

HUMAN GENETICS

Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases

Oscar Franzén,^{1,2*} Raili Ermel,^{3,4*} Ariella Cohain,^{1*} Nicholas K. Akers,¹ Antonio Di Narzo,¹ Husain A. Talukdar,⁵ Hassan Foroughi-Asl,⁵ Claudia Giambartolomei,⁶ John F. Fullard,⁶ Katyayani Sukhvasi,³ Sulev Kõks,³ Li-Ming Gan,⁷ Chiara Giannarelli,^{1,8} Jason C. Kovacic,⁸ Christer Betsholtz,^{9,10} Bojan Losic,¹ Tom Michoel,¹¹ Ke Hao,¹ Panos Roussos,^{1,6,12} Josefin Skogsberg,⁵ Arno Ruusalepp,^{2,3,4} Eric E. Schadt,¹ Johan L. M. Björkegren^{1,2,3,5†}

Genome-wide association studies (GWAS) have identified hundreds of cardiometabolic disease (CMD) risk loci. However, they contribute little to genetic variance, and most downstream gene-regulatory mechanisms are unknown. We genotyped and RNA-sequenced vascular and metabolic tissues from 600 coronary artery disease patients in the Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task study (STARNET). Gene expression traits associated with CMD risk single-nucleotide polymorphism (SNPs) identified by GWAS were more extensively found in STARNET than in tissue- and disease-unspecific gene-tissue expression studies, indicating sharing of downstream cis-/trans-gene regulation across tissues and CMDs. In contrast, the regulatory effects of other GWAS risk SNPs were tissue-specific; abdominal fat emerged as an important gene-regulatory site for blood lipids, such as for the low-density lipoprotein cholesterol and coronary artery disease risk gene *PCSK9*. STARNET provides insights into gene-regulatory mechanisms for CMD risk loci, facilitating their translation into opportunities for diagnosis, therapy, and prevention.

In 2012, cardiovascular disease accounted for 17.5 million deaths, nearly one-third of all deaths worldwide, and >80% (14.1 million) were from coronary artery disease (CAD) and stroke. CAD is preceded by cardiometabolic diseases (CMDs) such as hypertension, impaired lipid and glucose metabolism, and systemic inflammation (1, 2). Genome-wide association studies (GWAS) have identified hundreds of DNA variants associated with risk for CAD (3), hypertension (4), blood lipid levels (5), markers of plasma glucose metabolism (6–10), type 2 diabetes (6, 11), body mass index (12), rheumatoid arthritis (13), systemic lupus erythematosus (SLE) (14), ulcerative colitis (15), and Crohn's disease (16). However, identifying susceptibility genes responsible for these loci has proven difficult.

GWAS loci typically span large, noncoding, intergenic regions with numerous single-nucleotide polymorphisms (SNPs) in strong linkage disequilibrium. These regions are enriched in cis-regulatory elements (17) and expression quantitative trait loci (eQTLs) (18–20), suggesting that gene regulation is the principal mechanism by which

risk loci affect complex disease etiology. However, it is largely unknown whether this gene-regulatory effect includes one or several genes acting in one or multiple tissues and whether risk loci for different diseases share cis- and trans-gene regulation. A better understanding of gene regulation may also shed light on why known GWAS risk loci explain only ~10% of expected heritable variance in CMD risk (21). Possibly, multiple risk loci, acting through common cis- and trans-genes, contribute synergistically to heritability (22, 23).

In the Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task study (STARNET) (fig. S1), we recruited 600 well-characterized (table S1 and fig. S2) CAD patients; genotyped DNA (6,245,505 DNA variant calls with minor allele frequency >5%) (fig. S3); and sequenced RNA isolated from blood, atherosclerotic-lesion-free internal mammary artery (MAM), atherosclerotic aortic root (AOR), subcutaneous fat (SF), visceral abdominal fat (VAF), skeletal muscle (SKLM), and liver (LIV) (15 to 30 million reads per sample) (figs. S4 to S11 and table S2).

In total, ~8 million cis-eQTLs were identified, and nearly half were unique SNP-gene pairs (figs. S12 to S26 and tables S3 to S7). The STARNET cis-eQTLs were enriched in genetic associations established by GWAS for CAD, CMDs, and Alzheimer's disease (AD) (3–16, 24) (figs. S27 to S33) and were further enriched after epigenetic filtering (figs. S34 to S39). Of 3326 genome-wide significant-risk SNPs identified by GWAS to date (25), 2,047 (61%) had a matching cis-QTL in STARNET (Fig. 1A). Of the 54 lead risk SNPs verified in meta-analyses of CAD GWAS (3), 38 cis-eQTLs with a regulatory trait concordance score (RTC) >0.9 and at least one candidate gene were identified in STARNET (table S8 and fig. S27). Compared with large data sets of cis-eQTL isolated only from blood, cis-eQTLs across all tissues in STARNET matched >10-fold more CAD and CMD-related GWAS risk SNPs (Fig. 1B). STARNET cis-eQTLs isolated from CAD-affected tissues also matched several-fold more CAD and CMD-related GWAS risk SNPs than cis-eQTLs from corresponding tissues isolated from predominantly healthy individuals in the Genotype Tissue Expression (GTEx) study (18) (Fig. 1C). Thus, not all gene-regulatory effects of disease-risk SNPs are identifiable in blood or healthy tissues. This notion was further underscored by comparing the statistical significances of cis-eQTLs for GWAS risk SNPs in STARNET with corresponding associations in GTEx (Fig. 1D). In STARNET, gene fusions (table S9) and CAD-related loss of function mutations (table S10) were also detected.

The cis effects of disease-associated risk loci identified by GWAS are central for understanding downstream molecular mechanisms of disease. However, these cis-genes likely also affect downstream trans-genes. To identify possible trans effects, we ran a targeted analysis to call both cis- and trans-genes for lead risk SNPs identified by GWAS. After assigning cis-eQTLs for 562 risk SNPs for CAD, CMDs, and AD (3–16, 24), we used a causal inference test (26) to conservatively call causal correlations between the cis-genes and trans-genes by assessing the probability that an interaction was causal [SNP→cis-gene→trans-gene; false discovery rate (FDR) < 1%] and not reactive (SNP→trans-gene→cis-gene; $P > 0.05$) (26) (table S11). We found extensive sharing of cis- and trans-gene regulation by GWAS risk loci across tissues and CMDs. In CAD, 28 risk loci with at least one causal interaction (FDR < 1%, $P > 0.05$) had a total of 51 cis-genes and 1040 trans-genes. Of these, 26 risk loci, 37 cis-genes [including 27 key drivers (27)], and 994 trans-genes were connected in a main CAD regulatory gene network acting across all seven tissues

¹Department of Genetics and Genomic Sciences, The Icahn Institute for Genomics and Multiscale Biology Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. ²Clinical Gene Networks AB, Jungfrugatan 10, 114 44 Stockholm, Sweden. ³Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, University of Tartu, Biomeedikum, Ravila 19, 50411, Tartu, Estonia. ⁴Department of Cardiac Surgery, Tartu University Hospital, 1a Ludvig Puusepa Street, 50406 Tartu, Estonia. ⁵Division of Vascular Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Scheeles Väg 2, 171 77 Stockholm, Sweden. ⁶Division of Psychiatric Genomics, Department of Psychiatry and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. ⁷Cardiovascular and Metabolic Diseases, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Pepparedsleden 1, Mölndal, 431 83, Sweden. ⁸Cardiovascular Research Center Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. ⁹AstraZeneca-Karolinska Integrated CardioMetabolic Centre (ICMC), Karolinska Institutet, Novum, Blickagången 6, 141 57 Huddinge, Sweden. ¹⁰Department of Immunology, Genetics and Pathology Dag Hammarskjölds Väg 20, 751 85 Uppsala, Sweden. ¹¹Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Old College, South Bridge, Edinburgh EH8 9YL, UK. ¹²Department of Psychiatry, J. J. Peters VA Medical Center, Mental Illness Research Education and Clinical Center (MIRECC), 130 West Kingsbridge Road, Bronx, NY 10468, USA.

*These authors contributed equally to this work. †Corresponding author. Email: johan.bjorkegren@mssm.edu

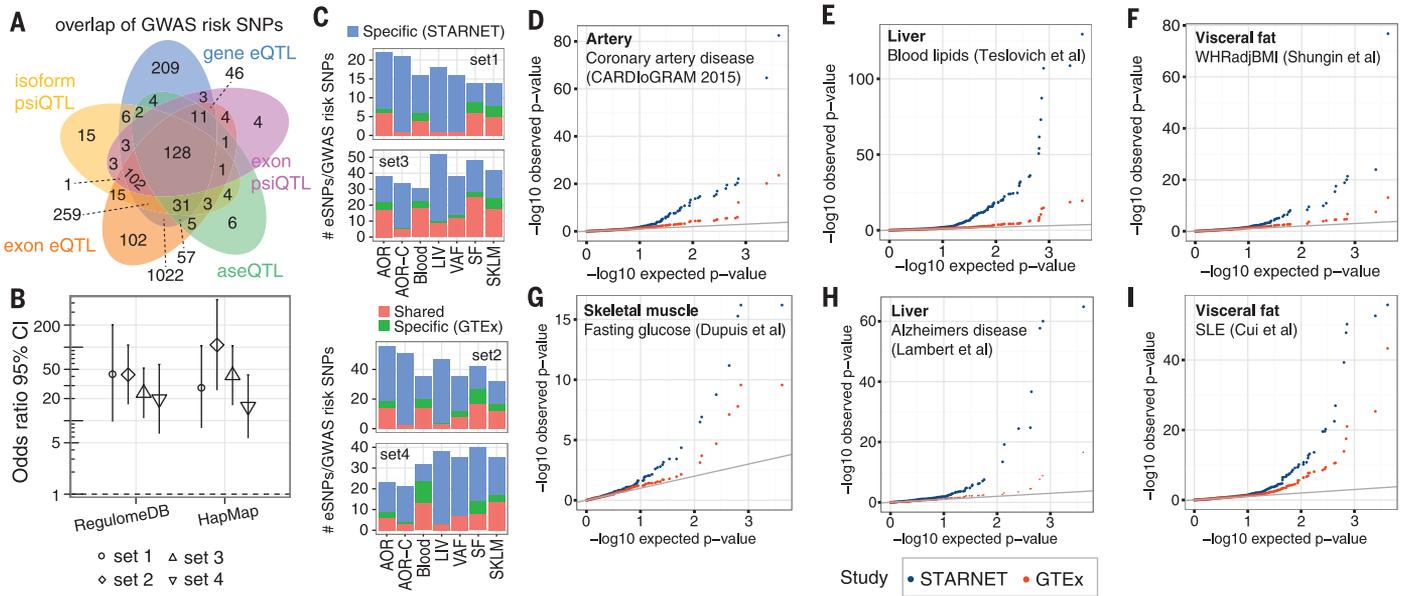


Fig. 1. QTLs and disease-associated risk SNPs identified by GWAS. (A) Venn diagram showing 2047 of 3326 disease-associated risk SNPs from the National Human Genome Research Institute GWAS catalog overlapping with at least one form of STARNET e/psi/aseQTLs. (B) Odds ratios that STARNET eQTLs coincide with CAD-associated risk SNPs (set 1, CARDIoGRAM-C4D, $n = 53$; set 2, CARDIoGRAM extended, $n = 150$) (3), blood lipids (set 3, $n = 35$) (5), and metabolic traits (set 4, $n = 132$) (6, 8, 10, 12) versus blood eQTLs from

RegulomeDB and HapMap. The y axis shows odds ratios. Error bars, 95% confidence intervals. (C) Stacked bar plots comparing tissue-specific eQTLs from STARNET and GTEx (18) coinciding with disease-associated risk SNPs in the same sets 1 to 4 as in (B). (D to I). Q-Q plots showing associations of tissue-specific STARNET (blue) and GTEx (18) (red) cis-eQTLs of disease-associated risk SNPs identified by GWAS for CAD (3) (D), blood lipids (5) (E), waist-hip ratio (12) (F), fasting glucose (6) (G), AD (24) (H), and SLE (14) (I).

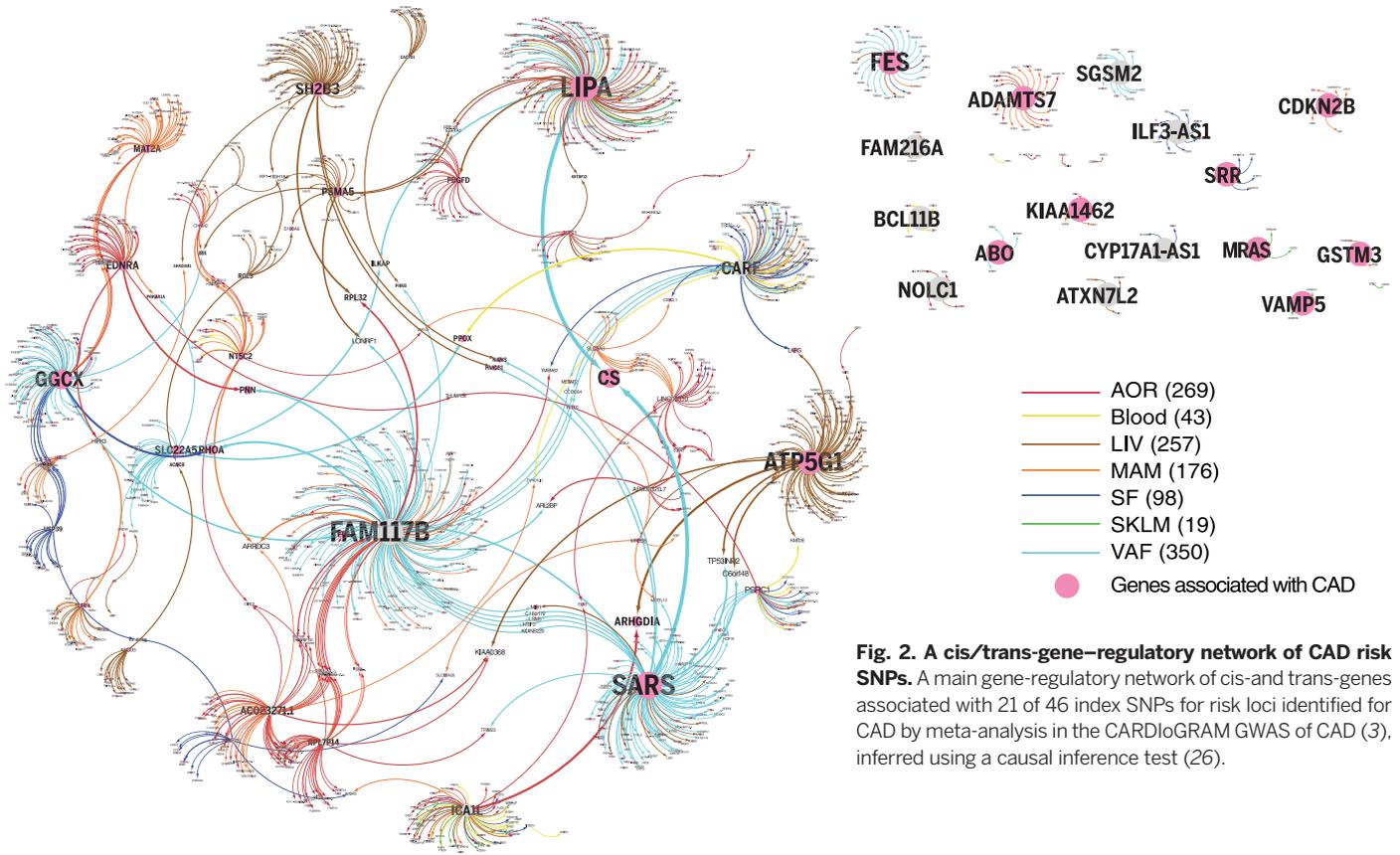


Fig. 2. A cis/trans-gene-regulatory network of CAD risk SNPs. A main gene-regulatory network of cis- and trans-genes associated with 21 of 46 index SNPs for risk loci identified for CAD by meta-analysis in the CARDIoGRAM GWAS of CAD (3), inferred using a causal inference test (26).

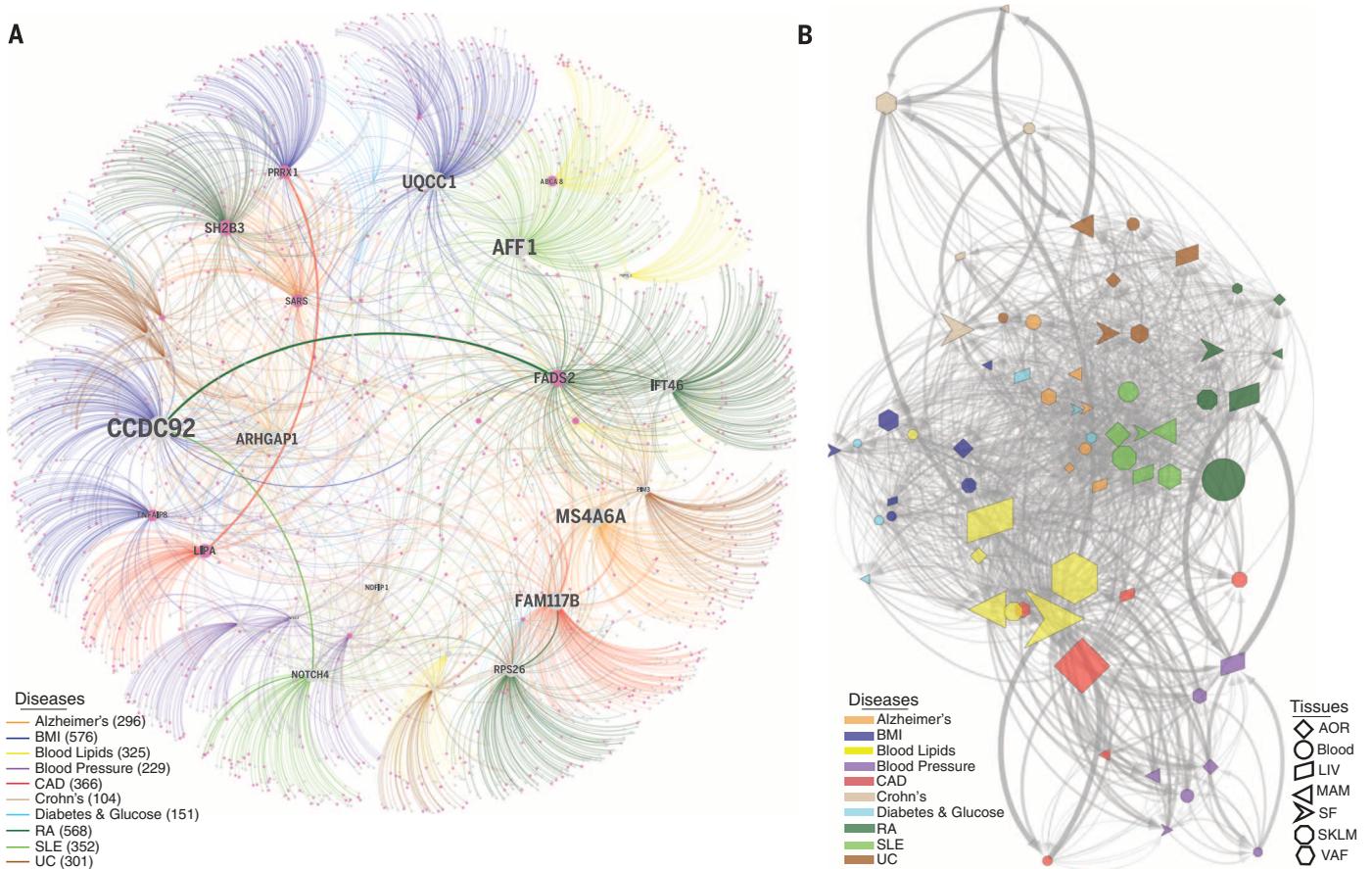


Fig. 3. Cis- and trans-gene regulation across CMDs and Alzheimer's disease. (A) A pan-disease risk SNP cis/trans-gene regulatory network. Thirty-six top key disease drivers, including 33 cis-genes for risk SNPs identified for CMDs including CAD and AD by GWAS (3–16, 24), were identified as having >100 downstream genes in any disease-specific network or belonging to the top five key drivers in the main regulatory gene network for each disease (table S11). Edge thickness reflects how frequent an edge is part of the shortest path between all pairs of network nodes. Node size reflects the number of downstream nodes in the network. RA, rheumatoid arthritis; UC, ulcerative colitis. (B) Cis- and

trans-gene regulation across disease-tissue pairs. Nodes represent unique disease-tissue pairs. Edges occur when a cis-gene in one node has downstream trans-genes present also in another node. Edge thickness defined as in (A). Node size reflects its centrality in the network: The position of the nodes in the network (i.e., layout) was derived from an edge-weighted spring layout algorithm. The “weight” is defined as the number of trans-genes that have a connection from the upstream node's cis-genes, normalized by the total number of trans-genes between two connecting nodes, with the result that highly connected nodes are positioned in the center of the network.

(Fig. 2). The trans-genes in this network were enriched with genes previously associated with CAD and atherosclerosis (Fisher's test, 1.54-fold; $P = 8 \times 10^{-10}$) (table S11). Sharing of cis/trans-genes downstream of complex disease risk loci also emerged for other CMDs and AD (3–16, 24) (fig. S40). In fact, we identified 33 cis-genes regulated by risk SNPs across all CMDs, including CAD and AD, acting as key drivers in a pan-disease cis/trans-gene regulatory network (Fig. 3A).

Among CMDs, cis/trans-genes of GWAS risk SNPs for blood lipid levels (5) emerged as central (Fig. 3B) where tissue-specific downstream effects were, besides LIV (46 cis- and 150 trans-genes), observed in the fat tissues (SF, 45 cis- and 372 trans-genes; VAF, 38 cis- and 465 trans-genes) (fig. S41 and table S11). Visceral abdominal fat examples included *ABCA8/ABCA5* (rs4148008) associated with 36 downstream trans-genes in VAF and HDL (high-density lipoprotein); *EVT5* (rs7515577) associated with 32 VAF trans-genes and total cholesterol; and *STARD3* (rs11869286)

associated with 7 VAF trans-genes and HDL. In addition, the cis-gene *TMEM258* (rs174546) with 22 trans-genes in abdominal fat surfaced as a parallel/alternative regulatory site of plasma low-density lipoprotein (LDL) to the proposed *FADS1,2,3* in LIV (5) (fig. S41). Other risk SNPs with VAF-specific cis-genes had few or even no trans-genes (fig. S41). For example, two risk SNPs—rs11206510 for CAD and rs12046679 for LDL cholesterol level (3, 5)—regulate *PCSK9* in VAF, not in LIV (Fig. 4, A and B). The VAF specificity of these eQTLs *PCSK9* was confirmed in an independent gene expression data set from morbidly obese patients (28) (Fig. 4C and fig. S30), suggesting that *PCSK9* is secreted from VAF into the portal vein to affect hepatic LDL receptor degradation, LDL plasma levels, and risk for CAD (29). Interestingly, and as previously suggested (30), we observed that STARNET patients in the upper, compared to the lower, 5th to 20th percentiles of waist-hip ratio (i.e., patients with and without “male fat”) had higher

levels of circulating *PCSK9* (Fig. 4D) and LDL/HDL ratio (Fig. 4E).

STARNET provides new insights into tissue-specific gene-regulatory effects of disease-associated risk SNPs identified by GWAS, as exemplified by abdominal fat for blood lipids, and will be a complementary resource for exploring GWAS findings moving forward. Furthermore, STARNET also revealed unexpected sharing of cis- and trans-genes downstream of risk loci for CMDs across both tissues and diseases. We anticipate that the identified cis/trans-gene regulatory networks will help elucidate the complex downstream effects of risk loci for common complex diseases, including possible epistatic effects that could shed light on the missing heritability of CMD risk. Given the detailed phenotypic data on STARNET patients, we can begin to identify how genetic variability interacts with environmental perturbations across tissues to cause pathophysiological alterations and complex diseases.

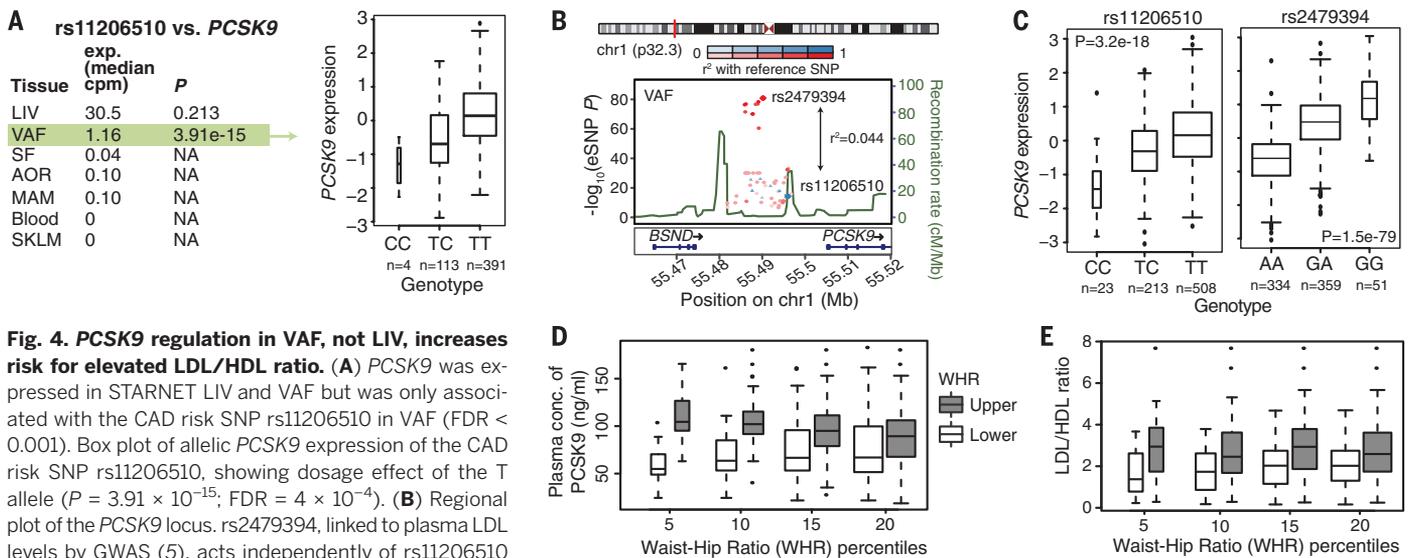


Fig. 4. PCSK9 regulation in VAF, not LIV, increases risk for elevated LDL/HDL ratio. (A) *PCSK9* was expressed in STARNET LIV and VAF but was only associated with the CAD risk SNP rs11206510 in VAF (FDR < 0.001). Box plot of allelic *PCSK9* expression of the CAD risk SNP rs11206510, showing dosage effect of the T allele ($P = 3.91 \times 10^{-15}$; FDR = 4×10^{-4}). (B) Regional plot of the *PCSK9* locus. rs2479394, linked to plasma LDL levels by GWAS (5), acts independently of rs11206510 as the lead eQTL of *PCSK9* expression in VAF. rs2479394 was not an eQTL of *PCSK9* in STARNET LIV. (C) Box plots of allelic *PCSK9* expression in VAF of rs11206510 and rs2479394 in a gene-tissue expression study of morbidly obese patients (fig. S29) (28). (D and E) Box plots of *PCSK9* levels (D) and ratios of LDL/HDL (E) in plasma isolated from the STARNET patients within the upper and lower 5th to 20th percentiles of waist-hip ratio (WHR) (*PCSK9*: 5th, $P = 8.0 \times 10^{-11}$; 10th, $P = 1.9 \times 10^{-11}$; 15th, $P = 5.9 \times 10^{-5}$; 20th, $P = 0.004$. LDL/HDL ratio: 5th, $P = 0.007$; 10th, $P = 0.001$; 15th, $P = 0.0005$; 20th, $P = 0.0009$).

REFERENCES AND NOTES

- G. K. Hansson, *N. Engl. J. Med.* **352**, 1685–1695 (2005).
- G. I. Shulman, *N. Engl. J. Med.* **371**, 2237–2238 (2014).
- M. Nikpay *et al.*, *Nat. Genet.* **47**, 1121–1130 (2015).
- G. B. Ehret *et al.*, *Nature* **478**, 103–109 (2011).
- T. M. Teslovich *et al.*, *Nature* **466**, 707–713 (2010).
- J. Dupuis *et al.*, *Nat. Genet.* **42**, 105–116 (2010).
- N. Soranzo *et al.*, *Diabetes* **59**, 3229–3239 (2010).
- R. J. Strawbridge *et al.*, *Diabetes* **60**, 2624–2634 (2011).
- A. K. Manning *et al.*, *Nat. Genet.* **44**, 659–669 (2012).
- P. An *et al.*, *Metabolism* **63**, 461–468 (2014).
- E. Zeggini *et al.*, *Nat. Genet.* **40**, 638–645 (2008).
- D. Shungin *et al.*, *Nature* **518**, 187–196 (2015).
- E. A. Stahl *et al.*, *Nat. Genet.* **42**, 508–514 (2010).
- Y. Cui, Y. Sheng, X. Zhang, *J. Autoimmun.* **41**, 25–33 (2013).
- C. A. Anderson *et al.*, *Nat. Genet.* **43**, 246–252 (2011).
- A. Franke *et al.*, *Nat. Genet.* **42**, 1118–1125 (2010).
- K. Musunuru *et al.*, *Nature* **466**, 714–719 (2010).
- K. G. Ardlie *et al.*, *Science* **348**, 648–660 (2015).
- H. Foroughi Asl *et al.*, *Circ. Cardiovasc. Genet.* **8**, 305–315 (2015).
- H. A. Talukdar *et al.*, *Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease Cell Syst* **2**, 196–208 (2016).
- P. M. Visscher, M. A. Brown, M. I. McCarthy, J. Yang, *Am. J. Hum. Genet.* **90**, 7–24 (2012).

- W. H. Wei, G. Hemani, C. S. Haley, *Nat. Rev. Genet.* **15**, 722–733 (2014).
- P. C. Phillips, *Nat. Rev. Genet.* **9**, 855–867 (2008).
- J. C. Lambert *et al.*, *Nat. Genet.* **45**, 1452–1458 (2013).
- D. Welter *et al.*, *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
- J. Millstein, B. Zhang, J. Zhu, E. E. Schadt, *BMC Genet.* **10**, 23 (2009).
- I. M. Wang *et al.*, *Mol. Syst. Biol.* **8**, 594 (2012).
- D. M. Greenawalt *et al.*, *Genome Res.* **21**, 1008–1016 (2011).
- K. Leander *et al.*, *Circulation* **133**, 1230–1239 (2016).
- S. Yusuf *et al.*, *Lancet* **366**, 1640–1649 (2005).

ACKNOWLEDGMENTS

The STARNET study was supported by the University of Tartu (SPIGVARENG to J.L.M.B.), the Estonian Research Council (ETF grant 8853 to A.R. and J.L.M.B.), the Astra-Zeneca Translational Science Centre-Karolinska Institutet (a joint research program in translational science, to J.L.M.B.), Clinical Gene Networks AB (CGN) as an SME of the FP6/FP7 EU-funded integrated project CVgenes@target (HEALTH-F2-2013-601456), the Leducq transatlantic networks, CAD Genomics (C.G., E.E.S., and J.L.M.B.), Sphingonet (C.B.), the Torsten and Ragnar Söderberg Foundation (C.B.), the Knut and Alice Wallenberg Foundation (C.B.), the American Heart Association (A14SFRN20840000 to J.C.K., E.E.S., and J.L.M.B.), the National Institutes of Health (NIH NHLBI R01HL125863 to J.L.M.B.; NIH NHLBI R01HL71207 to E.E.S.; R01AG050986 to P.R.; NIH

NHLBI K23HL11339 to C.G.; NIH NHLBI K08HL11330 to J.C.K.), and the Veterans Affairs (Merit grant BX002395 to P.R.). The DNA genotyping and RNA sequencing were in part performed by the SNP&SEQ technology platform at Science for Life Laboratory the National Genomics Infrastructure (NGI) in Uppsala and Stockholm supported by the Swedish Research Council (VR-RFI), the Knut and Alice Wallenberg Foundation, and Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). CGN has financially contributed to the STARNET study. J.L.M.B. is the founder and chairman of CGN. J.L.M.B., E.E.S., and A.R. are on the board of directors for CGN. J.L.M.B., T.M., and A.R. own equity in CGN and receive financial compensation from CGN. This work was supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. The STARNET data is accessible through the Database of Genotypes and Phenotypes (dbGAP).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/353/6301/827/suppl/DC1
Materials and Methods
Figs. S1 to S41
Tables S1 to S11
References (31–79)

25 October 2015; accepted 22 July 2016
10.1126/science.aad6970

Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases

Oscar Franzén, Raili Ermel, Ariella Cohain, Nicholas K. Akers, Antonio Di Narzo, Husain A. Talukdar, Hassan Foroughi-Asl, Claudia Giambartolomei, John F. Fullard, Katyayani Sukhvasi, Sulev Köks, Li-Ming Gan, Chiara Giannarelli, Jason C. Kovacic, Christer Betsholtz, Bojan Losic, Tom Michael, Ke Hao, Panos Roussos, Josefin Skogsberg, Arno Ruusalepp, Eric E. Schadt and Johan L. M. Björkegren

Science **353** (6301), 827-830.
DOI: 10.1126/science.aad6970

Genetic variation and coronary artery disease

Most genetic variants lie outside protein-coding genes, but their effects, especially in human health, are not well understood. Franzén *et al.* examined gene expression in tissues affected by coronary artery disease (CAD). They found that individuals with loci that have been associated with CAD in genome-wide analyses had different patterns of tissue-specific gene expression than individuals without these genetic variants. Similarly, tissues not associated with CAD did not have CAD-like expression patterns. Thus, tissue-specific data can be used to dissect the genetic effects that predispose individuals to CAD.

Science, this issue p. 827

ARTICLE TOOLS

<http://science.sciencemag.org/content/353/6301/827>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2016/08/17/353.6301.827.DC1>

REFERENCES

This article cites 89 articles, 12 of which you can access for free
<http://science.sciencemag.org/content/353/6301/827#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)