Edinburgh Research Explorer

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

Cardiometabolic Risk Loci Share Downstream *Cis*- and *Trans*-Gene Regulation Across Tissues and Diseases

The Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) Study

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 Sukhavasi[3], Sulev Köks[3], Li-Ming Gan[7], Chiara Gianarelli[1, 8], Jason C. Kovacic[8], Christer Betsholtz[5, 9], Bojan Losic[1], Tom Michoel[10], Ke Hao[1], Panos Roussos[1,6,11], Josefin Skogsberg[5], Arno Ruusalepp[2,3,4], Eric E. Schadt[1] and Johan L.M. Björkegren[1,2,3,5]

[1]Department of Genetics & Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York 10029, NY, USA

[2]Clinical Gene Networks AB, Jungfrugatan 10, 114 44 Stockholm, Sweden

[3]Department of Pathophysiology, Institute of Biomedicine and Translation Medicine, University of Tartu, Biomeedikum, Ravila 19, 50411, Tartu, Estonia

[4]Department of Cardiac Surgery, Tartu University Hospital, 1a L. Puusepa St., 50406 Tartu, Estonia

[5]Division of Vascular Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Scheeles väg 2, 171 77 Stockholm, Sweden

[6]Division of Psychiatric Genomics, Department of Psychiatry and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York 10029, NY, USA

[7]Cardiovascular and Metabolic Diseases, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Pepparedsleden 1, Mölndal, 431 83, Sweden

[8]Cardiovascular Research Centre, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York 10029, NY, USA

[9]Department of Immunology, Genetics and Pathology Dag Hammarskjölds Väg 20, 751 85 Uppsala, Sweden

[10]Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Old College, South Bridge, Edinburgh EH8 9YL, UK

[11]Department of Psychiatry, JJ Peters VA Medical Center, Mental Illness Research Education and Clinical Center (MIRECC), JJ Peters VA Medical Center, 130 West Kingsbridge Road, Bronx, NY 10468, USA

[*]Shared first authorship. Correspondence to: johan.bjorkegren@mssm.edu

Abstract

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 STARNET study. Gene expression traits associated with CMD risk 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 SNP were tissue-specific; abdominal fat emerged as an important gene-regulatory site for blood lipids, such as for the LDL-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.

One Sentence Summary

RNA-seq of vascular and metabolic tissues of coronary artery disease patients reveals *cis-* and *trans*-effects in disease

.

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 (*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, 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) (15–30 million read depth, figs. S4-S11, table S2).

In total, ~8 million *cis*-eQTLs were identified, and nearly half were unique SNP-gene pairs (figs. S12-S26, tables S3-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-S33) and were further enriched after epigenetic filtering (figs. S34-S39). Of 3,326 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, fig. S27). Compared to large datasets 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 GTEx (*18*) (Fig. 1C). Thus, several gene-regulatory effects of disease risk SNPs appear not to be 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 used a causal inference test (*26*) to conservatively call both *cis*- and *trans*-genes for lead risk SNPs identified by GWAS. After assigning *cis*-eQTLs for 562 risk SNPs for CMDs and AD (*3-16, 24*), we sought causal correlations between the *cis*-genes and *trans*-genes by assessing the probability that an interaction was causal (*cis*-SNP→*cis*-gene→*trans*-gene, FDR<1%) and not reactive (*cis*-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 *cis*-gene (FDR <1%) 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 7 tissues (Fig. 2). This network was enriched in genes associated with CAD and

atherosclerosis (Fisher's test, 1.54-fold, *P*=8E-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, common key driver genes regulated by risk SNPs across all CMDs, including CAD and AD, formed a pan-disease regulatory *cis/trans*-gene network in which 33/36 *cis*-genes were identified as key disease drivers (Fig. 3A).

Across CMDs and AZ, *cis*/*trans*-genes of GWAS risk SNPs for blood lipid levels (*5*) emerged as central (Fig. 3B) where beside LIV (n=46/150 *cis*/*trans*-genes), fat tissues harbored many downstream genes (45/372 *cis*/*trans*-genes in SF and 38/465 in VAF (fig. S41, table S11). Abdominal fat (VAF) examples included *ABCA8/ABCA5* (rs4148008) associated with 36 downstream *trans*-genes in VAF and HDL; *EVI5* (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 LDL to the proposed *FADS*-1,2,3 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. 4A, B). The VAF-specificity of these eQTLs were confirmed in an independent gene expression dataset from patients with morbidly obesity (*28*) (Fig. 4C, fig. S30) implicating 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 did observe that STARNET patients in the upper, compared to the lower 5th-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).

Thus, STARNET provides new insights into tissue-specific gene-regulatory effects of disease-associated risk SNPs identified by GWAS, exemplified with abdominal fat for blood lipids. We also detected 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. Thus, the STARNET dataset is a complementary resource for other studies to leverage the initial findings by GWAS, particularly of CAD and CMDs.

References and Notes

1.   G. K. Hansson, Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med* **352**, 1685-1695 (2005).

2.   G. I. Shulman, Ectopic fat in insulin resistance, dyslipidemia, and cardiometabolic disease. *N Engl J Med* **371**, 2237-2238 (2014).

3.   M. Nikpay *et al.*, A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* **47**, 1121-1130 (2015).

4.   G. B. Ehret *et al.*, Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-109 (2011).

5.   T. M. Teslovich *et al.*, Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713 (2010).

6.   J. Dupuis *et al.*, New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet.* **42**, 105-116 (2010).

7.   N. Soranzo *et al.*, Common variants at 10 genomic loci influence hemoglobin $A_1(C)$ levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229-3239 (2010).

8.   R. J. Strawbridge *et al.*, Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624-2634 (2011).

9.   A. K. Manning *et al.*, A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet.* **44**, 659-669 (2012).

10.   P. An *et al.*, Genome-wide association study identifies common loci influencing circulating glycated hemoglobin (HbA1c) levels in non-diabetic subjects: The Long Life Family Study (LLFS). *Metabolism* **63**, 461-468 (2014).

11.   E. Zeggini *et al.*, Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* **40**, 638-645 (2008).

12.   D. Shungin *et al.*, New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196 (2015).

13.   E. A. Stahl *et al.*, Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet.* **42**, 508-514 (2010).

14. Y. Cui, Y. Sheng, X. Zhang, Genetic susceptibility to SLE: Recent progress from GWAS. *J Autoimmun.* **41**, 25-33 (2013).

15. C. A. Anderson *et al.*, Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet.* **43**, 246-252 (2011).

16. A. Franke *et al.*, Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet.* **42**, 1118-1125 (2010).

17. K. Musunuru *et al.*, From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714-719 (2010).

18. The GTEx Consortium, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).

19. H. Foroughi Asl *et al.*, Expression quantitative trait Loci acting across multiple tissues are enriched in inherited risk for coronary artery disease. *Circ Cardiovasc Genet.* **8**, 305-315 (2015).

20. H. A. Talukdar *et al.*, Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease. *Cell Syst* **2**, 196-208 (2016).

21. P. M. Visscher, M. A. Brown, M. I. McCarthy, J. Yang, Five years of GWAS discovery. *Am J Hum Genet.* **90**, 7-24 (2012).

22. W. H. Wei, G. Hemani, C. S. Haley, Detecting epistasis in human complex traits. *Nat Rev Genet.* **15**, 722-733 (2014).

23. P. C. Phillips, Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet.* **9**, 855-867 (2008).

24. J. C. Lambert *et al.*, Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* **45**, 1452-1458 (2013).

25. D. Welter *et al.*, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001-1006 (2014).

26. J. Millstein, B. Zhang, J. Zhu, E. E. Schadt, Disentangling molecular relationships with a causal inference test. *BMC Genet.* **10**, 23 (2009).

27. I. M. Wang *et al.*, Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers. *Mol. Syst. Biol.* **8**, 594 (2012).

28. D. M. Greenawalt *et al.*, A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res.* **21**, 1008-1016 (2011).

29. K. Leander *et al.*, Circulating Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) Predicts Future Risk of Cardiovascular Events Independently of Established Risk Factors. *Circulation* **133**, 1230-1239 (2016).

30.   S. Yusuf *et al.*, Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet* **366**, 1640-1649 (2005).

Supplementary online materials:

Material and Methods
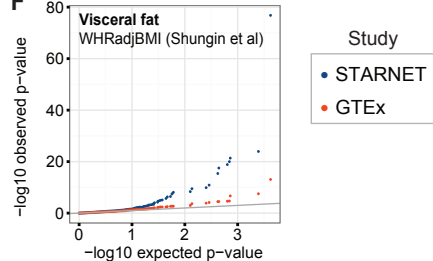
Figure S1-S41

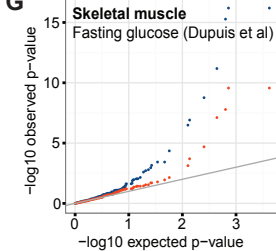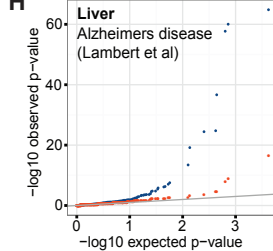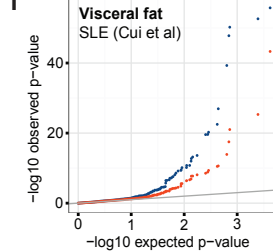Tables S1-S11

Figure Legends

Fig. 1. STARNET QTLs and disease-associated risk SNPs identified by GWAS. (**A**) Venn diagram showing 2,047/3,326 disease-associated risk SNPs from the NHGRI 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 isolated 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 1–4 as in (**B**). (**D-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**).
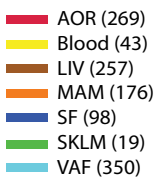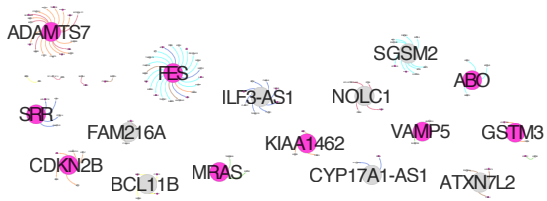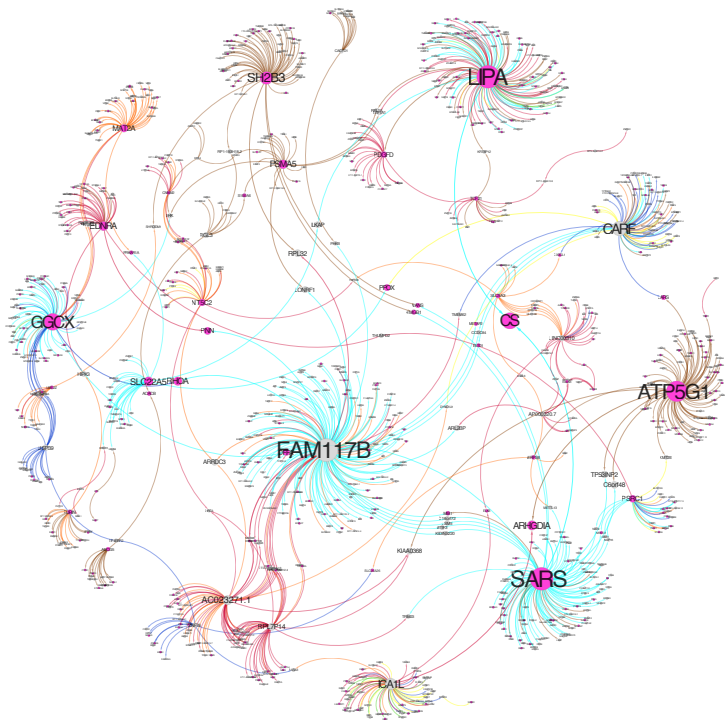
Fig. 2. A *cis/trans* gene-regulatory network of CAD risk SNPs. A main gene-regulatory network of *cis*-and *trans*-genes associated with 21/46 index SNPs for risk loci identified for CAD by meta-analysis in the CARDIoGRAM GWAS of CAD (*3*) inferred using a causal inference test (*26*).

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 5 key drivers in the main regulatory gene network for each disease (table S11). Node (gene) and edge color indicate disease belonging. Edge thickness represents how frequent an edge is the shortest path between all pairs of network nodes. Node size reflects the
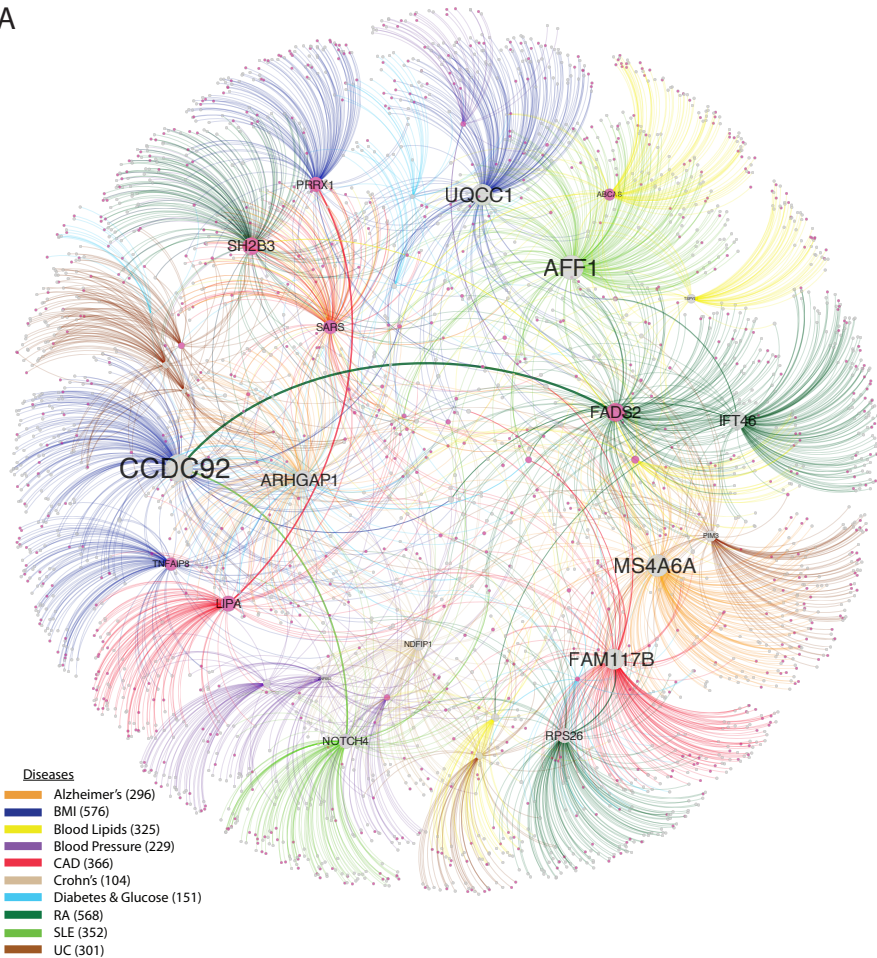
number of downstream nodes in the network. RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; UC, ulcerative colitis. (**B**) *cis* and *trans* gene regulation across disease/tissue pairs. Nodes represent unique disease-tissue pairs. Edges occur when a *cis*-gene in one node have downstream *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 — resulting in that highly connected nodes are positioned in the center of the network.

Fig. 4 *PCSK9* regulation in VAF, not liver, increases risk for elevated LDL/HDL ratio. (**A**) *PCSK9* was expressed in STARNET LIV and VAF but 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=3.91e-15; FDR=4e-04). (**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*). 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-20th percentile of waist-hip ratio (WHR) (PCSK9; 5th, P=8.0e-11; 10th, P=1.9e-11; 15th, P=5.9e-05; 20th, P=0.004: LDL/HDL ratio; 5th, P=0,007; 10th, P=0.001; 15th, P=0.0005; 20th, P=0.0009).

AOR (269)	
Blood (43)	
LIV (257)	
MAM (176)	
SF (98)	
SKLM (19)	
VAF (350)	

A

Diseases
- Alzheimer's (296)
- BMI (576)
- Blood Lipids (325)
- Blood Pressure (229)
- CAD (366)
- Crohn's (104)
- Diabetes & Glucose (151)
- RA (568)
- SLE (352)
- UC (301)

B

Diseases
- Alzheimer's
- BMI
- Blood Lipids
- Blood Pressure
- CAD
- Crohn's
- Diabetes & Glucose
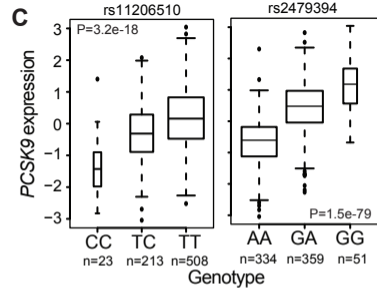- RA
- SLE
- UC

Tissues
- AOR
- Blood
- LIV
- MAM
- SF
- SKLM
- VAF
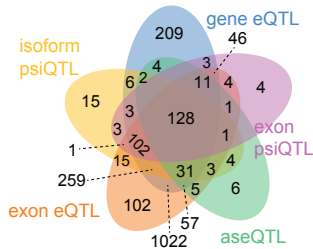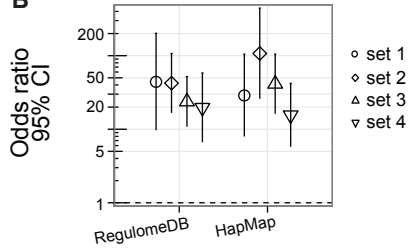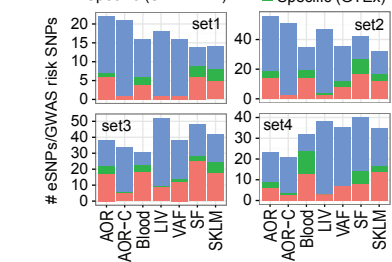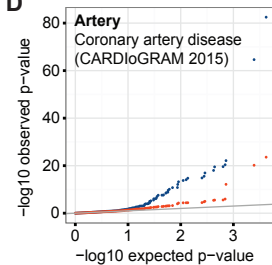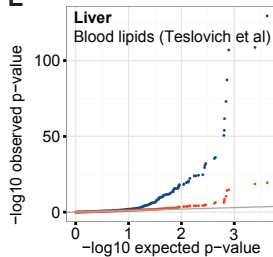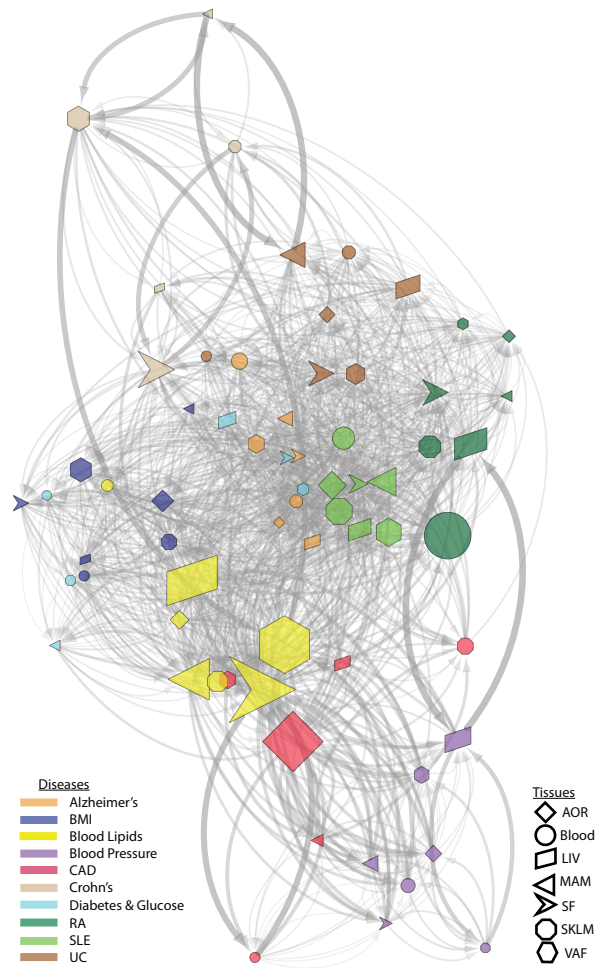
A rs11206510 vs. *PCSK9*

Tissue	exp. (median cpm)	*P*
LIV	30.5	0.213
VAF	1.16	3.91e-15
SF	0.04	NA
AOR	0.10	NA
MAM	0.10	NA
Blood	0	NA
SKLM	0	NA

PCSK9 expression

Genotype: CC (n=4), TC (n=113), TT (n=391)

B chr1 (p32.3)

chr1

r^2 with reference SNP

VAF

rs2479394

r^2=0.044

rs11206510

-log₁₀(eSNP *P*)

Recombination rate (cM/Mb)

BSND — *PCSK9*

Position on chr1 (Mb): 55.47 55.48 55.49 55.5 55.51 55.52

C

rs11206510   rs2479394

PCSK9 expression

P=3.2e-18

Genotype: CC (n=23), TC (n=213), TT (n=508)

P=1.5e-79

Genotype: AA (n=334), GA (n=359), GG (n=51)

D

Plasma conc. of PCSK9 (ng/ml)

WHR: Upper, Lower

Waist-Hip Ratio (WHR) percentiles: 5 10 15 20

E

LDL/HDL ratio

Waist-Hip Ratio (WHR) percentiles: 5 10 15 20