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,7 Hassan Foroughi-Asl,5 Claudia Giambartolomei,6 John F. Fullard,4 Katayani Sukhavasti,5 Sulev Kõks,2 Li-Ming Gan,7 Chiara Giannarelli,1,8 Jason C. Kovacic,8 Christer Betsholtz,1,9,10 Bojan Losic,1 Tom Michoel,1 Ke Hao,1 Panos Roussos,1,6,12 Josefin Skogsberg,5 Arno Ruusalepp,2,3,4 Eric E. Schadt,1 Johan L. M. Björkergren1,2,3,5

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 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 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 3,326 genome-wide significant-risk SNPs identified by GWAS to date (25), 2,047 (61%) had a matching cis-eQTL 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 (28) (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.
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 eQTLs. (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. 2). The trans-genes in this network were enriched with genes previously associated with CAD and atherosclerosis (Fisher’s test, 1.54-fold; $P = 8 \times 10^{-30}$) (table S11). Shar...
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 × 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 × 10⁻¹¹; 10th, P = 1.9 × 10⁻¹¹; 15th, P = 5.9 × 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
29. K. Leander et al., Circulation 133, 1230–1239 (2016).

ACKNOWLEDGMENTS
The STARnet study was supported by the University of Tartu (SPIGVAREJG 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-F2-2013-60456), the Leeduq transatlantic networks. CAD Genomics (C.G., E.E.S., and J.L.M.B.), Sphingonet (C.B.), the Torsten and Ragnar Siderberg Foundation (C.B.), the Knut and Alice Wallenberg Foundation (C.B.), the American Heart Association (A145RFN20840000 to J.C.K., E.E.S., and J.L.M.B.), the National Institutes of Health (NIH NHHLBI RO1HL25863 to J.L.M.B.; NIH NHLHLBI RO1HL73107 to E.E.S., RO1AG050986 to P.R.; NIH NHLHLBI K23HL11339 to C.G.; NIH NHHLBI RO1HL113310 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-RF1), 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/cgi/content/353/6310/827/suppl/DC1
Materials and Methods
Figs. S1 to S41
Tables S1 to S11
References (31–89)
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


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