Circulation Research

ORIGINAL RESEARCH

Single-Cell Epigenomics and Functional Fine-Mapping of Atherosclerosis GWAS Loci

Tiit Örd, Kadri Õunap,* Lindsey K. Stolze,* Redouane Aherrahrou,* Valtteri Nurminen, Anu Toropainen, Ilakya Selvarajan⬤, Tapio Lönnberg, Einari Aavik, Seppo Ylä-Herttuala⬤, Mete Civelek⬤, Casey E. Romanoski, Minna U. Kaikkonen⬤

RATIONALE: Genome-wide association studies have identified hundreds of loci associated with coronary artery disease (CAD). Many of these loci are enriched in cisregulatory elements but not linked to cardiometabolic risk factors nor to candidate causal genes, complicating their functional interpretation.

OBJECTIVE: Single-nucleus chromatin accessibility profiling of the human atherosclerotic lesions was used to investigate cell type–specific patterns of cisregulatory elements, to understand transcription factors establishing cell identity, and to interpret CAD-relevant, noncoding genetic variation.

METHODS AND RESULTS: We used single-nucleus ATAC-seq (assay for transposase-accessible chromatin with sequencing) to generate DNA accessibility maps in >7000 cells derived from human atherosclerotic lesions. We identified 5 major lesional cell types including endothelial cells, smooth muscle cells, monocyte/macrophages, natural killer/T cells, and B cells and further investigated subtype characteristics of macrophages and smooth muscle cells transitioning into fibromyocytes. We demonstrated that CAD-associated genetic variants are particularly enriched in endothelial and smooth muscle cell–specific open chromatin. Using single-cell coaccessibility and cis–expression quantitative trait loci information, we prioritized putative target genes and candidate regulatory elements for ≈30% of all known CAD loci. Finally, we performed genome-wide experimental fine-mapping of the CAD variants identified in genome-wide association studies using epigenetic quantitative trait loci analysis in primary human aortic endothelial cells and self-transcribing active regulatory region sequencing (STARR-Seq) massively parallel reporter assay in smooth muscle cells. This analysis identified potential causal single-nucleotide polymorphisms (SNPs) and the associated target gene for over 30 CAD loci. We present several examples where the chromatin accessibility and gene expression could be assigned to one cell type predicting the cell type of action for CAD loci.

CONCLUSIONS: These findings highlight the potential of applying single-nucleus ATAC-seq to human tissues in revealing relative contributions of distinct cell types to diseases and in identifying genes likely to be influenced by noncoding genome-wide association study variants.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: atherosclerosis ▪ coronary artery disease ▪ genetics ▪ genome-wide association study ▪ myocardial infarction

---

Editorial, see p 259 | In This Issue, see p 217

Atherosclerosis, a chronic inflammatory disease of the artery wall, is the underlying cause of coronary artery disease (CAD) and myocardial infarction (MI).

Recent transcriptomic analyses at single-cell level have demonstrated extensive heterogeneity and plasticity of cells within atherosclerotic lesions, with over 10 leukocyte

---

Correspondence to: Tiit Örd, PhD, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, 70211, Finland, Email tiit.ord@uef.fi; or Minna U Kaikkonen, PhD, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, 70211, Finland, Email minna.kaikkonen@uef.fi

*K. Õunap, L.K. Stolze, and R. Aherrahrou contributed equally as second authors.

Novelty and Significance

What Is Known?

- Single-cell RNA sequencing has uncovered marked heterogeneity of cell types within atherosclerotic lesions, but little is known how these differences manifest at the level of epigenetics.
- Coronary artery disease (CAD) genome-wide association study risk variants have been shown to be enriched in arterial tissue enhancers but little is known of the cell type–specific mechanisms of disease risk.
- Connecting variants to target genes and pinpointing the exact causal single-nucleotide polymorphism (SNP) among those in high linkage disequilibrium represents a major challenge for understanding the genetic basis of common diseases.

What New Information Does This Article Contribute?

- Our study provides the first single-cell map of chromatin accessibility in human atherosclerotic lesions.
- We present several CAD loci where chromatin accessibility and gene expression could be assigned to one cell type predicting the cell type of action for CAD loci.
- We prioritize and identify the most potential causal SNP(s) and the target gene(s) for over 30 CAD genome-wide association study loci.

In this study, we describe the cellular composition of atherosclerotic plaques based on chromatin accessibility. This allowed us to predict transcription factor binding at cis-regulatory elements that underlie the identity of the major lesional cell types. We use the genome-wide cisregulatory element maps to provide cell type–specific annotation and fine-mapping of common noncoding genome-wide association study variants associated with CAD. We predict interactions between risk variants and target genes using peak-promoter coaccessibility information and further prioritize such variants using cis–expression quantitative trait loci information from arterial tissues. Furthermore, we use analysis of epigenetic molecular trait loci and massively parallel reporter assay to functionally characterize all common CAD genome-wide association study variants located within endothelial and smooth muscle cell–specific cisregulatory elements, respectively. This strategy allowed us to identify the causal variant and the target gene for over 30 CAD loci. One-third of these loci were further associated with atherosclerosis-relevant phenotype in smooth muscle cells, supporting their functional importance. Altogether, our study represents an important resource to annotate CAD loci, predicting disease-relevant cell types as well as putative genes and variants that provide one step towards more comprehensive picture of the biology underlying atherosclerosis.

Nonstandard Abbreviations and Acronyms

CAD	coronary artery disease
EC	endothelial cell
FOX	Forkhead box
HOX	homeobox
LDL-C	low-density lipoprotein cholesterol
MEF2	myocyte enhancer factor 2)
MI	myocardial infarction
SMC	smooth muscle cell
Sn	single-nucleus
SOX	SRY-related HMG-box
TF	transcription factor

subtypes and nonimmune cell types identified in mice and humans.[1–4] However, the transcriptional mechanisms that define these cell states and identities remain elusive. As the unique gene expression profile of each cell type is controlled by the combined activity of promoters and enhancers, a deeper characterization of the transcriptional regulatory elements is needed. Although promoters are responsible for the binding of the factors that nucleate the

assembly of a functional preinitiation complex, enhancers contain the majority of binding sites for master TFs (transcription factors) responsible for the cell type specificity and dynamic patterns of gene expression.[5] Thus, the analysis of the TF recognition motifs associated with cell type–specific enhancer repertoires can provide unprecedented information on the lineage-determining TFs that control the cells' developmental and functional properties. To this end, recent single-nuclei (sn) chromatin accessibility studies have demonstrated the potential to interpret the TF regulatory grammar that underlies in vivo enhancer landscapes.[6]

Genome-wide association studies (GWAS) have discovered hundreds of common genetic variants significantly associated with CAD and MI.[7,8] However, the large majority of the GWAS variants are located in the noncoding regions of the genome with no known function.[9,10] It has been shown that a significant number of disease susceptibility regions fall within enhancer elements specific to disease-relevant cell types and states where they affect gene regulation mainly by altering TF binding.[11] For instance, hepatic and adipose gene expression-associated single-nucleotide polymorphisms (SNPs) are enriched for their associations with type 2 diabetes.[12] In addition, a landmark study focusing on enhancer variant at the chromosome 1p13, demonstrated that rs12740374

influences LDL-C (low-density lipoprotein cholesterol) level and MI risk via liver-specific transcriptional regulation of the *SORT1* gene with minimal effect in adipocytes and lymphocytes.[13] This highlights the need to identify the cell type–specific regulatory regions and their target genes to better understand the molecular mechanisms underlying the genetic basis of disease.

These findings have also propelled research towards the identification of active cisregulatory elements across atherosclerosis-relevant cell types in vascular tissue. To this end, large-scale functional genomics projects have advanced our understanding of cell- and tissue-specific enhancers.[14–16] Furthermore, there is extensive literature describing the enhancer repertoires of endothelial cells (ECs), smooth muscle cells (SMCs), macrophages, and T cells and the dynamic changes in regulatory element activity upon exposure to proatherogenic stimuli.[17–19] These include the description of superenhancers,[20] which are often found to regulate genes important for cell identity and to be enriched for disease-associated variants. Although these initial works have provided evidence that enhancers may play a significant role in orchestrating the gene expression changes observed during disease progression, we do not know how well the in vitro enhancer landscapes studied correlate with the in vivo tissue context.

To advance our knowledge of the in vivo enhancer repertoires of atherosclerosis-associated cell types and to identify the enhancer connectivity with target genes, we have profiled the cell type–specific cis-regulatory elements in human atherosclerotic lesions using single-nucleus assay for transposase-accessible chromatin with sequencing (snATAC-Seq). We use this data to (1) identify cisregulatory elements that differ between subtypes of cells, (2) compare the in vitro and in vivo enhancer repertoires, (3) to characterize the chromatin coaccessibility networks encompassing genetic variants associated with the risk of atherosclerosis (CAD, MI), and to (4) pinpoint potential target genes for these risk loci using chromatin coaccessibility scoring combined with cis–expression quantitative trait loci (eQTL) information. Finally, (5) we fine-map potentially causal genetic risk variants located within EC- and SMC-specific regulatory elements using molecular QTL information and massively parallel reporter assay, respectively.

METHODS

Data Availability

The data used for the analyses described in this article were obtained from the Genotype-Tissue Expression Portal on May 1, 2020 (phs000424.v8.p2).

Anonymized data and materials have been made publicly available at the figshare repository and can be accessed at https://doi.org/10.6084/m9.figshare.14501985.

Please see the Data Supplement for detailed methods.

RESULTS

snATAC-Seq Identifies the Major Constituent Cell Types of Atherosclerotic Lesions

To unravel the molecular composition of human atherosclerotic lesions, we optimized a protocol to obtain highly pure suspensions of nuclei from 3 endarterectomy samples and profiled chromatin accessibility at the level of individual cells using snATAC-Seq. In total, we obtained ≈7000 single-cell ATAC-Seq (assay for transposase-accessible chromatin with sequencing) profiles (sample 1: 1816 nuclei, sample 2: 2372 nuclei, sample 3: 2821 nuclei) with a median depth of 7129 fragments per nucleus. Projection of the combined dataset for visualization by t-distributed stochastic neighbor embedding followed by automated clustering identified 15 clusters (Figure 1A), which we manually annotated to 5 major nuclear populations based on known cell type–specific genes (Figure 1B and Methods in the Data Supplement) including ECs, smooth muscle cells (SMCs), monocyte/macrophages, natural killer/T cells, and B cells. To confirm the cell type identities, we calculated the top chromatin accessibility marker genes (Figure 1C) and analyzed genome-wide enrichment of TF motifs in accessible chromatin regions (Figure 1D). Importantly, the marker gene accessibility was highly consistent throughout the 3 samples (Figure IA in the Data Supplement). To investigate the how the chromatin accessibility based predictions correlate with expression of the predicted transcription factors in the corresponding cell types, we reprocessed the published single-cell RNA sequencing (scRNA-Seq) dataset from human atherosclerotic arteries (n=4) by Wirka et al[1] and annotated each cell cluster using known cell type marker genes (Figure II and III in the Data Supplement). A large majority of the TFs identified were also associated with cell type–specific gene expression in scRNA-seq (Figure IB in the Data Supplement), confirming their likely role in defining cell type–specific enhancer landscapes. For example, our analysis confirmed key EC-specific TF SOX (SRY-related HMG-box) 13, macrophage-specific TF SPI1 (Spi-1 proto-oncogene; encoding PU.1, purine-rich box-1), SMC-specific TF TEAD3 (TEA Domain Transcription Factor 3), and natural killer/T-cell–specific TF ETS1 (ETS proto-oncogene 1) (Figure 1E and Figure IB in the Data Supplement).

To discover the regulatory elements, we identified accessible chromatin regions in each major cell type using HOMER (Hypergeometric Optimization of Motif EnRichment).[5] Altogether, we identified 328,670 peaks with 30 000 to 98 000 peaks per cell type (Figure 1F, Tables I and II in the Data Supplement). Somewhat more peaks were identified from macrophages, T cells, and SMCs, possibly due to a higher overall number of nuclei in these clusters. Importantly, we found that a ≈4% fraction of all peaks (12 795) was shared across all 5 cell
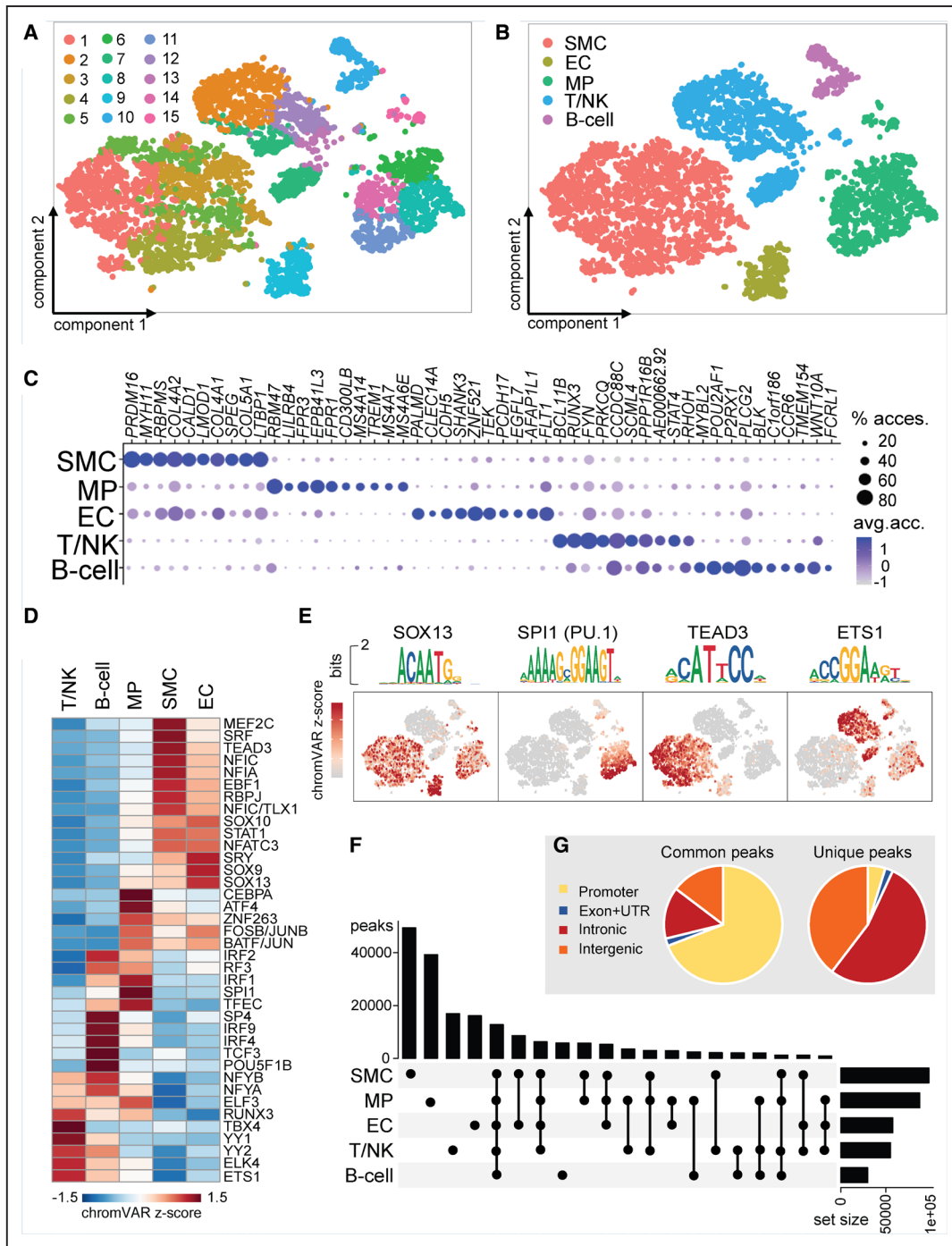
Figure 1. Clustering and identification of cell type clusters in human atherosclerosis single-nucleus assay for transposase-accessible chromatin with sequencing (snATAC-Seq) data.
A and **B**, t-distributed stochastic neighbor embedding (tSNE) projection of the 7009 snATAC-Seq profiles represented as (**A**) 15 clusters identified using automated clustering and (**B**) the 5 manually annotated clusters corresponding to smooth muscle cells (SMCs), endothelial cells (ECs), macrophages (MPs), T/natural killer (NK)-cells and B cells. **C**, Dot plot demonstrates the top chromatin accessibility marker genes of each atherosclerotic lesion cell type. Dot size corresponds to the proportion of cells within the cluster that displays gene accessibility (% access.), and dot color intensity corresponds to the average accessibility level (avg. acc.; counts in the gene body and promoter region, depth-normalized to 10,000 total counts per cell and log-transformed). For each cell type, the top 10 markers by fold change (provided false discovery rate [FDR]<0.05 by Wilcoxon rank-sum test) were selected for plotting. **D**, Heatmap showing differences in chromatin accessibility for the TF (transcription factor) motifs with the greatest accessibility variability between the 5 major atherosclerotic lesion cell populations. Color bar displays the chromVAR (Chromatin Variation Across Regions) accessibility deviation Z score, and motifs were required to satisfy FDR<0.05 (Wilcoxon rank-sum test) comparing one cell type vs all other cells. **E**, tSNE projection of selected TF motif accessibility Z scores. **F**, UpSet plot of intersected peaks among cell types. The 20 most populated intersection groups are presented. **G**, Pie chart representing the relative fraction of common or shared snATAC-Seq peaks in the different genomic annotations. ETS1 indicates ETS proto-oncogene 1, transcription factor; PU.1, purine-rich box 1; SOX, SRY-related HMG-box; TEAD3, TEA domain transcription factor 3; and UTR, untranscribed region.

type clusters, whereas ≈40% of the peaks (127 702) were unique to one single-cell cluster (Figure 1F, Table I in the Data Supplement). As expected, over 66% of the common peaks were promoter-associated (±1 kb of a transcription start site [TSS]), whereas 90% of the cell type–specific peaks were found at intergenic and intragenic enhancer sites, indicating that cell type specificity is mainly reflected by the enhancer repertoire (Figure 1G).

Analysis of Cell Subtype Chromatin Profiles in Relation to Their Gene Expression

In addition to the 5 major cell types discussed above, recent scRNA-Seq studies have identified several subtypes of these cells in humans.[1–4] In line with this, automated clustering of our dataset identified 15 subtypes of cells (Figure 1A). We, therefore, sought to analyze these cell subtypes in relation to the gene expression profiles. To achieve this, we performed coembedding of the snATAC-Seq with scRNA-seq[1] data (Figure 2A and 2B). This allowed the prediction of cell types for snATAC-Seq data based on distances to the preannotated cells in scRNA-seq data. We identified corresponding cell populations for 20 subtypes of cells (Figure 2A), with the largest subdivision seen among the myeloid cells (3 macrophage and 1 monocyte subgroup) and in the continuum of pericyte–SMC–fibroblast axis (8 subtypes). Comparison of the scRNA-Seq–derived annotations with the labels derived from only snATAC-Seq (Figure 1A and 1B) did not reveal major conflicts (Figure IV in the Data Supplement). Notably, our snATAC-Seq dataset was low in monocytes, fibroblasts, and pericytes, likely due to the endarterectomy samples having less complete representation of the vessel wall (eg, missing adventitia) compared to the whole artery sections used in the scRNA-Seq.

Genome-wide comparison of the accessibility profiles between the subtypes of cells suggested a high similarity of open chromatin regions (Figure 2C). We, therefore, focused our deeper characterization of the regulatory region differences on major macrophage (clusters 5, 11, and 13) and SMC lineages (clusters 3, 4, and 14) that were well represented in our snATAC-Seq data. In line with the current consensus,[21] 3 main macrophage subsets, proinflammatory (cluster 5, expressing *IL1B*, *CCL3*), anti-inflammatory lipid-associated (cluster 11, expressing *TREM2*, *SPP1*, *MS4A7,* and *APOE*), and resident-like (cluster 13, expressing *LYVE1*, *S100A9*) macrophages, were identified based on marker gene expression (Figure 2D; Figures V and VI in the Data Supplement). To identify the key TFs that occupy the cell type–specific open chromatin regions, we examined the patterns of motif enrichment throughout these 3 subtypes. As expected, the inflammatory macrophages showed enrichment of proinflammatory TF motifs, such as RELA (RELA proto-oncogene, NF-kB subunit) and AP-1 (activator protein 1) factors, whereas the lipid-associated macrophages

were enriched for IRF1 (interferon regulatory factor 1), SP4 (Sp4 transcription factor), and SP2 (Sp2 transcription factor). On the other hand, the regulatory elements of resident-like macrophages exhibited enrichment of KLF4 (Kruppel like factor 4) and NFATC2 (nuclear factor of activated T cells 2) motifs. These results are largely in line with our recent report, where the analysis was based on carotid endarterectomy samples (n=18) but where the proinflammatory macrophages were further subdivided into 2 clusters.[3]

Several recent studies have demonstrated SMC reprogramming to other cell types such as fibromyocytes, osteogenic cells, and macrophage-like cells.[1,4] We, therefore, applied Slingshot[22] to infer cell lineages in the snATAC-Seq and scRNA-Seq coembedded space (Figure 2A and 2B). Trajectories were identified within the 3 major clusters including the natural killer/T cells, myeloid cells, and the pericyte–SMC–fibroblast axis, whereas no links were predicted between them (Figure VII in the Data Supplement). This result, together with the evidence from recent SMC lineage tracing studies in mouse models of atherosclerosis demonstrating predominant SMC to fibromyocyte transformation,[1,4,23–25] motivated us to characterize this transition in more detail at the epigenomic level. Monocle 2[26] was used to perform the pseudo-temporal analysis and uncover DNA motifs that may be important in defining the differentiation path. Our analysis of the chromatin accessibility changes corroborated the path of SMC de-differentiation into fibromyocytes previously supported by scRNA-Seq.[1,4] As expected, the fibromyocytes were less enriched for NFY (nuclear transcription factor Y) and SRF (serum response factor) motifs that were highly enriched in the differentially accessible regions in SMCs (Figure 1D and Table III in the Data Supplement). In contrast, a significant genome-wide enrichment of AP-1, NFATC2, and STAT (signal transducer and activator of transcription) 1:STAT2 (and the similar motif of IRF1) motifs was seen along the SMC trajectory as these cells become more fibroblast-like. The activity of these TFs was further supported by the increased accessibility of the promoters encoding for NFATC1 and NFATC2 (Figures VIII and IX in the Data Supplement). Interestingly, the genome-wide accessibility of KLF4 motif only exhibited a nonsignificant trend of increase towards the end of the pseudotime, despite the important role of KLF4 in SMC reprogramming[24] (Figure X in the Data Supplement). Altogether, our pseudotemporal analysis identifies several transcription factors that could play a role in the process of SMC transdifferentiation that merit further research.

In Vivo Enhancer Landscapes Are Highly Similar to In Vitro Cultured Cells

As a plethora of studies has characterized the enhancer repertoires of the cultured vascular cells or blood-derived
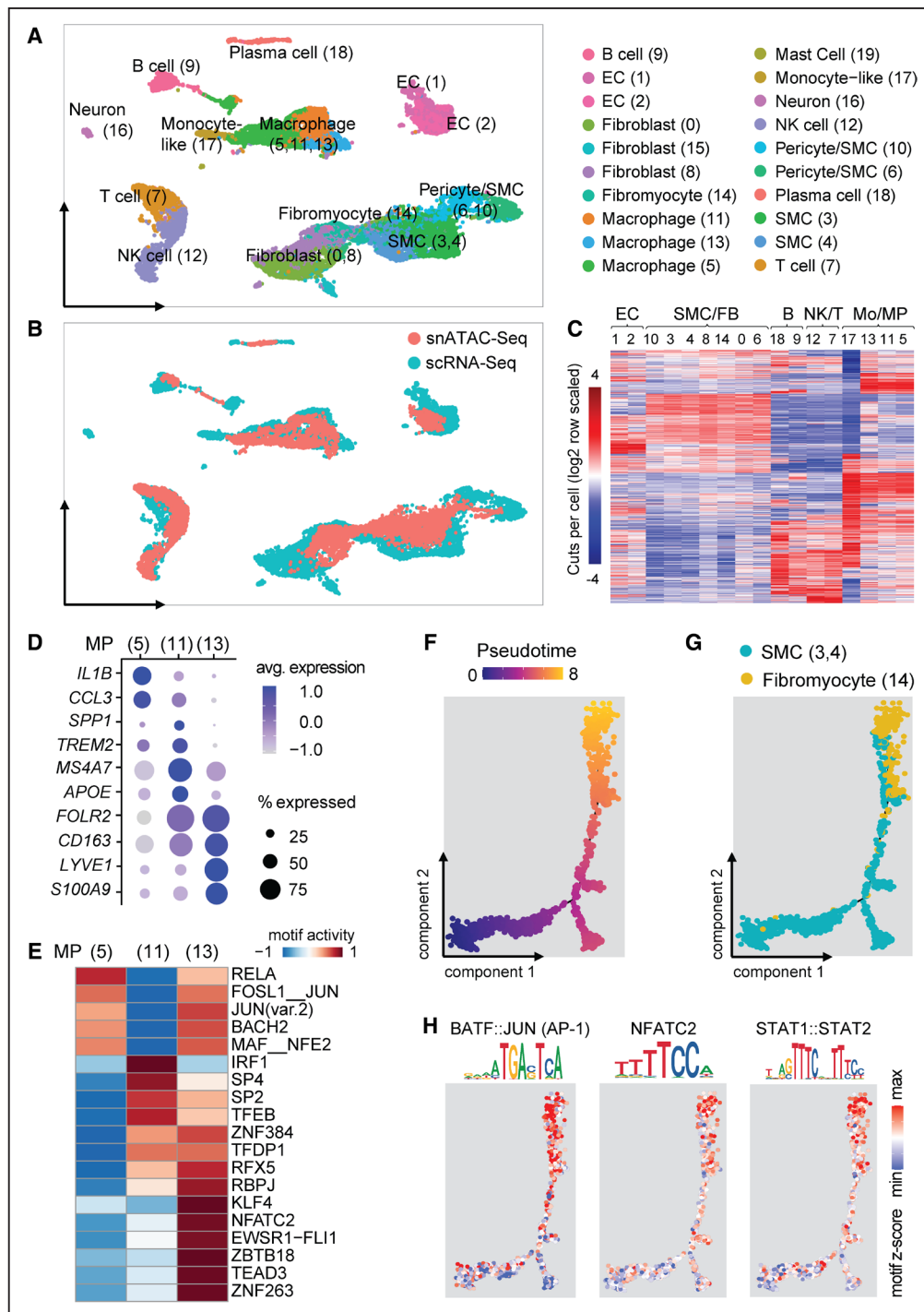
Figure 2. Identification of subtypes of macrophages (MPs) and smooth muscle cells (SMCs) based on the integration of single-cell gene expression and chromatin accessibility data.
A and **B**, Coembedded UMAP representation of both single-nucleus (sn)ATAC-Seq cells and single-cell RNA sequencing (scRNA-Seq) cells, showing (**A**) the cell type annotation and (**B**) the data type of origin. **C**, Row-normalized chromatin accessibility profiles for snATAC-Seq peaks reveals high similarity of cell subtypes. **D**, Dot plot demonstrating the top gene expression markers of the 3 MP subtypes. Dot size corresponds to the proportion of cells within the cluster that express the gene, and dot color intensity corresponds to average expression level (gene counts depth-normalized to 10 000 total counts per cell and log-transformed). **E**, Heatmap showing differences in chromatin accessibility for the TF (transcription factor) motifs with the greatest accessibility variability between the 3 MP subtypes. Color bar displays the chromVAR accessibility deviation Z score. **F**–**H**, Single-cell ATAC trajectory analysis of SMC lineage cells shown colored by (**F**) pseudotime, (**G**) cluster assignment, and (**H**) examples of differentially accessible motifs that vary significantly along the inferred trajectory. In **D** and **E**, results were required to pass Wilcoxon rank-sum test false discovery rate (FDR)<0.05 comparing one cell type vs all other cells. AP-1 indicates activator protein 1; BATF, basic leucine zipper ATF-like transcription factor; B, B cell; EC, endothelial cells; FB, fibroblast; JUN, Jun proto-oncogene, AP-1 transcription factor subunit; MO, monocyte; NFATC, nuclear factor of activated T cells 1; NK, natural killer; STAT, signal transducer and activator of transcription 1; and T, T cell.
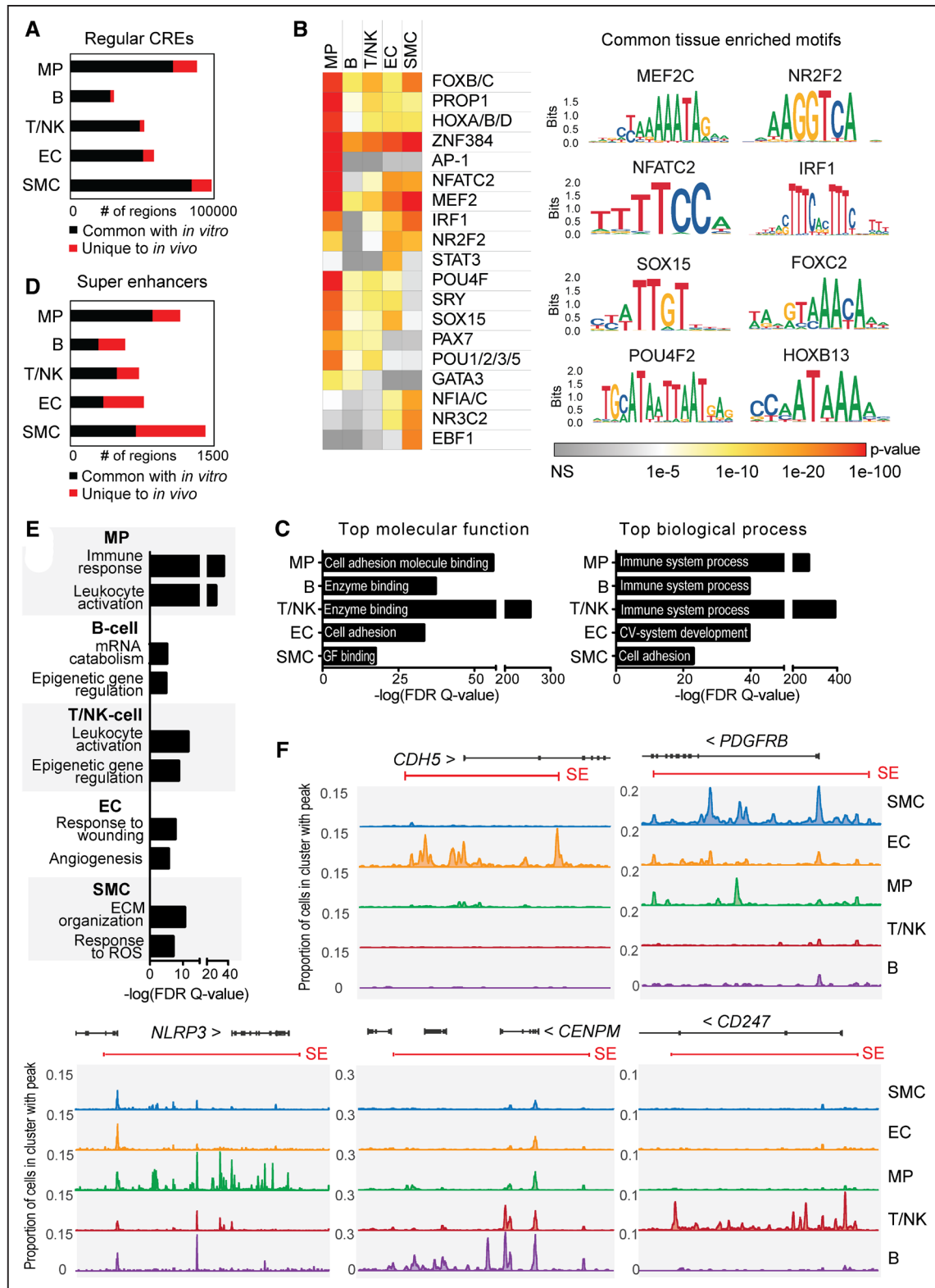
Figure 3. Comparison of in vivo–specific accessible regulatory elements to chromatin accessibility in cell culture models.
A, Overlap of in vitro and in vivo regular (nonsuperenhancer) cisregulatory element (CRE) counts by cell type. **B**, Motif enrichment within in vivo–specific single-nucleus (sn)ATAC-Seq peaks. Selected JASPAR (http://jaspar.genereg.net/) motif logos are presented. **C**, Top molecular function and biological process of the genes within 100 kb from the in vivo–specific CREs, analyzed using GREAT (Genomic Regions Enrichment of Annotations Tool). **D**, Overlap of in vitro and in vivo superenhancer (SE) counts by cell type. **E**, Top biological process of the nearby genes located within 1 MB of the superenhancer. **F**, Pseudobulk coverage track visualization of snATAC-Seq signal at the cell type–specific superenhancers. B indicates B cell; chr, chromosome; EC, endothelial cells; ECM, extracellular matrix; FOX, Forkhead box; HOX, homeobox; IRF1, interferon regulatory factor 1; MEF2C, myocyte enhancer factor 2C; MP, macrophage; NFATC2, nuclear factor of activated T cells 2; NK, natural killer; NR2F2, nuclear receptor subfamily 2 group F member 2; NS, not significant; POU4F2, POU class 4 homeobox 2; ROS, reactive oxygen species; SMC, smooth muscle cell; and T, T cell.

immune cells, we next sought to compare in vitro regulatory landscapes to those in tissue. To achieve this, we collected public bulk ATAC-Seq data from in vitro cells corresponding to the 5 major cell types identified in Figure 1, cultured in basal or proinflammatory conditions (Table IV in the Data Supplement), and compared them to the regulatory elements defined by snATAC-Seq. Our results demonstrate that the in vivo enhancer profiles are highly concordant with the in vitro profiles, with ≈70% to 90% of regions identified in vivo overlapping with those in corresponding cultured cells (Figure 3A). To gain insights into mechanisms driving tissue-specific enhancer landscapes, we subjected the in vivo–only regions to de novo motif analysis. The results (Figure 3B) identified TF recognition motifs associated with each cell type, and the corresponding TFs further exhibited cell type–specific expression pattern and motif accessibility (Figure IB in the Data Supplement, Figure 1D), as exemplified by AP-1 motif enrichment in macrophages and NFI (nuclear factor I) A/C in SMCs. These results could indicate a higher cellular heterogeneity and the existence of additional subtypes of cells in tissues, compared to the cell cultures commonly used as in vitro model systems. In addition, we identified several motifs that were enriched in several cell types in vivo compared to in vitro cultures, including MEF2 (myocyte enhancer factor 2), HOX (homeobox), SOX, POU (Pit-Oct-Unc) domain, FOX (Forkhead box), NFATC2, and NR2F2 (nuclear receptor subfamily 2 group F member 2) (COUP-TFII, COUP transcription factor 2) motifs, suggesting that developmental TFs could be more active in the tissues (Figure 3B). We extended these investigations by analyzing the annotations of the nearby genes of in vivo–specific peaks. A common gene ontology for nearby genes in macrophages, SMCs, and ECs was cell adhesion, whereas for immune cell in vivo–unique enhancers immune system process suggesting that the tissue context may have a major impact on effector functions associated with these cell types in atherosclerosis, compared to basic cell culture models (Figure 3C).

Previous studies have demonstrated that superenhancers, clusters of multiple enhancers, drive the expression of genes important for cell identity and function.[20] We, therefore, sought to identify in vivo–specific superenhancers and investigate if they were enriched around genes central to the cell type–specific functions. We detected a total of 2363 superenhancers in the 5 principal cell populations of which only 193 were shared between cell types (Tables I, V, and VI in the Data Supplement). Interestingly, the overlap of in vitro and in vivo superenhancers was smaller (average 57%) compared to that of regular enhancers (average 82%), suggesting that they could be more affected by the tissue context (Figure 3D), although this could also be affected by the representation of samples and treatment conditions in the superenhancer database[27] from which most of the

regions were selected (Table V in the Data Supplement). The superenhancers were associated with gene ontology terms representative of the cell type function in the tissue, such as EC superenhancers associating with genes implicated in angiogenesis and SMC superenhancers with genes related to extracellular matrix (ECM) organization (Figure 3E). Accordingly, near the strongest superenhancers were genes with important roles for the cell identity such as *ETS1*, *CDH5,* and *PODXL* for ECs, *CALD1* and *PDGFRB* for SMCs, *CEBPA*, *NLRP3,* and *FPR2* for macrophages, *CENPM* and *CCR6* for B cells and *CD247*, *BCL11B* and *CD8A/B* for T cells (Figure 3F and Figure XI in the Data Supplement). Altogether, this provides the first description of superenhancers in the lesional cell types that could be used to guide further research looking into the cellular control of atherogenesis.

Linking Cell Type–Specific Regulatory Regions to Target Genes Using Cis-Coaccessibility Networks

Recent studies have demonstrated that arterial tissue contributes the most to the identification of CAD/MI–associated genes (cis-eQTL discovery).[28] We were, therefore, interested to investigate which cell types of the arterial wall would show the highest enrichment of genetic variants associated with CAD/MI or related traits in their open chromatin regions. The results demonstrated that SMCs and ECs show the highest enrichment of GWAS SNPs for CAD and circulatory system-related traits such as pulse and blood pressure (Figure 4A; Table VII in the Data Supplement). Consistent with the overlap of CAD-related traits with hyperlipidemia and type 2 diabetes,[7] significant, although somewhat lower enrichment was also seen for these traits.

One of the significant challenges in understanding the underlying functional mechanisms of noncoding CAD/MI GWAS variants is the identification of their target genes, as the nearest gene approach is recognized to be imprecise. This has spurred studies based on chromosome conformation capture techniques to nominate target genes. This approach has been recently used to assign target genes of GWAS SNPs in cultures of human aortic ECs and coronary artery SMCs.[29,30] However, these approaches are limited by the notion that physical interaction does not necessarily mean functional interactions between an enhancer and a gene. Instead of direct functional interaction, the connections could also represent random interactions or bystander interactions (ie, DNA close to a direct interaction will also be close as a consequence of the former).[31] To mitigate this limitation, we adopted a recently published approach, Cicero,[32] which predicts cisregulatory DNA interactions from single-cell chromatin coaccessibility and further identifies cis-coaccessibility networks (CCANs) as modules of regulatory elements that are highly coaccessible with one another.
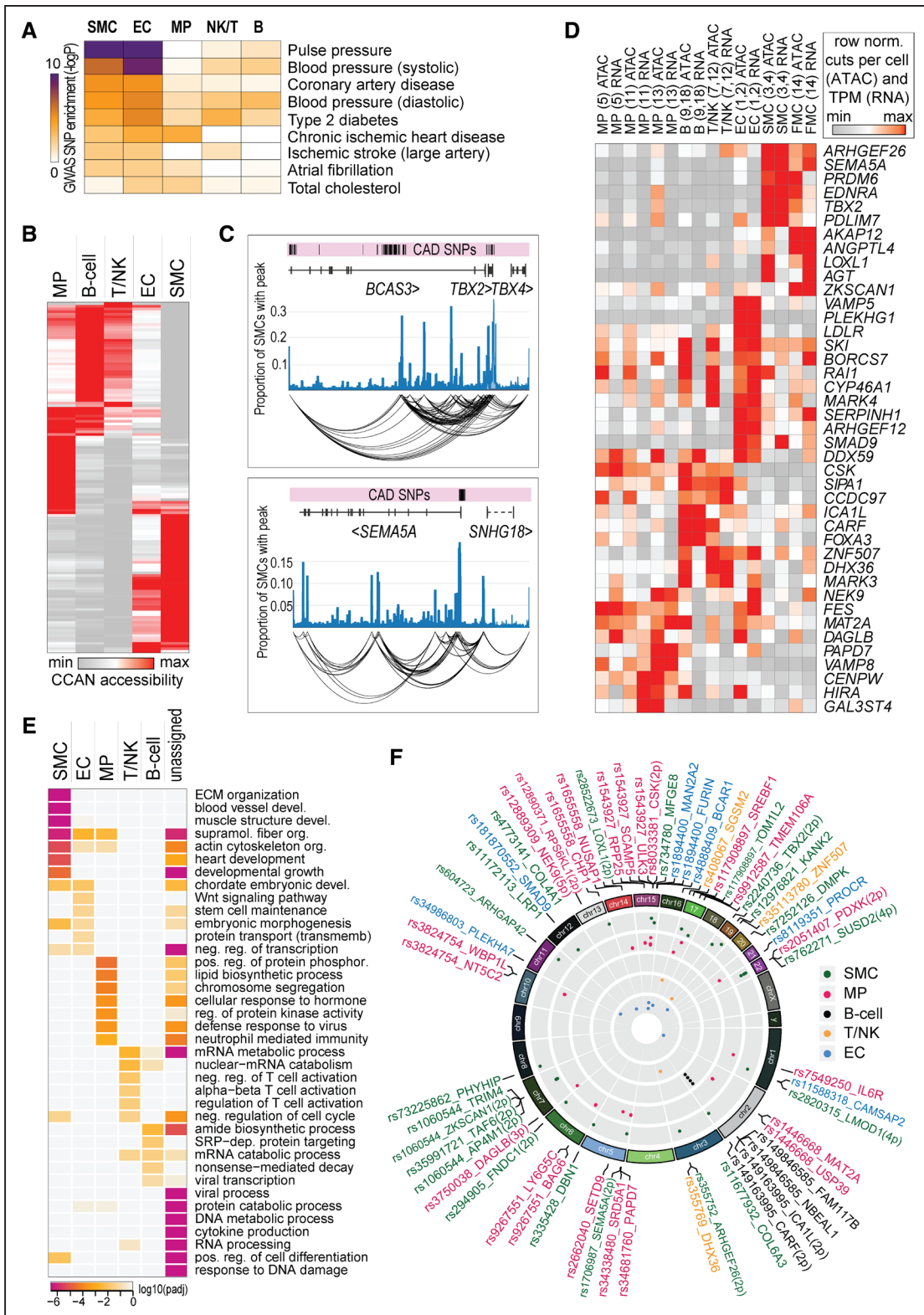
Figure 4. Enrichment of coronary artery disease (CAD)/myocardial infarction (MI) genome-wide association study (GWAS) signals in cell type–specific open chromatin and their linkage to potential target genes.
A, Enrichment of GWAS single-nucleotide polymorphisms (SNPs) for cardiometabolic traits in single-nucleus (sn)ATAC-Seq peaks of the 5 major cell types detected in atherosclerotic lesion samples. B, Mean accessibility of all the peaks within the 3427 cis-coaccessibility network (CCAN) displayed as row-normalized values by cell type. C, Pseudobulk snATAC-Seq coverage track (light blue) visualization of 2 smooth muscle cell (SMC)-specific CCANs centered around the *TBX2* (chr17:59018300-59570950) and the *SEMA5A* (chr5:9014150-9768900) genes. The peak-peak cis-coaccessibility is shown by arches for the pairs that exhibit Cicero score >0.5. D, Selected promoter-associated snATAC-Seq peaks that harbor CAD/MI GWAS SNPs and exhibit cell type–specific chromatin accessibility and gene expression. The row (*Continued*)

We identified 3427 CCANs with a median CCAN size of 400 kb (Table VIII in the Data Supplement). Twenty-seven percent (910/3427) of CCANs were shared between the 5 principal cell types suggesting that majority of them are cell type selective (Figure 4B). For example, *TBX2* and *SEMA5A* loci presented in CCANs that were highly accessible only in SMCs (Figures 4C and Figure XII in the Data Supplement). Notably, less CCANs were identified for B cells and ECs, possibly due to the lower number of nuclei in these clusters.

Next, we overlapped the CAD/MI GWAS SNPs with the snATAC-Seq peaks and the CCANs (Table IX and X in the Data Supplement). Altogether, 423 CAD/MI SNPs fell into 259 snATAC-Seq peaks, representing half of the candidate loci for CAD/MI and encompassing 4% (134/3427) of the CCANs detected across the atherosclerotic lesion cell types. Thirty-eight percent of the SNPs (98/259 SNPs) were in peaks proximal to promoters (corresponding to a total of 109 genes) and could thus be assigned to a particular gene (Table XI in the Data Supplement). These included several cell type–specific peaks which were also associated with cell type–specific gene expression based on scRNA-Seq,[1] including *AGT, EDNRA, TBX2, SEMA5A, PRDM6, PDLIM7, ARHGEF26* specific for human aortic smooth muscle cells (HASMCs), *LDLR, SERPINH1, ARHGEF12, VAMP5, PLEKHG1* specific for human aortic endothelial cells (HAECs), and *VAMP8, CENPW, DAGLB,* and *GAL3ST4* specific for macrophages (Figure 4D). Interestingly, subtype differences were also observed, as exemplified by higher accessibility and expression of *VAMP8* in resident-like macrophages and of *TBX2* in SMCs.

As the majority of the CAD SNP-containing enhancers were located outside of promoters, we used Cicero coaccessibility scoring to identify their putative target genes (promoters). Using a stringent coaccessibility threshold of >0.5 (Methods and Figure XIII in the Data Supplement), we identified 1182 significant peak pairs, including 883 promoters (Table XII in the Data Supplement). Interestingly, ≈40% of the genes (322/883) exhibited cell type specificity where both chromatin accessibility and gene expression were among the top 10% of all the studied cell types. Gene ontology analysis of these genes identified distinct cell type–specific functions that

could be implicated in disease cause (Figures 4E and Figures XIV and XV in the Data Supplement). For example, ECM organization and blood vessel development were enriched among the SMC-specific target genes, whereas EC-specific target genes were enriched for Wnt signaling pathway and embryonic morphogenesis. In contrast, the candidate genes which did not connect to cell type–specific regulatory mechanism (561/883) were enriched for more general cellular processes such as DNA and RNA metabolism (Figure 4E; unassigned). The enrichment for cytokine production and response to DNA damage was explained by extensive sharing of chromatin accessibility between the immune cells where the genes could not be reliably assigned to one cell type (Figure 4B). In summary, our analysis allowed the identification of an associated target gene for the majority (241/259 peaks) of the open chromatin regions harboring CAD risk SNPs (Table XII in the Data Supplement) and provide evidence of cell type–dependent and cell type–independent mechanisms.

Prioritization of CAD GWAS Loci That Affect Gene Expression in Arterial Tissues

eQTL analysis is an important tool for interpreting GWAS findings and allows us to explain regulatory polymorphism associations that influence the disease through altered gene expression in certain tissues. Using our new catalog of stringent Cicero-inferred chromatin coaccessibility connections, we next sought to prioritize the CAD GWAS SNPs by leveraging all the Genotype-Tissue Expression (GTEX) v8 significant variant–gene associations in arterial tissues (coronary, aorta, tibial). Altogether, 185 of the 259 snATAC-Seq peaks harboring CAD SNPs also had an eQTL supporting that SNP, representing a total of 264 genes. Importantly, at the chromatin accessibility level, ≈50% (129/264) of these SNP–gene pairs exhibited Cicero coaccessibility score >0.5 between the SNP-containing peak and the gene promoter, lending weight to the cis-eQTL connection (Figure XIIIB in the Data Supplement). Importantly, for 80 of these genes, the risk SNPs were located in one single-enhancer element (Figure XVI in the Data Supplement).

The majority of snATAC-Seq peaks (≈60%) were accessible in multiple cell types, and this was largely

---

Figure 4 Continued. normalization was performed separately for snATAC-Seq (peak cuts per cell) and single-cell RNA sequencing (scRNA-Seq) (transcripts per million; TPM) data. The top 10% row value is shown in red. **E**, Gene ontology enrichment for cell type–specific connections between a promoter and a peak containing a CAD/MI SNP (peak-promoter coaccessibility score >0.5). Cell type specificity was defined as peak accessibility and gene expression signal within the top 10% among all the cell types studied but only passed that threshold in one cell type. The resulting gene lists were profiled for over-representation of Gene Ontology Biological Process categories. The results were corrected for multiple testing to an experiment-wide threshold of a=0.05, and up to 7 most significant categories (provided P_{adj}<0.05) were picked for each gene list for plotting. **F**, The arterial cis–expression quantitative trait loci (eQTL) SNPs and the associated gene are shown for snATAC-Seq peaks that exhibit peak-promoter coaccessibility score >0.5. Peaks and genes were further filtered for cell type specificity by requiring peak accessibility and gene expression level to be within the top 10% among all the cell types studied but only one of the cell types (see Figure XVI in the Data Supplement). Only one SNP per peak is shown. If ≥2 peaks demonstrate cell type–specific accessibility, only one SNP is listed and p denotes the number of total cell type–specific peaks. For full list see Table XIII in the Data Supplement. B indicates B cell; chr, chromosome; EC, endothelial cells; dep., dependent; ECM, extracellular matrix; MP, macrophage; NK, natural killer; pos, position (coordinate) in chromosome; reg., regulation; SRP, signal recognition particle; and T, T cell; and Wnt, Wingless-related integration site.

explained by extensive sharing of peaks between immune cells. However, the remaining ≈40% of the peaks exhibited cell type–specific chromatin accessibility and frequently also cell type–specific expression of the gene associated by peak-promoter coaccessibility (Table XIII in the Data Supplement). Such gene-regulatory elements pairings are exemplified by macrophage-specific accessibility and expression of *IL6R, PDXK, and DAGLB,* HASMC-specific expression of *SEMA5A, LOXL1, CKB, FNDC1, LMOD1, LRP1, MFGE8, SUSD2, COL6A3,* and *COL4A1* as well as HAEC-specific expression of *PROCR, BCAR1, SLK*, and *PLEKHA7* (Figure 4F, Figures XVII and XVIII in the Data Supplement).

Altogether, our analysis allowed the prioritization and identification of the potential causal SNP(s) and the associated target gene(s) for ≈50% (124/259) of the atherosclerotic lesion open chromatin regions carrying CAD/MI–risk SNPs. We present several examples where the chromatin accessibility and gene expression could be assigned to one cell type allowing deeper understanding of the mechanisms by which the risk variants could act.

EXPERIMENTAL FINE-MAPPING OF THE CAUSAL VARIANTS IN ECS AND SMCS

Finally, we sought to fine map the CAD/MI–associated potential regulatory variants by experimental analysis of the SNP effect in primary human aortic ECs and SMCs by investigating if the genotypes are associated with differences in epigenetic molecular quantitative trait loci or if the genotypes display allele-specific enhancer activity under proatherogenic conditions, respectively (Figure 5A). To further link the functional variants to target genes, we used the peak-promoter coaccessibility and cis-eQTL evidence from GTEX v8 arterial tissues as described above. In addition, we analyzed correlation between the peak accessibility and the target gene expression in corresponding snATAC-Seq and scRNA-Seq cell populations (Figure 5A; Methods and Figure XIX in the Data Supplement) and investigated cis-eQTL associations in pure cell cultures of HAECs and HASMCs from genetically diverse donors.

We have previously identified thousands of genetic variants associated with differential read abundance (allelic bias) in ATAC-seq, histone 3 lysine 27 acetylation (H3K27ac) chromatin immunoprecipitation sequencing (ChIP-Seq), or TF ChIP-Seq data by collecting data from primary cultures of ECs originating from aortic explants and subsequently cultured in 2 different conditions: control (untreated) and proinflammatory cytokine IL (interleukin) 1β treatment.[34] We, therefore, sought to overlay this information with the CAD/MI GWAS SNPs that were located in snATAC-Seq peaks and for which both the SNP-containing peak and a coaccessible promoter were open (cuts per cell >0.1) in atherosclerotic lesion ECs.

Altogether, 32 SNPs with significant epigenetic molecular quantitative trait loci were identified, corresponding to 19 peaks (Figure 5B). Importantly, in the majority of the peaks, the SNP effect was context-specific, modulating the epigenetic signature in only one treatment condition (control or IL1β). For this set of genetic variants, 96 potential target genes were identified by analyzing peak-promoter coaccessibility, RNA level-ATAC accessibility correlation, and cis-eQTL association. Seven of these were further confirmed to represent cis-eQTLs in cultured HAECs including *FES, AXL, CLCN6, NEK9, ACYP1, EIF2B2,* and *MLH3* (Figure 5B and 5C and Figure XX in the Data Supplement). Taken together, these data demonstrate successful prioritization of functional regulatory variants that direct target gene expression in human ECs.

Secondly, we applied the massively parallel reporter assay method self-transcribing active regulatory region sequencing (STARR-Seq) in HASMCs to study allelic regulatory activities of SNPs.[35] First, we constructed a plasmid library encompassing accessible chromatin regions that overlap with CAD/MI GWAS SNPs (leads or proxies). From 3 to 5 of the most common haplotypes (minor allele frequency [MAF]>1%) in the European population were included for each region. The library was transfected to HASMCs, which were subsequently exposed to cholesterol to induce their phenotypic switching to cells that have been shown to resemble modulated SMCs found in atherosclerotic plaques.[36,37] In line with this, cholesterol loading led to a reduction in contractile markers *ACTA2, TAGLN, MYL9, CALD1* and the induction of synthetic and inflammatory markers *IL1B, IL1A, KLF4,* and *LGALS3* (Figure XXI in the Data Supplement). Read counts for the 3 biological replicates of STARR-Seq were highly correlated (r>0.93; Figure XXIIA in the Data Supplement). For analysis of differential allelic activity, all pairwise haplotype-versus-haplotype comparisons that were informative (ie, varying in genotype) for at least one CAD/MI SNP overlapping an snATAC-Seq peak were performed. This resulted in 959 pairwise haplotype comparisons across 351 STARR-Seq library regions, representing 376 CAD/MI–associated SNPs and a total of 238 snATAC-Seq peaks. A total of 34 CAD/MI SNPs, corresponding to 26 snATAC-Seq peaks, were significantly associated with an allelic effect (FDR<5%; Figure XXIIB in the Data Supplement). Of these, 27 SNPs had a GTEX cis-eQTL in arterial tissues across a total of 38 genes (Figure 6A), reproducing several of our predicted causal variants from Figure 4F. In addition, 2 genes represented further cis-eQTLs in cultured HASMCs (*FHL3* and *SNHG18*; Figure XXIII in the Data Supplement). Importantly, several genes represented high-confidence targets for allele-specific enhancer variants with the 3 lines of association evidence, as exemplified by *IL6R, COL4A1, DAGLB, FHL3, PDXK,* and highly SMC-specific genes *MFGE8, LOXL1, TBX2, SEMA5A, STOML1,*
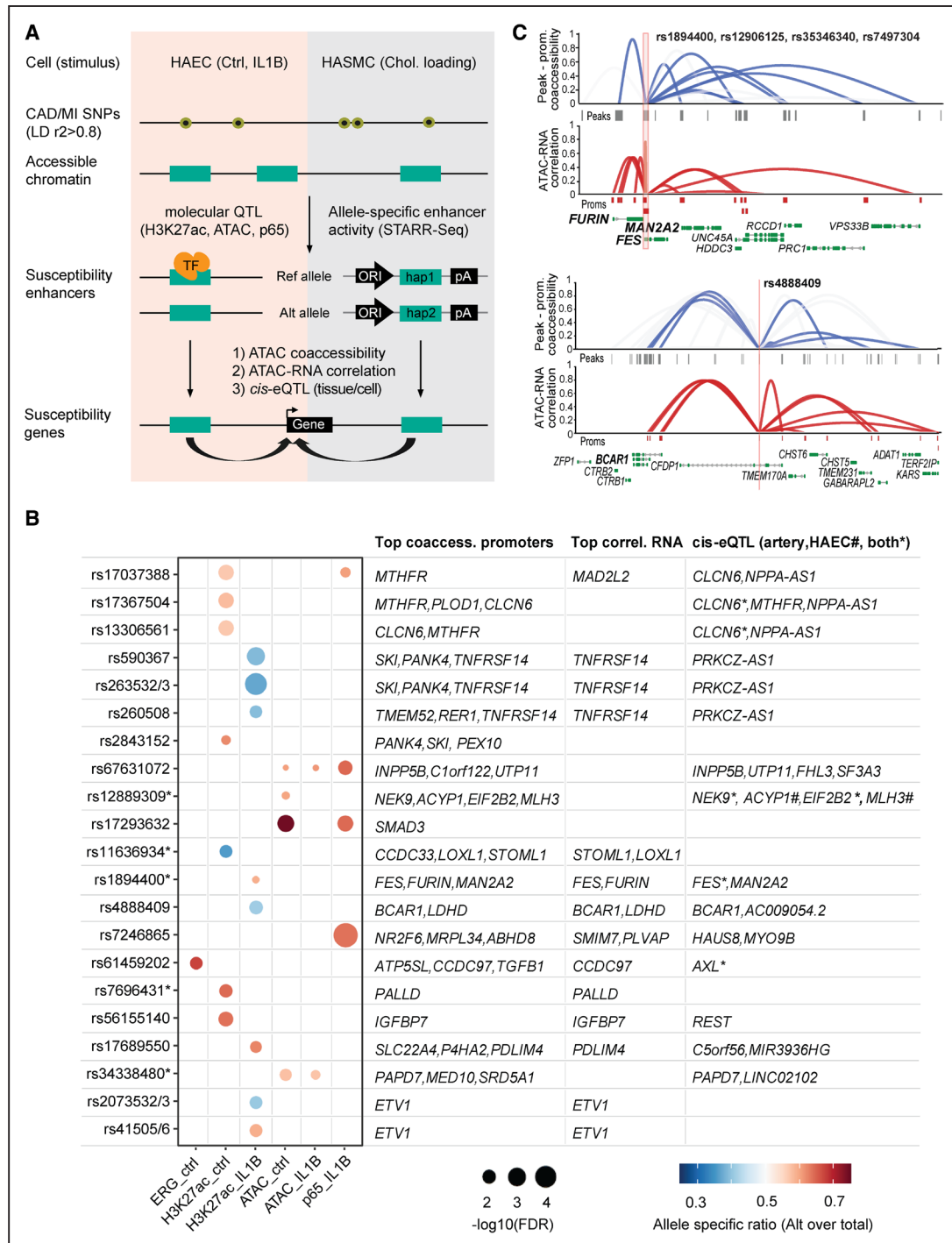
Figure 5. Identification of the coronary artery disease (CAD)/myocardial infarction (MI) functional variants using information from molecular quantitative trait loci (molQTLs).

A, Schematic representing the experimental fine-mapping of the causal variants in human aortic endothelial cells (HAECs) and human aortic smooth muscle cells (HASMCs). **B**, Dot plot summarizing the significant molQTLs that overlap CAD/MI single-nucleotide polymorphism (SNP)-containing single-nucleus (sn)ATAC-Seq peaks and where both the peak and the promoter are accessible in ECs. The epigenetic assays were carried out using bulk epigenomics methods on a panel of human aortic EC cultures derived from genetically diverse donors. Dot size corresponds to statistical significance (false discovery rate [FDR]<0.05, determined by RASQUAL,[33] n=21−44[34]) and dot color corresponds to the allele-specific ratio (blue, more reference allele reads; red, more alternative allele reads). The most potential predicted target genes, which demonstrate peak-promoter coaccessibility score >0.5, peak accessibility–gene expression correlation > 0.5, or cis−expression quantitative trait loci (eQTL) association in Genotype-Tissue Expression (GTEX) v8 arterial tissues (Q value≤0.05) or HAECs (FDR<0.05),[34] are shown. *Presence of other single-nucleotide polymorphisms (SNPs) with identical effect in the same snATAC-Seq peak. For the complete gene list, see Table XIV in the Data Supplement. **C**, Viewpoint plots centered at the molQTL SNPs for *FES* and *BCAR1* loci. Arcs depict peak-promoter coaccessibility (blue arcs) and the correlation between peak accessibility and gene expression (red arcs). The most highly supported target genes are highlighted in bold. Ctrl indicates control conditions; ERG, ETS transcription factor ERG; H3K27ac, histone 3 lysine 27 acetylation; hap, haplotype; IL, interleukin; LD, linkage disequilibrium; ORI, origin of replication with core promoter functionality; pA, poly(A) site; and TF, transcription factor.
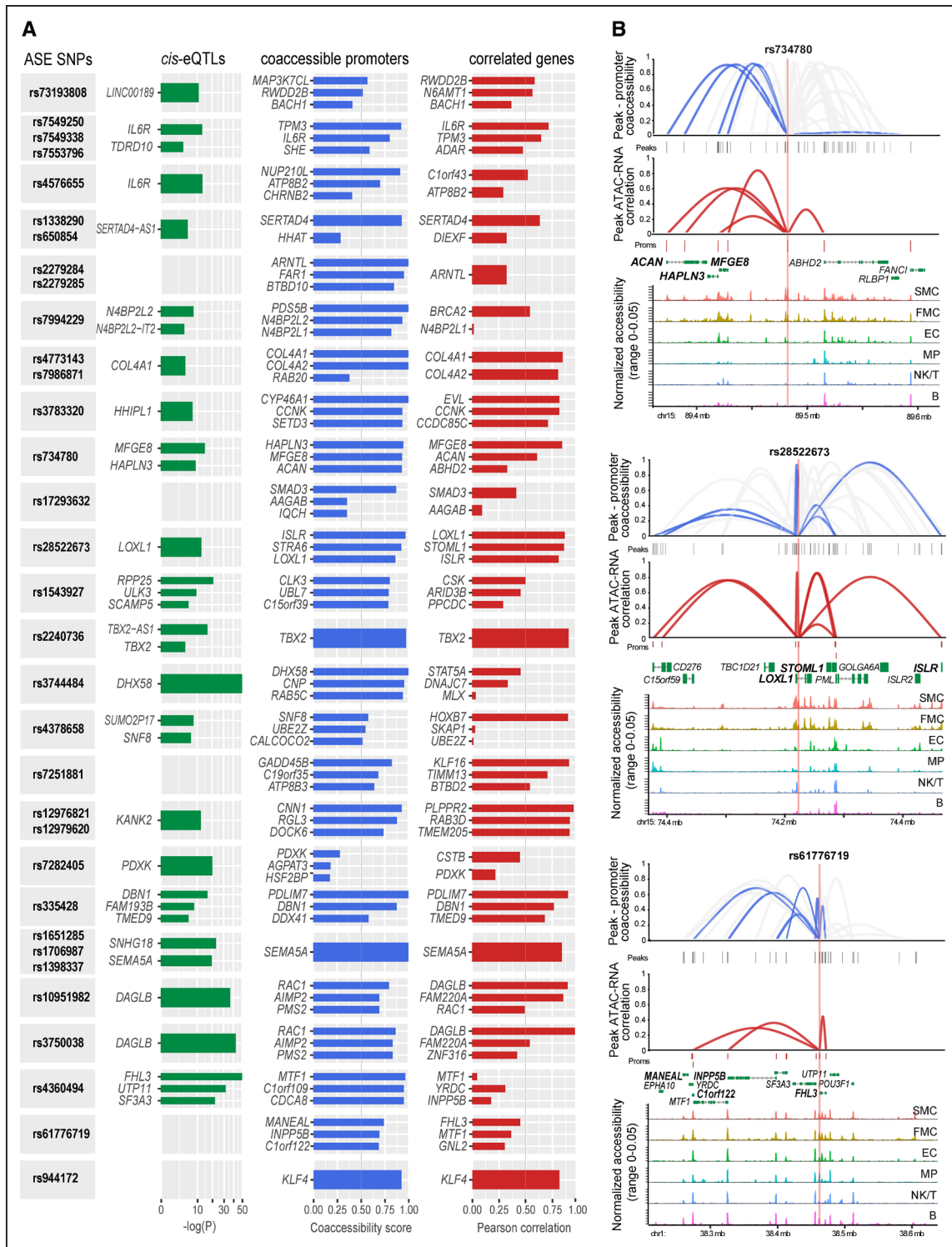
Figure 6. Identification of the coronary artery disease (CAD)/myocardial infarction (MI) functional variants using self-transcribing active regulatory region sequencing (STARR-Seq) in human aortic smooth muscle cells (HASMCs).
A, Bar plot summarizing the CAD/MI single-nucleotide polymorphisms (SNPs) that demonstrated significant allele-specific enhancer (ASE) activity in STARR-Seq performed in cholesterol-loaded primary HASMCs (false discovery rate [FDR] < 0.05 determined by mpralm,[38] n=3). Whenever available, the top 3 target genes predicted by peak-promoter coaccessibility, peak accessibility–gene expression correlation, and cis–expression quantitative trait loci (eQTL) association in Genotype-Tissue Expression (GTEX) v8 arterial tissues (Q value ≤0.05) are shown. Peak-promoter pairs with a peak-promoter coaccessibility score >0.5 and peak accessibility–gene expression correlation >0.5 were considered connected. rs6496126 was not shown as no target genes were predicted. For the complete list, see Table XV in the Data Supplement. **B**, Viewpoint plot of peak-promoter coaccessibility, peak accessibility–gene expression correlation, and the single-nucleus ATAC-Seq tracks centered at the STARR-Seq significant SNPs rs734780, rs28522673, and rs61776719. The predicted target genes are highlighted in bold. B indicates B cell; EC, endothelial cells; FMC, fibromyocyte; MP, macrophage; NK, natural killer; and SMC, smooth muscle cell; and T, T cell.
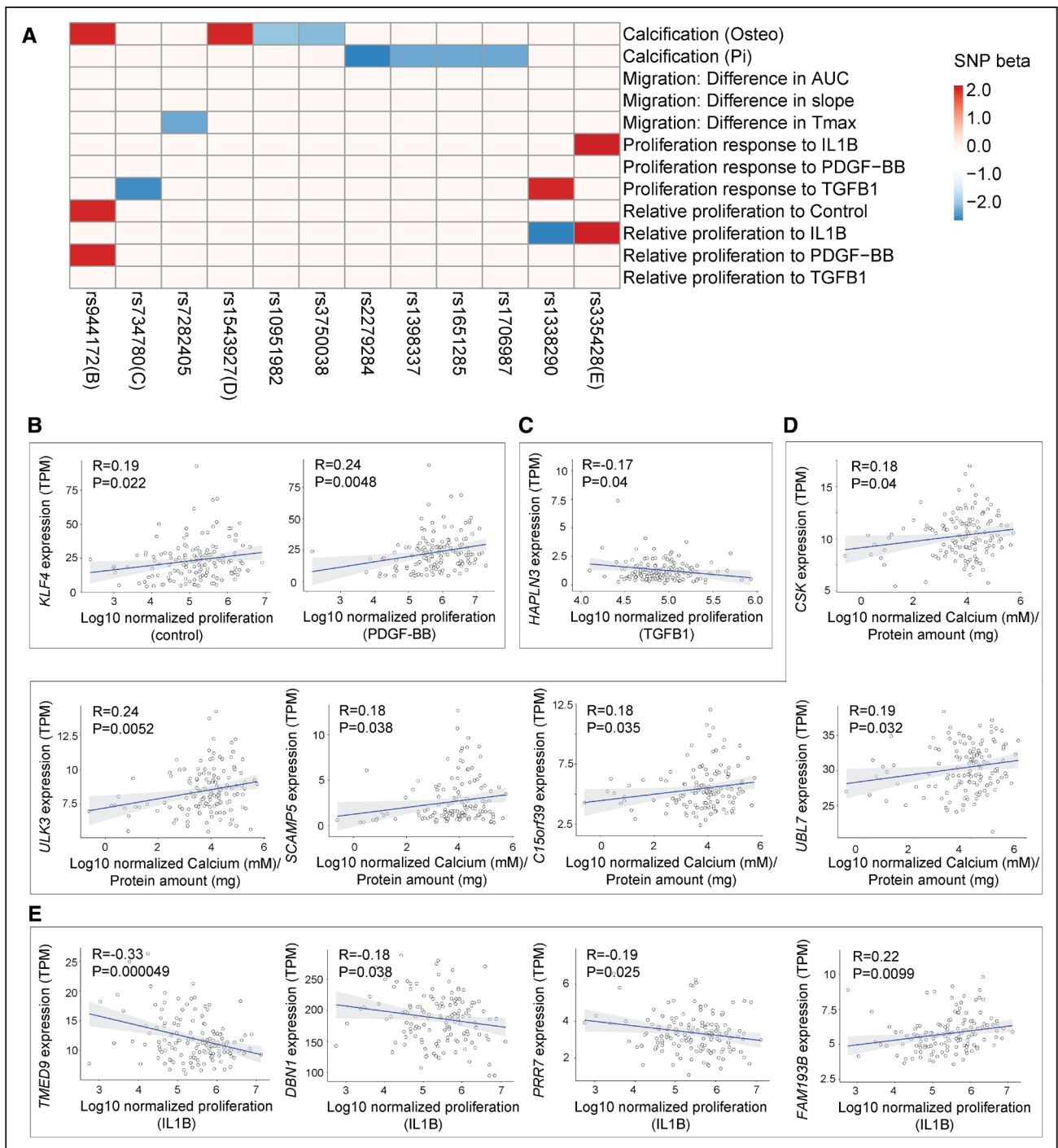
Figure 7. Association of the coronary artery disease (CAD)/myocardial infarction (MI) variants with smooth muscle cell (SMC) phenotypes.

A, Heatmap showing the effect sizes for significant associations between 12/34 CAD/MI variants and 12 SMCs phenotypes determined from 151 human aortic smooth muscle cell (HASMC) donors using FaST-LMM (Factored Spectrally Transformed Linear Mixed Models).[40] 12 of the 34 CAD/MI loci identified in genome-wide association study (GWAS) showed a nominal association ($P<0.05$) with at least one SMC phenotype. Rows show 12 SMC phenotypes, and columns show the index variants in the CAD/MI loci. The color key of the correlations is shown on the **left**. The colors refer to single-nucleotide polymorphism (SNP) weight (β) direction and magnitude, ranging from −2.5 (blue) to 2 (red). Significant associations ($P<0.05$) are indicated with a colored box. Negative effect sizes (blue) indicate that CAD/MI risk allele was associated with a lower SMC phenotype. In contrast, positive effect sizes (red) indicate that CAD/MI risk allele was associated with a higher SMC phenotype. Whenever the expression of a predicted target gene also correlates with the trait, the plot (**B–E**) is listed after the rsID. **B**, Correlation of *KLF4* (rs944172 locus) expression with SMC proliferation relative to control and to PDGF-BB (platelet-derived growth factor BB) stimulus. **C**, Correlation between *HAPLN3* (rs734780 locus) expression and the proliferation response to TGF (transforming growth factor) β1 (**D**) Correlation between *CSK*, *ULK3*, *SCAMP5*, *C15orf39* and *UBL7* (rs1543927 locus) and calcification under the osteogenic stimulus. **E**, Correlation between *TMED9*, *DBN1*, *PRR7,* and *FAM193B* (rs335428 locus) and proliferation response to IL (interleukin) 1β. AUC indicates area under the receiver operator characteristic curve; and TPM, transcripts per million.

*and KANK2 (*Figures 6A and 6B and Figure XXIV in the Data Supplement). As seen for ECs, often the same regulatory SNP was predicted to affect the expression of several genes within one genomic locus, as shown for rs734780 (*MFGE8*, *HAPLN3,* and *ACAN*), rs28522673 (*LOXL1*, *STOML1,* and *ISLR*), and rs61776719 (*FHL3*, *MANEAL*, *INPP5B*, *C1orf122*; Figure 6B and Figures XXV through XXVII in the Data Supplement).

Our analysis was able to pinpoint functional SNPs and their target genes in CAD-relevant cell types. However, to provide evidence of causality, it is important to study the impact of CAD/MI–associated variants on phenotypes relevant to atherosclerosis. To this end, we analyzed the association of the 34 fine-mapped CAD/MI SNPs from HASMCs with calcification, proliferation, and migration as previously described.[39] Our results demonstrate a nominal association ($P<0.05$) of 11 out of 34 CAD/MI variants with the SMC phenotypes (Figure 7A). Since the variants were shown to associate with the expression of several genes, we further correlated the gene expression with the trait to investigate which of the genes could mediate the association. We observed significant correlations between SMC phenotypes and gene expression for 11 CAD/MI–associated genes (Figure 7B). These were exemplified by significant correlation between proliferation and gene expression for *KLF4* and *TMED9* whereas the expression level of *CSK* correlated with calcification. Our integrative analyses thus identified several CAD/MI–associated genes that could modulate CAD risk by affecting SMC phenotypes that could serve as basis for future research exploring the molecular mechanisms by which the CAD/MI GWAS variants and associated target genes affect the disease development.

DISCUSSION

Here we investigated the chromatin accessibility in cell types of human atherosclerotic lesions and provided a resource of 7000 snATAC-seq profiles representing 5 major cell types, ECs, macrophages, B cells, T/natural killer cells, and SMCs. Our findings of cell subtypes are in line with recent scRNA-Seq based studies.[1–4] Importantly, our dataset provides one of the first high-resolution maps of cell type–specific regulatory elements, compared to prior bulk epigenomics-based maps from sorted or in vitro cultured cell populations. Our analysis confirmed the master regulators of lineage determination, such as PU.1 and CEBPA (CCAAT enhancer binding protein alpha) for macrophages[5] as well as TEAD3, MEF2C (myocyte enhancer factor 2C), and SRF for SMCs,[41–43] and also allowed the identification of novel cell subtype–specific TFs. To this end, we link the depletion of NFY and SRF motifs and the genome-wide enrichment of AP-1, NFAT, and STAT1:STAT2 motifs to the phenotypic switching of SMCs to fibromyocytes. Our

genome-wide findings are supported by several recent lines of evidence. First, AP-1 has been shown to bind jointly with TCF21 (transcription factor 21) to modulate H3K27Ac and chromatin state changes at CAD-linked loci.[44] TCF21 has been shown to be expressed in cells that migrate into the developing plaque, promote phenotypic modulation of SMCs to fibromyocytes and contribute to the protective fibrous cap.[1,45] Importantly, TCF21 has also been shown to suppress the expression of SRF and inhibit its association to myocardin and subsequent chromatin binding.[46] Second, locus-specific analyses have associated the binding of NFY, NFAT, and STAT1 to the suppression of SMC contractile genes.[47–49] Importantly, the inhibition of STAT1 has been shown to alleviate inflammation of the arterial wall, reduce plaque burden,[50] and reduce neointima formation after balloon injury,[51] making it a promising target for therapeutic intervention. To our surprise, the genome-wide accessibility of KLF4 motifs was not significantly increased in fibromyocytes, despite the established role of KLF4 in SMC reprogramming.[24] We speculate that this could be due to the KLF4 motif being relatively accessible throughout the SMC lineage and thus the KLF4 target genes not requiring the opening of many new genomic regions but rather activation of preexisting enhancer elements.

Numerous studies have found that cultured cells have a propensity to lose their tissue-specific properties,[52,53] which is expected to translate into changes at the level of cisregulatory elements. Here, we evaluate the extent of such change by comparing the in vivo chromatin accessibility to published studies of in vitro cultured or blood-derived immune cells in basal and proinflammatory conditions. We were surprised to note that the in vitro enhancer profiles were able to capture the majority of the cell type–specific chromatin accessibility regions, reaching on average over 80% overlap for regular enhancers and 60% for superenhancers. This lends supports to the use of primary cell cultures as surrogates for in vivo epigenetic studies in cases where tissue-extracted cells cannot be directly studied. Whether this holds true also for RNA, remains to be studied. Importantly, we identified several developmental TF motifs to be enriched in the thousands of in vivo–unique regulatory elements, including MEF2, HOX, SOX, FOX, POU domain, NFATC2, and NR2F2 (COUP-TFII) motifs. This could be explained by the different tissue origins, absence of blood flow, or reprogramming of epigenetic and transcriptional states of the in vitro cell cultures.[53–55] Surprisingly, only a few motifs were enriched in a cell type–specific manner, exemplified by AP-1 in tissue macrophages. In support of this finding, the environment has been shown to affect the enhancer selection in macrophages, as demonstrated by AP-1 motif enrichment occurring more frequently near PU.1-bound motifs in large peritoneal macrophages compared to microglia.[56] Still, even among the 15 members of the AP-1 family, the TFs binding can be divided

into those that play a role in the opening of the chromatin and to those TFs that participate in fine-tuning the optimal level of transcriptional response.[57] Future studies looking into TF-occupancy at a single-cell resolution are needed to provide an understanding of the collaborative interactions between the AP-1 family members in shaping the regulatory landscape in different environments.

The heritability of CAD has been estimated as 40%,[8] suggesting strong genetic contributions to disease. However, over 90% of the CAD GWAS variants are located in the noncoding regions of the genome with no known function to date.[9,10] Additionally, the lead SNP identified in GWAS may not be the causal SNP but instead linked genetically with a trait-associated SNP. This is explained by several SNPs being in high linkage disequilibrium within a haploblock, which makes the identification of the causal SNPs challenging. Here, we have taken advantage of the emerging evidence that common disease variants are enriched in the enhancers of disease-relevant cell types[58] to identify the cell type and genes through which the variant could act and to pinpoint the potentially causal variants. We demonstrate that half of the risk loci for CAD/MI loci harbor at least one risk SNP that falls within the open chromatin regions identified in this study. Importantly, ≈30% of the risk loci fell within cell type–specific snATAC-Seq peaks, suggesting cell type–selective mechanisms of action. The majority of the variants were accessible in all or many lesional cell types, indicating that multiple mechanisms acting through different cell types could exist.

There is ample evidence that enhancers and promoters that are located within same 3-dimensional chromatin architecture form domains of coordinated regulation.[59] Motivated by this, we also leveraged the profiles of chromatin coaccessibility across single cells to infer pairs of chromatin accessibility peaks that are likely to be in close physical proximity in 3-dimensional space.[32] To this end, we identified 134 CCANs that harbored risk variants for CAD/MI and exhibited coaccessibility correlation with ≈1700 promoters. This suggests that the cisregulatory networks affected by the CAD/MI SNPs could encompass a larger set of target genes than previously anticipated. In support of this, we and others have recently shown that perturbation of a single enhancer within such chromatin interaction hub can lead to significant changes in the expression of many connected genes.[60,61] In line with this, we provide several examples where coaccessibility with several promoters is further supported by the correlation between peak accessibility and gene expression as well as cis-eQTL association. An interesting example is represented by the rs734780-locus, where the SNP-containing enhancer is strongly predicted to regulate *MFGE8*, *HAPLN3*, and *ACAN* genes and is associated with altered proliferation in response to TGF (transforming growth factor) β1 stimulation. Importantly, 3 of these genes have been linked to important functional changes of SMCs that promote atherogenesis: *ACAN* has been associated with aortic stiffness and chondrogenesis,[62] *MFGE8* to the proinflammatory phenotypic shift of SMCs,[63] and *HAPLN3* to in pericellular ECM formation.[64] Still, we recognize that the statistical association between a peak/SNP and a gene/promoter activity does not necessarily mean that the regulatory element is necessary or sufficient to exert a functional effect and further experimental validations are needed. To this end, a recent study confirmed our prediction of rs12906125 regulating the expression of both *FES* and *FURIN* while showing that the variant affects endothelial phenotypes relevant to atherosclerosis such as presentation of E-selectin at the cell membrane.[65]

Finally, our data could help to understand the genetic complexity of CAD by providing experimental fine-mapping for a total of 65 of the CAD/MI−associated variants corresponding to 33 genomic risk loci. These include the previously identified candidate causal variants rs17293632 (*SMAD3*) and rs7549250 (*IL6R*) that were validated in enhancer trap assays in HCASMCs[66] and further validated in our STARR-Seq analysis in HASMCs. Our molecular quantitative trait loci analysis also demonstrated that rs17293632 (*SMAD3*) exhibits allele-specific chromatin accessibility and p65/NF-κB (nuclear factor-κB) binding in HAECs. Still, no other loci revealed similar concordant effects in both cell types, supporting high cell type or context specificity of the variant action. However, further studies comparing the results from the same technology (eg, massively parallel reporter assay) and conditions are needed to delineate the extent of this phenomenon in CAD loci. Future work should further extend to functional characterization of the identified target genes to understand their cell type−dependent and cell type−independent implications in disease. Our analysis provides further support to the genes associated with vascular remodeling (*COL4A1/2*, *FURIN*) and transcriptional regulation (*FHL3*, *KLF4*, *ARNTL*) but also highlights new candidates with potentially similar roles (*TBX2*, *LOXL1*, *SNHG18*) that have not been previously associated with CAD.[8] For example, TBX2 (T-box transcription factor 2) has been extensively studied in the development of myocardium but emerging evidence also supports its importance in differentiation and functionality of smooth muscle cells[67,68] that warrant further research in the context of atherosclerosis. We also identified several novel target genes with no reported biological function but rather additional associations that could provide mechanistic insight. To this end, we identify a potentially causal variant regulating the Scaffold Protein, Cas family member *BCAR1* gene (rs4888409) in a locus that has been further associated with carotid intima-media thickness as well as chloride voltage-gated channel 6 *CLCN6* gene and pyridoxal kinase *PDXK* (rs7282405)[69] loci that have both been associated with blood pressure. How the effects of the common genetic variants affect the

endothelial and smooth muscle cell function and explain the phenotype associations remains a central question to be answered.

Our study is limited by the number of replicates (n=3 males) and the fact that the samples originate from endarterectomy operations of diseased individuals. The inclusion of more patient samples with detailed clinical characteristics as well as healthy controls with complete aortic sections is expected to allow more robust subtype distinction and to also allow sex to be investigated as a biological variable. Furthermore, multiomics methods that can measure both transcriptome and open chromatin status in a single cell are expected to further improve the analysis of transcriptome networks in the future. Additionally, our analysis focused on the most prominent enhancers and might have missed weaker enhancers or enhancers that are only detected upon specific environmental cues. Still, we think our analysis provides a useful resource to the researchers looking to understand the cell type–specific gene-regulatory mechanisms in atherosclerosis-associated cell types. Our findings not only produce maps of regulatory elements but also predicts the cisregulatory interactions that can provide mechanistic understanding of the effect of noncoding genetic variants associated with CAD.

ARTICLE INFORMATION

Affiliations

A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio (T.Ö., K.Õ., V.N., A.T., I.S., E.A., S.Y.-H., M.U.K.). Department of Cellular and Molecular Medicine, The College of Medicine, The University of Arizona, Tucson, AZ (L.K.S., C.E.R.). Center for Public Health Genomics (R.A., M.C.) and Department of Biomedical Engineering (M.C.), University of Virginia, Charlottesville. Turku Bioscience Centre, University of Turku and Åbo Akademi University, Finland (T.L.).

Acknowledgments

Sources of Funding

Disclosures

None.

Supplemental Materials

Expanded Materials and Methods
Data Supplement Figures I–XXV
Data Supplement Tables I–XV
References 33,38,40,69–94

REFERENCES

1. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, Kundu R, Nagao M, Coller J, Koyano TK, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med*. 2019;25:1280–1289. doi: 10.1038/s41591-019-0512-5

2. Fernandez DM, Rahman AH, Fernandez NF, Chudnovskiy A, Amir ED, Amadori L, Khan NS, Wong CK, Shamailova R, Hill CA, et al. Single-cell immune landscape of human atherosclerotic plaques. *Nat Med*. 2019;25:1576–1588. doi: 10.1038/s41591-019-0590-4

3. Depuydt MAC, Prange KHM, Slenders L, Örd T, Elbersen D, Boltjes A, de Jager SCA, Asselbergs FW, de Borst GJ, Aavik E, et al. Microanatomy of the human atherosclerotic plaque by single-cell transcriptomics. *Circ Res*. 2020;127:1437–1455. doi: 10.1161/CIRCRESAHA.120.316770

4. Pan H, Xue C, Auerbach BJ, Fan J, Bashore AC, Cui J, Yang DY, Trignano SB, Liu W, Shi J, et al. Single-cell genomics reveals a novel cell state during smooth muscle cell phenotypic switching and potential therapeutic targets for atherosclerosis in mouse and human. *Circulation*. 2020;142:2060–2075.

5. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–589. doi: 10.1016/j.molcel.2010.05.004

6. Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filippova GN, Huang X, Christiansen L, DeWitt WS, et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell*. 2018;174:1309.e18–1324.e18. doi: 10.1016/j.cell.2018.06.052

7. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res*. 2018;122:433–443. doi: 10.1161/CIRCRESAHA.117.312086

8. Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res*. 2018;114:1241–1257. doi: 10.1093/cvr/cvy084

9. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009;106:9362–9367. doi: 10.1073/pnas.0903103106

10. Won HH, Natarajan P, Dobbyn A, Jordan DM, Roussos P, Lage K, Raychaudhuri S, Stahl E, Do R. Disproportionate contributions of select genomic compartments and cell types to genetic risk for coronary artery disease. *PLoS Genet*. 2015;11:e1005622. doi: 10.1371/journal.pgen.1005622

11. Heinz S, Romanoski CE, Benner C, Allison KA, Kaikkonen MU, Orozco LD, Glass CK. Effect of natural genetic variation on enhancer selection and function. *Nature*. 2013;503:487–492. doi: 10.1038/nature12615

12. Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, Weingarth DT, Zhang B, Greenawalt D, Dobrin R, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet*. 2010;6:e1000932. doi: 10.1371/journal.pgen.1000932

13. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010;466:714–719. doi: 10.1038/nature09266

14. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74. doi: 10.1038/nature11247

15. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–330. doi: 10.1038/nature14248

16. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507:455–461. doi: 10.1038/nature12787

17. Hogan NT, Whalen MB, Stolze LK, Hadeli NK, Lam MT, Springstead JR, Glass CK, Romanoski CE. Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells. *Elife*. 2017;6:e22536.

18. Liu B, Pjanic M, Wang T, Nguyen T, Gloudemans M, Rao A, Castano VG, Nurnberg S, Rader DJ, Elwyn S, et al. Genetic regulatory mechanisms of smooth muscle cells map to coronary artery disease risk loci. *Am J Hum Genet*. 2018;103:377–388. doi: 10.1016/j.ajhg.2018.08.001

19. Kuznetsova T, Prange KHM, Glass CK, de Winther MPJ. Transcriptional and epigenetic regulation of macrophages in atherosclerosis. *Nat Rev Cardiol*. 2020;17:216–228. doi: 10.1038/s41569-019-0265-3

20. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934–947. doi: 10.1016/j.cell.2013.09.053

21. Willemsen L, de Winther MP. Macrophage subsets in atherosclerosis as defined by single-cell technologies. *J Pathol*. 2020;250:705–714. doi: 10.1002/path.5392

22. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. 2018;19:477. doi: 10.1186/s12864-018-4772-0

23. Dobnikar L, Taylor AL, Chappell J, Oldach P, Harman JL, Oerton E, Dzierzak E, Bennett MR, Spivakov M, Jørgensen HF. Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nat Commun*. 2018;9:4567. doi: 10.1038/s41467-018-06891-x

24. Alencar GF, Owsiany KM, Karnewar S, Sukhavasi K, Mocci G, Nguyen AT, Williams CM, Shamsuzzaman S, Mokry M, Henderson CA, et al. Stem cell pluripotency genes Klf4 and Oct4 regulate complex SMC phenotypic changes critical in late-stage atherosclerotic lesion pathogenesis. *Circulation*. 2020;142:2045–2059. doi: 10.1161/CIRCULATIONAHA.120.046672

25. Kim JB, Zhao Q, Nguyen T, Pjanic M, Cheng P, Wirka R, Travisano S, Nagao M, Kundu R, Quertermous T. Environment-sensing aryl hydrocarbon receptor inhibits the chondrogenic fate of modulated smooth muscle cells in atherosclerotic lesions. *Circulation*. 2020;142:575–590. doi: 10.1161/CIRCULATIONAHA.120.045981

26. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017;14:979–982. doi: 10.1038/nmeth.4402

27. Jiang Y, Qian F, Bai X, Liu Y, Wang Q, Ai B, Han X, Shi S, Zhang J, Li X, et al. SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res*. 2019;47:D235–D243. doi: 10.1093/nar/gky1025

28. Hauberg ME, Zhang W, Giambartolomei C, Franzén O, Morris DL, Vyse TJ, Ruusalepp A, Sklar P, Schadt EE, Björkegren JLM, et al; CommonMind Consortium. Large-scale identification of common trait and disease variants affecting gene expression. *Am J Hum Genet*. 2017;101:157. doi: 10.1016/j.ajhg.2017.06.003

29. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Várnai C, Thiecke MJ, et al; BLUEPRINT Consortium. Lineage-Specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*. 2016;167:1369.e19–1384.e19. doi: 10.1016/j.cell.2016.09.037

30. Lalonde S, Codina-Fauteux VA, de Bellefon SM, Leblanc F, Beaudoin M, Simon MM, Dali R, Kwan T, Lo KS, Pastinen T, et al. Integrative analysis of vascular endothelial cell genomic features identifies AIDA as a coronary artery disease candidate gene. *Genome Biol*. 2019;20:133. doi: 10.1186/s13059-019-1749-5

31. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet*. 2013;14:390–403. doi: 10.1038/nrg3454

32. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell*. 2018;71:858–871.e8. doi: 10.1016/j.molcel.2018.06.044

33. Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat Genet*. 2016;48:206–213. doi: 10.1038/ng.3467

34. Stolze LK, Conklin AC, Whalen MB, López Rodríguez M, Õunap K, Selvarajan I, Toropainen A, Örd T, Li J, Eshghi A, et al. Systems genetics in human endothelial cells identifies non-coding variants modifying enhancers, expression, and complex disease traits. *Am J Hum Genet*. 2020;106:748–763. doi: 10.1016/j.ajhg.2020.04.008

35. Muerdter F, Boryń ŁM, Woodfin AR, Neumayr C, Rath M, Zabidi MA, Pagani M, Haberle V, Kazmar T, Catarino RR, et al. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods*. 2018;15:141–149. doi: 10.1038/nmeth.4534

36. Chattopadhyay A, Kwartler CS, Kaw K, Li Y, Kaw A, Chen J, LeMaire SA, Shen YH, Milewicz DM. Cholesterol-induced phenotypic modulation of smooth muscle cells to macrophage/fibroblast-like cells is driven by an unfolded protein response. *Arterioscler Thromb Vasc Biol*. 2021;41:302–316. doi: 10.1161/ATVBAHA.120.315164

37. Rong JX, Shapiro M, Trogan E, Fisher EA. Transdifferentiation of mouse aortic smooth muscle cells to a macrophage-like state after cholesterol loading. *Proc Natl Acad Sci USA*. 2003;100:13531–13536. doi: 10.1073/pnas.1735526100

38. Myint L, Avramopoulos DG, Goff LA, Hansen KD. Linear models enable powerful differential activity analysis in massively parallel reporter assays. *BMC Genomics*. 2019;20:209. doi: 10.1186/s12864-019-5556-x

39. Aherrahrou R, Guo L, Nagraj VP, Aguhob A, Hinkle J, Chen L, Yuhl Soh J, Lue D, Alencar GF, Boltjes A, et al. Genetic regulation of atherosclerosis-relevant phenotypes in human vascular smooth muscle cells. *Circ Res*. 2020;127:1552–1565. doi: 10.1161/CIRCRESAHA.120.317415

40. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods*. 2011;8:833–835. doi: 10.1038/nmeth.1681

41. Kaplan-Albuquerque N, Van Putten V, Weiser-Evans MC, Nemenoff RA. Depletion of serum response factor by RNA interference mimics the mitogenic effects of platelet derived growth factor-BB in vascular smooth muscle cells. *Circ Res*. 2005;97:427–433. doi: 10.1161/01.RES.0000179776.40216.a9

42. Wang Z, Wang DZ, Pipes GC, Olson EN. Myocardin is a master regulator of smooth muscle gene expression. *Proc Natl Acad Sci USA*. 2003;100:7129–7134. doi: 10.1073/pnas.1232341100

43. Firulli AB, Miano JM, Bi W, Johnson AD, Casscells W, Olson EN, Schwarz JJ. Myocyte enhancer binding factor-2 expression and activity in vascular smooth muscle cells. Association with the activated phenotype. *Circ Res*. 1996;78:196–204. doi: 10.1161/01.res.78.2.196

44. Zhao Q, Wirka R, Nguyen T, Nagao M, Cheng P, Miller CL, Kim JB, Pjanic M, Quertermous T. TCF21 and AP-1 interact through epigenetic modifications to regulate coronary artery disease gene expression. *Genome Med*. 2019;11:23. doi: 10.1186/s13073-019-0635-9

45. Nurnberg ST, Cheng K, Raiesdana A, Kundu R, Miller CL, Kim JB, Arora K, Carcamo-Oribe I, Xiong Y, Tellakula N, et al. Coronary artery disease associated transcription factor TCF21 regulates smooth muscle precursor cells that contribute to the fibrous cap. *PLoS Genet*. 2015;11:e1005155. doi: 10.1371/journal.pgen.1005155

46. Nagao M, Lyu Q, Zhao Q, Wirka RC, Bagga J, Nguyen T, Cheng P, Kim JB, Pjanic M, Miano JM, et al. Coronary disease-associated gene TCF21 inhibits smooth muscle cell differentiation by blocking the myocardin-serum response factor pathway. *Circ Res*. 2020;126:517–529. doi: 10.1161/CIRCRESAHA.119.315968

47. Halterman JA, Kwon HM, Zargham R, Bortz PD, Wamhoff BR. Nuclear factor of activated T cells 5 regulates vascular smooth muscle cell phenotypic modulation. *Arterioscler Thromb Vasc Biol*. 2011;31:2287–2296. doi: 10.1161/ATVBAHA.111.232165

48. Kirchmer MN, Franco A, Albasanz-Puig A, Murray J, Yagi M, Gao L, Dong ZM, Wijelath ES. Modulation of vascular smooth muscle cell phenotype by STAT-1 and STAT-3. *Atherosclerosis*. 2014;234:169–175. doi: 10.1016/j.atherosclerosis.2014.02.029

49. Silvestre-Roig C, Fernández P, Esteban V, Pello ÓM, Indolfi C, Rodríguez C, Rodríguez-Calvo R, López-Maderuelo MD, Bauriedel G, Hutter R, et al. Inactivation of nuclear factor-Y inhibits vascular smooth muscle cell proliferation and neointima formation. *Arterioscler Thromb Vasc Biol*. 2013;33:1036–1045. doi: 10.1161/ATVBAHA.112.300580

50. He C, Medley SC, Hu T, Hinsdale ME, Lupu F, Virmani R, Olson LE. PDGFRβ signalling regulates local inflammation and synergizes with hypercholesterolaemia to promote atherosclerosis. *Nat Commun*. 2015;6:7770. doi: 10.1038/ncomms8770

51. Torella D, Curcio A, Gasparri C, Galuppo V, De Serio D, Surace FC, Cavaliere AL, Leone A, Coppola C, Ellison GM, et al. Fludarabine prevents smooth muscle proliferation in vitro and neointimal hyperplasia in vivo through specific inhibition of STAT-1 activation. *Am J Physiol Heart Circ Physiol*. 2007;292:H2935–H2943. doi: 10.1152/ajpheart.00887.2006

52. Chamley-Campbell J, Campbell GR, Ross R. The smooth muscle cell in culture. *Physiol Rev*. 1979;59:1–61. doi: 10.1152/physrev.1979.59.1.1

53. Sabbagh MF, Nathans J. A genome-wide view of the of the-differentiation of central nervous system endothelial cells in culture. *Elife*. 2020;9:e51276.

54. Nestor CE, Ottaviano R, Reinhardt D, Cruickshanks HA, Mjoseng HK, McPherson RC, Lentini A, Thomson JP, Dunican DS, Pennings S, et al. Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. *Genome Biol*. 2015;16:11. doi: 10.1186/s13059-014-0576-y

55. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, Jung S, Amit I. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell*. 2014;159:1312–1326. doi: 10.1016/j.cell.2014.11.018

56. Gosselin D, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, Stender JD, Chun HB, Garner H, Geissmann F, et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell*. 2014;159:1327–1340. doi: 10.1016/j.cell.2014.11.023

57. Fonseca GJ, Tao J, Westin EM, Duttke SH, Spann NJ, Strid T, Shen Z, Stender JD, Sakai M, Link VM, et al. Diverse motif ensembles specify non-redundant DNA binding activities of AP-1 family members in macrophages. *Nat Commun*. 2019;10:414. doi: 10.1038/s41467-018-08236-0

58. Boix CA, James BT, Park YP, Meuleman W, Kellis M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature*. 2021;590:300–307. doi: 10.1038/s41586-020-03145-z

59. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376–380. doi: 10.1038/nature11082

60. Miguel-Escalada I, Bonàs-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, Javierre BM, Rolando DMY, Farabella I, Morgan CC, et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat Genet*. 2019;51:1137–1148. doi: 10.1038/s41588-019-0457-0

61. Selvarajan I, Toropainen A, Garske KM, López Rodríguez M, Ko A, Miao Z, Kaminska D, Õunap K, Örd T, Ravindran A, et al. Integrative analysis of liver-specific non-coding regulatory SNPs associated with the risk of coronary artery disease. *Am J Hum Genet*. 2021;108:411–430. doi: 10.1016/j.ajhg.2021.02.006

62. Yasmin, Maskari RA, McEniery CM, Cleary SE, Li Y, Siew K, Figg NL, Khir AW, Cockcroft JR, Wilkinson IB, et al. The matrix proteins aggrecan and fibulin-1 play a key role in determining aortic stiffness. *Sci Rep*. 2018;8:8550.

63. Chiang HY, Chu PH, Lee TH. MFG-E8 mediates arterial aging by promoting the proinflammatory phenotype of vascular smooth muscle cells. *J Biomed Sci*. 2019;26:61. doi: 10.1186/s12929-019-0559-0

64. Ogawa H, Oohashi T, Sata M, Bekku Y, Hirohata S, Nakamura K, Yonezawa T, Kusachi S, Shiratori Y, Ninomiya Y. Lp3/Hapln3, a novel link protein that co-localizes with versican and is coordinately up-regulated by platelet-derived growth factor in arterial smooth muscle cells. *Matrix Biol*. 2004;23:287–298. doi: 10.1016/j.matbio.2004.07.001

65. Wünnemann F, Tadjo TF, Beaudoin M, Lalonde S, Lo KS, Lettre G. CRISPR perturbations at many coronary artery disease loci impair vascular endothelial cell functions. *bioRxiv*. Preprint posted online February 11, 2021. doi: 10.1101/2021.02.10.430527

66. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Lee JD, Perisic L, Hedin U, Kundu RK, Majmudar D, et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat Commun*. 2016;7:12092. doi: 10.1038/ncomms12092

67. Aydogdu N, Rudat C, Trowe MO, Kaiser M, Ludtke TH, Taketo MM, Christoffels VM, Moon A, Kispert A. TBX2 and TBX3 act downstream of canonical WNT signaling in patterning and differentiation of the mouse ureteric mesenchyme. *Development*. 2018;145:dev171827.

68. Wojahn I, Lüdtke TH, Christoffels VM, Trowe MO, Kispert A. TBX2-positive cells represent a multi-potent mesenchymal progenitor pool in the developing lung. *Respir Res*. 2019;20:292. doi: 10.1186/s12931-019-1264-y

69. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47:D1005–D1012. doi: 10.1093/nar/gky1120

70. Stuart T, Srivastava A, Lareau C, Satija R. Multimodal single-cell chromatin analysis with Signac. *bioRxiv*. Preprint posted online November 10, 2020. doi: 10.1101/2020.11.09.373613

71. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177:1888.e21–1902.e21. doi: 10.1016/j.cell.2019.05.031

72. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods*. 2017;14:975–978. doi: 10.1038/nmeth.4401

73. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res*. 2018;46:D260–D266. doi: 10.1093/nar/gkx1126

74. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28:495–501. doi: 10.1038/nbt.1630

75. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi: 10.1038/nature15393

76. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795

77. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform*. 2013;14:144–161. doi: 10.1093/bib/bbs038

78. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44:955–959. doi: 10.1038/ng.2354

79. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods*. 2015;12:1061–1063.

80. Navab M, Imes SS, Hama SY, Hough GP, Ross LA, Bork RW, Valente AJ, Berliner JA, Drinkwater DC, Laks H. Monocyte transmigration induced by modification of low density lipoprotein in cocultures of human aortic wall cells is due to induction of monocyte chemotactic protein 1 synthesis and is abolished by high density lipoprotein. *J Clin Invest*. 1991;88:2039–2046. doi: 10.1172/JCI115532

81. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol*. 2020;38:276–278. doi: 10.1038/s41587-020-0439-x

82. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics*. 2012;28:1530–1532. doi: 10.1093/bioinformatics/bts196

83. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*. 2016;32:1479–1485. doi: 10.1093/bioinformatics/btv722

84. Charif DL, Lobry JR. SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M eds. Springer, Berlin, Heidelberg. *Structural Approaches to Sequence Evolution Biological and Medical Physics, Biomedical Engineering Springer*;2007:207–232.

85. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*. 2017;27:491–499. doi: 10.1101/gr.209601.116

86. Kalita CA, Brown CD, Freiman A, Isherwood J, Wen X, Pique-Regi R, Luca F. High-throughput characterization of genetic effects on DNA-protein binding and gene transcription. *Genome Res*. 2018;28:1701–1708. doi: 10.1101/gr.237534.118

87. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–930. doi: 10.1093/bioinformatics/btt656

88. Aherrahrou R, Aherrahrou Z, Schunkert H, Erdmann J. Coronary artery disease associated gene Phactr1 modulates severity of vascular calcification in vitro. *Biochem Biophys Res Commun*. 2017;491:396–402. doi: 10.1016/j.bbrc.2017.07.090

89. Frismantiene A, Philippova M, Erne P, Resink TJ. Smooth muscle cell-driven vascular diseases and molecular mechanisms of VSMC plasticity. *Cell Signal*. 2018;52:48–64. doi: 10.1016/j.cellsig.2018.08.019

90. Toma I, McCaffrey TA. Transforming growth factor-β and atherosclerosis: interwoven atherogenic and atheroprotective aspects. *Cell Tissue Res*. 2012;347:155–175. doi: 10.1007/s00441-011-1189-3

91. Blank RS, Owens GK. Platelet-derived growth factor regulates actin isoform expression and growth state in cultured rat aortic smooth muscle cells. *J Cell Physiol*. 1990;142:635–642. doi: 10.1002/jcp.1041420325

92. Alexander MR, Moehle CW, Johnson JL, Yang Z, Lee JK, Jackson CL, Owens GK. Genetic inactivation of IL-1 signaling enhances atherosclerotic plaque instability and reduces outward vessel remodeling in advanced atherosclerosis in mice. *J Clin Invest*. 2012;122:70–79. doi: 10.1172/JCI43713

93. Chen CN, Li YS, Yeh YT, Lee PL, Usami S, Chien S, Chiu JJ. Synergistic roles of platelet-derived growth factor-BB and interleukin-1beta in phenotypic modulation of human aortic smooth muscle cells. *Proc Natl Acad Sci USA*. 2006;103:2665–2670. doi: 10.1073/pnas.0510973103

94. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*. 2014;46:100–106. doi: 10.1038/ng.2876