randomly selected from the first and second level neighboring genes of *CAVI* in adipose network. Five negative controls were randomly selected

842 from the genes not connected to *CAVI* or its first level networks in adipose network. Statistical significance of genes was determined by linear
843 model, adjusting for batch effect and siRNA differences. N=612/siRNA group, mean \pm SEM, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. B) Overlap of
844 *CAVI* neighboring genes in the adipose tissue subnetwork at various distance levels with the differentially expressed genes in the gonadal adipose
845 tissue in *Cav1* knockout mice (N=3/group). Overlap p-value is determined by Fisher's exact test. *Overlap $p < 0.05$ after Bonferroni correction.

846

847 **Fig 6.** Associations of KDs and subnetworks with cardiometabolic traits in mice. (A) Association between KD expression and cardiometabolic
848 traits in adipose tissue from HF-HMDP (HF) and aorta tissue from atherogenic-HMDP (ATH) as determined by Pearson correlation. * $p < 0.05$;
849 ** $p < 0.05$ after Bonferroni correction for the KD number; *** $p < 0.05$ after Bonferroni correction for the number of KDs and traits. (B)
850 Transcriptomic-wide association of KD subnetworks and cardiometabolic traits in adipose tissue from HF-HMDP, and aorta tissue from
851 atherogenic-HMDP, as evaluated by MSEA. (C) Genome-wide association of KD subnetworks and cardiometabolic traits based on adipose eQTL
852 mapping in HF-HMDP, and aorta eQTL mapping in Atherogenic-HMDP, as determined by MSEA. $p < 0.05$, $p < 3.3 \times 10^{-3}$, and $p < 4.8 \times 10^{-4}$ correspond to
853 uncorrected and Bonferroni-corrected p-values (correcting for the number of KDs or for the number of KD and trait combinations).

854

855 **Tables**856 **Table 1.** Summary of top co-expression modules associated with CVD or T2D (FDR < 1% in Meta-MSEA, in column FDR_{meta})

Disease	Module ID	Tissue	Annotation	Gene No.	CAR+C4D/ DIAGRAM	JHS	FHS	WHI	WHI	WHI	P _{meta}	FDR _{meta}
					Mixed	AA	EA	EA	AA	HA		
CVD	4406	O1, O2, O5	NA	154	3.32E-10	NS	-	2.83E-02	4.41E-03	NS	5.73E-09	<0.01%
	4522	Adp, Lv, T	Signaling by FGFR mutants	2072	1.03E-04	1.62E-02	-	3.80E-02	5.53E-03	2.86E-02	3.39E-08	<0.01%
	4540	O4, O5	NA	1233	9.72E-04	NS	-	NS	1.50E-02	5.52E-04	5.07E-07	0.06%
	5242	Adr	Cholesterol Biosynthesis	306	4.19E-06	4.71E-02	-	NS	2.31E-02	NS	2.64E-06	0.08%
	4087	Adp, Dg	Carboxylic acid metabolic process	158	2.34E-06	NS	-	NS	8.63E-03	2.17E-02	4.24E-06	0.09%
	4019	Ly	Transmembrane transport of small molecules	2876	1.89E-03	4.46E-02	-	NS	NS	6.85E-04	7.91E-06	0.20%
	4941	O4, O5	Establishment of localization	908	8.97E-06	1.52E-02	-	NS	NS	3.94E-02	2.72E-06	0.21%
	5023	Ly	TCA cycle and respiratory electron transport	2890	NS	6.37E-05	-	1.53E-03	NS	1.50E-02	1.15E-05	0.22%
	blue	O2, O4	Cell cycle	657	1.08E-02	NS	-	NS	NS	1.77E-04	3.85E-06	0.30%
	5329	Adr	Biological oxidations	1028	NS	2.32E-02	-	5.01E-03	3.26E-02	2.26E-02	2.21E-05	0.35%
	124	O3, O4	NA	14	NS	1.48E-03	-	NS	7.05E-07	NS	4.86E-06	0.55%
	4656	O3, O4	Cellular protein complex assembly	371	NS	NS	-	NS	3.64E-03	2.27E-04	8.85E-06	0.67%
	4147	O5	NA	111	1.55E-02	2.06E-04	-	NS	8.85E-03	NS	5.72E-06	0.68%
	4989	Adr	Metabolism of amino acids and derivatives	453	1.86E-03	7.41E-03	-	NS	3.71E-04	NS	7.81E-05	0.82%
T2D	5323	Mn	NA	38	8.68E-04	NS	NS	2.25E-04	1.05E-03	NS	1.58E-07	0.02%
	5250	Adp, Dg, Mn	NA	37	4.78E-05	NS	NS	3.01E-02	3.46E-07	NS	4.32E-07	0.03%

4880	Mn	NA	141	8.96E-03	NS	1.18E-02	5.06E-04	NS	NS	1.61E-06	0.06%
6872	Mn	NA	119	NS	1.26E-03	7.44E-03	7.79E-03	NS	NS	1.26E-06	0.06%
4879	Ms	NA	376	3.18E-02	NS	5.88E-04	NS	2.66E-03	2.20E-03	1.19E-06	0.14%
6533	Mn	Cholesterol biosynthesis	48	NS	5.02E-03	NS	NS	NS	1.26E-06	1.06E-05	0.25%
6977	Bld, O3	NA	40	3.66E-02	NS	4.01E-05	NS	1.81E-02	4.05E-02	1.71E-06	0.39%
6675	Mn	Cholesterol biosynthesis	152	3.72E-03	3.35E-02	NS	NS	NS	2.06E-05	2.56E-05	0.52%
37	O2	NA	34	1.94E-03	5.53E-03	NS	NS	9.38E-04	NS	4.95E-06	0.57%
4302	Adp	NA	40	2.07E-03	NS	NS	4.80E-03	4.05E-06	NS	9.89E-06	0.71%
6690	Adr	Complement and coagulation cascades	641	1.93E-02	1.01E-04	NS	2.24E-02	NS	NS	1.36E-05	0.86%
4059	Dg	SLC mediated transmembrane transport	51	NS	3.05E-02	5.80E-03	NS	1.50E-02	NS	1.29E-05	0.86%
4937	Dg	Amino acid metabolic process	80	9.21E-03	NS	5.88E-03	NS	1.37E-03	NS	2.11E-05	0.89%
5059	Ve	TCA cycle and respiratory electron transport	164	7.31E-04	NS	2.74E-02	8.66E-04	NS	NS	6.64E-06	0.95%

857 Module IDs were randomly assigned IDs to co-expression modules. The annotation refers to the top functional category enriched in the co-
858 expression modules (Bonferroni-corrected $p < 0.05$ based on Fisher's exact test, number of direct overlapping genes > 5). Numbers in scientific
859 format were p-values from MSEA or Meta-MSEA analysis, and those reaching $FDR < 20\%$ in individual cohort analysis via MSEA (not the
860 FDR_{meta} in Meta-MSEA) are highlighted in bold. CAR+C4D: CARDIoGRAMplusC4D; Mixed: mixed ethnicities; JHS: Jackson Heart Study; FHS:
861 Framingham Heart Study; WHI: Women's Health Initiative; AA: African Americans; HA: Hispanic Americans; EA: European Americans; P_{meta}
862 and FDR_{meta} : p and FDR values from Meta-MSEA analysis across cohorts. Adp – adipose tissue; Adr - adrenal gland; Bld – Blood; Dg - digestive
863 tract; Lv – liver; Ly – lymphocyte; Ms – muscle; O1 – chromosomal distance mapping based on a 50kb window; O2 – ENCODE-based
864 Regulome SNPs; O3 – combining all tissue-specific eQTLs into a single multi-tissue eSNP set; O4 – merging eQTL sets with Regulome data; O5
865 – combined mapping (distance, eQTLs, ENCODE); T – thyroid gland; Ve – vascular endothelium.

866 **Table 2.** Summary of the 15 key drivers and their corresponding subnetworks shared by CVD and T2D

Key drivers	Gene name	Sub-net size	Tissues	P_{CVD}	FDR_{CVD}	P_{T2D}	FDR_{T2D}	No. of CVD module	No. of T2D module	Suggestive genetic effect direction (CVD/T2D)	Subnetwork function
<i>ACAT2</i>	Acetyl-CoA Acetyltransferase 2	192	Adp, Dg, Lv, Ms, T	1.24E-03	5.32%	5.37E-03	4.35%	6	7	uncertain	Cell cycle; Cholesterol biosynthesis
<i>ACLY</i>	ATP Citrate Lyase	129	Adp, Dg, Lv, Ms	5.96E-04	6.17%	5.78E-05	0.47%	5	6	consistent	Cholesterol biosynthesis; Steroid biosynthesis
<i>CAV1</i>	Caveolin 1	954	Adp, Adr, Art, Dg, Ms, T, Ve	1.24E-05	0.20%	3.96E-05	0.32%	7	4	consistent	Immune system; Focal adhesion
<i>COL6A2</i>	Collagen Type VI Alpha 2 Chain	294	Adp, Adr, Dg, Ms, T	2.47E-03	4.45%	4.97E-05	0.40%	2	1	consistent	Extracellular matrix
<i>COX7A2</i>	Cytochrome C Oxidase Subunit 7A2	152	Adp, Adr, Art, Bld, Dg, Lv, Ly	2.34E-04	3.79%	1.31E-04	1.85%	1	4	uncertain	Respiratory electron transport
<i>DBI</i>	Diazepam Binding Inhibitor	181	Adp, Art, Bld, Dg, Is, Lv, Ly, Ms	1.57E-03	7.70%	1.33E-02	6.75%	5	5	uncertain	Respiratory electron transport
<i>HMGCR</i>	3-Hydroxy-3-Methylglutaryl-CoA Reductase	75	Art, Dg, Lv, Ms	7.53E-03	9.09%	7.28E-03	4.87%	1	5	opposite	Cholesterol biosynthesis; Steroid biosynthesis
<i>IDII</i>	isopenentenyl-diphosphate delta isomerase 1	89	Adp, Art, Dg, Is, Lv, Ms, T	6.77E-03	8.95%	2.13E-03	3.46%	3	4	opposite	Cholesterol biosynthesis; Steroid biosynthesis
<i>IGF1</i>	insulin like growth	993	Adr, Ms	2.65E-03	5.37%	3.71E-04	1.20%	7	2	consistent	Immune system; Focal

factor 1										adhesion	
<i>MCAM</i>	melanoma cell adhesion molecule	183	Adp, Adr, Art, Ms, T	2.65E-03	7.16%	1.93E-03	5.22%	4	2	uncertain	Extracellular matrix
<i>MEST</i>	mesoderm specific transcript	132	Adp, Adr, Lv, Ms	1.66E-03	3.36%	6.84E-04	1.58%	4	2	uncertain	Fibroblast growth factor signaling
<i>MSMO1</i>	methylsterol monooxygenase 1	133	Adp, Art, Dg, Lv, Ms, T,	2.38E-03	7.70%	4.34E-05	0.63%	1	4	uncertain	Cholesterol biosynthesis; Steroid biosynthesis
<i>PCOLCE</i>	procollagen C-endopeptidase enhancer	307	Adp, Adr, Art, Hy, Lv, Ms	1.14E-03	6.17%	1.71E-06	0.03%	2	2	uncertain	Extracellular matrix
<i>SPARC</i>	secreted protein acidic and cysteine rich	482	Adp, Adr, Art, Dg, Lv, Ms, Ve	1.81E-03	9.63%	2.02E-03	8.18%	5	3	consistent	Extracellular matrix
<i>ZFP36</i>	ZFP36 ring finger protein	176	Adp, Adr, Art, Lv, Ly, Ms	1.42E-03	8.45%	1.64E-02	7.69%	3	3	uncertain	Hypoxia-inducible factors; CD40 signaling

867 P and FDR values were based on Meta-MSEA analysis of the KD subnetworks for enrichment of CVD or T2D GWAS signals across cohorts. The
868 subnetwork size indicates the number of neighboring genes directly connected to a KD when all the tissue-specific networks where the KD was
869 found are combined. No. of module columns indicate the number of CVD or T2D-associated co-expression modules from which each KD was
870 identified. Suggestive genetic effect direction was designated “consistent” or “opposite” if the proportion of variants having consistent or opposite
871 effect direction in CVD or T2D was statistically significant in either eQTL mapping or chromosomal distance mapping. Otherwise, “uncertain”
872 was called. Subnetwork function was annotated based on KEGG and Reactome databases. Adp – adipose tissue; Adr - adrenal gland; Art – artery;
873 Dg - digestive tract; Is – Islet; Hy – hypothalamus; Lv – liver; Ly – lymphocyte; Ms – muscle; T: thyroid gland; Ve: vascular endothelium.