Additional candidate genes for human atherosclerotic disease identified through annotation based on chromatin organization

Short title: Atherosclerosis genes and chromatin organization

Saskia Haitjema MD [1*], Claartje A. Meddens MD[2,3*], Sander W. van der Laan MSc[1], Daniel Kofink MSc[4], Magdalena Harakalova MD PhD[4,5], Vinicius Tragante PhD[1], Hassan Foroughi Asl PhD[6], Jessica van Setten PhD[1,5], Maarten M. Brandt MSc[7,8], Joshua C. Bis PhD[9], Christopher O'Donnell MD PhD[10,11] on behalf of the CHARGE Consortium, the METASTROKE Group of the International Stroke Genetics Consortium, Caroline Cheng PhD[7,8], Ivana Isgum PhD[12], Imo E. Hoefer MD PhD[13], Johannes Waltenberger MD PhD[14], Erik Biessen PhD[15], J. Wouter Jukema PhD[16], Pieter A.F.M. Doevendans MD PhD[4], Edward E.S. Nieuwenhuis MD PhD[2], Jeanette Erdmann MD PhD[17,18,19], Johan L. M. Björkegren MD PhD[6,20,21], Gerard Pasterkamp MD PhD[1,13], Folkert W. Asselbergs MD PhD[4,22,23], Hester M. den Ruijter PhD[1], Michal Mokry MD PhD[2,3,24]

[1] Laboratory of Experimental Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands

[2] Department of Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands

[3] Regenerative Medicine Center Utrecht, University Medical Center Utrecht, Utrecht, The Netherlands

[4] Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands

[5] Netherlands Heart Institute, Utrecht, the Netherlands

[6] Vascular Biology Unit, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

[7] Department of Nephrology & Hypertension, University Medical Center Utrecht, Utrecht, The Netherlands

[8] Department of Cardiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

[9] Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, United State of America

[10] Department of Cardiology, Boston Veterans Administration Healthcare, West Roxbury, Massachusetts, United States of America

[11] NHLBI's Framingham Heart Study, Framingham, Massachusetts, United States of America

[12] Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

[13] Laboratory of Clinical Chemistry and Haematology, University Medical Center Utrecht, Utrecht, The Netherlands

[14] Department of Cardiovascular Medicine, Universitätsklinikum Münster, Münster, Germany

[15] Department of Pathology, Maastricht University Medical Center, Maastricht, The Netherlands

[16] Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands

[17] Institute for Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany

[18] DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, Lübeck, Germany

[19] University Heart Center Lübeck, Lübeck, Germany

[20] Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, United States of America

[21] Department of Physiology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

[22] Durrer Center for Cardiogenetic Research, Netherlands Heart Institute, Utrecht, the Netherlands

[23] Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom

[24] Epigenomics facility, University Medical Center Utrecht, Utrecht, The Netherlands

*Authors contributed equally

Corresponding authors:

Hester M. den Ruijter

Laboratory of Experimental Cardiology

University Medical Center Utrecht

PO box 85500

3508 GA Utrecht

The Netherlands

T: +31 88 755 7654

E: h.m.denruijter-2@umcutrecht.nl

Michal Mokry

Wilhelmina Children's Hospital

Dr. Michal Mokry, Huispostnummer KC.03.046.0

PO Box 85090

3508 AB Utrecht

The Netherlands

T: +31 88 756 7697

E: m.mokry@umcutrecht.nl

Journal subject terms: Epigenetics; Gene Expression and Regulation; Atherosclerosis; Coronary Artery Disease; Ischemic Stroke

Background: As genome-wide association efforts, such as CARDIoGRAM and METASTROKE, are ongoing to reveal susceptibility loci for their underlying disease: atherosclerotic disease, identification of candidate genes explaining the associations of these loci has proven the main challenge. Many disease susceptibility loci co-localize with DNA regulatory elements, which influence gene expression through chromatin interactions. Therefore, the target genes of these regulatory elements can be considered candidate genes. Applying these biological principles, we used an alternative approach to annotate susceptibility loci and identify candidate genes for human atherosclerotic disease based on circular chromosome conformation capture followed by sequencing (4C-seq).

Methods and Results: In human monocytes and coronary endothelial cells we generated 63 chromatin interaction datasets for 37 active DNA regulatory elements that co-localize with known susceptibility loci for coronary artery disease (CARDIoGRAMplusC4D) and large artery stroke (METASTROKE). By 4C-seq we identified a physical 3D interaction with 326 candidate genes expressed in at least one of these cell types, of which 294 have not been reported before. We highlight 16 genes based on expression quantitative trait loci.

Conclusions: Our findings provide additional candidate-gene annotation for 37 disease susceptibility loci for human atherosclerotic disease that are of potential interest to better understand the complex pathophysiology of cardiovascular diseases.

Key words: Epigenetics; Gene Expression and Regulation; Atherosclerosis; Coronary Artery Disease; Ischemic Stroke

INTRODUCTION

Atherosclerosis is a chronic inflammatory disease of the lipid-rich vascular wall that underlies many cardiovascular diseases (CVD)(1). A large part of the disease burden of atherosclerosis can be traced back to coronary artery disease (CAD) and large artery stroke (LAS). Genome-wide association studies (GWAS) have helped to unravel the complex genomic background of these diseases, currently explaining about 10% of heritability(2,3). The current approach is to annotate a novel susceptibility locus with the gene at the nearest genomic position. Some alternative strategies also take into account gene expression or protein-protein interactions(4,5). A recent effort employing these bioinformatics-based approaches resulted in 98 new candidate genes for CAD(6). In the last few years, the evidence that variants identified by GWAS also contribute to the disease pathogenesis by affecting the regulatory DNA sequences they reside in is growing(7–9). These genetic variants may affect the activity of the DNA regulatory elements (DRE) and, under specific circumstances, lead to dysregulation of gene expression. This is mediated by long range 3D chromatin-chromatin interactions where the regulated candidate genes can be located up to ~1 MB away(10–12) – a distance much larger than is normally used to annotate candidate genes in GWAS. These candidate genes can be identified by capturing the physical chromatin-chromatin interaction between a known disease susceptibility locus and the promoter of the gene(s) it presumably regulates(13). Here we systematically apply this principle (study design is summarized in Fig. 1) to variants identified by large meta-analyses of GWAS for CAD and LAS; altogether assaying 47 previously identifiedsusceptibility loci(2,3). Atherosclerotic disease starts in the endothelial lining of the affected arteries and involves attraction and proliferation of monocytes(14). Therefore, we studied 37 loci that co-localize with active DRE in human monocytes and/or in cardiac endothelial cells. We used circular chromosome conformation capture sequencing (4C-seq) to identify candidate genes based on their physical interaction with one of the active DRE.

METHODS

Cell culture

Primary commercially available human cardiac endothelial cells (CEC) that were isolated by enzymatic detachment (Lonza Clonetics™) were cultured in RPMI-1640 with 10% FCS and standard supplements. Cells were harvested for 4C template preparation by trypsinisation at 60-80% confluence.

Monocyte isolation

Human peripheral blood was collected from a healthy donor in sodium-heparin tubes. Peripheral Blood Mononuclear Cells (PBMCs) were isolated by Ficoll-Paque gradient centrifugation. PMBCs were incubated with magnetic CD14+-microbeads (Milteny, order nr. 130-050-201) according to manufacturer's manual. Thereafter cells were magnetically separated by the AutoMACS™ Separator, the positive fraction (monocytes) was used for 4C template preparation.

Circular Chromosome Conformation Capture-Template preparation

The 4C template was prepared as previously described(13). Summarized, $10x10^6$ cells were used per cell type (monocytes and CEC). Cells were crosslinked in 2% formaldehyde. After chromatin isolation, the chromatin was digested with *DpnII* (NEB, # R0543L). Digestion was stopped through heat inactivation of the restriction enzyme. Samples were diluted and ligated by T4 DNA ligase. The second digestion was carried out with CviQI (NEB, #R069S) and inactivated by phenol:chloroform extraction. The chromatin was diluted, for the final T4 ligation and the chromatin was purified. The quality of digestion and ligation was assessed on agarose gels.

Viewpoint selection and primer design

All SNPs from table 1 and 2 and the young-CAD SNP (rs16986953) from the CARDIoGRAMplusC4D paper(2) (n=47) and the two replicated SNPs (rs2383207 and rs2107595) from METASTROKE(3) were considered for viewpoint design (Supplemental table 8). When the SNPs within the susceptibility locus were less than 15,000 bp apart (e.g. rs12740374, rs602633 and rs599839 in the *SORT1* region), only one SNP was selected as a viewpoint. Susceptibility loci that overlapped with active DRE were identified through FAIRE, the presence of H3K4Me1, H3K4Me3, H3K27Ac, H3K4Me2 or H3K9Ac, EP300 or CTCF binding sites or DNase hypersensitivity sites (Supplemental table 9). DRE falling within the susceptibility locus coordinates were considered overlapping with the susceptibility locus. The primers were designed as was described previously(13). Primer sequences are listed in Supplemental table 8. In summary,

primers were designed in a window of 5 kb up- and downstream from the associated SNP. Forward and reverse primers were designed at least 300 bp apart. Forward (reading) primers were designed on top of the first restriction enzyme site. The reverse (non-reading) primer was designed close to (max 100bp away from) the second restriction enzyme site. In case no primer pair could be designed within the initial window, the window was extended 5 kb up- and downstream (n=8). In the case of rs1561198 this did not result in a suitable primer, so a primer pair that was 299bp apart was selected for this viewpoint.

Circular Chromosome Conformation Capture- Sequencing (4C-seq) library preparation

4C-sequencing library preparation was performed as described previously(13), with minor adaptations in order to make the protocol compatible with the large number of viewpoints: the PCR of 4C template was performed with 600 ng (monocytes) or 1,6 µg (coronary endothelial cells) of 4C template per reaction. 8 to 10 primer pairs were multiplexed in the initial PCR reaction (primer sequences are listed in Supplemental table 8). Primers pairs were pooled according to primer efficiency (based on intensity on gel electrophoresis signal after PCR on test template). PCR products were purified after an initial PCR reaction of 6 cycles (reaction volume = 200 µL) and divided among 8-10 PCR reactions containing single primer pairs for another 26 cycles (reaction volume = 25 µL). Thereafter, PCR products derived from the same cells were pooled in equimolar amounts and a final 6 cycle PCR reaction containing 20 ng of pooled PCR product (reaction volume = 100 µL) was performed with primers that contained sequencing adaptor sequences. All fragments >700bp were removed using size selection on a 1% agarose gel follow by gel extraction of the selected products (Qiagen, #28704). Quality measures for the 4C library preparation and sequencing can be found in Supplemental figure 1.

Sequencing

Libraries were sequenced using the HiSeq2500 platform (Illumina), according to the manufacturer's protocol, producing 50 bp single end reads.

Data analysis

The raw sequencing reads were de-multiplexed based on viewpoint specific primer sequences. Reads were then trimmed to 16 bases and mapped to an *in silico* generated library of fragends (fragment ends) neighboring all *DpnII* sites in human genome (NCBI37/hg19), using the custom Perl scripts. No mismatches were allowed during the mapping and the reads mapping to only one possible fragend were used for further analysis.

Identification of the interacting genes

First, we calculated the number of covered fragends within a running window of *k* fragends throughout the whole chromosome where the viewpoint is located. The *k* was set separately for every viewpoint so it contains on average 20 covered fragends in the area around the viewpoint (+/- 100kb). Next, we compared the number of covered fragends in each running window to the random distribution. The windows with significantly higher number of covered fragends compared to random distribution ($p<10^{-8}$ based on binominal cumulative distribution function; R *pbinom*) were considered as significant 4C-seq signal. The following criteria were defined for the identification of the candidate genes; i) the Transcriptional Start Site (TSS) co-localizes with a significant 4C-seq signal ($P<10^{-8}$) within 5 kbp; ii) the susceptibility variant or other variant in linkage disequilibrium (LD) co-localizes with a DNA regulatory element identified though FAIRE, the presence of H3K4Me1, H3K4Me3, H3K27Ac, H3K4Me2 or H3K9Ac, EP300 or CTCF binding sites or DNase hypersensitivity sites (Supplemental table 9) in the cell type from which the 4C-seq signal originated and iii) the gene is expressed (RPKM > 0.5) in the assayed cell type.

Identification of gene expression

For monocyte expression, data from the ENCODE database were used (Supplemental table 10)(15). For coronary endothelial cell expression, HMVECs (Lonza) were cultured on gelatine coated plates in EGM2-MV (Lonza) supplemented with penicillin and streptomycin. Subsequently, HMVECs were cultured for 20 hours in low serum medium (EBM + 0.5% FCS), followed by cell lysis and RNA isolation using the RNeasy isolation kit (Qiagen). Polyadenylated mRNA was isolated using Poly(A) Beads (NEXTflex). Sequencing libraries were made using the Rapid Directional RNA-Seq Kit (NEXTflex) and sequenced on Illumina NextSeq500 to produce single-end 75 base long reads (Utrecht Sequencing Facility). Reads were aligned to the human reference genome GRCh37 using STAR version2.4.2a(16). Read groups were added to the BAM files with Picard's AddOrReplaceReadGroups (v1.98). The BAM files are sorted with Sambamba v0.4.5 and transcript abundances are quantified with HTSeq-count version 0.6.1p1(17) using the union mode. Subsequently, reads per kilobase of transcript per million reads sequenced (RPKM's) are calculated with edgeR's rpkm() function(18).

Pathway analysis

The interacting genes (with and without expressed CARDIoGRAMplusC4D/METASTROKE genes) were analyzed using QIAGEN's Ingenuity Pathway Analysis (IPA, 2015 winter version, QIAGEN Redwood City, www.qiagen.com/ingenuity). We used IPA to identify canonical biological pathways within the Ingenuity Knowledge Base to which the interacting genes were mapped. Limits were set to only direct relationships that were experimentally observed in humans. We performed six rounds of pathway analysis, three in each of the cell types: one with only CARDIoGRAMplusC4D/METASTROKE genes that were expressed in the cell type, one with the CARDIoGRAMplusC4D/METASTROKE genes supplemented by the newly identified genes and one with the novel genes only.

Tracks and plots

All tracks were accessed from the UCSC browser (hg19) (http://genome.ucsc.edu/). Regional plots were generated using LocusZoom version 1.3(19).

Gene-based tests

Data for CAD were downloaded from the CARDIoGRAMplusC4D website (http://www.cardiogramplusc4d.org). We obtained summary statistics from GWAS on body mass index (BMI), blood lipids including LDL, HDL, total cholesterol and triglycerides, systolic and diastolic blood pressure, coronary calcification, fasting glucose, smoking behavior, and type 2 diabetes from public online resources and data on intima-media thickness and plaque-presence via data request (Supplemental Table 11). We used a VErsatile Gene-based Association Study (VEGAS) to calculate gene-based association statistics from the summary statistics of each interacting gene for each trait. The details of the methods applied by VEGAS have been described elsewhere(20). In short, SNPs are mapped to the gene (in and around ±50kb from 5' and 3' gene borders), and using the GWAS p-value a gene-based test statistic is calculated corrected for the underlying population linkage disequilibrium structure. Finally using simulations an empirical gene-based *p*-value of association with the phenotype is calculated per gene. VEGAS results were considered multiple testing significant if they were $P < 6.97 \times 10^{-6}$ (0.05/22 phenotypes x 326 available genes in VEGAS).

eQTL analysis in STAGE

Within the STAGE study, patients undergoing coronary artery bypass grafting (CABG) surgery were sampled for seven different tissues, namely atherosclerotic arterial wall (AAW), internal mammary artery (IMA), liver, skeletal muscle (SM), subcutaneous fat (SF), visceral fat (VF), and fasting whole blood (WB) for RNA and DNA isolation(21). Patients that were eligible for CABG and had no other severe systemic diseases (e.g. widespread cancer or active systemic inflammatory disease) were included. For quality control in genotyping, SNPs filtered for minor allele frequency MAF < 5%, Hardy-Weinberg equilibrium (HWE) p-value < $1x10^{-6}$, and call rate of 100%. Imputation was carried out using IMPUTE2 with 1000 Genomes EUR as the reference(22). Quality control for imputed genotypes used additionally an IMPUTE2 INFO score filter < 0.3. After QC a total of 5,473,585 SNPs remained. The Ethical committee of the Karolinska Hospital approved the study, and all patients gave written informed consent after the nature and possible consequences of the study were explained. An expression trait was tested for association with each genotyped and imputed SNP using Kruskal-Wallis test and false discovery rate to correct for multiple testing as described before. First, all *cis*-pairs of SNPs within 50kb of the transcription start or end site for each gene were identified. Next, *cis* SNP-gene pairs were tested for association in all seven STAGE tissues using kruX(23). The p-value for eQTL inclusion in kruX was set at 0.05. Finally, an empirical FDR estimate for each eQTL-gene pair was calculated using ten permutations by shuffling patient IDs on genotype data. As a result, the most significant eQTL-gene association in each tissue was reported.

eQTL analysis in Haploreg

Data on eQTL in healthy individuals were extracted from Haploreg version 4.1 (http://www.broadinstitute.org/mammals/haploreg/haploreg.php). The viewpoint SNPs and all SNPs in LD ($r^2 > 0.8$) were used as input. From the output, for each interacting gene, the most significant eQTL within each tissue was extracted.

eQTL analysis in CTMM circulating cells

CTMM circulating cells is a Dutch cohort from four different hospitals comprising of 714 patients undergoing coronary angiography of whom blood was stored. Monocytes were isolated by density centrifugation followed by positive magnetic bead isolation (CD14) and expression was measured using the Illumina humanHT-12 v3 Gene Expression BeadChip Array. After removal of samples with a median

intensity of <50, 370 patients were included in the analysis. The data were quantile normalized and and log2 transformed using the lumi R package(24).

Genotyping was performed using a customized Affymetrix Axiom Tx array containing 767,203 genetic markers. Community standard quality control was performed, filtering out samples with missingness >5%, outlying heterozygosity (± 4SD from the cohort mean) or inconsistent sex. Samples of non-European descent or those that were out of Hardy-Weinberg equilibrium ($p<5x10^{-5}$) were removed. In total, 622 were used in the current analysis. Untyped variants were imputed using a combined reference panel of the 1000 Genomes Project(25) and Genome of the Netherlands(26) totaling more than 90 million genetic variants across the genome. We used the software packages SHAPEIT(27) for phasing and IMPUTE2(22) for imputation. Prior to *cis*-eQTL analysis we filtered the imputed genotype data from CTMM based on MAF > 0.5%, HWE P > $1x10^{-6}$, Info-metric > 0.9, and only focused on those variants in LD (r2 >= 0.8) with the CAD associated variants. We then used fastQTL (v2.184)(28) to perform I-eQTL analyses using a fixed range (based on the 4C interactions) around each probeID available on the expression array.

Mouse knockout models

Murine gene names were mapped to the genes as follows. First, a custom data file was downloaded from the HUGO Gene Nomenclature Committee (http://www.genenames.org/cgi-bin/download) including the Approved gene name and the Mouse Genome Database ID from the Mouse Genome Informatics database and a file containing all available phenotypic information for all knockout mice was downloaded from MGI (ftp://ftp.informatics.jax.org/pub/reports/index.html#pheno). Next, for all approved gene names of genes identified through 4C-seq, the mouse phenotypes were looked up by linking the MGI IDs. If no linkage could be made for the MGI ID, this was coded as no available mouse model. If a mouse model was available, but no phenotype was found, this was coded as no available phenotype. If a mouse model was specifically coded as not showing any phenotype upon knockout, this was coded as a gene not resulting in any phenotype. Murine cardiovascular phenotypes were defined as a phenotype resulting in any of the following: impaired blood coagulation or abnormal platelets, abnormal glucose levels or homeostasis, abnormal vascular morphology, vascular remodelling or arterial differentiation, abnormal blood pressure, abnormal vasoconstriction, vasodilatation or vascular permeability, abnormal stress response of the heart, myocardial infarction, abnormal (circulating) lipid levels, abnormal fat morphology or amount, abnormal body weight, abnormal lipid droplet or fat cell

size, abnormal macrophage response or inflammation, abnormal wound healing, arteritis, vasculitis, vascular occlusion or atherosclerosis.

Human knockout models

The interacting genes were extracted from the supplementary tables of the studies of Sulem *et al.* and MacArthur *et al.*(29,30). For each of the interacting genes, all SNPs and indels resulting in human functional knockouts were reported.

Drug targets

For the lookup of existing drugs that target any of the candidate genes, we used a custom built drug pipeline that searches for drug-gene interactions using DGIdb(31) , which merged the most known drug-gene interaction databases, such as DrugBank(32) and PharmGKB(33) . We removed redundant results using STITCH(34) and WHO's INN(35) . We tested overrepresentation of drug groups according to ATC codes(36) using Fisher exact tests.

RESULTS

4C-seq identifies additional candidate genes

We identified 37 active DNA regulatory elements that co-localize with susceptibility loci for CAD or LAS. Twenty-six were active in both monocytes and coronary endothelial cells, 5 were only active in monocytes and 6 were only active in coronary endothelial cells (Supplemental table 1). To identify the target genes of these active DRE, we generated 63 4C-seq interaction datasets. We applied the following criteria for the identification of candidate genes: I) the transcriptional start site (TSS) co-localizes with a significant 4C-seq signal ($P<10^{-8}$) within 5kb; II) the susceptibility variant or any other variant in LD (r^2 ≥0.8) co-localizes with an active DRE signal in the cell type from with the 4C-seq signal was obtained and III) the gene is expressed (RPKM > 0.5) in the studied cell type. With this approach, we identified 326 candidate genes (Supplemental table 1), 77 in human male coronary endothelial cells, 84 in human male monocytes and 165 in both cell types (Fig. 1). In total, we identified 294 candidate genes that were not previously reported by the CAD and LAS GWAS (Supplemental table 1). We replicated 235/242 (97.1%) of the chromatin interactions with expressed genes that were identified in male coronary endothelial cells in female coronary endothelial cells (Supplemental table 1).

4C-seq identifies candidate genes in novel pathways

We performed cell-type specific pathway analysis of the candidate genes identified by 4C-seq combined with the candidate genes that were previously identified by the GWAS on CAD and LAS (Supplemental table 2). Notably, these analyses revealed the *Hypoxia signaling in the cardiovascular system* pathway in monocytes (P = 0.01) and the *NRF-mediated oxidative stress response* pathway ($P = 4.68*10^{-4}$ and P = 0.026 in coronary endothelial cells and monocytes respectively, Supplemental table 2). These pathways are both involved in the cellular response to oxidative stress. Additionally, the 4C-seq approach revealed *PTEN* (a player in the *Hypoxia signaling in the cardiovascular system* pathway) as a novel candidate gene (Supplemental table 1). Although this gene was never reported via previous GWAS annotation, *PTEN* (phosphatase and tensin analog) was found to be a likely candidate gene based on dose-dependently upregulation by statins through higher peroxisome proliferator-activated receptor-gamma (PPAR gamma) activity(37). A mutation of *PTEN* led to inflammatory plaque characteristics in human atherosclerotic plaque(38) and increased stability of PTEN was found to ameliorate atherosclerosis(39). Furthermore, *PTEN* shares its upstream transcription regulator *ZEB2* with *CDKN2A* and *CDKN2B*

(enrichment *P* of overlap for ZEB2-regulated genes: 3.02×10^{-3} in monocytes and 3.06×10^{-3} in coronary endothelial cells). We reveal multiple novel pathways related to cardiovascular disease and we now show that *PTEN* physically interacts with a DRE at rs2246833 in monocytes (*P* interaction = 2.36×10^{-10}).

Expression of identified genes is genotype dependent

DRE exert their function through regulation of gene expression. We explored this mechanism by studying expression quantitative trait loci: the GWAS SNPs (or a SNP in LD; $r^2>0.8$) that significantly affected the expression of the candidate genes identified by 4C-seq in the studied tissues (Table 1). For the candidate genes identified by 4C-seq in coronary endothelial cells, we performed lookups within eQTL data of atherosclerotic artery wall and internal mammary artery in the STAGE cohort of patients undergoing cardiac bypass surgery. We identified two eQTL (FDR<0.1) in atherosclerotic artery wall (rs9818870: *MRAS* and rs2281727: *SRR*, Supplemental table 3a). The *SRR* gene, that has not been reported previously, encodes for the serine racemase enzyme that is an endogenous ligand of the glycine site of NDMA receptors in the brain. Blockage of this site was found to prevent stroke damage(40). Interestingly, a set of twice the number of genes from the same genetic locus that were not identified by 4C-seq as a target gene resulted in no significant eQTL in STAGE. Using the HaploReg tool, we additionally examined expression in aorta, coronary artery and tibial artery tissue and identified another seven eQTL for genes that we identified in coronary endothelial cells (Supplemental table 3b), of which *ARL3* and *FAM117B* were not reported before. Both genes are poorly studied in the context of cardiovascular disease. Within the *VAMP5-VAMP8-GGCX* locus we replicate rs1561198, that was previously reported to be an eQTL for *GGCX* in mammary artery by the CARDIoGRAMplusC4D investigators in the ASAP study(41), as an eQTL for *GGCX* in aorta and tibial artery.

For genes identified by 4C-seq in monocytes, we performed *cis*-eQTL analysis in monocytes from 370 patients undergoing coronary angiography for coronary artery atherosclerosis in the CTMM (Center for Translational Molecular Medicine) Circulating Cells cohort(42). We identified four eQTL (FDR<0.1) of which the genes overlap with genes identified by 4C-seq in monocytes of these patients (rs12740374: *PSRC1*, rs1561198: *VAMP8*, rs2246833: *LIPA*, rs12413409: *USMG5*, Supplemental table 3c). Previously, the CARDIoGRAMplusC4D investigators also identified rs1561198 as an eQTL for *VAMP8* in lymphoblastoid cells and skin in the MuTHER study(43). Inclusion of the previously published cardiovascular cohort of Zeller et al.(44) revealed five additional genes (Supplemental table 3d). The SNP that revealed the strongest association with gene expression of *PSRC1* in monocytes of CTMM

(rs7528419) is in perfect LD (1) with rs12740374 in the *SORT1* region. Interestingly, whereas the minor allele of the latter SNP is known to increase *SORT1* expression in liver, we found no such association between rs7528419 and *SORT1* expression in monocytes (nominal $P = 0.87$). In addition, we found an association between higher *PSRC1* expression in monocytes with a more severe atherosclerotic phenotype identified by a higher atherosclerotic burden, quantified using the SYNTAX score ($P = 0.003$). This association with high atherosclerosis burden could not solely be explained by LDL levels, the putative mechanism through which *SORT1* expression affects cardiovascular disease phenotypes (*P* when corrected for circulating LDL levels = 0.01). Expression of *PSRC1* in whole blood has previously been associated with cardiovascular disease in an Asian population(45). Largely because the functional significance of the minor allele of rs12740374 as a transcription factor binding site that increases *SORT1* expression directly, no further attention has been given to alternative candidate genes in the *SORT1* region. With our 4C-seq approach in monocytes, we here show first evidence that the expression of *PSRC1*, a candidate gene in the *SORT1* locus, is genotype-dependent expressed in monocytes and related to the severity of atherosclerosis. This example further supports the implication of our additionally identified candidate genes in cardiovascular disease.

Additional genetic annotation

We further explored current genetic knowledge for the candidate genes identified through 4C-seq (Table 1, Supplemental table 4-7).

First, if the candidate genes are effector genes of the DREs within CVD susceptibility loci, one would expect the genes to be enriched for (common) variants associated with CVD. Using the VEGAS algorithm, we concatenated GWAS p-values of all single-nucleotide polymorphisms (SNPs) in or within 50kb of a gene into a p-value for that particular gene. This way, we studied the genes identified by 4C-seq in published and unpublished GWAS data studying a total of 22 traits, either surrogate markers of atherosclerosis or known risk factors for cardiovascular disease (Supplemental table 4). Of all 326 candidate genes, 33 showed a significant association ($P < 0.05/(22x326) = 6.97x10^{-6}$) with coronary artery disease in the CARDIoGRAMplusC4D GWAS and 149 were nominally associated with coronary artery disease in the CARDIoGRAMplusC4D GWAS (significant enrichment, 149/326: binomial $P = 2.9 \times 10^{-102}$). Additionally, we found 7 genes that were significantly associated with BMI in GIANT and 29 genes that were significantly associated with at least one lipid trait in GLGC.

Second, we annotated the prioritized interacting genes from our 4C-seq experiment with phenotypic information of mouse models within the Mouse Genome Informatics database (MGI, www.informatics.jax.org). We found murine phenotypic information on 144 mouse homologues (Supplemental table 5). Knockout of 67 of them resulted in a phenotype related to cardiovascular disease (significant enrichment, 67/144: binomial $P = 1.36 \times 10^{-47}$), such as abnormal blood vessel morphology (*Col4a1, Cxcl12, Epor, Shc1, Tcf7l1*), altered circulating fatty acid levels (*Csf2, Kdm3a, Ldlr, Lipa, Pten*) and impaired vascular contractility (*Acta2*). Human variants in *ACTA2* are associated with early onset stroke and MI(46). Knockout of two candidate gene murine homologues affected development of atherosclerotic lesions, namely *Ldlr* (accelerated development of atherosclerosis) and *Shc1* (resistance to diet-induced atherosclerosis). The p66 isoform of human SHC1 is implicated in reactive oxygen species generation and its knockdown in endothelial cells of obese mice attenuated production of these radicals and of fatty acids oxidation(47). Third, we investigated the biological effect of human knockout variation of the candidate genes to study druggability. We queried two datasets of available information on SNPs and insertion/deletion variants that cause human functional knockouts(29,30). We found human knockouts, caused by nonsense, splice or frameshift variants, for 89 candidate genes (Supplemental table 6).

Fourth, using a custom-built drug discovery pipeline, we found available compounds to target 50 of the candidate genes (Supplemental table 7a). These drugs showed a relative overrepresentation for usage as immunomodulating agents ($P = 0.012$ in coronary endothelial cells, $P < 0.001$ in monocytes) (Supplemental table 7b).

Together, these findings provide further evidence that by using the 4C-seq method we identified additional candidate target genes for human atherosclerotic disease.

DISCUSSION

Based on 3D chromatin-chromatin interactions with DNA regulatory elements that co-localize with previously identified susceptibility loci, we present 294 additional candidate genes for CAD and LAS that are of potential interest in the pathophysiology of human atherosclerotic disease. This study is the first to systematically study the human chromatin interactions of the CARDIoGRAMplusC4D and METASTROKE loci. Many of the additional genes have not been implicated in atherosclerosis before. Our approach, from a DNA regulatory point of view, complements conventional methods for candidate gene identification of GWAS susceptibility, can help further unravel diseases with a complex genetic background, and pave the way for cell-type specific drug development.

We have highlighted the 4C candidate genes that we could annotate via additional analyses and that therefore have known or foreseeable effects on cardiovascular disease. Based on tissue-specific pathway analyses, we highlighted *PTEN* that is known to be upregulated by statins and to possess effects on atherosclerosis(37–39). Furthermore, based on eQTL studies, we identified *SRR*, the effect of which was previously implicated in stroke(40), and *USMG5*, that was previously associated with white matter hyperintensities in the brain(48). Of special interest is the finding of an alternative mechanism by which the susceptibility locus that contains rs7528419 (*SORT1* region) may exert its effect. Using 4C-sequencing we identified a physical interaction between an active regulatory element that overlaps rs7528419 and *PSRC1* in monocytes. Moreover, we found an association between rs7528419 and the expression of *PSRC1* in monocytes and an association between *PSRC1* expression and atherosclerosis severity. This association was independent of LDL levels, which is the putative mechanism of rs12740374, a SNP in perfect LD (1.0) with rs7528419 that was previously found to increase *SORT1* expression in liver.

Mapping the SNPs that identify susceptibility loci in GWAS to genes that affect a complex disease, such as cardiovascular disease, is a challenging task. By annotating the locus with the linearly closest gene, the 3D conformation of chromatin is inadvertently not taken into account. Many of the additional candidate genes we report are located outside the GWAS susceptibility loci. Using 5C (chromosome conformation capture carbon copy) the importance of studying 3D interactions has been highlighted previously; in a sample of 628 TSS from the ENCODE project only 7% of the over 1000 long-range looping interactions were with the nearest gene(10). In a previous effort to identify candidate genes based on DNA regulatory mechanisms, 33 enhancers in the 9p21 locus were scrutinized(49). Interestingly, the chromatin interaction between the enhancers identified by 3C (chromatin conformation capture) was

found to be remodeled upon treatment with interferon-γ in HUVECs. In our 4C-seq experiment, we confirmed the physical chromatin-chromatin interaction between the 9p21 susceptibility locus and several candidate genes, among which interferons, in human coronary endothelial cells and monocytes. However, we found that these genes were not actively expressed in these cell types and therefore did not consider them any further.

There are some limitations to this study. First, there is no consensus about the gold standard approach to analysis of 4C-seq data. For example, we used a conservative cut-off for calling a chromatin-chromatin interaction ($P < 10\text{-}8$). Altering this cut-off may result in more candidate genes. However, this likely leads to more false-positive results. We therefore report a quantitative measure for the p-value of the interaction of the DRE with the proposed candidate gene to enable the reader to take these considerations into account when interpreting the data. Second, while 4C-sequencing enables us to look at physical interactions, these interactions do not necessarily mean that the expression in the studied tissue is in fact regulated by the association locus or even expressed. We therefore decided to only report only genes that are actively expressed in the studied tissues. Furthermore, we found no eQTL association between the SNP of interest and any of the genes *in the vicinity* of the genes that were identified by 4C-sequencing, indicating that the resolution of the technique is sufficient to distinguish between candidate genes and less relevant genes within a genomic region.A more accurate cell type-specific mapping of susceptibility loci to candidate genes in humans is of paramount importance for the development of specific compounds in the pharmaceutical industry. The genes we identified display only partial overlap between coronary endothelial cells and monocytes. This finding stresses the importance of cell-specific approaches in order to grasp the complex biology of atherosclerotic disease. It also highlights the possibility to develop cell-specific compounds to target atherosclerotic disease. Our results therefore underline the need to investigate cell type-specific 3D chromatin conformation in future functional follow-up of GWAS data.

Acknowledgements

The help of Noortje van den Dungen and Nico Lansu with the 4C-seq experiments is greatly acknowledged.

Funding

Saskia Haitjema and Sander van der Laan are supported by the FP EU project CVgenes@target (HEALTH-F2-2013-601456). Claartje A. Meddens is supported by the Alexandre Suerman Stipendium (UMC Utrecht). Sander van der Laan is funded through grants from the Netherlands CardioVascular Research Initiative ("GENIUS", CVON2011-19) and the Interuniversity Cardiology Institute of the Netherlands (ICIN, 09.001). Magdalena Harakalova is supported by NIH Ro1 grant LM010098. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant HL105756. The work for STAGE was supported by PROCARDIS in the 6th EU-framework program (LSHM-CT-2007-037273), the Swedish Heart-Lung Foundation (JLMB), the King Gustaf V and Queen Victoria's Foundation of Freemasons (JLMB), the Swedish Society of Medicine (JLMB), the Swedish Heart Lung Foundation and Research Council (HFA and JLMB). This work was also supported by grant from University of Tartu (SP1GVARENG, JLMB), the Estonian Research Council (ETIS, JLMB) and by Clinical Gene Networks AB. Folkert W. Asselbergs is supported by a Dekker scholarship-Junior Staff Member 2014T001 – Netherlands Heart Foundation and UCL Hospitals NIHR Biomedical Research Centre. Michal Mokry is supported by OZF/2012 WKZ-fund.

Disclosures

All authors declare to have no competing interests. Johan L.M. Björkegren is founder, main shareholder and chairman of the board for Clinical Gene Networks AB (CGN). CGN has an invested interest in microarray data generated from the STAGE cohort. Cavadis B.V. financed genotyping of AEGS1. Gerard Pasterkamp is founder and stockholder of Cavadis B.V.

REFERENCES

1. Hansson GK, Libby P, Tabas I. Inflammation and plaque vulnerability. Journal of Internal Medicine. 2015. p. 483–93.

2. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet [Internet]. 2012 Dec 2;45(1):25–33. Available from: http://www.nature.com/doifinder/10.1038/ng.2480

3. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): A meta-analysis of genome-wide association studies. Lancet Neurol. 2012;11(11):951–62.

4. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun [Internet]. 2015;6:5890. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4420238&tool=pmcentrez&rendertype=abstract

5. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011;7(1).

6. Brnne I, Civelek M, Vilne B, Di Narzo A, Johnson AD, Zhao Y, et al. Prediction of causal candidate genes in coronary artery disease loci. Arterioscler Thromb Vasc Biol. 2015;35(10):2207–17.

7. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science (80-) [Internet]. 2012;337(6099):1190–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3771521&tool=pmcentrez&rendertype=abstract

8. Mokry M, Middendorp S, Wiegerinck CL, Witte M, Teunissen H, Meddens CA, et al. Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium. Gastroenterology. 2014;146(4):1040–7.

9. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet [Internet]. 2012;45(2):124–30. Available from: http://dx.doi.org/10.1038/ng.2504\nhttp://www.nature.com/doifinder/10.1038/ng.2504

10. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature [Internet]. 2012 Sep 5;489(7414):109–13. Available from: http://www.nature.com/doifinder/10.1038/nature11279

11. De Laat W, Klous P, Kooren J, Noordermeer D, Palstra RJ, Simonis M, et al. Three-Dimensional Organization of Gene Expression in Erythroid Cells. Current Topics in Developmental Biology. 2008. p. 117–39.

12. Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet [Internet]. 2014 Jan 12;46(2):205–12. Available from: http://www.nature.com/doifinder/10.1038/ng.2871

13. Van De Werken HJG, De Vree PJP, Splinter E, Holwerda SJB, Klous P, De Wit E, et al. 4C technology: Protocols and data analysis. Methods Enzymol. 2012;513:89–112.

14. Hansson GK, Libby P. The immune response in atherosclerosis: a double-edged sword. Nat Rev Immunol. 2006;6(7):508–19.

15. Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. An integrated encyclopedia of DNA elements in the human genome. Nature [Internet]. 2012;489(7414):57–74. Available from: http://www.nature.com/doifinder/10.1038/nature11247\npapers3://publication/doi/10.1038/nature11247

16. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

17. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.

18. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139–40.

19. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics [Internet]. 2010 Sep 15;26(18):2336–7. Available from: http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btq419

20. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010;87(1):139–45.

21. Hägg S, Skogsberg J, Lundström J, Noori P, Nilsson R, Zhong H, et al. Multi-organ expression profiling uncovers a gene module in coronary artery disease involving

transendothelial migration of leukocytes and LIM domain binding 2: The Stockholm Atherosclerosis Gene Expression (STAGE) study. PLoS Genet. 2009;5(12).

22. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. Schork NJ, editor. PLoS Genet [Internet]. 2009 Jun 19;5(6):e1000529. Available from: http://dx.plos.org/10.1371/journal.pgen.1000529

23. Qi J, Asl H, Björkegren J, Michoel T. kruX: matrix-based non-parametric eQTL discovery. BMC Bioinformatics [Internet]. 2014;15(1):11. Available from: http://www.biomedcentral.com/1471-2105/15/11

24. Du P, Kibbe WA, Lin SM. lumi: A pipeline for processing Illumina microarray. Bioinformatics. 2008;24(13):1547–8.

25. McVean GA, Altshuler (Co-Chair) DM, Durbin (Co-Chair) RM, Abecasis GR, Bentley DR, Chakravarti A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature [Internet]. 2012 Oct 31;491(7422):56–65. Available from: http://www.nature.com/doifinder/10.1038/nature11632

26. Francioli LC, Menelaou A, Pulit SL, van Dijk F, Palamara PF, Elbers CC, et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet [Internet]. 2014 Jun 29;46(8):818–25. Available from: http://www.nature.com/doifinder/10.1038/ng.3021

27. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods [Internet]. 2012 Dec 27;10(1):5–6. Available from: http://www.nature.com/doifinder/10.1038/nmeth.2307

28. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics [Internet]. 2015;1–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26708335

29. Sulem P, Helgason H, Oddson A, Stefansson H, Gudjonsson SA, Zink F, et al. Identification of a large set of rare complete human knockouts. Nat Genet [Internet]. 2015 Mar 25;47(5):448–52. Available from: http://www.nature.com/doifinder/10.1038/ng.3243

30. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. Science (80-) [Internet]. 2012 Feb 17;335(6070):823–8. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1215040

31. Griffith M, Griffith OL, Coffman AC, Weible J V, McMichael JF, Spies NC, et al. DGIdb: mining the druggable genome. Nat Methods [Internet]. 2013;10(12):1209–10. Available from: http://dx.doi.org/10.1038/nmeth.2689

32. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: A comprehensive resource for “Omics” research on drugs. Nucleic Acids Res. 2011;39(SUPPL. 1).

33. Altman RB. PharmGKB: a logical home for knowledge relating genotype to drug response phenotype. Nat Genet [Internet]. 2007;39:426. Available from: http://dx.doi.org/10.1038/ng0407-426\nhttp://www.nature.com/ng/journal/v39/n4/pdf/ng0407-426.pdf

34. Kuhn M, Szklarczyk D, Franceschini A, Von Mering C, Jensen LJ, Bork P. STITCH 3: Zooming in on protein-chemical interactions. Nucleic Acids Res. 2012;40(D1).

35. Van Bever E, Wirtz VJ, Azermai M, De Loof G, Christiaens T, Nicolas L, et al. Operational rules for the implementation of INN prescribing. Int J Med Inform. 2014;83(1):47–56.

36. Pahor M, Chrischilles EA, Guralnik JM, Brown SL, Wallace RB, Carbonin P. Drug data coding and analysis in epidemiologic studies. Eur J Epidemiol. 1994;10(4):405–11.

37. Teresi RE, Planchon SM, Waite KA, Eng C. Regulation of the PTEN promoter by statins and SREBP. Hum Mol Genet. 2008;17(7):919–28.

38. Muthalagan E, Ganesh RN, Sai Chandran B V, Verma SK. Phosphatase and tensin analog expression in arterial atherosclerotic lesions. Indian J Pathol Microbiol. 2014;57(3):427–30.

39. Dai XY, Cai Y, Mao DD, Qi YF, Tang C, Xu Q, et al. Increased stability of phosphatase and tensin homolog by intermedin leading to scavenger receptor A inhibition of macrophages reduces atherosclerosis in apolipoprotein E-deficient mice. J Mol Cell Cardiol. 2012;53(4):509–20.

40. De Miranda J, Santoro A, Engelender S, Wolosker H. Human serine racemase: Moleular cloning, genomic organization and functional analysis. Gene. 2000;256(1-2):183–8.

41. Folkersen L, Van’t Hooft F, Chernogubova E, Agardh HE, Hansson GK, Hedin U, et al. Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. Circ Cardiovasc Genet. 2010;3(4):365–73.

42. Hoefer IE, Sels JW, Jukema JW, Bergheanu S, Biessen E, McClellan E, et al. Circulating cells as predictors of secondary manifestations of cardiovascular disease: Design of the CIRCULATING CELLS study. Clin Res Cardiol. 2013;102(11):847–56.

43. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet [Internet]. 2012;44(10):1084–9. Available from: http://dx.doi.org/10.1038/ng.2394\nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3784328&tool=pmcentrez&rendertype=abstract

44. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility. Bochdanovits Z, editor. PLoS One [Internet]. 2010 May 18;5(5):e10693. Available from: http://dx.plos.org/10.1371/journal.pone.0010693

45. Arvind P, Nair J, Jambunathan S, Kakkar V V., Shanker J. CELSR2-PSRC1-SORT1 gene expression and association with coronary artery disease and plasma lipid levels in an Asian Indian cohort. J Cardiol. 2014;64(5):339–46.

46. Guo DC, Papke CL, Tran-Fadulu V, Regalado ES, Avidan N, Johnson RJ, et al. Mutations in Smooth Muscle Alpha-Actin (ACTA2) Cause Coronary Artery Disease, Stroke, and Moyamoya Disease, Along with Thoracic Aortic Disease. Am J Hum Genet. 2009;84(5):617–27.

47. Paneni F, Costantino S, Cosentino F. P66Shc-induced redox changes drive endothelial insulin resistance. Atherosclerosis. 2014;236(2):426–9.

48. Lopez LM, Hill WD, Harris SE, Valdes Hernandez M, Munoz Maniega S, Bastin ME, et al. Genes From a Translational Analysis Support a Multifactorial Nature of White Matter Hyperintensities. Stroke [Internet]. 2015 Feb;46(2):341–7. Available from: http://stroke.ahajournals.org/lookup/doi/10.1161/STROKEAHA.114.007649

49. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. Nature [Internet]. 2011 Feb 10;470(7333):264–8. Available from: http://www.nature.com/doifinder/10.1038/nature09753

TABLES

Table 1 Candidate genes identified by 4C-seq in human coronary endothelial cells and/or human monocytes

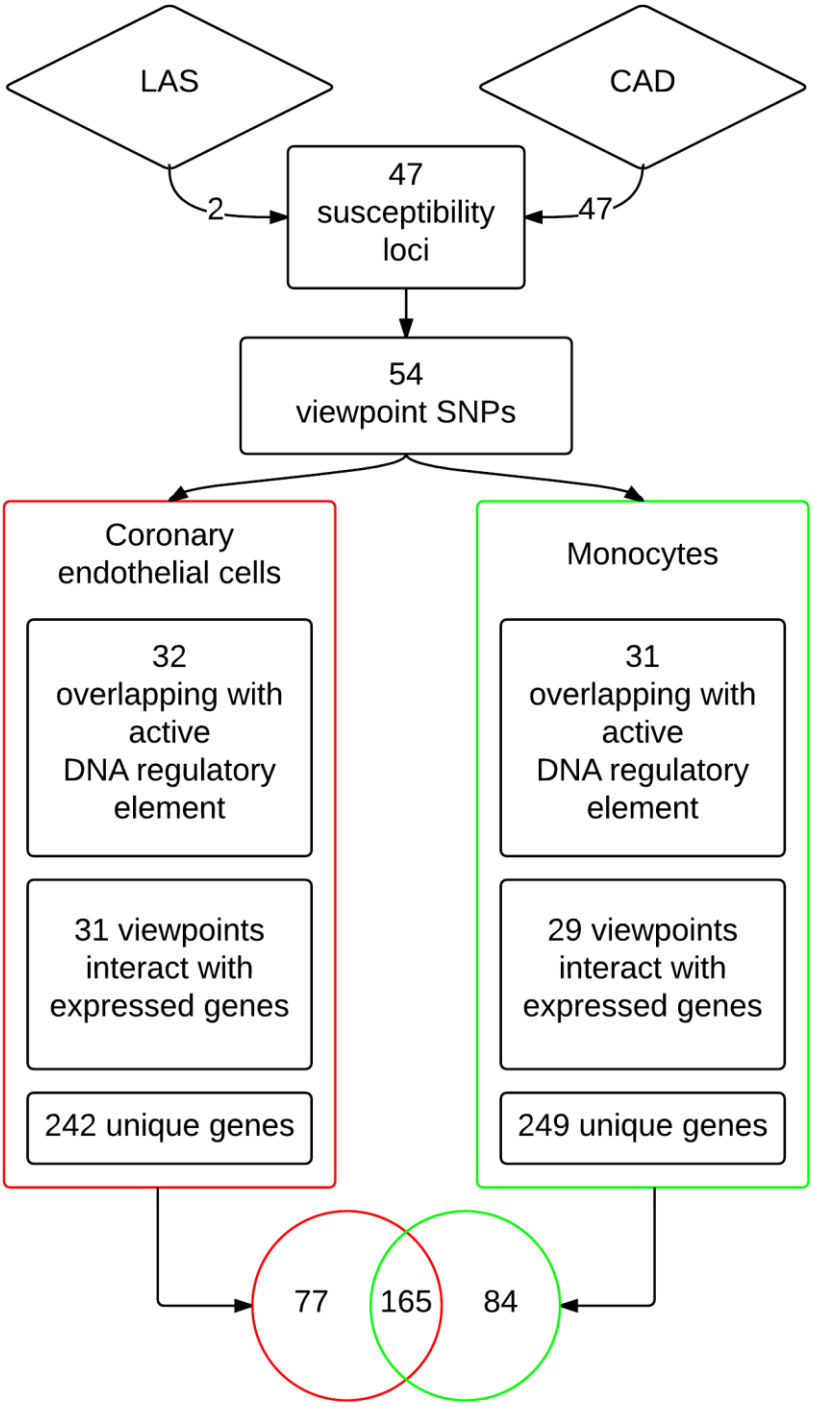
Chr	Susceptibility locus	4C-seq viewpoint(s)	Gene identified by 4C-seq	Cell type of identification		eQTL
				Coronary endothelial cells	Monocytes	
1	MIA3	rs17464857	*AIDA*	√	√	
			BROX	√	√	
			MARC1	√		
			MIA3	√	√	
			TAF1A	√	√	
1	SORT1	rs12740374	*AMPD2*	√	√	
			ATXN7L2	√	√	
			CYB561D1	√	√	
			GNAI3	√	√	
			GNAT2	√		
			GSTM2	√	√	
			GSTM4	√	√	
			PSMA5	√	√	
			PSRC1	√	√	Monocytes
			SARS	√	√	
			SORT1	√	√	
2	APOB	rs515135	*LDAH*	√	√	
2	VAMP5-VAMP8-GGCX	rs1561198	*GGCX*	√	√	CEC
			C2orf68	√	√	
			ELMOD3	√	√	
			MAT2A	√	√	
			RETSAT	√	√	
			RNF181	√	√	
			TGOLN2	√	√	
			TMEM150A	√	√	
			USP39	√	√	
			VAMP5	√	√	
			VAMP8	√	√	Monocytes
2	WDR12	rs6725887	*CARF*		√	
			FAM117B	√	√	CEC/ Monocytes
			NBEAL1	√	√	CEC
			WDR12	√	√	
3	MRAS	rs9818870	*MRAS*	√	√	CEC
6	ANKS1A	rs12205331	*C6orf106*	√	√	
			RPS10	√	√	

			SNRPC	√	√	
			UHRF1BP1	√	√	
6	PHACTR1	rs9369640, rs12526453	*MYLIP*		√	
			PHACTR1	√	√	CEC
8	LPL	rs264	*INTS10*	√	√	
8	TRIB1	rs2954029	*TRIB1*	√	√	
9	CDKN2BAS	rs1333049, rs3217992, rs2383207	*CDKN2A*	√		
			CDKN2B	√	√	
10	CYP17A1-CNNM2-NT5C2	rs12413409	*ARL3*	√	√	CEC
			USMG5	√	√	Monocytes
10	CNNM2	rs12413409	*BORCS7*	√	√	
			WBP1L	√	√	
10	KIAA1462	rs2505083	*KIAA1462*	√		CEC
10	LIPA	rs11203042, rs2246833	*LIPA*	√	√	CEC/ Monocytes
13	COL4A1-COL4A2	rs4773144	*COL4A1*	√		
			COL4A2	√		
17	RAI1-PEMT-RASD1	rs12936587	*PEMT*	√	√	Monocytes
			RASD1	√	√	Monocytes
17	SMG6	rs2281727	*SRR*	√	√	CEC
			SMG6	√	√	
17	UBE2Z	rs15563	*CALCOCO2*	√	√	Monocytes
			KPNB1		√	
			UBE2Z	√	√	Monocytes
19	LDLR	rs1122608	*C19orf52*		√	
			CARM1		√	
			LDLR		√	
			SMARCA4		√	
			TSPAN16		√	
			YIPF2		√	

* Only genes that have their interacting viewpoint as eQTL or genes with a significant gene-based *P* value ($P < 6.97 \times 10^{-6}$ by gene-based test using VEGAS (corrected for multiple-testing 0.05/22 phenotypes x 326 available genes) are depicted. The full list of genes identified by 4C-seq can be found in Supplemental table 1. Susceptibility locus: name of the locus as given by CARDIoGRAMplusC4D or METASTROKE; viewpoint: SNP used as the focus point for the primer design of the 4C experiment; Gene: gene physically interacting with viewpoint, determined by 4C-seq; Underlined genes: genes that have previously been reported by CARDIoGRAMplusC4D or METASTROKE; Chr: chromosome; eQTL: expression quantitative trait locus; GWAS: genome-wide association study CAD: coronary artery disease; BMI: body mass index; TC: total cholesterol; LDL: low-density lipoprotein; HDL: high-density lipoprotein; TG: triglycerides

FIGURES

Figure 1 Flowchart of identification of candidate genes



Susceptibility loci: SNPs associated with risk of disease in METASTROKE and/or CARDIoGRAM. Viewpoint SNPs: SNPs used as the focus point for the primer design of the 4C experiment.

Figure legend for Supplemental figure 1

Supplemental figure 1. Quality of 4C datasets of coronary endothelial cells and monocytes.
The graphs show the quality measures of the 4C datasets. Y-axes shows the percentage of fragends covered around the viewpoint (measure for the complexity of the dataset), x-axes show the percentage of reads in cis i.e. the percentage of reads that map to the chromosome of the viewpoint (measure for the specificity of the dataset). Each dot represents one viewpoint.